

Klausur zu Methoden der Statistik I (mit Kurzlösung)

Wintersemester 2018/19

Aufgabe 1

Bei den Bayerischen Landtagswahlen 2018 erlangten die Parteien folgende Sitze:

Partei	Anzahl der Sitze
CSU	85
Grüne	38
FW	27
AfD	22
SPD	22
FDP	11

- Bestimmen Sie Merkmalsträger, Merkmalstyp und Skalenniveau.
- Stellen Sie die relativen Häufigkeiten der Sitzverteilung tabellarisch und grafisch dar.
- Inwiefern ändert sich die grafische Darstellung aus Teilaufgabe b), wenn Sie statt der relativen die absoluten Häufigkeiten abtragen? [Hinweis: Zeichnung nicht notwendig.]
- Bestimmen und interpretieren Sie die normierte Entropie der Verteilung.

Aufgabe 2

Die Lebenszufriedenheit kann von vielen Faktoren abhängen. Eine davon ist sicher die Note in der Klausur Methoden der Statistik I. In der folgenden Tabelle sind von 7 Studierenden die Note in der Klausur Methoden der Statistik I (U) sowie die Lebenszufriedenheit auf einer zehnstufigen Skala (V) von 1 sehr unzufrieden bis 10 sehr zufrieden abgetragen:

Studierende	Klausurnote U	Zufriedenheit V
ν_1	1,3	7
ν_2	2,7	5
ν_3	2,3	9
ν_4	1,0	10
ν_5	1,7	8
ν_6	3,7	3
ν_7	5,0	6

- Welche Skalenniveaus weisen die beiden Variablen auf?
- Berechnen Sie den Rangkorrelationskoeffizienten nach Spearman und interpretieren Sie das Ergebnis.
- Wie hoch ist in diesem Fall der Unterschied des Ergebnisses zwischen den Ihnen bekannten Formel-Varianten des Rangkorrelationskoeffizienten nach Spearman? [Hinweis: keine Rechnung!]
- Berechnen Sie das obere und untere Quartil sowie den Median für die Variable Note in der Klausur Methoden der Statistik I (U).
- Für die univariate Analyse des Merkmals Lebenszufriedenheit (V) wird vorgeschlagen, das arithmetische Mittel zu berechnen. Warum genau könnte das problematisch sein?

Aufgabe 3

Das Statistische Bundesamt errechnet jährlich den Verbraucherpreisindex, mit dessen Hilfe es eine Aussage über die Höhe der Inflation trifft.

Gegeben ist der Verbraucherpreisindex (VPI) für Deutschland mit 2010 als Basisjahr:

Jahr	2018	2017	2016	2015	2014	2013	2012	2011	2010
VPI	111,4	109,3	107,4	106,9	106,6	105,7	104,1	102,1	100

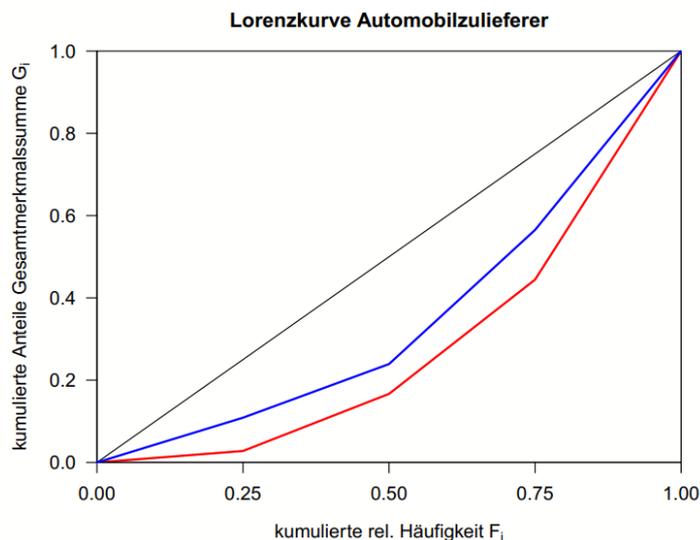
- a) Berechnen Sie die Wachstumsraten des Verbraucherpreisindex zum jeweiligen Vorjahr in Prozent.
- b) Um wie viel Prozent sind die Verbraucherpreise 2018 höher als 2010?
- c) Wie hoch ist die durchschnittliche Veränderung des Verbraucherpreises im gesamten Zeitraum?

Für das Jahr 2019 geht das Statistische Bundesamt von folgenden Preissteigerungen gegenüber dem Jahr 2018 aus:

Warenkorb	2018		2019	
	Preis in Euro	Menge	Preis in Euro	Menge
Nahrungsmittel	75	100	78	102
Wohnen	13	113	15	120
Bildung	29	35	25	34
Verkehr	25	76	26	85
Bekleidung	25	65	25	66

- d) Um wie viel Prozent wird sich das Preisniveau 2019 ändern? Verwenden Sie für die Berechnung den Ansatz nach Laspeyres.

Das Statistische Bundesamt arbeitet den Kartellbehörden zu. Die Grafik zeigt die Lorenzkurve aller ober- und unterfränkischen (rote und blaue Linie) Automobilzulieferer.



- e) Vergleichen Sie die beiden Lorenzkurven und treffen Sie eine Aussage darüber, in welcher Region eine höhere Konzentration ist. Begründen Sie Ihre Antwort kurz.

Aufgabe 4

Kleider machen Leute und Ausgaben für den Kauf einer Jeans können sehr unterschiedlich sein. In der Modebranche geht man schon länger länger davon aus, dass Männer teurere Jeans kaufen als Frauen. In der folgenden Tabelle sind die Kosten für den letzten Jeanskauf (J) in Euro sowie das Geschlecht des Käufers bzw. der Käuferin (S) verzeichnet:

Person	Geschlecht S	Kosten Jeans J
ν_1	männlich	150
ν_2	weiblich	70
ν_3	männlich	120
ν_4	männlich	90
ν_5	weiblich	90
ν_6	weiblich	50
ν_7	weiblich	20

- Welches ist die abhängige Variable und welches ist die unabhängige Variable im vorliegenden Fall?
- Berechnen Sie ein geeignetes Zusammenhangsmaß zwischen der Variable Kosten für den letzten Jeanskauf (J) und der Variable Geschlecht (S) und interpretieren Sie das Ergebnis.

Im Folgenden wird vermutet, dass die Jeanskaufausgaben in Euro (J) als Prädiktor (unabhängige Variable) für die abhängige Variable Friseur Ausgaben (F) dienen könnten. Verwenden Sie die berechneten Werte der Varianz für das Merkmal Kosten der letzten Jeans (J aus der Aufgabe b). [Hinweis: wenn Sie keine Ergebnisse aus Aufgabe b) haben, verwenden Sie für die Varianz des Merkmals Jeansausgaben $s_J^2 = 1500$]

Person	Kosten Jeans J	Friseur Ausgaben F
ν_1	150	9
ν_2	70	39
ν_3	120	17
ν_4	90	20
ν_5	90	22
ν_6	50	34
ν_7	20	40

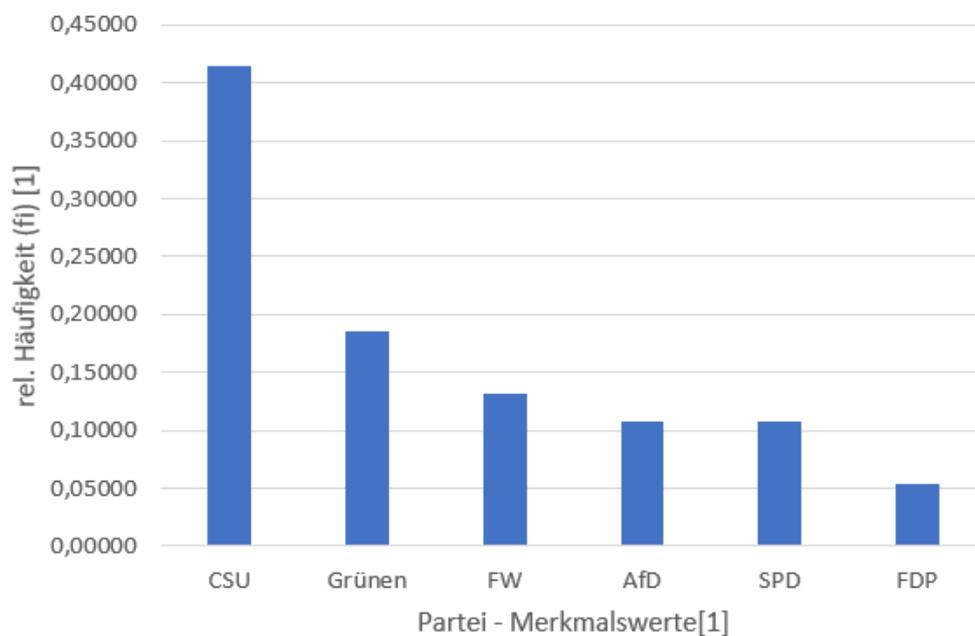
- Berechnen Sie die Parameter der ausgleichenden Regressionsgerade.
- Bestimmen Sie die Güte der Regression und interpretieren Sie das Bestimmtheitsmaß. Dazu ist Ihnen noch folgende Information gegeben: die Varianz des Merkmals Friseur Ausgaben beträgt $s_F^2 = 121,5532$.
- Berechnen Sie den Korrelationskoeffizienten nach Pearson auf Basis der Informationen aus Teilaufgabe c) und d).

Lösungen 1

- a) Merkmalsträger: Abgeordnete
Merkmalsträger: Qualitativ
Skalenniveau: Nominal

- b) Tabelle

Partei	Anteil	rel.Häufigkeit f_i
CSU	85	0,41463
Grüne	38	0,185373
FW	27	0,13171
AfD	22	0,10732
SPD	22	0,107327
FDP	11	0,053665
Σ	205	1,0



- c) keine Änderung; lediglich die Benennung der y-Achse ändert sich
d) Entropie

$$H_A = -\sum f_i \cdot \log_2(f_i) = -\frac{1}{\ln(2)} \sum f_i \cdot \ln(f_i) = 2,28013$$

Normierte Entropie

$$\log_2(K) = \log_2(6) = 2,58496$$

$$H_A^* = \frac{2,28013}{2,58496} = 0,88208$$

hohe Entropie = hohe Streuung

Lösung 2

a) Beide Variablen sind ordinal skaliert

b) Tabelle

U_v	$R_U(e_v)$	V_v	$R_v(e_v)$	d_v^2
1,0	1	10	1	0
1,3	2	7	4	4
1,7	3	8	3	0
2,3	4	9	2	4
2,7	5	5	6	1
3,7	6	3	7	1
5,0	7	6	5	4
Σ				14

$$C_{uv} = 1 - \frac{6 \cdot 14}{7 \cdot (7^2 - 1)} = 0,75$$

Sehr starker positiver Zusammenhang

c) kein Unterschied, da keine Bindung

d) Tabelle

U_v	n_v	f_v	F_v
1,0	1	0,1425	0,1425
1,3	1	0,1425	0,2858
1,7	1	0,1425	0,4287
2,3	1	0,1425	0,5716
2,7	1	0,1425	0,7145
3,7	1	0,1425	0,8574
5,0	1	0,1425	1,0000

Unteres Quantil $u_{0,25}=1,3$

Median $u_{0,5}=2,3$

Oberes Quantil $u_{0,75}=3,7$

e) Zufriedenheit hat keine Abstandsinformationen, weshalb das arithmetische Mittel kein geeignetes Maß der zentralen Tendenz für diese Variable ist.

Lösung 3

a) Tabelle

Jahr	2011	2012	2013	2014
$w_{t,t-1}$	2,1	1,95%=2%	1,5%	0,85%=0,9%
Jahr	2015	2016	2017	2018
$w_{t,t-1}$	0,3%	0,5%	1,8%	1,92%

b) $w_{18;10} = \frac{111,4}{100} - 1 = 0,114 = 11,4\%$

c) Tabelle

Jahr	2011	2012	2013	2014	2015	2016	2017	2018
$M_{t,t-1}$	1,021	1,0195	1,015	1,0085	1,003	1,0047	1,0177	1,0192

$$M_{t,t-1} = \prod_{t=18}^{t=10} M_t = 1,1137$$

$$\bar{w}_{18,10} = \sqrt[8]{1,1137} - 1 = 1,36\%$$

d) $p^2 = \frac{13.971}{13.509} = 1.03419$

Preissteigerung 3,4%

e) Lorenzkurve der unterfränkischen Automobilzulieferer liegt über der Lorenzkurve der oberfränkischen. Daher höhere Konzentration in Oberfranken.

Lösung 4

- a) Abhängige Variable: Kosten für Jeans J
 Unabhängige Variable: Geschlecht S

- b) Geeignetes Zusammenhangsmaß: η^2

$$\bar{x}_{\text{Frau}} = 57,5$$

$$\bar{x}_{\text{Mann}} = 120$$

$$\bar{x}_{\text{Gesamt}} = 84,2857$$

$$s_{\text{int}}^2 = \frac{1}{n} \sum_{i=1}^k s_i^2 \cdot n_i =$$

$$= \frac{1}{7} \cdot (1800 + 2675) = 639,2857$$

$$s_{\text{ext}}^2 = \frac{1}{7} \cdot [3826,5337 + 2865,8949] = 956,6327$$

$$\eta^2 = \frac{s_{\text{ext}}^2}{s_{\text{int}}^2 + s_{\text{ext}}^2} = 0,5994$$

Knapp 60% des abhängigen Merkmals Jeans wird durch das unabhängige Merkmal Geschlecht erklärt.

- c) $\beta_1 = \frac{s_{xy}}{s_x^2}$ $\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}$

x_v	F_v	
150	9	1345
70	39	2730
120	17	2040
90	20	1820
90	22	1980
50	34	1700
20	40	800
Σ		12395

$$\bar{F} = 25,8571$$

$$S_{JF} = \frac{12395}{7} - 84,2857 \cdot 25,8571 = -408,6695$$

$$\beta_1 = \frac{-408,6695}{1595,9184} = -0,2561$$

$$\beta_0 = 25,8571 + 0,2561 \cdot 84,2857 = 47,4427$$

- d) $S_F^2 = 750,1429 - 25,8571^2 = 121,5532$

$$v^2 = \frac{(-408,6655)^2}{1595,9184 \cdot 121,5532} = 0,8605$$

ca. 86% der Variable des abhängigen Merkmals wird durch die Regression erklärt

- e) $r_{XY} = -0,9279$

sehr stark negativer Zusammenhang