

Statistikexamen

Zeit: 90 Minuten

Name: _____

Matr. Nummer: _____

Hinweise:

1. Zugelassene Hilfsmittel: Open-Book: Aufschriebe, Formelsammlung, Skript, Taschenrechner (keine gespeicherten Formeln etc.!), Notizen.
2. Jede Antwort muss hinreichend begründet werden. Antworten ohne Begründung ergeben 0 Punkte.
3. Unleserliche Ergebnisse werden nicht gewertet. Nutzen Sie bei weiterem Platzbedarf bitte auch die Rückseiten der Klausurblätter!
4. Die geschätzte Bearbeitungszeit (in Minuten) für eine Aufgabe entspricht der Punktzahl. Somit sind die Aufgaben insgesamt 90 Punkte wert.
5. **Viel Glück!!!**

Frage	Punkte	Erreichte Punkte
1	10	
2	20	
3	10	
4	10	
5	20	
6	20	
Gesamt	90	

Aufgabe 1: Deskriptive Statistik – Skalenniveau (10 Punkte)

Geben Sie das Messniveau der folgenden Daten sowie die Lagemaßen, welche berechnet werden können, an.

Daten	Messniveau	Lagemaße, die berechnet werden können
Tierrassen		
Umsätze		
Einkommensklassen		
Postleitzahlen		
Worte		

Lösung:

Daten	Messniveau	Lagemaße, die berechnet werden können
Tierrassen	Nominal	Modalwert
Umsätze	Verhältnis	Modalwert, Median, Mittelwert
Einkommensklassen	Ordinal	Modalwert, Median
Postleitzahlen	Nominal	Modalwert
Worte	Nominal	Modalwert

Aufgabe 2: Deskriptive Statistik – Maße (20 Punkte)

Betrachten Sie den folgenden Datensatz mit Umsätzen und Marketingausgaben von vier Firmen:

Firma	1	2	3	4
Umsatz (in Tausend €)	4050	5670	6489	3611
Marketingausgaben (in Tausend €)	759	500	876	256

Geben Sie jeweils mindestens einen Rechenschritt an.

- Berechnen Sie den durchschnittlichen Umsatz und die durchschnittlichen Marketingausgaben.
(Hinweis: Sie können Ihre Ergebnisse auf ganze Zahlen runden)
- Berechnen Sie die Kovarianz von Marketingausgaben und Umsatz.
- Interpretieren Sie die Kovarianz in Bezug auf den Kontext.
- Berechnen Sie den Korrelationskoeffizienten und interpretieren Sie den Wert des Koeffizienten. Liegt eine starke Korrelation vor?
(Standardabweichung Umsatz = 1172 T€, Standardabweichung Marketing = 240T€)
- Auf welche Werte ändern sich Kovarianz und Korrelationskoeffizient, wenn Sie Umsatz in Millionen € (statt Tausend €) messen? Begründen Sie Ihre Antworten kurz.
(Hinweis: Sie müssen Maße hierzu nicht noch einmal komplett neu berechnen. Überlegen Sie sich, wie sich die Einheiten auf die Maße auswirken.)

Lösung:

a)

$$\text{Mittelw. Umsatz} = \frac{4050 + 5670 + 6489 + 3611}{4} = 4955$$

$$\text{Mittelw. Marketing} = \frac{759 + 500 + 876 + 256}{4} \approx 598$$

b)

$$\begin{aligned} \text{Kovarianz}_{\text{Umsatz, Marketing}} &= \\ &= \frac{(4050 - 4955)(759 - 598) + (5670 - 4955)(500 - 598) + (6489 - 4955)(876 - 598) + (3611 - 4955)(256 - 598)}{4} \\ &\approx 167'581 \end{aligned}$$

- Kovarianz > 0. Umsatz und Marketing sind positiv korreliert – Mehr Marketingausgaben führen zu mehr Umsatz. Keine Aussage über Stärke möglich, da nicht normiert.

- d) *Korrelationskoeffizient* $= \frac{167 \cdot 581}{1172 \cdot 240} \approx 0.6$. Relativ starke Korrelation (> 0.5).
- e) *Korrelationskoeffizient* ändert sich nicht, da normiert. Kovarianz reduziert sich um den Faktor 1000.

Aufgabe 3: Wahrscheinlichkeitstheorie – Chebyshev Ungleichung (10 Punkte)

Eine Firma führt Eignungstests für Bewerber für Managementpositionen durch. Im Schnitt erreichen die Bewerber 75 (von 100) Punkten (Standardabweichung 10 Punkte). Die Art der Verteilung sei unbekannt. Begründen Sie Ihre Antworten und geben Sie Rechenschritte an.

- a) Wie viel Prozent der Bewerber erreichen mindestens 50 bis 100 Punkte?
- b) In welchem Punkteintervall liegen mindestens 50% der Bewerber (runden Sie die Intervallgrenzen auf ganze Punkte)?

Lösung:

a)

- $k = \frac{100-75}{10} = 2.5$
- $\text{Prozent} = 1 - \frac{1}{k^2} = 1 - \frac{1}{6.25} = 84\%$

b)

- $1 - \frac{1}{k^2} = 50\% \Leftrightarrow k = \sqrt{2}$
- Intervall: $75 \pm \sqrt{2} \cdot 10 \approx \text{von 61 Punkten bis 89 Punkten}$

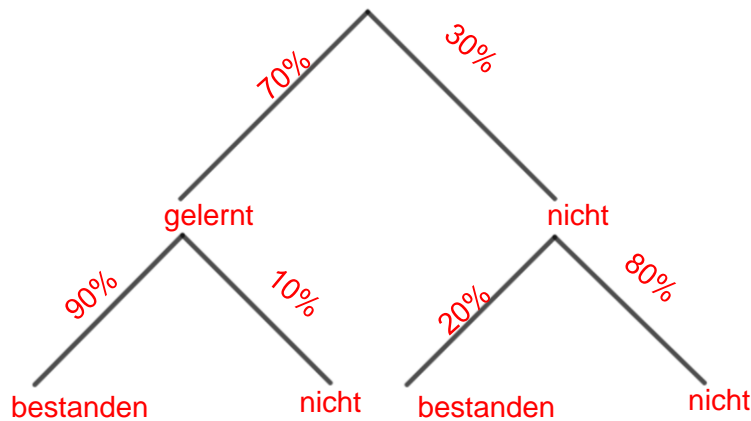
Aufgabe 4: Wahrscheinlichkeitstheorie – Bedingte Wahrscheinlichkeiten (10 Punkte)

70% der Studierenden lernen viel für Statistik (und 30% nicht). Von den 70% die viel lernen, bestehen 90% das Statistik-Examen (und 10% nicht). Von denjenigen die nicht viel lernen, bestehen nur 20% das Statistik-Examen (und 80% nicht).

- a) Stellen Sie diese Situation in einem Baumdiagramm dar.
- b) Berechnen Sie die (unbedingte) Wahrscheinlichkeit das Statistik-Examen zu bestehen.
- c) Einer Ihrer Kommilitonen behauptet, dass er nicht viel gelernt hat und trotzdem bestanden hat. Berechnen Sie die Wahrscheinlichkeit, dass er die Wahrheit sagt (d.h. berechnen Sie die bedingte Wahrscheinlichkeit, dass jemand wenig gelernt hat gegeben, dass er bestanden hat).

Lösung:

a)



b) Unbedingte Wahrscheinlichkeit:

$$P[\text{bestanden}] = 70\% \cdot 90\% + 30\% \cdot 20\% = 69\%$$

c) Bedingte Wahrscheinlichkeit:

$$\begin{aligned}
 P[\text{wenig gelernt} | \text{bestanden}] &= \frac{P[\text{bestanden} | \text{wenig gelernt}] \cdot P[\text{wenig gelernt}]}{P[\text{bestanden}]} \\
 &= \frac{20\% \cdot 30\%}{69\%} \approx 8.7\%
 \end{aligned}$$

Aufgabe 5: Inferenzstatistik – Schätzer und Hypothesentest (20 Punkte)

In einer Stichprobe mit 46 Studierenden der Hochschule Heilbronn ist der durchschnittliche Intelligenzquotient (IQ) 97 (langfristiger Durchschnitt: 102, Standardabweichung: 12).

- Bestimmen Sie den Punktschätzer für den durchschnittlichen IQ.
- Bestimmen Sie das 99%-Konfidenzintervall (den Intervallschätzer) für den durchschnittlichen IQ.

Nun möchten wir wissen, ob der durchschnittliche IQ der Studierenden im Vergleich zum langfristigen Durchschnitt gesunken ist und führen hierfür einen Hypothesentest durch.

- Schreiben Sie die Null- und Alternativhypothese des Hypothesentests in Bezug zum Aufgabenkontext auf.
- Handelt es sich um einen einseitigen oder zweiseitigen Test? Falls es sich um einen einseitigen Test handelt, ist dieser links- oder rechtsseitig? Begründen Sie Ihre Antworten.
- Handelt es sich um einen Einstichproben- oder Zweistichprobentest? Begründen Sie Ihre Antwort.
- Berechnen und interpretieren Sie den p-Wert. Liegt Evidenz für einen gesunkenen IQ vor?

Lösung:

a) Punktschätzer = Stichprobenmittelwert = 97

b)

- $99\% \rightarrow z_{0.5} = 2.575$
- $Standardfehler = \frac{Standardabweichung}{\sqrt{Stichprobengröße}} = \frac{12}{\sqrt{46}} \approx 1.77$
- $Error\ Bound = z_{0.5} \cdot Standardfehler = 4.56$
- $99\% - Konfidenzintervall = Punktschätzer \pm Error\ Bound:$
Intervall von ca. 92 bis ca. 102

c) Hypothesen:

$H_0: Durchschnitts - IQ \geq 102$

$H_a: Durchschnitts - IQ < 102$

d) Einseitig -> 'gesunken', Linksseitig -> 'gesunken' (Richtung zeigt nach links)

e) Einstichprobentest -> nur eine Gruppe (Studierende)

f) P-Wert

$$1. \ z - wert = \frac{x - \mu_0}{Standardfehler} = \frac{97 - 102}{1.77} = -2.82$$

2. Tabellenwert für $z = 2.82$: 0.4976

3. $P\text{-Wert} = 0.5 - 0.4976 = 0.24\%$

$P\text{-Wert} < 1\% \rightarrow$ Stark statistisch signifikant. Es liegt starke Evidenz vor, dass der IQ gesunken ist.

Aufgabe 6: Inferenzstatistik – Hypothesentest (20 Punkte)

Wir möchten wissen, ob Übergewicht Auswirkungen auf den Gesundheitszustand hat. Hierfür erheben wir eine Stichprobe mit 172 Personen und beobachten folgende Häufigkeiten:

Beobachtete Häufigkeiten

		Overweight?		Total
		Yes	No	
Health Condition	Good	41	43	84
	Bad	53	35	88
Total		94	78	172

Erwartete Häufigkeiten

		Overweight?		Total
		Yes	No	
Health Condition	Good	a	b	84
	Bad	c	d	88
Total		94	78	172

- Berechnen Sie die erwarteten Häufigkeiten (a,b,c und d). Geben Sie für jeden Wert einen Rechenschritt an.
- Schreiben Sie die Null- und Alternativhypothese des Chi2-Unabhängigkeitstest in Bezug auf den Aufgabenkontext auf.
- Berechnen Sie die Teststatistik des Chi2-Tests.
- Überprüfen Sie, ob der Chi2-Test auf einem 5%-Niveau ($\alpha = 5\%$) signifikant ist. Was können wir schlussfolgern?

Lösung:

g) Erwartete Häufigkeiten:

- $a = 84 * 94 / 172 \approx 45.91$
- $b = \frac{84 * 78}{172} = 38.1$
- $c = \frac{88 * 94}{172} = 48.09$
- $d = \frac{88 * 78}{172} = 39.9$

h) Hypothesen:

H0: Gesundheitszustand und Übergewicht sind stochastisch unabhängig.

Ha: Gesundheitszustand und Übergewicht sind stochastisch abhängig.

i) Teststatistik:

$$Chi2 = \frac{(41 - 45.91)^2}{45.91} + \frac{(38.1 - 43)^2}{38.1} + \frac{(48.09 - 53)^2}{48.09} + \frac{(39.9 - 35)^2}{39.9} \approx 2.26$$

j) Signifikanz:

- Kritischer Wert bei $df = 1$ und $\alpha = 5\%$: 3.84
- Also: $Chi2 = 2.26 < 3.84 = \text{kritischer Wert}$

Wir können die Nullhypothese nicht auf 5%-Niveau ablehnen. Keine Evidenz für einen Zusammenhang von Übergewicht und Gesundheitszustand.