Probeklausur IV

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)	Beförderungsklasse einer Fluggesellschaft („Economy“, „Economy Plus“, „Business“ oder „First“)	O
b)	Subjektive Bewertung der Inhalte eines Workshops („interessant“, „teilweise interessant“ oder „uninteressant“)	O
c)	Behälter von Erfrischungsgetränken (z.B. „Glasflasche“, „Dose“, „PET-Flasche“)	N
d)	Täglicher Stromverbrauch von Haushalten (in Kilowattstunden)	V
e)	ISBN von Büchern im Bestand einer Buchhandlung (z.B. „ISBN 978-3-642-12770-0“)	N

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

	Aussage	richtig	falsch
f)	Dichotome Variablen sind immer ordinalskaliert.		×
g)	Eine gerichtete Alternativhypothese lässt sich nur für Tests mit zwei Stichproben aufstellen.		×
h)	Je größer die Stichprobe, desto größer das Konfidenzintervall (bei gegebenem Konfidenzniveau).		×
i)	Die Auswahl z.B. jedes 100. Merkmalsträgers nennt man „systematische Stichprobe".	×	
j)	Ein Fehler 2. Art bedeutet, dass die Nullhypothese beibehalten wird, obwohl sie nicht stimmt.	×	

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-Unabhängigkeitstest
- 5 = Eindimensionaler χ^2-Test
- 6 = Korrelation/Regression

	Fragestellung	Testverfahren
k)	Variiert die Niederschlagsmenge in Niedersachsen signifikant stärker als in Mecklemburg-Vorpommern?	3
l)	Sind alle Geschmacksrichtungen von Kaugummis bei den Kund*innen gleich beliebt? Es liegen Verkaufszahlen vor.	5
m)	Sie möchten feststellen, ob es einen signifikanten Unterschied in der Varianz der Ticketpreise von zwei Reisebüros gibt.	3
n)	Es wird behauptet, Polizeikontrollen seien in jedem Stadtbezirk gleich wahrscheinlich. Sie möchten diese Behauptung anhand empirischer Zahlen überprüfen.	5
o)	Gibt es einen Zusammenhang zwischen Automarke und Stadt bzw. Landkreis der Zulassung?	4

Aufgabe 2

Der Pegelstand eines Schifffahrtskanals wird an sechs zufälligen Tagen im Monat überprüft. Für Mai und Juni diesen Jahres wurden folgende Werte erfasst:

Messung	Werte im Mai (in Metern)	Werte im Juni (in Metern)
1	11,35	10,68
2	14,19	11,46
3	12,70	10,64
4	11,58	12,05
5	13,93	12,10
6	14,21	10,45

a) Welcher der beiden Monate hatte im Durchschnitt den höheren Pegelstand? *(5 Punkte)*

Mai:

$$\begin{aligned}\bar{x} &= \frac{\sum\limits_{i=1}^{n} x_i}{n} \\ &= \frac{77{,}96}{6} \\ &\approx 12{,}99\end{aligned}$$

Juni:

$$\begin{aligned}\bar{x} &= \frac{\sum\limits_{i=1}^{n} x_i}{n} \\ &= \frac{67{,}38}{6} \\ &= 11{,}23\end{aligned}$$

Im Mai war der durchschnittliche Wasserstand mit 12,99 m höher als im Juni mit 11,23 m.

---

b) Ist der Median der Werte für Mai oder Juni größer? *(5 Punkte)*

Geordnete Listen:

(i)	$x_{(i)}$ für Mai	$x_{(i)}$ für Juni
1	11,35	10,45
2	11,58	10,64
3	12,70	10,68
4	13,93	11,46
5	14,19	12,05
6	14,21	12,10

Bestimmung des Medians bei gerader Stichprobengröße $n = 6$ für Mai:

$$\begin{aligned} Md &= \frac{x_{(3)} + x_{(4)}}{2} \\ &= \frac{12{,}70 + 13{,}93}{2} \\ &\approx 13{,}32 \end{aligned}$$

Für Juni:

$$\begin{aligned} Md &= \frac{x_{(3)} + x_{(4)}}{2} \\ &= \frac{10{,}68 + 11{,}46}{2} \\ &= 11{,}07 \end{aligned}$$

Der Median für Mai (13,32 m) ist größer als für Juni (11,07 m).

c) Welcher der beiden Monate weist den größeren Quartilsabstand der Pegelstände auf? *(5 Punkte)*

Berechnung Quartilsabstand für Mai:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= x_{(5)} - x_{(2)} \\ &= 14{,}19 - 11{,}58 \\ &= 2{,}61 \end{aligned}$$

Für Juni:

$$
\begin{aligned}
IQR &= Q_3 - Q_1 \\
&= x_{(5)} - x_{(2)} \\
&= 12{,}05 - 10{,}64 \\
&= 1{,}41
\end{aligned}
$$

Der Quartilsabstand der Mai-Stichprobe (2,61 m) ist größer als der der Juni-Stichprobe (1,41 m).

Aufgabe 3

Die wöchentlichen Besuchszahlen in einem Frankfurter Freibad sind während der Saison annähernd normalverteilt mit einem Mittelwert von 2098 und einer Standardabweichung von 279.

a) Mit welcher Wahrscheinlichkeit sind in einer zufälligen Woche mehr als 2400 Menschen im Freibad? *(5 Punkte)*

$$
\begin{aligned}
z_p &= \frac{x_p - \mu}{\sigma} \\
&= \frac{2400 - 2098}{279} \\
&\approx 1{,}08
\end{aligned}
$$

$$
\begin{aligned}
P(x \geq 2400) &\approx P(z \geq 1{,}08) \\
&= 1 - P(z < 1{,}08) \\
&\approx 1 - 0{,}8599 \\
&= 0{,}1401
\end{aligned}
$$

Die Marke von 2400 Besucher*innen wird mit einer Wahrscheinlichkeit von ca. 14% überschritten.

---

b) Welche Besuchszahl wird nur mit 1% Wahrscheinlichkeit unterschritten? *(5 Punkte)*

---

z-Perzentil:

$$
\begin{aligned}
z_{1\%} &= -z_{99\%} \\
&\approx -2{,}33
\end{aligned}
$$

Umgekehrte z-Transformation:

$$
\begin{aligned}
x_{1\%} &= \mu + z_{1\%} \cdot \sigma \\
&\approx 2098 - 2{,}33 \cdot 279 \\
&= 1447{,}93
\end{aligned}
$$

Die Marke von 1447 Besucher*innen wird nur zu 1% unterschritten.

---

c) In welchem Korridor liegen erwartungsgemäß die mittleren 90% der wöchentlichen Besuchszahlen? *(5 Punkte)*

---

z-Werte der Intervallgrenzen:

$$\begin{aligned} z_{(1-\alpha/2)} &= z_{95\%} \\ &\approx 1{,}65 \end{aligned}$$

$$\begin{aligned} z_{(\alpha/2)} &= z_{5\%} \\ &= -z_{95\%} \\ &\approx -1{,}65 \end{aligned}$$

Umgekehrte z-Transformation Obergrenze:

$$\begin{aligned} x_{95\%} &= \mu + z_{95\%} \cdot \sigma \\ &= 2098 + 1{,}65 \cdot 279 \\ &= 2558{,}35 \end{aligned}$$

$$\begin{aligned} x_{5\%} &= \mu + z_{5\%} \cdot \sigma \\ &= 2098 - 1{,}65 \cdot 279 \\ &= 1637{,}65 \end{aligned}$$

90% der Werte liegen zwischen 2558 und 1638 Besucher*innen.

Aufgabe 4

(Fortführung von Aufgabe 3)

Wegen Geschäftsaufgabe ist der Kiosk auf dem Freibadgelände seit vier Wochen geschlossen. Die Betreiber*innen des Freibads befürchten, dass sich dieser Umstand negativ auf die Besuchszahlen auswirkt. Für die betroffenen Wochen sind die folgenden Zahlen erhoben:

Kalenderwoche	Anzahl Besucher*innen
KW 20	1601
KW 21	2299
KW 22	1812
KW 23	1860

Prüfen Sie, ob die Besuchszahlen im angegebenen Zeitraum signifikant gesunken sind. Wählen Sie 5% als Signifikanzniveau. Gehen Sie weiterhin von einer normalverteilten Grundgesamtheit mit Mittelwert 2098 und Standardabweichung 279 aus.

---

1. Test wählen und Voraussetzungen prüfen

Es sollen die Mittelwerte einer Grundgesamtheit (Besuchszahlen insgesamt) mit einer Stichprobe (Besuchszahlen der letzten 4 Wochen) verglichen werden. Da zusätzlich σ bekannt ist, muss ein z-Test durchgeführt werden (sonst wäre es ein 1-Stichproben-t-Test).

2. Hypothesen formulieren

Gerichtete Alternativhypothese:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$

3. Signifikanzniveau entscheiden

$$\alpha = 0{,}05$$

4. Ablehnungsbereich bestimmen

$$\begin{aligned} z &\leq z_\alpha \\ z &\leq z_{5\%} \\ z &\leq -z_{95\%} \\ z &\leq -1{,}65 \end{aligned}$$

5. Prüfgröße berechnen

Arithmetisches Mittel $\bar{x}$:

$$\begin{aligned} \bar{x} &= \frac{\sum\limits_{i=1}^{n} x_i}{n} \\ &= \frac{7572}{4} \\ &= 1893 \end{aligned}$$

Prüfgröße z:

$$
\begin{aligned}
z &= \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma} \\
&= \sqrt{4} \cdot \frac{1893 - 2098}{279} \\
&\approx -1{,}47
\end{aligned}
$$

6. Ergebnis interpretieren

Der kritische Wert von -1,65 wurde nicht unterschritten, der Ablehnungsbereich nicht erreicht. Die Nullhypothese muss also beibehalten werden. Es konnte kein signifikanter Rückgang der Besuchszahlen festgestellt werden ($\alpha = 0{,}05$).
