

STATISTIK

Klausuren und Musterlösungen
aus zehn Jahren Grundlagenlehre

Klausuren aus den Vorlesungen Statistik I (deskriptive und explorative Statistik) und Statistik II (induktive Statistik und Wahrscheinlichkeitslehre) im berufsbegleitenden Bachelor-Studiengang Betriebswirtschaftslehre an der Hochschule Harz von 2016 – 2025

Christian Reinboth

Impressum

© Christian Reinboth

Erstveröffentlicht am 28.09.2025 über den OER Campus von <https://studip.hs-harz.de/>.

Sie dürfen:

Teilen — das Material in jedwedem Format oder Medium vervielfältigen und weiterverbreiten und zwar für beliebige Zwecke, sogar kommerziell.

Bearbeiten — das Material remixen, verändern und darauf aufbauen und zwar für beliebige Zwecke, sogar kommerziell.

Unter folgenden Bedingungen:

Namensnennung — Sie müssen angemessene Urheber- und Rechteangaben machen, einen Link zur Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Diese Angaben dürfen in jeder angemessenen Art und Weise gemacht werden, allerdings nicht so, dass der Eindruck entsteht, der Lizenzgeber unterstütze gerade Sie oder Ihre Nutzung besonders.

Keine weiteren Einschränkungen — Sie dürfen keine zusätzlichen Klauseln oder technische Verfahren einsetzen, die anderen rechtlich irgendetwas untersagen, was die Lizenz erlaubt.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.



<https://creativecommons.org/licenses/by/4.0/deed.de>

Einige Hinweise zur Nutzung dieser Aufgabensammlung

Seit dem Sommersemester 2016 habe ich das große Glück, die Einführungsvorlesung Statistik im berufsbegleitenden Bachelor-Studiengang Betriebswirtschaftslehre am Fachbereich Wirtschaftswissenschaften der Hochschule Harz sowie zeitweise auch weitere Statistik-Vorlesungen bei den angehenden Wirtschaftsingenieuren und die Methodenlehre bei den Konsumentenpsychologen unterrichten zu dürfen, was mir all diese Jahre über viel Freude bereitet hat. Im Rahmen unserer Vorlesungen haben wir uns neben Grundbegriffen und -ideen der Statistik und Wahrscheinlichkeitslehre unter anderem mit Mittelwerten, Streuungsmaßen, Korrelationskoeffizienten, der einfachen linearen Regressionsanalyse, statischen Hypothesentests, Konfidenzintervallen, mit dem Satz von Bayes oder auch mit der Bestimmung von angemessenen Stichprobengrößen beschäftigt.

Alle diese Themen finden sich natürlich auch in den Klausuren wieder, wobei in den meisten Semestern jeweils eine Klausur für Statistik I (deskriptive und explorative Statistik) und Statistik II (Inhalte aus der Statistik I sowie induktive Statistik und Wahrscheinlichkeitslehre) zu schreiben waren. In der Vorbereitungsphase auf diese Klausuren gab es seitens der Studierenden immer wieder den Wunsch nach Übungsaufgaben, Probeklausuren oder realen Klausuren aus zurückliegenden Semestern – natürlich mit entsprechenden Musterlösungen. Diesem Wunsch möchte ich nach zehn Jahren mit dieser Sammlung von Klausuren und Musterlösungen zukünftig besser als bislang nachkommen.

Die nachfolgende Sammlung ist naturgemäß über einen sehr langen Zeitraum entstanden. Den geneigten Leserinnen und Lesern möchte ich daher noch folgende Hinweise zur Nutzung mitgeben:

- Angesichts des großen Umfangs dieser Aufgabensammlung habe ich auf eine perfekte Vereinheitlichung der Formatierung (z.B. durch Angleichung aller Tabellen oder die immer gleiche Ausformulierung von Rechenwegen) verzichtet und die Klausuren und Musterlösungen (auch aus zeitlichen Gründen) mehr oder weniger in ihrem „Urzustand“ belassen.
- Die Musterlösungen habe ich (vor allem früher) in erster Linie für mich selbst und nicht für die Studierenden verfasst – auch wenn diese nach den Prüfungen auf Wunsch stets Einblick erhalten haben. Sie fallen daher nicht immer so umfangreich (etwa mit Formeln und zusätzlichen Erläuterungen) aus, wie man es sich für Selbstlernmaterialien wünschen würde, weil sie ursprünglich eben nicht für das Selbststudium, sondern lediglich zur Kontrolle der erarbeiteten Lösungen gedacht waren. Insbesondere sind Rechenwege teils abgekürzt oder verwenden unübliche Symbole wie etwa das * für die Multiplikation (was für zukünftige Versionen dieser Aufgabensammlung noch verbessert wird). Für das Selbststudium besser geeignetere Materialien biete ich unter <https://www.hs-harz.de/creinboth/lehre> zum Download an.
- Da es sich um Klausuren aus zehn Jahren und damit aus fast zwanzig Semestern Lehre handelt, werden einzelne Aufgaben sich ähneln oder sogar wiederholen – was ja aber bei der Vorbereitung nicht unbedingt schadet.
- Fast alle in den (späteren) Klausuren zur visuellen Auflockerung genutzten Cliparts stammen von der Seite Public Domain Vectors (<https://publicdomainvectors.org/>) und sind Copyright-frei nutzbar. Die Urheber anderer Abbildungen werden entsprechend benannt. Die Veröffentlichung dieser Aufgabensammlung unter einer freien Lizenz stellt selbstverständlich keine Übertragung dieser Lizenz auf von Dritten geschaffene Abbildungen dar.

- Ich möchte nicht ausschließen, dass sich in einzelnen Aufgaben noch ein Rechtschreibfehler, in einzelnen Musterlösungen vielleicht sogar auch noch ein Rechenfehler findet. Sehr gerne nehme ich Hinweise auf solche Fehler per E-Mail an creinboth@hs-harz.de entgegen. Da es sich bei dieser Aufgabensammlung um ein „lebendes“ Dokument handelt, kann ich Korrekturen problemlos in zukünftigen Versionen vornehmen und bedanke mich schon im Voraus für jede Rückmeldung, die mir bei der kontinuierlichen Verbesserung meiner Materialien hilft.
- Die Vergabe von 100 Punkten pro Klausur bzw. 50 Punkten pro Teilklausur, wird dem einen oder anderen erfahreneren Lehrenden möglicherweise sauer aufstoßen. Das kann ich zwar sehr gut nachvollziehen, habe mich aber dennoch bewusst für dieses vereinfachende System entschieden, weil ich nicht möchte, dass Studierende in den kurzen (60 und 120 Minuten) Klausuren wertvolle Zeit damit verschwenden, gedanklich nach möglichen „Bestehenspfaden“ zu suchen. Der unbestreitbare Nachteil einer weniger passgenauen Bewertbarkeit von Leistungen wird durch diesen Vorteil in meinen Augen klar ausgeglichen.
- Ja – die Musterlösungen der späteren Klausuren sind besser als die der früheren Prüfungen. Auch die Aufgaben selbst sind meiner Einschätzung nach mit der Zeit ausgereifter geworden und folgen heute meist einem gemeinsamen Thema, um den Prüfungen einen gewissen roten Faden zu verleihen. Wer also nur zwei oder drei Klausuren durchrechnen möchte, sollte mit den Prüfungen ab etwa 2020 beginnen. Der Lerneffekt zeigt sich – wie so oft – eben nicht nur beim Lernenden, sondern insbesondere auch beim Lehrenden.

Wernigerode, den 28. September 2025

Christian Reinboth

Inhalt

Sommersemester 2016 Klausur Statistik I.....	6
Sommersemester 2016 Klausur Statistik II.....	17
Wintersemester 2016/2017 Klausur Statistik I	34
Wintersemester 2016/2017 Klausur Statistik II	44
Sommersemester 2017 Klausur Statistik I.....	59
Sommersemester 2017 Klausur Statistik II.....	69
Sommersemester 2018 Klausur Statistik I.....	86
Sommersemester 2018 Klausur Statistik II.....	96
Wintersemester 2018/2019 Klausur Statistik I	112
Wintersemester 2018/2019 Klausur Statistik II	123
Sommersemester 2019 Klausur Statistik I.....	135
Sommersemester 2019 Klausur Statistik II.....	144
Wintersemester 2019/2020 Klausur Methodenlehre.....	162
Sommersemester 2020 Klausur Statistik I.....	170
Sommersemester 2020 Klausur Statistik II.....	183
Wintersemester 2020/2021 Klausur Statistik	204
Sommersemester 2021 Klausur Statistik I.....	214
Sommersemester 2021 Klausur Statistik II.....	227
Sommersemester 2022 Klausur Statistik I.....	245
Sommersemester 2022 Klausur Statistik II.....	256
Sommersemester 2023 Klausur Statistik I.....	275
Sommersemester 2023 Klausur Statistik II.....	285
Sommersemester 2024 Klausur Statistik I.....	308
Sommersemester 2024 Klausur Statistik II.....	321
Sommersemester 2025 Klausur Statistik I.....	334
Sommersemester 2025 Klausur Statistik II.....	344

Sommersemester 2016

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2016

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 11. Juni 2016

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2016

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Quartilskoeffizient der Schiefe lässt sich lediglich bei bimodalen Verteilungen interpretieren.

☐ richtig ☐ falsch

2) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

☐ richtig ☐ falsch

3) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

☐ richtig ☐ falsch

4) Bei der Klassenbildung gehören die Klassengrenzen jeweils zur Hälfte zu beiden Klassen.

☐ richtig ☐ falsch

5) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig ☐ falsch

6) Die Berechnung der Varianz setzt mindestens metrisch skalierte Daten voraus.

☐ richtig ☐ falsch

7) Bei einem IQR von Null kann der Quartilskoeffizient der Schiefe nicht berechnet werden.

☐ richtig ☐ falsch

8) Die Spannweite wird extrem von Ausreißern beeinflusst und ist daher nicht robust.

☐ richtig ☐ falsch

9) Der Quartilskoeffizient der Schiefe ist unempfindlich gegenüber Ausreißern.

☐ richtig ☐ falsch

10) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Bei einer Befragung von 15 Studierenden wurden auf die Frage nach den monatlichen Wohnkosten folgende Angaben gemacht.

Student Nr.	Monatliche Wohnkosten (in EUR)
1	220
2	250
3	310
4	175
5	230
6	340
7	300
8	240
9	260
10	265
11	185
12	325
13	220
14	250
15	230

Zeichnen Sie einen (erweiterten) Box-Plot für diese Verteilung (10 Punkte) und benennen Sie alle für die Konstruktion benötigten Größen (10 Punkte).

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Bei einer Mathematiklausur ergab sich die folgende Notenverteilung:

Note	Anzahl Schüler/innen
1	3
2	12
3	6
4	3
5	5
6	1

Maße der zentralen Tendenz:

- Bestimmen Sie das arithmetische Mittel. (4 Punkte)
- Bestimmen Sie den Median. (4 Punkte)
- Bestimmen Sie den Modus. (4 Punkte)

Streuungsmaße / Dispersionsparameter:

- Bestimmen Sie die Varianz. (5 Punkte)
- Bestimmen Sie die Spannweite. (4 Punkte)
- Bestimmen Sie den Interquartilsabstand. (4 Punkte)

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie den Quartilkoeffizienten der Schiefe. (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Ein Kinobetreiber verändert (*ceteris paribus*) an 10 Tagen die Eintrittspreise für den Abendfilm und zeichnet die Besucherzahlen auf. Es ergibt sich die folgende Tabelle:

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	836
2	9,00	741
3	9,50	798
4	10,00	740
5	10,50	685
6	11,00	673
7	11,50	710
8	12,00	625
9	12,50	531
10	13,00	330

Berechnen (15 Punkte) und interpretieren (5 Punkte)
Sie den Bravais-Pearson-Korrelationskoeffizienten.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Berechnen Sie für die bereits bekannte Verteilung der Eintrittspreise (unabhängige Variable x) und Besucherzahlen (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte) Diskutieren Sie, ob die Ergebnisse zum Bravais-Pearson-Regressionskoeffizienten aus Aufgabenteil IV passen. (5 Punkte)

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	836
2	9,00	741
3	9,50	798
4	10,00	740
5	10,50	685
6	11,00	673
7	11,50	710
8	12,00	625
9	12,50	531
10	13,00	330

LÖSUNG KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2016

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Quartilskoeffizient der Schiefe lässt sich lediglich bei bimodalen Verteilungen interpretieren.

☐ richtig ☒ falsch

2) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

☒ richtig ☐ falsch

3) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

☐ richtig ☒ falsch

4) Bei der Klassenbildung gehören die Klassengrenzen jeweils zur Hälfte zu beiden Klassen.

☐ richtig ☒ falsch

5) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☒ richtig ☐ falsch

6) Die Berechnung der Varianz setzt mindestens metrisch skalierte Daten voraus.

☒ richtig ☐ falsch

7) Bei einem IQR von Null kann der Quartilskoeffizient der Schiefe nicht berechnet werden.

☒ richtig ☐ falsch

8) Die Spannweite wird extrem von Ausreißern beeinflusst und ist daher nicht robust.

☒ richtig ☐ falsch

9) Der Quartilskoeffizient der Schiefe ist unempfindlich gegenüber Ausreißern.

☒ richtig ☐ falsch

10) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☒ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Bei einer Befragung von 15 Studierenden wurden auf die Frage nach den monatlichen Wohnkosten folgende Angaben gemacht.

Student Nr.	Monatliche Wohnkosten (in EUR)
1	220
2	250
3	310
4	175
5	230
6	340
7	300
8	240
9	260
10	265
11	185
12	325
13	220
14	250
15	230

Zeichnen Sie einen (erweiterten) Box-Plot für diese Verteilung (10 Punkte) und benennen Sie alle für die Konstruktion benötigten Größen (10 Punkte).

Schritt 1: Aufstellung der geordneten Verteilung.

175; 185; 220; 220; 230; 230; 240; 250; 250; 260; 265; 300; 310; 325; 340

Schritt 2: Berechnung der Quartilswerte sowie des IQR.

$(n * p) = (15 * 0,25) = 3,75 \rightarrow$ nicht ganzzahlig $\rightarrow k = 4 \rightarrow p_{0,25} = 220$

$(n * p) = (15 * 0,50) = 7,50 \rightarrow$ nicht ganzzahlig $\rightarrow k = 8 \rightarrow p_{0,50} = 250$

$(n * p) = (15 * 0,75) = 11,25 \rightarrow$ nicht ganzzahlig $\rightarrow k = 12 \rightarrow p_{0,75} = 300$

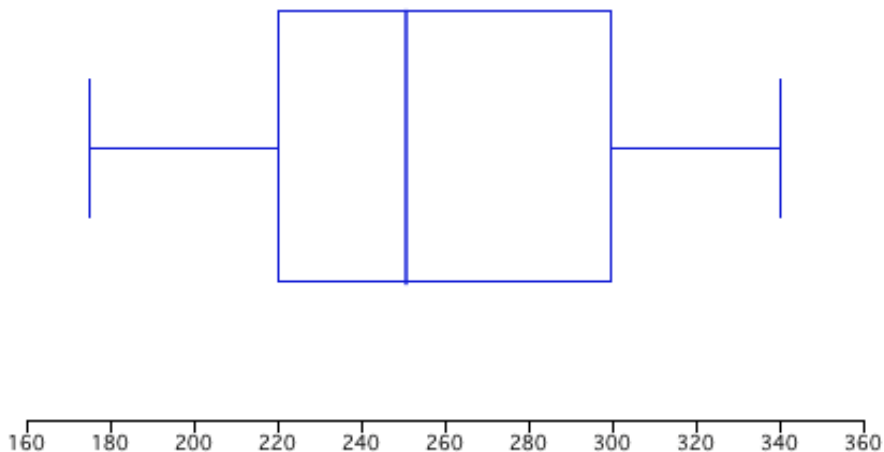
$IQR = 300 - 220 = 80$

$1,5 \text{ IQR} = 120$

$3,0 \text{ IQR} = 240$

Schritt 3: Konstruktion des Box-Plots.

- Die Box verläuft von 220 bis 300, der Median ist bei 250 einzuzeichnen.
- Der obere Zaun könnte maximal bis zur 420 ($300 + 120$) laufen, der größte nicht-extreme Wert in diesem Bereich ist 340 $\rightarrow$ oberer Zaun endet bei 340.
- Der untere Zaun könnte minimal bis zur 100 ($220 - 120$) laufen, der kleinste nicht-extreme Wert in diesem Bereich ist 175 $\rightarrow$ unterer Zaun endet bei 175.
- Da die 175 und die 340 den kleinsten bzw. den größten Wert im Datensatz darstellen, existieren in beide Richtungen keine Ausreißer.



(Grafik erstellt mit SSP: Smith's Statistical Package)

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Bei einer Mathematiklausur ergab sich die folgende Notenverteilung:

Note	Anzahl Schüler/innen
1	3
2	12
3	6
4	3
5	5
6	1

Maße der zentralen Tendenz:

- Bestimmen Sie das arithmetische Mittel. (4 Punkte)
- Bestimmen Sie den Median. (4 Punkte)
- Bestimmen Sie den Modus. (4 Punkte)

Arithmetisches Mittel: Nicht möglich, da es sich um ordinalskalierte Daten handelt.

Median:

Geordnete Verteilung:

1; 1; 1; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 3; 3; 3; 3; 3; 3; 4; 4; 4; 5; 5; 5; 5; 5; 6
 $(n * p) = (30 * 0,5) = 15 \rightarrow$ ganzzahlig $\rightarrow$ Mittel des 15. und 16. Werts = 2,5

Modus:

Der Modus beträgt 2.

Streuungsmaße / Dispersionsparameter:

- Bestimmen Sie die Varianz. (5 Punkte)
- Bestimmen Sie die Spannweite. (4 Punkte)
- Bestimmen Sie den Interquartilsabstand. (4 Punkte)

Varianz: Nicht möglich, da es sich um ordinalskalierte Daten handelt.

Spannweite: Nicht möglich, da es sich um ordinalskalierte Daten handelt.

Interquartilsabstand: Unteres Quartil: $(n * p) = 30 * 0,25 = 7,5 \rightarrow 8$. Wert = 2

Oberes Quartil: $(n * p) = 30 * 0,75 = 22,5 \rightarrow 23$. Wert = 4

$IQR = 4 - 2 = 2$

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie den Quartilkoeffizienten der Schiefe. (5 Punkte)

Quartilskoeffizient: $g_{0,25} = [(4-2,5) - (2,5-2)] / (4-2) = 1/2 = 0,5$

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Ein Kinobetreiber verändert (ceteris paribus) an 10 Tagen die Eintrittspreise für den Abendfilm und zeichnet die Besucherzahlen auf. Es ergibt sich die folgende Tabelle:

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	836
2	9,00	741
3	9,50	798
4	10,00	740
5	10,50	685
6	11,00	673
7	11,50	710
8	12,00	625
9	12,50	531
10	13,00	330

Berechnen (15 Punkte) und interpretieren (5 Punkte)

Sie den Bravais-Pearson-Korrelationskoeffizienten.

x	y	x^2	y^2	$(x*y)$
8,50	836	72,25	698896	7106,00
9,00	741	81,00	549081	6669,00
9,50	798	90,25	636804	7581,00
10,00	740	100,00	547600	7400,00
10,50	685	110,25	469225	7192,50
11,00	673	121,00	452929	7403,00
11,50	710	132,25	504100	8165,00
12,00	625	144,00	390625	7500,00
12,50	531	156,25	281961	6637,50
13,00	330	169,00	108900	4290,00
S: 107,50	S: 6669,00	S: 1176,25	S: 4640121	S: 69944,00
M: 10,75	M: 666,90			

$$r = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sqrt{\sum_{i=1}^n (x_i^2) - n * \bar{x}^2} * \sqrt{\sum_{i=1}^n (y_i^2) - n * \bar{y}^2}}$$

$$r = \frac{69944 - 10 \cdot 10,75 \cdot 666,90}{\sqrt{1176,25 - 10 \cdot 10,75^2} \cdot \sqrt{4640121 - 10 \cdot 6690^2}} = \frac{-1747,75}{1992,90} = -0,8769$$

Es existiert eine starke negative lineare Korrelation.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Berechnen Sie für die bereits bekannte Verteilung der Eintrittspreise (unabhängige Variable x) und Besucherzahlen (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 (5 Punkte). Passen die gefundenen Ergebnisse mit dem Bravais-Pearson-Regressionskoeffizienten aus Aufgabenteil IV zusammen? (5 Punkte)

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	836
2	9,00	741
3	9,50	798
4	10,00	740
5	10,50	685
6	11,00	673
7	11,50	710
8	12,00	625
9	12,50	531
10	13,00	330

x	y	x ²	(x*y)
8,50	836	72,25	7106,00
9,00	741	81,00	6669,00
9,50	798	90,25	7581,00
10,00	740	100,00	7400,00
10,50	685	110,25	7192,50
11,00	673	121,00	7403,00
11,50	710	132,25	8165,00
12,00	625	144,00	7500,00
12,50	531	156,25	6637,50
13,00	330	169,00	4290,00
S: 107,50	S: 6669,00	S: 1176,25	S: 69944,00
M: 10,75	M: 666,90		

$$b = (69944 - 10 \cdot 10,75 \cdot 666,9) / (1176,25 - 10 \cdot 10,75^2) = -84,7394$$

$$a = 666,9 - (-84,7394 \cdot 10,75) = 1577,8485$$

$$y = 1577,85 - 84,74 \cdot x$$

Berechnung des Gütemaßes R^2 :

$$\text{Möglichkeit 1: } R^2 = r^2 = 0,8769^2 = 0,7689 = 0,77$$

Möglichkeit 2: Berechnung über Formel und Hilfstabelle

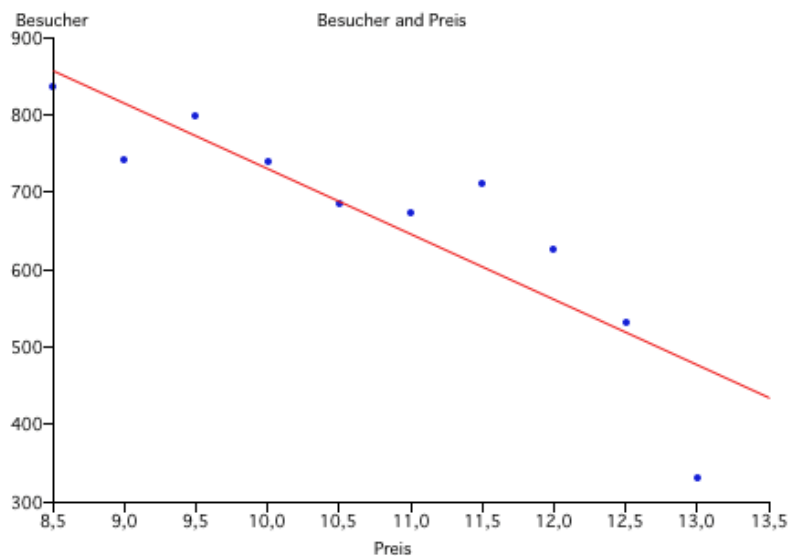
x	y	y*	(y*-y _m) ²	(y - y*) ²
8,50	836	857,56	36351,24	464,83
9,00	741	815,19	21989,92	5504,16
9,50	798	772,82	11219,05	634,03
10,00	740	730,45	4038,60	91,20
10,50	685	688,08	448,59	9,49
11,00	673	645,71	449,02	744,74
11,50	710	603,34	4039,87	11376,36
12,00	625	560,97	11221,16	4099,84
12,50	531	518,60	21992,89	153,76
13,00	330	476,23	36355,05	21383,21
//	S: 6669,00	//	148105,39	44461,62

Das arithmetische Mittel von y liegt bei 666,90.

Die Regressionsformel zur Berechnung von y* lautet $y = 1577,85 - 84,74 \cdot x$.

$TSS = 148105,39 + 44461,62 = 192567,01$

$R^2 = 148105,39 / 192567,01 = 0,7691 = 0,77$



(Grafik erstellt mit SSP: Smith's Statistical Package)

Sommersemester 2016

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Automatisierung und Informatik
bbgl. Bachelorstudiengang Betriebswirtschaftslehre
Sommersemester 2016

Prüfungsfach:

Statistik II

Name des Prüfers: Christian Reinboth

Prüfungstag: 16. Juli 2016

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung (auch eigene), Taschenrechner.
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars. Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK

Berufsbegleitender Studiengang Betriebswirtschaftslehre
Sommersemester 2016
Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

2) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☐ richtig ☐ falsch

3) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☐ falsch

4) Bei der Klassierung von Daten dürfen auch unterschiedlich breite Klassen gebildet werden.

☐ richtig ☐ falsch

5) Variablen mit nur zwei möglichen Ausprägungen werden als dichotom bezeichnet.

☐ richtig ☐ falsch

6) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des Y-Achsenschnittpunkts der linearen Regressionsgeraden.

☐ richtig ☐ falsch

7) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

8) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☐ richtig ☐ falsch

9) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

10) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Eine Umfrage unter 30 Studierenden der Hochschule Harz ergab u.a. die folgende Altersverteilung:

Nr.	Alter	Nr.	Alter	Nr.	Alter
1	21 Jahre	11	33 Jahre	21	19 Jahre
2	19 Jahre	12	26 Jahre	22	22 Jahre
3	23 Jahre	13	81 Jahre	23	29 Jahre
4	22 Jahre	14	19 Jahre	24	34 Jahre
5	21 Jahre	15	18 Jahre	25	22 Jahre
6	24 Jahre	16	22 Jahre	26	24 Jahre
7	26 Jahre	17	26 Jahre	27	20 Jahre
8	22 Jahre	18	30 Jahre	28	21 Jahre
9	31 Jahre	19	20 Jahre	29	25 Jahre
10	27 Jahre	20	24 Jahre	30	32 Jahre

Zeichnen Sie...

- a) ...ein Stamm-Blatt-Diagramm mit einer Stammbreite von 10. (5 Punkte)
- b) ...ein Stamm-Blatt-Diagramm mit einer Stammbreite von 5. (5 Punkte)

Aufgabenteil III: Verteilungsparameter (20 Punkte)

Um die maximal mögliche Wasserentnahmemenge aus der Kalten Bode für eine Beschneiungsanlage zu bestimmen, wird über ein Jahr jeweils am zweiten Dienstag jedes Monats der Wasserstand an einem vorab festgelegten Messpunkt erhoben. Es ergeben sich folgende Werte:

Monat	Wasserstand
Januar	1,17 m
Februar	1,29 m
März	1,23 m
April	1,18 m
Mai	1,25 m
Juni	1,04 m
Juli	1,08 m
August	0,98 m
September	1,12 m
Oktober	1,14 m
November	1,19 m
Dezember	1,21 m

Maße der zentralen Tendenz:

- Bestimmen Sie das arithmetische Mittel. (3 Punkte)
- Bestimmen Sie den Median. (3 Punkte)
- Bestimmen Sie den Modus. (3 Punkte)

Streuungsmaße / Dispersionsparameter:

- Bestimmen Sie die Varianz. (3 Punkte)
- Bestimmen Sie den Interquartilsabstand. (3 Punkte)

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie die Kurtosis / Exzeß. (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (10 Punkte)

Studierenden ist die Teilnahme an acht zusätzlichen Tutoriumsterminen grundsätzlich freigestellt. Im Anschluss an eine Klausur soll überprüft werden, ob zwischen der Teilnahme an diesen Tutorien und der späteren Klausurnote ein monotoner Zusammenhang besteht.

Student Nr.	Absolvierte Tutorien	Klausurnote
1	0	3,4
2	2	5,0
3	3	2,8
4	5	2,0
5	1	2,8
6	4	2,3
7	4	1,7
8	6	1,3
9	8	3,0
10	7	2,5

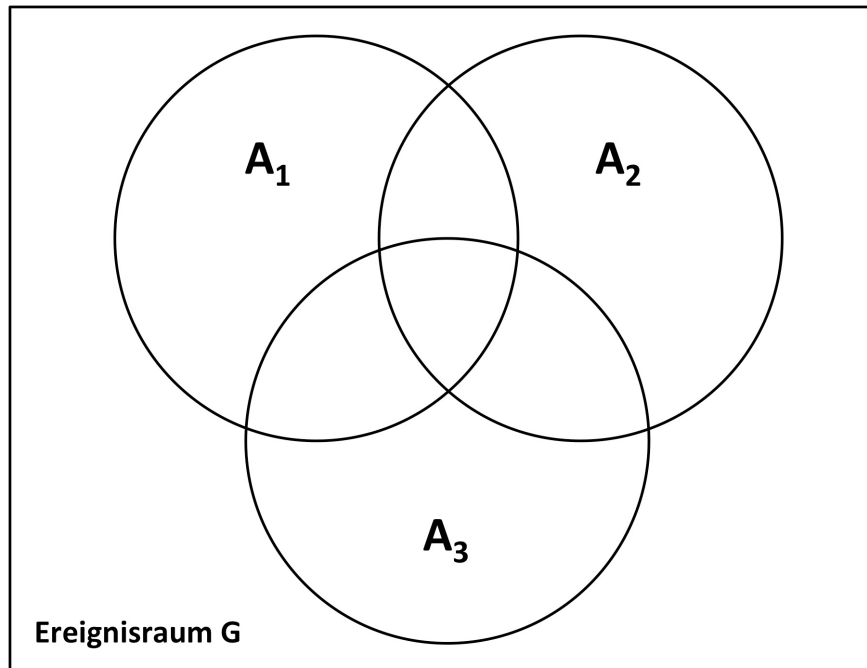
Berechnen und interpretieren Sie...

- a) ...den Konkordanzkoeffizienten nach Kendall. (5 Punkte)
- b) ...den Rangkorrelationskoeffizienten nach Spearman. (5 Punkte)

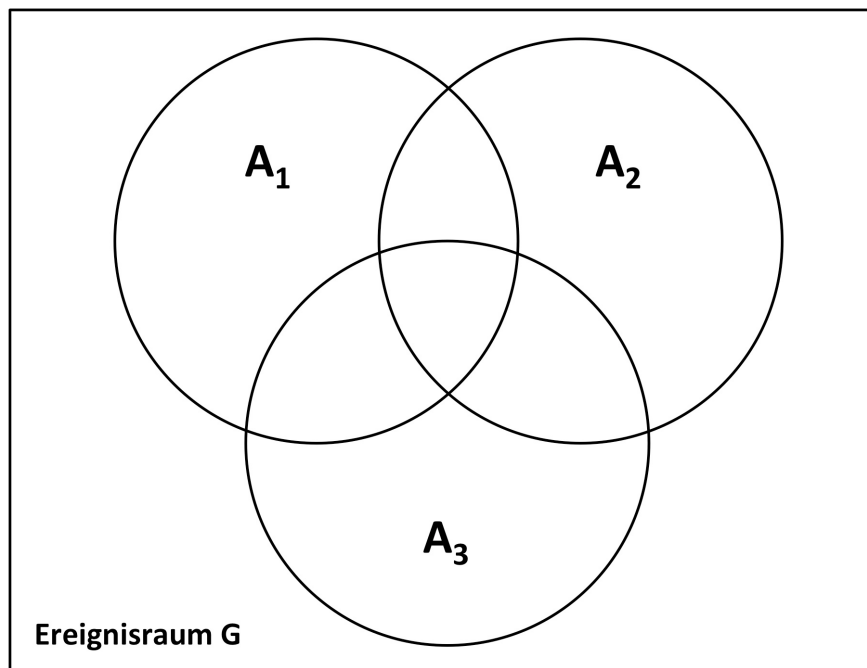
Aufgabenteil V: Interpretation von Venn-Diagrammen (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

$$(A_1 \cup A_2) \cap A_3$$



$$(A_1 \cup \overline{A_2}) \cap A_3$$



Aufgabenteil VI: Analyse mehrstufiger Zufallsexperimente (10 Punkte)

In einem geschlossenen Karton befinden sich 20 Fachbücher – 10 Physikbücher, 5 Mathematikbücher und 5 Statistikbücher. Ein Student zieht zufällig und mit geschlossenen Augen hintereinander drei Bücher aus dem Karton, ohne diese wieder zurückzulegen. Wie groß ist die Wahrscheinlichkeit dafür, dass er...

- a) ...ausschließlich Mathematikbücher zieht? (3 Punkte)
- b) ...ausschließlich Physikbücher zieht? (3 Punkte)
- c) ...genau zwei Statistikbücher zieht? (4 Punkte)

Aufgabenteil VII: Variationen und Kombinationen (10 Punkte)

Ein Student besitzt 8 T-Shirts, 3 Jogginghosen und 2 Paar Turnschuhe, die er zum Training immer wieder neu miteinander kombinieren möchte. Zu wie vielen Trainings muss er gehen, um jede Kombinationsmöglichkeit genau einmal ausprobiert zu haben?

Aufgabenteil VIII: Berechnung bedingter Wahrscheinlichkeiten (10 Punkte)

Ein Drogenspürhund am Flughafen schlägt beim Vorhandensein von Drogen im Gepäck mit einer Wahrscheinlichkeit von 98% an. Befinden sich keine Drogen im Gepäck, schlägt er allerdings mit einer Wahrscheinlichkeit von 4% ebenfalls an. An einem beliebigen Tag durchlaufen 300 Reisende die Sicherheitsschleuse mit dem Drogenspürhund, wobei 10 Reisende Drogen mit sich führen.

Wie groß ist die Wahrscheinlichkeit dafür, dass nach einem Anschlagen des Drogenspürhunds auch tatsächlich Drogen im Gepäck gefunden werden können?

Aufgabenteil IX: Chi-Quadrat-Unabhängigkeitstest (10 Punkte)

Im Rahmen einer Betriebsgesundheits-Untersuchung wird der Blutdruck von 100 Mitarbeitern/innen gemessen. Auf einem Self-Assessment-Bogen müssen die Mitarbeiter/innen außerdem angeben, ob sie in ihrer Freizeit gar keinen, wenig oder viel Sport betreiben. Es ergibt sich folgendes Bild:

	Hat keinen Bluthochdruck	Hat Bluthochdruck
Treibt keinen Sport	3	24
Treibt wenig Sport	12	16
Treibt viel Sport	35	10

Führen Sie auf Basis dieser Daten einen Chi-Quadrat-Test auf stochastische Unabhängigkeit der beiden Variablen (Bluthochdruck / Sportintensität) durch (7 Punkte) und interpretieren Sie das Ergebnis (3 Punkte). Als Irrtumswahrscheinlichkeit / Wahrscheinlichkeit für Fehler 1. Ordnung ist 1% zu wählen (kritischer Wert: 6,63).

LÖSUNGEN – KLAUSUR STATISTIK

Berufsbegleitender Studiengang Betriebswirtschaftslehre
Sommersemester 2016
Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☒ falsch

2) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☒ richtig ☐ falsch

3) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☒ falsch

4) Bei der Klassierung von Daten dürfen auch unterschiedlich breite Klassen gebildet werden.

☒ richtig ☐ falsch

5) Variablen mit nur zwei möglichen Ausprägungen werden als dichotom bezeichnet.

☒ richtig ☐ falsch

6) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des Y-Achsenschnittpunkts der linearen Regressionsgeraden.

☒ richtig ☐ falsch

7) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☒ richtig ☐ falsch

8) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☒ richtig ☐ falsch

9) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

☒ richtig ☐ falsch

10) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☒ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Eine Umfrage unter 30 Studierenden der Hochschule Harz ergab u.a. die folgende Altersverteilung:

Nr.	Alter	Nr.	Alter	Nr.	Alter
1	21 Jahre	11	33 Jahre	21	19 Jahre
2	19 Jahre	12	26 Jahre	22	22 Jahre
3	23 Jahre	13	81 Jahre	23	29 Jahre
4	22 Jahre	14	19 Jahre	24	34 Jahre
5	21 Jahre	15	18 Jahre	25	22 Jahre
6	24 Jahre	16	22 Jahre	26	24 Jahre
7	26 Jahre	17	26 Jahre	27	20 Jahre
8	22 Jahre	18	30 Jahre	28	21 Jahre
9	31 Jahre	19	20 Jahre	29	25 Jahre
10	27 Jahre	20	24 Jahre	30	32 Jahre

Zeichnen Sie...

- a) ...ein Stamm-Blatt-Diagramm mit einer Stammbreite von 10. (5 Punkte)
 b) ...ein Stamm-Blatt-Diagramm mit einer Stammbreite von 5. (5 Punkte)

18; 19; 19; 19

20; 20; 21; 21; 21; 22; 22; 22; 22; 22; 23; 24; 24; 24; 25; 26; 26; 26; 27; 29

30; 31; 32; 33; 34

81

Stammbreite von 10

1 | 8 9 9 9

2 | 00 1 1 1 2 2 2 2 2 3 4 4 4 5 6 6 6 7 9

3 | 0 1 2 3 4

Jedes Blatt ein Fall

Ein Ausreißer

Stammbreite von 5

1 | 8 9 9 9

2 | 00 1 1 1 2 2 2 2 2 3 4 4 4

2 | 5 6 6 6 7 9

3 | 0 1 2 3 4

Jedes Blatt ein Fall

Ein Ausreißer

Aufgabenteil III: Verteilungsparameter (20 Punkte)

Um die maximal mögliche Wasserentnahmemenge aus der Kalten Bode für eine Beschneiungsanlage zu bestimmen, wird über ein Jahr jeweils am zweiten Dienstag jedes Monats der Wasserstand an einem vorab festgelegten Messpunkt erhoben. Es ergeben sich folgende Werte:

Monat	Wasserstand
Januar	1,17 m
Februar	1,29 m
März	1,23 m
April	1,18 m
Mai	1,25 m
Juni	1,04 m
Juli	1,08 m
August	0,98 m
September	1,12 m
Oktober	1,14 m
November	1,19 m
Dezember	1,21 m

Maße der zentralen Tendenz:

- Bestimmen Sie das arithmetische Mittel. (3 Punkte)
- Bestimmen Sie den Median. (3 Punkte)
- Bestimmen Sie den Modus. (3 Punkte)

Streuungsmaße / Dispersionsparameter:

- Bestimmen Sie die Varianz. (3 Punkte)
- Bestimmen Sie den Interquartilsabstand. (3 Punkte)

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie die Kurtosis / Exzeß. (5 Punkte)

Arithmetisches Mittel

$$1,17 + 1,29 + 1,23 + 1,18 + 1,25 + 1,04 + 1,08 + 0,98 + 1,12 + 1,14 + 1,19 + 1,21 = 13,88$$

$$13,88 / 12 = 1,156666667 = 1,1567$$

Median

$$0,98; 1,04; 1,08; 1,12; 1,14; 1,17; 1,18; 1,19; 1,21; 1,23; 1,25; 1,29$$

$$\text{Mittel des 6. und 7. Werts: } (1,17 + 1,18) / 2 = 1,175$$

Modus

(nicht bestimmbar)

Varianz

x	$(x - x_q)^2$
1,17	0,000169
1,29	0,017689
1,23	0,005329
1,18	0,000529
1,25	0,008649
1,04	0,013689
1,08	0,005929
0,98	0,031329
1,12	0,001369
1,14	0,000289
1,19	0,001089
1,21	0,002809

Summe: 0,088868

 $0,088868 / 12 = 0,007405667 = 0,007$ Interquartilsabstand

0,98; 1,04; 1,08; 1,12; 1,14; 1,17; 1,18; 1,19; 1,21; 1,23; 1,25; 1,29

0,25 * 12 = 3 -> Mittel zwischen 3. und 4. Wert: $(1,08 + 1,12) / 2 = 1,10$ 0,75 * 12 = 9 -> Mittel zwischen 9. und 10. Wert: $(1,21 + 1,23) / 2 = 1,22$ IQR = $1,22 - 1,10 = 0,12$ Kurtosis/Exzeß

x	$(x - x_q)^2$	$(x - x_q)^4$
1,17	0,000169	0,0256
1,29	0,017689	0,00614656
1,23	0,005329	0,01336336
1,18	0,000529	0,02313441
1,25	0,008649	0,01048576
1,04	0,013689	0,07890481
1,08	0,005929	0,05764801
0,98	0,031329	0,12117361
1,12	0,001369	0,04100625
1,14	0,000289	0,03418801
1,19	0,001089	0,02085136
1,21	0,002809	0,01679616

s4: $0,088868 / 12 = 0,007405667 \rightarrow \text{Wurzel } (0,007)^4 = 0,0000548439$ m4: $0,0016268461 / 12 = 0,000135571$ $(m4/s4) - 3 = -0,528056903$

Aufgabenteil IV: Zusammenhangsmaße (10 Punkte)

Studierenden ist die Teilnahme an acht zusätzlichen Tutoriumsterminen grundsätzlich freigestellt. Im Anschluss an eine Klausur soll überprüft werden, ob zwischen der Teilnahme an diesen Tutorien und der späteren Klausurnote ein monotoner Zusammenhang besteht.

Student Nr.	Absolvierte Tutorien	Klausurnote
1	0	3,4
2	2	5,0
3	3	2,8
4	5	2,0
5	1	2,8
6	4	2,3
7	4	1,7
8	6	1,3
9	8	3,0
10	7	2,5

Berechnen und interpretieren Sie...

- a) ...den Konkordanzkoeffizienten nach Kendall. (5 Punkte)
 b) ...den Rangkorrelationskoeffizienten nach Spearman. (5 Punkte)

Kendall

Nr.	x	rg(x)	y	rg(y)	K	D
1	0	1	3,4	9	1	8
2	1	2	2,8	6,5	2	5
3	2	3	5	10	0	7
4	3	4	2,8	6,5	1	5
5	4	5,5	2,3	4	2	3
6	4	5,5	1,7	2	3	1
7	5	7	2	3	2	1
8	6	8	1,3	1	2	0
9	7	9	2,5	5	1	0
10	8	10	3	8	//	//

$$K = 14$$

$$D = 30$$

$$\tau = (2 * (14 - 30)) / (10 * 9) = -0,3556$$

Spearman

Nr.	x	rg(x)	y	rg(y)	d	d ²
1	0	1	3,4	9	-8	64
2	2	3	5,0	10	-7	49
3	3	4	2,8	6,5	-2,5	6,25
4	5	7	2,0	3	4	16
5	1	2	2,8	6,5	-4,5	20,25
6	4	5,5	2,3	4	1,5	2,25
7	4	5,5	1,7	2	3,5	12,25
8	6	8	1,3	1	7	49
9	8	10	3,0	8	2	4
10	7	9	2,5	5	4	16

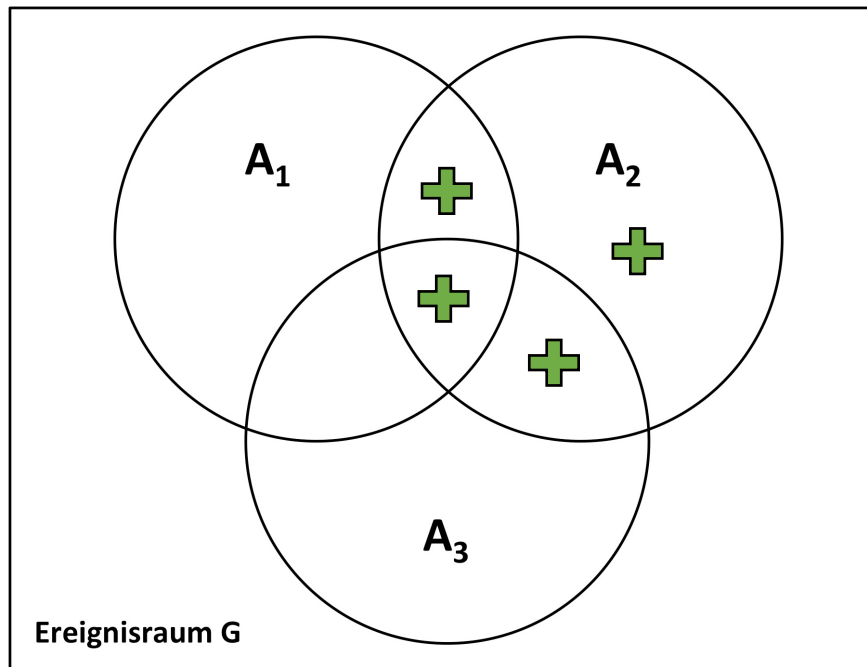
Summe d² = 239

$\rho = 1 - ((6 * 239) / ((10^2 - 1) * 10)) = -0,4485$

Aufgabenteil V: Interpretation von Venn-Diagrammen (10 Punkte)

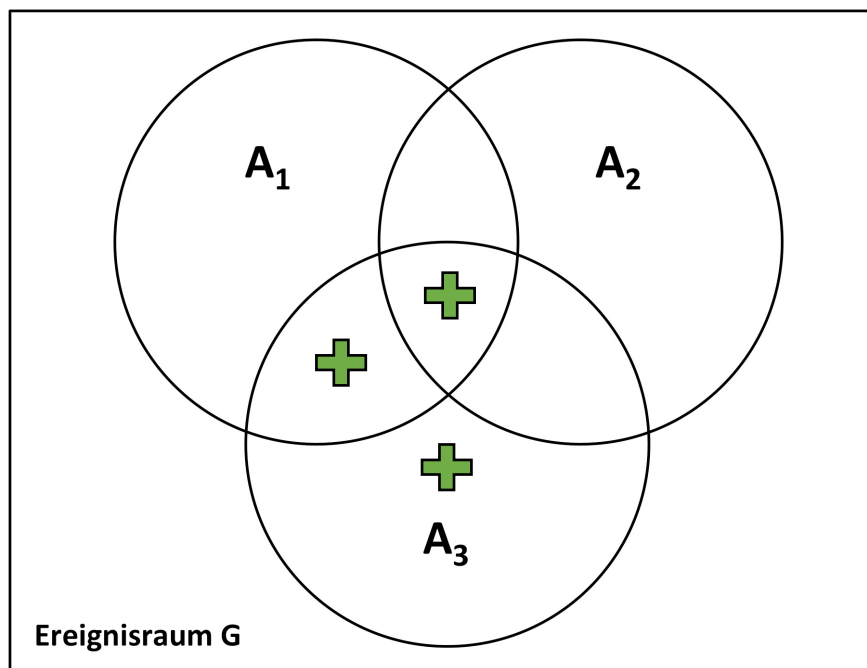
Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

$$(A_1 \cup A_2) \cap A_2 = (A_1 \cap A_2) \cup (A_2 \cap A_2) = A_2 \text{ komplett}$$



$$(A_1 \cup \overline{A_2}) \cap A_3 = (A_1 \cap A_3) \cup (\overline{A_2} \cap A_3)$$

= Schnittfläche von A_1 und A_3 plus A_3 ohne Schnittfläche mit A_2



Aufgabenteil VI: Analyse mehrstufiger Zufallsexperimente (10 Punkte)

In einem geschlossenen Karton befinden sich 20 Fachbücher – 10 Physikbücher, 5 Mathematikbücher und 5 Statistikbücher. Ein Student zieht zufällig und mit geschlossenen Augen hintereinander drei Bücher aus dem Karton, ohne diese wieder zurückzulegen. Wie groß ist die Wahrscheinlichkeit dafür, dass er...

- a) ...ausschließlich Mathematikbücher zieht? (3 Punkte)
- b) ...ausschließlich Physikbücher zieht? (3 Punkte)
- c) ...genau zwei Statistikbücher zieht? (4 Punkte)

Ausschließlich Mathematikbücher

$$5/20 * 4/19 * 3/18 = 0,00877193 = 0,87\%$$

Ausschließlich Physikbücher

$$10/20 * 9/19 * 8/18 = 0,105263158 = 10,53\%$$

Genau zwei Statistikbücher

Welche Pfade sind hier zu berücksichtigen?

S-S-M	$(5/20) * (4/19) * (5/18)$	0,014619883
S-S-P	$(5/20) * (4/19) * (10/18)$	0,029239766
S-M-S	$(5/20) * (5/19) * (4/18)$	0,014619883
S-P-S	$(5/20) * (10/19) * (4/18)$	0,029239766
M-S-S	$(5/20) * (5/19) * (4/18)$	0,014619883
P-S-S	$(10/20) * (5/19) * (4/18)$	0,029239766

$$\text{Summe: } 0,131578947 = 13,16\%$$

Aufgabenteil VIII: Variationen und Kombinationen (10 Punkte)

Ein Student besitzt 8 T-Shirts, 3 Jogginghosen und 2 Paar Turnschuhe, die er zum Training immer wieder neu miteinander kombinieren möchte. Zu wie vielen Trainings muss er gehen, um jede Kombinationsmöglichkeit genau einmal ausprobiert zu haben?

$$8 * 3 * 2 = 48$$

Aufgabenteil VIII: Berechnung bedingter Wahrscheinlichkeiten (10 Punkte)

Ein Drogenspürhund am Flughafen schlägt beim Vorhandensein von Drogen im Gepäck mit einer Wahrscheinlichkeit von 98% an. Befinden sich keine Drogen im Gepäck, schlägt er allerdings mit einer Wahrscheinlichkeit von 4% ebenfalls an. An einem beliebigen Tag durchlaufen 300 Reisende die Sicherheitsschleuse mit dem Drogenspürhund, wobei 10 Reisende Drogen mit sich führen.

Wie groß ist die Wahrscheinlichkeit dafür, dass nach einem Anschlag des Drogenspürhunds auch tatsächlich Drogen im Gepäck gefunden werden können?

Gefundene Drogen:	Fehlalarme:
$10 * 0,98 = 9,8$	$290 * 0,04 = 11,6$

Trefferwahrscheinlichkeit:
 $9,8 / (9,8 + 11,6) = 0,4579 = 45,79\%$

Aufgabenteil IX: Chi-Quadrat-Unabhängigkeitstest (10 Punkte)

Im Rahmen einer Betriebsgesundheits-Untersuchung wird der Blutdruck von 100 Mitarbeitern/innen gemessen. Auf einem Self-Assessment-Bogen müssen die Mitarbeiter/innen außerdem angeben, ob sie in ihrer Freizeit gar keinen, wenig oder viel Sport betreiben. Es ergibt sich folgendes Bild:

	Hat keinen Bluthochdruck	Hat Bluthochdruck
Treibt keinen Sport	3	24
Treibt wenig Sport	12	16
Treibt viel Sport	35	10

Führen Sie auf Basis dieser Daten einen Chi-Quadrat-Test auf stochastische Unabhängigkeit der beiden Variablen (Bluthochdruck / Sportintensität) durch (7 Punkte) und interpretieren Sie das Ergebnis (3 Punkte). Als Irrtumswahrscheinlichkeit / Wahrscheinlichkeit für Fehler 1. Ordnung ist 1% zu wählen (kritischer Wert: 6,63).

	Hat keinen Bluthochdruck	Hat Bluthochdruck	Randsumme
Treibt keinen Sport	3	24	27
Treibt wenig Sport	12	16	28
Treibt viel Sport	35	10	45
Randsumme	50	50	100

Tabelle der erwarteten Häufigkeiten

	Hat keinen Bluthochdruck	Hat Bluthochdruck	Randsumme
Treibt keinen Sport	13,5	13,5	27
Treibt wenig Sport	14	14	28
Treibt viel Sport	22,5	22,5	45
Randsumme	50	50	100

$$(3-13,5)^2/13,5 = 8,166666667$$

$$(12-14)^2/14 = 0,285714286$$

$$(35-22,5)^2/22,5 = 6,944444444$$

$$(24-13,5)^2/13,5 = 8,166666667$$

$$(16-14)^2/14 = 0,285714286$$

$$(10-22,5)^2/22,5 = 6,944444444$$

$$\text{Chi-Quadrat-Wert} = 30,79365079$$

Liegt über dem kritischen Wert: Die Nullhypothese kann verworfen werden, ein Zusammenhang zwischen den betrachteten Variablen ist wahrscheinlich.

Wintersemester 2016/2017

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Wintersemester 2016 / 2017

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 12. November 2016

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2016 / 2017

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für exponentielle Zusammenhänge.

☐ richtig ☐ falsch

2) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☐ falsch

3) Die Berechnung des Rangkorrelationskoeffizienten nach Spearman erfordert metrische Daten.

☐ richtig ☐ falsch

4) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☐ falsch

5) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen 0 und 1 annehmen.

☐ richtig ☐ falsch

6) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht der Wurzel aus dem Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

7) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

☐ richtig ☐ falsch

8) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☐ falsch

9) Variablen mit nur zwei möglichen Ausprägungen werden als bivariat bezeichnet.

☐ richtig ☐ falsch

10) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Bei einer Befragung von 10 Studierenden wurden auf die Frage nach den monatlichen Wohnkosten folgende Angaben gemacht.

Student Nr.	Monatliche Wohnkosten (in EUR)
1	270
2	210
3	315
4	330
5	190
6	240
7	270
8	310
9	180
10	230

Zeichnen Sie für diese Verteilung...

- einen erweiterten Box-Plot (10 Punkte).
- ein Stem-and-Leaf-Diagramm mit einer aus Ihrer Sicht geeigneten Stammbreite (10 Punkte).

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Bei einer Statistik-Klausur an der Hochschule Harz ergab sich die folgende Notenverteilung:

Note	Anzahl Studenten/innen
1	10
2	12
3	8
4	5
5	5

Maße der zentralen Tendenz:

- Bestimmen Sie das arithmetische Mittel. (4 Punkte)
- Bestimmen Sie den Median. (4 Punkte)
- Bestimmen Sie den Modus. (4 Punkte)

Streuungsmaße / Dispersionsparameter:

- Bestimmen Sie die Varianz. (5 Punkte)
- Bestimmen Sie die Spannweite. (4 Punkte)
- Bestimmen Sie den Interquartilsabstand. (4 Punkte)

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie den Quartilkoeffizienten der Schiefe. (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Ein Kinobetreiber verändert (*ceteris paribus*) an 10 Tagen die Eintrittspreise für den Abendfilm und zeichnet die Besucherzahlen auf. Es ergibt sich die folgende Tabelle:

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	760
2	9,00	735
3	9,50	841
4	10,00	700
5	10,50	673
6	11,00	510
7	11,50	645
8	12,00	531
9	12,50	421
10	13,00	413

Berechnen (15 Punkte) und interpretieren (5 Punkte)
Sie den Bravais-Pearson-Korrelationskoeffizienten.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Berechnen Sie für die bereits bekannte Verteilung der Eintrittspreise (unabhängige Variable x) und Besucherzahlen (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte) Diskutieren Sie, ob die Ergebnisse zum Bravais-Pearson-Regressionskoeffizienten aus Aufgabenteil IV passen. (5 Punkte)

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	760
2	9,00	735
3	9,50	841
4	10,00	700
5	10,50	673
6	11,00	510
7	11,50	645
8	12,00	531
9	12,50	421
10	13,00	413

MUSTERLÖSUNGEN KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2016 / 2017

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für exponentielle Zusammenhänge.

O richtig **X falsch**

2) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

O richtig **X falsch**

3) Die Berechnung des Rangkorrelationskoeffizienten nach Spearman erfordert metrische Daten.

O richtig **X falsch**

4) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

X richtig O falsch

5) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen 0 und 1 annehmen.

O richtig **X falsch**

6) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht der Wurzel aus dem Bravais-Pearson-Korrelationskoeffizienten.

O richtig **X falsch**

7) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

O richtig **X falsch**

8) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

O richtig **X falsch**

9) Variablen mit nur zwei möglichen Ausprägungen werden als bivariat bezeichnet.

X richtig O falsch

10) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

X richtig O falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Bei einer Befragung von 10 Studierenden wurden auf die Frage nach den monatlichen Wohnkosten folgende Angaben gemacht.

Student Nr.	Monatliche Wohnkosten (in EUR)
1	270
2	210
3	315
4	330
5	190
6	240
7	270
8	310
9	180
10	230

Zeichnen Sie für diese Verteilung...

- c) ...einen erweiterten Box-Plot (10 Punkte).
- d) ...ein Stem-and-Leaf-Diagramm mit einer aus Ihrer Sicht geeigneter Stammbreite (10 Punkte).

a) Erweiterter Box-Plot

Schritt 1: Alle Werte in eine geordnete Reihenfolge bringen:

180, 190, 210, 230, 240, 270, 270, 310, 315, 330

Schritt 2: Berechnung der Quartile sowie des IQR:

Unteres Quartil: $10 \cdot 0,25 = 2,5 \rightarrow 3$. Wert = 210

Mittleres Quartil: $10 \cdot 0,50 = 5,0 \rightarrow (240 + 270) / 2 = 255$

Oberes Quartil: $10 \cdot 0,75 = 7,5 \rightarrow 8$. Wert = 310

Interquartilsabstand: $310 - 210 = 100$

1,5-facher Interquartilsabstand: 150

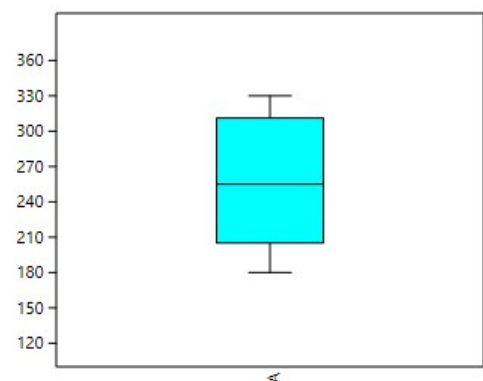
Schritt 3: Konstruktion des Box-Plots:

Die Box verläuft von 210 auf 310.

Der Median wird bei 255 eingezeichnet.

Der obere Zaun kann maximal bis zum Wert $310 + 150 = 460$ laufen. Der größte noch in diesem Bereich liegende Wert ist die 330. Da keine größeren Werte existieren, gibt es nach oben keine Ausreißer oder Extremwerte.

Der untere Zaun kann minimal bis zum Wert $210 - 150 = 60$ laufen. Der kleinste noch in diesem Bereich liegende Wert ist die 180. Da keine kleineren Werte existieren, gibt es nach unten keine Ausreißer oder Extremwerte.



b) Stem-and-Leaf-Diagramm mit geeigneter Stammbreite (z.B. 100)

```

1 | 8 9
2 | 1 3 4 7 7
3 | 1 1 3
  
```

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Bei einer Statistik-Klausur an der Hochschule Harz ergab sich die folgende Notenverteilung:

Note	Anzahl Studenten/innen
1	10
2	12
3	8
4	5
5	5

Maße der zentralen Tendenz:

- Bestimmen Sie das arithmetische Mittel. (4 Punkte)
- Bestimmen Sie den Median. (4 Punkte)
- Bestimmen Sie den Modus. (4 Punkte)

Streuungsmaße / Dispersionsparameter:

- Bestimmen Sie die Varianz. (5 Punkte)
- Bestimmen Sie die Spannweite. (4 Punkte)
- Bestimmen Sie den Interquartilsabstand. (4 Punkte)

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie den Quartilkoeffizienten der Schiefe. (5 Punkte)

a) Arithmetisches Mittel: Nicht möglich, da es sich um ordinalskalierte Daten handelt.

b) Median: $40 \cdot 0,5 = 20 \rightarrow (20. + 21. \text{ Wert}) / 2 \rightarrow (2 + 2) / 2 = 2$

c) Modus: Ein eindeutiger Modus lässt sich nicht identifizieren.

d) Varianz: Nicht möglich, da es sich um ordinalskalierte Daten handelt.

e) Spannweite: Nicht möglich, da es sich um ordinalskalierte Daten handelt.

f) Interquartilsabstand:

$$40 \cdot 0,25 = 10 \rightarrow (10. + 11. \text{ Wert}) / 2 \rightarrow (1 + 2) / 2 = 1,5$$

$$40 \cdot 0,75 = 30 \rightarrow (30. + 31. \text{ Wert}) / 2 \rightarrow (3 + 4) / 2 = 3,5$$

$$IQR = 3,5 - 1,5 = 2$$

g) Quartilkoeffizient der Schiefe:

$$g_{0,25} = [(3,5 - 2) - (2 - 1,5)] / 2$$

$$g_{0,25} = (1,5 - 0,5) / 2$$

$$g_{0,25} = 1 / 2 = 0,5$$

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Ein Kinobetreiber verändert (*ceteris paribus*) an 10 Tagen die Eintrittspreise für den Abendfilm und zeichnet die Besucherzahlen auf. Es ergibt sich die folgende Tabelle:

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	760
2	9,00	735
3	9,50	841
4	10,00	700
5	10,50	673
6	11,00	510
7	11,50	645
8	12,00	531
9	12,50	421
10	13,00	413

Berechnen (15 Punkte) und interpretieren (5 Punkte)
 Sie den Bravais-Pearson-Korrelationskoeffizienten.

Nr.	x	y	x ²	y ²	x*y
1	8,50	760,00	72,25	577600	6460,00
2	9,00	735,00	81,00	540225	6615,00
3	9,50	841,00	90,25	707281	7989,50
4	10,00	700,00	100,00	490000	7000,00
5	10,50	673,00	110,25	452929	7066,50
6	11,00	510,00	121,00	260100	5610,00
7	11,50	645,00	132,25	416025	7417,50
8	12,00	531,00	144,00	281961	6372,00
9	12,50	421,00	156,25	177241	5262,50
10	13,00	413,00	169,00	170569	5369,00
Summe	107,50	6229,00	1176,25	4073931	65162,00
Mittel	10,75	622,90	irrelevant	irrelevant	irrelevant

$$r = \frac{65162 - 10 * 10,75 * 622,9}{\sqrt{1176,25 - 10 * 10,75^2} * \sqrt{4073931 - 10 * 622,9^2}} = \frac{-1799,75}{4,54 * 440,32} = -0,89$$

Es liegt eine starke negative lineare Korrelation vor.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Berechnen Sie für die bereits bekannte Verteilung der Eintrittspreise (unabhängige Variable x) und Besucherzahlen (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte) Diskutieren Sie, ob die Ergebnisse zum Bravais-Pearson-Regressionskoeffizienten aus Aufgabenteil IV passen. (5 Punkte)

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	760
2	9,00	735
3	9,50	841
4	10,00	700
5	10,50	673
6	11,00	510
7	11,50	645
8	12,00	531
9	12,50	421
10	13,00	413

Nr.	x	y	x^2	$x \cdot y$
1	8,50	760,00	72,25	6460,00
2	9,00	735,00	81,00	6615,00
3	9,50	841,00	90,25	7989,50
4	10,00	700,00	100,00	7000,00
5	10,50	673,00	110,25	7066,50
6	11,00	510,00	121,00	5610,00
7	11,50	645,00	132,25	7417,50
8	12,00	531,00	144,00	6372,00
9	12,50	421,00	156,25	5262,50
10	13,00	413,00	169,00	5369,00
Summe	107,50	6229,00	1176,25	65162,00
Mittel	10,75	622,90	irrelevant	irrelevant

$$b = [65162 - 10 \cdot 10,75 \cdot 622,9] / [1176,25 - 10 \cdot 10,75^2] = -1799,75 / 20,625 = -87,26$$

$$a = 622,90 - (-87,26 \cdot 10,75) = 1560,95$$

$$y = 1560,95 - 87,26 \cdot x$$

Das Gütemaß R^2 entspricht dem quadrierten BPK und liegt bei 0,81.

Wintersemester 2016/2017

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Wintersemester 2016 / 2017

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 14. Januar 2017

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2016 / 2017

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

☐ richtig ☐ falsch

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

3) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

☐ richtig ☐ falsch

4) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☐ richtig ☐ falsch

5) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

☐ richtig ☐ falsch

6) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht der Wurzel aus dem Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

7) Liegt der Momentenkoeffizient der Schiefe nahe Null, ist die Verteilung rechtsschief.

☐ richtig ☐ falsch

8) Aus Stem-and-Leaf-Diagrammen kann man die Ursprungsdaten direkt ablesen.

☐ richtig ☐ falsch

9) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

10) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die nachfolgende Tabelle gibt die „Ärztedichte“ (die Anzahl der Einwohner/innen je berufstätigem Arzt / je berufstätiger Ärztin – unabhängig von dessen / deren Fachgebiet) für alle 16 Bundesländer wieder. (Quelle: Bundesärztekammer)

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

Zeichnen Sie einen (erweiterten) Box-Plot für diese Verteilung (5 Punkte)
und benennen Sie alle für die Konstruktion benötigten Größen (5 Punkte).

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Bestimmen Sie für die bereits bekannten Daten zur „Ärztendichte“...

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

- a) das arithmetische Mittel,
- b) den Median,
- c) die Spannweite und
- d) den Quartilkoeffizienten der Schiefe.
- e) Geben Sie die Fünf-Werte-Zusammenfassung an. (jeweils 3 Punkte)

Aufgabenteil IV: Lineare Regressionsanalyse (15 Punkte)

Stellen Sie für die Verteilung von Körpergröße (unabhängige Variable x) und Körpergewicht (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte).

Person Nr.	Körpergröße (in cm)	Körpergewicht (in kg)
1	172	81
2	174	87
3	161	73
4	158	71
5	173	88
6	181	91
7	183	95
8	170	82
9	172	82
10	166	76

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Das Gegenereignis zu mindestens 3 Augen beim einmaligen Wurf eines fairen Würfels lautet höchstens 2 Augen.

☐ richtig ☐ falsch

2) Ein korrekt durchgeführter statistischer Test gestattet eine definitive Aussage über die Korrektheit von Null- und Alternativhypothese.

☐ richtig ☐ falsch

3) Die Varianz $\text{Var}(X)$ kann ausschließlich für diskrete Zufallsvariablen berechnet werden.

☐ richtig ☐ falsch

4) Bei einer stetigen Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

☐ richtig ☐ falsch

5) Eine Stichprobe ist nur dann repräsentativ, wenn sie exakt den mit Hilfe der Cochran-Formel bestimmten Umfang aufweist.

☐ richtig ☐ falsch

6) Eine Menge kann beliebig viele Teilmengen enthalten.

☐ richtig ☐ falsch

7) Ist die Schnittmenge der Mengen A und B leer, so bedeutet dies, dass die Menge A die Komplementärmenge der Menge B (und umgekehrt) ist.

☐ richtig ☐ falsch

8) Eine assoziative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

☐ richtig ☐ falsch

9) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig ☐ falsch

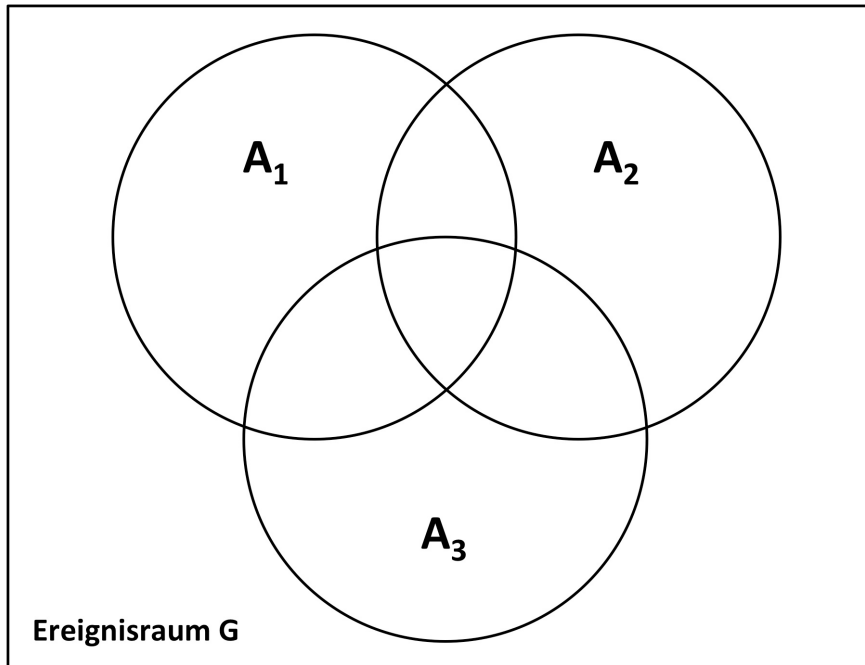
10) Elemente, die gleichzeitig der Menge A als auch der Menge B angehören, sind stets auch Bestandteil der Vereinigungsmenge der Mengen A und B.

☐ richtig ☐ falsch

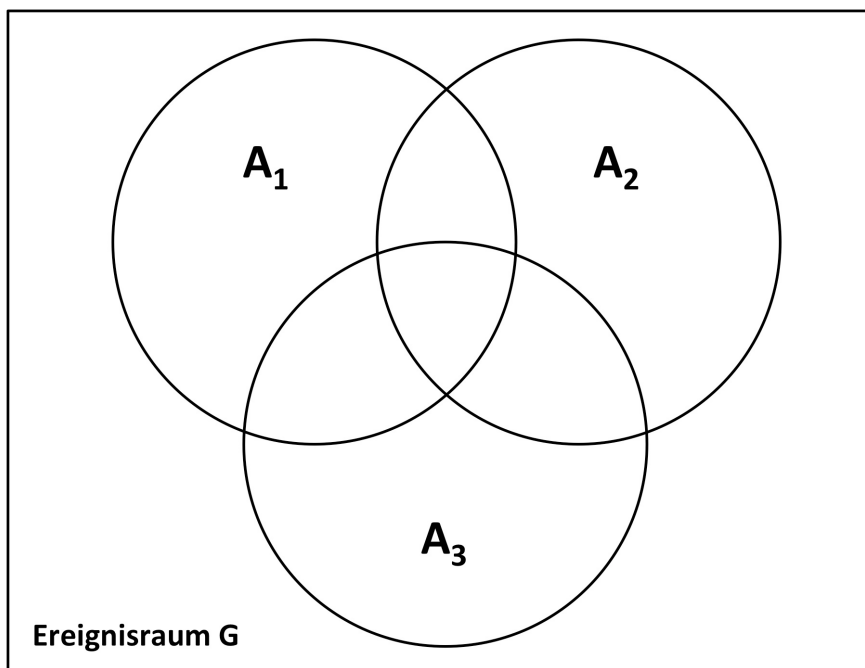
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

a) $(A_1 \cup A_2) \cap \bar{A}_3$



b) $(A_1 \cap A_2) \cup (A_2 \cap A_3)$



Aufgabenteil VII: Kombinatorik (15 Punkte)

Aus den fünf Ziffern 1, 2, 3, 4 und 5 soll eine Liste mit allen möglichen fünfstelligen Zahlen gebildet werden, also z.B. 12345, 21435, 54321 etc. pp. Zu dieser Liste sind zwei Fragen zu beantworten.

- 1) Wie viele Zahlen enthält die Liste insgesamt? (10 Punkte)
- 2) Wie viele dieser Zahlen beginnen mit einer 3? (5 Punkte)

Aufgabenteil VIII: Bedingte Wahrscheinlichkeiten (15 Punkte)

Eine Sicherheitssoftware für die Analyse von Video-Überwachungsaufnahmen an einer Flughafen-Sicherheitsschleuse kann das Gesicht einer gesuchten Person mit einer Wahrscheinlichkeit von 92% erkennen. Allerdings identifiziert die Software in 3% aller Fälle eine nicht gesuchte Person irrtümlich als gesucht.

Die Sicherheitsbehörden gehen davon aus, dass an einem bestimmten Tag eine Gruppe von 10 gesuchten Personen versuchen wird, die Schleuse zu passieren. Das übliche Personenaufkommen pro Tag liegt bei 10.000 Fluggästen. Mit der Präsenz weiterer gesuchter Personen ist am betrachteten Tag nicht zu rechnen.

- a) Mit wie vielen fälschlicherweise als "gesucht" identifizierten Personen ist zu rechnen? (5 Punkte)
- b) Die Software schlägt Alarm. Wie groß ist die Wahrscheinlichkeit dafür, dass tatsächlich eine gesuchte Person entdeckt wurde? (10 Punkte)

MUSTERLÖSUNG KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2016 / 2017

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

X richtig O falsch

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

X richtig O falsch

3) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

O richtig **X falsch**

4) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

X richtig O falsch

5) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

X richtig O falsch

6) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht der Wurzel aus dem Bravais-Pearson-Korrelationskoeffizienten.

O richtig **X falsch**

7) Liegt der Momentenkoeffizient der Schiefe nahe Null, ist die Verteilung rechtsschief.

O richtig **X falsch**

8) Aus Stem-and-Leaf-Diagrammen kann man die Ursprungsdaten direkt ablesen.

X richtig O falsch

9) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

O richtig **X falsch**

10) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

O richtig **X falsch**

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die nachfolgende Tabelle gibt die „Ärztedichte“ (die Anzahl der Einwohner/innen je berufstätigem Arzt / je berufstätiger Ärztin – unabhängig von dessen / deren Fachgebiet) für alle 16 Bundesländer wieder. (Quelle: Bundesärztekammer)

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

Zeichnen Sie einen (erweiterten) Box-Plot für diese Verteilung (5 Punkte) und benennen Sie alle für die Konstruktion benötigten Größen (5 Punkte).

$$X_{0,25} = 0,25 \cdot 16 = 4 \rightarrow \text{ganzzahlig, daher} \rightarrow (x_4 + x_5) / 2 = (205 + 208) / 2 = 206,5$$

$$X_{0,50} = 0,50 \cdot 16 = 8 \rightarrow \text{ganzzahlig, daher} \rightarrow (x_8 + x_9) / 2 = (223 + 225) / 2 = 224,0$$

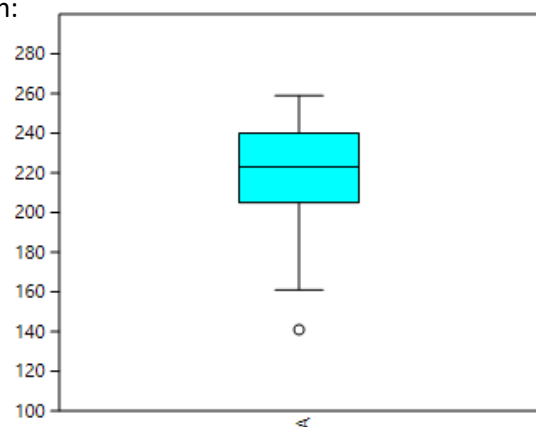
$$X_{0,75} = 0,75 \cdot 16 = 12 \rightarrow \text{ganzzahlig, daher} \rightarrow (x_{12} + x_{13}) / 2 = (235 + 240) / 2 = 237,5$$

$$\text{IQR} = 237,5 - 206,5 = 31$$

$$1,5 \cdot \text{IQR} = 46,5$$

Für die Konstruktion der Box ergeben sich folgende Größen:

- Die Box läuft von 206,5 auf 237,5 mit einer Median-Linie bei 224.
- Der obere Zaun könnte maximal bis $237,5 + 46,5 = 284$ verlaufen. Er endet daher beim größten Wert innerhalb des Intervalls: 259.
- Der untere Zaun könnte minimal bis $206,5 - 46,5 = 160$ verlaufen. Er endet daher beim kleinsten Wert innerhalb des Intervalls: 161.
- Nach unten ist ein Ausreißer (141) zu kennzeichnen.
- Weitere Ausreißer oder Extremwerte liegen nicht vor.



Aufgabenteil III: Verteilungsparameter (15 Punkte)

Bestimmen Sie für die bereits bekannten Daten zur „Ärztendichte“...

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

- f) das arithmetische Mittel,
- g) den Median,
- h) die Spannweite und
- i) den Quartilskoeffizienten der Schiefe.
- j) Geben Sie die Fünf-Werte-Zusammenfassung an. (jeweils 3 Punkte)

a) arithmetisches Mittel

$$259 + 249 + 249 + 240 + 235 + 229 + 226 + 225 + 223 + 223 + 219 + 208 + 205 + 169 + 161 + 141$$

$$= 3461 / 16 = 216,3125$$

b) Median

Wurde bereits in Aufgabenteil II berechnet: 224

c) Spannweite

$$259 - 141 = 118$$

d) Quartilskoeffizient der Schiefe

Alle erforderlichen Größen wurden in Aufgabenteil II berechnet.

$$((x_{0,75} - x_{0,50}) - (x_{0,50} - x_{0,25})) / IQR = ((237,5 - 224) - (224 - 206,5)) / 31 = -0,1290$$

e) Fünf-Werte-Zusammenfassung

$$[259; 237,5; 224; 206,5; 141]$$

Aufgabenteil IV: Lineare Regressionsanalyse (15 Punkte)

Stellen Sie für die Verteilung von Körpergröße (unabhängige Variable x) und Körpergewicht (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte).

Person Nr.	Körpergröße (in cm)	Körpergewicht (in kg)
1	172	81
2	174	87
3	161	73
4	158	71
5	173	88
6	181	91
7	183	95
8	170	82
9	172	82
10	166	76

Schritt 1: Aufstellung der linearen Regressionsfunktion

Nr.	X	Y	X ²	Y ²	(X*Y)
1	172	81	29584	6561	13932
2	174	87	30276	7569	15138
3	161	73	25921	5329	11753
4	158	71	24964	5041	11218
5	173	88	29929	7744	15224
6	181	91	32761	8281	16471
7	183	95	33489	9025	17385
8	170	82	28900	6724	13940
9	172	82	29584	6724	14104
10	166	76	27556	5776	12616
SUMME	1710	826	292964	68774	141781
MITTEL	171	82,6	//	//	//

$$b = (141781 - 10 * 171 * 82,6) / (292964 - 10 * 171^2) = 535 / 554 = 0,9657$$

$$a = 82,6 - 0,9657 * 171 = -82,5347$$

$$y = -82,5347 + 0,9657 * x$$

Schritt 2: Bestimmung der Güte des Regressionsmodells

$$\text{Term 1: } (141781 - 10 * 171 * 82,6) = 535$$

$$\text{Term 2: WURZEL } (292964 - 10 * 171^2) = \text{WURZEL } (554) = 23,5372$$

$$\text{Term 3: WURZEL } (68774 - 10 * 82,6^2) = \text{WURZEL } (546,4) = 23,3752$$

$$\text{BPK} = 535 / (23,5372 * 23,3752) = 9,501867$$

$$R^2 = 0,9732$$

Das Regressionsmodell erklärt einen Großteil der Streuung.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Das Gegenereignis zu mindestens 3 Augen beim einmaligen Wurf eines fairen Würfels lautet höchstens 2 Augen.

X richtig O falsch

2) Ein korrekt durchgeführter statistischer Test gestattet eine definitive Aussage über die Korrektheit von Null- und Alternativhypothese.

O richtig **X falsch**

3) Die Varianz $\text{Var}(X)$ kann ausschließlich für diskrete Zufallsvariablen berechnet werden.

O richtig **X falsch**

4) Bei einer stetigen Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

X richtig O falsch

5) Eine Stichprobe ist nur dann repräsentativ, wenn sie exakt den mit Hilfe der Cochran-Formel bestimmten Umfang aufweist.

O richtig **X falsch**

6) Eine Menge kann beliebig viele Teilmengen enthalten.

X richtig O falsch

7) Ist die Schnittmenge der Mengen A und B leer, so bedeutet dies, dass die Menge A die Komplementärmenge der Menge B (und umgekehrt) ist.

O richtig **X falsch**

8) Eine assoziative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

O richtig **X falsch**

9) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

O richtig **X falsch**

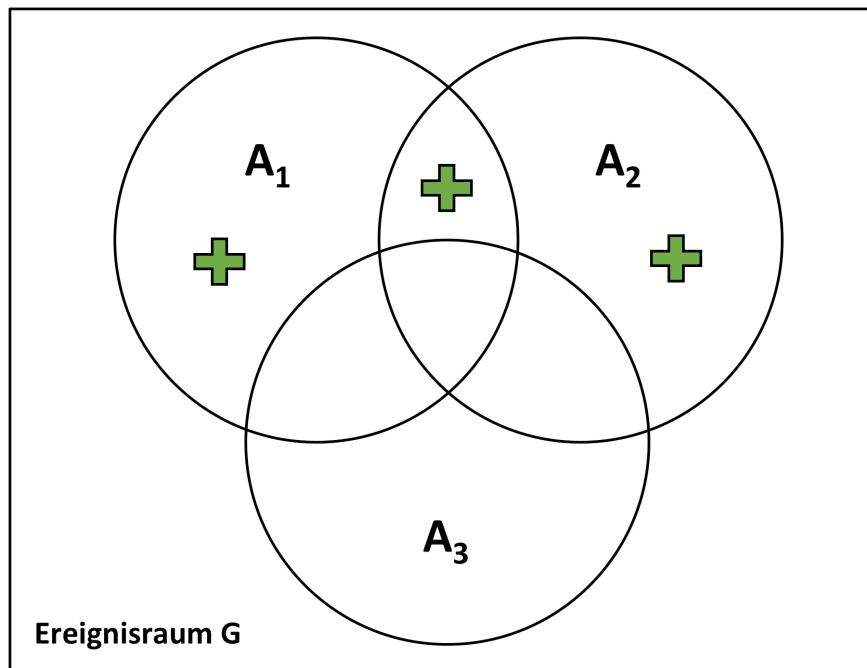
10) Elemente, die gleichzeitig der Menge A als auch der Menge B angehören, sind stets auch Bestandteil der Vereinigungsmenge der Mengen A und B.

X richtig O falsch

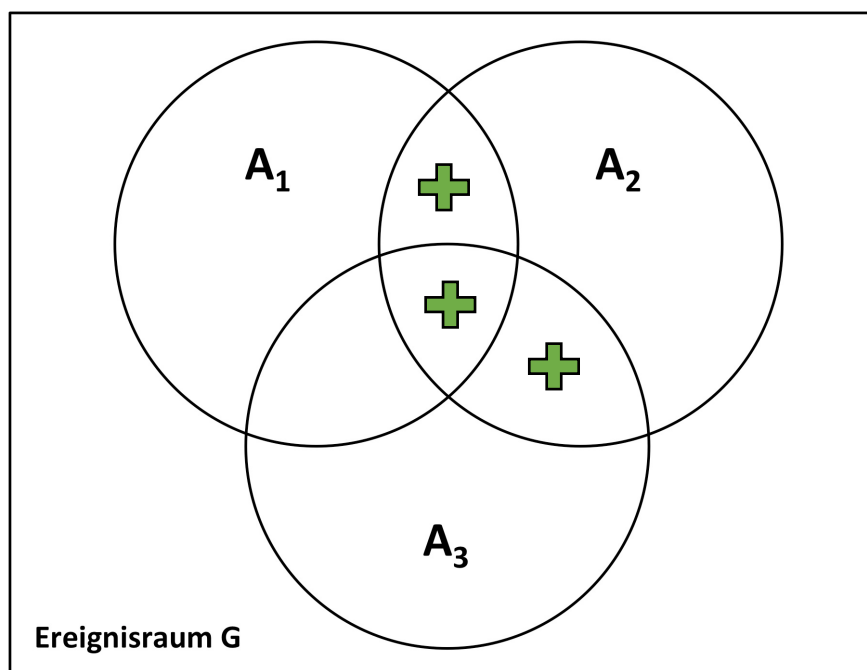
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

- a) $(A_1 \cup A_2) \cap \bar{A}_3$ (A_1 und A_2 komplett ohne alle Schnittflächen mit A_3)



- b) $(A_1 \cap A_2) \cup (A_2 \cap A_3)$ (Schnittmengen von A_1 und A_2 sowie A_2 und A_3)



Aufgabenteil VII: Kombinatorik (15 Punkte)

Aus den fünf Ziffern 1, 2, 3, 4 und 5 soll eine Liste mit allen möglichen fünfstelligen Zahlen gebildet werden, also z.B. 12345, 21435, 54321 etc. pp. Zu dieser Liste sind zwei Fragen zu beantworten.

- 1) Wie viele Zahlen enthält die Liste insgesamt? (10 Punkte)
- 2) Wie viele dieser Zahlen beginnen mit einer 3? (5 Punkte)

- 1) Die Liste enthält insgesamt $5! = 120$ Zahlen.
- 2) Von diesen beginnen $4! = 24$ mit einer 3.

Aufgabenteil VIII: Bedingte Wahrscheinlichkeiten (15 Punkte)

Eine Sicherheitssoftware für die Analyse von Überwachungsaufnahmen an einer Flughafen-Sicherheitsschleuse kann das Gesicht einer gesuchten Person mit einer Wahrscheinlichkeit von 92% erkennen. Allerdings identifiziert die Software in 3% aller Fälle eine nicht gesuchte Person irrtümlich als gesucht.

Die Sicherheitsbehörden gehen davon aus, dass an einem bestimmten Tag eine Gruppe von 10 gesuchten Personen versuchen wird, die Schleuse zu passieren. Das übliche Personenaufkommen pro Tag liegt bei 10.000 Fluggästen. Mit der Präsenz weiterer gesuchter Personen ist am betrachteten Tag nicht zu rechnen.

- a) Mit wie vielen fälschlicherweise als "gesucht" identifizierten Personen ist zu rechnen? (5 Punkte)
- b) Die Software schlägt Alarm. Wie groß ist die Wahrscheinlichkeit dafür, dass tatsächlich eine gesuchte Person entdeckt wurde? (10 Punkte)

- a) Anzahl der fälschlicherweise als „gesucht“ identifizierten Personen

$$10000 - 10 = 9990$$

$$9990 \cdot 0,03 = 299,7$$

- b) Wahrscheinlichkeit für eine gesuchte Person bei ausgelöstem Alarm.

Fälschlicherweise als gesucht identifizierte Personen: 299,7

Richtigerweise als gesucht identifizierte Personen: 9,2

Insgesamt als gesucht identifizierte Personen: 308,9

$$\text{Verhältnis: } 9,2 / 308,9 = 0,02978$$

Sommersemester 2017

Klausur Statistik I

Hinweis: Bei der Klausur Statistik I im Sommersemester 2017 handelte es sich um eine Prüfung, die – auf Wunsch der Studierenden – im Statistik-Labor unter Einsatz der Software PAST abgelegt wurde. Um sie zu replizieren, muss zunächst PAST installiert (<https://www.nhm.uio.no/english/research/re-sources/past/>) und anschließend der Datensatz aus der Musterlösung übertragen werden. Die Datei steht zudem (aber sicher nicht für immer) via https://www.hs-harz.de/user-mounts/322_m1036/dateien/Klausur_I_17.06.2017_PAST-Datei.dat zum Download zur Verfügung.

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2017

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 20. Mai 2017

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen. Dies gilt selbstverständlich nicht für den Rechner, an dem die Prüfung abgelegt wird.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: PC mit PAST
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars. Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2017

Christian Reinboth

Aufgabenteil I: Theorie (20 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (2 Punkte pro korrekter Beantwortung)

1) Der Regressionskoeffizient (b) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsenschnittpunkts der linearen Regressionsgeraden.

☐ richtig ☐ falsch

2) Die Standardabweichung berechnet sich als positive Wurzel aus der Varianz.

☐ richtig ☐ falsch

3) Ein Bravais-Pearson-Korrelationskoeffizient von $-0,85$ deutet auf eine schwache lineare Korrelation hin.

☐ richtig ☐ falsch

4) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig ☐ falsch

5) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig ☐ falsch

6) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☐ falsch

7) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

8) Kreisdiagramme eignen sich eher für diskrete als für stetige Daten.

☐ richtig ☐ falsch

9) Der statistische Ersatz fehlender Werte setzt mindestens metrisch skalierte Daten voraus.

☐ richtig ☐ falsch

10) Stetige Daten sollten vor der Erstellung von Säulendiagrammen klassiert werden.

☐ richtig ☐ falsch

Öffnen Sie zur Bearbeitung der nachfolgenden Aufgaben bitte die Datei mit der Bezeichnung „PAST-Datei zur Klausur vom 20.05.2017.dat“ in PAST.

Aufgabenteil II: Analyse eines Datensatzes mit PAST (80 Punkte)

- a) Bestimmen Sie – sofern möglich – für die sechs Variablen Gründungsjahr, Einwohner, Höhe, Sonnentage, Schneetage und Kfz-Kennzeichen jeweils einen geeigneten Mittelwert sowie ein geeignetes Maß für die Streuung. (12 Punkte)

Variable	Bezeichnung Mittelwert	Wert Mittelwert	Bezeichnung Streuungsmaß	Wert Streuungsmaß
Gründungsjahr				
Einwohner				
Höhe				
Sonnentage				
Schneetage				
Kfz-Kennzeichen				

- b) Lässt sich für die Variable „Gründungsjahr“ ein arithmetisches Mittel bestimmen? Und falls ja – wie ist es zu interpretieren? (8 Punkte)
- c) Stellen Sie für die gemeinsame Verteilung von Höhe (unabhängige Variable x) und Schneetagen (abhängige Variable y) die lineare Regressionsfunktion auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (20 Punkte)
- d) Welche Menge an Schneetagen würden Sie auf Basis des in c) erstellten Modells in einem Ort auf 223 m ü.N.N. erwarten? (5 Punkte) Welche Menge an Schneetagen würden Sie auf Basis des gleichen Modells in einem Ort auf 1.865 m ü.N.N. erwarten? (5 Punkte)
- e) Stellen Sie für die gemeinsame Verteilung von Höhe (unabhängige Variable x) und Sonnentagen (abhängige Variable y) die lineare Regressionsfunktion auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (10 Punkte)
- f) Vergleichen Sie die Güte der in c) und e) erstellten Regressionsmodelle. Worauf lässt sich der Güteunterschied zwischen beiden Modellen zurückführen? (10 Punkte)
- g) Wählen und berechnen Sie für die beiden Variablen Höhe und Sonnentage einen geeigneten Korrelationskoeffizienten und begründen Sie kurz Ihre Wahl. (10 Punkte)

KLAUSUR STATISTIK I MIT LÖSUNGEN

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2017

Christian Reinboth

Aufgabenteil I: Theorie (20 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (2 Punkte pro korrekter Beantwortung)

1) Der Regressionskoeffizient (b) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsenschnittpunkts der linearen Regressionsgeraden.

☐ richtig ☒ **falsch**

2) Die Standardabweichung berechnet sich als positive Wurzel aus der Varianz.

☒ **richtig** ☐ falsch

3) Ein Bravais-Pearson-Korrelationskoeffizient von -0,85 deutet auf eine schwache lineare Korrelation hin.

☐ richtig ☒ **falsch**

4) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☒ **richtig** ☐ falsch

5) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig ☒ **falsch**

6) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☒ **falsch**

7) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☒ **falsch**

8) Kreisdiagramme eignen sich eher für diskrete als für stetige Daten.

☒ **richtig** ☐ falsch

9) Der statistische Ersatz fehlender Werte setzt mindestens metrisch skalierte Daten voraus.

☒ **richtig** ☐ falsch

10) Stetige Daten sollten vor der Erstellung von Säulendiagrammen klassiert werden.

☒ **richtig** ☐ falsch

Öffnen Sie zur Bearbeitung der nachfolgenden Aufgaben bitte die Datei mit der Bezeichnung „PAST-Datei zur Klausur vom 20.05.2017.dat“ in PAST.

	Gründungsjahr	Einwohner (2015)	Höhe in m ü.N.N.	Sonnentage (2015)	Schneetage (2015)	Kfz-Kennzeichen	G	H
Wernigerode	1127	33108	240	243	63	HZ		
Halberstadt	804	43768	122	210	34	HZ		
Quedlinburg	922	24555	123	208	34	HZ		
Nordhausen	780	42217	208	245	51	NDH		
Goslar	979	50782	255	262	65	GS		
Bad Harzburg	916	21735	261	251	67	GS		
Thale	840	17639	156	217	39	HZ		
Merseburg	881	34052	88	125	29	SK		
Bernburg	782	33920	85	123	27	SLK		
Sangerhausen	802	27752	154	246	38	MSH		
Staßfurt	806	26634	73	117	23	SLK		
Eisleben	994	24198	114	221	33	MSH		
Stendal	1160	40269	32	321	11	SDL		
Weißenfels	899	40389	100	129	32	BLK		
Blankenburg	1123	20294	288	253	68	HZ		
16	•							
17	•							
18	•							
19	•							
20	•							

Aufgabenteil II: Analyse eines Datensatzes mit PAST (80 Punkte)

- a) Bestimmen Sie – sofern möglich – für die sechs Variablen Gründungsjahr, Einwohner, Höhe, Sonnentage, Schneetage und Kfz-Kennzeichen jeweils einen geeigneten Mittelwert sowie ein geeignetes Maß für die Streuung. (12 Punkte)

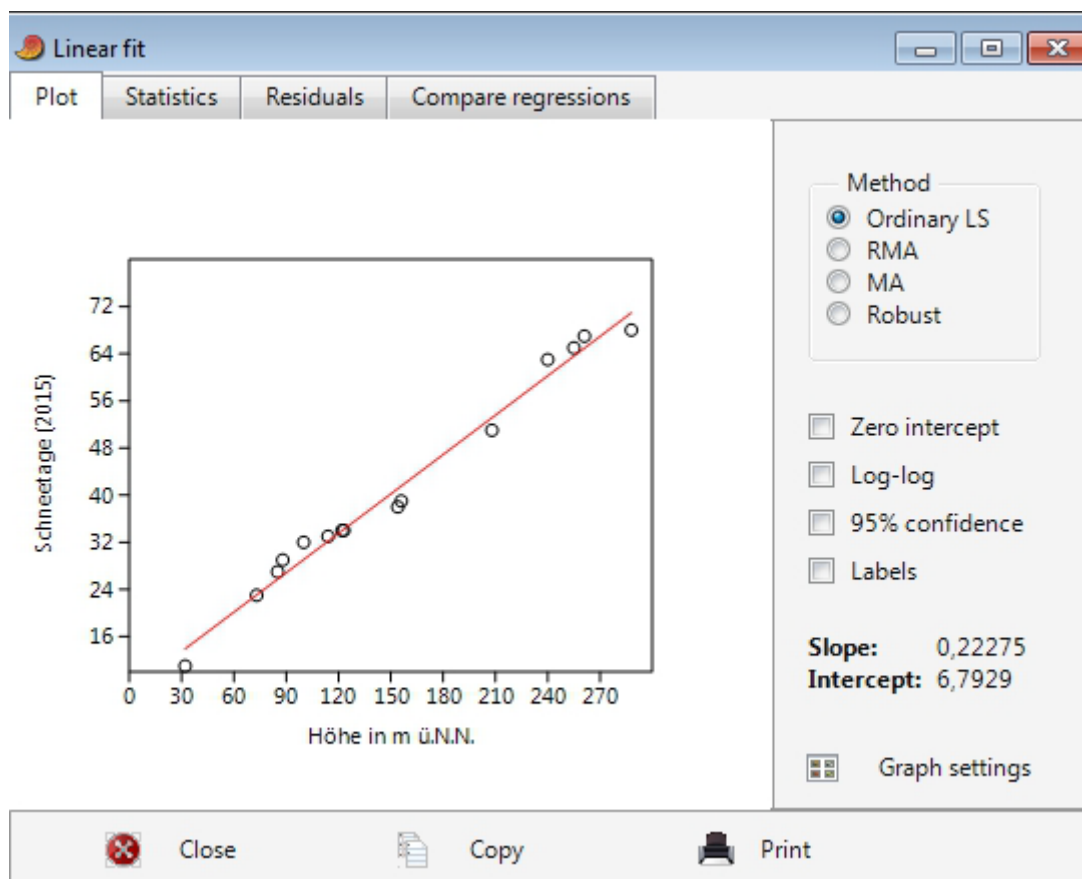
	Gründungsjahr	Einwohner (2015)	Höhe in m ü.N.N.	Sonnentage (2015)	Schneetage (2015)
N	15	15	15	15	15
Min	780	17639	32	117	11
Max	1160	50782	288	321	68
Sum	13815	481312	2299	3171	614
Mean	921	32087,47	153,2667	211,4	40,93333
Std. error	33,66516	2540,319	20,35131	15,78329	4,569533
Variance	17000,14	9,679827E07	6212,638	3736,686	313,2095
Stand. dev	130,3846	9838,611	78,82029	61,12844	17,69773
Median	899	33108	123	221	34
25 prcntil	804	24198	88	129	29
75 prcntil	994	40389	240	251	63
Skewness	0,7554623	0,2918871	0,4037742	-0,3768339	0,3619291
Kurtosis	-0,6720077	-0,9160178	-1,073643	-0,6028782	-0,900996
Geom. mean	912,8006	30665,59	132,5773	201,9926	37,05533
Coeff. var	14,15685	30,66185	51,4269	28,91601	43,23549

Variable	Bezeichnung Mittelwert	Wert Mittelwert	Bezeichnung Streuungsmaß	Wert Streuungsmaß
Gründungsjahr	nicht sinnvoll	//	nicht sinnvoll	//
Einwohner	arithm. Mittel	32.087	Standardabw.	9.838
Höhe	arithm. Mittel	153,26	Standardabw.	78,82
Sonnentage	arithm. Mittel	211,40	Standardabw.	61,13
Schneetage	arithm. Mittel	40,94	Standardabw.	17,69
Kfz-Kennzeichen	Modus	HZ	nicht möglich	//

- b) Lässt sich für die Variable „Gründungsjahr“ ein arithmetisches Mittel bestimmen? Und falls ja – wie ist es zu interpretieren? (8 Punkte)

Das arithmetische Mittel darf bestimmt werden, lässt sich aber als singulärer Wert („durchschnittliches Gründungsjahr“) nicht sinnvoll interpretieren. Möglich (wenn auch nicht unbedingt sehr viel sinnvoller) wäre aber z.B. die Subtraktion vom aktuellen Jahr („durchschnittliches Alter“).

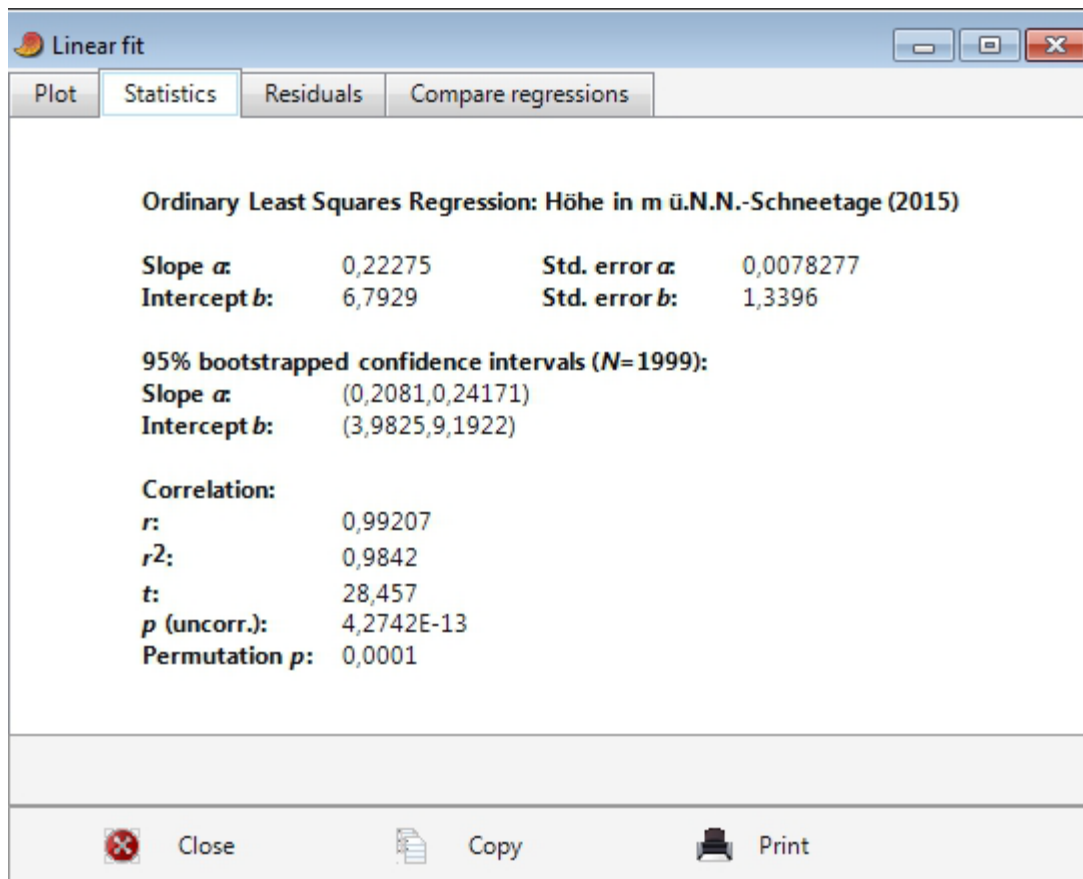
- c) Stellen Sie für die gemeinsame Verteilung von Höhe (unabhängige Variable x) und Schneetagen (abhängige Variable y) die lineare Regressionsfunktion auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (20 Punkte)



$$Y = 6,7929 + 0,22275 \cdot X$$

$$\text{Schneetage} = 6,7929 + 0,22275 \cdot \text{Höhe}$$

$R^2 = 0,9842 \rightarrow$ Das Modell erklärt 98,42% der Streuung, ist also äußerst aussagekräftig.



- d) Welche Menge an Schneetagen würden Sie auf Basis des in c) erstellten Modells in einem Ort auf 223 m ü.N.N. erwarten? (5 Punkte) Welche Menge an Schneetagen würden Sie auf Basis des gleichen Modells in einem Ort auf 1.865 m ü.N.N. erwarten? (5 Punkte)

$$\text{Schneetage} = 6,7929 + 0,22275 \cdot \text{Höhe}$$

$$6,7929 + 0,22275 \cdot 223 = 56,46615$$

In einem Ort auf 223 m ü.N.N. wären ~ 56 Schneetage zu erwarten.

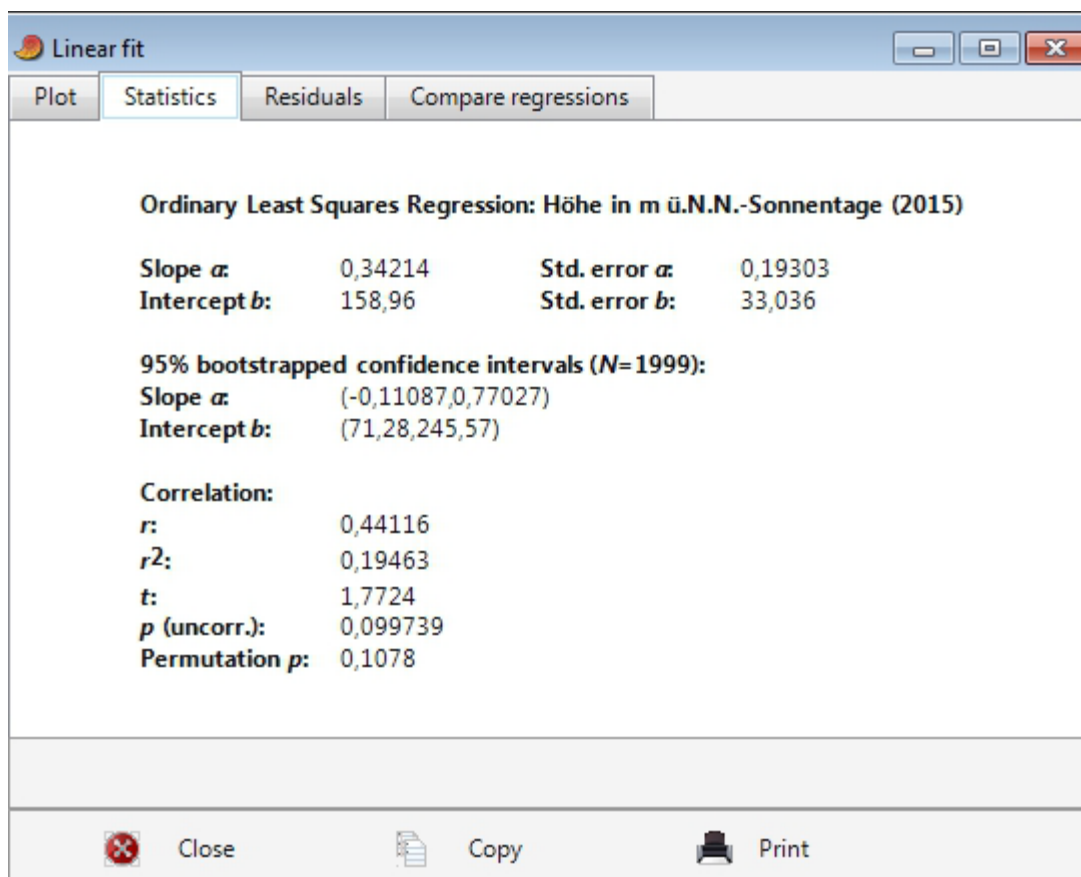
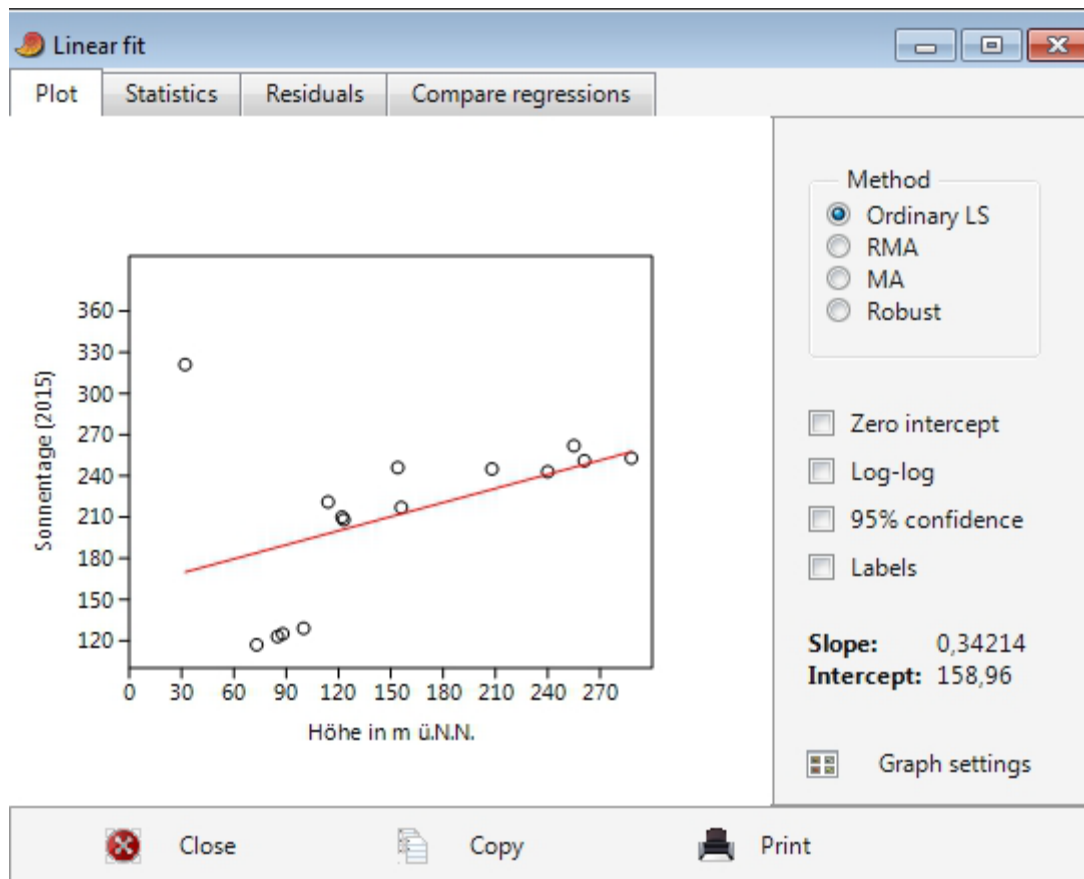
Für einen Ort auf 1.865 m ü.N.N. sollte auf Basis des vorliegenden Modells keine Prognose angestrengt werden, da die dem zugrundeliegenden Daten aus Kommunen zwischen 32 m ü.N.N. und 288 m ü.N.N. stammen. Dieser Wertebereich ist von 1.865 m ü.N.N. zu weit entfernt, um noch von einer Gültigkeit des linearen Zusammenhangs ausgehen zu können.

- e) Stellen Sie für die gemeinsame Verteilung von Höhe (unabhängige Variable x) und Sonnentagen (abhängige Variable y) die lineare Regressionsfunktion auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (10 Punkte)

$$Y = 158,96 + 0,34214 \cdot X$$

$$\text{Sonnentage} = 158,96 + 0,34214 \cdot \text{Höhe}$$

$R^2 = 0,19463 \rightarrow$ Das Modell erklärt nur 19,46% der Streuung und ist somit kaum aussagekräftig.

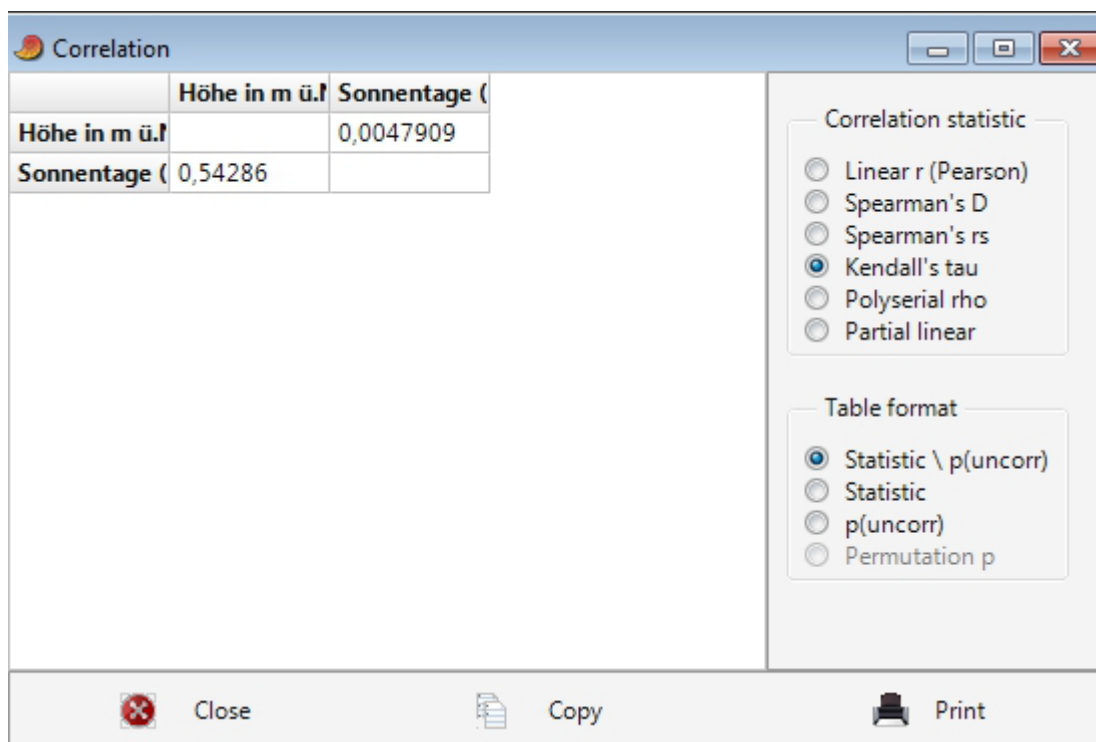


- f) Vergleichen Sie die Güte der in c) und e) erstellten Regressionsmodelle. Worauf lässt sich der Güteunterschied zwischen beiden Modellen zurückführen? (10 Punkte)

Ein wesentlicher Faktor dürfte der im Streudiagramm besonders auffällige Wert für Stendal sein, der erheblich zum schlechten Fit der Regressionsgleichung beitragen dürfte. Abgesehen davon ist erkennbar, dass ein monotoner Zusammenhang klar vorliegt, ein durchgehend linearer aber nicht vorzuliegen scheint (sondern vielmehr zwei teillineare Strecken sowie der bereits erwähnte aus der Reihe fallende Einzelwert).

- g) Wählen und berechnen Sie für die beiden Variablen Höhe und Sonnentage einen geeigneten Korrelationskoeffizienten und begründen Sie kurz Ihre Wahl. (10 Punkte)

Da angesichts der Ergebnisse aus f) eher nicht von einer linearen, wohl aber von einer monotonen Korrelation ausgegangen werden kann, empfiehlt sich an dieser Stelle der Rangkorrelationskoeffizient nach Spearman (0,56071) oder der Konkordanzkoeffizient nach Kendall (0,54286). Beide deuten auf eine mittelstarke monotone Korrelation hin, die nach Entfernung des Wertes für Stendal sicher noch deutlicher erkennbar wäre.



Sommersemester 2017

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2017

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 27. August 2017

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2017

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☐ falsch

2) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☐ falsch

3) Der Momentenkoeffizient der Schiefe lässt sich lediglich bei unimodalen Verteilungen interpretieren.

☐ richtig ☐ falsch

4) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☐ richtig ☐ falsch

5) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

☐ richtig ☐ falsch

6) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

7) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

☐ richtig ☐ falsch

8) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

☐ richtig ☐ falsch

9) Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus.

☐ richtig ☐ falsch

10) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die nachfolgende Tabelle gibt die „Ärztedichte“ (die Anzahl der Einwohner/innen je berufstätigem Arzt / je berufstätiger Ärztin – unabhängig von dessen / deren Fachgebiet) für alle 16 Bundesländer wieder. (Quelle: Bundesärztekammer)

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

Erstellen Sie ein Stem-and-Leaf-Diagramm mit einer Stammbreite von 100 (5 Punkte)
sowie ein Stem-and-Leaf-Diagramm mit einer Stammbreite von 10 (5 Punkte).

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Bestimmen Sie für die bereits bekannten Daten zur „Ärztendichte“...

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

- k) das arithmetische Mittel,
- l) den Modus,
- m) den Interquartilsabstand,
- n) die Varianz
- o) und die Standardabweichung. (jeweils 3 Punkte)

Aufgabenteil IV: Lineare Regressionsanalyse (15 Punkte)

Stellen Sie für die Verteilung von Körpergröße (unabhängige Variable x) und Körpergewicht (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte).

Person Nr.	Körpergröße (in cm)	Körpergewicht (in kg)
1	175	85
2	178	103
3	184	83
4	162	76
5	173	81
6	163	83
7	181	62
8	174	75
9	174	75
10	176	85

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

2) Um mit der Laplace-Wahrscheinlichkeit rechnen zu können, muss die Anzahl aller günstigen sowie die Anzahl aller möglichen Elementarereignisse bekannt sein.

☐ richtig ☐ falsch

3) Das Konfidenzniveau eines Konfidenzintervalls muss immer vor der Aufstellung des Intervalls festgelegt und darf keinesfalls nachträglich an das Ergebnis angepasst werden.

☐ richtig ☐ falsch

4) Mit dem Satz von Bayes kann nur gerechnet werden, wenn metrisch skalierte Merkmale vorliegen.

☐ richtig ☐ falsch

5) Bei einer normalverteilten Variablen ist der Modus stets kleiner als der Erwartungswert.

☐ richtig ☐ falsch

6) Die Wahrscheinlichkeit für fehlerhafte Testergebnisse nimmt mit der Anzahl durchgeführter Tests zu. Dieses Phänomen wird auch als alpha-Fehlerinflation bezeichnet.

☐ richtig ☐ falsch

7) Ist die Schnittmenge der Mengen A und B leer, so bedeutet dies, dass die Menge A die Komplementärmenge der Menge B (und umgekehrt) ist.

☐ richtig ☐ falsch

8) Wird eine inkorrekte Nullhypothese im Rahmen eines statistischen Tests nicht verworfen, so handelt es sich um einen Fehler 2. Art.

☐ richtig ☐ falsch

9) Bei einer diskreten Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

☐ richtig ☐ falsch

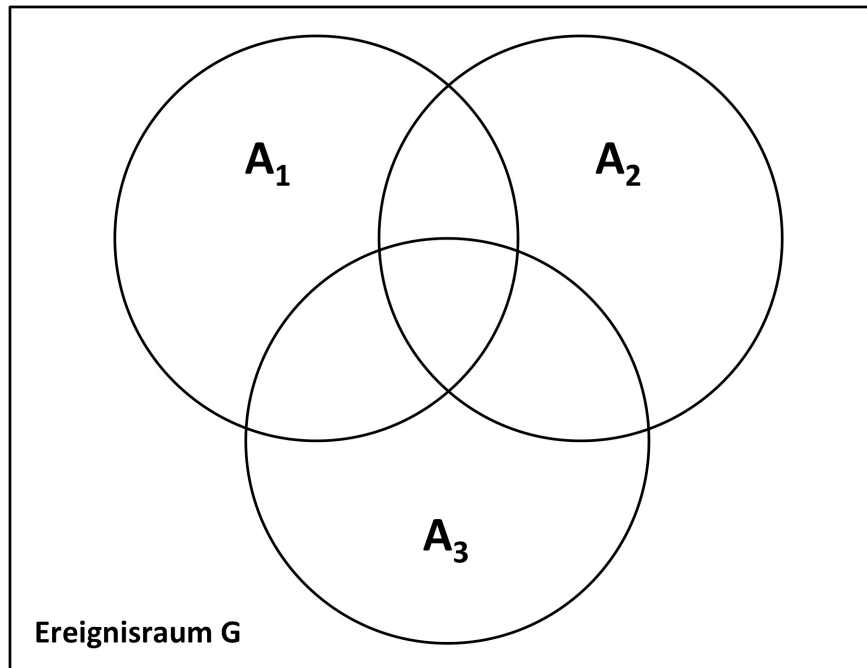
10) Die Varianz $\text{Var}(X)$ kann ausschließlich für diskrete Zufallsvariablen berechnet werden.

☐ richtig ☐ falsch

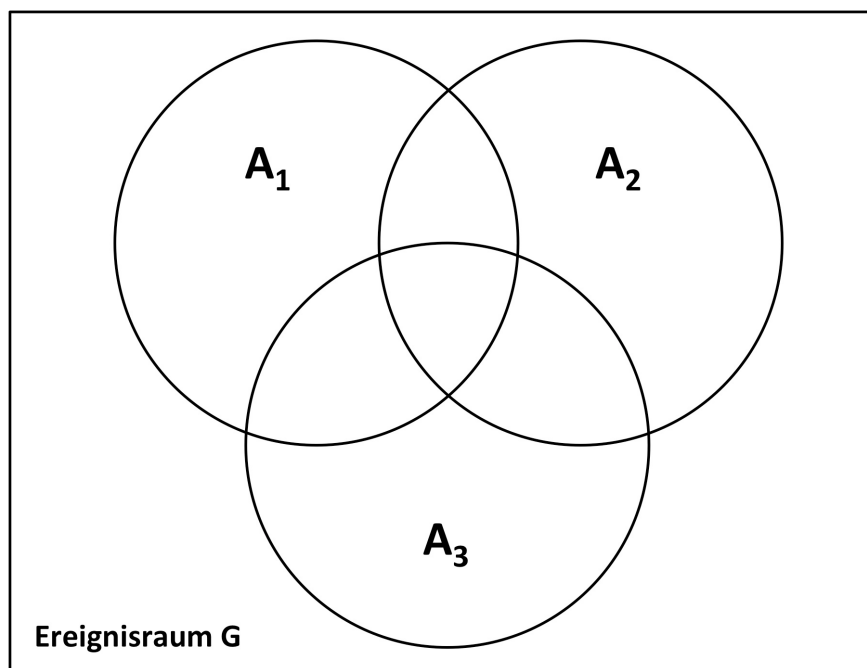
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

a) $(A_1 \cap A_2 \cap A_3)$



b) $(A_1 \cup A_1)$



Aufgabenteil VII: Kombinatorik (15 Punkte)

- a) In der Fußball-Kreisliga Südharz spielen derzeit 10 Mannschaften. In einem Turnier sollen alle Mannschaften im Spielmodus "jeder gegen jeden" gegeneinander antreten. Wie viele Spiele müssen insgesamt stattfinden? (10 Punkte)
- b) Eine bei einem ambulanten Pflegedienst beschäftigte Altenpflegekraft muss im Rahmen einer Schicht 9 verschiedene Patienten/innen aufsuchen, die über das Stadtgebiet von Halberstadt verteilt wohnen. Wie viele verschiedene Tourenpläne lassen sich erstellen? (5 Punkte)

Aufgabenteil VIII: Bedingte Wahrscheinlichkeiten (15 Punkte)

In der Telefonzentrale eines Großflughafens gehen in einem durchschnittlichen Monat 100 Bombendrohungen ein, die jedoch fast immer auf Scherzanrufer zurückzuführen sind. Zu einem Drohanruf, dem eine reale Bedrohung zugrunde liegt, kommt es lediglich einmal im Jahr. Um die Zahl der durch Fehlalarme ausgelösten Störungen im Betriebsablauf zu minimieren, investiert der Flughafen in eine Stimmanalyse-Software, die reale Bedrohungen mit einer Sicherheit von 98% korrekt erkennt. Allerdings ordnet die Software auch 5% aller harmlosen Anrufe fälschlicherweise als kritisch ein.

Wie groß ist die Wahrscheinlichkeit dafür, dass ein durch die Software als kritisch eingeordneter Anruf auch tatsächlich auf eine reale Bedrohungslage hinweist?

KLAUSUR STATISTIK I und II MIT LÖSUNGEN

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2017

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

O richtig **X falsch**

2) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

O richtig **X falsch**

3) Der Momentenkoeffizient der Schiefe lässt sich lediglich bei unimodalen Verteilungen interpretieren.

X richtig O falsch

4) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

X richtig O falsch

5) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

O richtig **X falsch**

6) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

O richtig **X falsch**

7) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

X richtig O falsch

8) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

X richtig O falsch

9) Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus.

O richtig **X falsch**

10) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

X richtig O falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die nachfolgende Tabelle gibt die „Ärztedichte“ (die Anzahl der Einwohner/innen je berufstätigem Arzt / je berufstätiger Ärztin – unabhängig von dessen / deren Fachgebiet) für alle 16 Bundesländer wieder. (Quelle: Bundesärztekammer)

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

Erstellen Sie ein Stem-and-Leaf-Diagramm mit einer Stammbreite von 100 (5 Punkte)
sowie ein Stem-and-Leaf-Diagramm mit einer Stammbreite von 10 (5 Punkte).

Stammbreite 100

1 | 4 6 6
2 | 0 0 1 2 2 2 2 2 3 4 4 4 5

Keine Ausreißer
Jedes Blatt ein Fall

Stammbreite 10

14 | 1
15 |
16 | 1 9
17 |
18 |
19 |
20 | 5 8
21 | 9
22 | 3 3 5 6 9
23 | 5
24 | 0 9 9
25 | 9

Keine Ausreißer
Jedes Blatt ein Fall

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Bestimmen Sie für die bereits bekannten Daten zur „Ärztendichte“...

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)
1	Brandenburg	259
2	Niedersachsen	249
3	Sachsen-Anhalt	249
4	Sachsen	240
5	Thüringen	235
6	Rheinland-Pfalz	229
7	Schleswig-Holstein	226
8	Baden-Württemberg	225
9	Hessen	223
10	Nordrhein-Westfalen	223
11	Mecklenburg-Vorpommern	219
12	Bayern	208
13	Saarland	205
14	Bremen	169
15	Berlin	161
16	Hamburg	141

- p) das arithmetische Mittel,
- q) den Modus,
- r) den Interquartilsabstand,
- s) die Varianz
- t) und die Standardabweichung. (jeweils 3 Punkte)

Arithmetisches Mittel

Summe: 3.461 $\rightarrow 3.461 / 16 = 216,31$

Modus

Nicht berechenbar, da keine unimodale Verteilung vorliegt.

Interquartilsabstand

Oberes Quartil: $0,75 \cdot 16 = 12 \rightarrow (12. + 13. \text{ Wert}) / 2 = (235 + 240) / 2 = 237,5$

Unteres Quartil: $0,25 \cdot 16 = 4 \rightarrow (4. + 5. \text{ Wert}) / 2 = (205 + 208) / 2 = 206,5$

Interquartilsabstand: $237,5 - 206,5 = 31$

Varianz

Arithmetisches Mittel: 216,31

Nr.	Bundesland	Einwohner je berufstätigem Arzt (2015)	$(x - 216,31)^2$
1	Brandenburg	259	1822,44
2	Niedersachsen	249	1068,64
3	Sachsen-Anhalt	249	1068,64
4	Sachsen	240	561,22
5	Thüringen	235	349,32
6	Rheinland-Pfalz	229	161,04
7	Schleswig-Holstein	226	93,90
8	Baden-Württemberg	225	75,52
9	Hessen	223	44,76
10	Nordrhein-Westfalen	223	44,76
11	Mecklenburg-Vorpommern	219	7,24
12	Bayern	208	69,06
13	Saarland	205	127,92
14	Bremen	169	2238,24
15	Berlin	161	3059,20
16	Hamburg	141	5671,60

Summe: 16.463,44

 $16.463,44 / 16 = 1.028,96$ Einwohner²Standardabweichung

Die Standardabweichung ist die positive Wurzel der Varianz: 32,07 Einwohner

Aufgabenteil IV: Lineare Regressionsanalyse (15 Punkte)

Stellen Sie für die Verteilung von Körpergröße (unabhängige Variable x) und Körpergewicht (abhängige Variable y) die lineare Regressionsfunktion (10 Punkte) auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 . (5 Punkte).

Nr.	x	y	x^2	y^2	(x*y)
1	175	85	30625	7225	14875
2	178	103	31684	10609	18334
3	184	83	33856	6889	15272
4	162	76	26244	5776	12312
5	173	81	29929	6561	14013
6	163	83	26569	6889	13529
7	181	62	32761	3844	11222
8	174	75	30276	5625	13050
9	174	75	30276	5625	13050
10	176	85	30976	7225	14960
SUM	1740	808	303196	66268	140617
MITTEL	174	80,8	//	//	//

$$b = (140617 - 10 * 174 * 80,8) / (303196 - 10 * 174^2) = 25 / 436 = 0,057$$

$$a = 80,8 - 0,057 * 174 = 70,882$$

$$y = 70,882 + 0,057 * x$$

Schritt 2: Bestimmung der Güte des Regressionsmodells

$$\text{Term 1: } (140617 - 10 * 174 * 80,8) = 25$$

$$\text{Term 2: WURZEL } (303196 - 10 * 174^2) = \text{WURZEL } (436) = 20,88$$

$$\text{Term 3: WURZEL } (66268 - 10 * 80,8^2) = \text{WURZEL } (981,6) = 31,33$$

$$\text{BPK} = 25 / (20,88 * 31,33) = 0,038$$

$$R^2 = 0,001444$$

Das Regressionsmodell erklärt die Streuung überhaupt nicht.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

X richtig ☐ falsch

2) Um mit der Laplace-Wahrscheinlichkeit rechnen zu können, muss die Anzahl aller günstigen sowie die Anzahl aller möglichen Elementarereignisse bekannt sein.

X richtig ☐ falsch

3) Das Konfidenzniveau eines Konfidenzintervalls muss immer vor der Aufstellung des Intervalls festgelegt und darf keinesfalls nachträglich an das Ergebnis angepasst werden.

X richtig ☐ falsch

4) Mit dem Satz von Bayes kann nur gerechnet werden, wenn metrisch skalierte Merkmale vorliegen.

☐ richtig **X falsch**

5) Bei einer normalverteilten Variablen ist der Modus stets kleiner als der Erwartungswert.

☐ richtig **X falsch**

6) Die Wahrscheinlichkeit für fehlerhafte Testergebnisse nimmt mit der Anzahl durchgeführter Tests zu. Dieses Phänomen wird auch als alpha-Fehlerinflation bezeichnet.

X richtig ☐ falsch

7) Ist die Schnittmenge der Mengen A und B leer, so bedeutet dies, dass die Menge A die Komplementärmenge der Menge B (und umgekehrt) ist.

☐ richtig **X falsch**

8) Wird eine inkorrekte Nullhypothese im Rahmen eines statistischen Tests nicht verworfen, so handelt es sich um einen Fehler 2. Art.

X richtig ☐ falsch

9) Bei einer diskreten Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

☐ richtig **X falsch**

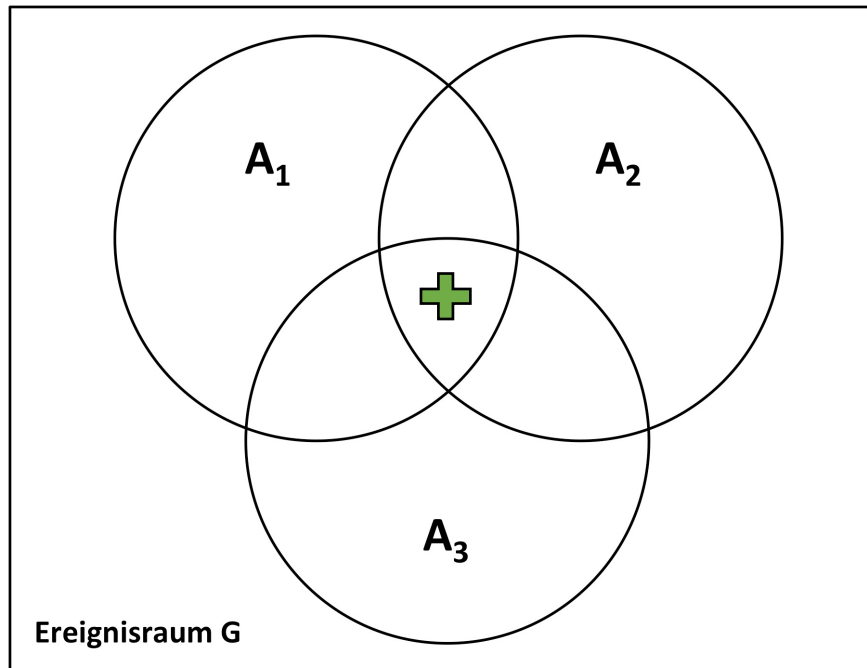
10) Die Varianz $\text{Var}(X)$ kann ausschließlich für diskrete Zufallsvariablen berechnet werden.

☐ richtig **X falsch**

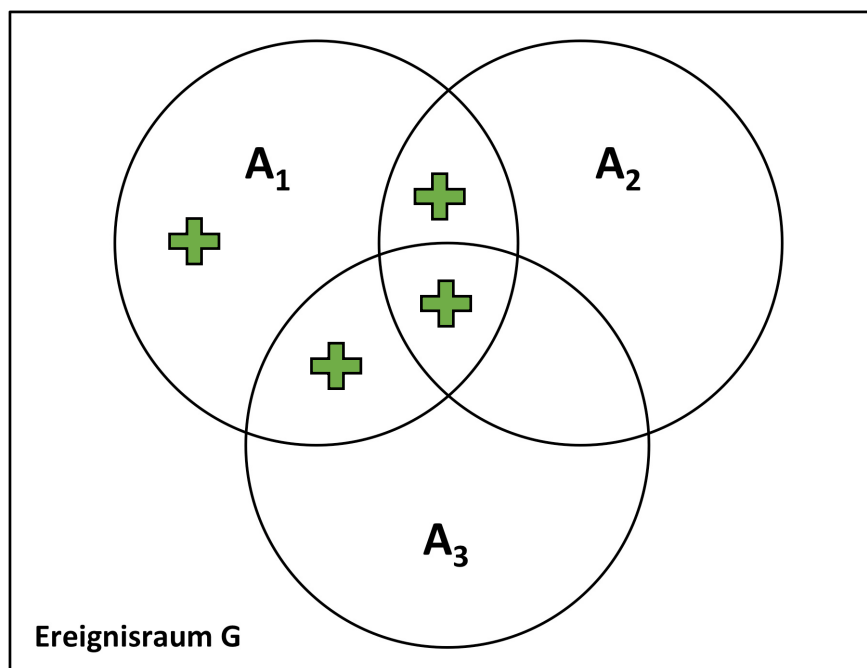
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

- a) $(A_1 \cap A_2 \cap A_3)$ (Schnittmenge aller drei Kreise)



- b) $(A_1 \cup A_1)$ (nur Kreis A_1)



Aufgabenteil VII: Kombinatorik (15 Punkte)

a) In der Fußball-Kreisliga Südharz spielen derzeit 10 Mannschaften. In einem Turnier sollen alle Mannschaften im Spielmodus "jeder gegen jeden" gegeneinander antreten. Wie viele Spiele müssen insgesamt stattfinden? (10 Punkte)

Es handelt sich um eine Kombination (die Reihenfolge des Ziehens spielt keine Rolle - das Spiel Mannschaft A gegen Mannschaft B ist identisch mit dem Spiel Mannschaft B gegen Mannschaft A) im Modell ohne Zurücklegen (keine Mannschaft kann im gleichen Pairing doppelt gezogen werden - sonst müsste sie gegen sich selbst antreten).

$$n! / (k! * (n-k)!)$$

$$n = 10$$

$$k = 2$$

$$10! / (2! * 8!) = 45$$

Es müssen insgesamt 45 Spiele stattfinden, bevor jede Mannschaft einmal gegen jede Mannschaft gespielt hat.

b) Eine bei einem ambulanten Pflegedienst beschäftigte Altenpflegekraft muss im Rahmen einer Schicht 9 verschiedene Patienten/innen aufsuchen, die über das Stadtgebiet von Halberstadt verteilt wohnen. Wie viele verschiedene Tourenpläne lassen sich erstellen? (5 Punkte)

Hier handelt sich um eine Variation (die Reihenfolge der Ereignisse spielt eine Rolle, da nach verschiedenen Tourenplänen gefragt wurde) im Modell ohne Zurücklegen (jeder Patient / jede Patientin wird nur einmal besucht) sowie den Sonderfall der Permutation (alle Patienten/innen werden aufgesucht), daher gilt:

$$n! = 9! = 362.880$$

Aufgabenteil VIII: Bedingte Wahrscheinlichkeiten (15 Punkte)

In der Telefonzentrale eines Großflughafens gehen in einem durchschnittlichen Monat 100 Bombendrohungen ein, die jedoch fast immer auf Scherzanrufer zurückzuführen sind. Zu einem Drohanruf, dem eine reale Bedrohung zugrunde liegt, kommt es lediglich einmal im Jahr. Um die Zahl der durch Fehlalarme ausgelösten Störungen im Betriebsablauf zu minimieren, investiert der Flughafen in eine Stimmanalyse-Software, die reale Bedrohungen mit einer Sicherheit von 98% korrekt erkennt. Allerdings ordnet die Software auch 5% aller harmlosen Anrufe fälschlicherweise als kritisch ein.

Wie groß ist die Wahrscheinlichkeit dafür, dass ein durch die Software als kritisch eingeordneter Anruf auch tatsächlich auf eine reale Bedrohungslage hinweist?

- In einem Jahr kommt es insgesamt zu $100 \cdot 12 = 1.200$ Anrufen.
- Davon sind 1.199 Anrufe keine Bedrohung, 1 Anruf ist eine Bedrohung.
- Dieser 1 gefährliche Anruf wird mit 98%iger Sicherheit auch enttarnt = 0,98.
- Von 1.199 harmlosen Anrufen werden 5% fälschlicherweise geflaggt = 59,95.
- Insgesamt werden also $0,98 + 59,95 = 60,93$ Anrufe als verdächtig eingestuft.
- Von diesen 60,93 Anrufen sind aber nur 0,98 wirklich kritisch = 0,016.
- Die Wahrscheinlichkeit liegt somit bei lediglich 1,6%.

Sommersemester 2018

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2018

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 23. Juni 2018

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2018

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für exponentielle Zusammenhänge.

☐ richtig ☐ falsch

2) Die Fünf-Werte-Zusammenfassung besteht aus den Quartilen, Minimum und Maximum.

☐ richtig ☐ falsch

3) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☐ falsch

4) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

☐ richtig ☐ falsch

5) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

6) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig ☐ falsch

7) Der Quartilkoeffizient der Schiefe ist unempfindlich gegenüber Ausreißern.

☐ richtig ☐ falsch

8) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☐ richtig ☐ falsch

9) Kreisdiagramme eignen sich eher für diskrete als für stetige Daten.

☐ richtig ☐ falsch

10) Der Modus lässt sich nur bestimmen, wenn eine unimodale Verteilung vorliegt.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Der Pegelstand des Harzer Gebirgsflusses Holtemme wird an zehn Messpunkten automatisch erhoben. Die folgende Tabelle zeigt die Pegelstände an diesen Messpunkten am Nachmittag des 23.05.2018.

Pegel	Pegelstand
01 – Steinerne Renne	172 cm
02 – Wernigerode – Hasserode	178 cm
03 – Wernigerode – Stadtmitte	183 cm
04 – Wernigerode – Minsleben	181 cm
05 – Wernigerode – Silstedt	168 cm
06 – Derenburg – Ortsmitte	163 cm
07 – Halberstadt – Mahndorf	172 cm
08 – Halberstadt – Poetengang	173 cm
09 – Halberstadt – Emersleben	178 cm
10 – Halberstadt - Nienhagen	182 cm

Konstruieren Sie für die Verteilung der Pegelstände ein Stamm-Blatt-Diagramm mit a) einer Stammbreite von 100 (10 Punkte) und b) einer Stammbreite von 10 (10 Punkte).

Aufgabenteil III: Verteilungsparameter (30 Punkte)

In einem Automotive-Zulieferbetrieb werden täglich mehrere zehntausend Spitzgussteile hergestellt. Eine Zufallsstichprobe bestehend aus zehn Bauteilen wird vermessen und gewogen.

Bauteil-Nr.	Gewicht in g
01	26
02	27
03	23
04	26
05	24
06	24
07	78
08	29
09	21
10	22

Bestimmen Sie die folgenden Größen:

- Arithmetisches Mittel (5 Punkte)
- Um 20% getrimmtes arithmetisches Mittel (5 Punkte)
- Median (5 Punkte)
- Interquartilsabstand (5 Punkte)
- Varianz (5 Punkte)
- Standardabweichung (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

In einer Rinderzucht werden zehn ausgewachsene Tiere verschiedenen Alters gewogen.

Rind-Nr.	Gewicht in kg	Alter in Jahren
01	563	15
02	701	21
03	523	13
04	521	12
05	618	16
06	633	18
07	571	16
08	671	20
09	710	22
10	512	10

- a) Berechnen Sie den Rangkorrelationskoeffizienten nach Spearman. (10 Punkte)
 b) Berechnen Sie den Bravais-Pearson-Korrelationskoeffizienten. (10 Punkte)

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Rind-Nr.	Gewicht in kg	Alter in Jahren
01	563	15
02	701	21
03	523	13
04	521	12
05	618	16
06	633	18
07	571	16
08	671	20
09	710	22
10	512	10

Stellen Sie für die vorliegenden Daten aus der Rinderzucht eine lineare Regressionsfunktion mit dem Alter der Tiere in Jahren als unabhängiger Variable (x) und ihrem Gewicht in kg als abhängiger Variable (y) auf (10 Punkte). Bestimmen Sie anschließend die Güte des Regressionsmodells und interpretieren Sie das Ergebnis (10 Punkte).

KLAUSUR STATISTIK I MIT LÖSUNGEN

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2018

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für exponentielle Zusammenhänge.

☐ richtig ☒ **falsch**

2) Die Fünf-Werte-Zusammenfassung besteht aus den Quartilen, Minimum und Maximum.

☒ **richtig** ☐ falsch

3) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☒ **falsch**

4) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

☒ **richtig** ☐ falsch

5) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☒ **falsch**

6) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig ☒ **falsch**

7) Der Quartilkoeffizient der Schiefe ist unempfindlich gegenüber Ausreißern.

☒ **richtig** ☐ falsch

8) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☒ **richtig** ☐ falsch

9) Kreisdiagramme eignen sich eher für diskrete als für stetige Daten.

☒ **richtig** ☐ falsch

10) Der Modus lässt sich nur bestimmen, wenn eine unimodale Verteilung vorliegt.

☒ **richtig** ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Der Pegelstand des Harzer Gebirgsflusses Holtemme wird an zehn Messpunkten automatisch erhoben. Die folgende Tabelle zeigt die Pegelstände an diesen Messpunkten am Nachmittag des 23.05.2018.

Pegel	Pegelstand
01 – Steinerne Renne	172 cm
02 – Wernigerode – Hasserode	178 cm
03 – Wernigerode – Stadtmitte	183 cm
04 – Wernigerode – Minsleben	181 cm
05 – Wernigerode – Silstedt	168 cm
06 – Derenburg – Ortsmitte	163 cm
07 – Halberstadt – Mahndorf	172 cm
08 – Halberstadt – Poetengang	173 cm
09 – Halberstadt – Emersleben	178 cm
10 – Halberstadt - Nienhagen	182 cm

Konstruieren Sie für die Verteilung der Pegelstände ein Stamm-Blatt-Diagramm mit a) einer Stammbreite von 100 (10 Punkte) und b) einer Stammbreite von 10 (10 Punkte).

Stamm-Blatt-Diagramm mit einer Stammbreite von 100

1|6 6 7 7 7 7 8 8 8

Stammbreite: 100

Keine Ausreißer

Jedes Blatt ein Fall

Stamm-Blatt-Diagramm mit einer Stammbreite von 10

16 | 3 8

17 | 2 2 3 8 8

18 | 1 2 3

Stammbreite: 10

Keine Ausreißer

Jedes Blatt ein Fall

Aufgabenteil III: Verteilungsparameter (30 Punkte)

In einem Automotive-Zulieferbetrieb werden täglich mehrere zehntausend Spitzgussteile hergestellt. Eine Zufallsstichprobe bestehend aus zehn Bauteilen wird vermessen und gewogen.

Bauteil-Nr.	Gewicht in g
01	26
02	27
03	23
04	26
05	24
06	24
07	78
08	29
09	21
10	22

Bestimmen Sie die folgenden Größen:

g) Arithmetisches Mittel (5 Punkte)

$$300 / 10 = 30$$

h) Um 20% getrimmtes arithmetisches Mittel (5 Punkte)

~~21, 22, 23, 24, 24, 26, 26, 27, 29, 78~~

$$150 / 6 = 25$$

i) Median (5 Punkte)

21, 22, 23, 24, 24, 26, 26, 27, 29, 78 Median = 25

j) Interquartilsabstand (5 Punkte)

21, 22, 23, 24, 24, 26, 26, 27, 29, 78 Oberes Quartil = 27 Unteres Quartil = 23 IQR = 4

k) Varianz (5 Punkte)

Bauteil-Nr.	Gewicht in g	$(x_i - 30)^2$
01	26	81
02	27	64
03	23	49
04	26	36
05	24	36
06	24	16
07	78	16
08	29	9
09	21	1
10	22	2304

$$2612 / 10 = 261,2$$

l) Standardabweichung (5 Punkte)

$$\text{Wurzel aus } 261,2 = 16,16$$

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

In einer Rinderzucht werden zehn ausgewachsenen Tiere verschiedenen Alters gewogen.

Rind-Nr.	Gewicht in kg	Alter in Jahren
01	563	15
02	701	21
03	523	13
04	521	12
05	618	16
06	633	18
07	571	16
08	671	20
09	710	22
10	512	10

a) Berechnen Sie den Rangkorrelationskoeffizienten nach Spearman. (10 Punkte)

b) Berechnen Sie den Bravais-Pearson-Korrelationskoeffizienten. (10 Punkte)

Konkordanzkoeffizient nach Spearman

Rind-Nr.	Gewicht	rg (x)	Alter	rg (y)	D	D ²
01	563	4	15	4	0	0
02	701	9	21	9	0	0
03	523	3	13	3	0	0
04	521	2	12	2	0	0
05	618	6	16	5,5	0,5	0,25
06	633	7	18	7	0	0
07	571	5	16	5,5	-0,5	0,25
08	671	8	20	8	0	0
09	710	10	22	10	0	0
10	512	1	10	1	0	0

$$\rho = 1 - [(6 \cdot 0,5) / (10^2 - 1) \cdot 10] = 1 - 3/990 = 0,9969$$

Bravais-Pearson-Korrelationskoeffizient

Rind-Nr.	Gewicht (x)	Alter (y)	x ²	y ²	(x*y)
01	563	15	316969	225	8445
02	701	21	491401	441	14721
03	523	13	273529	169	6799
04	521	12	271441	144	6252
05	618	16	381924	256	9888
06	633	18	400689	324	11394
07	571	16	326041	256	9136
08	671	20	450241	400	13420
09	710	22	504100	484	15620
10	512	10	262144	100	5120
Summe	6023	163	3678479	2799	100795
Mittel	602,3	16,3	//	//	//

$$r = \frac{100795 - 10 * 602,3 * 16,3}{\sqrt{3678479 - 10 * 602,3^2} * \sqrt{2799 - 10 * 16,3^2}} = \frac{2620,1}{\sqrt{50826,1} * \sqrt{142,1}}$$

$$= \frac{2620,1}{225,45 * 11,92} = \frac{2620,1}{2689,45} = 0,9749$$

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Rind-Nr.	Gewicht in kg	Alter in Jahren
01	563	15
02	701	21
03	523	13
04	521	12
05	618	16
06	633	18
07	571	16
08	671	20
09	710	22
10	512	10

Stellen Sie für die vorliegenden Daten aus der Rinderzucht eine lineare Regressionsfunktion mit dem Alter der Tiere in Jahren als unabhängiger Variable (x) und ihrem Gewicht in kg als abhängiger Variable (y) auf (10 Punkte). Bestimmen Sie anschließend die Güte des Regressionsmodells und interpretieren Sie das Ergebnis (10 Punkte).

Rind-Nr.	Alter (x)	Gewicht (y)	x ²	y ²	(x*y)
01	15	563	225	316969	8445
02	21	701	441	491401	14721
03	13	523	169	273529	6799
04	12	521	144	271441	6252
05	16	618	256	381924	9888
06	18	633	324	400689	11394
07	16	571	256	326041	9136
08	20	671	400	450241	13420
09	22	710	484	504100	15620
10	10	512	100	262144	5120
Summe	163	6023	2799	3678479	100795
Mittel	16,3	602,3	//	//	//

$$b = \frac{100795 - 10 * 16,3 * 602,3}{2799 - 10 * 16,3^2} = \frac{2620,1}{142,1} = 18,44$$

$$a = 602,3 - 18,44 * 16,3 = 301,73$$

$$y = 301,73 + 18,44 * x$$

Bewertung des Regressionsmodells

$$0,9749^2 = 0,95$$

Das Modell erklärt einen Großteil der Streuung.

Sommersemester 2018

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2018

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 07. Juli 2018

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2018

Christian Reinboth

Aufgabenteil I: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Der Quartilskoeffizient der Schiefe lässt sich lediglich bei bimodalen Verteilungen interpretieren.

☐ richtig ☐ falsch

2) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsen Schnittpunkts der linearen Regressionsgeraden.

☐ richtig ☐ falsch

3) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

4) Der Interquartilsabstand (IQR) ist ein robusteres Streuungsmaß als die Spannweite.

☐ richtig ☐ falsch

5) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☐ falsch

6) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☐ falsch

7) Der Median ist derjenige Wert, der in der geordneten Verteilung am häufigsten auftritt.

☐ richtig ☐ falsch

8) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☐ richtig ☐ falsch

9) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☐ falsch

10) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Ein Strandspaziergang erbringt eine Ausbeute von 20 Muscheln unterschiedlicher Art und Größe. Die spätere Vermessung des Funds führt zu folgenden Ergebnissen:

1	12 mm
2	16 mm
3	23 mm
4	17 mm
5	18 mm
6	20 mm
7	21 mm
8	16 mm
9	22 mm
10	41 mm
11	12 mm
12	13 mm
13	26 mm
14	21 mm
15	34 mm
16	38 mm
17	25 mm
18	16 mm
19	42 mm
20	33 mm

Stellen Sie ein Stamm-Blatt-Diagramm mit geeigneter Stammbreite auf.

Aufgabenteil III: Verteilungsparameter (20 Punkte)

Ein Fischkutter fängt mit dem Schleppnetz rund 1,2 Tonnen Thunfisch pro Fahrt. Um den Marktwert des Fangs zu taxieren, werden 10 zufällig ausgewählte Fische gewogen und in die Güteklassen A, B und C eingeteilt. Es ergibt sich folgendes Bild:

1	5,5 kg	A
2	4,3 kg	A
3	2,3 kg	C
4	3,6 kg	B
5	3,5 kg	A
6	5,1 kg	B
7	3,2 kg	A
8	4,4k g	B
9	6,0 kg	C
10	4,0 kg	C

Bestimmen Sie für beide Merkmale

- a) ein geeignetes Lagemaß. (5 Punkte je Lagemaß)
- b) ein geeignetes Streuungsmaß. (5 Punkte je Streuungsmaß)

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Eine Meeresbiologin untersucht 10 an der Flensburger Förde gestrandete Delfine. Dabei werden unter anderem Alter und Gewicht der Tiere festgehalten.

1	351 kg	12 Jahre
2	212 kg	7 Jahre
3	223 kg	8 Jahre
4	118 kg	5 Jahre
5	158 kg	12 Jahre
6	216 kg	17 Jahre
7	336 kg	16 Jahre
8	286 kg	11 Jahre
9	252 kg	10 Jahre
10	193 kg	9 Jahre

Berechnen Sie den Rangkorrelationskoeffizienten nach Spearman. Existiert ein monotoner Zusammenhang zwischen Alter und Gewicht? (10 Punkte) Würde es sich Ihrer Ansicht nach lohnen, auch den Bravais-Pearson-Korrelationskoeffizienten zu berechnen? Wenn ja, warum - wenn nein, warum nicht? (5 Punkte)

Aufgabenteil V: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Eine kommutative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

☐ richtig

☐ falsch

2) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

☐ richtig

☐ falsch

3) Die Durchführung des Chi-Quadrat-Unabhängigkeitstests setzt mindestens ordinal skalierte Merkmale voraus.

☐ richtig

☐ falsch

4) Bei einer stetigen Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

☐ richtig

☐ falsch

5) Die Wahrscheinlichkeiten aller möglichen Elementarereignisse eines Zufallsvorgangs ergeben zusammengerechnet stets den Wert 0.

☐ richtig

☐ falsch

6) Die Dichtefunktion der Normalverteilung ist glockenförmig und symmetrisch.

☐ richtig

☐ falsch

7) Die Teilmenge einer Menge kann nur maximal so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig

☐ falsch

8) Das logische NICHT wird auch als Negation bezeichnet.

☐ richtig

☐ falsch

9) Einzelne und unteilbare Mitglieder einer Menge werden als Elemente bezeichnet.

☐ richtig

☐ falsch

10) Das Gegenereignis zu mindestens 3 Augen beim einmaligen Wurf eines fairen Würfels lautet höchstens 4 Augen.

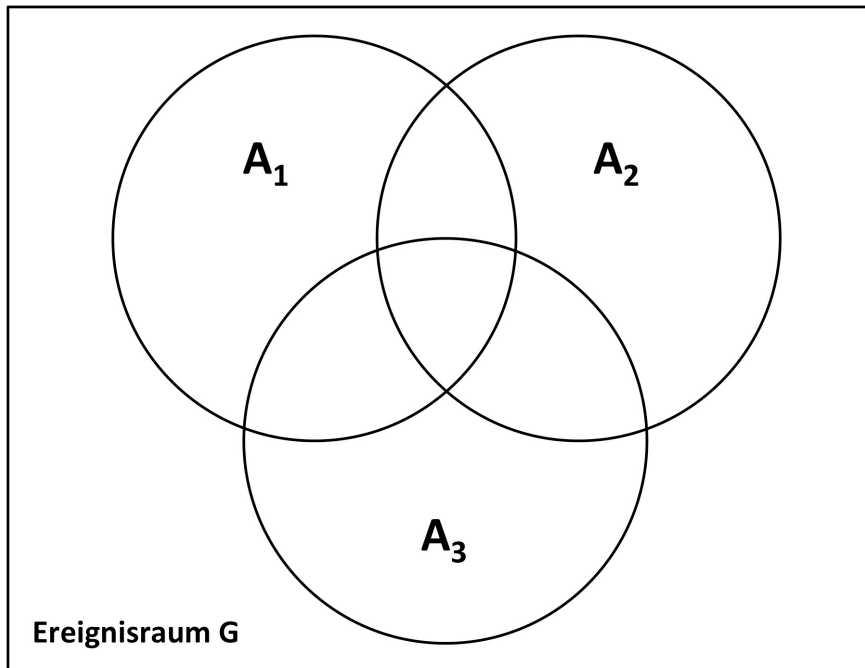
☐ richtig

☐ falsch

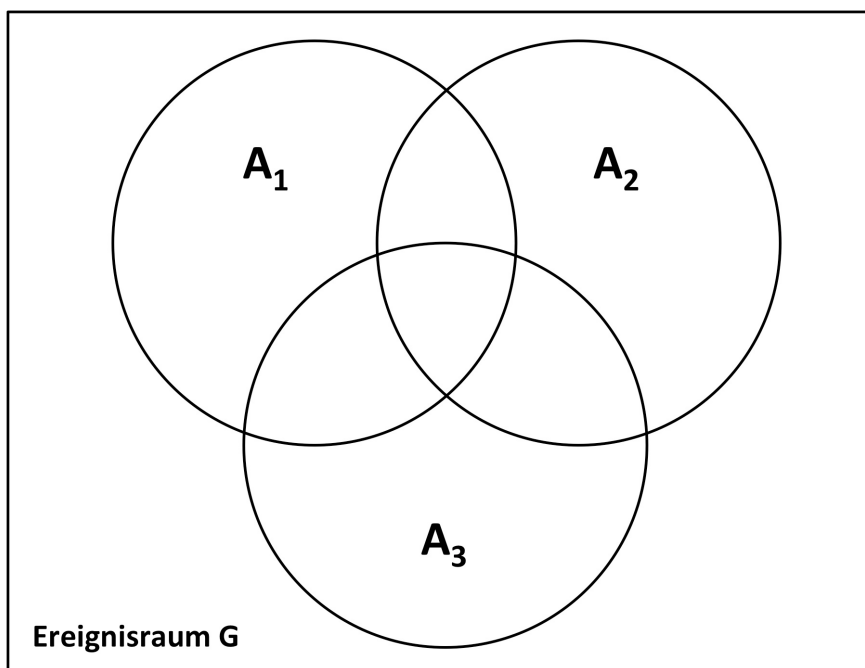
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

a) $(A_1 \cap A_2) \cup A_3$



b) $A_1 \cap \bar{A}_2$



Aufgabenteil VII: Kombinatorik (5 Punkte)

Vor einer Klappbrücke warten zwei Segelschiffe und drei Motorschiffe auf die Möglichkeit zum Auslaufen. In wie vielen verschiedenen Aneinanderreihungen von Segel- und Motorschiffen können die fünf Wasserfahrzeuge die Brücke passieren, sobald diese geöffnet wird?

Aufgabenteil VIII: Wahrscheinlichkeitsrechnung (10 Punkte)

Weil die zweite Klasse eines Ferienfliegers überbucht ist, sollen 3 der 200 wartenden Passagiere auf 3 freie Plätze in der ersten Klasse umgebucht werden. Unter den 200 wartenden Passagieren befinden sich 120 Männer, 60 Frauen und 20 Kinder. Wie groß ist bei zufälliger Auswahl der umzubuchenden Passagiere die Wahrscheinlichkeit dafür, dass

- a) ausschließlich Frauen (5 Punkte)
- b) mindestens zwei Männer (5 Punkte)

selektiert werden?

Aufgabenteil IX: Wahrscheinlichkeitsrechnung (10 Punkte)

Während 90% der an einem Strand zu findenden Quallen größtenteils harmlose Hautreizungen erzeugen, handelt es sich bei 10% um Feuerquallen, die bei Kontakt gefährliche Hautverletzungen verursachen können. Tritt man auf eine solche Feuerqualle, benötigt man mit einer Wahrscheinlichkeit von 95% sofortige ärztliche Hilfe - bei anderen Quallenarten liegt das Risiko auf eine schwerwiegende Verletzung dagegen bei nur 5%. Sie spüren nach einem Tritt auf eine Qualle schon das erste Jucken - wie groß ist die Wahrscheinlichkeit einer versorgungsbedürftigen Verletzung?

Aufgabenteil X: Bedingte Wahrscheinlichkeiten (10 Punkte)

Die Hafenpolizei geht davon aus, dass sich in 50 von 5.000 täglich an einem Hafen verschifften Containern Schmuggelware befindet. Eine neue Röntgenanlage soll dabei helfen, diese Container zuverlässiger zu identifizieren. Befindet sich Schmuggelware in einem Container, wird diese während des Röntgenvorgangs mit einer Sicherheit von 95% erkannt. Gleichzeitig werden aber auch 2% aller Container ohne Schmuggelware als kritisch identifiziert und einer genaueren Untersuchung unterzogen.

Sie stehen zufällig neben dem Röntgengerät, als dieses anschlägt. Wie groß ist die Wahrscheinlichkeit dafür, dass bei der nun folgenden Öffnung tatsächlich Schmuggelware gefunden wird?

KLAUSUR STATISTIK I UND II MIT LÖSUNGEN

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2018

Christian Reinboth

Aufgabenteil I: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Der Quartilskoeffizient der Schiefe lässt sich lediglich bei bimodalen Verteilungen interpretieren.

☐ richtig ☒ **falsch**

2) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsen Schnittpunkts der linearen Regressionsgeraden.

☐ richtig ☒ **falsch**

3) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☒ **richtig** ☐ falsch

4) Der Interquartilsabstand (IQR) ist ein robusteres Streuungsmaß als die Spannweite.

☒ **richtig** ☐ falsch

5) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☒ **richtig** ☐ falsch

6) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☒ **falsch**

7) Der Median ist derjenige Wert, der in der geordneten Verteilung am häufigsten auftritt.

☐ richtig ☒ **falsch**

8) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☒ **richtig** ☐ falsch

9) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☒ **falsch**

10) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

☒ **richtig** ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Ein Strandspaziergang erbringt eine Ausbeute von 20 Muscheln unterschiedlicher Art und Größe. Die spätere Vermessung des Funds führt zu folgenden Ergebnissen:

1	12 mm
2	16 mm
3	23 mm
4	17 mm
5	18 mm
6	20 mm
7	21 mm
8	16 mm
9	22 mm
10	41 mm
11	12 mm
12	13 mm
13	26 mm
14	21 mm
15	34 mm
16	38 mm
17	25 mm
18	16 mm
19	42 mm
20	33 mm

Stellen Sie ein Stamm-Blatt-Diagramm mit geeigneter Stammbreite auf.

1 | 2 2 3 6 6 6 7 8

2 | 0 1 1 2 3 5 6

3 | 3 4 8

4 | 1 2

Stammbreite 10

Jedes Blatt ein Fall

Keine Ausreißer

Aufgabenteil III: Verteilungsparameter (20 Punkte)

Ein Fischkutter fängt mit dem Schleppnetz rund 1,2 Tonnen Thunfisch pro Fahrt. Um den Marktwert des Fangs zu taxieren, werden 10 zufällig ausgewählte Fische gewogen und in die Güteklassen A, B und C eingeteilt. Es ergibt sich folgendes Bild:

1	5,5 kg	A
2	4,3 kg	A
3	2,3 kg	C
4	3,6 kg	B
5	3,5 kg	A
6	5,1 kg	B
7	3,2 kg	A
8	4,4 kg	B
9	6,0 kg	C
10	4,0 kg	C

Bestimmen Sie für beide Merkmale

- a) ein geeignetes Lagemaß. (5 Punkte je Lagemaß)
 b) ein geeignetes Streuungsmaß. (5 Punkte je Streuungsmaß)

a) Lagemaße

Gewicht: Arithmetisches Mittel: $\text{Summe aller Werte} / 10 = 41,9/10 = 4,19 \text{ kg}$
 Median: Arithmetisches Mittel aus 5. und 6. Wert (4,0 und 4,3) = 4,15 kg
 Güteklasse: Modus (ist aber nicht möglich, da Verteilung nicht unimodal)

b) Streuungsmaße

Gewicht: Spannweite: $6 - 2,3 = 3,7 \text{ kg}$
 Quartilkoeffizient der Schiefe: $(0,95 - 0,65) / 1,6 = 0,1875$

Unteres Quartil: 3. Wert = 3,5 Median: $(4,0 + 4,3) / 2 = 4,15$
 Oberes Quartil: 8. Wert = 5,1 IQR: $5,1 - 3,5 = 1,6$

Nr.	x	$(x - \text{arithmetisches Mittel})^2$
1	5,5 kg	1,7161
2	4,3 kg	0,0121
3	2,3 kg	3,5721
4	3,6 kg	0,3481
5	3,5 kg	0,4761
6	5,1 kg	0,8281
7	3,2 kg	0,9801
8	4,4 kg	0,0441
9	6,0 kg	3,2761
10	4,0 kg	0,0361

Summe = 11,289 Varianz: $11,289/10 = 1,289$ Standardabweichung: 1,062

Güteklasse: Es lässt sich kein sinnvolles Streuungsmaß bestimmen.

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Eine Meeresbiologin untersucht 10 an der Flensburger Förde gestrandete Delfine. Dabei werden unter anderem Alter und Gewicht der Tiere festgehalten.

1	351 kg	12 Jahre
2	212 kg	7 Jahre
3	223 kg	8 Jahre
4	118 kg	5 Jahre
5	158 kg	12 Jahre
6	216 kg	17 Jahre
7	336 kg	16 Jahre
8	286 kg	11 Jahre
9	252 kg	10 Jahre
10	193 kg	9 Jahre

Berechnen Sie den Rangkorrelationskoeffizienten nach Spearman. Existiert ein monotoner Zusammenhang zwischen Alter und Gewicht? (10 Punkte) Würde es sich Ihrer Ansicht nach lohnen, auch den Bravais-Pearson-Korrelationskoeffizienten zu berechnen? Wenn ja, warum - wenn nein, warum nicht? (5 Punkte)

Nr.	x	rg(x)	y	rg(y)	d	d ²
1	351	10	12	7,5	2,5	6,25
2	212	4	7	2	2	4
3	223	6	8	3	3	9
4	118	1	5	1	0	0
5	158	2	12	7,5	-5,5	30,25
6	216	5	17	10	-5	25
7	336	9	16	9	0	0
8	286	8	11	6	2	4
9	252	7	10	5	2	4
10	193	3	9	4	-1	1

Summe aller d² = 83,5

$$\rho = 1 - ((6 * 83,5) / ((10^2 - 1) * 10)) = 1 - (501 / 990) = 1 - 0,506 = 0,494$$

Aufgrund der relativen Schwäche dieses Koeffizienten für monotone Korrelationen würde sich die Berechnung eines Koeffizienten für lineare Korrelationen nicht lohnen.

Aufgabenteil V: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Eine kommutative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

X richtig O falsch

2) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

O richtig **X falsch**

3) Die Durchführung des Chi-Quadrat-Unabhängigkeitstests setzt mindestens ordinal skalierte Merkmale voraus.

O richtig **X falsch**

4) Bei einer stetigen Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

X richtig O falsch

5) Die Wahrscheinlichkeiten aller möglichen Elementarereignisse eines Zufallsvorgangs ergeben zusammengerechnet stets den Wert 0.

O richtig **X falsch**

6) Die Dichtefunktion der Normalverteilung ist glockenförmig und symmetrisch.

X richtig O falsch

7) Die Teilmenge einer Menge kann nur maximal so viele Elemente enthalten wie die Menge, zu der sie gehört.

X richtig O falsch

8) Das logische NICHT wird auch als Negation bezeichnet.

X richtig O falsch

9) Einzelne und unteilbare Mitglieder einer Menge werden als Elemente bezeichnet.

X richtig O falsch

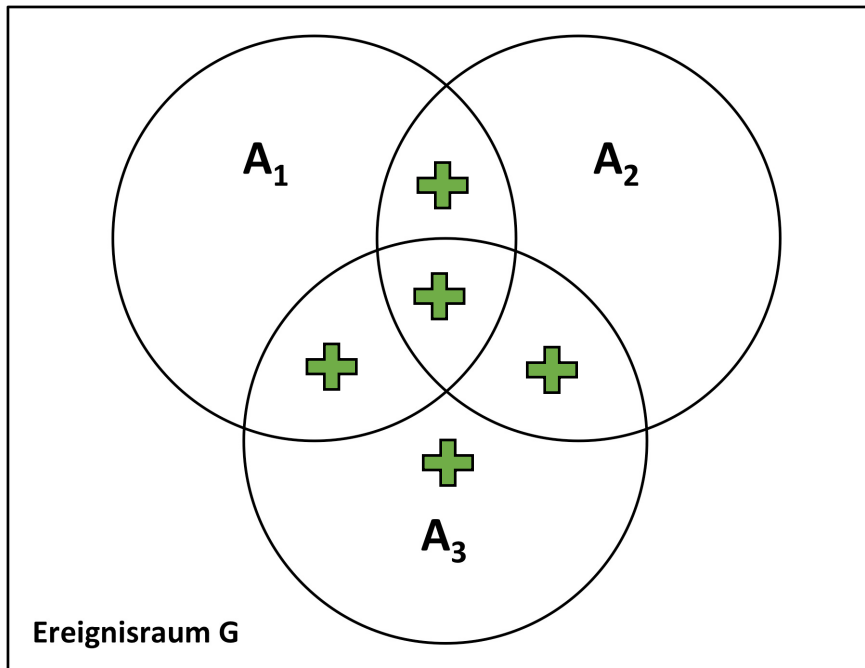
10) Das Gegenereignis zu mindestens 3 Augen beim einmaligen Wurf eines fairen Würfels lautet höchstens 4 Augen.

O richtig **X falsch**

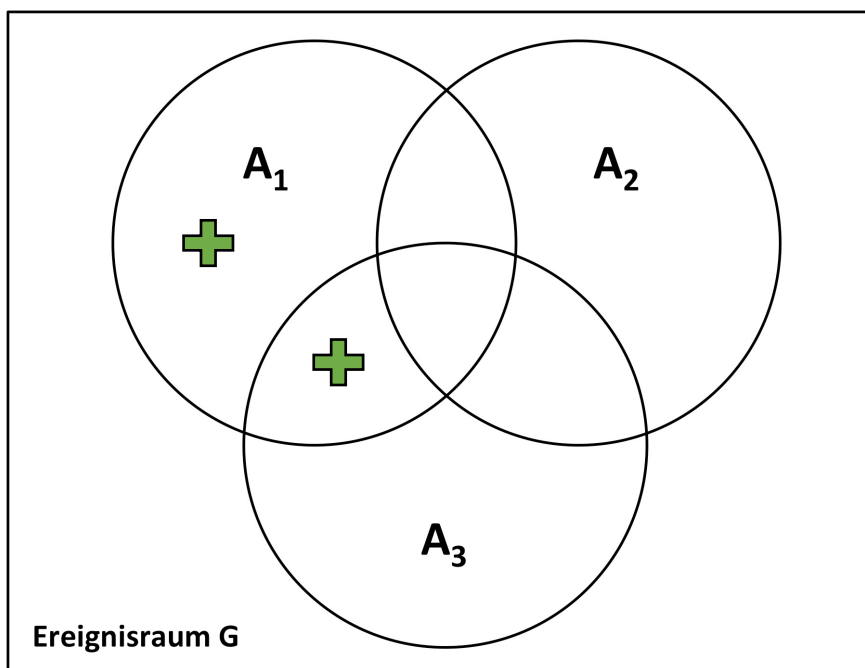
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

a) $(A_1 \cap A_2) \cup A_3$ = Gesamte Schnittfläche von A1 und A2 sowie A3 komplett



b) $A_1 \cap \bar{A}_2$ = Fläche von A1 ausgenommen der Schnittfläche mit A2



Aufgabenteil VII: Kombinatorik (5 Punkte)

Vor einer Klappbrücke warten zwei Segelschiffe und drei Motorschiffe auf die Möglichkeit zum Auslaufen. In wie vielen verschiedenen Aneinanderreihungen von Segel- und Motorschiffen können die fünf Wasserfahrzeuge die Brücke passieren, sobald diese geöffnet wird?

SSMMM	MSMMS	MSSMM	MSMSM
SMSMM	MMSMS	MMSSM	
SMMSM	MMMSS		
SMMMS			

Insgesamt existieren 10 Möglichkeiten der Anordnung.

Aufgabenteil VIII: Wahrscheinlichkeitsrechnung (10 Punkte)

Weil die zweite Klasse eines Ferienfliegers überbucht ist, sollen 3 der 200 wartenden Passagiere auf 3 freie Plätze in der ersten Klasse umgebucht werden. Unter den 200 wartenden Passagieren befinden sich 120 Männer, 60 Frauen und 20 Kinder. Wie groß ist bei zufälliger Auswahl der umzubuchenden Passagiere die Wahrscheinlichkeit dafür, dass

- a) ausschließlich Frauen (5 Punkte)
- b) mindestens zwei Männer (5 Punkte)

selektiert werden?

- a) Ein Pfad im Pfaddiagramm

$$F-F-F = 60/200 * 59/199 * 58/198 = 0,0261$$

Die Wahrscheinlichkeit liegt bei 2,61%.

- b) Sieben Pfade im Pfaddiagramm

$$M-M-M = 120/200 * 119/199 * 118/198 = 0,2138$$

$$M-M-F = 120/200 * 119/199 * 60/198 = 0,1087$$

$$M-M-K = 120/200 * 119/199 * 20/198 = 0,0362$$

$$M-F-M = 120/200 * 60/199 * 119/198 = 0,1087$$

$$M-K-M = 120/200 * 20/199 * 119/198 = 0,0362$$

$$F-M-M = 60/200 * 120/199 * 119/198 = 0,1087$$

$$K-M-M = 20/200 * 120/199 * 119/198 = 0,0362$$

$$\text{Summe: } 0,6487$$

Die Wahrscheinlichkeit liegt bei 64,87%.

Aufgabenteil IX: Wahrscheinlichkeitsrechnung (10 Punkte)

Während 90% der an einem Strand zu findenden Quallen größtenteils harmlose Hautreizungen erzeugen, handelt es sich bei 10% um Feuerquallen, die bei Kontakt gefährliche Hautverletzungen verursachen können. Tritt man auf eine solche Feuerqualle, benötigt man mit einer Wahrscheinlichkeit von 95% sofortige ärztliche Hilfe - bei anderen Quallenarten liegt das Risiko auf eine schwerwiegende Verletzung dagegen bei nur 5%. Sie spüren nach einem Tritt auf eine Qualle schon das erste Jucken - wie groß ist die Wahrscheinlichkeit einer versorgungsbedürftigen Verletzung?

Zwei Pfade im Pfaddiagramm:

Auf normale Qualle getreten + kritische Verletzung

Auf Feuerqualle getreten + kritische Verletzung

$$0,9 * 0,05 = 0,045$$

$$0,1 * 0,95 = 0,095$$

$$0,045 + 0,095 = 0,14 = 14\%$$

Kontrolle:

$$0,9 * 0,95 = 0,855$$

$$0,1 * 0,05 = 0,005$$

$$0,045 + 0,095 + 0,855 + 0,005 = 1 \rightarrow \text{passt}$$

Aufgabenteil X: Bedingte Wahrscheinlichkeiten (10 Punkte)

Die Hafenpolizei geht davon aus, dass sich in 50 von 5.000 täglich an einem Hafen verschifften Containern Schmuggelware befindet. Eine neue Röntgenanlage soll dabei helfen, diese Container zuverlässiger zu identifizieren. Befindet sich Schmuggelware in einem Container, wird diese während des Röntgenvorgangs mit einer Sicherheit von 95% erkannt. Gleichzeitig werden aber auch 2% aller Container ohne Schmuggelware als kritisch identifiziert und einer genaueren Untersuchung unterzogen.

Sie stehen zufällig neben dem Röntgengerät, als dieses anschlägt. Wie groß ist die Wahrscheinlichkeit dafür, dass bei der nun folgenden Öffnung tatsächlich Schmuggelware gefunden wird?

5.000 Container

4.950 harmlos - davon 2% false positives = 99 Container

50 kritisch - davon 95% korrekt identifiziert = 47,5 Container

Gesamtzahl der als kritisch identifizierten Container: $99 + 47,5 = 146,5$

davon tatsächlich kritisch: 47,5

Verhältnis: 32,42%

Die Wahrscheinlichkeit auf einen Treffer liegt bei 32,42%.

Alternativ über die Formel:

$$(0,01 * 0,95) / ((0,01 * 0,95) + (0,99 * 0,02)) = 0,0095 / (0,0095 + 0,0198) = 0,0095 / 0,0293 = 0,3242$$

Wintersemester 2018/2019

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Wintersemester 2018 / 2019

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 03. November 2018

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2018/2019

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Quartilskoeffizient der Schiefe lässt sich lediglich bei bimodalen Verteilungen interpretieren.

☐ richtig ☐ falsch

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

3) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

4) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

5) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für lineare Zusammenhänge.

☐ richtig ☐ falsch

6) Der Modus lässt sich nur bestimmen, wenn eine biomodale Verteilung vorliegt.

☐ richtig ☐ falsch

7) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

8) Bei der Klassierung von Daten dürfen auch unterschiedlich breite Klassen gebildet werden.

☐ richtig ☐ falsch

9) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☐ falsch

10) Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Die nachfolgende Tabelle enthält zehn Datensätze aus einer Studierendenbefragung, die auf dem Campus der Hochschule Harz durchgeführt wurde. Gefragt wurde (unter anderem) nach Alter, (biologischem) Geschlecht, monatlich verfügbarem Nettoeinkommen und gewähltem Studiengang.

Nr.	Alter	Geschlecht	Einkommen	Studiengang
1	21 Jahre	männlich	580,00 EUR	BWL
2	20 Jahre	männlich	420,00 EUR	BWL
3	23 Jahre	weiblich	230,00 EUR	Informatik
4	28 Jahre	männlich	450,00 EUR	Medienkonzeption
5	20 Jahre	weiblich	380,00 EUR	Medienkonzeption
6	27 Jahre	weiblich	450,00 EUR	Informatik
7	33 Jahre	männlich	320,00 EUR	Informatik
8	21 Jahre	weiblich	410,00 EUR	BWL
9	22 Jahre	weiblich	280,00 EUR	BWL
10	19 Jahre	weiblich	260,00 EUR	Medienkonzeption

Zeichnen Sie einen erweiterten Box-Plot für das Alter und geben Sie die für die Konstruktion benötigten Größen an (10 Punkte). Wie alt müsste ein zusätzlicher Proband mindestens sein, um im Box-Plot als Ausreißer am oberen Rand der Verteilung ausgewiesen zu werden? (10 Punkte)

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Berechnen Sie auf Basis der bereits bekannten Datensätze aus der Studierendenbefragung die weiter unten aufgeführten Mittelwerte und Streuungsmaße.

Nr.	Alter	Geschlecht	Einkommen	Studiengang
1	21 Jahre	männlich	580,00 EUR	BWL
2	20 Jahre	männlich	420,00 EUR	BWL
3	23 Jahre	weiblich	230,00 EUR	Informatik
4	28 Jahre	männlich	450,00 EUR	Medienkonzeption
5	20 Jahre	weiblich	380,00 EUR	Medienkonzeption
6	27 Jahre	weiblich	450,00 EUR	Informatik
7	33 Jahre	männlich	320,00 EUR	Informatik
8	21 Jahre	weiblich	410,00 EUR	BWL
9	22 Jahre	weiblich	280,00 EUR	BWL
10	19 Jahre	weiblich	260,00 EUR	Medienkonzeption

- Arithmetisches Mittel des Alters (5 Punkte).
- Median und Modus des Geschlechts (6 Punkte).
- Median und Modus des Studiengangs (6 Punkte).
- Spannweite, Interquartilsabstand und Standardabweichung des Einkommens (6 Punkte).
- Spannweite, Interquartilsabstand und Standardabweichung des Studiengangs (3 Punkte).
- Welche Unterschiede gibt es zwischen den Merkmalen "Geschlecht" und "Studiengang"? (4 Punkte) [Tipp: Es gibt drei Unterschiede, Sie müssen aber nicht alle finden.]

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Die Hochschule Harz ist bemüht, die Höhe der eingeworbenen öffentlichen Fördermittel für die Durchführung von Forschungsprojekten (die sogenannten Drittmittel) stetig zu steigern - unter anderem durch die Erarbeitung und Einreichung von zusätzlichen Förderanträgen. Die (fiktive) Tabelle zeigt, wie sich die Anzahl der eingereichten Förderanträge und die Summe der mit diesen Anträgen eingeworbenen Drittmittel über fünf Jahre entwickelt hat.

Jahr	Eingereichte Förderanträge	Eingeworbene Drittmittel
2010	3	1,1 Mio. Euro
2011	8	1,6 Mio. Euro
2012	10	1,8 Mio. Euro
2013	14	1,9 Mio. Euro
2014	19	1,9 Mio. Euro

Berechnen und interpretieren Sie den Bravais-Pearson-Regressionskoeffizienten.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Sehen Sie sich bitte erneut die aus Aufgabe IV bereits bekannte Tabelle an.

Jahr	Eingereichte Förderanträge	Eingeworbene Drittmittel
2010	3	1,1 Mio. Euro
2011	8	1,6 Mio. Euro
2012	10	1,8 Mio. Euro
2013	14	1,9 Mio. Euro
2014	19	1,9 Mio. Euro

a) Stellen Sie die lineare Regressionsgleichung auf, wobei die Anzahl der gestellten Förderanträge als unabhängige Variable und die Höhe der eingeworbenen Drittmittel als abhängige Variable zu betrachten sind (10 Punkte).

b) Bestimmen und interpretieren Sie die Güte der Regressionsgleichung (5 Punkte).

c) Werfen Sie nochmal einen Blick auf das Ergebnis von Aufgabe IV. Hätten Sie angesichts dieses Ergebnisses eine Regressionsanalyse durchgeführt, wenn es nicht Klausuraufgabe gewesen wäre? Wenn ja, warum - wenn nein, warum nicht? (5 Punkte)

MUSTERLÖSUNG ZUR KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2018/2019

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Quartilskoeffizient der Schiefe lässt sich lediglich bei bimodalen Verteilungen interpretieren.

O richtig **X falsch**

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

X richtig O falsch

3) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

O richtig **X falsch**

4) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

X richtig O falsch

5) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für lineare Zusammenhänge.

X richtig O falsch

6) Der Modus lässt sich nur bestimmen, wenn eine biomodale Verteilung vorliegt.

O richtig **X falsch**

7) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

X richtig O falsch

8) Bei der Klassierung von Daten dürfen auch unterschiedlich breite Klassen gebildet werden.

X richtig O falsch

9) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

O richtig **X falsch**

10) Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert.

X richtig O falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Die nachfolgende Tabelle enthält zehn Datensätze aus einer Studierendenbefragung, die auf dem Campus der Hochschule Harz durchgeführt wurde. Gefragt wurde (unter anderem) nach Alter, (biologischem) Geschlecht, monatlich verfügbarem Nettoeinkommen und gewähltem Studiengang.

Nr.	Alter	Geschlecht	Einkommen	Studiengang
1	21 Jahre	männlich	580,00 EUR	BWL
2	20 Jahre	männlich	420,00 EUR	BWL
3	23 Jahre	weiblich	230,00 EUR	Informatik
4	28 Jahre	männlich	450,00 EUR	Medienkonzeption
5	20 Jahre	weiblich	380,00 EUR	Medienkonzeption
6	27 Jahre	weiblich	450,00 EUR	Informatik
7	33 Jahre	männlich	320,00 EUR	Informatik
8	21 Jahre	weiblich	410,00 EUR	BWL
9	22 Jahre	weiblich	280,00 EUR	BWL
10	19 Jahre	weiblich	260,00 EUR	Medienkonzeption

Zeichnen Sie einen erweiterten Box-Plot für das Alter und geben Sie die für die Konstruktion benötigten Größen an (10 Punkte). Wie alt müsste ein zusätzlicher Proband mindestens sein, um im Box-Plot als Ausreißer am oberen Rand der Verteilung ausgewiesen zu werden? (10 Punkte)

Ordnen der Werte - 19, 20, 20, 21, 21, 22, 23, 27, 28, 33

Erster Schritt: Konstruktion der Box

Unteres Quartil: $0,25 \cdot 10 = 2,5 \rightarrow$ Wert an 3. Stelle $\rightarrow 20$

Mittleres Quartil: $0,50 \cdot 10 = 5 \rightarrow$ Mittel aus Werten an 5. und 6. Stelle $\rightarrow (21 + 22) / 2 = 21,5$

Oberes Quartil: $0,75 \cdot 10 = 7,5 \rightarrow$ Wert an 8. Stelle $\rightarrow 27$

IQR: $27 - 20 = 7$

Die Box verläuft somit von 20 auf 27, der Median wird bei 21,5 eingezeichnet.

Zweiter Schritt: Konstruktion der Zäune

1,5-facher IQR: $7 \cdot 1,5 = 10,5$

Unterer Zaun läuft minimal bis: $20 - 10,5 = 9,5 \rightarrow$ kleinster Wert ist die 19

Oberer Zaun läuft maximal bis: $27 + 10,5 = 37,5 \rightarrow$ größter Wert ist die 33

Der untere Zaun läuft bis zur 19, der obere Zaun bis zur 33.

Dritter Schritt: Identifikation der Ausreißer $\rightarrow$ entfällt

Wie alt müsste ein zusätzlicher Proband sein, um als Ausreißer gewertet zu werden?

Einfache Antwort für 7 von 10 Punkten (keine Ergänzung der Verteilung): Mindestens 38 Jahre.

Komplizierte Antwort für 10 von 10 Punkten: Neuberechnung des IQR und der Zäune erforderlich.

Unteres Quartil: $0,25 \cdot 11 = 2,75 \rightarrow$ Wert an 3. Stelle $\rightarrow 20$

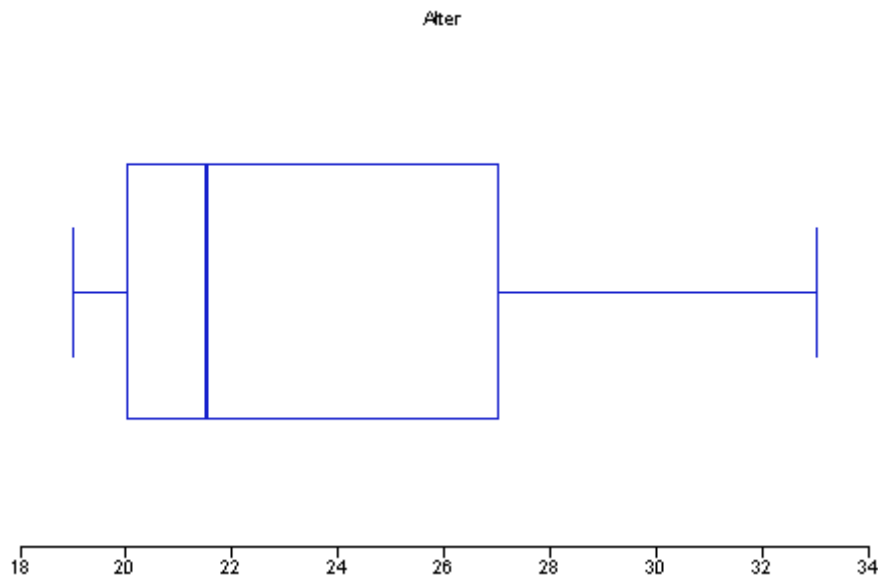
Oberes Quartil: $0,75 \cdot 11 = 8,25 \rightarrow$ Wert an 9. Stelle $\rightarrow 28$

IQR: $28 - 20 = 8$

1,5-facher IQR = $1,5 \cdot 8 = 12$

Der obere Zaun läuft somit maximal bis $28 + 12 = 40$.

Ein Ausreißer müsste mindestens 41 Jahre alt sein.



Aufgabenteil III: Verteilungsparameter (30 Punkte)

Berechnen Sie auf Basis der bereits bekannten Datensätze aus der Studierendenbefragung die weiter unten aufgeführten Mittelwerte und Streuungsmaße.

Nr.	Alter	Geschlecht	Einkommen	Studiengang
1	21 Jahre	männlich	580,00 EUR	BWL
2	20 Jahre	männlich	420,00 EUR	BWL
3	23 Jahre	weiblich	230,00 EUR	Informatik
4	28 Jahre	männlich	450,00 EUR	Medienkonzeption
5	20 Jahre	weiblich	380,00 EUR	Medienkonzeption
6	27 Jahre	weiblich	450,00 EUR	Informatik
7	33 Jahre	männlich	320,00 EUR	Informatik
8	21 Jahre	weiblich	410,00 EUR	BWL
9	22 Jahre	weiblich	280,00 EUR	BWL
10	19 Jahre	weiblich	260,00 EUR	Medienkonzeption

- Arithmetisches Mittel des Alters (5 Punkte).
 - Median und Modus des Geschlechts (6 Punkte).
 - Median und Modus des Studiengangs (6 Punkte).
 - Spannweite, Interquartilsabstand und Standardabweichung des Einkommens (6 Punkte).
 - Spannweite, Interquartilsabstand und Standardabweichung des Studiengangs (3 Punkte).
 - Welche Unterschiede gibt es zwischen den Merkmalen "Geschlecht" und "Studiengang"? (4 Punkte) [Tipp: Es gibt drei Unterschiede, Sie müssen aber nicht alle finden.]
- a) $21 + 20 + 23 + 28 + 20 + 27 + 33 + 21 + 22 + 19 = 234 \rightarrow 234 / 10 = 23,4$
- b) Median: Nicht möglich (Skalenniveau). Modus: weiblich (6 von 10)
- c) Median: Nicht möglich (Skalenniveau). Modus: Nicht möglich, da kein eindeutiges Maximum.
- d) Spannweite: $580 - 230 = 350$

Geordnete Reihenfolge: 230, 260, 280, 320, 380, 410, 420, 450, 450, 580

Unteres Quartil: $0,25 \cdot 10 = 2,5 \rightarrow$ Wert an 3. Stelle $\rightarrow 280$

Oberes Quartil: $0,75 \cdot 10 = 7,5 \rightarrow$ Wert an 8. Stelle $\rightarrow 450$

IQR: $450 - 280 = 170$

Berechnung von Varianz und Standardabweichung:

Arithmetisches Mittel: $230 + 260 + 280 + 320 + 380 + 410 + 420 + 450 + 450 + 580 = 3780$

$\rightarrow 3780 / 10 = 378$

$(230 - 378)^2 = 21904$	$(260 - 378)^2 = 13924$
$(280 - 378)^2 = 9604$	$(320 - 378)^2 = 3364$
$(380 - 378)^2 = 4$	$(410 - 378)^2 = 1024$
$(420 - 378)^2 = 1764$	$(450 - 378)^2 = 5184$
$(450 - 378)^2 = 5184$	$(580 - 378)^2 = 40804$

Summe: 102760

Varianz: $102760 / 10 = 10276$

Standardabweichung: Wurzel aus 10276 = 101,37

e) Alles nicht möglich.

f) Geschlecht: nicht häufbar	Studiengang: häufbar
Geschlecht: unimodal	Studiengang: multimodal
Geschlecht: dichotom/binär	Studiengang: nicht dichotom/binär

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Die Hochschule Harz ist bemüht, die Höhe der eingeworbenen öffentlichen Fördermittel für die Durchführung von Forschungsprojekten (die sogenannten Drittmittel) stetig zu steigern - unter anderem durch die Erarbeitung und Einreichung von zusätzlichen Förderanträgen. Die (fiktive) Tabelle zeigt, wie sich die Anzahl der eingereichten Förderanträge und die Summe der mit diesen Anträgen eingeworbenen Drittmittel über fünf Jahre entwickelt hat.

Jahr	Eingereichte Förderanträge	Eingeworbene Drittmittel
2010	3	1,1 Mio. Euro
2011	8	1,6 Mio. Euro
2012	10	1,8 Mio. Euro
2013	14	1,9 Mio. Euro
2014	19	1,9 Mio. Euro

Berechnen und interpretieren Sie den Bravais-Pearson-Regressionskoeffizienten.

Erster Schritt: Anlage der Hilfstabelle

Jahr	x	y	x^2	y^2	$(x \cdot y)$
2010	3	1,1	9	1,21	3,3
2011	8	1,6	64	2,56	12,8
2012	10	1,8	100	3,24	18,0
2013	14	1,9	196	3,61	26,6
2014	19	1,9	361	3,61	36,1
Summe	54	8,3	730	14,23	96,8
Durchschnitt	10,8	1,66	//	//	//

Zweiter Schritt: Durchführung der Berechnung

$$r = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sqrt{\sum_{i=1}^n (x^2) - n * \bar{x}^2} * \sqrt{\sum_{i=1}^n (y^2) - n * \bar{y}^2}}$$

$$\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y} = 7,16$$

$$\sqrt{\sum_{i=1}^n (x^2) - n * \bar{x}^2} = 12,12$$

$$\sqrt{\sum_{i=1}^n (y^2) - n * \bar{y}^2} = 0,67$$

$$\frac{7,16}{12,12 * 0,67} = 0,88$$

Dritter Schritt: Interpretation des Ergebnisses

Es liegt eine stark positive lineare Korrelation vor.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Sehen Sie sich bitte erneut die aus Aufgabe IV bereits bekannte Tabelle an.

Jahr	Eingereichte Förderanträge	Eingeworbene Drittmittel
2010	3	1,1 Mio. Euro
2011	8	1,6 Mio. Euro
2012	10	1,8 Mio. Euro
2013	14	1,9 Mio. Euro
2014	19	1,9 Mio. Euro

a) Stellen Sie die lineare Regressionsgleichung auf, wobei die Anzahl der gestellten Förderanträge als unabhängige Variable und die Höhe der eingeworbenen Drittmittel als abhängige Variable zu betrachten sind (10 Punkte).

b) Bestimmen und interpretieren Sie die Güte der Regressionsgleichung (5 Punkte).

c) Werfen Sie nochmal einen Blick auf das Ergebnis von Aufgabe IV. Hätten Sie angesichts dieses Ergebnisses eine Regressionsanalyse durchgeführt, wenn es nicht Klausuraufgabe gewesen wäre? Wenn ja, warum - wenn nein, warum nicht? (5 Punkte)

Aufstellung der Regressionsgleichung

Jahr	x	y	x ²	(x*y)
2010	3	1,1	9	3,3
2011	8	1,6	64	12,8
2012	10	1,8	100	18,0
2013	14	1,9	196	26,6
2014	19	1,9	361	36,1
Summe	54	8,3	730	96,8
Durchschnitt	10,8	1,66	//	//

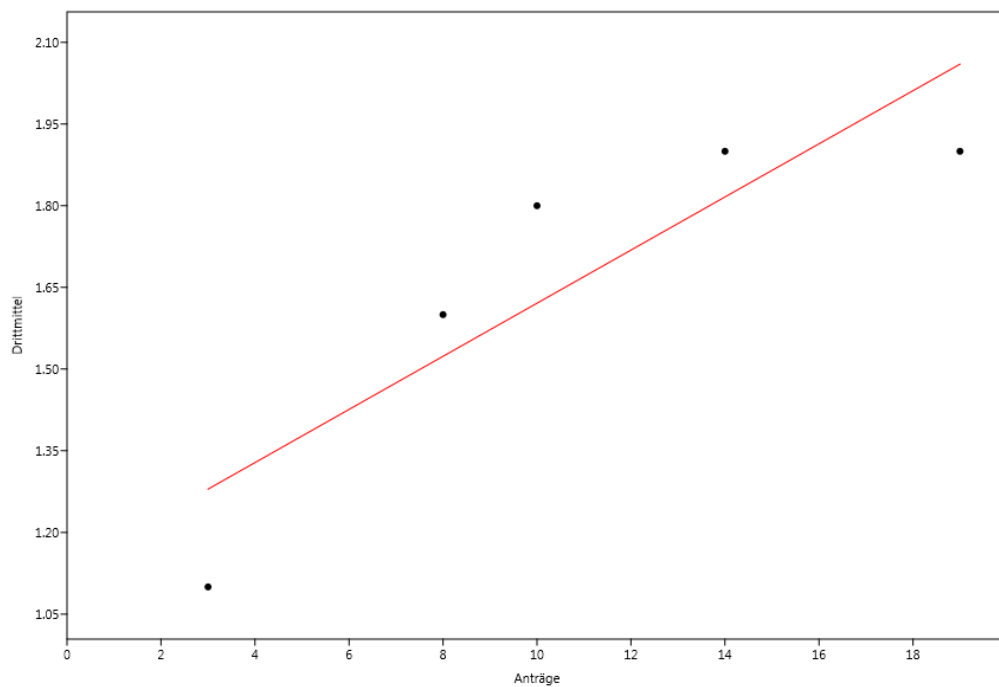
$$b = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sum_{i=1}^n (x^2) - n * \bar{x}^2}$$

$$b = 7,16 / 146,8 = 0,048$$

$$a = \bar{y} - b * \bar{x}$$

$$a = 1,66 - 0,0488 * 10,8 = 1,1329$$

$$y = 1,1329 + 0,0488 * x$$

Bestimmung und Interpretation der Modellgüte

Die Güte der Regressionsgleichung wurde über das Quadrieren des B-P-K ermittelt:

$$R^2 = 0,88^2 = 0,77$$

Das Modell leistet einen guten Beitrag zur Streuungsaufklärung (77%). Da der B-P-K bereits erkennen ließ, dass eine lineare Korrelation vorliegt, lag die Durchführung einer Regressionsanalyse nahe.

Wintersemester 2018/2019

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Wintersemester 2018 / 2019

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 20. Oktober 2018

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Wintersemester 2018/2019

Christian Reinboth

Aufgabenteil I: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Der Regressionskoeffizient (b) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsenschnittpunkts der linearen Regressionsgeraden.

☐ richtig ☐ falsch

2) Ist ein Merkmal häufbar, kann es mehrere Ausprägungen gleichzeitig annehmen.

☐ richtig ☐ falsch

3) Ordinalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☐ falsch

4) Die Fünf-Werte-Zusammenfassung kann nur für ordinalskalierte Werte berechnet werden.

☐ richtig ☐ falsch

5) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig ☐ falsch

6) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☐ falsch

7) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

8) Zwei Teilgesamtheiten einer Grundgesamtheit können sich niemals überschneiden.

☐ richtig ☐ falsch

9) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

10) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Ein Online-Shop für Schuhe erhebt unter anderem, wie viele Sekunden Besucher*innen auf den einzelnen Produktseiten verbleiben. Die nachfolgende Tabelle enthält die entsprechenden Daten von 20 Besucher*innen, die am 18.10.2018 die Produktseite des „Weidenstock Naturleder Gr. 37“ aufgerufen haben.

Nr.	Zeit in Sekunden
1	55
2	41
3	123
4	48
5	12
6	11
7	71
8	85
9	63
10	65
11	51
12	289
13	74
14	13
15	38
16	98
17	42
18	71
19	81
20	92

Stellen Sie ein Stamm-Blatt-Diagramm mit geeigneter Stammbreite auf.

Aufgabenteil III: Verteilungsparameter (20 Punkte)

Die nachfolgende Tabelle zeigt einen Auszug aus der Kundenkartei eines Feinkost-Versandhandels.

	Gesamtbestellwert April	Gesamtbestellwert Mai	Gesamtbestellwert Juni
Kunde A	127,00 EUR	138,00 EUR	192,00 EUR
Kunde B	98,00 EUR	210,00 EUR	37,00 EUR
Kunde C	143,00 EUR	0,00 EUR	123,00 EUR

Berechnen Sie

- arithmetisches Mittel (3 Punkte) und Median (3 Punkte) des Bestellwerts pro Kunde sowie
- Varianz (5 Punkte) und Standardabweichung (5 Punkte) der Bestellwerte insgesamt.
- Welches Maß der zentralen Tendenz ist hier aussagekräftiger – und warum? (4 Punkte)

Aufgabenteil IV: Regressionskoeffizienten (15 Punkte)

Ein Kinobetreiber verändert (ceteris paribus) an 10 Tagen die Eintrittspreise für den Abendfilm und zeichnet die Besucherzahlen auf. Es ergibt sich die folgende Tabelle:

Abend Nr.	Preis pro Karte (in Euro)	Besucher/innen pro Abend
1	8,50	836
2	9,00	741
3	9,50	798
4	10,00	740
5	10,50	685
6	11,00	673
7	11,50	710
8	12,00	625
9	12,50	531
10	13,00	330

Berechnen (12 Punkte) und interpretieren (3 Punkte)
Sie den Bravais-Pearson-Korrelationskoeffizienten.

Aufgabenteil V: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Mit dem Satz von Bayes kann nur gerechnet werden, wenn mindestens ordinal skalierte Merkmale vorliegen.

☐ richtig ☐ falsch

2) Die Differenzmenge zweier Mengen A und B ist definiert als die Menge derjenigen Elemente, die weder in Menge A noch in Menge B enthalten sind.

☐ richtig ☐ falsch

3) Der erste Schritt in jedem statistischen Testverfahren besteht in der Entscheidung, ob die Nullhypothese oder die Alternativhypothese getestet werden sollen.

☐ richtig ☐ falsch

4) Bei einer stetigen Verteilung beträgt die Punktwahrscheinlichkeit stets gleich Null.

☐ richtig ☐ falsch

5) Je höher das Konfidenzniveau eines Konfidenzintervalls ist, umso wahrscheinlicher ist, dass es den realen Wert in der Grundgesamtheit einschließt.

☐ richtig ☐ falsch

6) Eine Menge kann beliebig viele Teilmengen enthalten.

☐ richtig ☐ falsch

7) Ist die Schnittmenge der Mengen A und B leer, so bedeutet dies, dass die Menge A die Komplementärmenge der Menge B (und umgekehrt) ist.

☐ richtig ☐ falsch

8) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

9) Gerade und ungerade Zahlen sind ein passendes Beispiel für disjunkte Mengen.

☐ richtig ☐ falsch

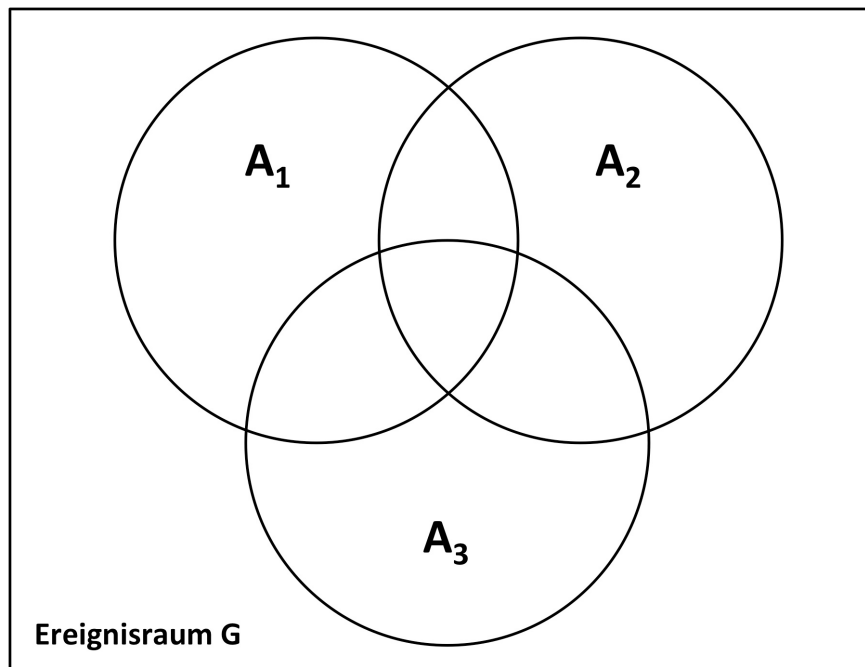
10) Eine kommutative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

☐ richtig ☐ falsch

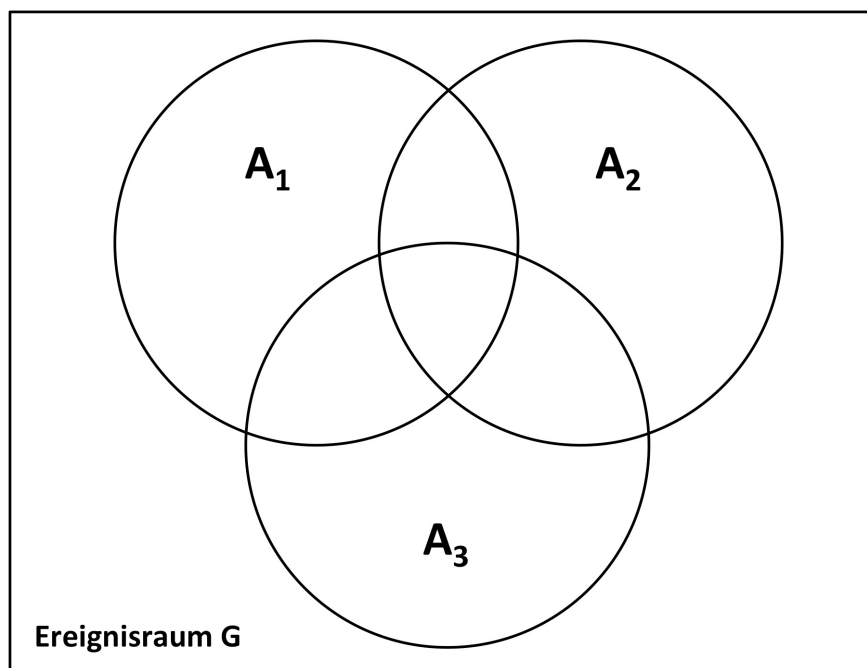
Aufgabenteil VI: Mengenlehre (10 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 5 Punkte)

a) $(A_1 \cap A_2) \cup A_3$



b) $\bar{A}_1 \cap \bar{A}_3$



Aufgabenteil VII: Kombinatorik (5 Punkte)

Vor der beschränkten Zufahrt eines Shopping-Centers warten zwei große LKW und drei kleine Lieferwagen auf die Möglichkeit der Einfahrt. In wie vielen verschiedenen Aneinanderreihungen von LKW und Lieferwagen können die fünf Fahrzeuge die Schranke passieren, sobald diese geöffnet wird?

Aufgabenteil VIII: Wahrscheinlichkeitsrechnung (10 Punkte)

In einer Ticketbörse werden Last Minute-Tickets für ein gefragtes Konzert vertrieben. Die Interessent*innen werden dabei in eine „Online-Warteschlange“ eingereiht und nach dem Zufallsprinzip selektiert. Derzeit sind noch zwei Tickets verfügbar, während sich 120 mögliche Käufer*innen in der Warteschlange befinden. Während 40 der Wartenden überzeugte Fans der auftretenden Musikgruppe sind, handelt es sich bei 80 Wartenden um Zwischenhändler, die die erstandenen Karten zu einem höheren Preis weiterverkaufen wollen. Wie groß ist bei zufälliger Auswahl der beiden letzten Ticketkäufer*innen die Wahrscheinlichkeit dafür, dass

- a) ausschließlich Fans (5 Punkte) sowie
- b) mindestens ein Wiederverkäufer (5 Punkte)

selektiert werden?

Aufgabenteil IX: Bedingte Wahrscheinlichkeiten (10 Punkte)

Ein Online-Versandhändler vertreibt gebrauchte Notebooks, die vor dem Weiterverkauf lediglich oberflächlich auf ihre Funktionstauglichkeit geprüft werden. Bei 10% dieser Notebooks treten beim Kunden noch vorab unerkannte Fehler auf, während 90% defektfrei ausgeliefert werden. Liegt ein reklamationfähiger Defekt vor, wird dieser durch den Käufer in 80% aller Fälle angezeigt. Auch bei 5% der defektfreien Geräte kommt es aufgrund von Benutzerfehlern zu Problemen, die in einer (dann allerdings erfolglosen) Reklamation münden. Wie groß ist nun die Wahrscheinlichkeit dafür, dass ein beliebiges reklamiertes Gerät tatsächlich defekt ist und umgetauscht werden muss?

Aufgabenteil X: Chi-Quadrat-Anpassungstest (10 Punkte)

Jeweils 10 männliche und 10 weibliche Probanden nehmen an einem Produkttest für ein neues Shampoo teil, dessen Ergebnis auf der bekannten Schulnotenskala von 1 bis 6 bewertet wurde. Die nachfolgende Tabelle zeigt die Verteilung dieser Bewertungen auf sehr gute bis gute (Noten 1 und 2), befriedigende (Noten 3 und 4) und ungenügende (Noten 5 und 6) Ergebnisse.

Ergebnis/Geschlecht	Männlich	Weiblich
Noten 1 und 2	2	6
Noten 3 und 4	4	3
Noten 5 und 6	4	1

Führen Sie auf Basis dieser Daten einen Chi-Quadrat-Test auf stochastische Unabhängigkeit der beiden Variablen (Geschlecht/Testergebnis) durch (7 Punkte) und interpretieren Sie das Ergebnis (3 Punkte). Als Irrtumswahrscheinlichkeit bzw. Wahrscheinlichkeit für einen Fehler 1. Ordnung ist 1% zu wählen (kritischer Wert: 6,63).

Kurzlösung zur Klausur Statistik I/II am 20.10.2018**Aufgabenteil I**

F
R
R
F
R
R
F
F
F
R

Aufgabenteil V

F
F
F
R
R
R
F
R
R
R

Aufgabenteil II

1 | 1 2 3
2 |
3 | 8
4 | 1 2 8
5 | 1 5
6 | 3 5
7 | 1 1 4
8 | 1 5
9 | 2 8

Jedes Blatt ein Fall
Stammbreite 10
2 Ausreißer

Aufgabenteil III

Kunde 1: Mittel = 152,33 EUR Median = 138,00 EUR
Kunde 2: Mittel = 115,00 EUR Median = 98,00 EUR
Kunde 3: Mittel = 88,66 EUR Median = 123,00 EUR

$$\begin{aligned}
 (127 - 118,66)^2 &= 641,61 \\
 (138 - 118,66)^2 &= 205,35 \\
 (192 - 118,66)^2 &= 1573,71 \\
 (98 - 118,66)^2 &= 2951,75 \\
 (210 - 118,66)^2 &= 3325,83 \\
 (37 - 118,66)^2 &= 13301,01 \\
 (143 - 118,66)^2 &= 87,05 \\
 (0 - 118,66)^2 &= 23204,43 \\
 (123 - 118,66)^2 &= 860,25 \\
 &= 46150,98 / 9 = 5127,89
 \end{aligned}$$

Positive Wurzel der Varianz: 71,61

Die größere Aussagekraft hat in diesem Fall das arithmetische Mittel.

Aufgabenteil IV

x	y	x ²	y ²	(x*y)
8,50	836	72,25	698896	7106,00
9,00	741	81,00	549081	6669,00
9,50	798	90,25	636804	7581,00
10,00	740	100,00	547600	7400,00
10,50	685	110,25	469225	7192,50
11,00	673	121,00	452929	7403,00
11,50	710	132,25	504100	8165,00
12,00	625	144,00	390625	7500,00
12,50	531	156,25	281961	6637,50
13,00	330	169,00	108900	4290,00
S: 107,50	S: 6669,00	S: 1176,25	S: 4640121	S: 69944,00
M: 10,75	M: 666,90			

$$r = \frac{69944 - 10 * 10,75 * 666,90}{\sqrt{1176,25 - 10 * 10,75^2} * \sqrt{4640121 - 10 * 66690^2}} = \frac{-1747,75}{1992,90} = -0,8769$$

Aufgabe VI

- a) Gesamte Schnittfläche von A1 und A2 sowie A3 komplett.
b) A2 ohne Anteile von A1 und A3 plus Ereignisraum.

Aufgabe VII

GGKKK KGKKG KGGKK KGK GK
GKGKK KKGKG KKG GK
GKKGK KKKGG
GKKKG

Insgesamt existieren 10 Möglichkeiten der Anordnung.

Aufgabe VIII

a) ausschließlich Fans = ein Pfad

$$F-F = 40/120 * 39/119 = 0,1092$$

(alternativ: 1 - 0,1092 = 0,8908)

b) mindestens ein Wiederverkäufer = drei Pfade

$$W-W = 80/120 * 79/119 = 0,4426$$

$$F-W = 40/120 * 80/119 = 0,2241$$

$$W-F = 80/120 * 40/119 = 0,2241$$

$$0,4426 + 0,2241 + 0,2241 = 0,8908$$

Aufgabe IX

Bei 100 Notebooks: 10 defekt -> 8 reklamiert

90 ok -> 4,5 reklamiert 8 zu Recht reklamiert / 12,5 gesamt reklamiert = 64%

Berechnung über den Satz von Bayes:

$$P(A_i | B) = \frac{P(B | A_i) * P(A_i)}{P(B)}$$

$$P(d | r) = \frac{P(r | d) * P(d)}{P(r)} \quad (d = \text{defekt}, r = \text{reklamiert})$$

$$P(d | r) = \frac{0,80 * 0,10}{(0,80 * 0,10) + (0,05 * 0,90)}$$

$$P(d | r) = \frac{0,08}{0,08 + 0,045} = \frac{0,08}{0,125} = 0,64$$

Aufgabe X

TABELLE: $\chi^2_{(q)}$ -Verteilung

q	$\alpha = 0.99$	$\alpha = 0.95$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.005$
1	0.000	0.004	0.016	2.706	3.841	6.635	7.879
2	0.020	0.103	0.211	4.605	5.991	9.210	10.597
3	0.115	0.352	0.584	6.251	7.815	11.345	12.838
4	0.297	0.711	1.064	7.779	9.488	13.277	14.860
5	0.554	1.145	1.610	9.236	11.070	15.086	16.750
6	0.872	1.635	2.204	10.645	12.592	16.812	18.548
7	1.239	2.167	2.833	12.017	14.067	18.475	20.278
8	1.646	2.733	3.490	13.362	15.507	20.090	21.955
9	2.088	3.325	4.168	14.684	16.919	21.666	23.589
10	2.558	3.940	4.865	15.987	18.307	23.209	25.188

Schritt 1: Berechnung der zu erwartenden Häufigkeiten bei völliger Unabhängigkeit

Ergebnis/Geschlecht	Männlich	Weiblich	Summe
Noten 1 und 2	2	6	8
Noten 3 und 4	4	3	7
Noten 5 und 6	4	1	5
Summe	10	10	20

Ergebnis/Geschlecht	Männlich	Weiblich	Summe
Noten 1 und 2	4	4	8
Noten 3 und 4	3,5	3,5	7
Noten 5 und 6	2,5	2,5	5
Summe	10	10	20

Schritt 2: Berechnung des Chi-Quadrat-Wertes

$$(2-4)^2/4=1$$

$$(6-4)^2/4=1$$

$$(4-3,5)^2/3,5=0,07$$

$$(3-3,5)^2/3,5=0,07$$

$$(4-2,5)^2/2,5=0,9$$

$$(1-2,5)^2/2,5=0,9$$

Summe: 3,94

Schritt 3: Vergleich mit dem kritischen Wert (9,21): Kritischer Wert wird nicht überschritten, daher kein Verwerfen von H_0 möglich → Unabhängigkeit wahrscheinlich

Chi-Square Tests

The observed values:

	M	W	
1-2	2	6	8
3-4	4	3	7
5-6	4	1	5
total	10	10	20

The expected values if the rows and columns are independent:

	M	W	
1-2	4,000	4,000	8
3-4	3,500	3,500	7
5-6	2,500	2,500	5
total	10	10	20

The chi-square value with 2 degrees of freedom: 3,9429

The p value: 0,1391

H_0 kleiner als krit. Wert = nicht verwerfen

Sommersemester 2019

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2019

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 25. Mai 2019

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2019

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der statistische Ersatz fehlender Werte setzt mindestens ordinalskalierte Daten voraus.

☐ richtig ☐ falsch

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

3) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert der Steigung der linearen Regressionsgeraden.

☐ richtig ☐ falsch

4) Nominalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☐ falsch

5) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

6) Die Berechnung des Bravais-Pearson-Korrelationskoeffizienten erfordert metrische Daten.

☐ richtig ☐ falsch

7) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

8) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

☐ richtig ☐ falsch

9) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

☐ richtig ☐ falsch

10) Die Fünf-Werte-Zusammenfassung besteht aus den Quartilen, Minimum und Maximum.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Zur Europawahl werden vor einem Wahllokal in Wernigerode zehn Wählerinnen und Wähler nach ihrem Wahlverhalten befragt. Bei dieser Gelegenheit werden auch einige demografische Merkmale erhoben, darunter das Alter der Befragten.

Nr.	Alter (in Jahren)
01	35
02	41
03	23
04	87
05	36
06	54
07	52
08	19
09	51
10	43

Stellen Sie für die Altersverteilung wahlweise

- a) einen erweiterten Box-Plot unter Angabe der für die Konstruktion benötigten Größen
 - b) ODER (!) ein Stamm-Blatt-Diagramm mit selbstgewählter und geeigneter Stammbreite
- auf.

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Neben der Angabe des Alters wurde im Rahmen der Wählerbefragung außerdem das monatlich individuell verfügbare Nettoeinkommen erhoben.

Nr.	Monatsnetto (in Euro)
01	1230
02	1110
03	1230
04	930
05	1890
06	2270
07	2430
08	850
09	2340
10	2150

Bestimmen Sie für die Verteilung der Einkommenswerte

- a) das arithmetische Mittel (5 Punkte),
- b) den Median (5 Punkte),
- c) den Modus (5 Punkte),
- d) die Spannweite (5 Punkte),
- e) den Interquartilsabstand (5 Punkte)
- f) und die Standardabweichung (5 Punkte).

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Im Rahmen der nächsten beiden Aufgaben soll die gemeinsame Verteilung der erhobenen Angaben zum Alter sowie zum monatlich verfügbaren Nettoeinkommen betrachtet werden.

Nr.	Alter (in Jahren)	Monatsnetto (in Euro)
01	35	1230
02	41	1110
03	23	1230
04	87	930
05	36	1890
06	54	2270
07	52	2430
08	19	850
09	51	2340
10	43	2150

Berechnen Sie den Rangkorrelationskoeffizienten nach Spearman um festzustellen, ob eine monotone Korrelation zwischen beiden Variablen existiert (15 Punkte) und interpretieren Sie das Ergebnis (5 Punkte).

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Stellen Sie für die bereits bekannte Verteilung der Alters- und Einkommensangaben ein lineares Regressionsmodell unter der Annahme auf, dass das Alter als unabhängige Variable (x) und das monatlich verfügbare Nettoeinkommen als abhängige Variable (y) zu betrachten sind (10 Punkte). Bestimmen und interpretieren Sie zudem die Güte des von Ihnen aufgestellten Modells (5 Punkte). Wie erklären Sie sich den von ihnen errechneten Grad der Streuungsaufklärung? (5 Punkte)

Nr.	Alter (in Jahren)	Monatsnetto (in Euro)
01	35	1230
02	41	1110
03	23	1230
04	87	930
05	36	1890
06	54	2270
07	52	2430
08	19	850
09	51	2340
10	43	2150

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2019

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

1 falsch	4 falsch	7 falsch	10 richtig
2 richtig	5 falsch	8 falsch	
3 falsch	6 richtig	9 richtig	

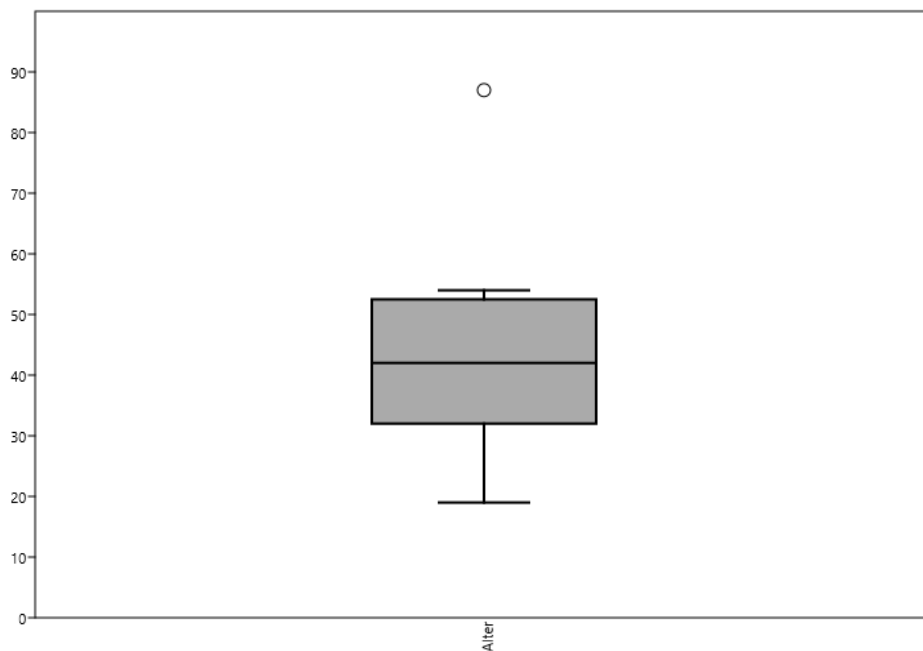
Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Lösungsvariante A: Box-Plot

Geordnete Reihe: 19; 23; 35; 36; 41; 43; 51; 52; 54; 87

 $0,25 * 10 = 2,5 \rightarrow$ Wert an 3. Stelle: 35 $0,50 * 10 = 5 \rightarrow$ Mittel der Werte an 5. und 6. Stelle: $(41 + 43) / 2 = \underline{42}$ $0,75 * 10 = 7,5 \rightarrow$ Wert an 8. Stelle: 52 $IQR = 52 - 35 = \underline{17}$ $1,5\text{-facher } IQR = 25,5$

- Die Box verläuft somit von 35 auf 52, der Median ist bei 42 einzuzeichnen.
- Der obere Zaun läuft maximal bis $52 + 25,5 = 77,5$. Der größte Nicht-Ausreißer im Datensatz und damit auch Grenze des oberen Zauns ist die 54. Der untere Zaun läuft minimal bis $35 - 25,5 = 9,5$. Der kleinste Wert im Datensatz und damit auch Grenze des unteren Zauns ist 19.
- Ausreißer nach unten kann es nicht geben, nach oben ist die 87 als Ausreißer zu markieren.



Lösungsvariante B: Stamm-Blatt-Diagramm

Geordnete Reihe: 19; 23; 35; 36; 41; 43; 51; 52; 54; 87

1 | 9
 2 | 3
 3 | 5 6
 4 | 1 3
 5 | 1 2 4

Jedes Blatt ein Wert
 Stammbreite 10
 1 Ausreißer (87)

Aufgabenteil III: Verteilungsparameter (30 Punkte)

a) Arithmetisches Mittel

$$1230 + 1110 + 1230 + 930 + 1890 + 2270 + 2430 + 850 + 2340 + 2150 = 16430$$

$$16430 / 10 = \underline{1643}$$

b) Median

Geordnete Reihe: 850; 930; 1110; 1230; 1230; 1890; 2150; 2270; 2340; 2430

$$0,50 * 10 = 5 \rightarrow \text{Mittel der Werte an 5. und 6. Stelle: } (1230 + 1890) / 2 = \underline{1560}$$

c) Modus

Hier nicht möglich, da keine unimodale Verteilung vorliegt.

d) Spannweite

$$2430 - 850 = \underline{1580}$$

e) Interquartilsabstand

$$0,25 * 10 = 2,5 \rightarrow \text{Wert an 3. Stelle: } \underline{1110}$$

$$0,75 * 10 = 7,5 \rightarrow \text{Wert an 8. Stelle: } \underline{2270}$$

$$IQR = 2270 - 1110 = \underline{1160}$$

f) Standardabweichung

Arithmetisches Mittel: 1643

$(1230 - 1643)^2 = 170569$	$(2270 - 1643)^2 = 393129$
$(1110 - 1643)^2 = 284089$	$(2430 - 1643)^2 = 619369$
$(1230 - 1643)^2 = 170569$	$(850 - 1643)^2 = 628849$
$(930 - 1643)^2 = 508369$	$(2340 - 1643)^2 = 485809$
$(1890 - 1643)^2 = 61009$	$(2150 - 1643)^2 = 257049$

Summe aller quadrierten Differenzen: 3578810

$$\text{Varianz: } 3578810 / 10 = \underline{357881}$$

$$\text{Standardabweichung: } \underline{598,23}$$

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Nr.	x	rg (x)	y	rg (y)	d	d ²
01	35	3	1230	4,5	-1,5	2,25
02	41	5	1110	3	2	4
03	23	2	1230	4,5	-2,5	6,25
04	87	10	930	2	8	64
05	36	4	1890	6	-2	4
06	54	9	2270	8	1	1
07	52	8	2430	10	-2	4
08	19	1	850	1	0	0
09	51	7	2340	9	-2	4
10	43	6	2150	7	-1	1

Summe aller quadrierten Abstände: 90,5

$$rho = 1 - \frac{6 * 90,5}{(10^2 - 1) * 10} = 1 - \frac{543}{99 * 10} = 1 - \frac{543}{990} = 1 - 0,548 = 0,452$$

Es liegt eine schwache, positive und monotone Korrelation vor.

Aufgabenteil V: Lineare Regressionsanalyse (20 Punkte)

Nr.	x	x ²	y	y ²	(x * y)
01	35	1225	1230	1512900	43050
02	41	1681	1110	1232100	45510
03	23	529	1230	1512900	28290
04	87	7569	930	864900	80910
05	36	1296	1890	3572100	68040
06	54	2916	2270	5152900	122580
07	52	2704	2430	5904900	126360
08	19	361	850	722500	16150
09	51	2601	2340	5475600	119340
10	43	1849	2150	4622500	92450
SUM	441	22731	16430	30573300	742680
MEAN	44,1	-//-	1643	-//-	-//-

Aufgabenteil A: Bestimmung der Regressionsfunktion

$$b = \frac{742680 - 10 * 44,1 * 1643}{22731 - 10 * 1944,81} = \frac{18117}{3282,9} = 5,52$$

$$a = 1643 - 5,52 * 44,1 = 1399,57$$

$$y = 1399,57 + 5,52 * x$$

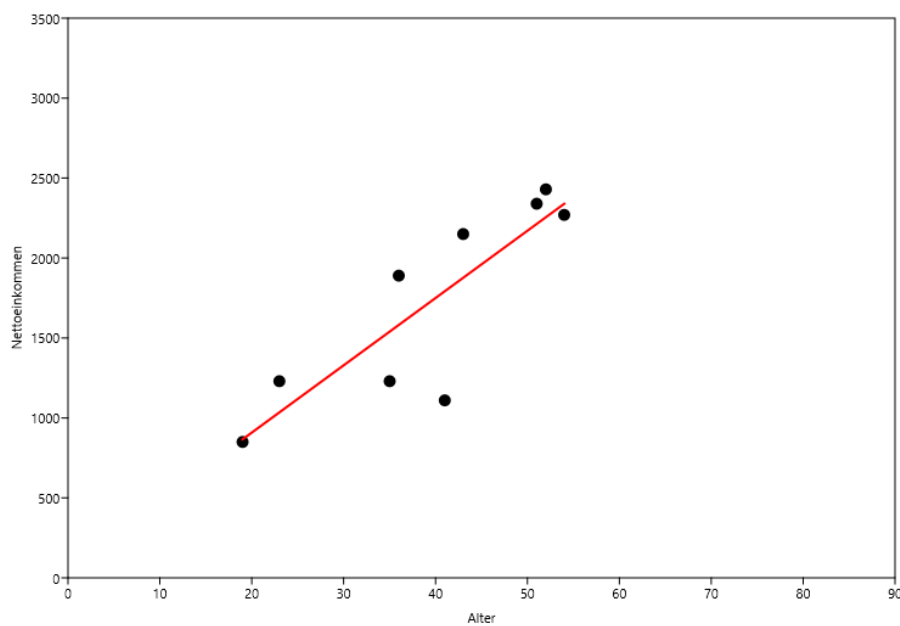
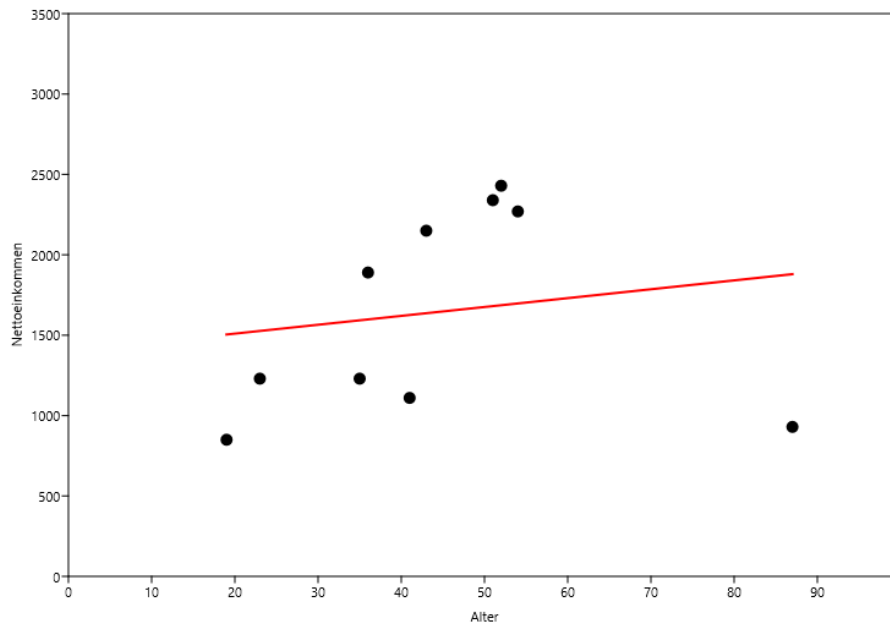
Aufgabenteil B: Bewertung der Regressionsgüte

$$r = \frac{742680 - 10 * 44,1 * 1643}{\sqrt{22731 - 10 * 1944,81} * \sqrt{30573300 - 10 * 2699449}} = \frac{18117}{\sqrt{3282,9} * \sqrt{3578810}}$$

$$= \frac{18117}{57,3 * 1891,77} = 0,1671$$

$$R^2 = 0,1671^2 = \underline{0,028}$$

Das Modell verfügt – durchaus kontraintuitiv – über nahezu keine Streuungsaufklärung. Grund hierfür ist der Leverage-Effekt (Ausreißer ist Befragter Nr. 4). Tatsächlich verbessert sich R^2 durch die Entfernung des Ausreißers von 0,028 auf 0,728.



Sommersemester 2019

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2019

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 13. Juli 2019

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2019

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der statistische Ersatz fehlender Werte setzt mindestens metrisch skalierte Daten voraus.

☐ richtig ☐ falsch

2) Kreisdiagramme eignen sich eher für stetige als für diskrete Daten.

☐ richtig ☐ falsch

3) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

4) Variablen mit nur zwei möglichen Ausprägungen werden als dichotom bezeichnet.

☐ richtig ☐ falsch5) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsen Schnittpunkts der linearen Regressionsgeraden.☐ richtig ☐ falsch

6) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

☐ richtig ☐ falsch

7) Eine repräsentative Zufallsauswahl setzt die genaue Definition der Grundgesamtheit voraus.

☐ richtig ☐ falsch

8) Verhältnisskala und Intervallskala unterscheiden sich durch den natürlichen Nullpunkt.

☐ richtig ☐ falsch

9) Die Fünf-Werte-Zusammenfassung kann nur für metrisch skalierte Werte berechnet werden.

☐ richtig ☐ falsch

10) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Bei einem Automobilzulieferer werden Metallspritzgussbauteile für den Motorbereich hergestellt. Aus einer Produktionscharge werden zufällig zehn solcher Bauteile entnommen und gewogen.

Bauteil Nr.	Gewicht (in g)
01	17,3
02	18,2
03	13,2
04	12,1
05	19,6
06	18,1
07	16,5
08	11,4
09	17,7
10	14,5

Zeichnen Sie einen Box-Plot (8 Punkte) und geben Sie die für dessen Konstruktion benötigten Größen an (2 Punkte).

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Der vor einer Woche neu eingerichtete Online-Shop einer Apotheke verzeichnete während dieser sieben Tage folgende Bestellsummen:

Bauteil Nr.	Summe aller Bestellungen (in EUR, gerundet)
Montag	136
Dienstag	142
Mittwoch	123
Donnerstag	486
Freitag	221
Samstag	185
Sonntag	87

Bestimmen Sie folgende Verteilungsparameter (jeweils 3 Punkte):

- a) Arithmetisches Mittel
- b) Modus
- c) Spannweite
- d) Standardabweichung
- e) Variationskoeffizient

Aufgabenteil IV: Zusammenhangsmaße (15 Punkte)

Eine Fast Food-Filiale bietet einen neuen Burger an und sucht für diesen noch nach einem rentablen und gleichzeitig von der Zielgruppe akzeptierten Zielpreis. Aus diesem Grund wird der Preis an zehn aufeinanderfolgenden Tagen (ceteris paribus) verändert und die Absatzmenge gemessen.

Tag	Preis (in EUR)	Absatzmenge (in Stück)
01	3,50	143
02	4,00	128
03	4,50	123
04	5,00	116
05	5,50	128
06	6,00	119
07	6,50	95
08	7,00	83
09	7,50	51
10	8,00	42

Korrelieren Preis und Absatzmenge – und wenn ja, in welcher Form? Berechnen und interpretieren Sie zur Beantwortung dieser Frage den Konkordanzkoeffizienten nach Kendall.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Im Rahmen eines statistischen Testverfahrens gibt der p-Wert / Signifikanzwert die Wahrscheinlichkeit dafür wieder, dass das Zurückweisen einer Nullhypothese falsch ist.

☐ richtig ☐ falsch

2) Die Wahrscheinlichkeit für fehlerhafte Testergebnisse nimmt mit der Anzahl durchgeführter Tests zu. Dieses Phänomen wird auch als alpha-Fehlerinflation bezeichnet.

☐ richtig ☐ falsch

3) Ein Bernoulli-Experiment kennt lediglich zwei Ausgänge (ja/nein bzw. 1/0).

☐ richtig ☐ falsch

4) Kann die Nullhypothese eines statistischen Tests nicht verworfen werden, gilt die Alternativhypothese damit als widerlegt.

☐ richtig ☐ falsch

5) Eine gültige Teilmenge muss stets aus mindestens zwei Elementen bestehen.

☐ richtig ☐ falsch

6) Eine Stichprobe ist nur dann repräsentativ, wenn sie exakt den mit Hilfe der Cochran-Formel bestimmten Umfang aufweist.

☐ richtig ☐ falsch

7) Das logische UND wird auch als Disjunktion bezeichnet.

☐ richtig ☐ falsch

8) Je höher das Konfidenzniveau eines Konfidenzintervalls ist, umso wahrscheinlicher ist, dass es den realen Wert in der Grundgesamtheit einschließt.

☐ richtig ☐ falsch

9) Ein Würfel wird dann als fair bezeichnet, wenn die Wahrscheinlichkeit des Wurfs für jede Augenzahl genau gleich groß ist.

☐ richtig ☐ falsch

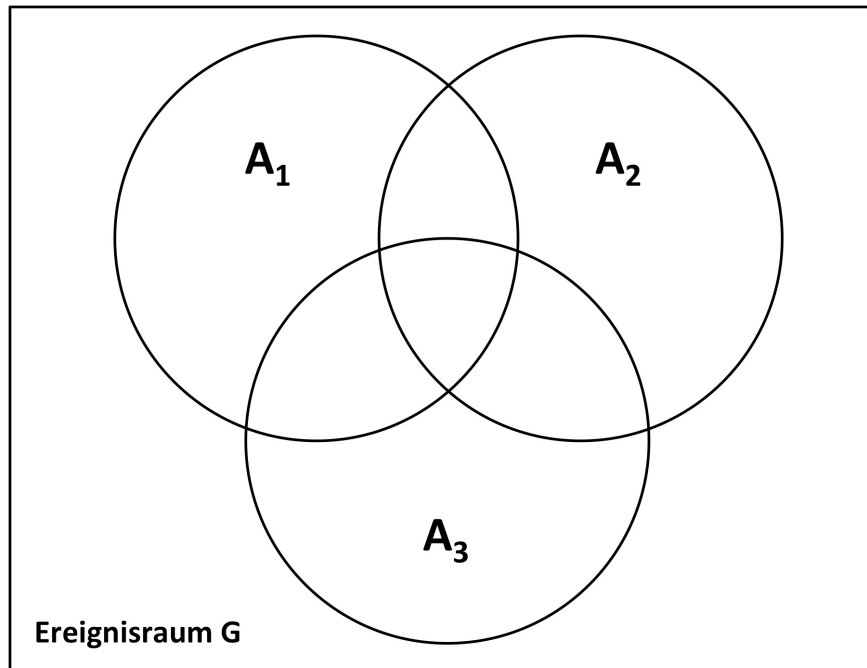
10) Das Konfidenzniveau eines Konfidenzintervalls muss immer vor der Aufstellung des Intervalls festgelegt und darf keinesfalls nachträglich an das Ergebnis angepasst werden.

☐ richtig ☐ falsch

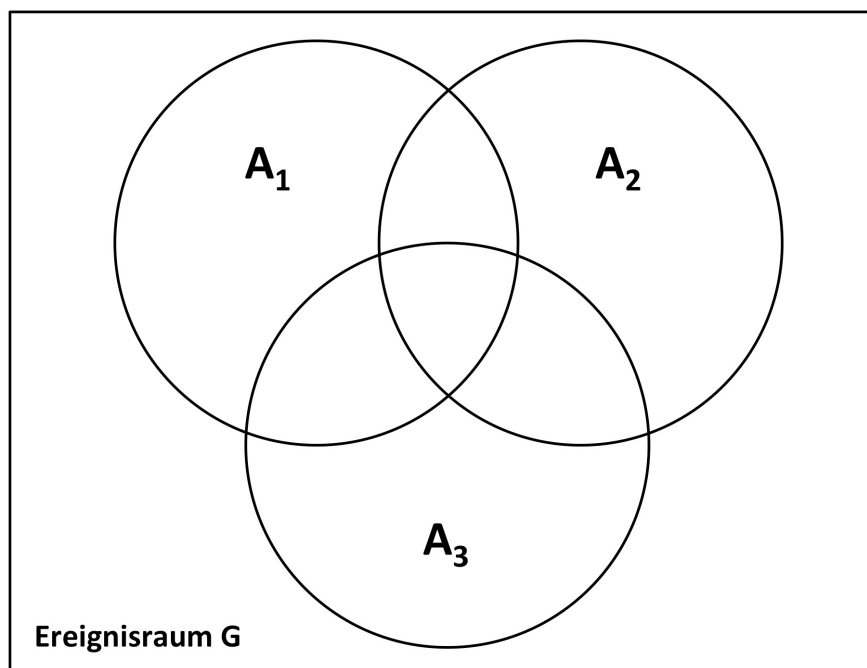
Aufgabenteil VI: Mengenlehre (6 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 3 Punkte)

a) $(A_1 \cap A_2 \cap A_3) \cup A_3$



b) $(A_3 \cup A_3)$



Aufgabenteil VII: Kombinatorik (9 Punkte)

Das Passwort für den Zugang zu einem Online-Banking-System kann sich beliebig aus den 26 Großbuchstaben von A bis Z, den 26 Kleinbuchstaben von a bis z sowie den 10 Ziffern von 0 bis 9 zusammensetzen. Mögliche Kombinationen wären damit beispielsweise:

De3zf564sz

HTDD675eMX

343GZW24yxx3

Wie viele unterschiedliche Passwörter können generiert werden, wenn:

- a) Die Länge des Passworts unbekannt ist? (3 Punkte)
- b) Bekannt ist, dass das Passwort aus 5 Stellen besteht? (3 Punkte)
- c) Bekannt ist, dass es mit zwei aufeinanderfolgenden „y“ endet? (3 Punkte)

Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

In einem Betrieb wird in zwei aufeinanderfolgenden Jahren jeweils eine Betriebsprüfung durch das Finanzamt vorgenommen. Angenommen die Wahrscheinlichkeit dafür, im Rahmen einer solchen Prüfung auf Unregelmäßigkeiten oder Formfehler zu stoßen, läge bei 20% - wie groß ist dann die Wahrscheinlichkeit dafür, dass mindestens eine der Prüfungen zu einer Nachzahlung führt?

Aufgabenteil IX: Bedingte Wahrscheinlichkeiten (10 Punkte)

Eine Softwarefirma arbeitet an einem neuartigen Virens Scanner für E-Mails, welcher die Anhänge von E-Mails durchsucht, bevor diese durch den Anwender geöffnet werden können. Der aktuelle Prototyp dieser Software erkennt einen virenverseuchten Anhang bereits mit einer sehr hohen Sicherheit von 97,5% - stuft allerdings auch noch 6% der unkritischen Anhänge fälschlicherweise als virenbelastet ein. Es wird davon ausgegangen, dass etwa 3% aller insgesamt mit Anhängen verschickten E-Mails einen Virus transportieren.

Vor einer Entscheidung über die Marktreife der Software möchte der CEO wissen, mit welcher Wahrscheinlichkeit ein späterer Anwender davon ausgehen kann, dass ein als virenverseucht gekennzeichnete E-Mail-Anhang auch tatsächlich einen Virus enthält.

Berechnen Sie diese Wahrscheinlichkeit mit Hilfe des Satzes von Bayes.

Aufgabenteil X: Chi-Quadrat-Unabhängigkeitstest (10 Punkte)

Ein Online-Versandhandel testet ein neues Abo-Modell, bei dem Kundinnen und Kunden in regelmäßigen Abständen eine Auswahl von hochpreisigen Lebensmitteln zugeschickt wird. Das Unternehmen unterscheidet intern in A-Kunden und B-Kunden, wobei sich A-Kunden durch eine höhere Kaufbereitschaft und Markentreue auszeichnen und daher von wirtschaftlich größerem Interesse sind. Ein Werbeangebot für das neue Abo-Modell wurde an 100 zufällig ausgewählte Kundinnen und Kunden verschickt, die zu etwa gleichen Teilen der A- und B-Kundschaft zuzuordnen sind. Dieser Versuch führte zu folgenden Ergebnissen:

	Abo abgeschlossen	Kein Abo abgeschlossen
A-Kunden	36	17
B-Kunden	19	28

Ermitteln Sie mit Hilfe des Chi-Quadrat-Tests auf stochastische Unabhängigkeit bei einem Fehlniveau α von 5% (kritischer Wert: 3,84), ob ein Zusammenhang zwischen der Kundeneinstufung und der Kaufbereitschaft zu vermuten ist.

MUSERLÖSUNG KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2019

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der statistische Ersatz fehlender Werte setzt mindestens metrisch skalierte Daten voraus.

X richtig ☐ falsch

2) Kreisdiagramme eignen sich eher für stetige als für diskrete Daten.

☐ richtig **X falsch**

3) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig **X falsch**

4) Variablen mit nur zwei möglichen Ausprägungen werden als dichotom bezeichnet.

X richtig ☐ falsch

5) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsenschnittpunkts der linearen Regressionsgeraden.

☐ richtig **X falsch**

6) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

X richtig ☐ falsch

7) Eine repräsentative Zufallsauswahl setzt die genaue Definition der Grundgesamtheit voraus.

X richtig ☐ falsch

8) Verhältnisskala und Intervallskala unterscheiden sich durch den natürlichen Nullpunkt.

X richtig ☐ falsch

9) Die Fünf-Werte-Zusammenfassung kann nur für metrisch skalierte Werte berechnet werden.

☐ richtig **X falsch**

10) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig **X falsch**

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Bei einem Automobilzulieferer werden Metallspritzgussbauteile für den Motorbereich hergestellt. Aus einer Produktionscharge werden zufällig zehn solcher Bauteile entnommen und gewogen.

Bauteil Nr.	Gewicht (in g)
01	17,3
02	18,2
03	13,2
04	12,1
05	19,6
06	18,1
07	16,5
08	11,4
09	17,7
10	14,5

Zeichnen Sie einen Box-Plot (8 Punkte) und geben Sie die für dessen Konstruktion benötigten Größen an (2 Punkte).

Geordnete Reihenfolge

11,4; 12,1; 13,2; 14,5; 16,5; 17,3; 17,7; 18,1; 18,2; 19,6

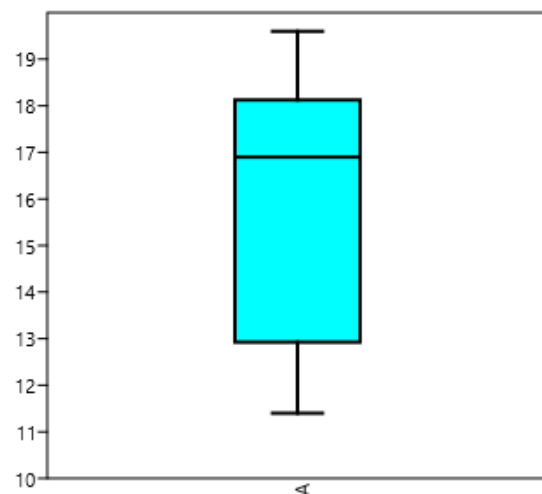
Berechnung der erforderlichen Größen

Unteres Quartil:	$0,25 \cdot 10 = 2,5$	Wert an 3. Stelle	13,2
Mittleres Quartil:	$0,50 \cdot 10 = 5,0$	Mittel der Werte an 5. und 6. Stelle	16,9
Oberes Quartil:	$0,75 \cdot 10 = 7,5$	Wert an 8. Stelle	18,1

Interquartilsabstand	4,9
1,5-facher Interquartilsabstand	7,35

Unterer Zaun läuft minimal bis zu	5,85
Unterer Zaun läuft tatsächlich bis zu	11,4
Oberer Zaun läuft maximal bis zu	25,45
Oberer Zaun läuft tatsächlich bis zu	19,6

Keine Ausreißer, keine Extremwerte



Aufgabenteil III: Verteilungsparameter (15 Punkte)

Der vor einer Woche neu eingerichtete Online-Shop einer Apotheke verzeichnete während dieser sieben Tage folgende Bestellsummen:

Bauteil Nr.	Summe aller Bestellungen (in EUR, gerundet)
Montag	136
Dienstag	142
Mittwoch	123
Donnerstag	486
Freitag	221
Samstag	185
Sonntag	87

Bestimmen Sie folgende Verteilungsparameter (jeweils 3 Punkte):

- a) Arithmetisches Mittel
- b) Modus
- c) Spannweite
- d) Standardabweichung
- e) Variationskoeffizient

Summe aller Werte: 1380

Arithmetisches Mittel: $1380 / 7 = 197,14$

Modus: unbestimmbar

Spannweite: $486 - 87 = 399$

$$(136 - 197,14)^2 = 3738,10$$

$$(142 - 197,14)^2 = 3040,42$$

$$(123 - 197,14)^2 = 5496,74$$

$$(486 - 197,14)^2 = 83440,10$$

$$(221 - 197,14)^2 = 569,30$$

$$(185 - 197,14)^2 = 147,38$$

$$(87 - 197,14)^2 = 12130,82$$

Varianz: $108562,86 / 7 = 15508,98$

Standardabweichung: Wurzel der Varianz = 124,54

Variationskoeffizient: $124,54 / 197,14 = 0,63$

Aufgabenteil IV: Zusammenhangsmaße (15 Punkte)

Eine Fast Food-Filiale bietet einen neuen Burger an und sucht für diesen noch nach einem rentablen und gleichzeitig von der Zielgruppe akzeptierten Zielpreis. Aus diesem Grund wird der Preis an zehn aufeinanderfolgenden Tagen (ceteris paribus) verändert und die Absatzmenge gemessen.

Tag	Preis (in EUR)	Absatzmenge (in Stück)
01	3,50	143
02	4,00	128
03	4,50	123
04	5,00	116
05	5,50	128
06	6,00	119
07	6,50	95
08	7,00	83
09	7,50	51
10	8,00	42

Korrelieren Preis und Absatzmenge – und wenn ja, in welcher Form? Berechnen und interpretieren Sie zur Beantwortung dieser Frage den Konkordanzkoeffizienten nach Kendall.

x	Rang x	y	Rang y	K	D
3,5	1,0	143	10,0	0	9
4,0	2,0	128	8,5	0	7
4,5	3,0	123	7,0	1	6
5,0	4,0	116	5,0	2	4
5,5	5,0	128	8,5	0	5
6,0	6,0	119	6,0	0	4
6,5	7,0	95	4,0	0	3
7,0	8,0	83	3,0	0	2
7,5	9,0	51	2,0	0	1
8,0	10,0	42	1,0	0	0

Summe der konkordanten Paare: 3

Summe der diskordanten Paare: 41

$$\tau = (2 \cdot (3 - 41)) / (10 \cdot (10 - 1)) = -0,84$$

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Im Rahmen eines statistischen Testverfahrens gibt der p-Wert / Signifikanzwert die Wahrscheinlichkeit dafür wieder, dass das Zurückweisen einer Nullhypothese falsch ist.

☐ richtig ☒ falsch

2) Die Wahrscheinlichkeit für fehlerhafte Testergebnisse nimmt mit der Anzahl durchgeführter Tests zu. Dieses Phänomen wird auch als alpha-Fehlerinflation bezeichnet.

☒ richtig ☐ falsch

3) Ein Bernoulli-Experiment kennt lediglich zwei Ausgänge (ja/nein bzw. 1/0).

☒ richtig ☐ falsch

4) Kann die Nullhypothese eines statistischen Tests nicht verworfen werden, gilt die Alternativhypothese damit als widerlegt.

☐ richtig ☒ falsch

5) Eine gültige Teilmenge muss stets aus mindestens zwei Elementen bestehen.

☐ richtig ☒ falsch

6) Eine Stichprobe ist nur dann repräsentativ, wenn sie exakt den mit Hilfe der Cochran-Formel bestimmten Umfang aufweist.

☐ richtig ☒ falsch

7) Das logische UND wird auch als Disjunktion bezeichnet.

☐ richtig ☒ falsch

8) Je höher das Konfidenzniveau eines Konfidenzintervalls ist, umso wahrscheinlicher ist, dass es den realen Wert in der Grundgesamtheit einschließt.

☒ richtig ☐ falsch

9) Ein Würfel wird dann als fair bezeichnet, wenn die Wahrscheinlichkeit des Wurfs für jede Augenzahl genau gleich groß ist.

☒ richtig ☐ falsch

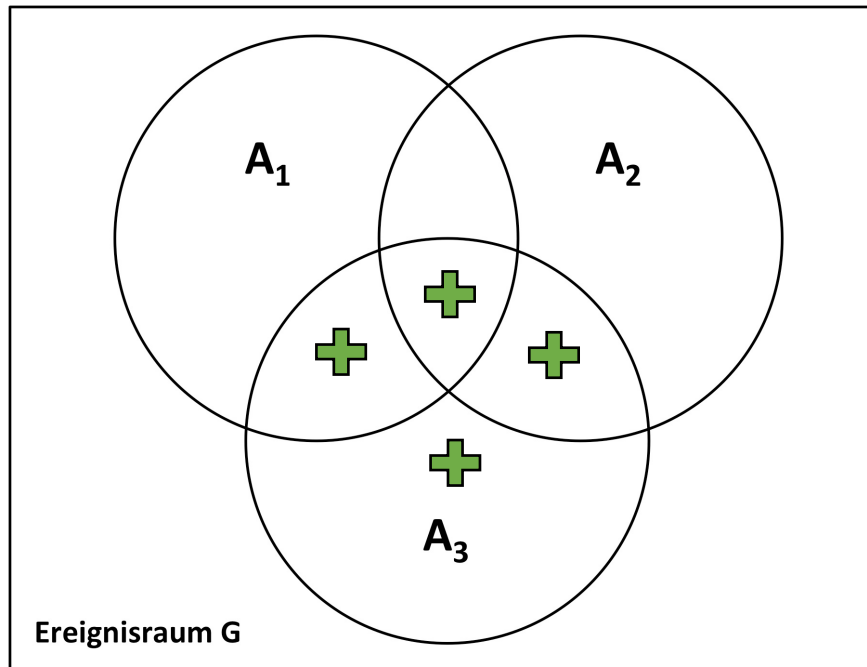
10) Das Konfidenzniveau eines Konfidenzintervalls muss immer vor der Aufstellung des Intervalls festgelegt und darf keinesfalls nachträglich an das Ergebnis angepasst werden.

☒ richtig ☐ falsch

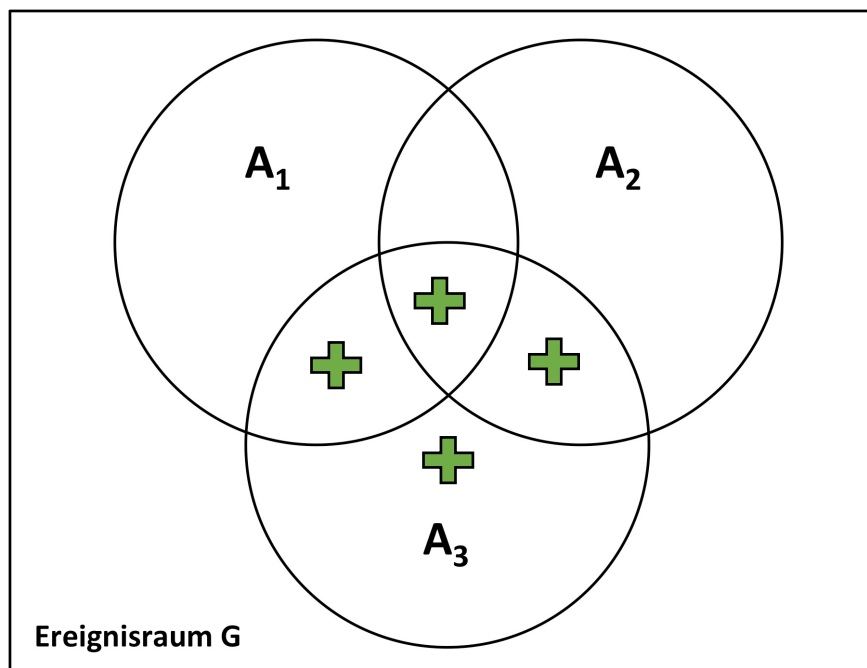
Aufgabenteil VI: Mengenlehre (6 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 3 Punkte)

a) $(A_1 \cap A_2 \cap A_3) \cup A_3 \rightarrow A_3$ komplett



b) $(A_3 \cup A_3) \rightarrow A_3$ komplett



Aufgabenteil VII: Kombinatorik (9 Punkte)

Das Passwort für den Zugang zu einem Online-Banking-System kann sich beliebig aus den 26 Großbuchstaben von A bis Z, den 26 Kleinbuchstaben von a bis z sowie den 10 Ziffern von 0 bis 9 zusammensetzen. Mögliche Kombinationen wären damit beispielsweise:

De3zf564sz

HTDD675eMX

343GZW24yxx3

Wie viele unterschiedliche Passwörter können generiert werden, wenn:

- a) Die Länge des Passworts unbekannt ist? (3 Punkte)
- b) Bekannt ist, dass das Passwort aus 5 Stellen besteht? (3 Punkte)
- c) Bekannt ist, dass es mit zwei aufeinanderfolgenden „y“ endet? (3 Punkte)

Grundsätzlich handelt es sich um eine Variation mit Zurücklegen.

- a) Kann nicht berechnet werden.
- b) $62^5 = 916.132.832$
- c) $62^3 = 238.328$

Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

In einem Betrieb wird in zwei aufeinanderfolgenden Jahren jeweils eine Betriebsprüfung durch das Finanzamt vorgenommen. Angenommen die Wahrscheinlichkeit dafür, im Rahmen einer solchen Prüfung auf Unregelmäßigkeiten oder Formfehler zu stoßen, läge bei 20% - wie groß ist dann die Wahrscheinlichkeit dafür, dass mindestens eine der Prüfungen zu einer Nachzahlung führt?

Wahrscheinlichkeit für gar keinen Fund: $0,8 * 0,8 = 0,64$

Wahrscheinlichkeit für mindestens einen Fund: $1 - 0,64 = 0,36$

Aufgabenteil IX: Bedingte Wahrscheinlichkeiten (10 Punkte)

Eine Softwarefirma arbeitet an einem neuartigen Virenschanner für E-Mails, welcher die Anhänge von E-Mails durchsucht, bevor diese durch den Anwender geöffnet werden können. Der aktuelle Prototyp dieser Software erkennt einen virenverseuchten Anhang bereits mit einer sehr hohen Sicherheit von 97,5% - stuft allerdings auch noch 6% der unkritischen Anhänge fälschlicherweise als virenbelastet ein. Es wird davon ausgegangen, dass etwa 3% aller insgesamt mit Anhängen verschickten E-Mails einen Virus transportieren.

Vor einer Entscheidung über die Marktreife der Software möchte der CEO wissen, mit welcher Wahrscheinlichkeit ein späterer Anwender davon ausgehen kann, dass ein als virenverseucht gekennzeichnete E-Mail-Anhang auch tatsächlich einen Virus enthält.

Berechnen Sie diese Wahrscheinlichkeit mit Hilfe des Satzes von Bayes.

E-Mails	100
davon mit Viren	3
davon ohne Viren	97
E-Mails mit Viren	3
davon entdeckt: 97,5%	2,925
davon unentdeckt: 2,5%	0,075
E-Mails ohne Viren	97
davon "entdeckt": 6%	5,82
davon unentdeckt: 94%	91,18
Insgesamt entdeckte E-Mails	8,745
davon tatsächlich mit Viren	2,925
Trefferwahrscheinlichkeit	0,334

Aufgabenteil X: Chi-Quadrat-Unabhängigkeitstest (10 Punkte)

Ein Online-Versandhandel testet ein neues Abo-Modell, bei dem Kundinnen und Kunden in regelmäßigen Abständen eine Auswahl von hochpreisigen Lebensmitteln zugeschickt wird. Das Unternehmen unterscheidet intern in A-Kunden und B-Kunden, wobei sich A-Kunden durch eine höhere Kaufbereitschaft und Markentreue auszeichnen und daher von wirtschaftlich größerem Interesse sind. Ein Werbeangebot für das neue Abo-Modell wurde an 100 zufällig ausgewählte Kundinnen und Kunden verschickt, die zu etwa gleichen Teilen der A- und B-Kundschaft zuzuordnen sind. Dieser Versuch führte zu folgenden Ergebnissen:

	Abo abgeschlossen	Kein Abo abgeschlossen
A-Kunden	36	17
B-Kunden	19	28

Ermitteln Sie mit Hilfe des Chi-Quadrat-Tests auf stochastische Unabhängigkeit bei einem Fehlerniveau α von 5% (kritischer Wert: 3,84), ob ein Zusammenhang zwischen der Kundeneinstufung und der Kaufbereitschaft zu vermuten ist.

Berechnung der Zeilen- und Spaltensummen

	Abo	Kein Abo	
A-Kunden	36	17	53
B-Kunden	19	28	47
	55	45	100

Berechnung der zu erwartenden Wahrscheinlichkeiten

	Abo	Kein Abo
A-Kunden	29,15	23,85
B-Kunden	25,85	21,15

Berechnung des Chi-Quadrat-Wertes

$$(36 - 29,15)^2 / 29,15 = 1,609691252$$

$$(19 - 25,85)^2 / 25,85 = 1,815183752$$

$$(17 - 23,85)^2 / 23,85 = 1,967400419$$

$$(28 - 21,15)^2 / 21,15 = 2,21855792$$

Summe: 7,610833343

Kritischer Wert: 3,84

Die Nullhypothese kann verworfen werden – ein Zusammenhang ist wahrscheinlich.

Wintersemester 2019/2020

Klausur Methodenlehre

Hinweis: Bei der Klausur Methodenlehre im Wintersemester 2019/2020 handelte es sich um eine Prüfung, die – auf Wunsch der Studierenden – im Statistik-Labor unter Einsatz der Software R abgelegt wurde. Um sie zu replizieren, muss zunächst R installiert (<https://www.r-project.org/>) und anschließend der Datensatz neu initialisiert oder heruntergeladen werden. Die entsprechende Datei steht (aber sicher nicht für immer) via [https://www.hs-harz.de/user-mounts/322_m1036/dateien/Klausur I_17.06.2017_PAST-Datei.dat](https://www.hs-harz.de/user-mounts/322_m1036/dateien/Klausur_I_17.06.2017_PAST-Datei.dat) zum Download zur Verfügung.

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
Master-Studiengang Business Consulting
Wintersemester 2019/2020

Prüfungsfach:

Wissenschaftliche Methodenlehre / Statistik

Name des Prüfers: Christian Reinboth

Prüfungstag: 14.01.2020

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen. Dies gilt selbstverständlich nicht für den Rechner, an dem die Prüfung abgelegt wird.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: PC/eigener Laptop mit PAST und R/RStudio.
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars. Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR METHODENLEHRE

Master-Studiengang Business Consulting
Wintersemester 2019/2020
Christian Reinboth

Aufgabenteil I: Theorie (20 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (2 Punkte pro korrekter Beantwortung)

1) Die Berechnung des Rangkorrelationskoeffizienten nach Spearman erfordert metrische Daten.

☐ richtig ☐ falsch

2) Zwei Teilgesamtheiten einer Grundgesamtheit können sich beliebig überschneiden.

☐ richtig ☐ falsch

3) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☐ falsch

4) Der Regressionskoeffizient (b) der einfachen linearen Regressionsanalyse entspricht dem Wert des Y-Achsenschnittpunkts der linearen Regressionsgeraden.

☐ richtig ☐ falsch

5) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

6) Je mehr Hypothesen man an einem Datensatz testet, desto höher wird die Wahrscheinlichkeit, dass eine davon fehlerhaft als zutreffend angenommen wird.

☐ richtig ☐ falsch

7) Beim Chi-Quadrat-Unabhängigkeitstest handelt es sich um einen Test auf stochastische Unabhängigkeit zweier nominal skalierten Merkmale.

☐ richtig ☐ falsch

8) Eine kommutative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

☐ richtig ☐ falsch

9) Je höher das Konfidenzniveau eines Konfidenzintervalls ist, umso wahrscheinlicher ist, dass es den realen Wert in der Grundgesamtheit einschließt.

☐ richtig ☐ falsch

10) Kann die Nullhypothese eines statistischen Tests nicht verworfen werden, gilt die Alternativhypothese damit als widerlegt.

☐ richtig ☐ falsch

Öffnen Sie zur Bearbeitung der nachfolgenden Aufgaben bitte die Datei mit der Bezeichnung „Statistiklausur vom 14.01.2020“ in R/RStudio (R-Datei) oder in PAST (DAT-Datei).

Aufgabenteil II: Analyse eines Datensatzes (80 Punkte)

Der vorliegende Datensatz enthält die Daten von 20 Studierenden aus einem Statistik-Kurs. Diese wurden vor dem Ablegen einer Statistiklausur befragt, wie viele Stunden Zeit sie (ungefähr) in die Vorbereitung auf diese Klausur investiert haben. Im Datensatz erfasst wurden außerdem die in der Klausur erreichte Punktzahl (minimal 0, maximal 100) sowie Geschlecht und Alter der Studierenden.

Beantworten Sie zu diesem Datensatz (unter Zuhilfenahme von Software) die nachfolgenden Fragen.

1) Verfügt die Variable „Geschlecht“ über einen Modus – und falls ja, wie lautet dieser? (5 Punkte)

2) Bestimmen Sie das arithmetische Mittel und den Median der Altersverteilung. (5 Punkte)

Arithmetisches Mittel	
Median	

3) Benennen und bestimmen Sie alle Größen, die für die Konstruktion eines erweiterten Box-Plots der Punkteverteilung benötigt würden. (10 Punkte)

4) Weist dieser Box-Plot einen oder mehrere Ausreißer auf? Falls ja, um welchen Wert bzw. um welche Werte handelt es sich? (5 Punkte)

5) Bestimmen Sie Spannweite, Varianz und Standardabweichung der Punkteverteilung. (5 Punkte)

Spannweite	
Varianz	
Standardabweichung	

6) Besteht eine monotone Korrelation zwischen der in die Klausurvorbereitung investierten Zeit und den erreichten Punkten? Bestimmen Sie zur Beantwortung dieser Frage einen geeigneten Korrelationskoeffizienten und interpretieren Sie diesen. (10 Punkte)

7) Stellen Sie ein lineares Regressionsmodell mit den erreichten Punkten als abhängiger und den in die Vorbereitung der Klausur investierten Stunden als unabhängiger Variable auf und bewerten Sie die Güte dieses Modells. Welche Punktzahl wäre bei einer Vorbereitungszeit von 10 Stunden zu erwarten? (15 Punkte)

8) Welche Punktzahl wäre bei einer Vorbereitungszeit von 40 Stunden zu erwarten? Welche Schlussfolgerungen über das Modell ergeben sich daraus? (10 Punkte)

9) Studierende, die weniger als 50 Punkte erreicht haben, haben den Test nicht bestanden. Wie viele Stunden Vorbereitungszeit hat die Subgruppe dieser Studierenden durchschnittlich in die Klausurvorbereitung investiert? (10 Punkte)

10) Wie viele Stunden Vorbereitungszeit haben die Top 3-Studierenden durchschnittlich in die Klausurvorbereitung investiert? (5 Punkte)

KLAUSUR METHODENLEHRE - KURZLÖSUNGEN

Master-Studiengang Business Consulting

Wintersemester 2019/2020

Christian Reinboth

1) Verfügt die Variable „Geschlecht“ über einen Modus – und falls ja, wie lautet dieser?

```
table(Statistiklausur$Geschlecht)
barplot(table(Statistiklausur$Geschlecht))
```

Nein.

2) Bestimmen Sie das arithmetische Mittel und den Median der Altersverteilung.

```
mean(Statistiklausur$Alter)
23.1
```

```
median(Statistiklausur$Alter)
23
```

3) Benennen und bestimmen Sie alle Größen, die für die Konstruktion eines erweiterten Box-Plots der Punkteverteilung benötigt würden.

```
quantile(Statistiklausur$Punkte,0.25)
53
```

```
quantile(Statistiklausur$Punkte,0.50)
66,5
```

```
quantile(Statistiklausur$Punkte,0.75)
77
```

```
IQR(Statistiklausur$Punkte)
24
```

```
summary(Statistiklausur$Punkte)
```

4) Weist dieser Box-Plot einen oder mehrere Ausreißer auf? Falls ja, um welchen Wert bzw. um welche Werte handelt es sich?

```
boxplot(Statistiklausur$Punkte)
Nein.
```

5) Bestimmen Sie Spannweite, Varianz und Standardabweichung der Punkteverteilung.

```
max(Statistiklausur$Punkte)-min(Statistiklausur$Punkte)
57
```

```
var(Statistiklausur$Punkte)
299,21
```

```
sd(Statistiklausur$Punkte)
17,30
```

6) Besteht eine monotone Korrelation zwischen der in die Klausurvorbereitung investierte Zeit und den erreichten Punkten? Bestimmen Sie zur Beantwortung dieser Frage einen geeigneten Korrelationskoeffizienten und interpretieren Sie diesen.

```
cor(Statistiklausur$Vorbereitung,Statistiklausur$Punkte,method="spearman")  
0,93
```

```
cor(Statistiklausur$Vorbereitung,Statistiklausur$Punkte,method="kendall")  
0,83
```

7) Stellen Sie ein lineares Regressionsmodell mit den erreichten Punkten als abhängiger und den in die Vorbereitung der Klausur investierten Stunden als unabhängiger Variable auf und bewerten Sie die Güte dieses Modells. Welche Punktzahl wäre bei einer Vorbereitungszeit von 10 Stunden zu erwarten?

```
cor(Statistiklausur$Vorbereitung,Statistiklausur$Punkte,method="pearson")  
0,93
```

```
Modell<-lm(Statistiklausur$Punkte~Statistiklausur$Vorbereitung)  
summary(Modell)
```

```
Punkte = 42,5701 + 1,7542 * Vorbereitung  
42.5701+1.7542*10 = 60,11
```

8) Welche Punktzahl wäre bei einer Vorbereitungszeit von 40 Stunden zu erwarten? Welche Schlussfolgerungen über das Modell ergeben sich daraus?

```
42.5701+1.7542*40 = 112,7381
```

Das Modell ist offensichtlich nur in einem bestimmten Wertebereich gültig und somit teillinear.

9) Studierende, die weniger als 50 Punkte erreicht haben, haben den Test nicht bestanden. Wie viele Stunden Vorbereitungszeit hat diese Subgruppe der Studierenden durchschnittlich in die Klausurvorbereitung investiert?

```
Auswahl<-Statistiklausur$Vorbereitung[Statistiklausur$Punkte<50]  
mean(Auswahl)  
2
```

10) Wie viele Stunden Vorbereitungszeit haben die Top 3-Studierenden durchschnittlich in die Klausurvorbereitung investiert?

```
table(Statistiklausur$Punkte)  
Auswahl2<-Statistiklausur$Vorbereitung[Statistiklausur$Punkte>85]  
mean(Auswahl2)  
24,67
```

Verwendeter Datensatz

	▲ Geschlecht ▲	Alter ▲	Vorbereitung ▲	Punkte ▲
1	m	21	0	41
2	m	21	2	49
3	m	23	4	53
4	w	23	7	55
5	w	24	12	67
6	m	24	28	73
7	w	24	21	77
8	w	24	23	84
9	w	25	23	91
10	m	26	27	97
11	m	21	20	76
12	m	22	11	66
13	m	23	9	61
14	w	20	6	54
15	w	26	3	40
16	m	21	11	69
17	m	22	8	53
18	w	20	24	88
19	w	23	20	77
20	m	29	3	40

R-Code zur Re-Initialisierung dieses Datensatzes

Geschlecht

```
<-c ("m","m","m","w","w","m","w","w","w","m","m","m","m","w","w","m","m","w","w","m")
```

```
Alter <- c (21,21,23,23,24,24,24,24,25,26,21,22,23,20,26,21,22,20,23,29)
```

```
Vorbereitung <- c (0,2,4,7,12,28,21,23,23,27,20,11,9,6,3,11,8,23,20,3)
```

```
Punkte <- c (41,49,53,55,67,73,77,84,91,97,76,66,61,54,40,69,53,88,77,40)
```

```
Statistiklausur <- data.frame(Geschlecht,Alter,Vorbereitung,Punkte)
```

Sommersemester 2020

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang Betriebswirtschaftslehre
Sommersemester 2020

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 09. Juli 2020

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2020

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☐ falsch

2) Der Momentenkoeffizient der Schiefe lässt sich lediglich bei unimodalen Verteilungen interpretieren.

☐ richtig ☐ falsch

3) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☐ richtig ☐ falsch

4) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für lineare Zusammenhänge.

☐ richtig ☐ falsch

5) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

☐ richtig ☐ falsch

6) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

☐ richtig ☐ falsch

7) Ein Bravais-Pearson-Korrelationskoeffizient von -0,85 deutet auf eine schwache lineare Korrelation hin.

☐ richtig ☐ falsch

8) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert der Steigung der linearen Regressionsgeraden.

☐ richtig ☐ falsch

9) Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus.

☐ richtig ☐ falsch

10) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

In einem Online-Shop gingen am 07.05.2022 Bestellungen von insgesamt 15 Kundinnen und Kunden ein, die Altersangaben in ihrem Kundenprofil hinterlegt haben. Diese 15 Altersangaben sind in der nachfolgenden Tabelle enthalten.

Kundennummer	Alter (in ganzen Jahren)
17231	25
23471	31
23232	23
15327	27
17333	38
27443	19
51338	41
16207	28
32110	27
46558	68
27339	36
13237	31
72112	25
12613	25
92121	30

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Altersangaben und benennen Sie die hierfür erforderlichen Größen (15 Punkte). Wie jung müsste ein zusätzlicher (!) Kunde sein, um im Box-Plot als Ausreißer im unteren Bereich sichtbar zu werden? (5 Punkte)

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Für die am 07.05.2022 im Online-Shop aus Aufgabe II getätigten Bestellungen liegen ebenfalls die Bestellwerte (jeweils auf- oder abgerundet auf ganze Eurobeträge) vor.

Kundennummer	Bestellwert (in Euro)
17231	24
23471	30
23232	23
15327	45
17333	30
27443	23
51338	31
16207	12
32110	17
46558	138
27339	46
13237	58
72112	69
12613	6
92121	27

Berechnen Sie für die Bestellwerte die folgenden Lage- und Streuungsmaße:

- a) Arithmetisches Mittel (5 Punkte)
- b) Um 5% getrimmtes arithmetisches Mittel (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Varianz (10 Punkte)
- f) Standardabweichung (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Der Online-Shop aus den vorangegangenen Aufgaben experimentiert mit Online-Werbung und schaltet über einen Zeitraum von 10 Tagen jeweils eine zunehmende Anzahl von Werbeanzeigen auf einer vielbesuchten Nachrichtenseite. Es wird jeweils aufgezeichnet, wie viele Personen, die an jedem Tag über diese Nachrichtenseite in den Online-Shop gekommen sind, dort einen Kauf getätigt haben. Die nachfolgende Tabelle zeigt die erfassten Daten.

Datum	Anzahl der geschalteten Werbeanzeigen	Anzahl der über die Werbeklicks gewonnenen Kundinnen und Kunden
01.04.2021	5	0
02.04.2021	10	2
03.04.2021	15	3
04.04.2021	20	1
05.04.2021	25	4
06.04.2021	30	6
07.04.2021	35	6
08.04.2021	40	5
09.04.2021	45	6
10.04.2021	50	5

Berechnen und interpretieren Sie für die vorliegenden Daten a) den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) sowie b) den Bravais-Pearson-Korrelationskoeffizienten (20 Punkte).

MUSTERLÖSUNG KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2020

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Metrisch skalierte Daten können in eine natürliche Reihenfolge gebracht werden.

X richtig

O falsch

2) Der Momentenkoeffizient der Schiefe lässt sich lediglich bei unimodalen Verteilungen interpretieren.

X richtig

O falsch

3) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

X richtig

O falsch

4) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für lineare Zusammenhänge.

X richtig

O falsch

5) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

X richtig

O falsch

6) Der Median wird von Ausreißern erheblich beeinflusst und gilt daher als nicht robust.

O richtig

X falsch

7) Ein Bravais-Pearson-Korrelationskoeffizient von -0,85 deutet auf eine schwache lineare Korrelation hin.

O richtig

X falsch

8) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert der Steigung der linearen Regressionsgeraden.

O richtig

X falsch

9) Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus.

O richtig

X falsch

10) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

O richtig

X falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

In einem Online-Shop gingen am 07.05.2022 Bestellungen von insgesamt 15 Kundinnen und Kunden ein, die Altersangaben in ihrem Kundenprofil hinterlegt haben. Diese 15 Altersangaben sind in der nachfolgenden Tabelle enthalten.

Kundennummer	Alter (in ganzen Jahren)
17231	25
23471	31
23232	23
15327	27
17333	38
27443	19
51338	41
16207	28
32110	27
46558	68
27339	36
13237	31
72112	25
12613	25
92121	30

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Altersangaben und benennen Sie die hierfür erforderlichen Größen (15 Punkte). Wie jung müsste ein zusätzlicher (!) Kunde sein, um im Box-Plot als Ausreißer im unteren Bereich sichtbar zu werden? (5 Punkte)

Bildung der geordneten Reihe:

19, 23, 25, 25, 25, 27, 27, 28, 30, 31, 31, 36, 38, 41, 68

Bestimmung der Quartilswerte:

$0,25 \cdot 15 = 3,75 \rightarrow$ Wert an der 4. Stelle der geordneten Verteilung $\rightarrow 25$

$0,50 \cdot 15 = 7,50 \rightarrow$ Wert an der 8. Stelle der geordneten Verteilung $\rightarrow 28$

$0,75 \cdot 15 = 11,25 \rightarrow$ Wert an der 12. Stelle der geordneten Verteilung $\rightarrow 36$

Bestimmung des IQR:

$IQR \rightarrow 36 - 25 = 11$

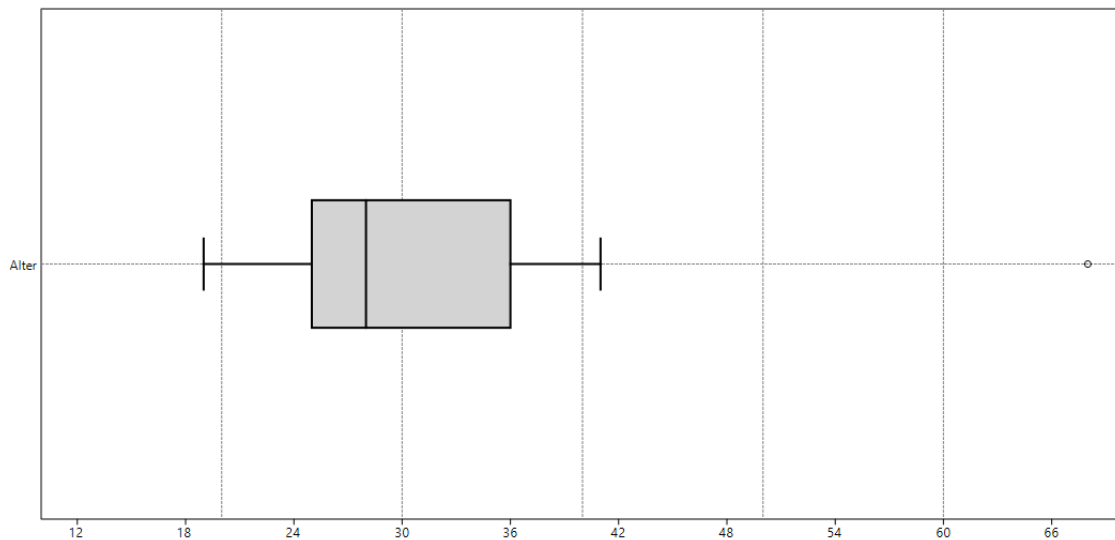
1,5-facher IQR $\rightarrow 16,5$

Bestimmung der Zaungrenzen

$25 - 16,5 = 8,5 \rightarrow$ kleinster Wert im Datensatz über dieser Grenze $\rightarrow 19$

$36 + 16,5 = 52,5 \rightarrow$ größter Wert im Datensatz unter dieser Grenze $\rightarrow 41$

Die Box verläuft von 25 auf 36, der Median wird bei 28 eingezeichnet. Der untere Zaun verläuft bis zu 19, der obere Zaun bis zur 41, ein einzelner Ausreißer wird bei 68 eingezeichnet.



Wie jung müsste ein zusätzlicher (!) Kunde sein, um im Box-Plot als Ausreißer im unteren Bereich sichtbar zu werden?

Einfache Lösung:

Ein zusätzlicher Kunde müsste 8 Jahre oder jünger sein.

Komplexe Lösung für maximale Punktzahl:

Bei einem zusätzlichen Kunden gäbe es insgesamt 16 Werte.

Bildung der geordneten Reihe:

X, 19, 23, 25, 25, 25, 27, 27, 28, 30, 31, 31, 36, 38, 41, 68

Bestimmung der Quartilswerte:

$0,25 \cdot 16 = 4 \rightarrow$ Mittel der Werte an 4. und 5. Stelle $\rightarrow 25$

$0,75 \cdot 16 = 12 \rightarrow$ Mittel der Werte an 12. und 13. Stelle $\rightarrow 33,5$

Bestimmung des IQR:

$33,5 - 25 = 8,5$

1,5-facher IQR = 12,75

Bestimmung der Zaungrenzen

$25 - 12,75 = 12,25$

Da das Alter in ganzen Jahren erfasst wird, dürfte eine zusätzliche Person höchstens 12 Jahre alt sein, um als Ausreißer im unteren Teil des Box-Plots zu erscheinen.

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Für die am 07.05.2022 im Online-Shop aus Aufgabe II getätigten Bestellungen liegen ebenfalls die Bestellwerte (jeweils auf- oder abgerundet auf ganze Eurobeträge) vor.

Kundennummer	Bestellwert (in Euro)
17231	24
23471	30
23232	23
15327	45
17333	30
27443	23
51338	31
16207	12
32110	17
46558	138
27339	46
13237	58
72112	69
12613	6
92121	27

Berechnen Sie für die Bestellwerte die folgenden Lage- und Streuungsmaße:

- a) Arithmetisches Mittel (5 Punkte)
- b) Um 5% getrimmtes arithmetisches Mittel (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Varianz (10 Punkte)
- f) Standardabweichung (5 Punkte)

a) Arithmetisches Mittel

$$24 + 30 + 23 + 45 + 30 + 23 + 31 + 12 + 17 + 138 + 46 + 58 + 69 + 6 + 27 = 579$$

$$579 / 15 = 38,6$$

b) Um 5% getrimmtes arithmetisches Mittel

Bildung der geordneten Reihe:

6, 12, 17, 23, 23, 24, 27, 30, 30, 31, 45, 46, 58, 69, 138

$0,05 \cdot 15 = 0,75 \rightarrow$ Entfernung jeweils des größten und des kleinsten Wertes

Neue geordnete Reihe:

12, 17, 23, 23, 24, 27, 30, 30, 31, 45, 46, 58, 69

$$12 + 17 + 23 + 23 + 24 + 27 + 30 + 30 + 31 + 45 + 46 + 58 + 69 = 435$$

$$435 / 13 = 33,46$$

c) Modus

Es ist kein eindeutiges Maximum erkennbar.

d) Spannweite

$$138 - 6 = 132$$

e) Varianz

Arithmetisches Mittel: 38,6

$$(24-38,6)^2 = 213,16$$

$$(30-38,6)^2 = 73,96$$

$$(23-38,6)^2 = 243,36$$

$$(45-38,6)^2 = 40,96$$

$$(30-38,6)^2 = 73,96$$

$$(23-38,6)^2 = 243,36$$

$$(31-38,6)^2 = 57,76$$

$$(12-38,6)^2 = 707,56$$

$$(17-38,6)^2 = 466,56$$

$$(138-38,6)^2 = 9880,36$$

$$(46-38,6)^2 = 54,76$$

$$(58-38,6)^2 = 376,36$$

$$(69-38,6)^2 = 924,16$$

$$(6-38,6)^2 = 1062,76$$

$$(27-38,6)^2 = 134,56$$

Summe aller quadrierten Differenzen: 14553,6

$$14553,6 / 15 = 970,24 \text{ (Euro}^2\text{)}$$

f) Standardabweichung

Wurde aus 970,24 -> 31,15 (Euro)

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Der Online-Shop aus den vorangegangenen Aufgaben experimentiert mit Online-Werbung und schaltet über einen Zeitraum von 10 Tagen jeweils eine zunehmende Anzahl von Werbeanzeigen auf einer vielbesuchten Nachrichtenseite. Es wird jeweils aufgezeichnet, wie viele Personen, die an jedem Tag über diese Nachrichtenseite in den Online-Shop gekommen sind, dort einen Kauf getätigt haben. Die nachfolgende Tabelle zeigt die erfassten Daten.

Datum	Anzahl der geschalteten Werbeanzeigen	Anzahl der über die Werbeklicks gewonnenen Kundinnen und Kunden
01.04.2021	5	0
02.04.2021	10	2
03.04.2021	15	3
04.04.2021	20	1
05.04.2021	25	4
06.04.2021	30	6
07.04.2021	35	6
08.04.2021	40	5
09.04.2021	45	6
10.04.2021	50	5

Berechnen und interpretieren Sie für die vorliegenden Daten a) den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) sowie b) den Bravais-Pearson-Korrelationskoeffizienten (20 Punkte).

a) Rangkorrelationskoeffizient nach Spearman

x	rg (x)	y	rg (y)	d	d ²
5	1	0	1	0	0
10	2	2	3	-1	1
15	3	3	4	-1	1
20	4	1	2	2	4
25	5	4	5	0	0
30	6	6	9	-3	9
35	7	6	9	-2	4
40	8	5	6,5	1,5	2,25
45	9	6	9	0	0
50	10	5	6,5	3,5	12,25

Summe der quadrierten Rangplatzdifferenzen: 33,5

$$\rho = 1 - ((6 \cdot 33,5) / ((10^2 - 1) \cdot 10)) = 1 - (201 / 990) = 1 - 0,2030 = 0,797$$

Es zeigt sich eine mittlere bis starke positive monotone Korrelation. Die Aussagekraft des Ergebnisses wird allerdings durch die Vielzahl von verbundenen Rängen (50% der Werte) stark eingeschränkt.

b) Bravais-Pearson-Korrelationskoeffizient

	x	y	x²	y²	(x*y)
1	5	0	25	0	0
2	10	2	100	4	20
3	15	3	225	9	45
4	20	1	400	1	20
5	25	4	625	16	100
6	30	6	900	36	180
7	35	6	1225	36	210
8	40	5	1600	25	200
9	45	6	2025	36	270
10	50	5	2500	25	250
Summe	275	38	9625	188	1295
Mittel	27,5	3,8	//	//	//

$n = 10$

Arith. Mittel $x = 27,5$

Arith. Mittel $y = 3,8$

Summe $x^2 = 9625$

Summe $y^2 = 188$

Summe $(x*y) = 1295$

$\rightarrow 250 / (45,41 * 6,6) = 250 / 299,71 = 0,834$

Es zeigt sich eine starke positive lineare Korrelation.

Sommersemester 2020

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2020

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 19. Juli 2020

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2020

Christian Reinboth

Aufgabenteil I: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

2) Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert.

☐ richtig ☐ falsch

3) Die Fünf-Werte-Zusammenfassung kann nur für metrisch skalierte Werte berechnet werden.

☐ richtig ☐ falsch

4) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☐ falsch

5) Variablen mit nur zwei möglichen Ausprägungen werden als dichotom bezeichnet.

☐ richtig ☐ falsch

6) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Outlier-Effekt bezeichnet.

☐ richtig ☐ falsch

7) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

8) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☐ falsch

9) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

10) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Eine Befragung von 20 per Zufall ausgewählten Schülerinnen und Schülern auf dem Schulhof eines Wernigeröder Gymnasiums erbracht unter anderem die nachfolgenden Angaben zum Lebensalter.

Proband/in Nr.	Lebensalter (in Jahren)	Proband/in Nr.	Lebensalter (in Jahren)
1	12	11	12
2	17	12	17
3	13	13	16
4	19	14	14
5	20	15	14
6	14	16	13
7	16	17	17
8	12	18	21
9	11	19	17
10	19	20	16

Erstellen Sie das Stamm-Blatt-Diagramm der Altersverteilung

- a) mit einer Stammbreite von 10 (5 Punkte) sowie
- b) mit einer Stammbreite von 5 (5 Punkte).

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Die monatlichen Kita-Gebühren variieren von Kommune zu Kommune. Eine zufällige Auswahl von Kommunen in Sachsen-Anhalt erbrachte die nachfolgenden Angaben zur Gebührenhöhe für eine achtstündige Betreuung in einer kommunalen Kindertagesstätte (reale Werte aus 2017¹).

Name der Kommune	Höhe der monatlichen Kitagebühr
Falkenstein	150,00 EUR
Gerbstedt	160,00 EUR
Wittenberg	155,00 EUR
Tangermünde	254,00 EUR
Huy	260,00 EUR
Skopau	180,00 EUR
Wolmirstedt	270,00 EUR
Haldensleben	165,00 EUR
Osterburg	153,00 EUR
Klötze	150,00 EUR

Bestimmen Sie für die Verteilung der monatlichen Kita-Gebühren

- a) das arithmetische Mittel (2 Punkte),
- b) das um 10% getrimmte arithmetische Mittel (3 Punkte),
- c) die Spannweite (2 Punkte),
- d) den Interquartilsabstand (2 Punkte) sowie
- e) die Standardabweichung (6 Punkte).

¹ <https://www.mdr.de/sachsen-anhalt/kita-preise-100.html>

Aufgabenteil IV: Regressionskoeffizienten (20 Punkte)

Ein Bildungsforscher vermutet einen Zusammenhang zwischen der Helligkeit der Beleuchtung in Klassenräumen und der Konzentrationsfähigkeit von Schülerinnen und Schülern. Um diese Hypothese zu überprüfen, wird eine Gruppe von zehn Schülerinnen und Schülern über einen Zeitraum von mehreren Wochen bei wechselnder Beleuchtungsstärke wiederholt (*ceteris paribus*) einem Konzentrations-test unterzogen, bei dem zwischen 0 und 10 Punkten erzielt werden können. Die nachfolgende Tabelle gibt die kumulierte Gesamtpunktzahl aller zehn Probandinnen und Probanden sowie die zugehörige Beleuchtungsstärke für den jeweiligen Testdurchlauf (in der metrischen Einheit Lux) wieder.

Test Nr.	Beleuchtungsstärke (in lx)	Kumulierte Punktzahl
1	450	35
2	500	42
3	550	43
4	600	51
5	650	44
6	700	52
7	750	64
8	800	59
9	850	67
10	900	68

Berechnen und interpretieren Sie

- a) den Rangkorrelationskoeffizienten nach Spearman (10 Punkte) sowie
- b) den Bravais-Pearson-Korrelationskoeffizienten (10 Punkte).

Aufgabenteil V: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

2) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

☐ richtig ☐ falsch

3) Die zwei Axiome von Kolmogoroff definieren die mathematischen Eigenschaften des Modells mit bzw. ohne Zurücklegen.

☐ richtig ☐ falsch

4) Ein Würfel wird dann als fair bezeichnet, wenn er bei (in etwa) der Hälfte aller Würfe eine gerade Augenzahl anzeigt.

☐ richtig ☐ falsch

5) Das logische NICHT wird auch als Negation bezeichnet.

☐ richtig ☐ falsch

6) Sowohl bei Variationen als auch bei Kombinationen wird in Modelle mit und ohne Zurücklegen unterschieden.

☐ richtig ☐ falsch

7) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsen Schnittpunkts der linearen Regressionsgeraden.

☐ richtig ☐ falsch

8) Die Wahrscheinlichkeiten aller möglichen Elementarereignisse eines Zufallsvorgangs ergeben zusammengerechnet stets den Wert 0.

☐ richtig ☐ falsch

9) Eine Teilmenge einer leeren Menge kann immer wieder nur selbst eine leere Menge sein.

☐ richtig ☐ falsch

10) Bei der einfachen linearen Regressionsanalyse darf nur die abhängige (Y) Variable die unabhängige (X) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

Aufgabenteil VI: Lineare Regression (10 Punkte)

Aus Aufgabe IV ist Ihnen bereits der Versuch zur möglichen Auswirkung der Beleuchtungsstärke auf das Konzentrationsvermögen bekannt. Die Ergebnisse finden sich in der nachfolgenden Tabelle.

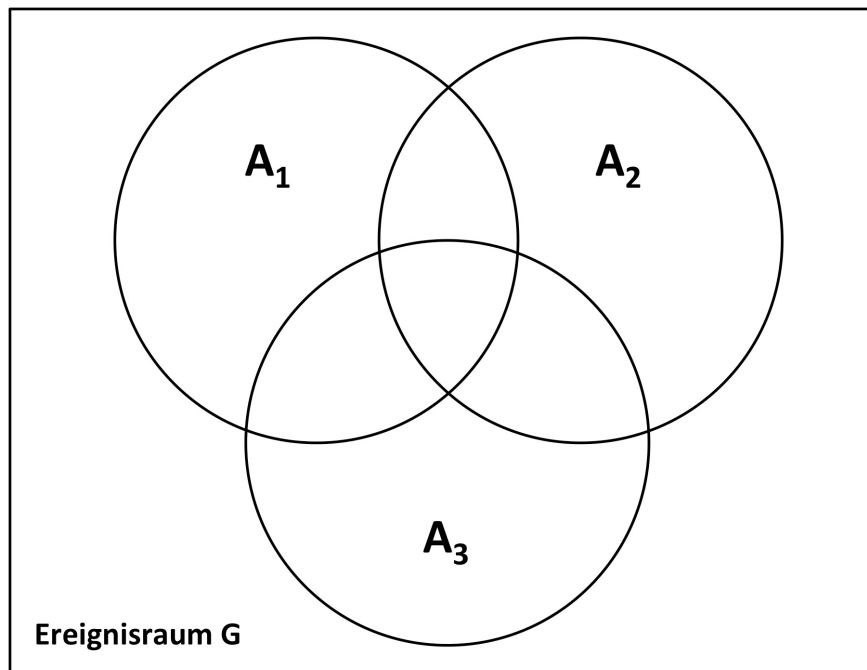
Test Nr.	Beleuchtungsstärke (in lx)	Kumulierte Punktzahl
1	450	35
2	500	42
3	550	43
4	600	51
5	650	44
6	700	52
7	750	64
8	800	59
9	850	67
10	900	68

Stellen Sie ein lineares Regressionsmodell mit der Beleuchtungsstärke als unabhängiger Variablen (x) und der kumulierten Punktzahl als abhängiger Variablen (y) auf (5 Punkte) und bestimmen und interpretieren Sie dessen Güte (5 Punkte).

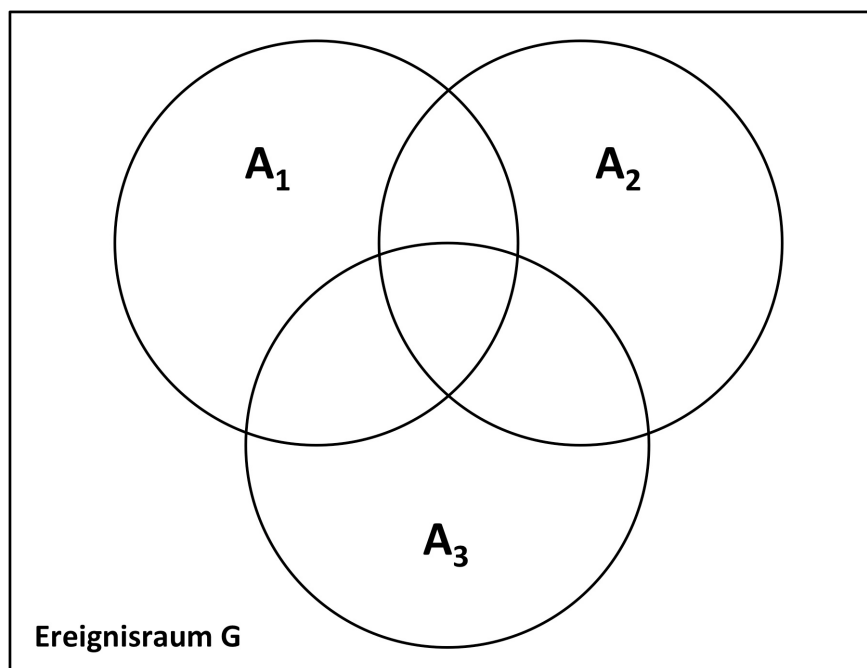
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 2,5 Punkte)

a) $(A_1 \cap A_2) \cup A_3$



b) $(A_1 \cup A_3) \cap A_3$



Aufgabenteil VIII: Kombinatorik (10 Punkte)

In einer Kita-Klasse werden 6 Jungen und 7 Mädchen betreut. Für eine neue Aktivität soll eine Gruppe bestehend aus 3 Jungen und 3 Mädchen gebildet werden. Wie viele mögliche Zusammenstellungen dieser Gruppe gibt es?

Aufgabenteil IX: Wahrscheinlichkeitsrechnung (10 Punkte)

Die Klasse 9b einer örtlichen Gesamtschule will ein Freilichtmuseum besuchen. Damit der Ausflug wie geplant stattfinden kann, darf es am Ausflugstag nicht regnen – und außerdem muss mindestens eine der beiden den Ausflug begleitenden Lehrerinnen gesund und einsatzfähig sein. Die Wahrscheinlichkeit für Regenwetter liegt laut Wettervorhersage bei 10%, die Wahrscheinlichkeit für den krankheitsbedingten Ausfall einer Lehrkraft bei jeweils 5%.

Mit welcher Wahrscheinlichkeit wird der Ausflug stattfinden?

Aufgabenteil X: Bedingte Wahrscheinlichkeiten (10 Punkte)

Eine Kommune stellt die Berechnung der Kita-Gebühren auf eine Formel um, die sich an den Einkommensverhältnissen der Eltern orientiert. Da diese von der Kommune nicht erfasst werden, geschieht die Gebührenberechnung auf Basis einer freiwilligen Selbstauskunft. Ausgehend von Erfahrungswerten in anderen Bereichen wird damit gerechnet, dass rund 2% aller Eltern absichtlich ein zu geringes Einkommen angeben, um die Gebührenlast zu senken.

Die Leiterin einer Kita, in der Kinder aus 120 Familien betreut werden, glaubt, eine falsche Angabe mit einer hohen Sicherheit von 99% als solche erkennen zu können. Nehmen wir nachfolgend an, dass dies zutrifft und dass sie zudem korrekte Angaben mit einer geringen Wahrscheinlichkeit von lediglich 3% irrtümlich als falsch einstuft. Mit welcher Wahrscheinlichkeit ist die Leiterin im Recht, wenn sie einen entsprechenden Verdachtsfall an die Stadtverwaltung meldet?

KLAUSUR STATISTIK I und II MIT LÖSUNGEN

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2020

Christian Reinboth

Aufgabenteil I: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☒ richtig ☐ falsch

2) Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert.

☒ richtig ☐ falsch

3) Die Fünf-Werte-Zusammenfassung kann nur für metrisch skalierte Werte berechnet werden.

☐ richtig ☒ falsch

4) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☒ falsch

5) Variablen mit nur zwei möglichen Ausprägungen werden als dichotom bezeichnet.

☒ richtig ☐ falsch

6) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Outlier-Effekt bezeichnet.

☐ richtig ☒ falsch

7) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☒ falsch

8) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☒ falsch

9) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☒ falsch

10) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☒ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Eine Befragung von 20 per Zufall ausgewählten Schülerinnen und Schülern auf dem Schulhof eines Wernigeröder Gymnasiums erbracht unter anderem die nachfolgenden Angaben zum Lebensalter.

Proband/in Nr.	Lebensalter (in Jahren)	Proband/in Nr.	Lebensalter (in Jahren)
1	12	11	12
2	17	12	17
3	13	13	16
4	19	14	14
5	20	15	14
6	14	16	13
7	16	17	17
8	12	18	21
9	11	19	17
10	19	20	16

Erstellen Sie das Stamm-Blatt-Diagramm der Altersverteilung

- a) mit einer Stammbreite von 10 (5 Punkte) sowie
 b) mit einer Stammbreite von 5 (5 Punkte).

a) Stammbreite von 10

1 | 122233444666777799

2 | 01

Stammbreite 10

Keine Ausreißer

Jedes Blatt ein Fall

b) Stammbreite von 5

1 | 122233444

1 | 666777799

2 | 01

Stammbreite 5

Keine Ausreißer

Jedes Blatt ein Fall

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Die monatlichen Kita-Gebühren variieren von Kommune zu Kommune. Eine zufällige Auswahl von Kommunen in Sachsen-Anhalt erbrachte die nachfolgenden Angaben zur Gebührenhöhe für eine achttündige Betreuung in einer kommunalen Kindertagesstätte (reale Werte aus 2017²).

Name der Kommune	Höhe der monatlichen Kitagebühr
Falkenstein	150,00 EUR
Gerbstedt	160,00 EUR
Wittenberg	155,00 EUR
Tangermünde	254,00 EUR
Huy	260,00 EUR
Skopau	180,00 EUR
Wolmirstedt	270,00 EUR
Haldensleben	165,00 EUR
Osterburg	153,00 EUR
Klötze	150,00 EUR

Bestimmen Sie für die Verteilung der monatlichen Kita-Gebühren

- a) das arithmetische Mittel (2 Punkte),
- b) das um 10% getrimmte arithmetische Mittel (3 Punkte),
- c) die Spannweite (2 Punkte),
- d) den Interquartilsabstand (2 Punkte) sowie
- e) die Standardabweichung (6 Punkte).

a) Arithmetisches Mittel

$$150 + 160 + 155 + 254 + 260 + 180 + 270 + 165 + 153 + 150 = 1897$$

$$1897 / 10 = \underline{\underline{189,70 \text{ (EUR)}}}$$

b) Um 10% getrimmtes arithmetisches Mittel

Bildung der geordneten Reihe:

150, 150, 153, 155, 160, 165, 180, 254, 260, 270

Kürzung um 10% - Kürzung um größten und kleinsten Wert:

150, 153, 155, 160, 165, 180, 254, 260

Berechnung des arithmetischen Mittels:

$$150 + 153 + 155 + 160 + 165 + 180 + 254 + 260 = 1477$$

$$1477 / 8 = \underline{\underline{184,63 \text{ (EUR)}}}$$

c) Spannweite

$$270 - 150 = \underline{\underline{120 \text{ (EUR)}}}$$

² <https://www.mdr.de/sachsen-anhalt/kita-preise-100.html>

d) Interquartilsabstand

Die geordnete Reihe wurde bereits gebildet:

150, 150, 153, 155, 160, 165, 180, 254, 260, 270

$10 \cdot 0,25 = 2,5 \rightarrow$ Wert an 3. Stelle $\rightarrow$ 153

$10 \cdot 0,75 = 7,5 \rightarrow$ Wert an 8. Stelle $\rightarrow$ 254

$$\text{IQR} = 254 - 153 = \underline{\underline{101 \text{ (EUR)}}}$$

e) Standardabweichung

Arithmetisches Mittel = 189,70 EUR

$$(150 - 189,70)^2 = 1.576,09$$

$$(160 - 189,70)^2 = 882,09$$

$$(155 - 189,70)^2 = 1204,09$$

$$(254 - 189,70)^2 = 4134,49$$

$$(260 - 189,70)^2 = 4942,09$$

$$(180 - 189,70)^2 = 94,09$$

$$(270 - 189,70)^2 = 6448,09$$

$$(165 - 189,70)^2 = 610,09$$

$$(153 - 189,70)^2 = 1346,89$$

$$(150 - 189,70)^2 = 1576,09$$

Summe aller quadrierten Abstände = 22814,10

$$\text{Varianz} = 22814,10 / 10 = 2.281,41 \text{ (EUR}^2\text{)}$$

$$\text{Standardabweichung} = \text{positive Wurzel der Varianz} = \underline{\underline{47,76 \text{ (EUR)}}}$$

Aufgabenteil IV: Regressionskoeffizienten (20 Punkte)

Ein Bildungsforscher vermutet einen Zusammenhang zwischen der Helligkeit der Beleuchtung in Klassenräumen und der Konzentrationsfähigkeit von Schülerinnen und Schülern. Um diese Hypothese zu überprüfen, wird eine Gruppe von zehn Schülerinnen und Schülern über einen Zeitraum von mehreren Wochen bei wechselnder Beleuchtungsstärke wiederholt (*ceteris paribus*) einem Konzentrations-test unterzogen, bei dem zwischen 0 und 10 Punkten erzielt werden können. Die nachfolgende Tabelle gibt die kumulierte Gesamtpunktzahl aller zehn Probandinnen und Probanden sowie die zugehörige Beleuchtungsstärke für den jeweiligen Testdurchlauf (in der metrischen Einheit Lux) wieder.

Test Nr.	Beleuchtungsstärke (in lx)	Kumulierte Punktzahl
1	450	35
2	500	42
3	550	43
4	600	51
5	650	44
6	700	52
7	750	64
8	800	59
9	850	67
10	900	68

Berechnen und interpretieren Sie

- a) den Rangkorrelationskoeffizienten nach Spearman (10 Punkte) sowie
 b) den Bravais-Pearson-Korrelationskoeffizienten (10 Punkte).

a) Rangkorrelationskoeffizient nach Spearman

Nr.	Lux	Rang (x)	Punkte	Rang (y)	d	d ²
1	450	1	35	1	0	0
2	500	2	42	2	0	0
3	550	3	43	3	0	0
4	600	4	51	5	-1	1
5	650	5	44	4	1	1
6	700	6	52	6	0	0
7	750	7	64	8	-1	1
8	800	8	59	7	1	1
9	850	9	67	9	0	0
10	900	10	68	10	0	0

Summe der quadrierten Rangplatzdifferenzen: 4

$$\rho = 1 - ((6 \cdot 4) / ((10^2 - 1) \cdot 10))$$

$$\rho = 1 - 24 / 990 = 1 - 0,024 = \underline{\underline{0,975}}$$

Es existiert eine starke, positive, monotone Korrelation.

b) Bravais-Pearson-Korrelationskoeffizient

Nr.	Lux (x)	Punkte (y)	x^2	y^2	$(x \cdot y)$
1	450	35	202500	1225	15750
2	500	42	250000	1764	21000
3	550	43	302500	1849	23650
4	600	51	360000	2601	30600
5	650	44	422500	1936	28600
6	700	52	490000	2704	36400
7	750	64	562500	4096	48000
8	800	59	640000	3481	47200
9	850	67	722500	4489	56950
10	900	68	810000	4624	61200
S	6750	525	4762500	28769	369350
D	675	52,5	//	//	//

$$= 14975 / (454,15 \cdot 34,73) = \underline{\underline{0,949}}$$

Es existiert eine starke, positive, lineare Korrelation.

Aufgabenteil V: Theorie (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1/2 Punkt pro korrekter Beantwortung)

1) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

X richtig ☐ **O falsch**

2) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

X richtig ☐ **O falsch**

3) Die zwei Axiome von Kolmogoroff definieren die mathematischen Eigenschaften des Modells mit bzw. ohne Zurücklegen.

☐ **O richtig** **X falsch**

4) Ein Würfel wird dann als fair bezeichnet, wenn er bei (in etwa) der Hälfte aller Würfe eine gerade Augenzahl anzeigt.

☐ **O richtig** **X falsch**

5) Das logische NICHT wird auch als Negation bezeichnet.

X richtig ☐ **O falsch**

6) Sowohl bei Variationen als auch bei Kombinationen wird in Modelle mit und ohne Zurücklegen unterschieden.

X richtig ☐ **O falsch**

7) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert des X-Achsen Schnittpunkts der linearen Regressionsgeraden.

☐ **O richtig** **X falsch**

8) Die Wahrscheinlichkeiten aller möglichen Elementarereignisse eines Zufallsvorgangs ergeben zusammengerechnet stets den Wert 0.

☐ **O richtig** **X falsch**

9) Eine Teilmenge einer leeren Menge kann immer wieder nur selbst eine leere Menge sein.

X richtig ☐ **O falsch**

10) Bei der einfachen linearen Regressionsanalyse darf nur die abhängige (Y) Variable die unabhängige (X) Variable beeinflussen – und nicht umgekehrt.

☐ **O richtig** **X falsch**

Aufgabenteil VI: Lineare Regression (10 Punkte)

Aus Aufgabe IV ist Ihnen bereits der Versuch zur möglichen Auswirkung der Beleuchtungsstärke auf das Konzentrationsvermögen bekannt. Die Ergebnisse finden sich in der nachfolgenden Tabelle.

Test Nr.	Beleuchtungsstärke (in lx)	Kumulierte Punktzahl
1	450	35
2	500	42
3	550	43
4	600	51
5	650	44
6	700	52
7	750	64
8	800	59
9	850	67
10	900	68

Stellen Sie ein lineares Regressionsmodell mit der Beleuchtungsstärke als unabhängiger Variablen (x) und der kumulierten Punktzahl als abhängiger Variablen (y) auf (5 Punkte) und bestimmen und interpretieren Sie dessen Güte (5 Punkte).

1) Aufstellung des Regressionsmodells

Nr.	Lux (x)	Punkte (y)	x^2	y^2	$(x \cdot y)$
1	450	35	202500	1225	15750
2	500	42	250000	1764	21000
3	550	43	302500	1849	23650
4	600	51	360000	2601	30600
5	650	44	422500	1936	28600
6	700	52	490000	2704	36400
7	750	64	562500	4096	48000
8	800	59	640000	3481	47200
9	850	67	722500	4489	56950
10	900	68	810000	4624	61200
S	6750	525	4762500	28769	369350
D	675	52,5	//	//	//

$$n = 10$$

$$\bar{x} = 675$$

$$\bar{y} = 52,5$$

$$\sum x^2 = 4762500$$

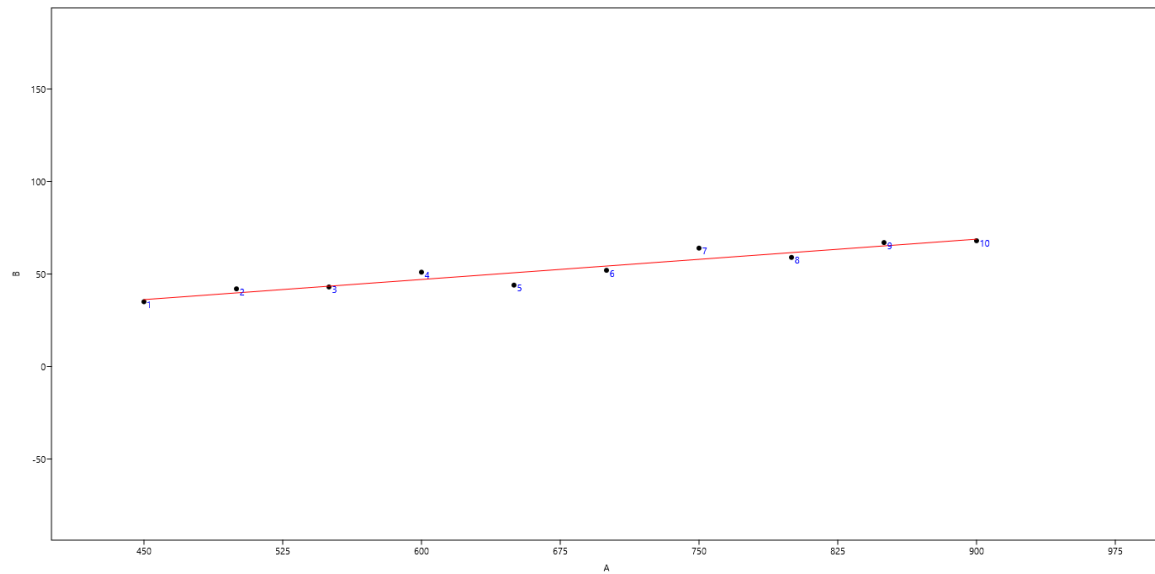
$$\sum y^2 = 28769$$

$$\sum (x \cdot y) = 369350$$

$$b = 14975 / 206250 = 0,07$$

$$a = 52,5 - 0,07 \cdot 675 = 5,25$$

$$\mathbf{y = 5,25 + 0,07 \cdot x}$$



2) Bestimmung der Güte des Regressionsmodells

Der Bravais-Pearson-Korrelationskoeffizient wurde bereits berechnet: $r = 0,949$

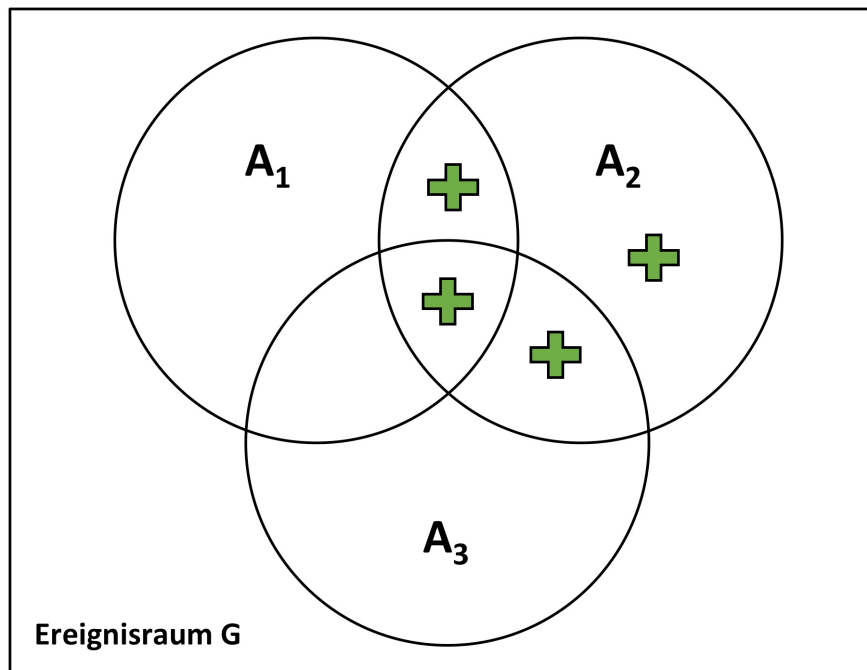
$R^2 = 0,901$

Das Modell liefert eine Streuungsaufklärung von ~90% und ist damit als hervorragend zu betrachten.

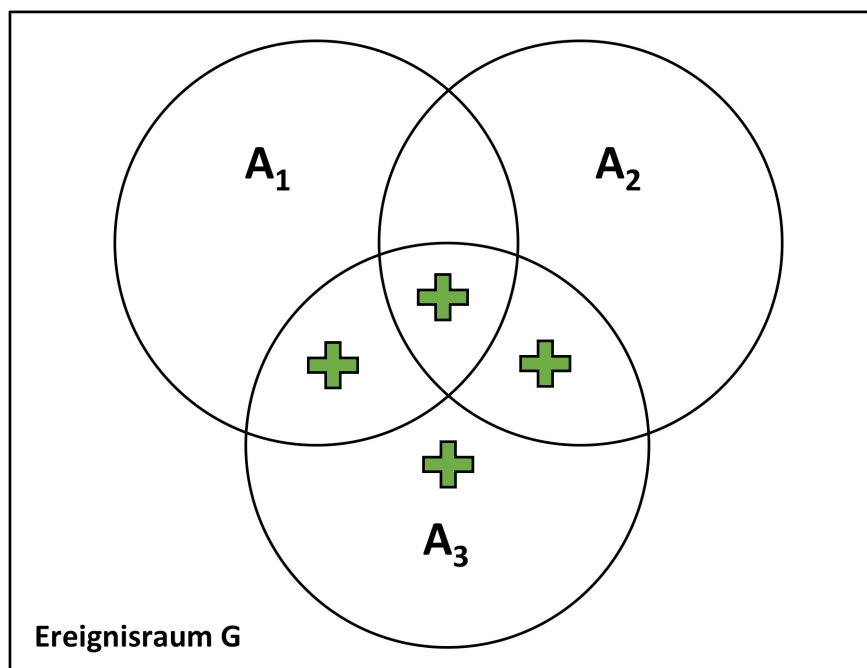
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 2,5 Punkte)

a) $(A_1 \cap A_2) \cup A_2$ -> Schnittfläche von A_1 und A_2 sowie A_2 komplett -> **A_2 komplett**



b) $(A_1 \cup A_3) \cap A_3$ -> **A_3 komplett**



Aufgabenteil VIII: Kombinatorik (10 Punkte)

In einer Kita-Klasse werden 6 Jungen und 7 Mädchen betreut. Für eine neue Aktivität soll eine Gruppe bestehend aus 3 Jungen und 3 Mädchen gebildet werden. Wie viele mögliche Zusammenstellungen dieser Gruppe gibt es?

Kombination (Reihenfolge spielt keine Rolle -> Sven und Mark ist identisch mit Mark und Sven)

Modell ohne Wiederholung -> Sven kann nur maximal einmal gezogen werden – nie mehrfach

$$n! / (k! * (n-k)!)$$

Auswahl von 3 Jungen aus 6 Jungen:

$$6! / (3! * (6-3)!) = 720 / (6 * 6) = 720 / 36 = 20$$

Auswahl von 3 Mädchen aus 7 Mädchen:

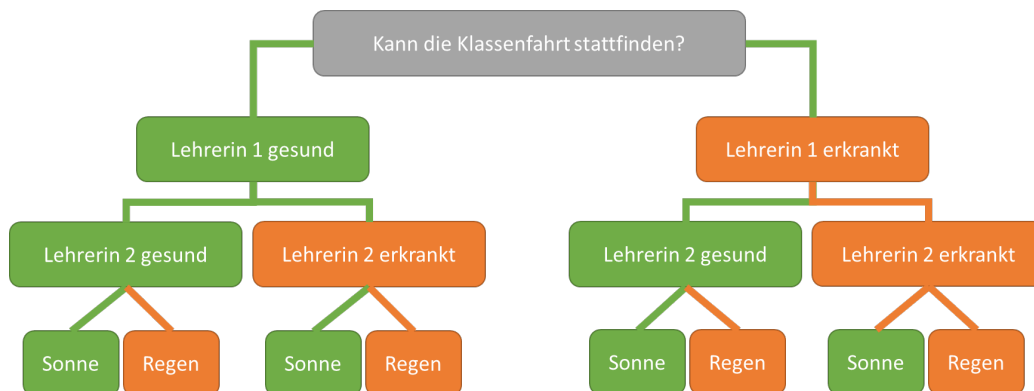
$$7! / (3! * (7-3)!) = 5040 / 144 = 35$$

Möglichkeiten insgesamt: $20 * 35 = \underline{700}$

Aufgabenteil IX: Wahrscheinlichkeitsrechnung (10 Punkte)

Die Klasse 9b einer örtlichen Gesamtschule will ein Freilichtmuseum besuchen. Damit der Ausflug wie geplant stattfinden kann, darf es am Ausflugstag nicht regnen – und außerdem muss mindestens eine der beiden den Ausflug begleitenden Lehrerinnen gesund und einsatzfähig sein. Die Wahrscheinlichkeit für Regenwetter liegt laut Wettervorhersage bei 10%, die Wahrscheinlichkeit für den krankheitsbedingten Ausfall einer Lehrkraft bei jeweils 5%.

Mit welcher Wahrscheinlichkeit wird der Ausflug stattfinden?



Drei der acht Pfade führen zu einem positiven Ausgang:

Lehrerin 1 gesund – Lehrerin 2 gesund – kein Regenwetter

Lehrerin 1 gesund – Lehrerin 2 erkrankt – kein Regenwetter

Lehrerin 1 erkrankt – Lehrerin 2 gesund – kein Regenwetter

$$L1G - L2G - KR \quad 0,95 * 0,95 * 0,9 = 0,81$$

$$L1G - L2K - KR \quad 0,95 * 0,05 * 0,9 = 0,04$$

$$L1K - L2G - KR \quad 0,05 * 0,95 * 0,9 = 0,04$$

$0,81 + 0,04 + 0,04 = 0,89$ -> Der Ausflug wird mit einer Wahrscheinlichkeit von 89% stattfinden.

Aufgabenteil X: Bedingte Wahrscheinlichkeiten (10 Punkte)

Eine Kommune stellt die Berechnung der Kita-Gebühren auf eine Formel um, die sich an den Einkommensverhältnissen der Eltern orientiert. Da diese von der Kommune nicht erfasst werden, geschieht die Gebührenberechnung auf Basis einer freiwilligen Selbstauskunft. Ausgehend von Erfahrungswerten in anderen Bereichen wird damit gerechnet, dass rund 2% aller Eltern absichtlich ein zu geringes Einkommen angeben, um die Gebührenlast zu senken.

Die Leiterin einer Kita, in der Kinder aus 120 Familien betreut werden, glaubt, eine falsche Angabe mit einer hohen Sicherheit von 99% als solche erkennen zu können. Nehmen wir nachfolgend an, dass dies zutrifft und dass sie zudem korrekte Angaben mit einer geringen Wahrscheinlichkeit von lediglich 3% irrtümlich als falsch einstuft. Mit welcher Wahrscheinlichkeit ist die Leiterin im Recht, wenn sie einen entsprechenden Verdachtsfall an die Stadtverwaltung meldet?

120 Familien

2 % machen falsche Angaben = 2,4 falsche Angaben

98% machen korrekte Angaben = 117,6 korrekte Angaben

Von 2,4 falschen Angaben werden 99% korrekt erkannt = 2,38

Von 117,6 korrekten Angaben werden 3% falsch erkannt = 3,53

Insgesamt werden damit $2,38 + 3,53 = 5,91$ Angaben als falsch erkannt.

Von diesen 5,91 als falsch erkannten Angaben sind 2,38 wirklich falsch = 0,40

Die Wahrscheinlichkeit für einen zutreffenden Verdacht liegt bei **40%**.

Wintersemester 2020/2021

Klausur Statistik

Hinweis: Bei dieser Klausur handelt es sich um eine Prüfung, die im Rahmen des Studiengangs Wirtschaftsingenieurwesen (nicht Betriebswirtschaftslehre) abgelegt wurde. In diesem Studiengang entfällt die Aufteilung in Statistik I und Statistik II zugunsten einer einzigen Prüfung.

Hochschule Harz
Wernigerode
Fachbereich Automatisierung und Informatik
bbgl. Bachelorstudiengang Wirtschaftsingenieurwesen
Wintersemester 2020 / 2021

Prüfungsfach:

Statistik

Name des Prüfers: Christian Reinboth

Prüfungstag: 28. November 2020

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK

Berufsbegleitender Studiengang Wirtschaftsingenieurwesen

Wintersemester 2020 / 2021

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☐ richtig☐ falsch

2) Bei der einfachen linearen Regressionsanalyse darf nur die abhängige (Y) Variable die unabhängige (X) Variable beeinflussen – und nicht umgekehrt.

☐ richtig☐ falsch

3) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig☐ falsch

4) Zwei Teilgesamtheiten einer Grundgesamtheit können sich beliebig überschneiden.

☐ richtig☐ falsch

5) Bei der Klassierung von Daten dürfen auch unterschiedlich breite Klassen gebildet werden.

☐ richtig☐ falsch

6) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleichbleibenden Rahmenbedingungen.

☐ richtig☐ falsch

7) Um repräsentativ zu sein, muss eine Zufallsstichprobe sowohl einen geeigneten Umfang als auch eine angemessen hohe Rücklaufquote aufweisen.

☐ richtig☐ falsch

8) Um mit der Laplace-Wahrscheinlichkeit rechnen zu können, muss die Anzahl aller günstigen sowie die Anzahl aller möglichen Elementarereignisse bekannt sein.

☐ richtig☐ falsch

9) Ordinalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig☐ falsch

10) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☐ richtig☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die IT-Abteilung eines Großunternehmens registriert in den ersten 14 Tagen nach Einführung einer neuen Kundenmanagement-Software die Anzahl der Support-Anfragen an den internen Helpdesk.

Datum	Anfragen
10.11.	87
11.11.	89
12.11.	83
13.11.	78
14.11.	77
15.11.	72
16.11.	68
17.11.	75
18.11.	63
19.11.	61
20.11.	58
21.11.	60
22.11.	61
23.11.	55

Zeichnen Sie einen Box-Plot für diese Verteilung (5 Punkte) und benennen Sie alle für die Konstruktion benötigten Größen (5 Punkte).

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Auf einer Spitzgussmaschine werden bewegliche Bauelemente für MRTs hergestellt. Aus einer laufenden Produktionsreihe werden 100 Elemente entnommen und gewogen, um festzustellen, wie weit sie vom Zielgewicht von 5,5 g abweichen. Die nachfolgende Tabelle zeigt die Verteilung der festgestellten Abweichungen.

Anzahl der Bauteile	Abweichung in g
58	0,0
33	0,1
6	0,2
2	0,5
1	0,8

Bestimmen Sie

- u) das arithmetische Mittel,
- v) das um 5% getrimmte arithmetische Mittel,
- w) den Modus,
- x) die Spannweite und
- y) den Interquartilsabstand.

(jeweils 3 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (15 Punkte)

Im Rahmen der Einführung eines Qualitätsmanagementsystems wird evaluiert, ob sich die Zeit, die in Gespräche mit Kundinnen und Kunden investiert wird, die sich bei der Hotline beschweren, auf die spätere Kundenzufriedenheit mit der Bearbeitung (erhoben in Punkten von 1 bis 100) auswirkt. Die nachfolgende Tabelle enthält einen Auszug aus den erfassten Daten, wobei darauf geachtet wurde, dass alle in der Tabelle zusammengeführten Kundenanfragen Probleme von vergleichbarem Arbeitsaufwand betrafen.

Kundennummer	Gesprächsdauer (min)	Zufriedenheit (Punkte)
OA17B	12	67
NA21C	21	81
CR23C	23	83
OA21D	13	70
OA19V	17	75
NC41D	41	46
NA44Y	25	85
CX17D	15	71
OA29C	17	80
DV13D	28	85

Berechnen (10 Punkte) und interpretieren (5 Punkte) Sie den Rangkorrelationskoeffizienten nach Spearman für die Variablen „Gesprächsdauer“ und „Zufriedenheit“.

Aufgabenteil V: Lineare Regressionsanalyse (15 Punkte)

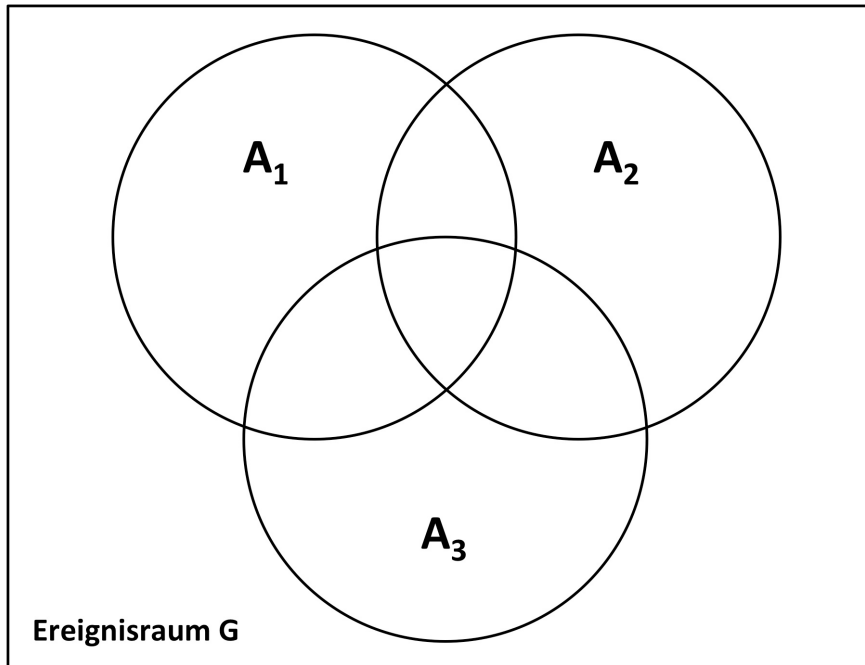
Stellen Sie für die bereits bekannte Verteilung der Gesprächsdauer und der Kundenzufriedenheit die lineare Regressionsgleichung mit der Gesprächsdauer als unabhängiger (x) Variable und der Kundenzufriedenheit als abhängiger (y) Variable (10 Punkte) auf und bewerten Sie deren Güte über das Bestimmtheitsmaß R^2 (5 Punkte).

Kundennummer	Gesprächsdauer (min)	Zufriedenheit (Punkte)
OA17B	12	67
NA21C	21	81
CR23C	23	83
OA21D	13	70
OA19V	17	75
NC41D	41	46
NA44Y	25	85
CX17D	15	71
OA29C	17	80
DV13D	28	85

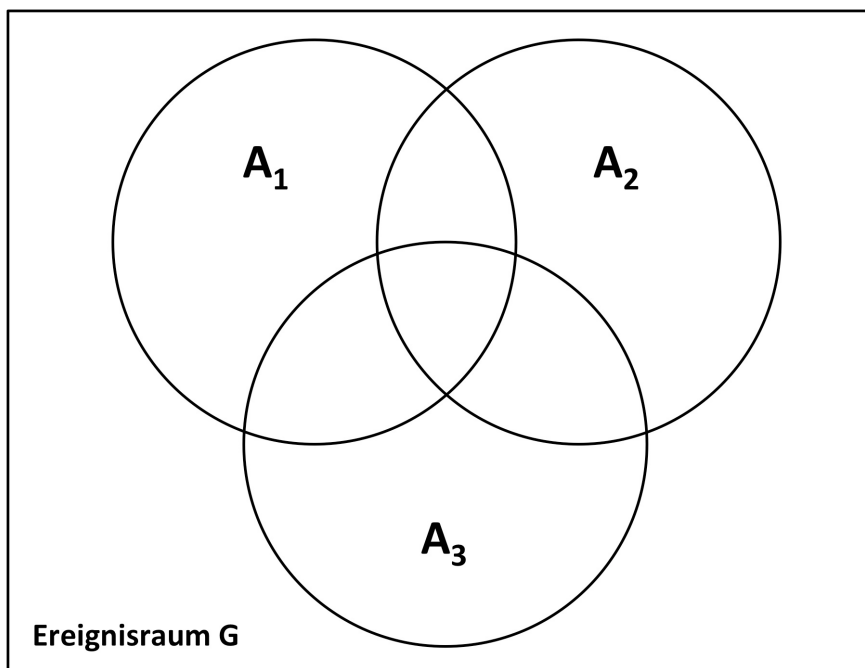
Aufgabenteil VI: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 2,5 Punkte)

a) $(A_1 \cup A_2) \cap \bar{A}_3$



b) $(A_1 \cap A_2) \cup (A_2 \cap A_3)$



Aufgabenteil VII: Pfaddiagramme (10 Punkte)

Die interne Revisionsabteilung eines Bankhauses prüft regelmäßig die Reisekostenabrechnungen von Mitarbeiterinnen und Mitarbeitern. Abrechnungen von Mitgliedern des Vorstands werden stets von zwei Personen geprüft. Dabei fällt der Prüferin Frau Schulze ein Fehler in der Abrechnung mit einer Wahrscheinlichkeit von 70% auf, während der Prüfer Herr Meier diesen lediglich mit einer Wahrscheinlichkeit von 50% bemerkt.

Die fehlerhafte Abrechnung eines Vorstandsmitglieds wird parallel durch Frau Schulze und Herrn Meier überprüft. Wie groß ist die Wahrscheinlichkeit dafür, dass beide den Fehler nicht bemerken?

Aufgabenteil VIII: Kombinatorik (5 Punkte)

Ein Außendienstmitarbeiter im Kundendienst soll im Rahmen eines Arbeitstages Vor-Ort-Anfragen von 7 verschiedene Kundinnen und Kunden bearbeiten, die über das Stadtgebiet von Halberstadt verteilt wohnen. In wie vielen verschiedenen Reihenfolgen könnte er diese aufsuchen?

Aufgabenteil IX: Bedingte Wahrscheinlichkeiten (15 Punkte)

Die IT-Abteilung eines Unternehmens scannt die Anhänge jeder E-Mail an den Kundensupport mit einer Anti-Viren-Software. Diese erkennt gefährliche Anhänge mit einer hohen Wahrscheinlichkeit von 99%, stuft fälschlicherweise aber auch 2% der harmlosen Anhänge als potentielle Gefahr ein. Pro Tag gehen rund 5.000 E-Mails beim Kundensupport dieses Unternehmens ein, von denen etwa 800 über einen Anhang verfügen. Aus den Erfahrungen der letzten Jahre weiß das Unternehmen, dass rund 5% der eingehenden Anhänge virenverseucht sind.

Sie arbeiten im Kundensupport dieses Unternehmens und erhalten eine automatisch vom Virens Scanner generierte E-Mail, dass eine an Sie gerichtete E-Mail soeben als kritisch eingestuft und gelöscht wurde. Mit welcher Wahrscheinlichkeit war diese Löschung berechtigt?

KLAUSUR STATISTIK MIT LÖSUNGEN

Berufsbegleitender Studiengang Wirtschaftsingenieurwesen

Wintersemester 2020 / 2021

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

- | | |
|----------------------|-----------------------|
| 1) RICHTIG (1 Punkt) | 6) RICHTIG (1 Punkt) |
| 2) FALSCH (1 Punkt) | 7) RICHTIG (1 Punkt) |
| 3) FALSCH (1 Punkt) | 8) RICHTIG (1 Punkt) |
| 4) RICHTIG (1 Punkt) | 9) RICHTIG (1 Punkt) |
| 5) RICHTIG (1 Punkt) | 10) RICHTIG (1 Punkt) |

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

55, 58, 60, 61, 61, 63, 68, 72, 75, 77, 78, 83, 87, 89

 $0,25 * 14 = 3,5 \rightarrow$ Wert an 4. Stelle $\rightarrow$ 61 $0,50 * 14 = 7 \rightarrow$ Mittel aus Werten an 7. und 8. Stelle $\rightarrow (68 + 72) / 2 =$ 70 $0,75 * 14 = 10,5 \rightarrow$ Wert an 11. Stelle $\rightarrow$ 78 $IQR = 78 - 61 =$ 17

1,5-facher IQR = 25,5

Die Box läuft von 61 auf 78, der Median wird bei 70 eingezeichnet.

Der obere Zaun läuft bis maximal zur $78 + 25,5 = 103,5 \rightarrow$ real bis zur 89.Der untere Zaun läuft bis minimal zur $61 - 25,5 = 35,5 \rightarrow$ real bis zur 55.

Es existieren keine Ausreißer oder Extremwerte im Datensatz.

5 Punkte für die korrekte Berechnung der Werte, 5 Punkte für eine einigermaßen saubere Zeichnung.

Aufgabenteil III: Verteilungsparameter (15 Punkte)

a) $(58 * 0) + (33 * 0,1) + (6 * 0,2) + (2 * 0,5) + (1 * 0,8) = 0 + 3,3 + 1,2 + 1 + 0,8 = 6,3$
 $6,3 / 100 =$ 0,063 (3 Punkte)

b) $100 * 0,05 = 5 \rightarrow$ Die 5 größten und die 5 kleinsten Werte werden aus dem Datensatz entfernt.

$(53 * 0) + (33 * 0,1) + (4 * 0,2) = 0 + 3,3 + 0,8 = 4,1$
 $4,1 / 90 =$ 0,046 (3 Punkte)

c) Der Modus liegt bei einer Abweichung von 0,0 g. (3 Punkte)

d) Die Spannweite liegt bei $0,8 \text{ g} - 0,0 \text{ g} =$ 0,8 g. (3 Punkte)

e) $0,25 * 100 = 25 \rightarrow$ Mittel aus Wert an 25. und 26. Stelle $\rightarrow (0,0 + 0,0) / 2 = 0,0$
 $0,75 * 100 = 75 \rightarrow$ Mittel aus Wert an 75. und 76. Stelle $\rightarrow (0,1 + 0,1) / 2 = 0,1$

 $IQR = 0,1 - 0,0 =$ 0,1 (3 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (15 Punkte)

Gesprächsdauer (min)	rg(x)	Zufriedenheit (Punkte)	rg(y)	d	d ²
12	1	67	2	-1	1
21	6	81	7	-1	1
23	7	83	8	-1	1
13	2	70	3	-1	1
17	4,5	75	5	-0,5	0,25
41	10	46	1	9	81
25	8	85	9,5	-1,5	2,25
15	3	71	4	-1	1
17	4,5	80	6	-1,5	2,25
28	9	85	9,5	-0,5	0,25

Summe aller quadrierten Rangplatzdifferenzen: 91

$$\rho = 1 - (6 * 91) / [(10^2 - 1) * 10] = 1 - 546 / 990 = 1 - 0,5516 = 0,4484 \text{ (10 Punkte)}$$

Interpretation: Es liegt eine positive und monotone Korrelation mittlerer Stärke vor. (5 Punkte)

Aufgabenteil V: Lineare Regressionsanalyse (15 Punkte)

	Gesprächsdauer (x)	Zufriedenheit (y)	x ²	y ²	(x*y)
	12	67	144	4489	804
	21	81	441	6561	1701
	23	83	529	6889	1909
	13	70	169	4900	910
	17	75	289	5625	1275
	41	46	1681	2116	1886
	25	85	625	7225	2125
	15	71	225	5041	1065
	17	80	289	6400	1360
	28	85	784	7225	2380
Summe	212	743	5176	56471	15415
Mittel	21,2	74,3			

$$b = (15415 - 10 * 21,2 * 74,3) / (5176 - 10 * 21,2^2) = -336,6 / 681,6 = -0,49$$

$$a = 74,3 + 0,49 * 21,2 = 84,69$$

$$y = 84,69 - 0,49 * x \quad (10 \text{ Punkte})$$

$$r = (15415 - 10 * 21,2 * 74,3) / [\text{WURZEL}(5176 - 10 * 21,2^2) * \text{WURZEL}(56471 - 10 * 74,3^2)] = -0,36$$

$$R^2 = 0,36^2 = 0,13$$

Das Modell verfügt über keine gute Streuungsaufklärung. Ursächlich hierfür dürfte der Ausreißer (41/46) sein. Dieser erklärt auch, warum sich zwar eine positive monotone Korrelation (wenn auch lediglich eine mittelstarke) in den Daten ermitteln ließ, die lineare Korrelation aber negativ ausfällt. Ursächlich ist hierfür der sogenannte Leverage-Effekt, der die Modellgüte erheblich verringert.

Aufgabenteil VI: Mengenlehre (5 Punkte)

- a) Fläche von A1, die sich nicht mit A3 überschneidet. (2,5 Punkte)
b) A1 komplett sowie die Schnittfläche von A2 und A3. (2,5 Punkte)

Aufgabenteil VII: Pfaddiagramme (10 Punkte)

$$0,3 * 0,5 = \underline{0,15}$$

Die Wahrscheinlichkeit dafür, dass beide den Fehler nicht bemerken, liegt bei 15%. (10 Punkte)

Aufgabenteil VIII: Kombinatorik (5 Punkte)

Es handelt sich um eine Variation im Modell ohne Zurücklegen (Spezialfall Permutation):

$$7! = \underline{5040}$$

Es gibt 5.040 Reihenfolgen, in denen der Mitarbeiter die Kunden aufsuchen kann. (5 Punkte)

Aufgabenteil IX: Bedingte Wahrscheinlichkeiten (15 Punkte)

5.000 E-Mails insgesamt

-> davon 800 mit Anhang

-> davon 95% unkritisch (760)

-> davon 5% kritisch (40)

Von 760 unkritischen E-Mails lösen 2% versehentlich den Alarm aus -> 15,2

Von 40 kritischen E-Mails lösen 99% zu Recht den Alarm aus -> 39,6

Insgesamt wird der Alarm also $15,2 + 39,6 = 54,8$ -mal ausgelöst, davon 39,6-mal zu Recht.

$$39,6 / 54,8 = \underline{72,3}$$

Die Wahrscheinlichkeit für einen echten Virus im Anhang der E-Mail liegt bei 72,3%. (15 Punkte)

Sommersemester 2021

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang Betriebswirtschaftslehre
Sommersemester 2021

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 08. Mai 2021

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2021

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus.

O richtig

O falsch

2) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

O richtig

O falsch

3) Ordinalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

O richtig

O falsch

4) Die Berechnung der Varianz setzt mindestens ordinalskalierte Daten voraus.

O richtig

O falsch

5) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

O richtig

O falsch

6) Im Gegensatz zum arithmetischen Mittel ist der Median gegenüber Ausreißern robust.

O richtig

O falsch

7) Die Fünf-Werte-Zusammenfassung besteht aus den Quartilen, Minimum und Maximum.

O richtig

O falsch

8) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

O richtig

O falsch

9) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

O richtig

O falsch

10) Das arithmetische Mittel kann nur für metrisch skalierte Daten gebildet werden.

O richtig

O falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

In einem Online-Shop gingen am 07.05.2021 Bestellungen von insgesamt 15 Kundinnen und Kunden ein, die Altersangaben in ihrem Kundenprofil hinterlegt haben. Diese 15 Altersangaben sind in der nachfolgenden Tabelle enthalten.

Kundennummer	Alter (in ganzen Jahren)
17231	25
23471	31
23232	23
15327	27
17333	38
27443	19
51338	41
16207	28
32110	27
46558	68
27339	36
13237	31
72112	25
12613	25
92121	30

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Altersangaben und benennen Sie die hierfür erforderlichen Größen (15 Punkte). Wie jung müsste ein zusätzlicher (!) Kunde sein, um im Box-Plot als Ausreißer im unteren Bereich sichtbar zu werden? (5 Punkte)

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Für die am 07.05.2021 im Online-Shop aus Aufgabe II getätigten Bestellungen liegen ebenfalls die Bestellwerte (jeweils auf- oder abgerundet auf ganze Eurobeträge) vor.

Kundennummer	Bestellwert (in Euro)
17231	24
23471	30
23232	23
15327	45
17333	30
27443	23
51338	31
16207	12
32110	17
46558	138
27339	46
13237	58
72112	69
12613	6
92121	27

Berechnen Sie für die Bestellwerte die folgenden Lage- und Streuungsmaße:

- a) Arithmetisches Mittel (5 Punkte)
- b) Um 5% getrimmtes arithmetisches Mittel (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Varianz (10 Punkte)
- f) Standardabweichung (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Der Online-Shop aus den vorangegangenen Aufgaben experimentiert mit Online-Werbung und schaltet über einen Zeitraum von 10 Tagen jeweils eine zunehmende Anzahl von Werbeanzeigen auf einer vielbesuchten Nachrichtenseite. Es wird jeweils aufgezeichnet, wie viele Personen, die an jedem Tag über diese Nachrichtenseite in den Online-Shop gekommen sind, dort einen Kauf getätigt haben. Die nachfolgende Tabelle zeigt die erfassten Daten.

Datum	Anzahl der geschalteten Werbeanzeigen	Anzahl der über die Werbeklicks gewonnenen Kundinnen und Kunden
01.04.2021	5	0
02.04.2021	10	2
03.04.2021	15	3
04.04.2021	20	1
05.04.2021	25	4
06.04.2021	30	6
07.04.2021	35	6
08.04.2021	40	5
09.04.2021	45	6
10.04.2021	50	5

Berechnen und interpretieren Sie für die vorliegenden Daten a) den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) sowie b) den Bravais-Pearson-Korrelationskoeffizienten (20 Punkte).

MUSTERLÖSUNG KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2021

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus.

☐ richtig ☒ **falsch**

2) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☒ **richtig** ☐ falsch

3) Ordinalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☒ **richtig** ☐ falsch

4) Die Berechnung der Varianz setzt mindestens ordinalskalierte Daten voraus.

☐ richtig ☒ **falsch**

5) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

☐ richtig ☒ **falsch**

6) Im Gegensatz zum arithmetischen Mittel ist der Median gegenüber Ausreißern robust.

☒ **richtig** ☐ falsch

7) Die Fünf-Werte-Zusammenfassung besteht aus den Quartilen, Minimum und Maximum.

☒ **richtig** ☐ falsch

8) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☒ **richtig** ☐ falsch

9) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☒ **falsch**

10) Das arithmetische Mittel kann nur für metrisch skalierte Daten gebildet werden.

☒ **richtig** ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

In einem Online-Shop gingen am 07.05.2021 Bestellungen von insgesamt 15 Kundinnen und Kunden ein, die Altersangaben in ihrem Kundenprofil hinterlegt haben. Diese 15 Altersangaben sind in der nachfolgenden Tabelle enthalten.

Kundennummer	Alter (in ganzen Jahren)
17231	25
23471	31
23232	23
15327	27
17333	38
27443	19
51338	41
16207	28
32110	27
46558	68
27339	36
13237	31
72112	25
12613	25
92121	30

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Altersangaben und benennen Sie die hierfür erforderlichen Größen (15 Punkte). Wie jung müsste ein zusätzlicher (!) Kunde sein, um im Box-Plot als Ausreißer im unteren Bereich sichtbar zu werden? (5 Punkte)

Bildung der geordneten Reihe:

19, 23, 25, 25, 25, 27, 27, 28, 30, 31, 31, 36, 38, 41, 68

Bestimmung der Quartilswerte:

$0,25 \cdot 15 = 3,75 \rightarrow$ Wert an der 4. Stelle der geordneten Verteilung $\rightarrow 25$

$0,50 \cdot 15 = 7,50 \rightarrow$ Wert an der 8. Stelle der geordneten Verteilung $\rightarrow 28$

$0,75 \cdot 15 = 11,25 \rightarrow$ Wert an der 12. Stelle der geordneten Verteilung $\rightarrow 36$

Bestimmung des IQR:

$\text{IQR} \rightarrow 36 - 25 = 11$

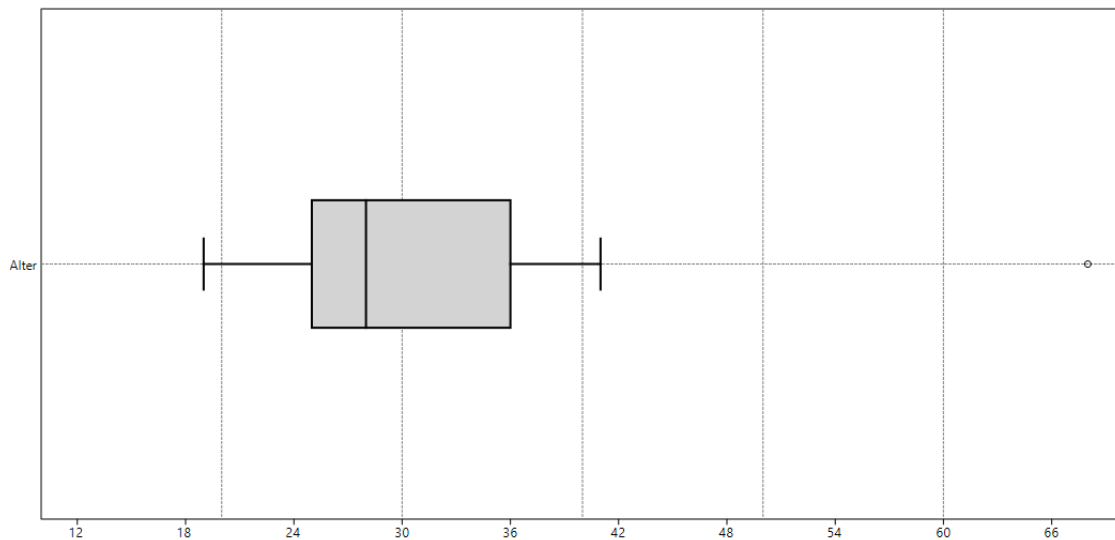
1,5-facher IQR $\rightarrow 16,5$

Bestimmung der Zaungrenzen

$25 - 16,5 = 8,5 \rightarrow$ kleinster Wert im Datensatz über dieser Grenze $\rightarrow 19$

$36 + 16,5 = 52,5 \rightarrow$ größter Wert im Datensatz unter dieser Grenze $\rightarrow 41$

Die Box verläuft von 25 auf 36, der Median wird bei 28 eingezeichnet. Der untere Zaun verläuft bis zu 19, der obere Zaun bis zur 41, ein einzelner Ausreißer wird bei 68 eingezeichnet.



Wie jung müsste ein zusätzlicher (!) Kunde sein, um im Box-Plot als Ausreißer im unteren Bereich sichtbar zu werden?

Einfache Lösung:

Ein zusätzlicher Kunde müsste 8 Jahre oder jünger sein.

Komplexe Lösung für maximale Punktzahl:

Bei einem zusätzlichen Kunden gäbe es insgesamt 16 Werte.

Bildung der geordneten Reihe:

X, 19, 23, 25, 25, 25, 27, 27, 28, 30, 31, 31, 36, 38, 41, 68

Bestimmung der Quartilswerte:

$0,25 \cdot 16 = 4 \rightarrow$ Mittel der Werte an 4. und 5. Stelle $\rightarrow 25$

$0,75 \cdot 16 = 12 \rightarrow$ Mittel der Werte an 12. und 13. Stelle $\rightarrow 33,5$

Bestimmung des IQR:

$33,5 - 25 = 8,5$

1,5-facher IQR = 12,75

Bestimmung der Zaungrenzen

$25 - 12,75 = 12,25$

Da das Alter in ganzen Jahren erfasst wird, dürfte eine zusätzliche Person höchstens 12 Jahre alt sein, um als Ausreißer im unteren Teil des Box-Plots zu erscheinen.

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Für die am 07.05.2021 im Online-Shop aus Aufgabe II getätigten Bestellungen liegen ebenfalls die Bestellwerte (jeweils auf- oder abgerundet auf ganze Eurobeträge) vor.

Kundennummer	Bestellwert (in Euro)
17231	24
23471	30
23232	23
15327	45
17333	30
27443	23
51338	31
16207	12
32110	17
46558	138
27339	46
13237	58
72112	69
12613	6
92121	27

Berechnen Sie für die Bestellwerte die folgenden Lage- und Streuungsmaße:

- a) Arithmetisches Mittel (5 Punkte)
- b) Um 5% getrimmtes arithmetisches Mittel (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Varianz (10 Punkte)
- f) Standardabweichung (5 Punkte)

a) Arithmetisches Mittel

$$24 + 30 + 23 + 45 + 30 + 23 + 31 + 12 + 17 + 138 + 46 + 58 + 69 + 6 + 27 = 579$$

$$579 / 15 = 38,6$$

b) Um 5% getrimmtes arithmetisches Mittel

Bildung der geordneten Reihe:

6, 12, 17, 23, 23, 24, 27, 30, 30, 31, 45, 46, 58, 69, 138

$0,05 \cdot 15 = 0,75 \rightarrow$ Entfernung jeweils des größten und des kleinsten Wertes

Neue geordnete Reihe:

12, 17, 23, 23, 24, 27, 30, 30, 31, 45, 46, 58, 69

$$12 + 17 + 23 + 23 + 24 + 27 + 30 + 30 + 31 + 45 + 46 + 58 + 69 = 435$$

$$435 / 13 = 33,46$$

c) Modus

Es ist kein eindeutiges Maximum erkennbar.

d) Spannweite

$$138 - 6 = 132$$

e) Varianz

Arithmetisches Mittel: 38,6

$$(24-38,6)^2 = 213,16$$

$$(30-38,6)^2 = 73,96$$

$$(23-38,6)^2 = 243,36$$

$$(45-38,6)^2 = 40,96$$

$$(30-38,6)^2 = 73,96$$

$$(23-38,6)^2 = 243,36$$

$$(31-38,6)^2 = 57,76$$

$$(12-38,6)^2 = 707,56$$

$$(17-38,6)^2 = 466,56$$

$$(138-38,6)^2 = 9880,36$$

$$(46-38,6)^2 = 54,76$$

$$(58-38,6)^2 = 376,36$$

$$(69-38,6)^2 = 924,16$$

$$(6-38,6)^2 = 1062,76$$

$$(27-38,6)^2 = 134,56$$

Summe aller quadrierten Differenzen: 14553,6

$$14553,6 / 15 = 970,24 \text{ (Euro}^2\text{)}$$

f) Standardabweichung

Wurde aus 970,24 -> 31,15 (Euro)

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Der Online-Shop aus den vorangegangenen Aufgaben experimentiert mit Online-Werbung und schaltet über einen Zeitraum von 10 Tagen jeweils eine zunehmende Anzahl von Werbeanzeigen auf einer vielbesuchten Nachrichtenseite. Es wird jeweils aufgezeichnet, wie viele Personen, die an jedem Tag über diese Nachrichtenseite in den Online-Shop gekommen sind, dort einen Kauf getätigt haben. Die nachfolgende Tabelle zeigt die erfassten Daten.

Datum	Anzahl der geschalteten Werbeanzeigen	Anzahl der über die Werbeklicks gewonnenen Kundinnen und Kunden
01.04.2021	5	0
02.04.2021	10	2
03.04.2021	15	3
04.04.2021	20	1
05.04.2021	25	4
06.04.2021	30	6
07.04.2021	35	6
08.04.2021	40	5
09.04.2021	45	6
10.04.2021	50	5

Berechnen und interpretieren Sie für die vorliegenden Daten a) den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) sowie b) den Bravais-Pearson-Korrelationskoeffizienten (20 Punkte).

a) Rangkorrelationskoeffizient nach Spearman

x	rg (x)	y	rg (y)	d	d ²
5	1	0	1	0	0
10	2	2	3	-1	1
15	3	3	4	-1	1
20	4	1	2	2	4
25	5	4	5	0	0
30	6	6	9	-3	9
35	7	6	9	-2	4
40	8	5	6,5	1,5	2,25
45	9	6	9	0	0
50	10	5	6,5	3,5	12,25

Summe der quadrierten Rangplatzdifferenzen: 33,5

$$\rho = 1 - ((6 \cdot 33,5) / ((10^2 - 1) \cdot 10)) = 1 - (201 / 990) = 1 - 0,2030 = 0,797$$

Es zeigt sich eine mittlere bis starke positive monotone Korrelation. Die Aussagekraft des Ergebnisses wird allerdings durch die Vielzahl von verbundenen Rängen (50% der Werte) stark eingeschränkt.

b) Bravais-Pearson-Korrelationskoeffizient

	x	y	x²	y²	(x*y)
1	5	0	25	0	0
2	10	2	100	4	20
3	15	3	225	9	45
4	20	1	400	1	20
5	25	4	625	16	100
6	30	6	900	36	180
7	35	6	1225	36	210
8	40	5	1600	25	200
9	45	6	2025	36	270
10	50	5	2500	25	250
Summe	275	38	9625	188	1295
Mittel	27,5	3,8	//	//	//

$n = 10$

Arith. Mittel $x = 27,5$

Arith. Mittel $y = 3,8$

Summe $x^2 = 9625$

Summe $y^2 = 188$

Summe $(x*y) = 1295$

$\rightarrow 250 / (45,41 * 6,6) = 250 / 299,71 = 0,834$

Es zeigt sich eine starke positive lineare Korrelation.

Sommersemester 2021

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang Betriebswirtschaftslehre
Sommersemester 2021

Prüfungsfach:

Statistik I und II

Name des Prüfers: Christian Reinboth

Prüfungstag: 10. Juli 2021

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2021

Christian Reinboth

Aufgabenteil I: Theorie Statistik I (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (0,5 Punkte pro korrekter Beantwortung)

1) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

2) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☐ falsch

3) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

4) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☐ richtig ☐ falsch

5) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

6) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

7) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☐ falsch

8) Aus Stem-and-Leaf-Diagrammen kann man die Ursprungsdaten direkt ablesen.

☐ richtig ☐ falsch

9) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☐ falsch

10) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die Ostfriesland Tee GmbH testet den Absatz der neuen Teemischung „Leuchtturmstimmung“ in zehn Verkaufsstellen in Norddeich und erhebt nach drei Tagen die Verkaufszahlen. Es ergeben sich die nachfolgend aufgeführten Absatzmengen.

Verkaufsstelle	Absatzmenge (in g)
Rosis Teestube	175
Kaufland Norddeich	90
Teespezialitäten Meyer	230
Aldi Norddeich	180
N&P Norddeich	30
Netto Norddeich	210
Shop Teemuseum	340
Nordmarkt	275
Friesentee Müller	190
Lidl Norddeich	245



Zeichnen Sie einen Box-Plot für die Verteilung der Absatzmengen (7 Punkte) und benennen Sie alle hierfür erforderlichen Kennzahlen (3 Punkte).

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Bestimmen Sie für die bereits bekannte Verteilung der Absatzmengen

Verkaufsstelle	Absatzmenge (in g)
Rosis Teestube	175
Kaufland Norddeich	90
Teespezialitäten Meyer	230
Aldi Norddeich	180
N&P Norddeich	30
Netto Norddeich	210
Shop Teemuseum	340
Nordmarkt	275
Friesentee Müller	190
Lidl Norddeich	245



- das arithmetische Mittel (3 Punkte),
- das um 10% getrimmte arithmetische Mittel (3 Punkte),
- den Modus (2 Punkte),
- die Spannweite (2 Punkte) und
- die Standardabweichung (5 Punkte).

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Um den Absatz der bereits bekannten Teemischung „Leuchtturmstimmung“ zu steigern, testet die Ostfriesland Tee GmbH über zwei Monate den Einsatz von mobilen Probierständen in der Fußgängerzone von Aurich. Es wird wochenweise erfasst, wie viele Verkaufsstände im Einsatz waren und wie viele Packungen der neuen Teesorte in den kooperierenden Geschäften jeweils verkauft wurden.

Woche	Probierstände	Verkaufte Packungen
1. Woche	1	73
2. Woche	2	85
3. Woche	3	86
4. Woche	4	185
5. Woche	5	181
6. Woche	6	185
7. Woche	7	217
8. Woche	8	223

Gibt es zwischen beiden Variablen eine Korrelation? Berechnen (15 Punkte) und interpretieren (5 Punkte) Sie zur Beantwortung dieser Frage den Konkordanzkoeffizienten nach Kendall.

Aufgabenteil V: Theorie Statistik II (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (0,5 Punkte pro korrekter Beantwortung)

1) Eine Menge, in der keinerlei Elemente enthalten sind, wird in der Mengenlehre als Nullmenge bezeichnet.

☐ richtig ☐ falsch

2) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

☐ richtig ☐ falsch

3) Bei der einfachen linearen Regressionsanalyse darf nur die abhängige (Y) Variable die unabhängige (X) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

4) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

☐ richtig ☐ falsch

5) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Leverage-Effekt bezeichnet.

☐ richtig ☐ falsch

6) Eine kommutative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

☐ richtig ☐ falsch

7) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig ☐ falsch

8) Beim Chi-Quadrat-Unabhängigkeitstest handelt es sich um einen Test auf Gleichheit der Mittelwerte zweier metrisch skalierten Merkmale.

☐ richtig ☐ falsch

9) Gerade und ungerade Zahlen sind ein passendes Beispiel für disjunkte Mengen.

☐ richtig ☐ falsch

10) Die Durchführung des Chi-Quadrat-Unabhängigkeitstests setzt mindestens ordinal skalierte Merkmale voraus.

☐ richtig ☐ falsch

Aufgabenteil VI: Univariate lineare Regressionsanalyse (15 Punkte)

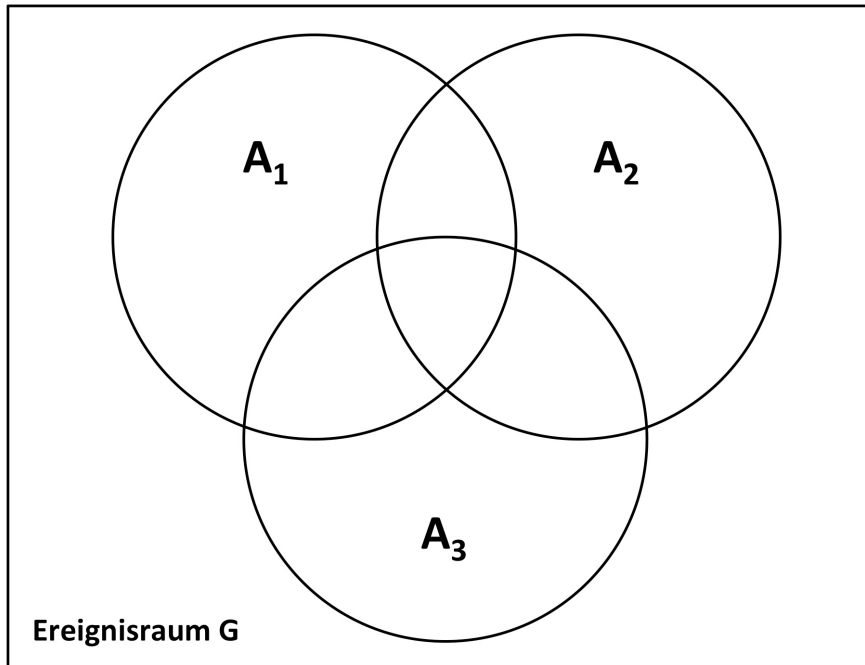
Stellen Sie für die bereits bekannte Verteilung der Anzahl an Probierständen sowie der Anzahl an verkauften Packungen der Teesorte „Leuchtturmstimmung“ ein lineares Regressionsmodell mit der Anzahl an Probierständen als unabhängiger Variablen (x) und der Anzahl an verkauften Packungen als abhängiger Variablen (y) auf (10 Punkte) und bewerten Sie dessen Güte (5 Punkte).

Woche	Probierstände	Verkaufte Packungen
1. Woche	1	73
2. Woche	2	85
3. Woche	3	86
4. Woche	4	185
5. Woche	5	181
6. Woche	6	185
7. Woche	7	217
8. Woche	8	223

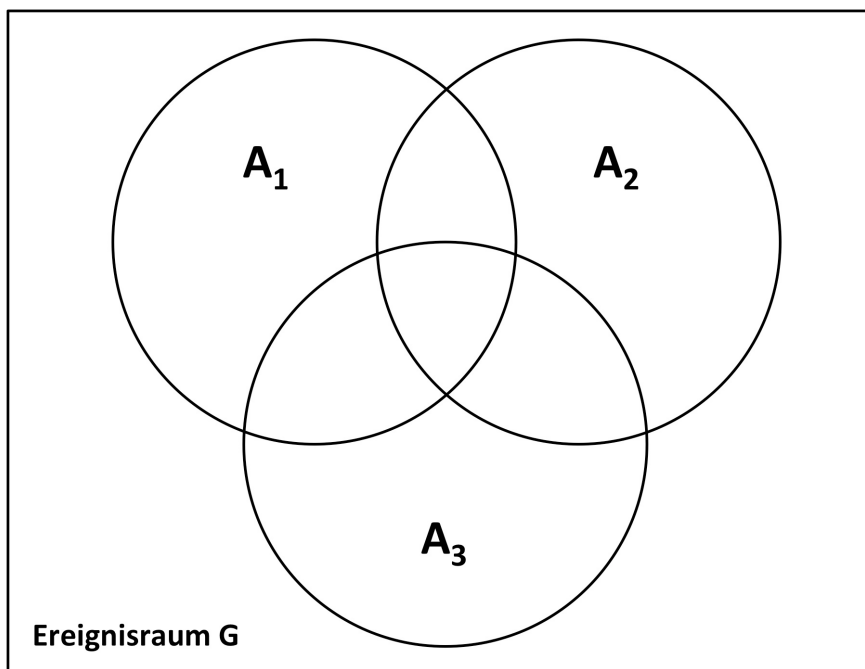
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 2,5 Punkte)

a) $(A_2 \cup A_3) \cap \bar{A}_2$



b) $A_2 \cap A_3 \cap \bar{A}_1$

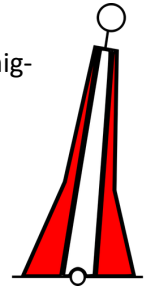


Aufgabenteil VIII: Kombinatorik (5 Punkte)

Eine Yacht, zwei Fischerboote und ein Schlauchboot warten an einer Hebebrücke auf Durchlass. Fünf Minuten vor Öffnung der Brücke ertönt ein Signal, nach dem sich alle Schiffe mit Sicherheitsabstand vor der Brücke aufzureihen haben. In wie vielen verschiedenen Reihenfolgen von Schiffstypen können sich die vier Schiffe nun vor der Hebebrücke in Position bringen?

**Aufgabenteil IX: Pfaddiagramme (5 Punkte)**

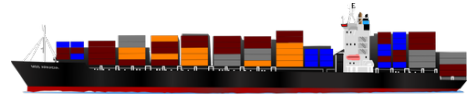
Eine Signalboje ist so konstruiert, dass sie auch bei schwerer See nicht in ihrer Funktionsfähigkeit eingeschränkt wird. Dennoch kann es zu einer solchen Einschränkung kommen, wenn mindestens zwei von drei Faktoren gleichzeitig eintreten: Das Reißen der Verankerung (A), der Ausfall des Signallichts (L) oder das Einreißen des Schwimmkörpers (S). Bei schwerem Seegang liegt die Wahrscheinlichkeit für das Reißen der Verankerung bei 5%, für den Ausfall des Signallichts ebenfalls bei 5% und für das Einreißen des Schwimmkörpers bei 0,5%.



Es herrscht schwere See. Wie hoch ist die Wahrscheinlichkeit für den Ausfall einer spezifischen Signalboje?

Aufgabenteil X: Satz von Bayes (15 Punkte)

Die Hafenpolizei geht davon aus, dass sich in 50 von 5.000 täglich an einem Hafen verschifften Containern Schmuggelware befindet. Eine neue Röntgenanlage soll dabei helfen, diese Container zuverlässiger zu identifizieren. Befindet sich Schmuggelware in einem Container, wird diese während des Röntgenvorgangs mit einer Sicherheit von 95% erkannt. Gleichzeitig werden aber auch 2% aller Container ohne Schmuggelware als kritisch identifiziert und einer genaueren Untersuchung unterzogen.



Sie stehen zufällig neben dem Röntgengerät, als dieses anschlägt. Wie groß ist die Wahrscheinlichkeit dafür, dass bei der nun folgenden Öffnung tatsächlich Schmuggelware gefunden wird?

LÖSUNGEN – KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2021

Christian Reinboth

Aufgabenteil I: Theorie Statistik I (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (0,5 Punkte pro korrekter Beantwortung)

1) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☒ **falsch**

2) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☒ **falsch**

3) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☒ **falsch**

4) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☒ **richtig** ☐ falsch

5) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

☒ **richtig** ☐ falsch

6) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☒ **richtig** ☐ falsch

7) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☒ **falsch**

8) Aus Stem-and-Leaf-Diagrammen kann man die Ursprungsdaten direkt ablesen.

☒ **richtig** ☐ falsch

9) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☒ **falsch**

10) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☒ **richtig** ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Die Ostfriesland Tee GmbH testet den Absatz der neuen Teemischung „Leuchtturmstimmung“ in zehn Verkaufsstellen in Norddeich und erhebt nach drei Tagen die Verkaufszahlen. Es ergeben sich die nachfolgend aufgeführten Absatzmengen.

Verkaufsstelle	Absatzmenge (in g)
Rosis Teestube	175
Kaufland Norddeich	90
Teespezialitäten Meyer	230
Aldi Norddeich	180
N&P Norddeich	30
Netto Norddeich	210
Shop Teemuseum	340
Nordmarkt	275
Friesentee Müller	190
Lidl Norddeich	245



Zeichnen Sie einen Box-Plot für die Verteilung der Absatzmengen (7 Punkte) und benennen Sie alle hierfür erforderlichen Kennzahlen (3 Punkte).

Schritt 1: Ordnung der Werte

30; 90; 175; 180; 190; 210; 230; 245; 275; 340

Schritt 2: Berechnung von Quartilen und IQR

Unteres Quartil: $10 \cdot 0,25 = 2,5 \rightarrow$ Wert an der 3. Stelle der geordneten Liste = 175

Median: $10 \cdot 0,5 = 5 \rightarrow$ Mittel der Werte an 5. und 6. Stelle der geordneten Liste = $(190+210)/2 = 200$

Oberes Quartil: $10 \cdot 0,75 = 7,5 \rightarrow$ Wert an der 8. Stelle der geordneten Liste = 245

$IQR = 245 - 175 = 70$

1,5-facher IQR = $70 \cdot 1,5 = 105$

Schritt 3: Konstruktion des Box-Plots

- Die Box läuft von 175 auf 245, der Median wird bei 200 eingezeichnet
- Der untere Zaun läuft minimal bis $175 - 105 = 70 \rightarrow$ kleinster nicht-extremer Wert ist 90
- Der obere Zaun läuft maximal bis $245 + 105 = 350 \rightarrow$ größter nicht-extremer Wert ist 340
- Es gibt einen Ausreißer (30)

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Bestimmen Sie für die bereits bekannte Verteilung der Absatzmengen

Verkaufsstelle	Absatzmenge (in g)
Rosis Teestube	175
Kaufland Norddeich	90
Teespezialitäten Meyer	230
Aldi Norddeich	180
N&P Norddeich	30
Netto Norddeich	210
Shop Teemuseum	340
Nordmarkt	275
Friesentee Müller	190
Lidl Norddeich	245



- a) das arithmetische Mittel (3 Punkte),
- b) das um 10% getrimmte arithmetische Mittel (3 Punkte),
- c) den Modus (2 Punkte),
- d) die Spannweite (2 Punkte) und
- e) die Standardabweichung (5 Punkte).

a) Arithmetisches Mittel

$$175 + 90 + 230 + 180 + 30 + 210 + 340 + 275 + 190 + 245 = 1965$$

$$1965 / 10 = 196,5$$

b) Um 10% getrimmtes arithmetisches Mittel

Streichung des größten (340) und kleinsten (30) Werts im Datensatz

$$175 + 90 + 230 + 180 + 210 + 275 + 190 + 245 = 1595$$

$$1595 / 8 = 199,375$$

c) Modus

Die Verteilung weist keinen Modus auf

d) Spannweite

$$340 - 30 = 310$$

e) Standardabweichung

$$\begin{array}{ll} (175 - 196,5)^2 = 462,25 & (210 - 196,5)^2 = 182,25 \\ (90 - 196,5)^2 = 11.342,25 & (340 - 196,5)^2 = 20.592,25 \\ (230 - 196,5)^2 = 1.122,25 & (275 - 196,5)^2 = 6.162,25 \\ (180 - 196,5)^2 = 272,25 & (190 - 196,5)^2 = 42,25 \\ (30 - 196,5)^2 = 27.722,25 & (245 - 196,5)^2 = 2.352,25 \end{array}$$

Summe aller quadrierten Abweichungen: 70.252,50

Varianz: 70.252,50 / 10 = 7.025,25

Standardabweichung = positive Wurzel aus 7.025,25 = 83,82

Aufgabenteil IV: Zusammenhangsmaße (20 Punkte)

Um den Absatz der bereits bekannten Teemischung „Leuchtturmstimmung“ zu steigern, testet die Ostfriesland Tee GmbH über zwei Monate den Einsatz von mobilen Probierständen in der Fußgängerzone von Aurich. Es wird wochenweise erfasst, wie viele Verkaufsstände im Einsatz waren und wie viele Packungen der neuen Teesorte in den kooperierenden Geschäften jeweils verkauft wurden.

Woche	Probierstände	Verkaufte Packungen
1. Woche	1	73
2. Woche	2	85
3. Woche	3	86
4. Woche	4	185
5. Woche	5	181
6. Woche	6	185
7. Woche	7	217
8. Woche	8	223

Gibt es zwischen beiden Variablen eine Korrelation? Berechnen (15 Punkte) und interpretieren (5 Punkte) Sie zur Beantwortung dieser Frage den Konkordanzkoeffizienten nach Kendall.

Woche	x	y	rg(y)	K	D
1. Woche	1	73	1	7	0
2. Woche	2	85	2	6	0
3. Woche	3	86	3	5	0
4. Woche	4	185	5,5	2	1
5. Woche	5	181	4	3	0
6. Woche	6	185	5,5	2	0
7. Woche	7	217	7	1	0
8. Woche	8	223	8	/	/

Summe der konkordanten Paare K = 26

Summe der diskordanten Paare D = 1

$$\tau = [2 \cdot (K - D)] / [n \cdot (n - 1)] = 50 / 56 = 0,8929$$

Es liegt eine starke, positive, monotone Korrelation vor. Die Aussagekraft des Koeffizienten ist allerdings durch den verbundenen Rang leicht abgeschwächt.

Aufgabenteil V: Theorie Statistik II (5 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (0,5 Punkte pro korrekter Beantwortung)

1) Eine Menge, in der keinerlei Elemente enthalten sind, wird in der Mengenlehre als Nullmenge bezeichnet.

☐ richtig ☒ falsch

2) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

☒ richtig ☐ falsch

3) Bei der einfachen linearen Regressionsanalyse darf nur die abhängige (Y) Variable die unabhängige (X) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☒ falsch

4) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

☐ richtig ☒ falsch

5) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Leverage-Effekt bezeichnet.

☒ richtig ☐ falsch

6) Eine kommutative Rechenoperation zeichnet sich dadurch aus, dass ihre Argumente miteinander vertauscht werden können, ohne dass sich das Ergebnis ändert.

☒ richtig ☐ falsch

7) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig ☒ falsch

8) Beim Chi-Quadrat-Unabhängigkeitstest handelt es sich um einen Test auf Gleichheit der Mittelwerte zweier metrisch skalierten Merkmale.

☐ richtig ☒ falsch

9) Gerade und ungerade Zahlen sind ein passendes Beispiel für disjunkte Mengen.

☒ richtig ☐ falsch

10) Die Durchführung des Chi-Quadrat-Unabhängigkeitstests setzt mindestens ordinal skalierte Merkmale voraus.

☐ richtig ☒ falsch

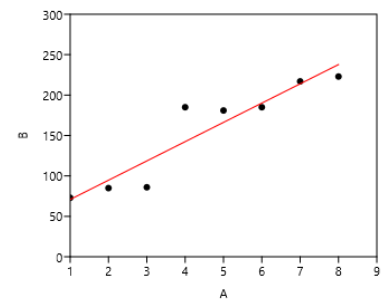
Aufgabenteil VI: Univariate lineare Regressionsanalyse (15 Punkte)

Stellen Sie für die bereits bekannte Verteilung der Anzahl an Probierständen sowie der Anzahl an verkauften Packungen der Teesorte „Leuchtturmstimmung“ ein lineares Regressionsmodell mit der Anzahl an Probierständen als unabhängiger Variablen (x) und der Anzahl an verkauften Packungen als abhängiger Variablen (y) auf (10 Punkte) und bewerten Sie dessen Güte (5 Punkte).

Woche	Probierstände	Verkaufte Packungen
1. Woche	1	73
2. Woche	2	85
3. Woche	3	86
4. Woche	4	185
5. Woche	5	181
6. Woche	6	185
7. Woche	7	217
8. Woche	8	223

Schritt 1: Aufstellung der Regressionsgleichung

Woche	x	y	x ²	y ²	(x*y)
1. Woche	1	73	1	5329	73
2. Woche	2	85	4	7225	170
3. Woche	3	86	9	7396	258
4. Woche	4	185	16	34225	740
5. Woche	5	181	25	32761	905
6. Woche	6	185	36	34225	1110
7. Woche	7	217	49	47089	1519
8. Woche	8	223	64	49729	1784
Summe	36	1235	204	217979	6559
Mittel	4,5	154,375	/	/	/



$$b = (6559 - 8 * 4,5 * 154,375) / (204 - 8 * 4,5^2) = 1001,5/42 = 23,845$$

$$a = 154,375 - 23,845 * 4,5 = 47,073$$

$$y = 47,073 + 23,845 * x$$

Schritt 2: Bestimmung der Modellgüte

$$r = (6559 - 8 * 4,5 * 154,375) / \text{WURZEL}(204 - 8 * 4,5^2) * \text{WURZEL}(217979 - 8 * 154,375^2) \\ = 1001,5 / 6,481 * 165,306 = 0,935$$

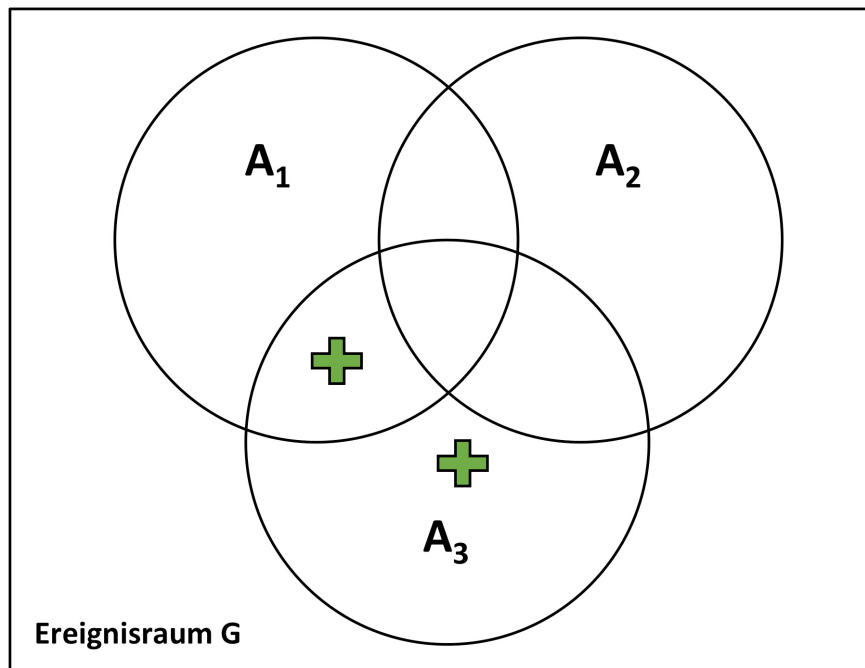
$$R^2 = 0,935^2 = 0,874$$

Das Modell hat einen hohen Grad an Streuungsaufklärung.

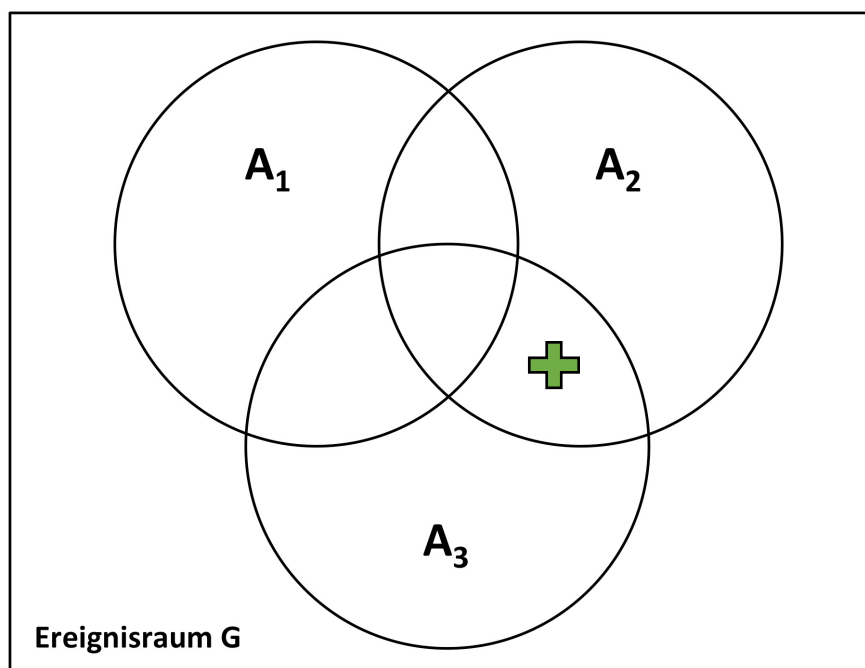
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm. (je 2,5 Punkte)

- a) $(A_2 \cup A_3) \cap \bar{A}_2$ -> Teile von A3 ohne Schnittmenge zu A2



- b) $A_2 \cap A_3 \cap \bar{A}_1$ -> Schnittmenge von A2 und A3 ohne A1



Aufgabenteil VIII: Kombinatorik (5 Punkte)

Eine Yacht, zwei Fischerboote und ein Schlauchboot warten an einer Hebebrücke auf Durchlass. Fünf Minuten vor Öffnung der Brücke ertönt ein Signal, nach dem sich alle Schiffe mit Sicherheitsabstand vor der Brücke aufzureihen haben. In wie vielen verschiedenen Reihenfolgen von Schiffstypen können sich die vier Schiffe nun vor der Hebebrücke in Position bringen?

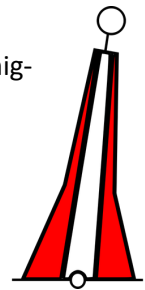


YSFF SYFF FFSY FFYS
YFSF SFYF FSFY FYFS
YFFS SFFY FSYF FYSF

Alternativ: Würde jedes Schiff einzeln zählen gäbe es $4! = 24$ Möglichkeiten.
Dadurch, dass F1 und F2 beliebig austauschbar sind, halbiert sich diese Anzahl.

Aufgabenteil IX: Pfaddiagramme (5 Punkte)

Eine Signalboje ist so konstruiert, dass sie auch bei schwerer See nicht in ihrer Funktionsfähigkeit eingeschränkt wird. Dennoch kann es zu einer solchen Einschränkung kommen, wenn mindestens zwei von drei Faktoren gleichzeitig eintreten: Das Reißen der Verankerung (A), der Ausfall des Signallichts (L) oder das Einreißen des Schwimmkörpers (S). Bei schwerem Seegang liegt die Wahrscheinlichkeit für das Reißen der Verankerung bei 5%, für den Ausfall des Signallichts ebenfalls bei 5% und für das Einreißen des Schwimmkörpers bei 0,5%.

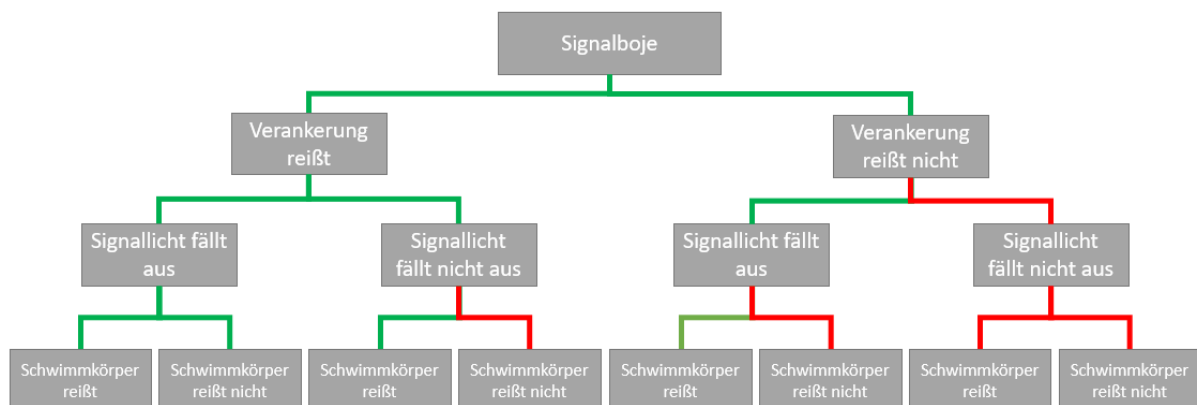


Es herrscht schwere See. Wie hoch ist die Wahrscheinlichkeit für den Ausfall einer spezifischen Signalboje?

Gesucht werden alle Pfade im Pfaddiagramm, in denen mindestens zwei Ereignisse eintreten.

A-L-S	$0,05 * 0,05 * 0,005 = 0,0000125$
A-L-nichtS	$0,05 * 0,05 * 0,995 = 0,0024875$
A-nichtL-S	$0,05 * 0,95 * 0,005 = 0,0002375$
nichtA-L-S	$0,95 * 0,05 * 0,005 = 0,0002375$

Summe = 0,002975 -> Die Wahrscheinlichkeit für einen Ausfall liegt bei nur ~ 0,3%.



Aufgabenteil X: Satz von Bayes (15 Punkte)

Die Hafenpolizei geht davon aus, dass sich in 50 von 5.000 täglich an einem Hafen verschifften Containern Schmuggelware befindet. Eine neue Röntgenanlage soll dabei helfen, diese Container zuverlässiger zu identifizieren. Befindet sich Schmuggelware in einem Container, wird diese während des Röntgenvorgangs mit einer Sicherheit von 95% erkannt. Gleichzeitig werden aber auch 2% aller Container ohne Schmuggelware als kritisch identifiziert und einer genaueren Untersuchung unterzogen.



Sie stehen zufällig neben dem Röntgengerät, als dieses anschlägt. Wie groß ist die Wahrscheinlichkeit dafür, dass bei der nun folgenden Öffnung tatsächlich Schmuggelware gefunden wird?

5.000 Container

4.950 harmlos - davon 2% false positives = 99 Container

50 kritisch - davon 95% korrekt identifiziert = 47,5 Container

Gesamtzahl der als kritisch identifizierten Container: $99 + 47,5 = 146,5$

davon tatsächlich kritisch: 47,5

Verhältnis: 32,42%

Die Wahrscheinlichkeit auf einen Treffer liegt somit bei 32,42%.

Alternativ über die Formel: $(0,01 * 0,95) / ((0,01 * 0,95) + (0,99 * 0,02)) = 0,0095 / (0,0095 + 0,0198)$
 $= 0,0095 / 0,0293 = 0,3242$

Sommersemester 2022

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2022

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 21. Mai 2022

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2022

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☐ richtig ☐ falsch

2) Eine repräsentative Zufallsauswahl setzt die genaue Definition der Grundgesamtheit voraus.

☐ richtig ☐ falsch

3) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

☐ richtig ☐ falsch

4) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☐ falsch

5) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

6) Der Median ist derjenige Wert, der in der geordneten Verteilung am häufigsten auftritt.

☐ richtig ☐ falsch

7) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für lineare Zusammenhänge.

☐ richtig ☐ falsch

8) Bei der Klassenbildung gehören die Klassengrenzen jeweils zur Hälfte zu beiden Klassen.

☐ richtig ☐ falsch

9) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig ☐ falsch

10) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Eine Teemanufaktur im ostfriesischen Aurich vertreibt verschiedene hausgemachte Teemischungen in 200g-Packungen. Zur Optimierung des eigenen Werbekonzepts wird über den Zeitraum einer Woche erfasst, wie viele Packungen dieser Größe jeweils verkauft werden.

Teesorte	Verkaufte Packungen
Apfel-Birke	28
Birne-Apfel	36
Sanddorn-Mango	23
Sanddorn-Vanille	12
Weißtee-Mango	8
Ananas-Mango	11
Tomate-Minze	9
Rote Beete	21
Mango-Birne	58
Honig-Apfel	17



Konstruieren Sie für die gegebene Verteilung der Verkaufszahlen jeweils ein Stamm-Blatt-Diagramm mit Stammbreite 10 (15 Punkte) und Stammbreite 5 (5 Punkte).

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Teesorte	Verkaufte Packungen
Apfel-Birke	28
Birne-Apfel	36
Sanddorn-Mango	23
Sanddorn-Vanille	12
Weißtee-Mango	8
Ananas-Mango	11
Tomate-Minze	9
Rote Beete	21
Mango-Birne	58
Honig-Apfel	17

Berechnen Sie für die bereits bekannte Verteilung der Verkaufszahlen

- | | |
|--|------------------|
| a) das arithmetische Mittel, | (5 Punkte) |
| b) das um 5% getrimmte arithmetische Mittel, | (5 Punkte) |
| c) den Median, | (5 Punkte) |
| d) den Modus, | (5 Punkte) |
| e) das 30%-Perzentil | sowie (5 Punkte) |
| f) die Standardabweichung. | (10 Punkte) |

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Die bereits bekannte Teehandlung experimentiert mit Werbeaufstellern, die an verschiedenen Punkten der Fußgängerzone von Aurich platziert werden. Über einen Zeitraum von zehn Wochen wird die Anzahl dieser Aufsteller (ceteris paribus) erhöht und beobachtet, ob sich eine Änderung bei den Verkaufszahlen einstellt. Diese Beobachtung führte zu den nachfolgend tabellierten Ergebnissen.

Kalenderwoche	Werbeaufsteller	Produktverkäufe
17. KW	0	217
18. KW	1	221
19. KW	2	213
20. KW	3	245
21. KW	4	311
22. KW	5	288
23. KW	6	392
24. KW	7	465
25. KW	8	582
26. KW	9	623



Bestimmen und interpretieren Sie für die oben angegebene bivariable Verteilung der Anzahl an genutzten Werbeaufstellern und der Produktverkäufe

- a) den Rangkorrelationskoeffizienten nach Spearman sowie (15 Punkte)
- b) den Bravais-Pearson-Korrelationskoeffizienten. (20 Punkte)

MUSTERLÖSUNG KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2022

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

X richtig O falsch

2) Eine repräsentative Zufallsauswahl setzt die genaue Definition der Grundgesamtheit voraus.

X richtig O falsch

3) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen -1 und 1 annehmen.

X richtig O falsch

4) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

O richtig **X falsch**

5) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

O richtig **X falsch**

6) Der Median ist derjenige Wert, der in der geordneten Verteilung am häufigsten auftritt.

O richtig **X falsch**

7) Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für lineare Zusammenhänge.

X richtig O falsch

8) Bei der Klassenbildung gehören die Klassengrenzen jeweils zur Hälfte zu beiden Klassen.

O richtig **X falsch**

9) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

X richtig O falsch

10) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

O richtig **X falsch**

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Eine Teemanufaktur im ostfriesischen Aurich vertreibt verschiedene hausgemachte Teemischungen in 200g-Packungen. Zur Optimierung des eigenen Werbekonzepts wird über den Zeitraum einer Woche erfasst, wie viele Packungen dieser Größe jeweils verkauft werden.

Teesorte	Verkaufte Packungen
Apfel-Birke	28
Birne-Äpfel	36
Sanddorn-Mango	23
Sanddorn-Vanille	12
Weißtee-Mango	8
Ananas-Mango	11
Tomate-Minze	9
Rote Beete	21
Mango-Birne	58
Honig-Äpfel	17

Konstruieren Sie für die gegebene Verteilung der Verkaufszahlen jeweils ein Stamm-Blatt-Diagramm mit Stammbreite 10 (15 Punkte) und Stammbreite 5 (5 Punkte).

0 | 8 9
 1 | 1 2 7
 2 | 1 3 8
 3 | 6
 4 |
 5 | 8

Stammbreite 10
 Keine Ausreißer
 Jedes Blatt 1 Fall

0 | 8 9
 1 | 1 2
 1 | 7
 2 | 1 3
 2 | 8
 3 |
 3 | 6
 4 |
 4 |
 5 |
 5 | 8

Stammbreite 5
 Keine Ausreißer
 Jedes Blatt 1 Fall

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Teesorte	Verkaufte Packungen
Apfel-Birke	28
Birne-Apple	36
Sanddorn-Mango	23
Sanddorn-Vanille	12
Weißtee-Mango	8
Ananas-Mango	11
Tomate-Minze	9
Rote Beete	21
Mango-Birne	58
Honig-Apple	17

Berechnen Sie für die bereits bekannte Verteilung der Verkaufszahlen

- a) das arithmetische Mittel, (5 Punkte)
- b) das um 5% getrimmte arithmetische Mittel, (5 Punkte)
- c) den Median, (5 Punkte)
- d) den Modus, (5 Punkte)
- e) das 30%-Perzentil sowie (5 Punkte)
- f) die Standardabweichung. (10 Punkte)

a) Arithmetisches Mittel

$$28 + 36 + 23 + 12 + 8 + 11 + 9 + 21 + 58 + 17 = 223$$

$$223 / 10 = \underline{22,3}$$

b) Um 5% getrimmtes arithmetisches Mittel

$$[8; 9; 11; 12; 17; 21; 23; 28; 26; 58]$$

$10 * 0,05 = 0,5 \rightarrow$ Kürzung von 1 Wert am oberen und unteren Ende der geordneten Verteilung

$$[9; 11; 12; 17; 21; 23; 28; 26]$$

$$9 + 11 + 12 + 17 + 21 + 23 + 28 + 26 = 157$$

$$157 / 8 = \underline{19,63}$$

c) Median

$10 * 0,5 = 5 \rightarrow$ Mittel der Werte an 5. und 6. Stelle der geordneten Verteilung

$$[8; 9; 11; 12; \underline{17; 21}; 23; 28; 26; 58]$$

$$(17 + 21) / 2 = \underline{19}$$

d) Modus

Da keine unimodale Verteilung vorliegt, existiert kein Modus.

e) 30%-Perzentil

$10 * 0,3 = 3 \rightarrow$ Mittel der Werte an 3. und 4. Stelle der geordneten Verteilung

$$[8; 9; \underline{11; 12}; 17; 21; 23; 28; 26; 58]$$

$$(11 + 12) / 2 = \underline{11,5}$$

f) Standardabweichung

Arithmetisches Mittel = 22,3

$$(28 - 22,3)^2 = 32,49$$

$$(36 - 22,3)^2 = 187,69$$

$$(23 - 22,3)^2 = 0,49$$

$$(12 - 22,3)^2 = 106,09$$

$$(8 - 22,3)^2 = 204,49$$

$$(11 - 22,3)^2 = 127,69$$

$$(9 - 22,3)^2 = 176,89$$

$$(21 - 22,3)^2 = 1,69$$

$$(58 - 22,3)^2 = 1274,49$$

$$(17 - 22,3)^2 = 28,09$$

$$32,49 + 187,69 + 0,49 + 106,09 + 204,49 + 127,69 + 176,89 + 1,69 + 1274,49 + 28,09 = 2140,1$$

$$2140,1 / 10 = 214,01$$

$$\text{pos. Wurzel aus } 214,01 = \underline{14,63}$$

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Die bereits bekannte Teehandlung experimentiert mit Werbeaufstellern, die an verschiedenen Punkten der Fußgängerzone von Aurich platziert werden. Über einen Zeitraum von zehn Wochen wird die Anzahl dieser Aufsteller (ceteris paribus) erhöht und beobachtet, ob sich eine Änderung bei den Verkaufszahlen einstellt. Diese Beobachtung führte zu den nachfolgend tabellierten Ergebnissen.

Kalenderwoche	Werbeaufsteller	Produktverkäufe
17. KW	0	217
18. KW	1	221
19. KW	2	213
20. KW	3	245
21. KW	4	311
22. KW	5	288
23. KW	6	392
24. KW	7	465
25. KW	8	582
26. KW	9	623

Bestimmen und interpretieren Sie für die oben angegebene bivariable Verteilung der Anzahl an genutzten Werbeaufstellern und der Produktverkäufe

a) den Rangkorrelationskoeffizienten nach Spearman sowie (15 Punkte)

b) den Bravais-Pearson-Korrelationskoeffizienten. (20 Punkte)

a) Rangkorrelationskoeffizient nach Spearman

KW	x	rg(x)	y	rg(y)	d	d ²
17. KW	0	1	217	2	-1	1
18. KW	1	2	221	3	-1	1
19. KW	2	3	213	1	2	4
20. KW	3	4	245	4	0	0
21. KW	4	5	311	6	-1	1
22. KW	5	6	288	5	1	1
23. KW	6	7	392	7	0	0
24. KW	7	8	465	8	0	0
25. KW	8	9	582	9	0	0
26. KW	9	10	623	10	0	0

Summe der quadrierten Rangplatzdifferenzen: 8

$$rho = 1 - \frac{6 * \sum d_i^2}{(n^2 - 1) * n} = 1 - \frac{6 * 8}{99 * 10} = 1 - \frac{48}{990} = 1 - 0,048 = 0,952$$

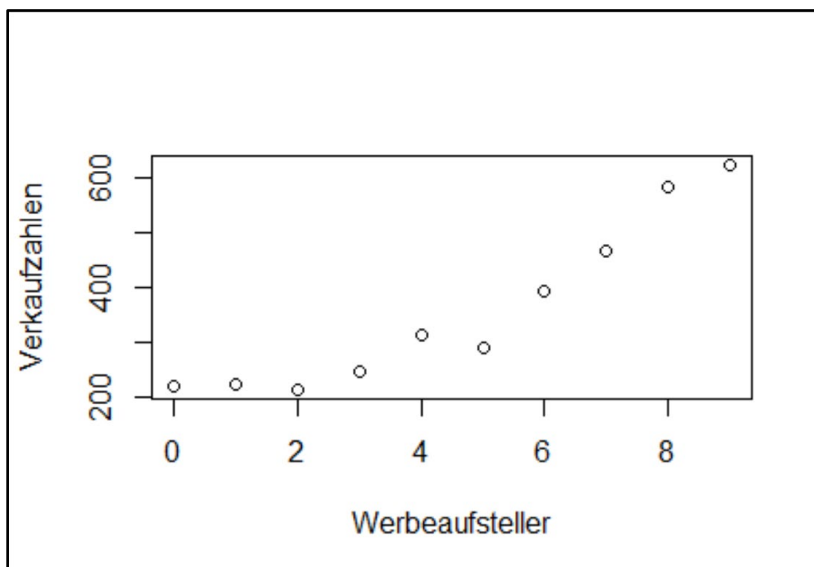
Es existiert eine starke positive monotone Korrelation.

b) Bravais-Pearson-Korrelationskoeffizient

KW	x	y	x ²	y ²	(x*y)
17. KW	0	217	0	47089	0
18. KW	1	221	1	48841	221
19. KW	2	213	4	45369	426
20. KW	3	245	9	60025	735
21. KW	4	311	16	96721	1244
22. KW	5	288	25	82944	1440
23. KW	6	392	36	153664	2352
24. KW	7	465	49	216225	3255
25. KW	8	582	64	338724	4656
26. KW	9	623	81	388129	5607
Summe	45	3557	285	1477731	19936
Mittel	4,5	355,7	-/-	-/-	-/-

$$r = \frac{\sum(x * y) - n * \bar{x} * \bar{y}}{\sqrt{\sum x^2 - n * \bar{x}^2} * \sqrt{\sum y^2 - n * \bar{y}^2}} = \frac{3929,5}{9,08 * 460,98} = 0,9388$$

Es existiert eine starke positive lineare Korrelation.



Sommersemester 2022

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2022

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 09. Juli 2022

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2022

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

- 1) Der Median ist derjenige Wert, der in der geordneten Verteilung am häufigsten auftritt.
☐ richtig ☐ falsch
- 2) Bei der Klassierung von Daten müssen stets unterschiedlich breite Klassen gebildet werden.
☐ richtig ☐ falsch
- 3) Bei der Klassenbildung gehören die Klassengrenzen jeweils zur Hälfte zu beiden Klassen.
☐ richtig ☐ falsch
- 4) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.
☐ richtig ☐ falsch
- 5) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.
☐ richtig ☐ falsch
- 6) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.
☐ richtig ☐ falsch
- 7) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.
☐ richtig ☐ falsch
- 8) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Outlier-Effekt bezeichnet.
☐ richtig ☐ falsch
- 9) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.
☐ richtig ☐ falsch
- 10) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.
☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Der Standard-Güterwaggontyp eines Bahnunternehmens kann eine maximale Nutzlast von 75 Tonnen transportieren. Im Rahmen einer betriebswirtschaftlichen Untersuchung ermittelt das Unternehmen die durchschnittliche Auslastung der Waggons in Abhängigkeit von unterschiedlichen Frachttypen. Dabei ergeben sich die nachfolgend tabellierten Werte.

Frachttyp	Durchschnittliche Nutzlast (in t)
Holz	62
Stahl	27
Schotter	63
Aluminium	73
Schüttgut	55
Saatgut	68
Papier	47
Kohle	69
Eisen	70
Sonstiges	67

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der durchschnittlichen Auslastungen und benennen Sie die hierfür erforderlichen Größen.

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Um zu evaluieren, ob an einem bislang unbeschränkten Bahnübergang eine Schranke eingerichtet werden sollte, wird über einen Zeitraum von 14 Tagen aufgezeichnet, wie viele Fahrzeuge den Bahnübergang jeweils in der Zeit zwischen 13:00 und 17:00 Uhr passieren. Es ergibt sich das nachfolgend tabellierte Bild.

Tag	Anzahl der Fahrzeuge
1	7
2	9
3	11
4	8
5	7
6	3
7	2
8	7
9	5
10	7
11	7
12	10
13	4
14	1

Bestimmen Sie für die hier dargelegte Verteilung

- das arithmetische Mittel (3 Punkte),
- den 35%-Perzentilwert (3 Punkte),
- den Modus (3 Punkte),
- die Spannweite (3 Punkte) und
- die Standardabweichung (3 Punkte).

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Ein privater Bahnbetreiber experimentiert auf einer von vielen Pendlerinnen und Pendlern genutzten Kurzstrecke mit neuen Preismodellen und bietet über zwei Monate jeweils (*ceteris paribus*) auf den Zeitraum einer Woche begrenzte Karten zu unterschiedlichen Preisen an. Die nachfolgende Tabelle zeigt die Entwicklung der Preise sowie die zugehörigen Ticketverkäufe.

Woche	Ticketpreis (in EUR)	Verkaufte Tickets
1	5,00	387
2	5,50	372
3	6,00	356
4	6,50	333
5	7,00	345
6	7,50	256
7	8,00	243
8	8,50	192

Berechnen (10 Punkte) und interpretieren (5 Punkte) Sie für die beiden Variablen „Ticketpreis“ und „verkaufte Tickets“ den Konkordanzkoeffizienten nach Kendall.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☐ falsch

2) Ein Würfel wird dann als fair bezeichnet, wenn die Wahrscheinlichkeit des Wurfs für jede Augenzahl genau gleich groß ist.

☐ richtig ☐ falsch

3) Kann die Nullhypothese eines statistischen Tests nicht verworfen werden, gilt die Alternativhypothese damit als widerlegt.

☐ richtig ☐ falsch

4) Das logische NICHT wird auch als Negation bezeichnet.

☐ richtig ☐ falsch

5) Der erste Schritt in jedem statistischen Testverfahren besteht in der Entscheidung, ob die Nullhypothese oder die Alternativhypothese getestet werden soll.

☐ richtig ☐ falsch

6) Ein Zufallsvorgang ist ein Vorgang, der in einem von mehreren möglichen disjunkten Ereignissen mündet. Welches Ereignis eintritt, kann vorab nicht mit Sicherheit ausgesagt werden.

☐ richtig ☐ falsch

7) Die zwei Axiome von Kolmogoroff definieren die mathematischen Eigenschaften des Modells mit bzw. ohne Zurücklegen.

☐ richtig ☐ falsch

8) Mengen, Schnittmengen und Vereinigungsmengen lassen sich grafisch mit Hilfe sogenannter Venn-Diagramme visualisieren.

☐ richtig ☐ falsch

9) Eine kommutative Gleichung zeichnet sich dadurch aus, dass sie mit -1 multipliziert werden kann, ohne dass sich ihr Ergebnis ändert.

☐ richtig ☐ falsch

10) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleichbleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

Aufgabenteil VI: Lineare Regression (15 Punkte)

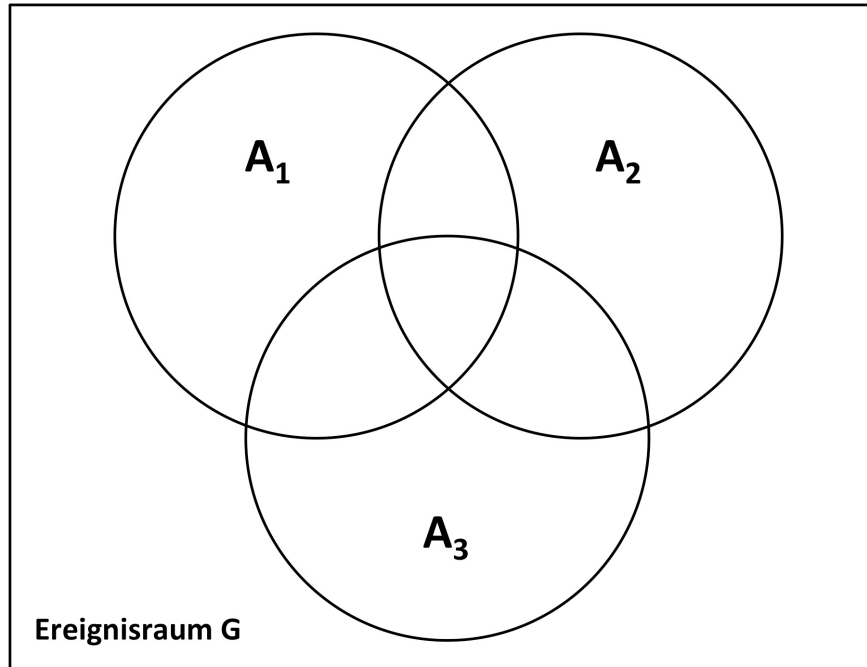
Woche	Ticketpreis (in EUR)	Verkaufte Tickets
1	5,00	387
2	5,50	372
3	6,00	356
4	6,50	333
5	7,00	345
6	7,50	256
7	8,00	243
8	8,50	192

Stellen Sie für die bereits bekannte Verteilung von Ticketpreisen und Ticketverkäufen ein lineares Regressionsmodell mit den Ticketpreisen als unabhängiger Variablen (x) und den Ticketverkäufen als abhängiger Variablen (y) auf (10 Punkte) und bestimmen und interpretieren Sie dessen Güte (5 Punkte).

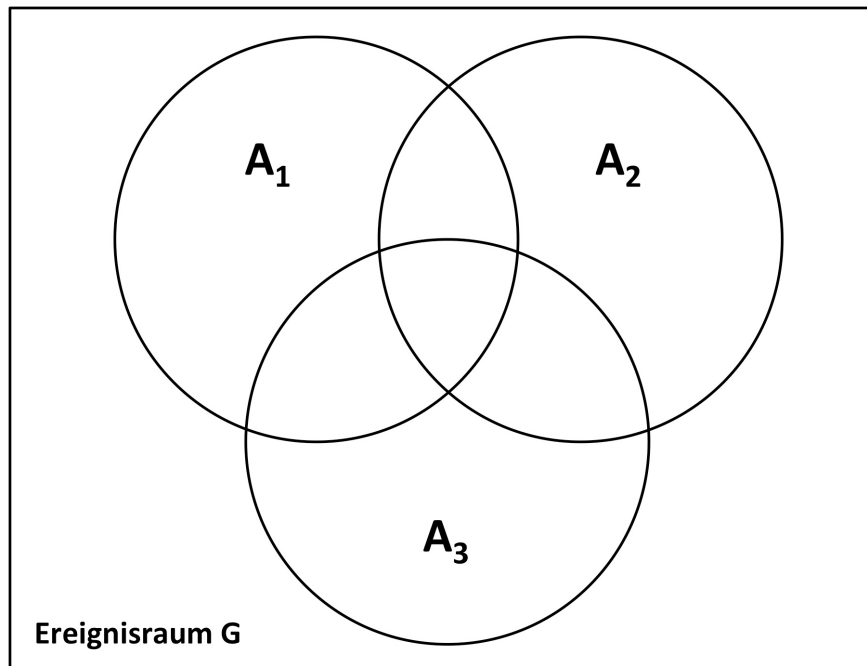
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_1 \cup A_2) \cap A_3$ (3 Punkte)



b) $(A_1 \cap \overline{A_1})$ (2 Punkte)



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

Ein Bahnreisender kann auf drei verschiedenen Wegen ein Ticket erwerben:

- Erste Wahl: Kauf am Bahnsteigautomaten
- Zweite Wahl bei defektem Automaten: Kauf per Handy-App
- Notlösung nur falls beides scheitert: Kauf beim Begleitpersonal im Zug

Sowohl der Kauf per Handy-App als auch der Kauf beim Begleitpersonal sind von einer funktionierenden mobilen Internetverbindung abhängig, die nicht immer gegeben ist. Grundsätzlich lassen sich zu den Chancen auf den reibungslosen Ablauf eines Fahrkartenkaufs folgende Feststellungen treffen:

- Beim Kauf am Bahnsteigautomaten besteht eine Wahrscheinlichkeit von 10% für einen Defekt, der den Abschluss des Kaufvorgangs unmöglich macht.
- Beim Kauf per Handy-App besteht eine Wahrscheinlichkeit von 20%, dass eine schlechte mobile Internetverbindung den Abschluss des Kaufvorgangs unmöglich macht.
- Beim Kauf beim Begleitpersonal im Zug ist dieses ebenfalls auf eine funktionierende mobile Internetverbindung angewiesen. Grundsätzlich liegt die Wahrscheinlichkeit für einen Ausfall auch hier bei 20% - hat aber schon der Kauf per Handy-App nicht funktioniert, steigt sie auf 50% an.

Beantworten Sie die folgenden zwei Fragen zu diesem Szenario:

- 1) Wie groß ist die Wahrscheinlichkeit dafür, gar keine Fahrkarte kaufen zu können? (2 Punkte)
- 2) Wie groß ist die Wahrscheinlichkeit dafür, das Begleitpersonal im Zug zwecks Kaufs einer Fahrkarte ansprechen zu müssen? (3 Punkte)

Aufgabenteil IX: Kombinatorik (5 Punkte)

An einer Bushaltestelle warten 8 Personen auf den gleichen Bus. Als dieser anhält, müssen sie aufgrund eines Defekts an der Hintertür alle durch die Vordertür einsteigen. In wie vielen verschiedenen Reihenfolgen könnten sich die Fahrgäste hierfür sortieren?

Aufgabenteil X: Satz von Bayes (10 Punkte)

Ein in einer U-Bahn-Station wird ein Videoüberwachungssystem installiert, das die Gesichter der Fahrgäste beim Ein- und Ausstieg erfasst und mit in einer Datenbank gespeicherten Fotos polizeilich gesuchter Personen vergleicht. Passiert eine gesuchte Person das Kamerasystem, wird sie mit einer sehr hohen Wahrscheinlichkeit von 98% korrekt identifiziert. Gleichzeitig werden aber auch 0,5% der Personen, deren Fotos nicht in der Datenbank auftauchen, fälschlicherweise als potentielle Gesuchte ausgewiesen. Man geht davon aus, dass unter den täglich rund 90.000 Personen, die diese U-Bahn-Station nutzen, etwa 6 gesuchte Personen zu erwarten sind.

In der Zentrale der Videoüberwachung ertönt ein Alarm. Wie hoch ist die Wahrscheinlichkeit dafür, dass es sich bei dem Systemtreffer wirklich um eine gesuchte Person handelt?

KLAUSUR STATISTIK I und II mit Lösungen

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2022

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Median ist derjenige Wert, der in der geordneten Verteilung am häufigsten auftritt.

O richtig **X falsch**

2) Bei der Klassierung von Daten müssen stets unterschiedlich breite Klassen gebildet werden.

O richtig **X falsch**

3) Bei der Klassenbildung gehören die Klassengrenzen jeweils zur Hälfte zu beiden Klassen.

O richtig **X falsch**

4) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

X richtig O falsch

5) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

X richtig O falsch

6) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

X richtig O falsch

7) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

X richtig O falsch

8) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Outlier-Effekt bezeichnet.

O richtig **X falsch**

9) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

O richtig **X falsch**

10) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

O richtig **X falsch**

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Der Standard-Güterwaggontyp eines Bahnunternehmens kann eine maximale Nutzlast von 75 Tonnen transportieren. Im Rahmen einer betriebswirtschaftlichen Untersuchung ermittelt das Unternehmen die durchschnittliche Auslastung der Waggons in Abhängigkeit von unterschiedlichen Frachttypen. Dabei ergeben sich die nachfolgend tabellierten Werte.

Frachttyp	Durchschnittliche Nutzlast (in t)
Holz	62
Stahl	27
Schotter	63
Aluminium	73
Schüttgut	55
Saatgut	68
Papier	47
Kohle	69
Eisen	70
Sonstiges	67

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der durchschnittlichen Auslastungen und benennen Sie die hierfür erforderlichen Größen.

Bildung einer geordneten Reihe:

27; 47; 55; 62; 63; 67; 68; 69; 70; 73

$x_{0,25} = 0,25 \cdot 10 = 2,5 \rightarrow$ Wert an der 3. Stelle der geordneten Verteilung $\rightarrow 55$

$x_{0,50} = 0,50 \cdot 10 = 5 \rightarrow$ Mittel der Werte an der 5. und 6. Stelle $\rightarrow (63 + 67) / 2 = 65$

$x_{0,75} = 0,75 \cdot 10 = 7,5 \rightarrow$ Wert an der 8. Stelle der geordneten Verteilung $\rightarrow 69$

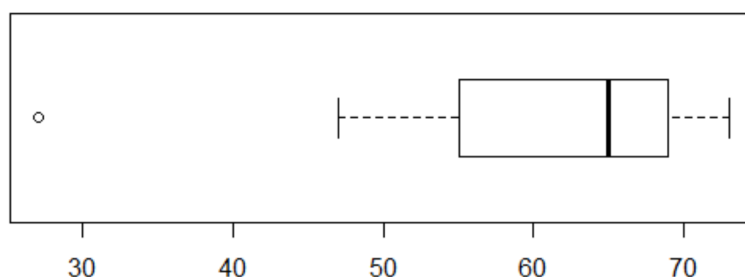
$IQR = 69 - 55 = 14$

1,5-facher IQR = 21

Der obere Zaun läuft maximal bis $69 + 21 = 90 \rightarrow$ größter Wert in diesem Bereich $\rightarrow 73$

Der untere Zaun läuft minimal bis $55 - 21 = 34 \rightarrow$ kleinster Wert in diesem Bereich $\rightarrow 47$

Es existiert ein Ausreißer (27)



Aufgabenteil III: Verteilungsparameter (15 Punkte)

Um zu evaluieren, ob an einem bislang unbeschränkten Bahnübergang eine Schranke eingerichtet werden sollte, wird über einen Zeitraum von 14 Tagen aufgezeichnet, wie viele Fahrzeuge den Bahnübergang jeweils in der Zeit zwischen 13:00 und 17:00 Uhr passieren. Es ergibt sich das nachfolgend tabellierte Bild.

Tag	Anzahl der Fahrzeuge
1	7
2	9
3	11
4	8
5	7
6	3
7	2
8	7
9	5
10	7
11	7
12	10
13	4
14	1

Bestimmen Sie für die hier dargelegte Verteilung

- a) das arithmetische Mittel (3 Punkte),
- b) den 35%-Perzentilwert (3 Punkte),
- c) den Modus (3 Punkte),
- d) die Spannweite (3 Punkte) und
- e) die Standardabweichung (3 Punkte).

a) $7 + 9 + 11 + 8 + 7 + 3 + 2 + 7 + 5 + 7 + 7 + 10 + 4 + 1 = 88 \rightarrow 88 / 14 = 6,29$

b) $0,35 \cdot 14 = 4,9 \rightarrow$ Wert an der 5. Stelle der geordneten Verteilung $\rightarrow 5$

c) 7

d) $11 - 1 = 10$

e) Arithmetisches Mittel: 6,29

$(7 - 6,29)^2 = 0,50$

$(9 - 6,29)^2 = 7,34$

$(11 - 6,29)^2 = 22,18$

$(8 - 6,29)^2 = 2,92$

$(7 - 6,29)^2 = 0,50$

$(3 - 6,29)^2 = 10,82$

$(2 - 6,29)^2 = 18,40$

$(7 - 6,29)^2 = 0,50$

$(5 - 6,29)^2 = 1,66$

$(7 - 6,29)^2 = 0,50$

$(7 - 6,29)^2 = 0,50$

$(10 - 6,29)^2 = 13,76$

$(4 - 6,29)^2 = 5,24$

$(1 - 6,29)^2 = 27,98$

Summe: 112,86

$112,86 / 14 = 8,06$

Wurzel aus 8,06 = 2,84

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Ein privater Bahnbetreiber experimentiert auf einer von vielen Pendlerinnen und Pendlern genutzten Kurzstrecke mit neuen Preismodellen und bietet über zwei Monate jeweils (*ceteris paribus*) auf den Zeitraum einer Woche begrenzte Karten zu unterschiedlichen Preisen an. Die nachfolgende Tabelle zeigt die Entwicklung der Preise sowie die zugehörigen Ticketverkäufe.

Woche	Ticketpreis (in EUR)	Verkaufte Tickets
1	5,00	387
2	5,50	372
3	6,00	356
4	6,50	333
5	7,00	345
6	7,50	256
7	8,00	243
8	8,50	192

Berechnen (10 Punkte) und interpretieren (5 Punkte) Sie für die beiden Variablen „Ticketpreis“ und „verkaufte Tickets“ den Konkordanzkoeffizienten nach Kendall.

Woche	Ticketpreis (in EUR)	Verkaufte Tickets	Rang	K	D
1	5,00	387	8	0	7
2	5,50	372	7	0	6
3	6,00	356	6	0	5
4	6,50	333	4	1	3
5	7,00	345	5	0	3
6	7,50	256	3	0	2
7	8,00	243	2	0	1
8	8,50	192	1	-	-

Summe konkordanter Paare: 1

Summe diskordanter Paare: 27

$$\tau = \frac{2 * (K - D)}{n * (n - 1)} = \frac{2 * (-26)}{8 * 7} = -0,9286$$

Es liegt eine monotone, negative und starke Korrelation vor.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die abhängige (Y) sowie die unabhängige (X) Variable der einfachen linearen Regressionsanalyse dürfen sich gegenseitig beeinflussen (interdependente Beziehung).

☐ richtig ☒ **falsch**

2) Ein Würfel wird dann als fair bezeichnet, wenn die Wahrscheinlichkeit des Wurfs für jede Augenzahl genau gleich groß ist.

☒ **richtig** ☐ falsch

3) Kann die Nullhypothese eines statistischen Tests nicht verworfen werden, gilt die Alternativhypothese damit als widerlegt.

☐ richtig ☒ **falsch**

4) Das logische NICHT wird auch als Negation bezeichnet.

☒ **richtig** ☐ falsch

5) Der erste Schritt in jedem statistischen Testverfahren besteht in der Entscheidung, ob die Nullhypothese oder die Alternativhypothese getestet werden soll.

☐ richtig ☒ **falsch**

6) Ein Zufallsvorgang ist ein Vorgang, der in einem von mehreren möglichen disjunkten Ereignissen mündet. Welches Ereignis eintritt, kann vorab nicht mit Sicherheit ausgesagt werden.

☒ **richtig** ☐ falsch

7) Die zwei Axiome von Kolmogoroff definieren die mathematischen Eigenschaften des Modells mit bzw. ohne Zurücklegen.

☐ richtig ☒ **falsch**

8) Mengen, Schnittmengen und Vereinigungsmengen lassen sich grafisch mit Hilfe sogenannter Venn-Diagramme visualisieren.

☒ **richtig** ☐ falsch

9) Eine kommutative Gleichung zeichnet sich dadurch aus, dass sie mit -1 multipliziert werden kann, ohne dass sich ihr Ergebnis ändert.

☐ richtig ☒ **falsch**

10) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleichbleibenden Rahmenbedingungen.

☒ **richtig** ☐ falsch

Aufgabenteil VI: Lineare Regression (15 Punkte)

Woche	Ticketpreis (in EUR)	Verkaufte Tickets
1	5,00	387
2	5,50	372
3	6,00	356
4	6,50	333
5	7,00	345
6	7,50	256
7	8,00	243
8	8,50	192

Stellen Sie für die bereits bekannte Verteilung von Ticketpreisen und Ticketverkäufen ein lineares Regressionsmodell mit den Ticketpreisen als unabhängiger Variablen (x) und den Ticketverkäufen als abhängiger Variablen (y) auf (10 Punkte) und bestimmen und interpretieren Sie dessen Güte (5 Punkte).

	x	y	x ²	y ²	(x*y)
	5,00	387	25,00	149769	1935
	5,50	372	30,25	138384	2046
	6,00	356	36,00	126736	2136
	6,50	333	42,25	110889	2164,50
	7,00	345	49,00	119025	2415
	7,50	256	56,25	65536	1920
	8,00	243	64,00	59049	1944
	8,50	192	72,25	36864	1632
Summe	54	2484	375	806252	16192,50
Mittel	6,75	310,5			

Aufstellung des Regressionsmodells

$$b = \frac{\sum(x_i * y_i) - n * \bar{x} * \bar{y}}{\sum x_i^2 - n * \bar{x}^2} = \frac{16192,5 - 8 * 6,75 * 310,5}{375 - 8 * 6,75^2} = \frac{-574,5}{10,5} = -54,71$$

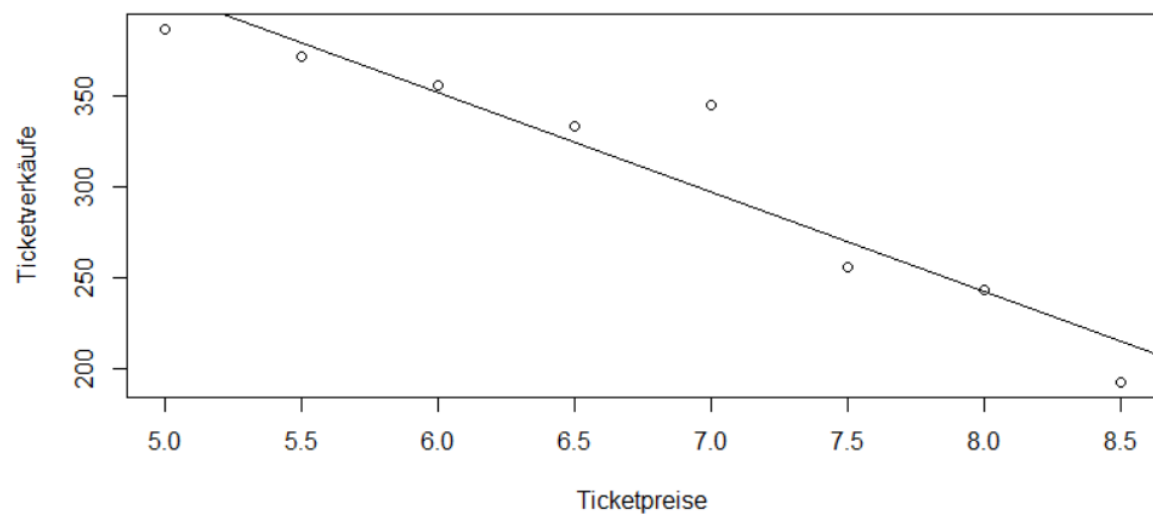
$$a = \bar{y} - b * \bar{x} = 310,5 + 54,71 * 6,75 = 679,82$$

$$\text{Tickets} = 679,82 - 54,71 * \text{Preis}$$

Prüfung der Güte

$$r = \frac{\sum(x_i * y_i) - n * \bar{x} * \bar{y}}{\sqrt{\sum x_i^2 - n * \bar{x}^2} * \sqrt{\sum y_i^2 - n * \bar{y}^2}} = \frac{16192,5 - 8 * 6,75 * 310,5}{3,24 * 187} = \frac{-574,5}{605,96} = -0,95$$

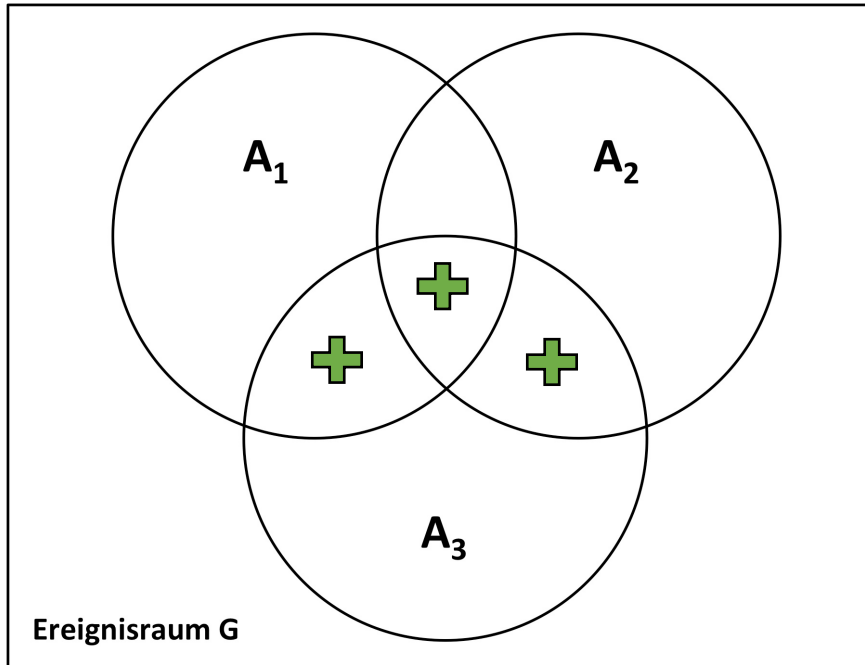
$$R^2=0,9$$



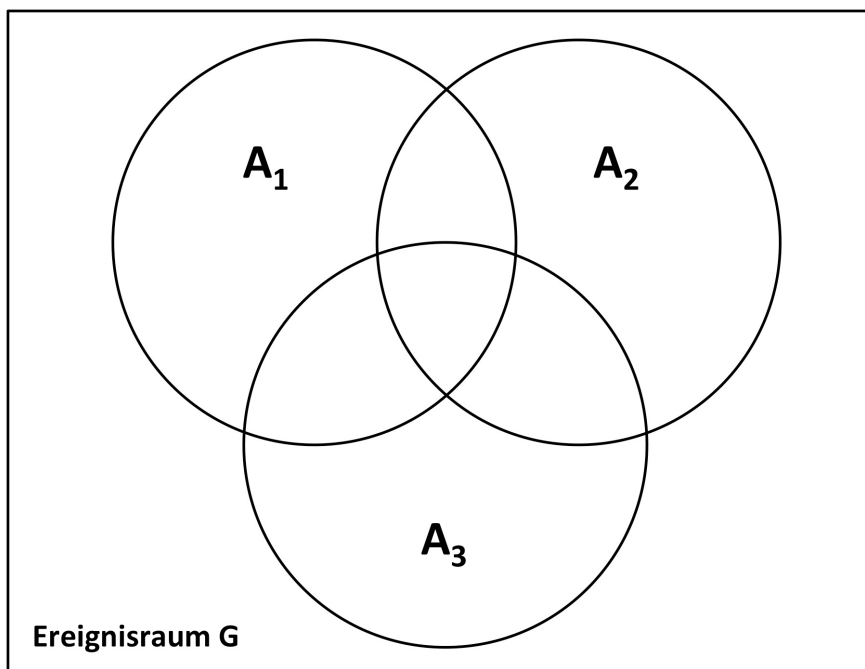
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_1 \cup A_2) \cap A_3$ (3 Punkte) -> Alle drei Schnittflächen von A_3 mit A_1 und A_2



b) $(A_1 \cap \overline{A_1})$ (2 Punkte) -> gar nichts



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

Ein Bahnreisender kann auf drei verschiedenen Wegen ein Ticket erwerben:

- Erste Wahl: Kauf am Bahnsteigautomaten
- Zweite Wahl bei defektem Automaten: Kauf per Handy-App
- Notlösung nur falls beides scheitert: Kauf beim Begleitpersonal im Zug

Sowohl der Kauf per Handy-App als auch der Kauf beim Begleitpersonal sind von einer funktionierenden mobilen Internetverbindung abhängig, die nicht immer gegeben ist. Grundsätzlich lassen sich zu den Chancen auf den reibungslosen Ablauf eines Fahrkartenkaufs folgende Feststellungen treffen:

- Beim Kauf am Bahnsteigautomaten besteht eine Wahrscheinlichkeit von 10% für einen Defekt, der den Abschluss des Kaufvorgangs unmöglich macht.
- Beim Kauf per Handy-App besteht eine Wahrscheinlichkeit von 20%, dass eine schlechte mobile Internetverbindung den Abschluss des Kaufvorgangs unmöglich macht.
- Beim Kauf beim Begleitpersonal im Zug ist dieses ebenfalls auf eine funktionierende mobile Internetverbindung angewiesen. Grundsätzlich liegt die Wahrscheinlichkeit für einen Ausfall auch hier bei 20% - hat aber schon der Kauf per Handy-App nicht funktioniert, steigt sie auf 50% an.

Beantworten Sie die folgenden zwei Fragen zu diesem Szenario:

- 1) Wie groß ist die Wahrscheinlichkeit dafür, gar keine Fahrkarte kaufen zu können? (2 Punkte)
- 2) Wie groß ist die Wahrscheinlichkeit dafür, das Begleitpersonal im Zug zwecks Kaufs einer Fahrkarte ansprechen zu müssen? (3 Punkte)

1) Fehler – Fehler – Fehler $\rightarrow 0,10 * 0,20 * 0,50 = 0,01$

2) Fehler – Fehler $\rightarrow 0,10 * 0,20 = 0,02$

Aufgabenteil IX: Kombinatorik (5 Punkte)

An einer Bushaltestelle warten 8 Personen auf den gleichen Bus. Als dieser anhält, müssen sie aufgrund eines Defekts an der Hintertür alle durch die Vordertür einsteigen. In wie vielen verschiedenen Reihenfolgen könnten sich die Fahrgäste hierfür sortieren?

Spielt die Reihenfolge eine Rolle? $\rightarrow$ JA $\rightarrow$ Variation

Handelt es sich um ein Modell mit oder ohne Zurücklegen? $\rightarrow$ Ohne Zurücklegen

Werden alle Elemente oder nur eine Auswahl gezogen? $\rightarrow$ Alle Elemente $\rightarrow$ Permutation

$n!$

$8! = 40.320$

Aufgabenteil X: Satz von Bayes (10 Punkte)

Ein in einer U-Bahn-Station wird ein Videoüberwachungssystem installiert, das die Gesichter der Fahrgäste beim Ein- und Ausstieg erfasst und mit in einer Datenbank gespeicherten Fotos polizeilich gesuchter Personen vergleicht. Passiert eine gesuchte Person das Kamerasystem, wird sie mit einer sehr hohen Wahrscheinlichkeit von 98% korrekt identifiziert. Gleichzeitig werden aber auch 0,5% der Personen, deren Fotos nicht in der Datenbank auftauchen, fälschlicherweise als potentielle Gesuchte ausgewiesen. Man geht davon aus, dass unter den täglich rund 90.000 Personen, die diese U-Bahn-Station nutzen, etwa 6 gesuchte Personen zu erwarten sind.

In der Zentrale der Videoüberwachung ertönt ein Alarm. Wie hoch ist die Wahrscheinlichkeit dafür, dass es sich bei dem Systemtreffer wirklich um eine gesuchte Person handelt?

90.000 Personen

-> davon 6 gesuchte

-> davon 89.994 weitere

Von den 6 gesuchten Personen

-> werden 98% erkannt -> 5,88

-> werden 2% nicht erkannt -> 0,12

Von den 89.994 weiteren Personen

-> werden 0,5% erkannt -> 449,97

-> werden 99,5% nicht erkannt -> 89.544,03

Insgesamt werden also $449,97 + 5,88 = 455,85$ Personen „erkannt“, davon 5,88 korrekt (1,3%).

Sommersemester 2023

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang Betriebswirtschaftslehre
Sommersemester 2023

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 29. April 2023

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung (mit Anmerkungen), Taschenrechner.
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2023

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

2) Zwei Teilgesamtheiten einer Grundgesamtheit können sich niemals überschneiden.

☐ richtig ☐ falsch

3) Qualitative und quantitative Methoden der Datenerhebung und Datenanalyse können grundsätzlich nicht miteinander kombiniert werden.

☐ richtig ☐ falsch

4) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☐ falsch

5) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig ☐ falsch

6) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☐ richtig ☐ falsch

7) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☐ richtig ☐ falsch

8) Die Berechnung der Varianz setzt mindestens ordinalskalierte Daten voraus.

☐ richtig ☐ falsch

9) Objektivität, Reliabilität und Validität sind die einzigen relevanten Kriterien für die Bewertung der Güte erhobener Daten.

☐ richtig ☐ falsch

10) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Eine Umfrage unter Professor*innen am Fachbereich Biowissenschaften einer kleineren Universität erbrachte folgende Angaben zur Anzahl der Jahre an persönlicher Forschungserfahrung.

Anonymisierte Teilnehmerkennung	Bisherige Forschungserfahrung (in Jahren)
OV37KS	17
OV28ER	8
FG13WS	3
DV66AZ	21
TF45WR	12
GH73TG	20
UF84FB	17
OV33ZU	16
ED85FT	7
FV75DD	5

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Jahresangaben und benennen Sie die hierfür erforderlichen Größen (15 Punkte). Wie viel Forschungserfahrung müsste ein/e zusätzliche/r (!) Professor*in aufweisen, um im Box-Plot als Ausreißer im oberen Bereich sichtbar zu werden? (5 Punkte)

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Die schon aus der vorherigen Aufgabe bekannten Daten sollen für das Forschungsberichtswesen der Hochschule weiter aufbereitet werden.

Anonymisierte Teilnehmerkennung	Bisherige Forschungserfahrung (in Jahren)
OV37KS	17
OV28ER	8
FG13WS	3
DV66AZ	21
TF45WR	12
GH73TG	20
UF84FB	17
OV33ZU	16
ED85FT	7
FV75DD	5

Berechnen Sie für die Angaben zur Forschungserfahrung die folgenden Lage- und Streuungsmaße:

- Arithmetisches Mittel (6 Punkte)
- Um 5% getrimmtes arithmetisches Mittel (6 Punkte)
- Modus (4 Punkte)
- Spannweite (4 Punkte)
- Varianz (7 Punkte)
- Standardabweichung (3 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (40 Punkte)

Neben der bisherigen Forschungserfahrung wird unter anderem auch erhoben, an wie vielen begutachteten (sog. peer-reviewten) Publikationen die jeweiligen Professor*innen im Vorjahr beteiligt waren.

Anonymisierte Teilnehmerkennung	Bisherige Forschungserfahrung (in Jahren)	Peer-reviewte Publikationen im Vorjahr
OV37KS	17	11
OV28ER	8	5
FG13WS	3	2
DV66AZ	21	15
TF45WR	12	7
GH73TG	20	17
UF84FB	17	8
OV33ZU	16	9
ED85FT	7	3
FV75DD	5	3

Berechnen und interpretieren Sie für die vorliegenden Daten a) den Rangkorrelationskoeffizienten nach Spearman (20 Punkte) sowie b) den Bravais-Pearson-Korrelationskoeffizienten (20 Punkte).

MUSTERLÖSUNG KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2023

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☒ **falsch**

2) Zwei Teilgesamtheiten einer Grundgesamtheit können sich niemals überschneiden.

☐ richtig ☒ **falsch**

3) Qualitative und quantitative Methoden der Datenerhebung und Datenanalyse können grundsätzlich nicht miteinander kombiniert werden.

☐ richtig ☒ **falsch**

4) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☒ **richtig** ☐ falsch

5) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☒ **richtig** ☐ falsch

6) Die kumulierte empirische Verteilungsfunktion lässt sich grafisch als Treppenfunktion darstellen.

☒ **richtig** ☐ falsch

7) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☒ **richtig** ☐ falsch

8) Die Berechnung der Varianz setzt mindestens ordinalskalierte Daten voraus.

☐ richtig ☒ **falsch**

9) Objektivität, Reliabilität und Validität sind die einzigen relevanten Kriterien für die Bewertung der Güte erhobener Daten.

☐ richtig ☒ **falsch**

10) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☒ **falsch**

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Eine Umfrage unter Professor*innen am Fachbereich Biowissenschaften einer kleineren Universität erbrachte folgende Angaben zur Anzahl der Jahre an persönlicher Forschungserfahrung.

Anonymisierte Teilnehmerkennung	Bisherige Forschungserfahrung (in Jahren)
OV37KS	17
OV28ER	8
FG13WS	3
DV66AZ	21
TF45WR	12
GH73TG	20
UF84FB	17
OV33ZU	16
ED85FT	7
FV75DD	5

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Jahresangaben und benennen Sie die hierfür erforderlichen Größen (15 Punkte). Wie viel Forschungserfahrung müsste ein/e zusätzliche/r (!) Professor*in aufweisen, um im Box-Plot als Ausreißer im oberen Bereich sichtbar zu werden? (5 Punkte)

Konstruktion des Box-Plots

Geordnete Wertereihe: 3, 5, 7, 8, 12, 16, 17, 17, 20, 21

$x_{0,25} \rightarrow 0,25 \cdot 10 = 2,5 \rightarrow$ Wert an der 3. Stelle der geordneten Verteilung $\rightarrow$ 7

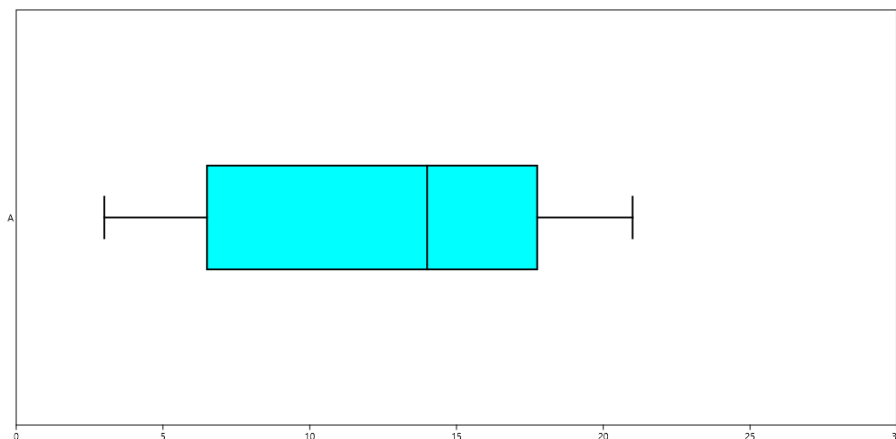
$x_{0,50} \rightarrow 0,50 \cdot 10 = 5,0 \rightarrow$ Mittel der Werte an der 5. und 6. Stelle $\rightarrow (12+16)/2 =$ 14

$x_{0,75} \rightarrow 0,75 \cdot 10 = 7,5 \rightarrow$ Wert an der 8. Stelle der geordneten Verteilung $\rightarrow$ 17

$IQR = 17 - 7 =$ 10

1,5-facher IQR = 15

Die Box läuft damit von 7 bis 17, der Median wird bei 14 eingezeichnet. Der untere Zaun könnte minimal bis $7-15=-8$ laufen, der kleinste Wert in diesem Bereich ist die 3. Nach unten gibt es somit keine Ausreißer oder Extremwerte. Der obere Zaun könnte maximal bis $17+15=32$ laufen, der größte Wert in diesem Bereich ist die 21. Auch nach oben gibt es somit keine Ausreißer oder Extremwerte.



Zusatzfrage

Wie viel Forschungserfahrung müsste ein/e zusätzliche/r (!) Professor*in aufweisen, um im Box-Plot als Ausreißer im oberen Bereich sichtbar zu werden?

Einfache Lösung ohne Berücksichtigung der Veränderungen in der Größe des Datensatzes:

Ein/e solche/r Professor*in müsste mindestens **33** (oder zumindest mehr als 32) Jahre an Forschungserfahrung aufweisen, um als Ausreißer im oberen Bereich sichtbar zu werden.

Komplexe Lösung unter Berücksichtigung der Veränderungen in der Größe des Datensatzes:

Bei einem zusätzlichen Professor hätte der Datensatz 11 Werte, der zusätzliche Wert soll am oberen Ende liegen. Die neue geordnete Reihe sieht somit wie folgt aus:

3, 5, 7, 8, 12, 16, 17, 17, 20, 21, a

Unabhängig von der Größe von a lassen sich neue Werte für die Ausreißeridentifikation bestimmen:

$X_{0,25} \rightarrow 0,25 \cdot 11 = 2,75 \rightarrow$ Wert an der 3. Stelle der geordneten Verteilung $\rightarrow 7$

$x_{0,75} \rightarrow 0,75 \cdot 11 = 8,25 \rightarrow$ Wert an der 9. Stelle der geordneten Verteilung $\rightarrow 20$

$IQR = 20 - 7 = 13$ 1,5-facher IQR = 19,5

Der obere Zaun würde damit maximal bis $(20 + 19,5) = 39,5$ laufen. Ein zusätzlicher (!) Professor müsste damit mindestens **40** (oder zumindest mehr als 39,5) Jahre an Forschungserfahrung aufweisen, um als Ausreißer im oberen Bereich sichtbar zu werden.

Aufgabenteil III: Verteilungsparameter (30 Punkte)

Die schon aus der vorherigen Aufgabe bekannten Daten sollen für das Forschungsberichtswesen der Hochschule weiter aufbereitet werden.

Anonymisierte Teilnehmerkennung	Bisherige Forschungserfahrung (in Jahren)
OV37KS	17
OV28ER	8
FG13WS	3
DV66AZ	21
TF45WR	12
GH73TG	20
UF84FB	17
OV33ZU	16
ED85FT	7
FV75DD	5

Berechnen Sie für die Angaben zur Forschungserfahrung die folgenden Lage- und Streuungsmaße:

- Arithmetisches Mittel (6 Punkte)
- Um 5% getrimmtes arithmetisches Mittel (6 Punkte)
- Modus (4 Punkte)
- Spannweite (4 Punkte)
- Varianz (7 Punkte)
- Standardabweichung (3 Punkte)

Arithmetisches Mittel

$$17 + 8 + 3 + 21 + 12 + 20 + 17 + 16 + 7 + 5 = 126$$

$$126/10 = \underline{12,6}$$

Um 5% getrimmtes arithmetisches Mittel

0,05 * 10 = 0,5 -> Kürzung um jeweils den größten und kleinsten Wert im Datensatz (3 und 21)

$$17 + 8 + 12 + 20 + 17 + 16 + 7 + 5 = 102$$

$$102/8 = \underline{12,75}$$

Modus

Die Verteilung ist nicht unimodal, weshalb sie auch **keinen Modus** aufweist.

Spannweite

$$21 - 3 = \underline{18}$$

Varianz

Arithmetisches Mittel: 12,6

$$(17-12,6)^2 = 19,36$$

$$(20-12,6)^2 = 54,76$$

$$(8-12,6)^2 = 21,16$$

$$(17-12,6)^2 = 19,36$$

$$(3-12,6)^2 = 92,16$$

$$(16-12,6)^2 = 11,56$$

$$(21-12,6)^2 = 70,56$$

$$(7-12,6)^2 = 31,36$$

$$(12-12,6)^2 = 0,36$$

$$(5-12,6)^2 = 57,76$$

Summe der quadrierten Differenzen vom arithmetischen Mittel: 378,4

$$378,4 / 10 = \underline{37,84}$$

Standardabweichung

Positive Wurzel aus der Varianz -> **6,15**

Aufgabenteil IV: Zusammenhangsmaße (40 Punkte)

Neben der bisherigen Forschungserfahrung wird unter anderem auch erhoben, an wie vielen begutachteten (sog. peer-reviewten) Publikationen die jeweiligen Professor*innen im Vorjahr beteiligt waren.

Anonymisierte Teilnehmerkennung	Bisherige Forschungserfahrung (in Jahren)	Peer-reviewte Publikationen im Vorjahr
OV37KS	17	11
OV28ER	8	5
FG13WS	3	2
DV66AZ	21	15
TF45WR	12	7
GH73TG	20	17
UF84FB	17	8
OV33ZU	16	9
ED85FT	7	3
FV75DD	5	3

Berechnen und interpretieren Sie für die vorliegenden Daten a) den Rangkorrelationskoeffizienten nach Spearman (20 Punkte) sowie b) den Bravais-Pearson-Korrelationskoeffizienten (20 Punkte).

Spearman

x	rg(x)	y	rg(y)	d	d ²
17	7,5	11	8	-0,5	0,25
8	4	5	4	0	0
3	1	2	1	0	0
21	10	15	9	1	1
12	5	7	5	0	0
20	9	17	10	-1	1
17	7,5	8	6	1,5	2,25
16	6	9	7	-1	1
7	3	3	2,5	0,5	0,25
5	2	3	2,5	-0,5	0,25

Summe aller quadrierten Rangplatzdifferenzen: 6

$$\rho = 1 - (6 \cdot 6) / [(10^2 - 1) \cdot 10] = 1 - 36 / 990 = \underline{\underline{0,9636}}$$

Es liegt eine **starke, positive, monotone** Korrelation vor.

Zwei **verbundene Ränge** verringern die Aussagekraft des Koeffizienten.

Bravais-Pearson

n	x	y	x ²	y ²	(x*y)
1	17	11	289	121	187
2	8	5	64	25	40
3	3	2	9	4	6
4	21	15	441	225	315
5	12	7	144	49	84
6	20	17	400	289	340
7	17	8	289	64	136
8	16	9	256	81	144
9	7	3	49	9	21
10	5	3	25	9	15
SUMME	126	80	1966	876	1288
MITTEL	12,6	8	-//-	-//-	-//-

$$1288 - 10 \cdot 12,6 \cdot 8 = 280$$

$$\sqrt{1966 - 10 \cdot 12,6^2} = 19,45$$

$$\sqrt{876 - 10 \cdot 8^2} = 15,36$$

$$1288 / 19,45 \cdot 15,36 = \underline{\underline{0,9369}}$$

Es liegt eine **starke, positive, lineare** Korrelation vor.

Sommersemester 2023

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2023

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 08. Juli 2023

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2023

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig☐ falsch

2) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig☐ falsch

3) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig☐ falsch

4) Eine linkssteile Verteilung ist immer auch rechtsschief, eine rechtssteile Verteilung ist immer auch linksschief.

☐ richtig☐ falsch

5) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen 0 und 1 annehmen.

☐ richtig☐ falsch

6) Nominalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig☐ falsch

7) Der Interquartilsabstand (IQR) ist ein robusteres Streuungsmaß als die Spannweite.

☐ richtig☐ falsch

8) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

☐ richtig☐ falsch

9) Aus Stem-and-Leaf-Diagrammen kann man die Ursprungsdaten direkt ablesen.

☐ richtig☐ falsch

10) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☐ richtig☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Im Rahmen einer archäologischen Ausgrabung werden 15 Tonscherben von Vasen aus der spätminoischen Phase (1.600 bis 1.450 v. Christus) gesichert, untersucht und katalogisiert.

Es ergibt sich folgende Übersicht:

Objektnummer	Höhe in mm	Breite in mm
A2312-001	53	76
A2312-002	55	23
A2312-003	23	78
A2312-004	76	45
A2312-005	25	37
A2312-006	34	82
A2312-007	55	21
A2312-008	28	75
A2312-009	69	63
A2312-010	43	53
A2312-011	40	57
A2312-012	16	43
A2312-013	23	28
A2312-014	128	91
A2312-015	51	60



Stellen Sie für die Verteilung der Angaben zur Höhe der Bruchstücke ein Stamm-Blatt-Diagramm mit einer aus Ihrer Sicht geeigneten Stammbreite auf. (7 Punkte) Zeichnen Sie zusätzlich noch das Stamm-Blatt-Diagramm mit einer Stammbreite von 50. (3 Punkte)

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Wir werfen erneut einen Blick auf die uns bereits bekannten Objektdaten.

Objektnummer	Höhe in mm	Breite in mm
A2312-001	53	76
A2312-002	55	23
A2312-003	23	78
A2312-004	76	45
A2312-005	25	37
A2312-006	34	82
A2312-007	55	21
A2312-008	28	75
A2312-009	69	63
A2312-010	43	53
A2312-011	40	57
A2312-012	16	43
A2312-013	23	28
A2312-014	128	91
A2312-015	51	60

Berechnen Sie für die Angaben zur Breite der Bruchstücke folgende Kennzahlen:

- a) Arithmetisches Mittel (3 Punkte)
- b) Um 10% getrimmtes arithmetisches Mittel (2 Punkte)
- c) Median (3 Punkte)
- d) Interquartilsabstand (2 Punkte)
- e) Varianz (3 Punkte)
- f) Standardabweichung (2 Punkte)

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Ein Museum senkt (ceteris paribus) über einen Zeitraum von 10 Tagen die Preise für den Eintritt in die Dauerausstellung um jeweils 50 Cent pro Tag und erfasst die Anzahl der täglichen Besucher*innen. Es ergeben sich folgende Werte.

Tag	Eintritt in EUR	Anzahl Besucher*innen
03.07.2023	7,50	81
04.07.2023	7,00	92
05.07.2023	6,50	94
06.07.2023	6,00	108
07.07.2023	5,50	107
08.07.2023	5,00	123
10.07.2023	4,50	154
11.07.2023	4,00	211
12.07.2023	3,50	227
13.07.2023	3,00	245

Berechnen und interpretieren Sie sowohl der Rangkorrelationskoeffizienten nach Spearman (7 Punkte) als auch den Bravais-Pearson-Korrelationskoeffizienten (8 Punkte) für die beiden hier aufgeführten Variablen Eintrittspreis und Besuchszahlen.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

2) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

☐ richtig ☐ falsch

3) Das logische UND wird auch als Konjunktion bezeichnet.

☐ richtig ☐ falsch

4) Eine gültige Teilmenge muss stets aus mindestens zwei Elementen bestehen.

☐ richtig ☐ falsch

5) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

☐ richtig ☐ falsch

6) Ein Würfel wird dann als fair bezeichnet, wenn die Wahrscheinlichkeit des Wurfs für jede Augenzahl genau gleich groß ist.

☐ richtig ☐ falsch

7) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig ☐ falsch

8) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht der Wurzel aus dem Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

9) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

10) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert der Steigung der linearen Regressionsgeraden.

☐ richtig ☐ falsch

Aufgabenteil VI: Lineare Regression (15 Punkte)

Wir werfen erneut einen Blick auf die uns bereits bekannten Daten zu den Museumsbesuchen.

Tag	Eintritt in EUR	Anzahl Besucher*innen
03.07.2023	7,50	81
04.07.2023	7,00	92
05.07.2023	6,50	94
06.07.2023	6,00	108
07.07.2023	5,50	107
08.07.2023	5,00	123
10.07.2023	4,50	154
11.07.2023	4,00	211
12.07.2023	3,50	227
13.07.2023	3,00	245

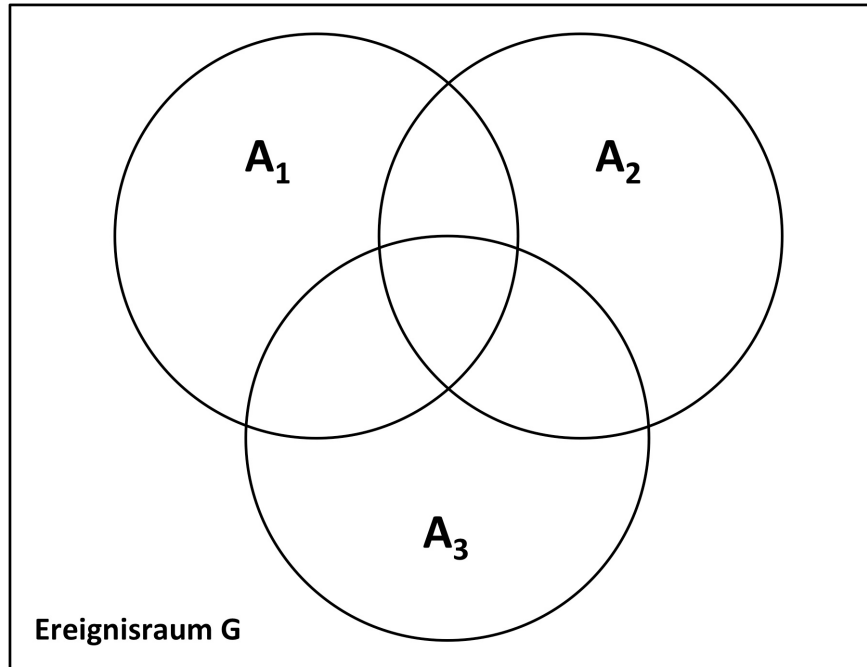
Stellen Sie ein lineares Regressionsmodell mit dem Eintrittspreis als unabhängiger Variable (x) und der Anzahl Besucher*innen als abhängiger Variable (y) auf. (7 Punkte) Berechnen und interpretieren Sie außerdem das Gütemaß R^2 für das von Ihnen aufgestellte Modell. (5 Punkte)

Hätte man das Modell mathematisch wie auch inhaltlich betrachtet auch umgekehrt mit der Anzahl Besucher*innen als unabhängiger Variable (x) und den Eintrittspreisen als abhängiger Variable (y) aufstellen können? (3 Punkte)

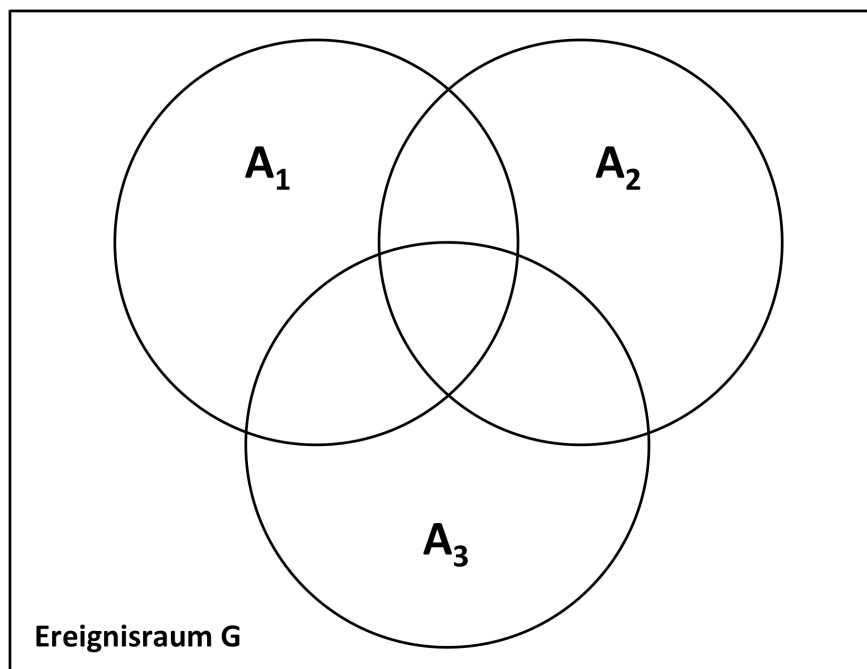
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_1 \cap A_2) \cap A_3$ (3 Punkte)



b) $\overline{A_2}$ (2 Punkte)



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

Ein Einbrecher versucht, eine diamantbesetzte Krone aus dem Fundus eines Museums zu entwenden. Damit ihm dies gelingt, muss er zwei Sicherheitssysteme nacheinander überwinden: Einen Fingerabdruck-Sensor an der Eingangstür des Lagerraums sowie eine Lichtschranke in der Mitte des Raumes. Dies gelingt ihm bei dem Fingerabdruck-Sensor mit einer Wahrscheinlichkeit von 65% sowie bei der Lichtschranke mit einer Wahrscheinlichkeit von 80%, wenn er über seine volle Konzentrationsfähigkeit verfügt. Streitet er sich jedoch am Abend vor dem Einbruch mit seiner Lebensgefährtin, sinkt die Wahrscheinlichkeit auf 55% beim Fingerabdruck-Sensor sowie auf 70% bei der Lichtschranke. Die Wahrscheinlichkeit auf einen Streit liegt an jedem beliebigen Abend bei 50%. Mit welcher Wahrscheinlichkeit gelingt der Diebstahl der Krone?

**Aufgabenteil IX: Kombinatorik (5 Punkte)**

Eine Besuchergruppe bestehend aus 12 Personen besucht die Gemäldeausstellung eines Museums. Vor einem besonders beeindruckenden Gemälde möchten alle 12 auf einer Bank Platz nehmen, um sich in Ruhe den Audioguide anzuhören. Leider bietet die Bank nur 4 Sitzplätze.

- a) In wie vielen möglichen Zusammenstellungen können 4 der 12 Personen auf der Bank Platz nehmen? (3 Punkte)
- b) Angenommen, die Gruppe bestünde aus 6 Frauen und 6 Männern. Wie viele Zusammenstellungen von 4 Personen gibt es, denen genau 2 Männer und 2 Frauen angehören? (2 Punkte)

**Aufgabenteil X: Satz von Bayes (10 Punkte)**

Eine Archäologin begeht eine in 380 Flächensegmente unterteilte potentielle Fundstelle auf der Suche nach im Boden lagernden Metallfragmenten mit einer Bodensonde. Die Bodensonde kann solche Metallfragmente innerhalb eines Flächensegments mit einer sehr hohen Wahrscheinlichkeit von 97% detektieren. Allerdings meldet sie auch für 8% aller Flächensegmente ohne spätere Fundstücke das Vorhandensein von Metallteilen. Aus Erfahrungen mit ähnlichen Fundstellen lässt sich ableiten, dass in etwa 5% aller Flächensegmente echte Funde zu erwarten sind.

Die Bodensonde hat gerade angeschlagen.

Wie hoch ist die Wahrscheinlichkeit auf einen echten Treffer?

KLAUSUR STATISTIK I und II mit Lösungen

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2023

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Das arithmetische Mittel kann nur für ordinalskalierte Daten gebildet werden.

☐ richtig ☒ **falsch**

2) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☒ **falsch**

3) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☒ **richtig** ☐ falsch

4) Eine linkssteile Verteilung ist immer auch rechtsschief, eine rechtssteile Verteilung ist immer auch linksschief.

☒ **richtig** ☐ falsch

5) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen 0 und 1 annehmen.

☐ richtig ☒ **falsch**

6) Nominalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.

☐ richtig ☒ **falsch**

7) Der Interquartilsabstand (IQR) ist ein robusteres Streuungsmaß als die Spannweite.

☒ **richtig** ☐ falsch

8) Die Korrelationskoeffizienten nach Spearman und Kendall stellen Maße für die monotone Korrelation dar.

☒ **richtig** ☐ falsch

9) Aus Stem-and-Leaf-Diagrammen kann man die Ursprungsdaten direkt ablesen.

☒ **richtig** ☐ falsch

10) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☒ **richtig** ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Im Rahmen einer archäologischen Ausgrabung werden 15 Tonscherben von Vasen aus der spätminoischen Phase (1.600 bis 1.450 v. Christus) gesichert, untersucht und katalogisiert.

Es ergibt sich folgende Übersicht:

Objektnummer	Höhe in mm	Breite in mm
A2312-001	53	76
A2312-002	55	23
A2312-003	23	78
A2312-004	76	45
A2312-005	25	37
A2312-006	34	82
A2312-007	55	21
A2312-008	28	75
A2312-009	69	63
A2312-010	43	53
A2312-011	40	57
A2312-012	16	43
A2312-013	23	28
A2312-014	128	91
A2312-015	51	60



Stellen Sie für die Verteilung der Angaben zur Höhe der Bruchstücke ein Stamm-Blatt-Diagramm mit einer aus Ihrer Sicht geeigneten Stammbreite auf. (7 Punkte) Zeichnen Sie zusätzlich noch das Stamm-Blatt-Diagramm mit einer Stammbreite von 50. (3 Punkte)

16
23; 23; 25; 28
34
40; 43
51; 53; 55; 55
69
76
128

1 6	0 1 2 2 2 3 4 4
2 3 3 5 8	0 5 5 5 5 6 7
3 4	1 2
4 0 3	
5 1 3 5 5	
6 9	
7 6	

Stammbreite 50
Jedes Blatt 1 Fall
Kein Ausreißer

Stammbreite 10
Jedes Blatt 1 Fall
1 Ausreißer (128)

Aufgabenteil III: Verteilungsparameter (15 Punkte)

Wir werfen erneut einen Blick auf die uns bereits bekannten Objektdaten.

Objektnummer	Höhe in mm	Breite in mm
A2312-001	53	76
A2312-002	55	23
A2312-003	23	78
A2312-004	76	45
A2312-005	25	37
A2312-006	34	82
A2312-007	55	21
A2312-008	28	75
A2312-009	69	63
A2312-010	43	53
A2312-011	40	57
A2312-012	16	43
A2312-013	23	28
A2312-014	128	91
A2312-015	51	60

Berechnen Sie für die Angaben zur Breite der Bruchstücke folgende Kennzahlen:

- g) Arithmetisches Mittel (3 Punkte)
- h) Um 10% getrimmtes arithmetisches Mittel (2 Punkte)
- i) Median (3 Punkte)
- j) Interquartilsabstand (2 Punkte)
- k) Varianz (3 Punkte)
- l) Standardabweichung (2 Punkte)

Arithmetisches Mittel

$$76 + 23 + 78 + 45 + 37 + 82 + 21 + 75 + 63 + 53 + 57 + 43 + 28 + 91 + 60 = 832$$

$$832 / 15 = \underline{55,47 \text{ mm}}$$

Um 10% getrimmtes arithmetisches Mittel

$15 * 0,1 = 1,5 \rightarrow$ Streichung der beiden größten und kleinsten Werte, Division durch 11

$$28 + 37 + 43 + 45 + 53 + 57 + 60 + 63 + 75 + 76 + 78 = 615$$

$$615 / 11 = \underline{55,91 \text{ mm}}$$

Median

Geordnete Reihe:

21; 23; 28; 37; 43; 45; 53; 57; 60; 63; 75; 76; 78; 82; 91

$0,5 * 15 = 7,5 \rightarrow$ Wert an der 8. Stelle der geordneten Reihe: 57 mm

Interquartilsabstand

Geordnete Reihe:

21; 23; 28; 37; 43; 45; 53; 57; 60; 63; 75; 76; 78; 82; 91

$0,25 \cdot 15 = 3,75 \rightarrow$ Wert an der 4. Stelle der geordneten Reihe: 37

$0,75 \cdot 15 = 11,25 \rightarrow$ Wert an der 12. Stelle der geordneten Reihe: 76

$$76 - 37 = \underline{39 \text{ mm}}$$

Varianz

Arithmetisches Mittel: 55,47

$$(21 - 55,47)^2 = 1188,18$$

$$(23 - 55,47)^2 = 1054,30$$

$$(28 - 55,47)^2 = 754,60$$

$$(37 - 55,47)^2 = 341,14$$

$$(43 - 55,47)^2 = 155,50$$

$$(45 - 55,47)^2 = 109,62$$

$$(53 - 55,47)^2 = 6,10$$

$$(57 - 55,47)^2 = 2,34$$

$$(60 - 55,47)^2 = 20,52$$

$$(63 - 55,47)^2 = 56,70$$

$$(75 - 55,47)^2 = 381,42$$

$$(76 - 55,47)^2 = 421,48$$

$$(78 - 55,47)^2 = 507,60$$

$$(82 - 55,47)^2 = 703,84$$

$$(91 - 55,47)^2 = 1262,38$$

Summe: 6965,73

$$6965,73 / 15 = \underline{464,38}$$

Standardabweichung

$$\sqrt{464,38} = \underline{21,55}$$

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Ein Museum senkt (ceteris paribus) über einen Zeitraum von 10 Tagen die Preise für den Eintritt in die Dauerausstellung um jeweils 50 Cent pro Tag und erfasst die Anzahl der täglichen Besucher*innen. Es ergeben sich folgende Werte.

Tag	Eintritt in EUR	Anzahl Besucher*innen
03.07.2023	7,50	81
04.07.2023	7,00	92
05.07.2023	6,50	94
06.07.2023	6,00	108
07.07.2023	5,50	107
08.07.2023	5,00	123
10.07.2023	4,50	154
11.07.2023	4,00	211
12.07.2023	3,50	227
13.07.2023	3,00	245

Berechnen und interpretieren Sie sowohl der Rangkorrelationskoeffizienten nach Spearman (7 Punkte) als auch den Bravais-Pearson-Korrelationskoeffizienten (8 Punkte) für die beiden hier aufgeführten Variablen Eintrittspreis und Besuchszahlen.

Rangkorrelationskoeffizient nach Spearman

Tag	Eintritt in EUR	rg(x)	Anzahl Besucher*innen	rg(y)	d	d ²
03.07.2023	7,50	10	81	1	9	81
04.07.2023	7,00	9	92	2	7	49
05.07.2023	6,50	8	94	3	5	25
06.07.2023	6,00	7	108	5	2	4
07.07.2023	5,50	6	107	4	2	4
08.07.2023	5,00	5	123	6	-1	1
10.07.2023	4,50	4	154	7	-3	9
11.07.2023	4,00	3	211	8	-5	25
12.07.2023	3,50	2	227	9	-7	49
13.07.2023	3,00	1	245	10	-9	81

Summe aller quadrierten Rangplatzdifferenzen: 328

$$1 - \frac{6 \sum d_i^2}{(n^2 - 1)n} = 1 - \frac{1968}{990} = 1 - 1,9879 = -0,9879$$

Der Spearman-Koeffizient liegt bei -0,9879.

Dies zeigt eine starke, negative, monotone Korrelation an.

Bravais-Pearson-Koeffizient

n	x	y	x ²	y ²	(x*y)
1	7,50	81	56,25	6561	607,50
2	7,00	92	49,00	8464	644,00
3	6,50	94	42,25	8836	611,00
4	6,00	108	36,00	11664	648,00
5	5,50	107	30,25	11449	588,50
6	5,00	123	25,00	15129	615,00
7	4,50	154	20,25	23716	693,00
8	4,00	211	16,00	44521	844,00
9	3,50	227	12,25	51529	794,50
10	3,00	245	9,00	60025	735,00
SUMME	52,5	1442	296,25	241894	6780,50
MITTEL	5,25	144,2	-//-	-//-	-//-

$$n = 10$$

$$\bar{x} = 5,25$$

$$\bar{y} = 144,2$$

$$\sum x^2 = 296,25$$

$$\sum y^2 = 241894$$

$$\sum (x \cdot y) = 6780,50$$

$$r = \frac{-790}{4,54 * 184,27} = -0,94$$

Der Bravais-Pearson-Korrelationskoeffizient liegt bei -0,94.

Dies zeigt eine starke, negative, lineare Korrelation an.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleich bleibenden Rahmenbedingungen.

X richtig ☐ falsch

2) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

X richtig ☐ falsch

3) Das logische UND wird auch als Konjunktion bezeichnet.

X richtig ☐ falsch

4) Eine gültige Teilmenge muss stets aus mindestens zwei Elementen bestehen.

☐ richtig **X falsch**

5) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

☐ richtig **X falsch**

6) Ein Würfel wird dann als fair bezeichnet, wenn die Wahrscheinlichkeit des Wurfs für jede Augenzahl genau gleich groß ist.

X richtig ☐ falsch

7) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig **X falsch**

8) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht der Wurzel aus dem Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig **X falsch**

9) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

X richtig ☐ falsch

10) Das konstante Glied (a) der einfachen linearen Regressionsanalyse entspricht dem Wert der Steigung der linearen Regressionsgeraden.

☐ richtig **X falsch**

Aufgabenteil VI: Lineare Regression (15 Punkte)

Wir werfen erneut einen Blick auf die uns bereits bekannten Daten zu den Museumsbesuchen.

Tag	Eintritt in EUR	Anzahl Besucher*innen
03.07.2023	7,50	81
04.07.2023	7,00	92
05.07.2023	6,50	94
06.07.2023	6,00	108
07.07.2023	5,50	107
08.07.2023	5,00	123
10.07.2023	4,50	154
11.07.2023	4,00	211
12.07.2023	3,50	227
13.07.2023	3,00	245

Stellen Sie ein lineares Regressionsmodell mit dem Eintrittspreis als unabhängiger Variable (x) und der Anzahl Besucher*innen als abhängiger Variable (y) auf. (7 Punkte) Berechnen und interpretieren Sie außerdem das Gütemaß R^2 für das von Ihnen aufgestellte Modell. (5 Punkte)

Hätte man das Modell mathematisch wie auch inhaltlich betrachtet auch umgekehrt mit der Anzahl Besucher*innen als unabhängiger Variable (x) und den Eintrittspreisen als abhängiger Variable (y) aufstellen können? (3 Punkte)

Aufstellung des Regressionsmodells

Wir können hier auf die Tabelle aus der Aufgabe zum B-P-K zurückgreifen:

n	x	y	x^2	y^2	$(x \cdot y)$
1	7,50	81	56,25	6561	607,50
2	7,00	92	49,00	8464	644,00
3	6,50	94	42,25	8836	611,00
4	6,00	108	36,00	11664	648,00
5	5,50	107	30,25	11449	588,50
6	5,00	123	25,00	15129	615,00
7	4,50	154	20,25	23716	693,00
8	4,00	211	16,00	44521	844,00
9	3,50	227	12,25	51529	794,50
10	3,00	245	9,00	60025	735,00
SUMME	52,5	1442	296,25	241894	6780,50
MITTEL	5,25	144,2	-/-	-/-	-/-

$$b = \frac{-790}{20,625} = -38,30$$

$$a = 144,2 + 38,30 \cdot 5,25 = 345,275$$

Das Regressionsmodell lautet demnach: $y = 345,28 - 38,3 \cdot x$

Bewertung der Modellgüte

Das Gütemaß R^2 kann aus dem bereits berechneten B-P-K gewonnen werden: $-0,94^2 = 0,8836$

Das Modell leistet eine sehr gute Streuungsaufklärung von 88,36%.

Frage zum „umgekehrten“ Modell

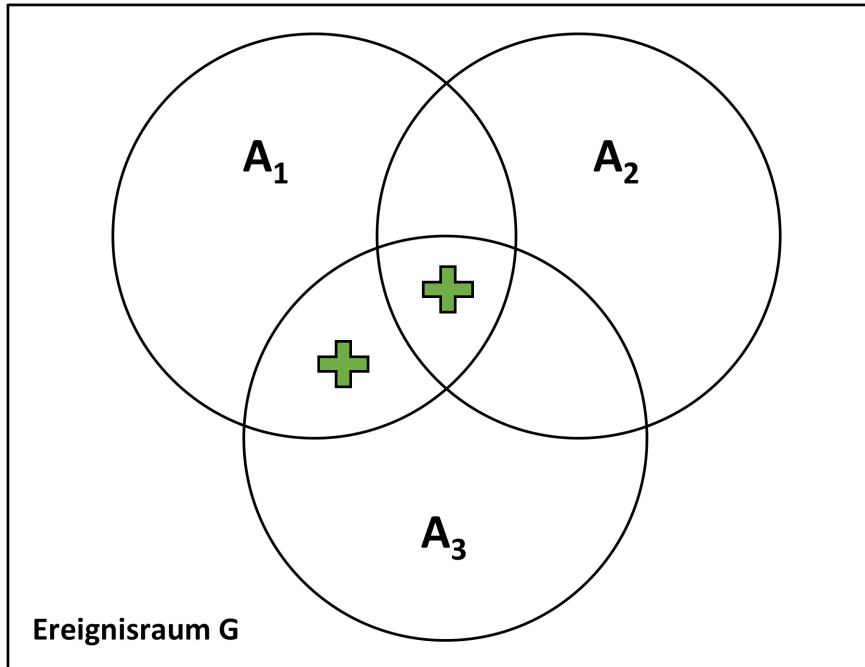
Rein mathematisch betrachtet wäre die Aufstellung dieses Modells möglich.

Unter inhaltlichen Gesichtspunkten ergäbe es jedoch keinen Sinn, da die Zahl der Besucher*innen kaum rückwirkend den Eintrittspreis für den vergangenen Tag beeinflussen könnte.

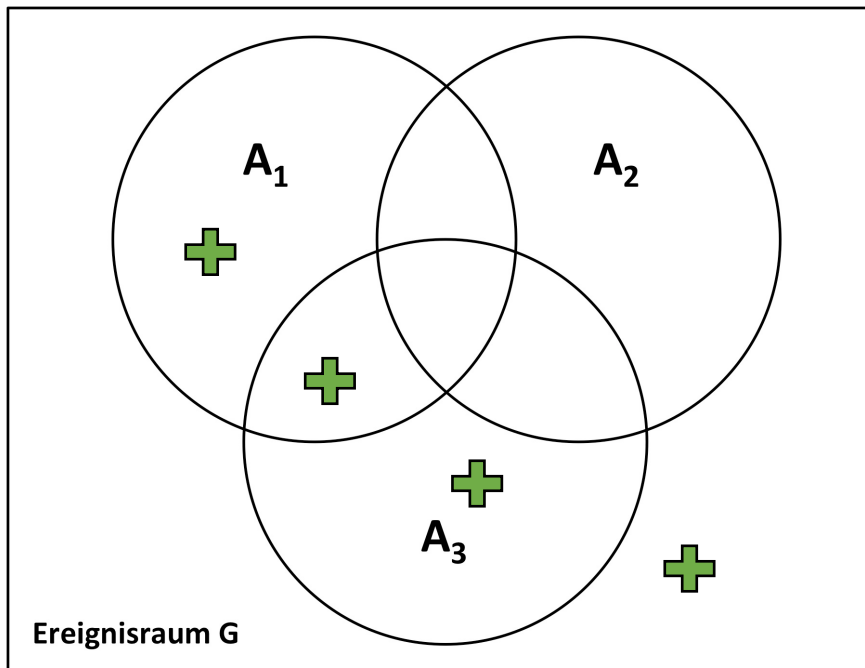
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_1 \cap A_2) \cap A_3$ (3 Punkte)



b) $\overline{A_2}$ (2 Punkte)



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

Ein Einbrecher versucht, eine diamantbesetzte Krone aus dem Fundus eines Museums zu entwenden. Damit ihm dies gelingt, muss er zwei Sicherheitssysteme nacheinander überwinden: Einen Fingerabdruck-Sensor an der Eingangstür des Lagerraums sowie eine Lichtschranke in der Mitte des Raumes. Dies gelingt ihm bei dem Fingerabdruck-Sensor mit einer Wahrscheinlichkeit von 65% sowie bei der Lichtschranke mit einer Wahrscheinlichkeit von 80%, wenn er über seine volle Konzentrationsfähigkeit verfügt. Streitet er sich jedoch am Abend vor dem Einbruch mit seiner Lebensgefährtin, sinkt die Wahrscheinlichkeit auf 55% beim Fingerabdruck-Sensor sowie auf 70% bei der Lichtschranke. Die Wahrscheinlichkeit auf einen Streit liegt an jedem beliebigen Abend bei 50%. Mit welcher Wahrscheinlichkeit gelingt der Diebstahl der Krone?



Für die Lösung sind zwei Pfade im Pfaddiagramm von Bedeutung:

Ehestreit -> Fingerabdruck-Sensor wird überwunden -> Lichtschranke wird überwunden

Kein Ehestreit -> Fingerabdruck-Sensor wird überwunden -> Lichtschranke wird überwunden

Auf allen anderen Pfaden wird mindestens eines der beiden Sicherheitssysteme ausgelöst.

Die Eintrittswahrscheinlichkeiten beider Pfade lauten wie folgt:

$$0,50 * 0,65 * 0,80 = 0,26$$

$$0,50 * 0,55 * 0,70 = 0,1925$$

$$0,26 + 0,1925 = 0,4525$$

Der Diebstahl gelingt mit einer Wahrscheinlichkeit von 45,25%.

Aufgabenteil IX: Kombinatorik (5 Punkte)

Eine Besuchergruppe bestehend aus 12 Personen besucht die Gemäldeausstellung eines Museums. Vor einem besonders beeindruckenden Gemälde möchten alle 12 auf einer Bank Platz nehmen, um sich in Ruhe den Audioguide anzuhören. Leider bietet die Bank nur 4 Sitzplätze.

- c) In wie vielen möglichen Zusammenstellungen können 4 der 12 Personen auf der Bank Platz nehmen? (3 Punkte)
- d) Angenommen, die Gruppe bestünde aus 6 Frauen und 6 Männern. Wie viele Zusammenstellungen von 4 Personen gibt es, denen genau 2 Männer und 2 Frauen angehören? (2 Punkte)

4 aus 12

Es handelt sich um eine Kombination (die Reihenfolge, in der die Personen auf der Bank sitzen, spielt keine Rolle, lediglich die Zusammenstellung der Gruppe) im Modell ohne Zurücklegen (Personen können nicht mehrfach gezogen werden).

$$\frac{n!}{k!(n-k)!} = \frac{12!}{4!8!} = \frac{12 * 11 * 10 * 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1}{4 * 3 * 2 * 1 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1} = \frac{12 * 11 * 10 * 9}{4 * 3 * 2 * 1} = \frac{11880}{24}$$

Es gibt 495 mögliche Zusammenstellungen.

2 x 2 aus 6

Auch hier handelt es sich um eine Kombination im Modell ohne Zurücklegen, allerdings werden zunächst 2 von 6 Frauen und dann 2 von 6 Männern gezogen und die Ergebnisse multipliziert, um die Gesamtzahl aller möglichen Zusammenstellungen zu erhalten.

$$\frac{n!}{k!(n-k)!} = \frac{6!}{2!4!} = \frac{6 * 5 * 4 * 3 * 2 * 1}{2 * 1 * 4 * 3 * 2 * 1} = \frac{6 * 5}{2 * 1} = 15$$

$$15 * 2 = 30$$

Es gibt 30 mögliche Zusammenstellungen.

Aufgabenteil X: Satz von Bayes (10 Punkte)

Eine Archäologin begeht eine in 380 Flächensegmente unterteilte potentielle Fundstelle auf der Suche nach im Boden lagernden Metallfragmenten mit einer Bodensonde. Die Bodensonde kann solche Metallfragmente innerhalb eines Flächensegments mit einer sehr hohen Wahrscheinlichkeit von 97% detektieren. Allerdings meldet sie auch für 8% aller Flächensegmente ohne spätere Fundstücke das Vorhandensein von Metallteilen. Aus Erfahrungen mit ähnlichen Fundstellen lässt sich ableiten, dass in etwa 5% aller Flächensegmente echte Funde zu erwarten sind.

Die Bodensonde hat gerade angeschlagen.

Wie hoch ist die Wahrscheinlichkeit auf einen echten Treffer?

Von 380 Flächensegmenten...

... sind auf 19 (5%) Funde zu erwarten

...sind auf 361 (95%) keine Funde zu erwarten

Von 19 Flächensegmenten mit Funden...

...werden 18,43 (97%) korrekt identifiziert (true positives)

...werden 0,57 (3%) nicht korrekt identifiziert

Von 361 Flächensegmenten ohne Funde...

...wird für 332,12 (92%) korrekterweise kein Fund angezeigt

...wird für 28,88 (8%) fälschlicherweise ein Fund angezeigt (false positives)

Es kommt somit in $18,43 + 28,88 = 47,31$ Fällen zur Anzeige eines Treffers.

Davon sind 18,43 Trefferanzeigen berechtigt.

Die Trefferquote liegt somit bei $18,43/47,31 = 0,3896 \rightarrow \underline{38,96\%}$.

Sommersemester 2024

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2024

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 04. Mai 2024

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre
Sommersemester 2024
Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Stochastik und Wahrscheinlichkeitstheorie sind Teilgebiete der Statistik.

☐ richtig ☐ falsch

2) Die Grundgesamtheit kann beliebig viele Teilgesamtheiten umfassen.

☐ richtig ☐ falsch

3) Ist ein Merkmal häufbar, kann es für eine statistische Einheit mehrere Ausprägungen annehmen.

☐ richtig ☐ falsch

4) Bei einer willkürlichen Auswahl hat jede statistische Einheit in der Grundgesamtheit die exakt gleiche Chance, in die Stichprobe aufgenommen zu werden.

☐ richtig ☐ falsch

5) Die Berechnung des getrimmten arithmetischen Mittels setzt mindestens ordinalskalierte Werte voraus.

☐ richtig ☐ falsch

6) Die Spannweite ist definiert als der Abstand zwischen dem größten und dem kleinsten Wert.

☐ richtig ☐ falsch

7) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

☐ richtig ☐ falsch

8) Der Interquartilsabstand (IQR) ist robust, d.h. unempfindlich gegenüber Ausreißern.

☐ richtig ☐ falsch

9) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☐ falsch

10) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Im Wald bei Wernigerode finden archäologische Ausgrabungen auf dem Gelände des ehemaligen Augustiner-Eremitenklosters Himmelpforte (*Porta coeli*, gegründet 1253, aufgegeben um 1560) statt. Dabei werden u.a. Überreste von dreizehn mittelalterlichen Reitersporen gefunden und mit folgenden Gewichtsangaben verzeichnet.

Objektnummer	Gewicht
HP2024-0128	117 g
HP2024-0325	82 g
HP2024-0235	123 g
HP2024-0532	94 g
HP2024-0382	212 g
HP2024-0376	18 g
HP2024-0137	76 g
HP2024-0694	94 g
HP2024-1238	114 g
HP2024-0732	120 g
HP2024-0027	131 g
HP2024-0573	128 g
HP2024-0673	71 g

Erstellen Sie zwei Stamm-Blatt-Diagramme mit einer Stammbreite von 100 (10 Punkte) sowie einer Stammbreite von 10 (10 Punkte). Vergessen Sie dabei nicht die zugehörigen Angaben.



Reitersporn aus dem 13. Jahrhundert, Stadtmuseum Hagen / Heike Wipperm

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Neben dem Gewicht der dreizehn Reitersporen wurde auch deren Länge tabellarisch erfasst.

Objektnummer	Länge
HP2024-0128	13,5 cm
HP2024-0325	12,1 cm
HP2024-0235	12,3 cm
HP2024-0532	11,0 cm
HP2024-0382	10,4 cm
HP2024-0376	10,2 cm
HP2024-0137	13,1 cm
HP2024-0694	12,5 cm
HP2024-1238	8,6 cm
HP2024-0732	17,4 cm
HP2024-0027	13,9 cm
HP2024-0573	11,8 cm
HP2024-0673	10,9 cm

Bestimmen Sie für die Verteilung der Längenangaben folgende statistische Kennzahlen:

- a) Arithmetisches Mittel (5 Punkte)
- b) Median (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Interquartilsabstand (5 Punkte)
- f) Varianz (5 Punkte)
- g) Standardabweichung (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Das archäologische Team findet bei den Grabungen auch zahlreiche Bruchstücke von Dachziegeln, die aus einer Ziegelei aus dem 19. Jahrhundert zu stammen scheinen, welche nach dem Abriss des Klosters auf einem Teil des Geländes errichtet wurde. Es wird angenommen, dass diese Ziegelei sich auf der heutigen Grabungs-Parzelle A38 befunden hat. Bei Richtigkeit dieser Annahme wird vermutet, dass die Anzahl der Bruchstücke mit der Entfernung zur Parzelle korrelieren müsste – je dichter eine Parzelle am angenommenen Standort liegt, umso mehr Produktionsabfall im Sinne von Bruch müsste sich dort finden lassen. Als Maß für die Entfernung zwischen zwei Parzellen wurde die Strecke zwischen den jeweiligen Parzellenmitten genutzt.

Berechnen Sie für die Variablen „Entfernung zu Parzelle A38“ und „Aufgefundene Bruchstücke“

- den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) sowie
- den Konkordanzkoeffizienten nach Kendall (15 Punkte) und
- vergleichen und interpretieren Sie die Ergebnisse mit Blick auf die Ausgangshypothese (5 Punkte).

Parzelle	Entfernung zu Parzelle A38	Aufgefundene Bruchstücke
A17	123 cm	78
A21	179 cm	82
A23	237 cm	71
A04	258 cm	62
A11	333 cm	64
A18	375 cm	56
A25	476 cm	64
A31	584 cm	44
A20	634 cm	17
A19	864 cm	3

KLAUSUR STATISTIK I MIT LÖSUNGEN

Berufsbegleitender Studiengang Betriebswirtschaftslehre Sommersemester 2024
Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Stochastik und Wahrscheinlichkeitstheorie sind Teilgebiete der Statistik.

☐ richtig ☒ falsch

2) Die Grundgesamtheit kann beliebig viele Teilgesamtheiten umfassen.

☒ richtig ☐ falsch

3) Ist ein Merkmal häufbar, kann es für eine statistische Einheit mehrere Ausprägungen annehmen.

☒ richtig ☐ falsch

4) Bei einer willkürlichen Auswahl hat jede statistische Einheit in der Grundgesamtheit die exakt gleiche Chance, in die Stichprobe aufgenommen zu werden.

☐ richtig ☒ falsch

5) Die Berechnung des getrimmten arithmetischen Mittels setzt mindestens ordinalskalierte Werte voraus.

☐ richtig ☒ falsch

6) Die Spannweite ist definiert als der Abstand zwischen dem größten und dem kleinsten Wert.

☒ richtig ☐ falsch

7) Erhebt man Daten in drei Phasen, ist von einer tertiärstatistischen Erhebung die Rede.

☐ richtig ☒ falsch

8) Der Interquartilsabstand (IQR) ist robust, d.h. unempfindlich gegenüber Ausreißern.

☒ richtig ☐ falsch

9) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☒ falsch

10) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☒ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Im Wald bei Wernigerode finden archäologische Ausgrabungen auf dem Gelände des ehemaligen Augustiner-Eremitenklosters Himmelpforte (*Porta coeli*, gegründet 1253, aufgegeben um 1560) statt. Dabei werden u.a. Überreste von dreizehn mittelalterlichen Reitersporen gefunden und mit folgenden Gewichtsangaben verzeichnet.

Objektnummer	Gewicht
HP2024-0128	117 g
HP2024-0325	82 g
HP2024-0235	123 g
HP2024-0532	94 g
HP2024-0382	212 g
HP2024-0376	18 g
HP2024-0137	76 g
HP2024-0694	94 g
HP2024-1238	114 g
HP2024-0732	120 g
HP2024-0027	131 g
HP2024-0573	128 g
HP2024-0673	71 g

Erstellen Sie zwei Stamm-Blatt-Diagramme mit einer Stammbreite von 100 (10 Punkte) sowie einer Stammbreite von 10 (10 Punkte). Vergessen Sie dabei nicht die die zugehörigen Angaben.



Reitersporn aus dem 13. Jahrhundert, Stadtmuseum Hagen / Heike Wippermann

Bildung einer geordneten Reihe:

18; 71; 76; 82; 94; 94; 114; 117; 120; 123; 128; 131; 212

Stamm-Blatt-Diagramm mit Stammbreite 100:

0 | 1 7 7 8 9 9

1 | 1 1 2 2 2 3

2 | 1

Stammbreite 100

Jedes Blatt 1 Fall

Keine Ausreißer

Stamm-Blatt-Diagramm mit Stammbreite 10:

7 | 1 6

8 | 2

9 | 4 4

10 |

11 | 4 7

12 | 0 3 8

13 | 1

Stammbreite 10

Jedes Blatt 1 Fall

2 Ausreißer (18, 212)

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Neben dem Gewicht der dreizehn Reitersporen wurde auch deren Länge tabellarisch erfasst.

Objektnummer	Länge
HP2024-0128	13,5 cm
HP2024-0325	12,1 cm
HP2024-0235	12,3 cm
HP2024-0532	11,0 cm
HP2024-0382	10,4 cm
HP2024-0376	10,2 cm
HP2024-0137	13,1 cm
HP2024-0694	12,5 cm
HP2024-1238	8,6 cm
HP2024-0732	17,4 cm
HP2024-0027	13,9 cm
HP2024-0573	11,8 cm
HP2024-0673	10,9 cm

Bestimmen Sie für die Verteilung der Längenangaben folgende statistische Kennzahlen:

- a) Arithmetisches Mittel (5 Punkte)
- b) Median (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Interquartilsabstand (5 Punkte)
- f) Varianz (5 Punkte)
- g) Standardabweichung (5 Punkte)

Grundsätzliche Empfehlung; Umwandlung der cm in mm -> Ergebnis k

- a) Arithmetisches Mittel

$$135 + 121 + 123 + 110 + 104 + 102 + 131 + 125 + 86 + 174 + 139 + 118 + 109 = 1577$$

$$1577 / 13 = \underline{121,31 \text{ mm} \rightarrow 12,13 \text{ cm}}$$

- b) Median

Bildung einer geordneten Reihe:

86; 102; 104; 109; 110; 118; 121; 123; 125; 131; 135; 139; 174

Berechnung des Medians:

$$0,5 * 13 = 6,5 \rightarrow \text{Wert an der 7. Stelle der geordneten Verteilung}$$

$$x_{\text{med}} = \underline{121 \text{ mm} \rightarrow 12,10 \text{ cm}}$$

- c) Modus

Da keine unimodale Verteilung vorliegt, ist kein Modus bestimmbar.

- d) Spannweite

$$d_s = 174 - 86 = \underline{88 \text{ mm} \rightarrow 8,80 \text{ cm}}$$

e) Interquartilsabstand

Unterer Quartilswert: $0,25 * 13 = 3,25 \rightarrow$ Wert an der 4. Stelle der geordneten Verteilung $X_{0,25} = 109$

Oberer Quartilswert: $0,75 * 13 = 9,75 \rightarrow$ Wert an der 10. Stelle der geordneten Verteilung $X_{0,75} = 131$

IQR = $131 - 109 = \underline{22 \text{ mm} \rightarrow 2,20 \text{ cm}}$

f) Varianz

Arithmetisches Mittel: 121,31

$$(86-121,31)^2 = 1246,80$$

$$(102-121,31)^2 = 372,88$$

$$(104-121,31)^2 = 299,64$$

$$(109-121,31)^2 = 151,54$$

$$(110-121,31)^2 = 127,92$$

$$(118-121,31)^2 = 10,96$$

$$(121-121,31)^2 = 0,10$$

$$(123-121,31)^2 = 2,86$$

$$(125-121,31)^2 = 13,62$$

$$(131-121,31)^2 = 93,90$$

$$(135-121,31)^2 = 187,42$$

$$(139-121,31)^2 = 312,94$$

$$(174-121,31)^2 = 2776,24$$

Summe aller quadrierten Abweichungen:

$$5596,82 \quad 5596,82 / 13 = \underline{430,52 \text{ mm}^2}$$

g) Standardabweichung

Positive Wurzel aus der Varianz: $\underline{20,75 \text{ mm} \rightarrow 2,08 \text{ cm}}$

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Das archäologische Team findet bei den Grabungen auch zahlreiche Bruchstücke von Dachziegeln, die aus einer Ziegelei aus dem 19. Jahrhundert zu stammen scheinen, welche nach dem Abriss des Klosters auf einem Teil des Geländes errichtet wurde. Es wird angenommen, dass diese Ziegelei sich auf der heutigen Grabungs-Parzelle 38 befunden hat. Bei Richtigkeit dieser Annahme wird vermutet, dass die Anzahl der Bruchstücke mit der Entfernung zur Parzelle korrelieren müsste – je dichter eine Parzelle am angenommenen Standort liegt, umso mehr Produktionsabfall im Sinne von Bruch müsste sich dort finden lassen. Als Maß für die Entfernung zwischen zwei Parzellen wurde die Strecke zwischen den jeweiligen Parzellenmitten genutzt.

Berechnen Sie für die Variablen „Entfernung zu Parzelle A38“ und „Aufgefundene Bruchstücke“

- den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) sowie
- den Konkordanzkoeffizienten nach Kendall (15 Punkte) und
- vergleichen und interpretieren Sie die Ergebnisse mit Blick auf die Ausgangshypothese (5 Punkte).

Parzelle	Entfernung zu Parzelle A38	Aufgefundene Bruchstücke
A17	123 cm	78
A21	179 cm	82
A23	237 cm	71
A04	258 cm	62
A11	333 cm	64
A18	375 cm	56
A25	476 cm	64
A31	584 cm	44
A20	634 cm	17
A19	864 cm	3

a) Rangkorrelationskoeffizient nach Spearman

Parzelle	x	rg(x)	y	rg(y)	d	d ²
A17	123 cm	1	78	9	-8	64
A21	179 cm	2	82	10	-8	64
A23	237 cm	3	71	8	-5	25
A04	258 cm	4	62	5	-1	1
A11	333 cm	5	64	6,5	-1,5	2,25
A18	375 cm	6	56	4	2	4
A25	476 cm	7	64	6,5	0,5	0,25
A31	584 cm	8	44	3	5	25
A20	634 cm	9	17	2	7	49
A19	864 cm	10	3	1	9	81

Summe der quadrierten Rangplatzdifferenzen: 315,5

$$\rho = 1 - (6 \cdot 315,5) / ((10^2 - 1) \cdot 10) = 1 - 1893 / 990 = 1 - 1,91 = \underline{-0,91}$$

b) Konkordanzkoeffizient nach Kendall

Parzelle	x	rg(x)	y	rg(y)	K	D
A17	123 cm	1	78	9	1	8
A21	179 cm	2	82	10	0	8
A23	237 cm	3	71	8	0	7
A04	258 cm	4	62	5	2	4
A11	333 cm	5	64	6,5	0	4
A18	375 cm	6	56	4	1	3
A25	476 cm	7	64	6,5	0	3
A31	584 cm	8	44	3	0	2
A20	634 cm	9	17	2	0	1
A19	864 cm	10	3	1	0	0

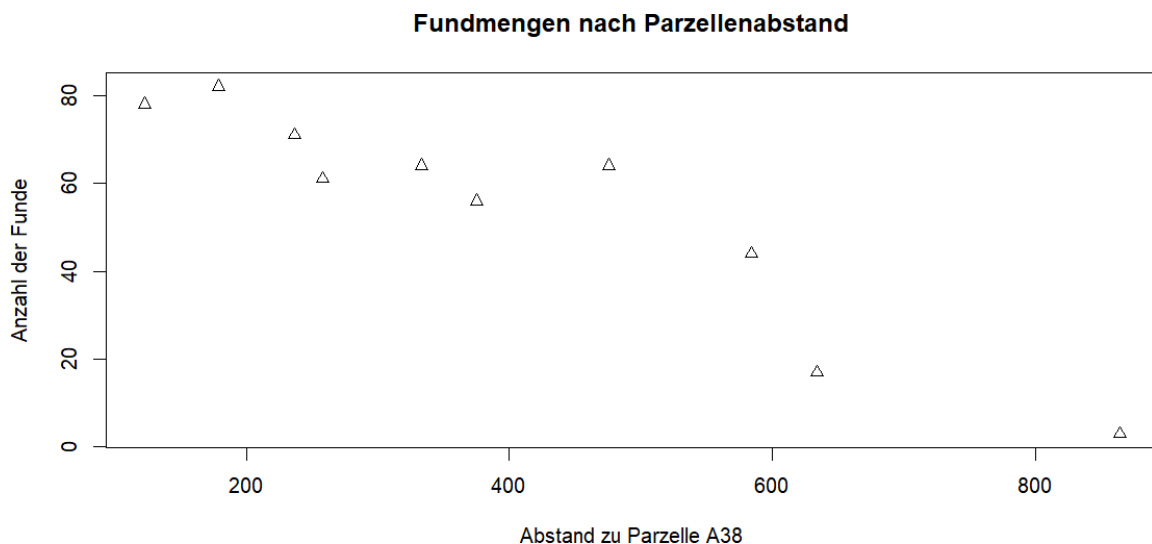
Summe der konkordanten Paare (K): 4

Summe der diskordanten Paare (D): 40

$$\tau = (2 * (4-40)) / (10 * (10-1)) = -72 / 90 = \underline{-0,8}$$

c) Interpretation der Ergebnisse

Beide Koeffizienten zeigen übereinstimmend eine starke, negative und monotone Korrelation. Dies kann als Indiz für die Richtigkeit der Arbeitshypothese gewertet werden, die eine negative Korrelation (große Entfernungen zum vermuteten Standort gehen mit geringen Fundmengen, hohe Fundmengen mit geringen Entfernungen zum vermuteten Standort der Ziegelei) erwarten lassen würde.



Sommersemester 2024

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2024

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 22. Juni 2024

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars. Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre
Sommersemester 2024
Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Rangplatzbindungen erhöhen die Aussagekraft des Spearman-Rangkorrelationskoeffizienten.

☐ richtig ☐ falsch

2) Die Spannweite wird nicht von Ausreißern beeinflusst und ist daher äußerst robust.

☐ richtig ☐ falsch

3) Die Größe der Box im Box-Plot wird durch das obere sowie das untere Quartil definiert.

☐ richtig ☐ falsch

4) Eine Grundgesamtheit kann in beliebig viele Teilgesamtheiten unterteilt werden.

☐ richtig ☐ falsch

5) Im Gegensatz zum arithmetischen Mittel ist der Median gegenüber Ausreißern robust.

☐ richtig ☐ falsch

6) Der Modus lässt sich nur bestimmen, wenn eine bimodale Verteilung vorliegt.

☐ richtig ☐ falsch

7) 2D-Streudiagramme stellen die gemeinsame Verteilung der Werte zweier Variablen dar.

☐ richtig ☐ falsch

8) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig ☐ falsch

9) Der statistische Ersatz fehlender Werte durch Mittelwerte oder lineare Trendwerte kann zu Verzerrungen bei der Varianz und bei Zusammenhangsmaßen führen.

☐ richtig ☐ falsch

10) Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen 0 und 1 annehmen.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

Für eine statistische Erhebung wird eine zufällige Auswahl von zehn Prozessen aus allen im vergangenen Monat vor einem Arbeitsgericht geführten Prozessen vorgenommen, die mit einem monetären Vergleich zu Gunsten des Beklagten abgeschlossen wurden. Die nachfolgende Tabelle enthält die Prozessaktennummern sowie die jeweils zugehörigen Vergleichssummen.

Prozessakte	Vergleichssumme
AQ-23-2314	5.320 EUR
AQ-23-3353	3.000 EUR
AQ-23-4424	2.300 EUR
AQ-24-1434	4.630 EUR
AQ-24-2243	23.570 EUR
AQ-25-4132	3.600 EUR
AQ-25-5324	4.500 EUR
AQ-25-6214	6.400 EUR
AQ-25-6455	5.890 EUR
AQ-25-7343	2.900 EUR

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Vergleichssummen und benennen Sie die hierfür erforderlichen Größen.

Aufgabenteil III: Verteilungsparameter (10 Punkte)

Prozessakte	Vergleichssumme
AQ-23-2314	5.320 EUR
AQ-23-3353	3.000 EUR
AQ-23-4424	2.300 EUR
AQ-24-1434	4.630 EUR
AQ-24-2243	23.570 EUR
AQ-25-4132	3.600 EUR
AQ-25-5324	4.500 EUR
AQ-25-6214	6.400 EUR
AQ-25-6455	5.890 EUR
AQ-25-7343	2.900 EUR

Bestimmen Sie für die bereits bekannte Verteilung der Vergleichssummen

- a) das arithmetische Mittel (2 Punkte),
- b) das um 10% getrimmte arithmetische Mittel (2 Punkte),
- c) den 35%-Perzentilwert (2 Punkte),
- d) den Modus (2 Punkte) und
- e) die Spannweite (2 Punkte).

Aufgabenteil IV: Korrelationskoeffizienten (20 Punkte)

Eine bekannte wirtschaftspsychologische Studie, die im Jahr 2011 von der Ben-Gurion-Universität und der Columbia University durchgeführt wurde, konnte anhand von mehr als tausende Urteilen belegen, dass Richter*innen zu strengeren Urteilen tendieren, je hungriger sie sind (ja, wirklich).

Diese Studie soll an der HS Harz mit einem kleinen Sample von zehn Fällen aus dem lokalen Verkehrsgericht nachvollzogen werden. Erfasst wurde jeweils die Zeit zwischen der letzten Pause des Richters zum Zeitpunkt des Verhandlungsbeginns und die Höhe des verhängten Bußgeldes für die verhandelten, nahezu identischen Verkehrsdelikte.

Fallnummer	Minuten seit letzter Pause	Höhe des Bußgeldes
B-17238	20 Min.	60 EUR
B-22163	28 Min.	80 EUR
B-23231	37 Min.	75 EUR
B-45437	49 Min.	70 EUR
B-63423	62 Min.	75 EUR
B-67612	78 Min.	90 EUR
B-75323	96 Min.	130 EUR
B-43634	121 Min.	125 EUR
B-45256	138 Min.	130 EUR
B-64654	140 Min.	155 EUR

Berechnen und interpretieren Sie den Konkordanzkoeffizienten nach Kendall (10 Punkte) sowie den Bravais-Pearson-Korrelationskoeffizienten (10 Punkte).

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Leverage-Effekt bezeichnet.

☐ richtig ☐ falsch

2) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleichbleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

3) Die Wahrscheinlichkeit für das Eintreten eines Zufallsereignisses darf minimal Null betragen, kann aber nie negativ werden.

☐ richtig ☐ falsch

4) Das logische UND wird auch als Konjunktion bezeichnet.

☐ richtig ☐ falsch

5) Eine zufällig gezogene Stichprobe mit hoher Rücklaufquote ist unabhängig von ihrem Umfang stets repräsentativ.

☐ richtig ☐ falsch

6) Das sogenannte Taxi-Beispiel ist eine bekannte Beispielaufgabe für die Anwendung des Satz von Bayes.

☐ richtig ☐ falsch

7) Je breiter ein Konfidenzintervall ist, umso größer ist die Wahrscheinlichkeit, dass der gesuchte Wert exakt einer der beiden Intervallgrenzen entspricht.

☐ richtig ☐ falsch

8) Die Teilmenge einer Menge kann nie so viele Elemente enthalten wie die Menge, zu der sie gehört.

☐ richtig ☐ falsch

9) Das Bestimmtheitsmaß der einfachen linearen Regressionsanalyse entspricht dem quadrierten Wert des Bravais-Pearson-Korrelationskoeffizienten.

☐ richtig ☐ falsch

10) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

Aufgabenteil VI: Lineare Regression (15 Punkte)

Die Untersuchung der aus Aufgabenteil IV bereits bekannten Daten soll mit der Aufstellung eines linearen Regressionsmodells fortgeführt werden. In diesem Modell soll die Variable „Minuten seit letzter Pause“ als unabhängige Variable x , die Variable „Höhe des Bußgeldes“ dagegen als abhängige Variable y aufgenommen werden.

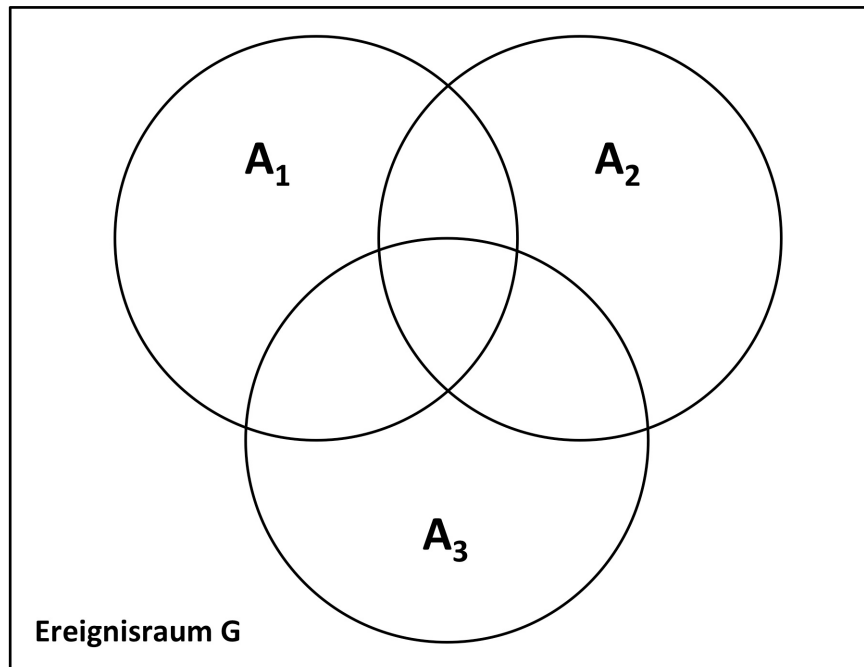
Fallnummer	Minuten seit letzter Pause	Höhe des Bußgeldes
B-17238	20 Min.	60 EUR
B-22163	28 Min.	80 EUR
B-23231	37 Min.	75 EUR
B-45437	49 Min.	70 EUR
B-63423	62 Min.	75 EUR
B-67612	78 Min.	90 EUR
B-75323	96 Min.	130 EUR
B-43634	121 Min.	125 EUR
B-45256	138 Min.	130 EUR
B-64654	140 Min.	155 EUR

Stellen Sie das gewünschte Regressionsmodell auf (10 Punkte) und interpretieren Sie das Ergebnis unter Berücksichtigung des Bestimmtheitsmaßes R^2 (5 Punkte).

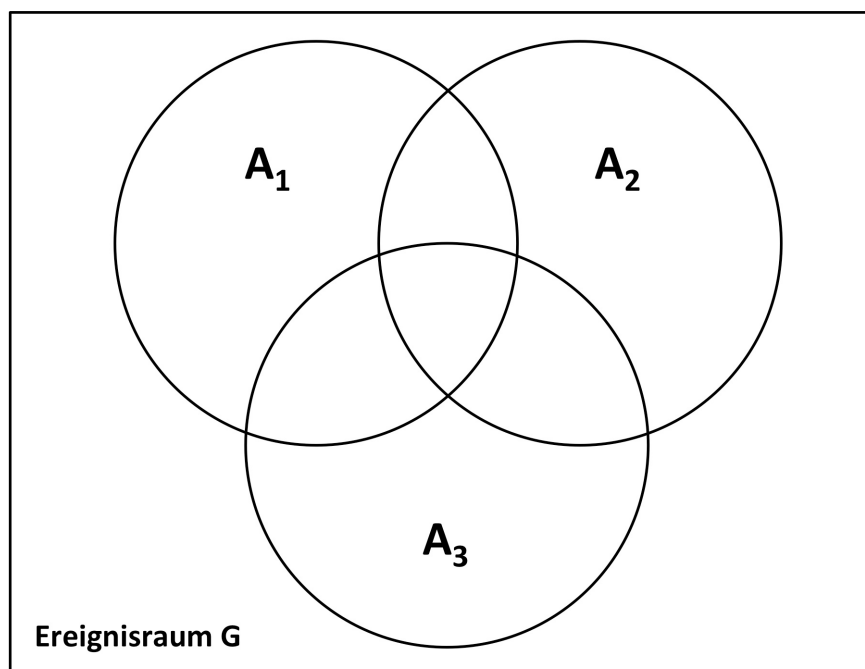
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_1 \cup A_2) \cap A_3$ (3 Punkte)



b) $(A_1 \cup \overline{A_1})$ (2 Punkte)



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

In einer Steuerbehörde findet eine Prüfung der von einem Bürger eingereichten Steuerunterlagen statt. An diesem Prüfungsvorgang sind zwei Steuerprüfer beteiligt, wobei die Unterlagen zunächst von einem der beiden Steuerprüfer gesichtet und beurteilt werden. Dieser teilt das Prüfergebnis anschließend dem anderen Steuerprüfer mit, der daraufhin eine zweite Prüfung vornimmt, um die Korrektheit des Ergebnisses abzusichern. Im vorliegenden Fall enthalten die eingereichten Unterlagen unstrittig einen Hinweis auf eine vertuschte Steuerhinterziehung, der allerdings von den Steuerprüfern nicht unbedingt entdeckt wird, da beide nur mit einer Trefferquote von 0,4 arbeiten.

Dieses an sich einfache Beispiel wird nun leider dadurch verkompliziert, dass sich die beiden Prüfer charakterlich voneinander unterscheiden. Prüfer A ist eher behäbig. Führt er die zweite Prüfung durch, wird er das Ergebnis des ersten Prüfers einfach immer bestätigen, unabhängig davon ob dieser eine Steuerhinterziehung aufdeckt oder nicht. Prüfer B ist dagegen besonders ehrgeizig. Führt er die zweite Prüfung durch und kam die erste Prüfung zu dem Ergebnis, dass keine Steuerhinterziehung vorliegt, wird er besonders aufmerksam arbeiten, wodurch seine Trefferquote von den sonst üblichen 0,4 auf 0,5 ansteigt. Welcher Prüfer die erste und welcher die zweite Prüfung durchführt, wird zu Beginn des Prüfverfahrens per fairem Los entschieden.

- a) Mit welcher Wahrscheinlichkeit geht der Steuerhinterzieher straffrei aus? (2 Punkte)
- b) Gibt es eine Prüfreihefolge, die für den Steuerhinterzieher günstiger ist – d.h. sollte er sich wünschen, dass ein bestimmter Prüfer die Erstprüfung durchführt, oder spielt dies aus der Perspektive des Geprüften keine Rolle? (3 Punkte)

Aufgabenteil IX: Kombinatorik (5 Punkte)

Pflichtverteidiger Müller werden von einem Amtsgericht per Zufallsverfahren mittellose Klienten zugewiesen. In diesem Monat finden dort 12 Prozesse gegen mittellose Klienten statt, die einen Pflichtverteidiger beantragt haben. Vier Klienten sollen von Herrn Müller vertreten werden. Wie viele unterschiedliche Klientengruppen könnten sich durch das Losverfahren ergeben?

Aufgabenteil X: Satz von Bayes (10 Punkte)

Eine neue gesetzliche Verordnung zwingt die Betreiber einer Online-Videoplattform dazu, jedes von den Nutzern hochgeladene Video auf mögliche Urheberrechtsverletzungen zu prüfen. Hierfür soll eine Prüfsoftware zum Einsatz kommen, die eine Urheberrechtsverletzung mit einer sehr hohen Wahrscheinlichkeit von 99,5% korrekt als solche erkennt. Allerdings stuft sie auch etwa 4% aller urheberrechtlich unkritischen Uploads fälschlicherweise als Urheberrechtsverletzungen ein. Pro Tag werden auf der Plattform etwa 40.000 Videos hochgeladen, in der Vergangenheit waren darunter rund 2% mit urheberrechtlich kritischen Inhalten.

Wie groß ist die Wahrscheinlichkeit dafür, dass ein von der Software als problematisch eingestuftes Upload sich in der Nachprüfung als echte Urheberrechtsverletzung erweist?

KLAUSUR STATISTIK I und II – KURZLÖSUNG

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2024

Christian Reinboth

Aufgabenteil I

F F R R R F R F R F

Aufgabenteil II

[2300; 2900; 3000; 3600; 4500; 4630; 5320; 5890; 6400; 23570]

 $10 \cdot 0,25 = 2,5 \rightarrow$ Wert an der 3. Stelle der geordneten Verteilung $\rightarrow 3.000$ $10 \cdot 0,50 = 5 \rightarrow$ Mittel der Werte an der 5. und 6. Stelle $\rightarrow (4500 + 4630) / 2 = 4565$ $10 \cdot 0,75 = 7,5 \rightarrow$ Wert an der 8. Stelle der geordneten Verteilung $\rightarrow 5890$ IQR = $5890 - 3000 = 2890$

1,5-facher IQR = 4335

Die Box läuft von 3.000 auf 5.890, der Median wird bei 4.565 eingezeichnet. Der obere Zaun läuft bis $(5.890 + 4.335 = 10.225)$ 6.400, der untere Zaun bis $(3.000 - 4.335 = -1.335)$ 2.300. Die 23.570 ist ein offensichtlicher Ausreißer, der sogar $(10.225 + 4.335 = 14.560)$ als Extremwert einzutragen ist.

Aufgabenteil IIIa) $62.110 / 10 = 6.211$ b) $36.240 / 8 = 4.530$ c) $0,35 \cdot 10 = 3,5 \rightarrow$ Wert an der 4. Stelle der geordneten Verteilung $\rightarrow 3.600$

d) Der Modus ist nicht bestimmbar, da es sich nicht um eine unimodale Verteilung handelt.

e) $23.570 - 2.300 = 21.270$ Aufgabenteil IV

a) Konkordanzkoeffizient nach Kendall

Fallnummer	Minuten seit letzter Pause (x)	rg (x)	Höhe des Bußgeldes (y)	rg (y)	K	D
B-17238	20 Min.	1	60 EUR	1	9	0
B-22163	28 Min.	2	80 EUR	5	5	3
B-23231	37 Min.	3	75 EUR	3,5	5	1
B-45437	49 Min.	4	70 EUR	2	6	0
B-63423	62 Min.	5	75 EUR	3,5	5	0
B-67612	78 Min.	6	90 EUR	6	4	0
B-75323	96 Min.	7	130 EUR	8,5	1	1
B-43634	121 Min.	8	125 EUR	7	2	0
B-45256	138 Min.	9	130 EUR	8,5	1	0
B-64654	140 Min.	10	155 EUR	10	0	0

Summe konkordanter Paare: 38

Summe diskordanter Paare: 5

 $\tau = (2 \cdot (K-D)) / (n \cdot (n-1)) = (2 \cdot 33) / 90 = 66 / 90 = 0,734$

Es liegt eine mittlere bis starke positive monotone Korrelation vor.

b) Bravais-Pearson-Korrelationskoeffizient

n	x	y	x ²	y ²	(x*y)
1	20	60	400	3600	1200
2	28	80	784	6400	2240
3	37	75	1369	5625	2775
4	49	70	2401	4900	3430
5	62	75	3844	5625	4650
6	78	90	6084	8100	7020
7	96	130	9216	16900	12480
8	121	125	14641	15625	15125
9	138	130	19044	16900	17940
10	140	155	19600	24025	21700
SUMME	769	990	77383	107700	88560
MITTEL	76,9	99	//	//	//

$$r = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sqrt{\sum_{i=1}^n (x_i^2) - n * \bar{x}^2} * \sqrt{\sum_{i=1}^n (y_i^2) - n * \bar{y}^2}}$$

$$= 12429 / (135,08 * 98,44) = 0,9347$$

Es liegt eine starke positive lineare Korrelation vor.

Aufgabenteil V

R R R R F R F F R R

Aufgabenteil VI

n	x	y	x ²	(x*y)
1	20	60	400	1200
2	28	80	784	2240
3	37	75	1369	2775
4	49	70	2401	3430
5	62	75	3844	4650
6	78	90	6084	7020
7	96	130	9216	12480
8	121	125	14641	15125
9	138	130	19044	17940
10	140	155	19600	21700
SUMME	769	990	77383	88560
MITTEL	76,9	99	//	//

$$b = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sum_{i=1}^n (x_i^2) - n * \bar{x}^2}$$

$$b = 0,6812$$

$$a = 99 - 0,6812 \cdot 76,9 = 46,6157$$

$$y = 46,6157 + 0,6812 \cdot x \rightarrow y = 46,62 + 0,68 \cdot x$$

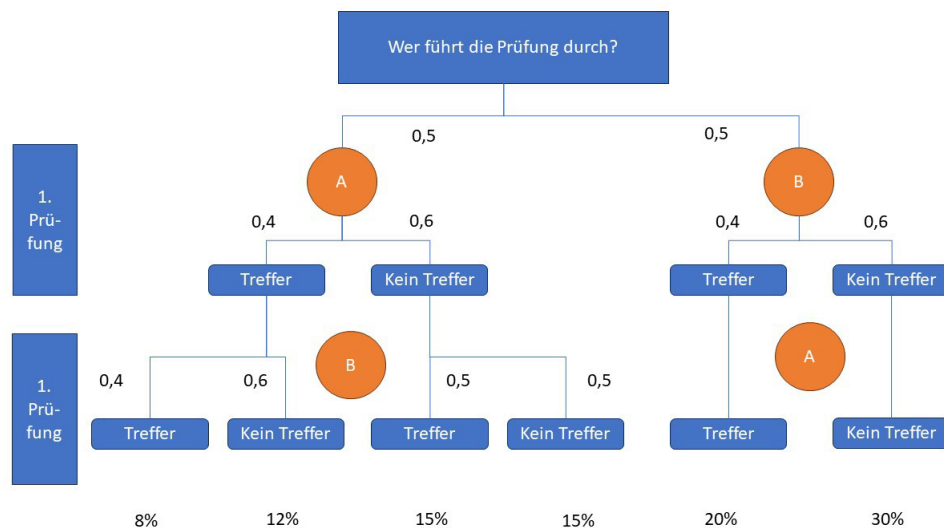
$$R^2 = 0,9347^2 = 0,8737$$

Das Modell hat eine hervorragende Streuungsaufklärung.

Aufgabenteil VII

- a) Nur die Schnittflächen von A1 und A3
- b) Gesamtes Diagramm einschließlich Ereignisraum

Aufgabenteil VIII



- a) Pfade, bei denen der Steuerhinterzieher nicht erwischt wird

$$0,5 \cdot 0,6 \cdot 0,5 = 15\%$$

$$0,5 \cdot 0,6 = 30\%$$

Der Steuerhinterzieher wird mit einer Wahrscheinlichkeit von 45% nicht erwischt.

- b) Vorteilhaft für den Steuerhinterzieher ist eine Erstprüfung durch B. Da A sich dem Urteil von B einfach nur anschließt, wird er nicht erwischt, sobald B nicht fündig wird. Bei einer Erstprüfung durch A findet dagegen im Falle eines ausbleibenden Treffers eine zweite ernsthafte Prüfung durch B statt.

Aufgabenteil IX

Es handelt sich um eine Kombination (Reihenfolge der Ereignisse spielt keine Rolle, wichtig ist nur, welche Personen in der zu vertretenen Gruppe sind) im Modell ohne Zurücklegen (jede Person kann nur höchstens einmal in die Gruppe gezogen werden). Somit ist zu rechnen:

$$n! / (k! \cdot (n-k)!) \quad n = 12$$

$$k = 4$$

$$12! = 479.001.600$$

$$4! = 24$$

$$8! = 40.320$$

$$479.001.600 / (24 * 40.320) = 479.001.600 / 967.680 = 495$$

Aufgabenteil X

40.000 Uploads

...davon 98% unkritisch: 39.200

...davon 2% kritisch: 800

Von 800 kritischen Videos

...werden 99,5% korrekt als kritisch erkannt: 796 (true positives)

...werden 0,5% fälschlicherweise nicht als kritisch erkannt: 4

Von 39.200 unkritischen Videos

...werden 96% korrekterweise als nicht kritisch erkannt: 37.632

...werden 4% fälschlicherweise als kritisch erkannt: 1.568 (false positives)

Insgesamt werden also $796 + 1.568 = 2.364$ Videos als kritisch erkannt, davon 796 zu Recht: 0,3367.

Sommersemester 2025

Klausur Statistik I

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang Betriebswirtschaftslehre
Sommersemester 2025

Prüfungsfach:

Statistik Teil 1

Name des Prüfers: Christian Reinboth

Prüfungstag: 10. Mai 2025

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich auf dem Klausurexemplar! Dafür wurde ausreichend Platz vorgesehen. Benutzen Sie auch die Blattrückseiten!
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2025

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☐ falsch

2) Verhältnisskala und Intervallskala unterscheiden sich durch den natürlichen Nullpunkt.

☐ richtig ☐ falsch

3) Der Median ist unempfindlich gegenüber Ausreißern und gilt daher als „robust“.

☐ richtig ☐ falsch

4) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☐ falsch

5) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☐ falsch

6) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig ☐ falsch

7) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig ☐ falsch

8) Qualitative und quantitative Methoden der Datenerhebung und Datenanalyse können grundsätzlich nicht miteinander kombiniert werden.

☐ richtig ☐ falsch

9) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☐ falsch

10) Die Größe der Box im Box-Plot wird durch oberen und unteren Quartilswert definiert.

☐ richtig ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Im Rahmen eines Vintage-Videospiel-Wettbewerbs sollen die Teilnehmenden über einen möglichst langen Zeitraum im Arcade-Spiel „Racing Master 1991“ ein Rennauto steuern. Zur Ermittlung des Gewinners wird die Zeit bis zum ersten Unfall in Sekunden erfasst. Die nachfolgenden Werte stammen von den zehn Gamerinnen und Gamern, die es bis ins Finale geschafft haben.

Teilnehmer	01	02	03	04	05	06	07	08	09	10
Spielzeit (in Sek.)	364	283	165	241	316	425	312	113	175	223

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Spielzeiten und benennen Sie die für die Konstruktion erforderlichen Größen. (15 Punkte) Wie groß müsste ein Wert in dieser Verteilung sein, um im oberen Wertebereich als Extremwert klassifiziert zu werden? (5 Punkte)

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Der Umfang von Software wird gelegentlich in der Einheit „Lines of Code“ (LOC) bzw. „Source Lines of Code“ (SLOC) gemessen – der Anzahl der Zeilen des Quelltextes, aus denen das Programm besteht. Ein Videospiel-Hersteller gibt für die von ihm in den letzten Jahren entwickelten Softwareprodukte die folgenden Gesamtgrößen (in Tausend LOC – TLOC) an.

Titel des Videospiels	Erscheinungsjahr	Umfang in TLOC
Teatris	2014	16
Super Müsli Bros.	2015	37
Tomb Raver	2015	42
Grand Theft Alpaca	2016	45
Devil May Fry	2018	61
Masters of Teräs Käsi	2019	85
League of Lettuce	2019	117
Candy Clash	2020	123
Uncharted Waters	2021	140
Mega Manatee	2022	145



Berechnen Sie für die TLOC-Angaben die folgenden Lage- und Streuungsmaße:

- Arithmetisches Mittel (5 Punkte)
- Um 5% getrimmtes arithmetisches Mittel (5 Punkte)
- Modus (5 Punkte)
- Spannweite (5 Punkte)
- Varianz (10 Punkte)
- Standardabweichung (5 Punkte)

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Die Aktualisierungsrate eines Bildschirms gemessen in Hertz (Hz) bestimmt, wie oft das Display pro Sekunde aktualisiert wird. Höhere Aktualisierungsraten können zu flüssigeren Bewegungen führen, was es Spielerinnen und Spielern ermöglicht, schneller auf Änderungen im Spiel zu reagieren. Ein Spieleentwickler will diesen möglichen Zusammenhang zwischen der Aktualisierungsrate und der Reaktionszeit untersuchen. Zu

diesem Zweck wird ein Experiment durchgeführt, bei dem die Aktualisierungsrate langsam gesteigert und die jeweils beste Reaktionszeit eines Spielers zur jeweiligen Rate aufgezeichnet wird.

Aktualisierungsrate in Hertz (Hz)	Beste Reaktionszeit in Millisekunden
60	280
75	273
90	266
120	254
144	251
165	254
180	223
240	236
300	191
360	170

Bestimme für die gemeinsame Verteilung von Aktualisierungsrate und Reaktionszeit sowohl den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) als auch den Konkordanzkoeffizienten nach Kendall (15 Punkte). Interpretiere beide Werte einzeln und im Vergleich zueinander (5 Punkte).

MUSTERLÖSUNG KLAUSUR STATISTIK I

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2025

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) In Stem-and-Leaf-Diagrammen repräsentiert ein Blatt stets einen Fall.

☐ richtig ☒ **falsch**

2) Verhältnisskala und Intervallskala unterscheiden sich durch den natürlichen Nullpunkt.

☒ **richtig** ☐ falsch

3) Der Median ist unempfindlich gegenüber Ausreißern und gilt daher als „robust“.

☒ **richtig** ☐ falsch

4) Die Zäune im Box-Plot umfassen stets eine Strecke mit Länge des 1,5-fachen IQR.

☐ richtig ☒ **falsch**

5) Der Modus ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig ☒ **falsch**

6) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☒ **richtig** ☐ falsch

7) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☒ **richtig** ☐ falsch

8) Qualitative und quantitative Methoden der Datenerhebung und Datenanalyse können grundsätzlich nicht miteinander kombiniert werden.

☐ richtig ☒ **falsch**

9) Eine per Zufallsauswahl zusammengestellte Stichprobe ist stets repräsentativ.

☐ richtig ☒ **falsch**

10) Die Größe der Box im Box-Plot wird durch oberen und unteren Quartilswert definiert.

☒ **richtig** ☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (20 Punkte)

Im Rahmen eines Vintage-Videospiel-Wettbewerbs sollen die Teilnehmenden über einen möglichst langen Zeitraum im Arcade-Spiel „Racing Master 1991“ ein Rennauto steuern. Zur Ermittlung des Gewinners wird die Zeit bis zum ersten Unfall in Sekunden erfasst. Die nachfolgenden Werte stammen von den zehn Gamerinnen und Gamern, die es bis ins Finale geschafft haben.

Teilnehmer	01	02	03	04	05	06	07	08	09	10
Spielzeit (in Sek.)	364	283	165	241	316	425	312	113	175	223

Zeichnen Sie einen erweiterten Box-Plot für die Verteilung der Spielzeiten und benennen Sie die für die Konstruktion erforderlichen Größen. (15 Punkte) Wie groß müsste ein Wert in dieser Verteilung sein, um im oberen Wertebereich als Extremwert klassifiziert zu werden? (5 Punkte)

Schritt 1: Bildung einer geordneten Reihe

113; 165; 175; 223; 241; 283; 312; 316; 364; 425 (n = 10)

Schritt 2: Berechnung von Quartilen und IQR

$x_{0,25} = 10 \cdot 0,25 = 2,5 \rightarrow 3$. Wert der geordneten Verteilung $\rightarrow 175$

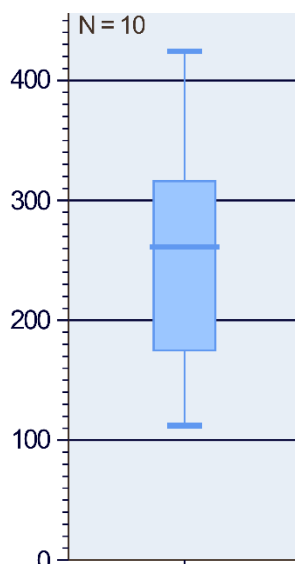
$x_{0,50} = 10 \cdot 0,5 \rightarrow 5 \rightarrow$ Mittel des 5. und 6. Wertes $\rightarrow (241 + 283) / 2 = 262$

$x_{0,75} = 10 \cdot 0,75 \rightarrow 7,5 \rightarrow 8$. Wert der geordneten Verteilung $\rightarrow 316$

$IQR = 316 - 175 = 141$

1,5-facher IQR = 211,5

Die Box läuft somit von 175 bis 316, der Median wird bei 262 eingezeichnet. Der obere Zaun könnte maximal bis zum Wert $316 + 211,5 = 527,5$ laufen, der größte Wert in diesem Bereich (und damit das Zaunende) ist die 425. Der untere Zaun könnte minimal bis zum Wert $175 - 211,5 = -36,5$ laufen, der kleinste Wert in diesem Bereich (und damit das untere Zaunende) ist die 113. Die Zäune laufen damit bis zur 425 bzw. bis zur 113. Ausreißer oder Extremwerte gibt es keine. Um als Extremwert im oberen Bereich klassifiziert zu werden, müsste ein Wert größer als $316 + 211,5 + 211,5 = 739$ sein.



Grafik erstellt mit SOFA Statistics
(<https://sourceforge.net/projects/sofastatistics/>).

Aufgabenteil III: Verteilungsparameter (35 Punkte)

Der Umfang von Software wird gelegentlich in der Einheit „Lines of Code“ (LOC) bzw. „Source Lines of Code“ (SLOC) gemessen – der Anzahl der Zeilen des Quelltextes, aus denen das Programm besteht. Ein Videospiel-Hersteller gibt für die von ihm in den letzten Jahren entwickelten Softwareprodukte die folgenden Gesamtgrößen (in Tausend LOC – TLOC) an.

Titel des Videospiels	Erscheinungsjahr	Umfang in TLOC
Teatris	2014	16
Super Müsli Bros.	2015	37
Tomb Raver	2015	42
Grand Theft Alpaca	2016	45
Devil May Fry	2018	61
Masters of Teräs Käsi	2019	85
League of Lettuce	2019	117
Candy Clash	2020	123
Uncharted Waters	2021	140
Mega Manatee	2022	145



Berechnen Sie für die TLOC-Angaben die folgenden Lage- und Streuungsmaße:

- a) Arithmetisches Mittel (5 Punkte)
- b) Um 5% getrimmtes arithmetisches Mittel (5 Punkte)
- c) Modus (5 Punkte)
- d) Spannweite (5 Punkte)
- e) Varianz (10 Punkte)
- f) Standardabweichung (5 Punkte)

a) Arithmetisches Mittel

$$n = 10$$

$$16 + 37 + 42 + 45 + 61 + 85 + 117 + 123 + 140 + 145 = 811$$

$$811 / 10 = \underline{81,1 \text{ TLOC}}$$

b) Um 5% getrimmtes arithmetisches Mittel

$10 \cdot 0,05 = 0,5 \rightarrow$ Kürzung um einen Wert am oberen und unteren Ende der geordneten Verteilung.

$$16; 37; 42; 45; 61; 85; 117; 123; 140; 145$$

$$37 + 42 + 45 + 61 + 85 + 117 + 123 + 140 = 650$$

$$650 / 8 = \underline{81,25 \text{ TLOC}}$$

Da der Datensatz keine Ausreißer enthält, ist die Berechnung des getrimmten arithmetischen Mittels überflüssig (und führt daher auch offensichtlich zu einem nahezu identischen Ergebnis).

c) Modus

Die Verteilung weist kein eindeutiges Maximum auf, somit lässt sich kein Modus bestimmen.

d) Spannweite

$$145 - 16 = \underline{129 \text{ TLOC}}$$

e) Varianz

Das arithmetische Mittel liegt bei 81,1.

$$(16 - 81,1)^2 = 4238,01$$

$$(37 - 81,1)^2 = 1944,81$$

$$(42 - 81,1)^2 = 1528,81$$

$$(45 - 81,1)^2 = 1303,21$$

$$(61 - 81,1)^2 = 404,01$$

$$(85 - 81,1)^2 = 15,21$$

$$(117 - 81,1)^2 = 1288,81$$

$$(123 - 81,1)^2 = 1755,61$$

$$(140 - 81,1)^2 = 3469,21$$

$$(145 - 81,1)^2 = 4083,21$$

Die Summe aller quadrierten Abweichungen der Werte vom arithmetischen Mittel beträgt 20.030,90.

$$20.030,90 / 10 = \underline{2.003,09 \text{ TLOC}^2}$$

f) Standardabweichung

Die positive Wurzel der Varianz beträgt 44,76 TLOC.

Aufgabenteil IV: Zusammenhangsmaße (35 Punkte)

Die Aktualisierungsrate eines Bildschirms gemessen in Hertz (Hz) bestimmt, wie oft das Display pro Sekunde aktualisiert wird. Höhere Aktualisierungsraten können zu flüssigeren Bewegungen führen, was es Spielerinnen und Spielern ermöglicht, schneller auf Änderungen im Spiel zu reagieren. Ein Spieleentwickler will diesen möglichen Zusammenhang zwischen der Aktualisierungsrate und der Reaktionszeit untersuchen. Zu diesem Zweck wird ein Experiment durchgeführt, bei dem die Aktualisierungsrate langsam gesteigert und die jeweils beste Reaktionszeit eines Spielers zur jeweiligen Rate aufgezeichnet wird.

Aktualisierungsrate in Hertz (Hz)	Beste Reaktionszeit in Millisekunden
60	280
75	273
90	266
120	254
144	251
165	254
180	223
240	236
300	191
360	170

Bestimme für die gemeinsame Verteilung von Aktualisierungsrate und Reaktionszeit sowohl den Rangkorrelationskoeffizienten nach Spearman (15 Punkte) als auch den Konkordanzkoeffizienten nach Kendall (15 Punkte). Interpretiere beide Werte einzeln und im Vergleich zueinander (5 Punkte).

Berechnung des Rangkorrelationskoeffizienten nach Spearman

Werte von x	Ränge von x	Werte von y	Ränge von y	Differenzen d	d ²
60	1	280	10	-9	81
75	2	273	9	-7	49
90	3	266	8	-5	25
120	4	254	6,5	-2,5	6,25
144	5	251	5	0	0
165	6	254	6,5	-0,5	0,25
180	7	223	3	4	16
240	8	236	4	4	16
300	9	191	2	7	49
360	10	170	1	9	81

Die Summe der quadrierten Rangplatzdifferenzen liegt somit bei 323,5.

Eingesetzt in die Formel für Spearmans rho ergibt sich: $1 - [(6 * 323,5) / ((10^2 - 1) * 10)]$
 $= 1 - (1941 / 990) = 1 - 1,96 = -0,96$

Berechnung des Konkordanzkoeffizienten nach Kendall

Werte von x	Ränge von x	Werte von y	Ränge von y	Konkordante Paare	Diskordante Paare
60	1	280	10	0	9
75	2	273	9	0	8
90	3	266	8	0	7
120	4	254	6,5	0	5
144	5	251	5	1	4
165	6	254	6,5	0	4
180	7	223	3	1	2
240	8	236	4	0	2
300	9	191	2	0	1
360	10	170	1	//	//

Die Summe der konkordanten Paare liegt somit bei 2, die Summe der diskordanten Paare bei 42.

Eingesetzt in die Formel für Kendalls tau ergibt sich: $[2 * (2 - 42)] / [10 * (10 - 1)] = -80 / 90 = -0,89$

Vergleichende Interpretation der beiden Koeffizienten

Beide Koeffizienten zeigen eine (sehr) starke, negative und monotone Korrelation zwischen der Aktualisierungsrate und der besten Reaktionszeit. Diese Korrelation sollte in der Praxis näher untersucht, jedoch nicht unmittelbar kausal interpretiert werden, auch wenn ein inhaltlicher Zusammenhang naheliegt. Der verbundene Rang schwächt die Aussagekraft beider Koeffizienten – da aber im Datensatz bei 20 Datenpunkten insgesamt nur ein verbundener Rang vorkommt, ist die Berechnung und Interpretation durchaus zulässig (obwohl der Datensatz in der Praxis natürlich zu klein wäre, um eine sinnvolle Korrelationsanalyse durchführen zu können).

Sommersemester 2025

Klausur Statistik II

Hochschule Harz
Wernigerode
Fachbereich Wirtschaftswissenschaften
bbgl. Bachelorstudiengang BWL B.A.
Sommersemester 2025

Prüfungsfach:

Statistik Teil 1 und 2

Name des Prüfers: Christian Reinboth

Prüfungstag: 10. Juli 2025

Matrikelnummer: _____

Vom Prüfer auszufüllen:

erreichte Punktzahl:

Note:

Bemerkungen:

Es ist nicht erlaubt, ein Handy mit in die Prüfung zu nehmen. Die Nutzung von Mobiltelefonen, PDAs und ähnlichen Speichermedien in der Prüfung ist unzulässig. Alle mitgebrachten Geräte sind auszuschalten und vom Prüfungstische zu räumen.

Täuschungsversuche führen zur Bewertung der Prüfungsleistung als „nicht ausreichend“ (5,0).

- Zugelassene Hilfsmittel: Formelsammlung, Taschenrechner
- Bitte überprüfen Sie sofort die Vollständigkeit Ihres Klausurexemplars.
Bei Unvollständigkeit wenden Sie sich bitte an den Aufsichtführenden.
- Beschriften Sie dann das Deckblatt mit Ihrer Matrikelnummer.
- Bearbeiten Sie die Aufgaben ausschließlich aus den ausgeteilten Seiten! Falls Sie weitere Seiten benötigen sollten, melden Sie sich bitte beim Aufsichtführenden.
- Zerlegen Sie das Klausurexemplar nicht in einzelne Blätter!

VIEL ERFOLG!

KLAUSUR STATISTIK I und II

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2025

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

☐ richtig☐ falsch

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

☐ richtig☐ falsch

3) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

☐ richtig☐ falsch

4) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

☐ richtig☐ falsch

5) Eine repräsentative Zufallsauswahl setzt die genaue Definition der Grundgesamtheit voraus.

☐ richtig☐ falsch

6) Die Fünf-Werte-Zusammenfassung kann nur für metrisch skalierte Werte berechnet werden.

☐ richtig☐ falsch

7) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

☐ richtig☐ falsch

8) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

☐ richtig☐ falsch

9) Verhältnisskala und Intervallskala unterscheiden sich durch den natürlichen Nullpunkt.

☐ richtig☐ falsch

10) Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert.

☐ richtig☐ falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

In einer Regionalliga wird statistisch erfasst, wie viele Gelbe Karten und Rote Karten ein Verein pro Saison erhält. Die nachfolgende Tabelle zeigt die Verteilung der Gelben Karten für die Saison 2024.

Verein	Gelbe Karten (Saison 2024)
FC Böhnshausen	12
Sportclub Wacker	11
Sportfreunde Harz	13
FC Fallingbostal	8
MSV Kleinhausen	5
FC Mühlenbach	11
SF Meierburg	16
Sportclub Burgdorf	9
MSV Großdendorf	21



Erstellen Sie ein Stem-and-Leaf-Diagramm mit einer aus Ihrer Sicht geeigneten Stammbreite für die Verteilung der Gelben Karten.

Aufgabenteil III: Verteilungsparameter (15 Punkte)

In der besagten Regionalliga werden auch Ablösesummen für Spieler gezahlt, die den Verein wechseln – wenn auch nicht so exorbitant hohe. Seit 2020 kam es zu insgesamt sieben solchen Zahlungen.

Spieler	Ablösesumme (in EUR)
Heinrich Meier	11.500
Klaus Weinrich	8.300
Hans Meiser	2.300
Peter Karstens	6.100
Tarik Aydinli	8.200
Merten Horstmann	17.000
Klaus Müller	6.000



Berechnen Sie für die Verteilung der Ablösesummen

- das arithmetische Mittel (3 Punkte),
- den Median (3 Punkte),
- den 10%-Perzentilwert (3 Punkte) und
- die Spannweite (2 Punkte).
- Wie groß müsste eine Ablösesumme sein, damit Sie diese in der vorliegenden Verteilung als Ausreißer klassifizieren würden? Begründen Sie Ihre Entscheidung. (4 Punkte)

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Der Redakteur einer Sportzeitschrift möchte überprüfen, ob es zwischen dem Ballbesitz (der Anteil der Spielzeit, in der sich der Ball im Besitz einer Mannschaft befindet) und den Spielerfolgen einen Zusammenhang geben könnte. Für die bekannten neun Kreisliga-Mannschaften hat er daher den durchschnittlichen Ballbesitz (in Minuten) über alle Spiele der vergangenen Saison sowie die Zahl der gewonnenen Spiele in dieser Saison recherchiert.

Verein	Ø Ballbesitz	Saisonsiege
FC Böhnhausen	32	0
Sportclub Wacker	38	1
Sportfreunde Harz	43	2
FC Fallingbostal	46	2
MSV Kleinhausen	48	4
FC Mühlenbach	51	3
SF Meierburg	55	5
Sportclub Burgdorf	57	5
MSV Großdendorf	61	8

Berechnen und interpretieren Sie für die beiden Variablen „durchschnittlicher Ballbesitz“ und „Saisonsiege“ den Rangkorrelationskoeffizienten nach Spearman (6 Punkte) sowie den Bravais-Pearson-Korrelationskoeffizienten (9 Punkte).

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die Wahrscheinlichkeit für fehlerhafte Testergebnisse nimmt mit der Anzahl durchgeführter Tests zu. Dieses Phänomen wird auch als alpha-Fehlerinflation bezeichnet.

☐ richtig ☐ falsch

2) Eine zufällig gezogene Stichprobe mit hoher Rücklaufquote ist unabhängig von ihrem Umfang stets repräsentativ.

☐ richtig ☐ falsch

3) Sowohl bei Variationen als auch bei Kombinationen wird in Modelle mit und ohne Zurücklegen unterschieden.

☐ richtig ☐ falsch

4) Eine Teilmenge einer leeren Menge kann immer wieder nur selbst eine leere Menge sein.

☐ richtig ☐ falsch

5) Ein Würfel wird dann als fair bezeichnet, wenn er bei (in etwa) der Hälfte aller Würfe eine gerade Augenzahl anzeigt.

☐ richtig ☐ falsch

6) Das logische NICHT wird auch als Negation bezeichnet.

☐ richtig ☐ falsch

7) Die Axiome von Kolmogoroff definieren die mathematischen Eigenschaften des Modells mit bzw. ohne Zurücklegen.

☐ richtig ☐ falsch

8) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Leverage-Effekt bezeichnet.

☐ richtig ☐ falsch

9) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleichbleibenden Rahmenbedingungen.

☐ richtig ☐ falsch

10) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

☐ richtig ☐ falsch

Aufgabenteil VI: Lineare Regression (15 Punkte)

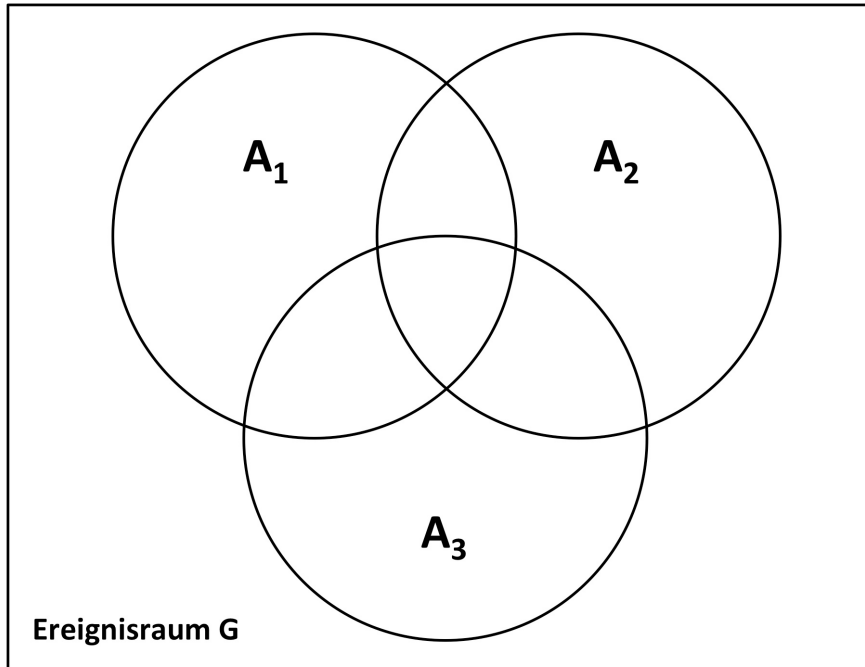
Verein	Ø Ballbesitz	Saisonsiege
FC Böhnhausen	32	0
Sportclub Wacker	38	1
Sportfreunde Harz	43	2
FC Fallingbostal	46	2
MSV Kleinhausen	48	4
FC Mühlenbach	51	3
SF Meierburg	55	5
Sportclub Burgdorf	57	5
MSV Großdendorf	61	8

Stellen Sie für die bereits bekannte Verteilung von durchschnittlichem Ballbesitz und Saisonsiegen ein lineares Regressionsmodell mit dem Ballbesitz (in Minuten) als unabhängiger Variablen (x) und den Saisonsiegen als abhängiger Variablen (y) auf (10 Punkte) und bestimmen und interpretieren Sie dessen Güte (5 Punkte).

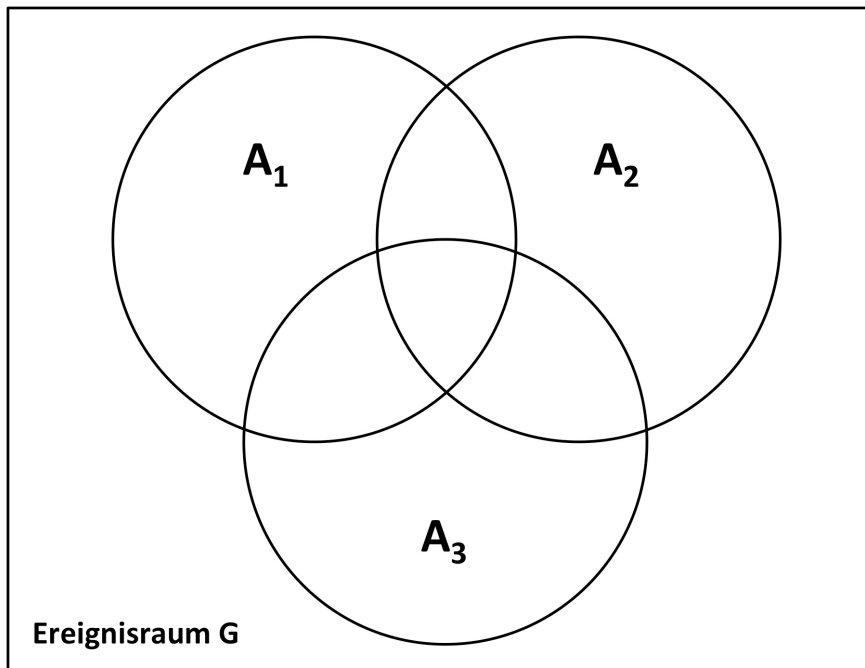
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_3 \cup A_2) \cup A_1$ (3 Punkte)



b) $(A_1 \cup \overline{A_1})$ (2 Punkte)



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

Eine Fußballmannschaft geht in einem Spiel mit einer Wahrscheinlichkeit von 60% während der ersten Halbzeit in Führung. Startet diese Mannschaft mit einer Führung in die zweite Halbzeit, so ist sie besonders motiviert und gewinnt das Spiel mit einer Wahrscheinlichkeit von 70%. Befindet sie sich zu Beginn der zweiten Halbzeit dagegen im Rückstand, sorgt dies für zusätzliche Nervosität und die Wahrscheinlichkeit eines Sieges sinkt auf 50%.

- a) Mit welcher Wahrscheinlichkeit gewinnt diese Mannschaft ihr nächstes Spiel? (3 Punkte)
- b) Kurz vor dem nächsten Spiel verletzt sich der beste Stürmer dieser Mannschaft, wodurch sich die Wahrscheinlichkeit sowohl einer Führung während der ersten Halbzeit als auch eines Sieges in der zweiten Halbzeit jeweils halbieren. Mit welcher Wahrscheinlichkeit wird sie nun gewinnen? (2 Punkte)

Aufgabenteil IX: Kombinatorik (5 Punkte)

In der Harzer Fußball-Kreisliga spielen derzeit 9 Mannschaften. In einem Wettbewerb sollen alle Mannschaften im Spielmodus "jeder gegen jeden" gegeneinander antreten.

Wie viele Spiele müssen insgesamt stattfinden?

**Aufgabenteil X: Satz von Bayes (10 Punkte)**

In einem Fußballstadion kommt es regelmäßig zu Problemen mit gewalttätigen Fans. Diese haben zwar bereits ein Hausverbot erhalten, tauchen aber dennoch regelmäßig bei Spielen auf. Eine neue Video-Überwachungssoftware soll nun zukünftig dabei helfen, diesen Personenkreis aus dem Stadion fernzuhalten. Ein Kamerasystem scannt die Gesichter aller Personen, die das Gelände betreten und leitet diese an ein KI-Tool weiter, welches die Scans mit Fotos der Personen vergleicht, die bereits ein Hausverbot erhalten haben. Dabei kann das Tool eine gesuchte Person mit einer Genauigkeit von 93% korrekt identifizieren, wenn sie eine Kamera passiert. Leider ordnet es auch in 2% aller Fälle eine nicht gesuchte Person fälschlicherweise als gesucht ein.

Zu einem bedeutenden Spiel werden rund 5.000 Besucherinnen und Besucher erwartet. Auf der Liste der Gewalttäter mit Hausverbot stehen derzeit 80 Namen und es ist davon auszugehen, dass etwa ein Viertel dieser Personen versuchen wird, unerkannt ins Stadion zu gelangen.

Das KI-Tool meldet, dass eine gesuchte Person erkannt wurde. Mit welcher Wahrscheinlichkeit wird eine Personenkontrolle nun zu dem Ergebnis führen, dass es sich auch tatsächlich um eine Person mit Hausverbot handelt?



KLAUSUR STATISTIK I und II mit Lösungen

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2025

Christian Reinboth

Aufgabenteil I: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine starke lineare Korrelation hin.

X richtig

O falsch

2) Eine mittels eines Korrelationskoeffizienten identifizierte Korrelation sollte näher untersucht, nicht jedoch unmittelbar inhaltlich interpretiert werden.

X richtig

O falsch

3) Der Median ist derjenige Wert, der genau in der Mitte der geordneten Verteilung liegt.

X richtig

O falsch

4) Der Interquartilsabstand (IQR) ist der Abstand zwischen oberem und unterem Quartil.

X richtig

O falsch

5) Eine repräsentative Zufallsauswahl setzt die genaue Definition der Grundgesamtheit voraus.

X richtig

O falsch

6) Die Fünf-Werte-Zusammenfassung kann nur für metrisch skalierte Werte berechnet werden.

O richtig

X falsch

7) Fehlende Werte können nur bei Personenbefragungen auftreten, da sie stets auf die Verweigerung von Auskünften durch Befragte zurückzuführen sind.

O richtig

X falsch

8) Die einzige in der induktiven Statistik zulässige Definition eines Ausreißers stammt aus der Konstruktionsvorschrift für den Box-Plot.

O richtig

X falsch

9) Verhältnisskala und Intervallskala unterscheiden sich durch den natürlichen Nullpunkt.

X richtig

O falsch

10) Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert.

X richtig

O falsch

Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)

In einer Regionalliga wird statistisch erfasst, wie viele Gelbe Karten und Rote Karten ein Verein pro Saison erhält. Die nachfolgende Tabelle zeigt die Verteilung der Gelben Karten für die Saison 2024.

Verein	Gelbe Karten (Saison 2024)
FC Böhnshausen	12
Sportclub Wacker	11
Sportfreunde Harz	13
FC Fallingbostal	8
MSV Kleinhausen	5
FC Mühlenbach	11
SF Meierburg	16
Sportclub Burgdorf	9
MSV Großdendorf	21



Erstellen Sie ein Stem-and-Leaf-Diagramm mit einer aus Ihrer Sicht geeigneten Stammbreite für die Verteilung der Gelben Karten.

```
0 | 5 8 9
1 | 1 1 2 3 6
2 | 1
```

Stammbreite 10

Jedes Blatt 1 Fall

Keine Ausreißer

Aufgabenteil III: Verteilungsparameter (15 Punkte)

In der besagten Regionalliga werden auch Ablösesummen für Spieler gezahlt, die den Verein wechseln – wenn auch nicht so exorbitant hohe. Seit 2020 kam es zu insgesamt sieben solchen Zahlungen.

Spieler	Ablösesumme (in EUR)
Heinrich Meier	11.500
Klaus Weinrich	8.300
Hans Meiser	2.300
Peter Karstens	6.100
Tarik Aydinli	8.200
Merten Horstmann	17.000
Klaus Müller	6.000



Berechnen Sie für die Verteilung der Ablösesummen

- das arithmetische Mittel (3 Punkte),
- den Median (3 Punkte),
- den 10%-Perzentilwert (3 Punkte) und
- die Spannweite (2 Punkte).
- Wie groß müsste eine Ablösesumme sein, damit Sie diese in der vorliegenden Verteilung als Ausreißer klassifizieren würden? Begründen Sie Ihre Entscheidung. (4 Punkte)

a) Arithmetisches Mittel

$$11.500 + 8.300 + 2.300 + 6.100 + 8.200 + 17.000 + 6.000 = 59.400$$

$$59.400 / 7 = 8.485,71$$

b) Median

Geordnete Verteilung der Werte:

2.300; 6.000; 6.100; 8.200; 8.300; 11.500; 17.000

$0,5 * 7 = 3,5 \rightarrow 4$. Wert der geordneten Verteilung $\rightarrow 8.200$

c) 10%-Perzentilwert

$0,1 * 7 = 0,7 \rightarrow 1$. Wert der geordneten Verteilung $\rightarrow 2.300$

d) Spannweite

$$17.000 - 2.300 = 14.700$$

e) Ausreißer

Möglich wäre eine Klassifizierung auf Basis der Konstruktionsvorschriften des Box-Plots.

$0,25 * 7 = 1,75 \rightarrow 2$. Wert der geordneten Verteilung $\rightarrow 6.000$

$0,75 * 7 = 5,25 \rightarrow 6$. Wert der geordneten Verteilung $\rightarrow 11.500$

IQR $\rightarrow 11.500 - 6.000 = 5.500$

1-5-facher IQR $\rightarrow 8.250$

$$11.500 + 8.250 = 19.750$$

Aufgabenteil IV: Korrelationskoeffizienten (15 Punkte)

Der Redakteur einer Sportzeitschrift möchte überprüfen, ob es zwischen dem Ballbesitz (der Anteil der Spielzeit, in der sich der Ball im Besitz einer Mannschaft befindet) und den Spielerfolgen einen Zusammenhang geben könnte. Für die bekannten neun Kreisliga-Mannschaften hat er daher den durchschnittlichen Ballbesitz (in Minuten) über alle Spiele der vergangenen Saison sowie die Zahl der gewonnenen Spiele in dieser Saison recherchiert.

Verein	Ø Ballbesitz	Saisonsiege
FC Böhnhäusen	32	0
Sportclub Wacker	38	1
Sportfreunde Harz	43	2
FC Fallingb.otel	46	2
MSV Kleinhausen	48	4
FC Mühlenbach	51	3
SF Meierburg	55	5
Sportclub Burgdorf	57	5
MSV Großdendorf	61	8

Berechnen und interpretieren Sie für die beiden Variablen „durchschnittlicher Ballbesitz“ und „Saisonsiege“ den Rangkorrelationskoeffizienten nach Spearman (6 Punkte) sowie den Bravais-Pearson-Korrelationskoeffizienten (9 Punkte).

a) Rangkorrelationskoeffizient nach Spearman

Verein	x	rg(x)	y	rg(y)	d	d ²
FC Böhnhäusen	32	1	0	1	0	0
Sportclub Wacker	38	2	1	2	0	0
Sportfreunde Harz	43	3	2	3,5	-0,5	0,25
FC Fallingb.otel	46	4	2	3,5	0,5	0,25
MSV Kleinhausen	48	5	4	6	-1	1
FC Mühlenbach	51	6	3	5	1	1
SF Meierburg	55	7	5	7,5	-0,5	0,25
Sportclub Burgdorf	57	8	5	7,5	0,5	0,25
MSV Großdendorf	61	9	8	9	0	0
Summe	//	//	//	//	//	3

$$\rho = 1 - ((6 \cdot 3) / ((9^2 - 1) \cdot 9)) = 1 - 0,025 = 0,975$$

Es liegt eine starke, positive, monotone Korrelation vor.

b) Bravais-Pearson-Korrelationskoeffizient

Verein	x	y	x ²	y ²	(x*y)
FC Böhnhäusen	32	0	1024	0	0
Sportclub Wacker	38	1	1444	1	38
Sportfreunde Harz	43	2	1849	4	86
FC Fallingb.otel	46	2	2116	4	92
MSV Kleinhausen	48	4	2304	16	192
FC Mühlenbach	51	3	2601	9	153
SF Meierburg	55	5	3025	25	275
Sportclub Burgdorf	57	5	3249	25	285
MSV Großdendorf	61	8	3721	64	488
Summe	431	30	21333	148	1609
Durchschnitt	47,89	3,33	//	//	//

$$r = 173,74 / (26,3 * 6,94) = 0,951$$

Es liegt eine starke, positive, lineare Korrelation vor.

Aufgabenteil V: Theorie (10 Punkte)

Sind die nachfolgenden Aussagen richtig oder falsch? (1 Punkt pro korrekter Beantwortung)

1) Die Wahrscheinlichkeit für fehlerhafte Testergebnisse nimmt mit der Anzahl durchgeführter Tests zu. Dieses Phänomen wird auch als alpha-Fehlerinflation bezeichnet.

X richtig ☐ O falsch

2) Eine zufällig gezogene Stichprobe mit hoher Rücklaufquote ist unabhängig von ihrem Umfang stets repräsentativ.

☐ O richtig **X falsch**

3) Sowohl bei Variationen als auch bei Kombinationen wird in Modelle mit und ohne Zurücklegen unterschieden.

X richtig ☐ O falsch

4) Eine Teilmenge einer leeren Menge kann immer wieder nur selbst eine leere Menge sein.

X richtig ☐ O falsch

5) Ein Würfel wird dann als fair bezeichnet, wenn er bei (in etwa) der Hälfte aller Würfe eine gerade Augenzahl anzeigt.

X richtig ☐ O falsch

6) Das logische NICHT wird auch als Negation bezeichnet.

X richtig ☐ O falsch

7) Die Axiome von Kolmogoroff definieren die mathematischen Eigenschaften des Modells mit bzw. ohne Zurücklegen.

☐ O richtig **X falsch**

8) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Leverage-Effekt bezeichnet.

X richtig ☐ O falsch

9) Ein Zufallsexperiment ist die (beliebig häufige) Wiederholung eines Zufallsvorgangs unter kontrollierten, gleichbleibenden Rahmenbedingungen.

X richtig ☐ O falsch

10) Bei der einfachen linearen Regressionsanalyse darf nur die unabhängige (X) Variable die abhängige (Y) Variable beeinflussen – und nicht umgekehrt.

X richtig ☐ O falsch

Aufgabenteil VI: Lineare Regression (15 Punkte)

Verein	Ø Ballbesitz	Saisonsiege
FC Böhnhausen	32	0
Sportclub Wacker	38	1
Sportfreunde Harz	43	2
FC Fallingbostal	46	2
MSV Kleinhausen	48	4
FC Mühlenbach	51	3
SF Meierburg	55	5
Sportclub Burgdorf	57	5
MSV Großdendorf	61	8

Stellen Sie für die bereits bekannte Verteilung von durchschnittlichem Ballbesitz und Saisonsiegen ein lineares Regressionsmodell mit dem Ballbesitz (in Minuten) als unabhängiger Variablen (x) und den Saisonsiegen als abhängiger Variablen (y) auf (10 Punkte) und bestimmen und interpretieren Sie dessen Güte (5 Punkte).

Verein	x	y	x ²	y ²	(x*y)
FC Böhnhausen	32	0	1024	0	0
Sportclub Wacker	38	1	1444	1	38
Sportfreunde Harz	43	2	1849	4	86
FC Fallingbostal	46	2	2116	4	92
MSV Kleinhausen	48	4	2304	16	192
FC Mühlenbach	51	3	2601	9	153
SF Meierburg	55	5	3025	25	275
Sportclub Burgdorf	57	5	3249	25	285
MSV Großdendorf	61	8	3721	64	488
Summe	431	30	21333	148	1609
Durchschnitt	47,89	3,33	//	//	//

a) Aufstellung des Regressionsmodells

$$b = 173,74 / 691,93 = 0,2511$$

$$a = 3,33 - 0,2511 * 47,89 = -8,6952$$

$$y = -8,6952 + 0,2511 * x$$

b) Gütebestimmung

$$r = 173,74 / (26,3 * 6,94) = 0,951$$

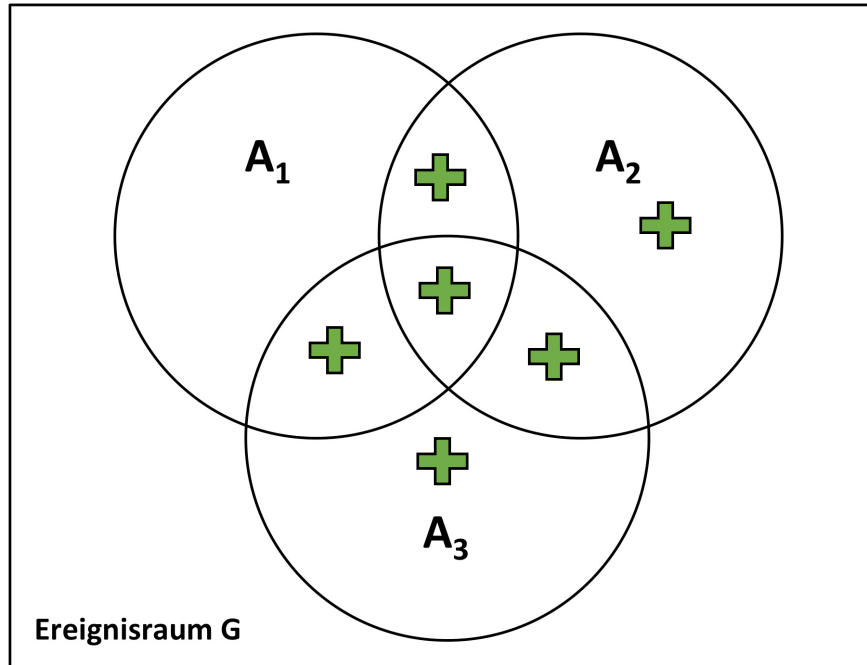
$$R^2 = 0,951^2 = 0,9044$$

Die Streuungsaufklärung des Modells ist sehr gut.

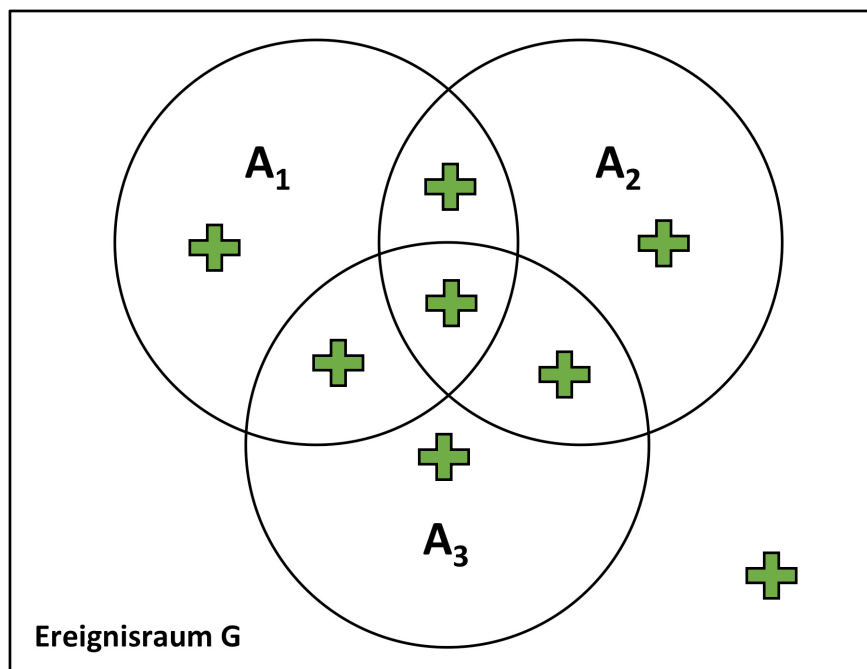
Aufgabenteil VII: Mengenlehre (5 Punkte)

Markieren Sie die jeweils bezeichneten Flächen deutlich sichtbar im Venn-Diagramm.

a) $(A_3 \cup A_2) \cup A_3$ (3 Punkte) -> A3 und A2 komplett



b) $(A_1 \cup \overline{A_1})$ (2 Punkte) -> komplettes Diagramm einschließlich Ereignisraum



Aufgabenteil VIII: Zufallsexperimente (5 Punkte)

Eine Fußballmannschaft geht in einem Spiel mit einer Wahrscheinlichkeit von 60% während der ersten Halbzeit in Führung. Startet diese Mannschaft mit einer Führung in die zweite Halbzeit, so ist sie besonders motiviert und gewinnt das Spiel mit einer Wahrscheinlichkeit von 70%. Befindet sie sich zu Beginn der zweiten Halbzeit dagegen im Rückstand, sorgt dies für zusätzliche Nervosität und die Wahrscheinlichkeit eines Sieges sinkt auf 50%.

- a) Mit welcher Wahrscheinlichkeit gewinnt diese Mannschaft ihr nächstes Spiel? (3 Punkte)
b) Kurz vor dem nächsten Spiel verletzt sich der beste Stürmer dieser Mannschaft, wodurch sich die Wahrscheinlichkeit sowohl einer Führung während der ersten Halbzeit als auch eines Sieges in der zweiten Halbzeit jeweils halbieren. Mit welcher Wahrscheinlichkeit wird sie nun gewinnen? (2 Punkte)

a) Unverletzter Stürmer

F = Führung
R = Rückstand
S = Sieg
N = Niederlage

Folgende Spielverläufe sind denkbar:

F-S
F-N
R-S
R-N

Von diesen Spielverläufen enden zwei mit einem Sieg:

$F-S \rightarrow 0,6 \cdot 0,7 = 0,42$
 $R-S \rightarrow 0,4 \cdot 0,5 = 0,20$
 $0,42 + 0,20 = 0,62 \rightarrow 62\% \text{ Gewinnwahrscheinlichkeit}$

b) Verletzter Stürmer

Nach wie vor gibt es zwei Spielverläufe, die zu einem Sieg führen:

$F-S \rightarrow 0,3 \cdot 0,35 = 0,105$
 $R-S \rightarrow 0,7 \cdot 0,25 = 0,175$
 $0,105 + 0,175 = 0,28 \rightarrow 28\% \text{ Gewinnwahrscheinlichkeit}$

Aufgabenteil IX: Kombinatorik (5 Punkte)

In der Harzer Fußball-Kreisliga spielen derzeit 9 Mannschaften. In einem Wettbewerb sollen alle Mannschaften im Spielmodus "jeder gegen jeden" gegeneinander antreten.

Wie viele Spiele müssen insgesamt stattfinden?



Es handelt sich um eine Kombination (die Reihenfolge des Ziehens spielt keine Rolle – das Spiel Mannschaft A gegen Mannschaft B ist identisch mit dem Spiel Mannschaft B gegen Mannschaft A) im Modell ohne Zurücklegen (keine Mannschaft kann im gleichen Pairing doppelt gezogen werden – sonst müsste sie gegen sich selbst antreten). Zu verwenden ist also die Formel:

$$n! / (k! * (n-k)!)$$

$$n = 9$$

$$k = 2$$

$$9! / (2! * 7!) = 362880 / 10080 = 36$$

Es müssen insgesamt 36 Spiele stattfinden, bevor jede Mannschaft einmal gegen jede Mannschaft gespielt hat.

Aufgabenteil X: Satz von Bayes (10 Punkte)

In einem Fußballstadion kommt es regelmäßig zu Problemen mit gewalttätigen Fans. Diese haben zwar bereits ein Hausverbot erhalten, tauchen aber dennoch regelmäßig bei Spielen auf. Eine neue Video-Überwachungssoftware soll nun zukünftig dabei helfen, diesen Personenkreis aus dem Stadion fernzuhalten. Ein Kamerasystem scannt die Gesichter aller Personen, die das Gelände betreten und leitet diese an ein KI-Tool weiter, welches die Scans mit Fotos der Personen vergleicht, die bereits ein Hausverbot erhalten haben. Dabei kann das Tool eine gesuchte Person mit einer Genauigkeit von 93% korrekt identifizieren, wenn sie eine Kamera passiert. Leider ordnet es auch in 2% aller Fälle eine nicht gesuchte Person fälschlicherweise als gesucht ein.

Zu einem bedeutenden Spiel werden rund 5.000 Besucherinnen und Besucher erwartet. Auf der Liste der Gewalttäter mit Hausverbot stehen derzeit 80 Namen und es ist davon auszugehen, dass etwa ein Viertel dieser Personen versuchen wird, unerkannt ins Stadion zu gelangen.

Das KI-Tool meldet, dass eine gesuchte Person erkannt wurde. Mit welcher Wahrscheinlichkeit wird eine Personenkontrolle nun zu dem Ergebnis führen, dass es sich auch tatsächlich um eine Person mit Hausverbot handelt?



5.000 Besucher

4.980 Fußballfans

20 Hooligans

20 Hooligans -> 93% korrekte Erkennungen -> 18,6 true positives

4.980 Fußballfans -> 2% falsche Erkennungen -> 99,6 false positives

$$18,6 / 118,2 = 0,1574$$

Die reale Trefferquote des Systems liegt bei 15,74%.