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. Sofern nicht anders angegeben, runden Sie Dezimalzahlen auf vier und Prozentzahlen auf zwei Nachkommastellen.
6. **Viel Glück!!!**

Aufgabe	**Punkte**	**Erreichte Punkte**
1	10	
2	25	
3	15	
4	20	
5	20	
Gesamt	**90**	

Aufgabe 1: Deskriptive Statistik – Skalenniveau und Lagemaße (10 Punkte)

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

Daten	**Messniveau**	**Lagemaße, die berechnet werden können**
Social-Media-Kanäle		
Performance-Bewertung (gut, mittel, schlecht)		
Kontostand		
Zustimmungswerte (trifft zu, neutral, trifft nicht zu)		
Arbeitseinkommen		

Lösung:

Daten	**Messniveau**	**Lagemaße, die berechnet werden können**
Social-Media-Kanäle	Nominal	Modalwert
Performance-Bewertung (gut, mittel, schlecht)	Ordinal	Modalwert, Median
Kontostand	Intervall	Modalwert, Median, Mittelwert
Zustimmungswerte (trifft zu, neutral, trifft nicht zu)	Ordinal	Modalwert, Median
Arbeitseinkommen	Verhältnis	Modalwert, Median, Mittelwert

Aufgabe 2: Deskriptive Statistik – Lage- und Zusammenhangsmaße (25 Punkte)

Betrachten Sie den folgenden Datensatz mit Miethöhe und Wohnraum von fünf Wohnungen:

Wohnung	1	2	3	4	5
Miete (in €)	1220	540	2300	950	1200
Wohnraum (in m^2)	75	45	125	80	125

Geben Sie jeweils mindestens einen Rechenschritt an.

a) Berechnen Sie den Median und Modalwert der Wohnraumgrößen.

b) Berechnen Sie die durchschnittliche Miethöhe und Wohnraumgröße.

c) Berechnen Sie die Kovarianz von Miete und Wohnraum.

d) Welche der Wohnungen verringert die Kovarianz? Erklären Sie kurz warum dies so ist unter Bezugnahme auf die Mittelwerte.

Standardabweichungen: Miete: 583 €, Wohnraum: 31 m^2.

e) Berechnen Sie den Korrelationskoeffizienten auf 2 Nachkommastellen und interpretieren Sie den Wert des Koeffizienten. Ist die Korrelation stark? Begründen Sie Ihre Antworten.

f) Ändert sich der Korrelationskoeffizient, wenn Miete statt in € in tausend € gemessen wird? Begründen Sie Ihre Antwort.

Lösung:

a) Median: 80, Modalwert: 125

b)

$$Mittelw_{Miete} = 1242$$

$$Mittelw_{Wohnraum} = 90$$

c) Kovarianz = 14'080

d) Wohnung 5: Miete ist unterdurchschnittlich, Wohnraum ist überdurchschnittlich hoch.

e) Korrelationskoeffizient = 0,78 -> starke Korrelation: 61% der Unterschiede in der Miete können durch Wohnraum erklärt werden.

f) Nein, keine Änderung. Der Korrelationskoeffizient ist eine normierte Kovarianz und hängt nicht von den Einheiten ab.

Aufgabe 3: Wahrscheinlichkeitstheorie (15 Punkte)

Die Wahrscheinlichkeit die Klausur in

- Statistik (S) zu bestehen ist: P(S) = 78%.
- BWL (B) zu bestehen ist: P(B) = 85%.

Gegeben „Statistik bestehen" und „BWL bestehen" sind **unabhängige** Ereignisse.

Berechnen Sie

a) die Wahrscheinlichkeit gleichzeitig Statistik **und** BWL zu bestehen $P(S \cap B)$.

b) die Wahrscheinlichkeit Statistik **oder** BWL zu bestehen $P(S \cup B)$. Benutzen Sie hierfür Ihr Ergebnis aus Teil a).

Gegeben „Statistik bestehen" und „BWL bestehen" sind **abhängige** Ereignisse.

Die Wahrscheinlichkeit Statistik und BWL gleichzeitig zu bestehen sei $P(S \cap B) = 81\%$.

Berechnen Sie

c) die Wahrscheinlichkeit mindestens eine der Klausuren nicht zu bestehen.

d) die bedingte Wahrscheinlichkeiten P(S | B), d.h. die Wahrscheinlichkeit Statistik zu bestehen gegeben, dass BWL bestanden wurde.

Tipp: Benutzen Sie folgende Formel. Bedingte Wahrscheinlichkeit von E gegeben F:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Lösung:

a) $P(S \cap B) = P(S) \cdot P(B) = 66{,}3\%$

b) $P(S \cup B) = P(S) + P(B) - P(S \cap B) = 96.7\%$

c) $P(nicht\ S \cup nicht\ B) = 1 - P(S \cap B) = 19\%$

d) Bedingte Wahrscheinlichkeit:

$$P(S \mid B) = P(S \cap B)/P(S) = 81\%/85\% \approx 95{,}29\%$$

Aufgabe 4: Inferenzstatistik – Hypothesentest (20 Punkte)

Der langfristige Durchschnitt der Monatsmiete von Studierenden an der Hochschule Heilbronn ist 231 € (Standardabweichung: 56 €).

In einer Stichprobe mit 52 Studierenden ist die durchschnittliche Monatsmiete 245 €. Wir möchten wissen, ob Studierende inzwischen mehr Miete bezahlen als in der Vergangenheit und führen hierzu einen Hypothesentest durch.

a) Handelt es sich um einen Ein- oder Zweistichprobentest. Begründen Sie Ihre Antwort.

b) Ist der Test einseitig oder zweiseitig. Falls einseitig, ist der Test rechts- oder linksseitig. Begründen Sie Ihre Antworten.

c) Schreiben Sie die Null- und Alternativhypothese des Hypothesentests in Bezug zum Aufgabenkontext auf.

d) Berechnen Sie den Standardfehler auf 2 Nachkommastellen.

e) Berechnen Sie die Teststatistik und den p-Wert. Interpretieren Sie den p-Wert. Liegt Evidenz vor, dass die monatlichen Mietzahlungen angestiegen sind?

f) Wie verändert sich der Standardfehler, wenn die Stichprobengröße ansteigt? Begründen Sie Ihre Antwort.

Lösung:

a) Einstichprobentest <- nur eine Gruppe

b) Einseitig <- „mehr Miete bezahlen“; Rechtsseitig <- Ablehnungsbereich rechts

c) Hypothesen:

H0: $\mu_{Miete} \leq 231$

Ha: $\mu_{Miete} > 231$

d) $Standardfehler = \frac{\sigma}{\sqrt{n}} = \frac{56}{\sqrt{52}} \approx 7{,}77$

e) $Teststatistik = \frac{x - \mu_0}{Standardfehler} = \frac{245-231}{7{,}77} \approx$ 1,8 -> p-Wert = 0,5 – 0,4641 = 3,59%

p-Wert > 5% -> Evidenz für einen Mietanstieg

f) Standardfehler sinkt. Extremwerte rücken näher bzw. Wurzel (n) wird größer -> SE wird kleiner, da unter dem Bruch.

Aufgabe 5: Inferenzstatistik – Hypothesentest Chi2 (20 Punkte)

Wir möchten wissen, ob die Sportpräferenz abhängig ist vom Geschlecht. Hierfür erheben wir eine Stichprobe mit 181 Personen und beobachten folgende Häufigkeiten:

Beobachtete Häufigkeiten:

	Wandern	**Fußball**	**Joggen**	Gesamt
Mann	12	52	20	84
Frau	45	31	21	97
Gesamt	57	83	41	181

Erwartete Häufigkeiten

	Wandern	**Fußball**	**Joggen**
Mann	a	b	c
Frau	d	e	f

a) Schreiben Sie die Null- und Alternativhypothese des Chi2-Unabhängigkeitstest in Bezug auf den Aufgabenkontext auf.

b) Berechnen Sie die erwarteten Häufigkeiten (a,b,c,d, e und f) auf 2 Nachkommastellen. Geben Sie für jede Häufigkeit einen Rechenschritt an.

c) Berechnen Sie die Teststatistik des Chi2-Tests auf 2 Nachkommastellen.

d) Berechnen Sie die Freiheitsgrade des Chi2-Tests.

e) Überprüfen Sie, ob der Chi2-Test auf einem 5%-Niveau ($\alpha = 5\%$) signifikant ist. Was können wir schlussfolgern?

Lösung:

a) Hypothesen:

H0: Geschlecht und Sportpräferenz sind stochastisch unabhängig.

Ha: Geschlecht und Sportpräferenz sind stochastisch abhängig.

b) Erwartete Häufigkeiten:

	Wandern	**Fußball**	**Joggen**
Mann	26,45	38,52	19,03
Frau	30,55	44,48	21,97

c) Teststatistik:

$$Chi2 = \frac{(12 - 26{,}45)^2}{26{,}45} + \cdots + \frac{(21 - 21{,}97)^2}{21{,}97} \approx 23{,}63$$

d) Freiheitsgrade: df = (2-1)*(3-1)= 2

e)

- Kritischer Wert bei df = 2 und $\alpha = 5\%$: 5,99146
- Also: Chi2 = $23{,}63$ > 5,99146 = kritischer Wert

  Wir können die Nullhypothese auf 5%-Niveau ablehnen. Evidenz für einen Zusammenhang von Geschlecht und Sportpräferenz.