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.

O richtig **X** falsch

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

X richtig O falsch

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

O richtig **X** falsch

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

O richtig **X** falsch

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

X richtig O falsch

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

X richtig O falsch

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

X richtig O falsch

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

X richtig O falsch

9) Der Quartilskoeffizient der Schiefe ist unempfindlich gegenüber Ausreißern.

X richtig O falsch

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

O richtig **X** 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.

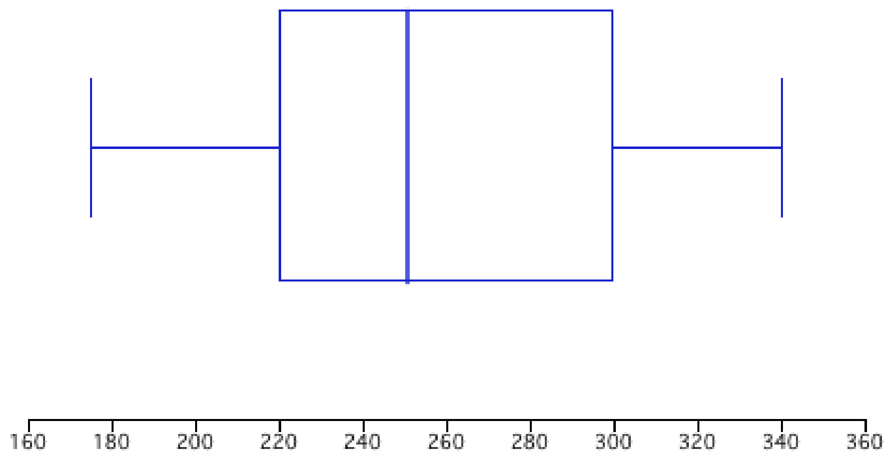
(n * p) = (15 * 0,25) = 3,75 -> nicht ganzzahlig -> k = 4 -> $p_{0,25}$ = 220
(n * p) = (15 * 0,50) = 7,50 -> nicht ganzzahlig -> k = 8 -> $p_{0,25}$ = 250
(n * p) = (15 * 0,75) = 11,25 -> nicht ganzzahlig -> k = 12 -> $p_{0,75}$ = 300

IQR = 300 – 220 = 80

1,5 IQR = 120
3,0 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 -> 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 -> 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 Mathematikklausur ergab sich die folgende Notenverteilung:

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

<u>Maße der zentralen Tendenz:</u>

- 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 -> ganzzahlig -> Mittel des 15. und 16. Werts = <u>2,5</u>

Modus: Der Modus beträgt <u>2</u>.

<u>Streuungsmaße / Dispersionsparameter:</u>

- 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 -> 8. Wert = 2
Oberes Quartil: (n * p) = 30 * 0,75 = 22,5 -> 23. Wert = 4
IQR = 4 – 2 = 2

Verteilungsmaße / Schiefe und Wölbung:

- Bestimmen Sie den Quartilskoeffizienten 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	x²	y²	(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			

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

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

Es existiert eine starke negative 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		

$b = (69944 – 10 * 10{,}75 * 666{,}9) / (1176{,}25 – 10 * 10{,}75^2) = -84{,}7394$
$a = 666{,}9 - -84{,}7394 * 10{,}75 = 1577{,}8485$

$y = 1577{,}85 – 84{,}74 * x$

Berechnung des Gütemaßes R^2:

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

Möglichkeit 2: Berechnung über Formel und Hilfstabelle

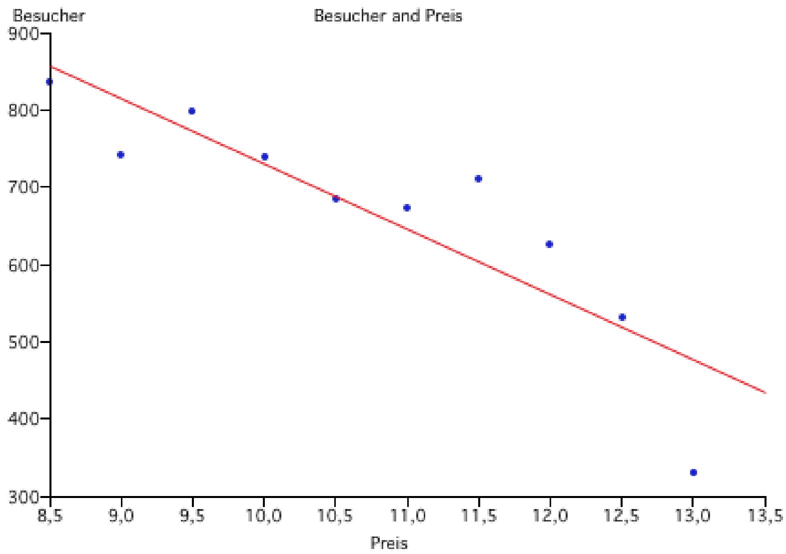
x	y	y*	$(y^*-y_m)^2$	$(y-y^*)^2$
8,50	836	857,56	36351,24	464,83
9,00	741	815,19	21989,92	5504,16
9,50	798	772,82	11219,05	634,03
10,00	740	730,45	4038,60	91,20
10,50	685	688,08	448,59	9,49
11,00	673	645,71	449,02	744,74
11,50	710	603,34	4039,87	11376,36
12,00	625	560,97	11221,16	4099,84
12,50	531	518,60	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 * x.

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

R2 = 148105,39 / 192567,01 = 0,7691 = 0,77



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