Statistik Vorlesung

10. März 2025

Dauer der Prüfung: 100 Minuten

VORNAME:		MATR.NR.:	
NACHNAME:			

ERLAUBT: **Skriptum des Instituts, Taschenrechner gemäß Liste**

VERBOTEN: **alle sonstigen Unterlagen, Handys**

Bei den Fragen aus Beispiel 1 und 2 bringt eine richtige Antwort 2 Punkte und eine falsche 1 Punkt Abzug. Es gibt keine negative Punktemitnahme in ein anderes Beispiel.

Berechnungen müssen nachvollziehbar aufgeschrieben sein.

Aufgabe	max. Punkte	erreichte Punkte
1	12	
2	10	
3	20	
4	20	
5	30	
6	8	
Summe	100	
Note		

1. (12 Punkte)

   a) Wenn der Gini-Koeffizient der Vermögensverteilung eines Landes steigt, bedeutet das, dass die Einkommens-Ungleichheit abnimmt.

      ◯ Richtig ◯ Falsch

   b) Bei der Auswahl eines geeigneten Zusammenhangsmaßes für unterschiedlich skalierte Merkmale orientiert man sich an dem Merkmal mit dem ________________ Skalenniveau.

   c) Wenn zwei Ereignisse A und B disjunkt sind, dann ist ihre gemeinsame Wahrscheinlichkeit $P(A \cap B)$ gleich 0.

      ◯ Richtig ◯ Falsch

   d) Eine Dichtefunktion kann keine Werte größer als 1 annehmen.

      ◯ Richtig ◯ Falsch

   e) Ein größerer Stichprobenumfang führt tendenziell zu einem ________________ Konfidenzintervall.

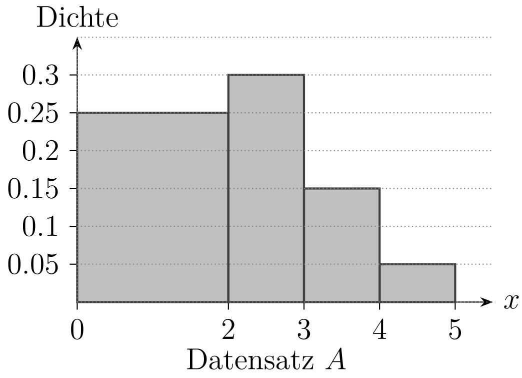
   f) Ein hoher p-Wert eines Regressionskoeffizienten deutet darauf hin, dass die unabhängige Variable einen starken Einfluss auf die abhängige Variable hat.

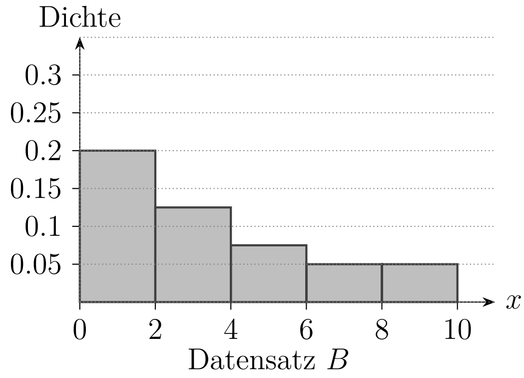
      ◯ Richtig ◯ Falsch

Lösung:
F
niedrigeren
R
F
kleineren, schmäleren
F

2. (10 Punkte)

Gegeben sind die Histogramme für den Datensatz A und für den Datensatz B.





Welche der folgenden Aussagen treffen zu?

a) Bei Datensatz B liegen in der Klasse von 0 bis 2 am meisten Beobachtungen.

◯ Richtig ◯ Falsch

b) Bei Datensatz A liegen in der Klasse von 2 bis 3 am meisten Beobachtungen.

◯ Richtig ◯ Falsch

c) Datensatz B ist linksschief.

◯ Richtig ◯ Falsch

d) Die Dichte in Datensatz A ist im Allgemeinen größer als die in Datensatz B. Daran kann man erkennen, dass die Anzahl der Beobachtungen bei A größer war als bei B.

◯ Richtig ◯ Falsch

e) Wie viel Prozent der Beobachtungen liegen bei Datensatz B im Intervall zwischen 6 und 10?

Lösung: R F F F 20

3. In einer Statistik-Übungsgruppe wurden im vergangenen Semester folgende Mitarbeitspunkte erzielt:

Mitarbeitspunkte	0	1	2	3	4	5
Absolute Häufigkeit	12	9	16	1	14	3

a) (7 Punkte) Zeichnen Sie eine geeignete Verteilungsfunktion zur Veranschaulichung dieser Daten! Wieviel Prozent der Studierenden haben maximal die Hälfte der erzielbaren Mitarbeitspunkte erreicht? Bestimmen Sie den entsprechenden Wert und markieren Sie ihn in Ihrer Zeichnung!

b) (13 Punkte) Zu Semesterende möchten die Studierenden einer Lerngruppe herausfinden, ob es einen Zusammenhang zwischen den aus den beiden Kurztests erzielten Punkten und ihren Mitarbeitspunkten gibt. Die erzielten Punkte der Studierenden aus der Lerngruppe sind in nachfolgender Tabelle gegeben.

	Herr A.	Frau B.	Frau C.	Herr D.
Kurztest-Punkte	18	20	12	14
Mitarbeitspunkte	4	3	2	1

i. Mit welcher Kennzahl kann ein möglicher linearer Zusammenhang zwischen den beiden Merkmale am besten beschrieben werden? Begründen Sie ihre Wahl!

ii. Berechnen Sie diese Kennzahl!

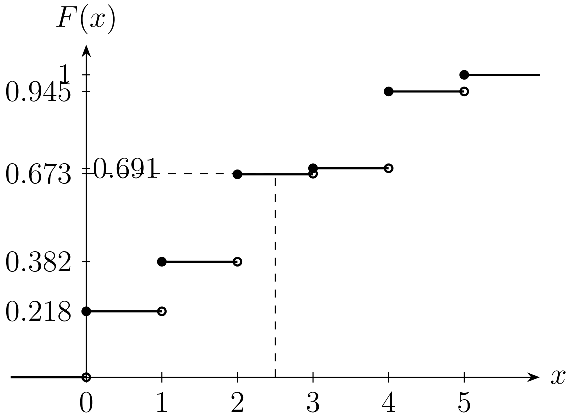
iii. Interpretieren Sie Ihr Ergebnis!

Ausführung Beispiel 3:

Ausführung Beispiel 3:

Lösung:

a)



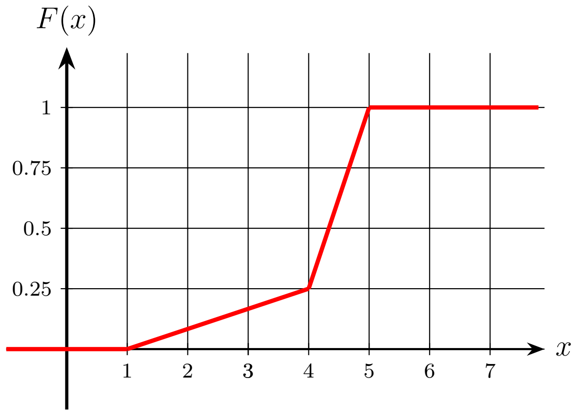
$F(2{,}5) = 0{,}673$

b) i. zwei kardinal skalierte Merkmale: Korrelationskoeffizient

ii. $\rho(X,Y) = 0{,}707$

iii. Es besteht eine starke positive Korrelation zwischen der Höhe der erzielten Kurztest-Punkte und den Mitarbeitspunkten; je höher die Kurztest-Punkte, desto tendenziell höher sind die Mitarbeitspunkte.

4. a) Gegeben ist die Verteilungsfunktion $F(x)$ einer Zufallsvariablen X:



Bestimmen Sie

i. (3 Punkte) $P(X \leq 3)$ und $P(4 \leq X \leq 5)$

ii. (2 Punkte) die Fläche unter der dazugehörigen Dichtefunktion $f(x)$ zwischen den Grenzen von 2,5 bis 4

iii. (3 Punkte) die Zahl c, für die $P(X \geq c) = 0{,}5$ gilt.

b) (12 Punkte) Es ist bekannt, dass 80 % aller Äpfel bei der Ernte zwischen 125 und 175 Gramm schwer sind. Berechnen Sie unter den Annahmen, dass das Gewicht der Äpfel normalverteilt ist und das Intervall $[125; 175]$ symmetrisch zum Erwartungswert liegt, die Wahrscheinlichkeit, dass ein beliebiger Apfel schwerer als 160 Gramm ist!

Hinweis: Erstellen Sie zunächst eine Skizze des Sachverhalts und bestimmen Sie den Erwartungswert und die Standardabweichung der Verteilung!

Ausführung Beispiel 4:

Ausführung Beispiel 4:

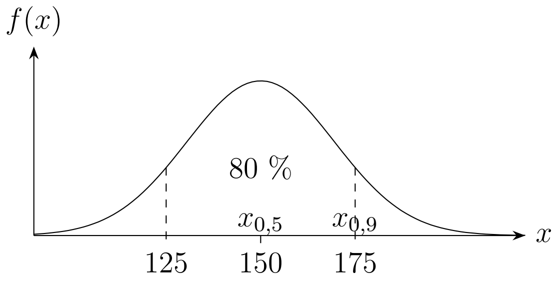
Lösung:

a) i. Es gilt $P(X \leq 3) = F(3) = \frac{1}{6}$ und $P(4 \leq X \leq 5) = F(5) - F(4) = \frac{3}{4}$.

ii. Wir suchen $F(4) - F(2{,}5) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8}$.

iii. Die Zahl c muss die Gleichung $F(c) = 0{,}5$ erfüllen, daher folgt $c = \frac{13}{3}$.

b)



Es gilt $\mu = 150$ wegen der Symmetrie der Normalverteilung und mit

$$P(X \leq 175) = \Phi\left(\frac{175 - 150}{\sigma}\right) = 0{,}9$$

folgt

$$\frac{175 - 150}{\sigma} = 1{,}282$$

beziehungsweise $\sigma \approx 19{,}501$.

Jetzt berechnen wir noch

$$P(X > 160) = 1 - P(X \leq 160) = 1 - \Phi\left(\frac{160 - 150}{19{,}501}\right) = 1 - \Phi(0{,}51) = 0{,}309.$$

5. Die zwei vierten Schulklassen, 4a und 4b, eines Gymnasiums fahren gemeinsam auf Schikurs. Zum Abschluss findet ein Wettrennen statt, an dem per Losentscheid je fünf Kinder ihre jeweilige Klasse vertreten. Beim Schirennen ergaben sich die folgenden Laufzeiten (in Sekunden) für die Kinder der beiden Schulklassen:

Schulklasse 4a (X)	36	38	41	33	37
Schulklasse 4b (Y)	38	40	37	34	33

Es kann angenommen werden, dass die Zeiten beim Schirennen normalverteilt sind.

a) (8 Punkte) Berechnen Sie ein zweiseitiges 90-%-Konfidenzintervall für die durchschnittliche Laufzeit der Schulklasse 4a!

b) (2 Punkte) Interpretieren Sie dieses Intervall inhaltlich mit Bezug auf die Aufgabenstellung!

c) (5 Punkte) Fahren die Kinder der Schulklasse 4b beim Schirennen signifikant schneller Schi als die Kinder der Schulklasse 4a? Welche Tests kommen für die Überprüfung der Fragestellung infrage? Begründen Sie Ihre Wahl! Formulieren Sie die Null- (H_0) und die Alternativhypothese (H_1) dieser Tests! Wie können Sie entscheiden? Begründen Sie Ihre Entscheidung verbal, ohne die Tests zu rechnen!

d) (15 Punkte) Angenommen, die Laufzeiten beim Schirennen wären nicht normalverteilt. Welcher Test wäre anzuwenden, wenn auf dem 0,1-Signifikanzniveau gezeigt werden soll, dass die Kinder der Schulklasse 4b beim Schirennen schneller Schifahren als die Kinder der Schulklasse 4a? Formulieren Sie die Null- und die Alternativhypothese und führen Sie den Test durch! Interpretieren Sie das Ergebnis inhaltlich mit Bezug auf die Aufgabenstellung!

Ausführung Beispiel 5:

Ausführung Beispiel 5:

Lösung:

a) [34,220 39,780]

b) Die durchschnittliche Laufzeit aller Kinder der Schulklasse 4a, d.h. der wahre Parameter der Grundgesamtheit, liegt mit einer Wahrscheinlichkeit von 90 % im Intervall [34,220; 39,780].

c) 2 Stichproben, Test auf Lage, normalverteilt, unverbunden
Zweistichproben-t-Test oder Welch-Test
F-Test zur Prüfung der Varianzhomogenität
bei Varianzhomogenität: Zweistichproben-t-Test
bei Varianzheterogenität: Welch-Test
$H_0 : \mu_y - \mu_x \geq 0$, $H_1 : \mu_y - \mu_x < 0$

d) Wilcoxon-Rangsummentest; $H_0 : c \geq 0$ $H_1 : c < 0$; $t_0 = 0{,}209$; $K =]1{,}282; \infty[$ $t_0 \notin K$; H_1 nicht bestätigt
Es kann nicht gezeigt werden (mit $\alpha = 0{,}1$), dass die Kinder von 4b beim Schirennen schneller Schifahren als die Kinder von 4a.

6. Zur Untersuchung von Faktoren, die die Hotelpreise in Salzburg beeinflussen, wurde untenstehende Regressionsanalyse durchgeführt. Das Modell basiert auf Daten von 70 Vier-Sterne-Hotels in der Stadt und schätzt den Preis für ein Doppelzimmer pro Nacht (in €) in Abhängigkeit von folgenden Einflussgrößen: Zimmergröße (in m^2), beabsichtigter Urlaubszeitraum (Sommer oder Winter), Verpflegung (mit oder ohne Frühstück) sowie Entfernung vom Stadtzentrum (in km).

Variable	Koeffizienten	Standardfehler	t-Statistik	P-Wert
Schnittpunkt	170,38	46,81	3,64	0,00
Zimmergröße	3,75	2,04	1,83	0,07
Winter	20,06	16,67	1,20	0,23
Frühstück inkludiert	57,94	16,83	3,44	0,00
Entfernung vom Stadtzentrum	−27,65	6,34	−4,36	0,00

a) (3 Punkte) Schätzen Sie den Übernachtungspreis für ein 20 m^2 großes Zimmer in einem Hotel, das 1,5 km vom Salzburger Stadtzentrum entfernt liegt. Nehmen Sie außerdem an, dass es sich um eine Buchung für die Sommerferien handelt und Verpflegung mit Frühstück erwünscht ist.

b) (1 Punkt) Wie groß ist der geschätzte Preisunterschied, wenn eine Nächtigung im gleichen Hotel im Winter statt im Sommer erfolgt?

c) (2 Punkte) Interpretieren Sie den Wert -27,65, der in der Tabelle angegeben ist, inhaltlich.

d) (2 Punkte) Welche Modellparameter sind auf einem Signifikanzniveau von $\alpha = 0{,}05$ statistisch signifikant?

Ausführung Beispiel 6:

Lösung:

a) 261,845

b) 20,06

c) Für jeden zusätzlichen Kilometer Entfernung vom Stadtzentrum sinkt der Preis bei sonst gleichen Bedingungen um 27,65 Euro.

d) Schnittpunkt, Frühstück, Entfernung