

Lösungen Probeklausur III

Statistische Verfahren in der Geographie

Till Straube <straube@geo.uni-frankfurt.de>

Institut für Humangeographie
Goethe-Universität Frankfurt

Aufgabe 1

Bestimmen Sie das Skalenniveau der folgenden Variablen. (5 Punkte)

Kürzen Sie ab: N = Nominalskala; O = Ordinalskala; I = Intervallskala; V = Verhältnisskala

Variable	Skalenniveau
a) Geburtsjahr einer Probandin (z.B. „1994“)	I
b) Selbstauskunft zum Drogenkonsum (kann die Werte „regelmäßig“, „selten“ und „nie“ annahmen)	O
c) Entgeltgruppe im öffentlichen Dienst (z.B. „E9“, „E10“, „E11“)	O
d) Baumart (z.B. „Buche“, „Linde“, „Birke“, ...)	N
e) Restaurant-Ratings auf einer Online-Plattform (1 Stern bis 5 Sterne)	I (aber O würde auch akzeptiert)

Sind die folgenden Aussagen richtig oder falsch? Kreuzen Sie das entsprechende Feld an. (5 Punkte)

Aussage	richtig	falsch
f) Die Alternativhypothese bei zweiseitigen Tests lautet $\mu_1 = \mu_2$.		×
g) Ordinaldaten können in Intervalldaten transformiert werden.		×
h) Je stärker die Werte der Variablen streuen, desto kleiner sollte die Stichprobe sein.		×
i) Der Verlauf der t -Verteilung hängt ab von der Anzahl der Freiheitsgrade.	×	
j) Die Standardabweichung ist definiert als die Quadratwurzel aus der Varianz.	×	

Geben Sie an, welches statistische Verfahren zur Beantwortung der unten stehenden Fragestellungen bzw. Untersuchungsabsichten angemessen ist. (5 Punkte)

Verwenden Sie dafür folgende Zahlen: 1 = z -Test bzw. 1-Stichproben- t -Test; 2 = 2-Stichproben- t -Test; 3 = F -Test; 4 = χ^2 -Test; 5 = Korrelation/Regression

Fragestellung	Testverfahren
k) Sie fragen sich, ob sich die durchschnittliche jährliche Sonnenscheindauer von Hamburg signifikant von der Sonnenscheindauer in München unterscheidet.	2
l) Variieren die durchschnittlichen Niederschlagsmengen in Deutschland stärker als die in Schweden?	3
m) Im Zusammenhang mit dem geplanten Ausbau eines Kohlekraftwerks wollen Sie untersuchen, ob der CO ₂ -Ausstoß an Ihrer Messstation zwischen 2000 und 2008 signifikant zugenommen hat.	2
n) Unterscheidet sich die durchschnittliche Lebenserwartung von Frauen* von der durchschnittlichen Lebenserwartung aller Personen in Deutschland?	1
o) Bestehen Brillenträger*innen eher die erste Fahrprüfung als Menschen, die keine Brille tragen?	4

Aufgabe 2

Eine stichprobenartige Bodenprobe ergibt folgende Werte für den Phosphorgehalt eines Grundstücks:

Phosphorgehalt (in mg/kg)
61
66
64
66
70
68
62
64
69
65

- a) Berechnen Sie das arithmetische Mittel und die Varianz der Stichprobe. (5 Punkte)

Arithmetisches Mittel:

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i}{n} \\ &= \frac{655}{10} \\ &= 65,5\end{aligned}$$

Varianz:

$$\begin{aligned}s^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \\ &= \frac{76,5}{9} \\ &= 8,5\end{aligned}$$

Das arithmetische Mittel beträgt $\bar{x} = 65,5$ und die Varianz $s^2 = 8,5$.

- b) Berechnen Sie den Variationskoeffizienten der Stichprobe. (5 Punkte)

Standardabweichung:

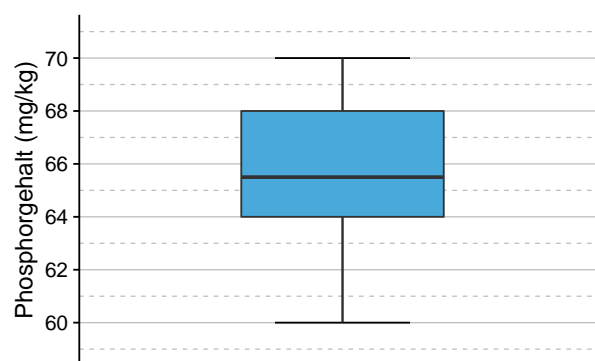
$$\begin{aligned}
 s &= \sqrt{s^2} \\
 &= \sqrt{8,5} \\
 &\approx 2,92
 \end{aligned}$$

Variationskoeffizient:

$$\begin{aligned}
 v &= \frac{s}{|\bar{x}|} \cdot 100\% \\
 &\approx \frac{2,92}{65,5} \cdot 100\% \\
 &\approx 4,46\%
 \end{aligned}$$

Der Variationskoeffizient der Stichprobe beträgt ca. 4,5%.

- c) Prüfen Sie, ob der folgende Boxplot die Stichprobe akkurat wiedergibt. Bestimmen Sie Median, Quartilsabstand sowie Spannweite und vergleichen Sie. (5 Punkte)



Sortierte Liste:

(i)	Phosphorgehalt (in mg/kg)
(1)	61
(2)	62
(3)	64
(4)	64
(5)	65
(6)	66
(7)	66
(8)	68
(9)	69
(10)	70

Bei Stichprobengröße $n = 10$ berechnet sich der Median durch:

$$\begin{aligned}Md &= \frac{x_{(5)} + x_{(6)}}{2} \\ &= \frac{65 + 66}{2} \\ &= 65,5\end{aligned}$$

Quartilsabstand:

$$\begin{aligned}Q_1 &= x_{(3)} \\ &= 64 \\ Q_3 &= x_{(8)} \\ &= 68 \\ IQR &= Q_3 - Q_1 \\ &= 68 - 64 \\ &= 4\end{aligned}$$

Spannweite:

$$\begin{aligned}R &= x_{(n)} - x_{(1)} \\ &= 70 - 61 \\ &= 9\end{aligned}$$

Der Minimalwert $x_{(1)} = 61$ ist nicht korrekt im Boxplot eingetragen, sonst sind die Werte korrekt.

Aufgabe 3

Eine Stadtverwaltung möchte die Mietpreisentwicklung für Gewerbeimmobilien in der innerstädtischen Einkaufspassage abschätzen. Sie folgt dabei der These: Entscheidend für die Höhe der monatlichen Mietpreise (in Euro pro Quadratmeter) sei die Entfernung zur nächstgelegenen Haltestelle des ÖPNV: Je näher an der Haltestation gelegen, desto höher der Mietpreis.

Für Aussagen über den angenommenen Zusammenhang stehen die Daten von sechs zufällig ausgewählten Gewerbeimmobilien in der Einkaufspassage zur Verfügung.

Immobilie	Entfernung (in Meter)	Quadratmeterpreis (in Euro pro Monat)
1	1141	30
2	850	49
3	862	40
4	1000	39
5	783	51
6	890	42

Die (gerundeten) arithmetischen Mittel betragen $\bar{x} = 921,0$ Meter und $\bar{y} = 41,8$ Euro, und die (gerundeten) Standardabweichungen liegen bei $s_x = 129,0$ Meter und $s_y = 7,6$ Euro.

- a) Wie groß ist der Zusammenhang zwischen der Entfernung zur nächstgelegenen Haltestelle und dem gemessenen Mietpreis pro Quadratmeter? Berechnen Sie den angemessenen Korrelationskoeffizienten und interpretieren Sie das Ergebnis. (5 Punkte)

Berechnung der Kovarianz mit Hilfe der Tabelle:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1141	30	220	-11,8	-2596,0
850	49	-71	7,2	-511,2
862	40	-59	-1,8	106,2
1000	39	79	-2,8	-221,2
783	51	-138	9,2	-1269,6
890	42	-31	0,2	-6,2

$$\begin{aligned}
 s_{xy} &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n - 1} \\
 &= \frac{-4498}{5} \\
 &= -899,6
 \end{aligned}$$

Berechnung Korrelationskoeffizient:

$$\begin{aligned}
 r &= \frac{s_{xy}}{s_x \cdot s_y} \\
 &\approx \frac{-899,6}{129 \cdot 7,6} \\
 &\approx -0,92
 \end{aligned}$$

Mit dem Korrelationskoeffizienten $r \approx -0,92$ konnte eine starke negative Korrelation festgestellt werden. Je kleiner die Entfernung zum ÖPNV, desto höher der Mietpreis.

-
- b) Die Stadtverwaltung hat unter Rückgriff auf diese Daten ein einfaches lineares Modell entwickelt, das eine Prognose der Mietpreise der Gewerbeimmobilien in Abhängigkeit von ihrer Entfernung zur nächstgelegenen Haltestelle des ÖPNV erlaubt. Wie lautet die Regressionsgleichung? (5 Punkte)
-

Steigung:

$$\begin{aligned} b &= \frac{s_{xy}}{s_x^2} \\ &\approx \frac{-899,6}{129,0^2} \\ &\approx -0,0541 \end{aligned}$$

Achsenabschnitt:

$$\begin{aligned} a &= \bar{y} - b \cdot \bar{x} \\ &\approx 41,8 + 0,0541 \cdot 921,0 \\ &\approx 91,63 \end{aligned}$$

Regressionsgerade:

$$\begin{aligned} y &= a + b \cdot x \\ y &= 91,63 - 0,0541 \cdot x \end{aligned}$$

Die Regressionsgleichung lautet: $y = 91,63 - 0,0541 \cdot x$

- c) Wie hoch fällt laut Modell der Mietpreis pro Quadratmeter für eine 500 Meter von der nächstgelegenen ÖPNV-Haltestelle entfernte Gewerbeimmobilie aus? (5 Punkte)
-

$$\begin{aligned} \hat{y} &= a + b \cdot x_i \\ &\approx 91,63 - 0,0541 \cdot 500 \\ &= 64,58 \end{aligned}$$

Laut Modell beträgt der Mietpreis 64,58 Euro.

Aufgabe 4

Sie führen eine Untersuchung zum Konsumverhalten von Studierenden mit und ohne Nebenjob in Hinblick auf Bio-Produkte durch. Eine Umfrage ergibt folgendes Ergebnis:

	Studierende mit Nebenjob	Studierende ohne Nebenjob	
Regelmäßiger Kauf von Bioprodukten	141	70	211
Kein regelmäßiger Kauf von Bioprodukten	253	149	402
	394	219	613

- a) Überprüfen Sie anhand dieser Daten, ob ein signifikanter positiver Zusammenhang zwischen der Ausübung eines Nebenjobs und dem regelmäßigen Konsum von Bio-Produkten besteht. Wählen Sie 0,05 als Signifikanzniveau. (10 Punkte)

1. Test wählen und Voraussetzungen prüfen

Überprüfung eines Zusammenhangs von nominalskalierten Variablen, deshalb: χ^2 -Unabhängigkeitstest.

2. Hypothesen formulieren

H_0 : Es besteht kein Zusammenhang zwischen Nebenjob und Bio-Konsum. H_1 : Es besteht ein positiver Zusammenhang zwischen Nebenjob und Bio-Konsum.
Achtung: H_1 ist eine *gerichtete* Alternativhypothese.

3. Signifikanzniveau entscheiden

$$\alpha = 0,05$$

4. Ablehnungsbereich ermitteln

Freiheitsgrade:

$$\begin{aligned} df &= (k - 1) \cdot (\ell - 1) \\ &= 1 \end{aligned}$$

Bei gerichteter Alternativhypothese lautet der Ablehnungsbereich

$$\chi^2 \geq \chi_{df;(1-2\cdot\alpha)}^2$$

$$\chi^2 \geq \chi_{1;90\%}^2$$

$$\chi^2 \geq 2,706$$

5. Prüfgröße berechnen

Berechnung in Tabelle anhand der Formeln

$$m_{ij} = \frac{n_{i \cdot} \cdot n_{\cdot j}}{n}$$

für Erwartungswerte (in Klammern) und

$$\frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

für χ^2 -Teilwerte (in blau):

	Studierende mit Nebenjob	Studierende ohne Nebenjob	
Regelmäßiger Kauf von Bioprodukten	141 (135,62) 0,213	70 (75,38) 0,384	211
Kein regelmäßiger Kauf von Bioprodukten	253 (258,38) 0,112	149 (143,62) 0,202	402
	394	219	613

Die beobachtete Häufigkeit der Kombination „Nebenjob“ und „Bio-Konsum“ ($n_{11} = 141$) übersteigt ihren Erwartungswert ($m_{11} =$). Die gerichtete Alternativhypothese ist daher zunächst haltbar.

Prüfgröße χ^2 :

$$\begin{aligned} \chi^2 &= \sum_{i=1}^k \sum_{j=1}^{\ell} \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \\ &\approx 0,213 + 0,384 + 0,112 + 0,202 \\ &= 0,911 \end{aligned}$$

6. Ergebnis interpretieren

Der Ablehnungsbereich wurde nicht erreicht, die Nullhypothese muss beibehalten werden.

Ein Zusammenhang zwischen Ausübung eines Nebenjobs und regelmäßigem Kauf von Bio-Produkten konnte nicht nachgewiesen werden ($\alpha = 0,05$).

b) Berechnen Sie ein Kennzahl, die aussagt, wie stark der Zusammenhang ausfällt. (5 Punkte)

Bei 2×2 -Tabellen muss hierfür der ϕ -Koeffizient berechnet werden.

$$\begin{aligned}\phi &= \sqrt{\frac{\chi^2}{n}} \\ &\approx \sqrt{\frac{0,911}{613}} \\ &\approx 0,04\end{aligned}$$

Der ϕ -Koeffizient liegt mit 0,04 sehr nah am Minimalwert 0, es ist also auch mit diesem Kennwert keine Korrelation festzustellen.
