

**LÖSUNGEN – PROBEKLAUSUR STATISTIK**

Berufsbegleitender Studiengang Betriebswirtschaftslehre

Sommersemester 2016

Christian Reinboth

**Aufgabenteil I: Theorie (5 Punkte)**

Sind die nachfolgenden Aussagen richtig oder falsch? (0,5 Punkte pro korrekter Beantwortung)

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

richtig                       falsch

2) Der Quartilkoeffizient der Schiefe ist unempfindlich gegenüber Ausreißern.

richtig                       falsch

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

richtig                       falsch

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

richtig                       falsch

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

richtig                       falsch

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

richtig                       falsch

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

richtig                       falsch

8) Die Verzerrung einer linearen Regressionsfunktion durch einen Ausreißer wird als Outlier-Effekt bezeichnet.

richtig                       falsch

9) Das arithmetische Mittel kann nur für metrisch skalierte Daten gebildet werden.

richtig                       falsch

10) Zwei Teilgesamtheiten einer Grundgesamtheit können sich niemals überschneiden.

richtig                       falsch

**Aufgabenteil II: Grafische Darstellungsformen (10 Punkte)**

Eine Erhebung unter 20 Studierenden zu ihren Lebenshaltungskosten erbrachte folgendes Ergebnis:

324 EUR	131 EUR
227 EUR	193 EUR
123 EUR	167 EUR
412 EUR	254 EUR
201 EUR	227 EUR
212 EUR	193 EUR
321 EUR	321 EUR
372 EUR	372 EUR
193 EUR	193 EUR
227 EUR	212 EUR

Zeichnen Sie ein Stem-and-Leaf-Diagramm...

- a)... mit einer Stammbreite von 100. (5 Punkte)
- b)... mit einer Stammbreite von 10. (5 Punkte)

Geordnete Verteilung:

123; 131; 167; 193; 193; 193; 193  
 201; 212; 212; 227; 227; 254  
 321; 321; 372; 372;  
 412

a) Stammbreite von 100:

```

1 | 2 3 6 9 9 9 9
2 | 0 1 1 2 2 5
3 | 2 2 7 7
4 | 1
    
```

Jedes Blatt 1 Fall  
 Keine Ausreißer

b) Stammbreite von 10:

```

                20 | 1          30 |          40 |
                21 | 2 2       31 |          41 | 2
12 | 3          22 | 7 7       32 | 1 1
13 | 1          23 |
14 |           24 |          34 |
15 |           25 | 4         35 |
16 | 7          26 |          36 |
17 |           27 |          37 | 2 2
18 |           28 |          38 |
19 | 3 3 3 3    29 |          39 |
    
```

Jedes Blatt 1 Fall  
 Keine Ausreißer

**Aufgabenteil III: Verteilungsparameter (15 Punkte)**

Auf dem Campus der Hochschule Harz wurden 10 willkürlich ausgewählte Personen nach ihrem monatlich verfügbaren Nettoeinkommen befragt. Dies führte zu folgenden Ergebnissen:

Nr.	Nettoeinkommen (in EUR)
1	1.330
2	1.340
3	1.230
4	2.110
5	1.050
6	2.460
7	1.730
8	1.620
9	1.390
10	2.310

Maße der zentralen Tendenz:

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

Streuungsmaße / Dispersionsparameter:

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

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie den Quartilkoeffizienten der Schiefe. (3 Punkte)

Bestimmung des arithmetischen Mittels

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$= (1330 + 1340 + 1230 + 2110 + 1050 + 2460 + 1730 + 1620 + 1390 + 2310) / 10 = \underline{\underline{1657}}$$

Bestimmung des Median

Bildung einer geordneten Reihe:

1050; 1230; 1330; 1340; 1390; 1620; 1730; 2110; 2310; 2460

$0,5 * 10 = 5$  -> ganzzahliger Wert, daher:

$$x_p = \frac{1}{2}(x_{(k)} + x_{(k+1)})$$

$$= (1390 + 1620) / 2 = \underline{\underline{1505}}$$

Bestimmung der Standardabweichung

Schritt 1: Berechnung der Varianz über die Summen der quadrierten Differenzen der einzelnen Werte zum arithmetischen Mittel (1657)

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

368449  
 + 182329  
 + 106929  
 + 100489  
 + 71289  
 + 1369  
 + 5329  
 + 205209  
 + 426409  
 + 644809

= 2112610

2112610 / 10 = 211261

$$s = +\sqrt{s^2}$$

Schritt 2: Berechnung der Standardabweichung als positive Wurzel der Varianz: **459,6314**

Bestimmung des Interquartilsabstands

Bildung einer geordneten Reihe:

1050; 1230; 1330; 1340; 1390; 1620; 1730; 2110; 2310; 2460

$$x_p = x_{(k)}$$

Unteres Quartil:  $0,25 * 10 = 2,5$  -> nicht ganzzahlig, daher  $k = 3$ . Wert -> 1330

Oberes Quartil:  $0,75 * 10 = 7,5$  -> nicht ganzzahlig, daher  $k = 8$ . Wert -> 2110

Interquartilsabstand:  $2110 - 1330 = \mathbf{780}$

Bestimmung des Quartilkoeffizienten der Schiefe

Alle für die Formel benötigten Größen sind bereits bekannt:

Unteres Quartil: 1330

Mittleres Quartil: 1505

Oberes Quartil: 2110

Interquartilsabstand: 780

In die Formel für den Quartilkoeffizienten eingesetzt ergibt sich:

$$g_{0,25} = \frac{(x_{0,75} - x_{med}) - (x_{med} - x_{0,25})}{x_{0,75} - x_{0,25}}$$

$[(2110-1505)-(1505-1330)] / 780 = \mathbf{0,5513}$

**Aufgabenteil IV: Zusammenhangsmaße (10 Punkte)**

Ein Unternehmen führt seit vielen Jahren Assessment Center mit Bewerberinnen und Bewerbern durch, bei denen zwischen 0 und 100 Punkten (auf einer metrischen Skala) erzielt werden können. Nach jeweils fünf Jahren im Unternehmen durchläuft jede/r Mitarbeiter/in ein Performance Review, in dessen Rahmen ebenfalls zwischen 0 und 100 Punkte (ebenfalls auf einer metrischen Skala) erzielt werden können. In dieser und in der nächsten Aufgabe soll der mögliche Zusammenhang zwischen dem Ergebnis im Assessment Center und im späteren Performance Review untersucht werden.

Mitarbeiter/in	Bewertung im Assessment Center	Bewertung im Performance Review
1	78	65
2	82	78
3	85	75
4	96	92
5	87	62
6	95	88
7	79	85
8	97	88
9	83	61
10	75	86

Berechnen und interpretieren Sie...

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

Rangkorrelationskoeffizient nach Spearman

MA	x	rg(x)	y	rg(y)	d	d <sup>2</sup>
1	78	2	65	3	-1	1
2	82	4	78	5	-1	1
3	85	6	75	4	2	4
4	96	9	92	10	-1	1
5	87	7	62	2	5	25
6	95	8	88	8,5	-0,5	0,25
7	79	3	85	6	-3	9
8	97	10	88	8,5	1,5	2,25
9	83	5	61	1	4	16
10	75	1	86	7	-6	36

Summe aller d<sup>2</sup>: 95,5

Eingesetzt in die Formel:

$$rho = 1 - \frac{6 * \sum d_i^2}{(n^2 - 1) * n}$$

$$1 - [(6 * 95,5) / (99 * 10)] = 1 - 0,5788 = \underline{\underline{0,4212}}$$

Konkordanzkoeffizient nach Kendall

MA	X	rg(x)	y	rg(y)	K	D
1	75	1	86	7	3	6
2	78	2	65	3	6	2
3	79	3	85	6	3	4
4	82	4	78	5	3	3
5	83	5	61	1	5	0
6	85	6	75	4	3	1
7	87	7	62	2	3	0
8	95	8	88	8,5	1	0
9	96	9	92	10	0	1
10	97	10	88	8,5	//	//

Summe aller konkordanten Paare (K): 27

Summe aller diskordanten Paare (D): 17

Eingesetzt in die Formel:

$$\tau = \frac{2 * (K - D)}{n * (n - 1)}$$

$$(2 * 10) / (10 * 9) = 20 / 90 = \underline{\underline{0,2222}}$$

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

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

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

Diese Fläche lässt sich am einfachsten dadurch bestimmen, dass man sich ins Gedächtnis ruft, dass eine Negation (Linie über dem Term) eine Invertierung bedeutet, d.h. es ist die komplette Fläche zu markieren, die durch den Term unter der Negation nicht beschrieben wird. In diesem Fall also die komplette Fläche (einschließlich des Ereignisraums G) außer der Schnittfläche von A1, A2 und A3.

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

Dieser Term lässt sich mit Hilfe des Distributivgesetzes stark vereinfachen. Zu markieren sind hier die Schnittflächen von A1 und A3 sowie von A2 und A3.

**Aufgabenteil VI: Analyse mehrstufiger Zufallsexperimente (10 Punkte)**

In einem Unternehmen werden Bauteile auf drei unabhängig voneinander arbeitenden Spritzgussmaschinen (A, B und C) hergestellt, wobei jede dieser Maschinen im Rahmen einer Schicht mit einer vorab bekannten Wahrscheinlichkeit ausfällt:

- Ausfallwahrscheinlichkeit von Maschine A: 0,10
- Ausfallwahrscheinlichkeit von Maschine B: 0,05
- Ausfallwahrscheinlichkeit von Maschine C: 0,01

Wie groß ist die Wahrscheinlichkeit dafür, dass in einer zufällig ausgewählten Schicht...

- a) ... keine der drei Maschinen ausfällt? (5 Punkte)
- b) ... mindestens eine der drei Maschinen ausfällt? (5 Punkte)

#### Wahrscheinlichkeit für den Ausfall keiner Maschine

Aus den Ausfallwahrscheinlichkeiten für die drei Maschinen

Ausfallwahrscheinlichkeit von Maschine A: 0,10

Ausfallwahrscheinlichkeit von Maschine B: 0,05

Ausfallwahrscheinlichkeit von Maschine C: 0,01

sollten der Einfachheit halber die Nichtausfallwahrscheinlichkeiten bestimmt werden:

Nichtausfallwahrscheinlichkeit von Maschine A: 0,90

Nichtausfallwahrscheinlichkeit von Maschine B: 0,95

Nichtausfallwahrscheinlichkeit von Maschine C: 0,99

Die Wahrscheinlichkeit dafür, dass während einer Schicht keine der drei Maschinen ausfällt (lediglich ein einziger Pfad im Pfaddiagramm) lässt sich anschließend über den Multiplikationssatz für disjunkte Ereignisse bestimmen, d.h.

$$P(\text{kein Ausfall}) = 0,90 * 0,95 * 0,99 = \mathbf{0,84645}$$

#### Wahrscheinlichkeit für den Ausfall mindestens einer Maschine

Der Ausfall mindestens einer Maschine (d.h. von einer, zwei oder drei Maschinen) stellt das Gegenereignis zum zuvor berechneten Ausfall keiner Maschine dar (d.h. alle anderen Pfade im Pfaddiagramm). Gemäß dem zweiten Axiom von Kolmogoroff (alle Pfade müssen gemeinsam immer 1 ergeben), lässt sich daher an dieser Stelle ganz einfach rechnen:

$$P(\text{mindestens ein Ausfall}) = 1 - 0,84645 = \mathbf{0,15355}$$

### **Aufgabenteil VII: Variationen und Kombinationen (15 Punkte)**

Ein Koffer ist mit einem dreistelligen Zahlenschloss gesichert, wobei jede Stelle auf die Ziffern 0 bis 9 eingestellt werden kann und sich die Ziffern wiederholen dürfen.

a) Wie viele potentiell korrekte Ziffernkombinationen gibt es, wenn...

- a1) ...über die korrekte Ziffernkombination nichts bekannt ist? (5 Punkte)
- a2) ...bekannt ist, dass die korrekte Ziffernkombination nur aus Ziffern größer als 5 besteht? (5 Punkte)

b) Das Zahlenschloss sperrt sich bei einer falschen Eingabe. Wie groß ist die Wahrscheinlichkeit, im Fall a1) (in dem über die Zusammensetzung der korrekten Ziffernkombination nichts bekannt ist) durch das Ausprobieren einer rein zufällig gewählten Ziffernkombinationen das Schloss beim ersten Versuch öffnen zu können? (5 Punkte)

Anzahl der Kombinationen bei fehlenden Informationen

Es handelt sich um eine Variation (bei einer PIN spielt die Reihenfolge der Ziffern eine Rolle) mit Zurücklegen (Ziffern können mehrfach auftreten). Wenn über die korrekte Kombination nichts bekannt ist, kommen für jede Stelle der PIN alle 10 Ziffern in Frage, daher gilt:

$$10 * 10 * 10 = \underline{1000}$$

Anzahl der Kombinationen bei Ziffern > 5

Wenn bekannt ist, dass die Kombination nur aus Ziffern größer 5 (6, 7, 8, 9) besteht, berechnet sich die Zahl der möglichen Kombinationen wie folgt:

$$4 * 4 * 4 = \underline{64}$$

Wahrscheinlichkeit für einen Volltreffer beim ersten Versuch

Hier ist die Formel für die klassische Wahrscheinlichkeitsdefinition nach Pierre de Laplace zu beachten:

$$P(\text{Volltreffer}) = 1 / 1000 = \underline{0,001}$$

**Aufgabenteil VIII: Berechnung bedingter Wahrscheinlichkeiten (10 Punkte)**

Eine Hochschule prüft alle eingereichten Bachelor-Arbeiten mit einer eigens entwickelten Software auf Plagiate. Diese werden von der Software mit einer Sicherheit von 95% korrekt erkannt. Pro Semester reichen 800 Studierende an dieser Hochschule Bachelor-Arbeiten zur Kontrolle ein, wobei davon auszugehen ist, dass in 3% der eingereichten Arbeiten Plagiate enthalten sind.

Wie groß ist die Wahrscheinlichkeit dafür, dass eine durch die Software als plagiatsverdächtig identifizierte Arbeit auch tatsächlich ein Plagiat enthält?

Diese Aufgabe ist über den Satz von Bayes zu lösen, was nachfolgend anhand der absoluten Zahlen demonstriert werden soll.

Von 800 eingereichten Bachelor-Arbeiten enthalten...

... 776 keine Plagiate (97%)

... 24 Plagiate (3%)

Von 776 Arbeiten ohne Plagiate...

... werden 737,2 korrekterweise als sauber klassifiziert (95%)

... werden 38,8 fälschlicherweise als unsauber klassifiziert (5%)

Von 24 Arbeiten mit Plagiaten...

... werden 22,8 korrekterweise als unsauber klassifiziert (95%)

... werden 1,2 fälschlicherweise als sauber klassifiziert (5%)



Insgesamt werden also  $38,8 + 22,8 = 61,6$  Arbeiten als Plagiate eingestuft. Von diesen 61,6 Arbeiten sind 22,8 Arbeiten wirklich Plagiate. Die Wahrscheinlichkeit dafür, dass eine als plagiatsverdächtig eingestufte Arbeit auch wirklich Plagiate enthält liegt somit bei **37,01%**.

### Aufgabenteil IX: Chi-Quadrat-Unabhängigkeitstest (15 Punkte)

Bei einem Einstufungstest für 30 FH-Erstsemester müssen zum Bestehen mindestens 50% der gestellten Fragen korrekt beantwortet werden. Rückblickend soll überprüft werden, ob zwischen der Zugangsqualifikation der Probanden und dem Abschneiden im Test ein Zusammenhang besteht.

	Bestanden	Nicht bestanden
Realschulabschluss	7	5
Fachabitur	6	3
Abitur	6	3

Führen Sie einen Chi-Quadrat-Test durch (10 Punkte) und interpretieren Sie diesen (5 Punkte).

Schritt 1: Berechnung der Randsummen:

	Bestanden	Nicht bestanden	Rand
Realschulabschluss	7	5	12
Fachabitur	6	3	9
Abitur	6	3	9
Rand	19	11	30

Schritt 2: Berechnung der zu erwartenden Verteilung bei völliger Unabhängigkeit:

	Bestanden	Nicht bestanden	Rand
Realschulabschluss	7,6	4,4	12
Fachabitur	5,7	3,3	9
Abitur	5,7	3,3	9
Rand	19	11	30

Schritt 3: Summierung der quadrierten Differenzen geeilt durch die zu erwartenden Werte

$$(7-7,6)^2/7,6 = 0,0474$$

$$(6-5,7)^2/5,7 = 0,0158$$

$$(6-5,7)^2/5,7 = 0,0158$$

$$(5-4,4)^2/4,4 = 0,0818$$

$$(3-3,3)^2/3,3 = 0,0273$$

$$(3-3,3)^2/3,3 = 0,0273$$

Summe: 0,2157

Es ergibt sich ein Chi-Quadrat-Wert von **0,2157**, welcher den kritischen Wert von 3,814 nicht überschreitet. Die Nullhypothese, nach der die beiden Variablen als unabhängig voneinander einzustufen sind, kann demzufolge **nicht verworfen** werden, d.h. ein Zusammenhang ist nicht wahrscheinlich.