Statistik Vorlesung

19. Mai 2025

Dauer der Prüfung: 100 Minuten

VORNAME:		MATR.NR.:	
NACHNAME:			

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

VERBOTEN: **alle sonstigen Unterlagen, Handys**

Berechnungen müssen nachvollziehbar aufgeschrieben sein.

Aufgabe	max. Punkte	erreichte Punkte
1	12	
2	10	
3	20	
4	20	
5	25	
6	13	
Summe	100	
Note		

1. (12 Punkte) Bei den Richtig-Falsch-Fragen bringt eine richtige Antwort 2 Punkte und eine falsche 1 Punkt Abzug. Es gibt keine negative Punktemitnahme in ein anderes Beispiel.

   a) Die Berechnung der Varianz setzt mindestens metrisch skalierte Daten voraus.

      ◯ Richtig ◯ Falsch

   b) Der Rangkorrelationskoeffizient nach Spearman ist ausschließlich für ordinal skalierte Merkmale anwendbar.

      ◯ Richtig ◯ Falsch

   c) Wenn X und Y stochastisch unabhängig sind, so gilt $P(X|Y) = P(X)$

      ◯ Richtig ◯ Falsch

   d) In einem Regressionsmodell wird immer vorausgesetzt, dass die unabhängige Variable X normalverteilt ist.

      ◯ Richtig ◯ Falsch

   e) Welche der folgenden Aussagen trifft auf den Graphen einer Dichtefunktion einer stetigen Zufallsvariable zu? Kreuzen Sie die zutreffende Antwort an.

      ◯ Er ist immer monoton steigend und nimmt nur Werte zwischen 0 und 1 an.

      ◯ Er ist immer monoton fallend und besitzt nur positive Funktionswerte.

      ◯ Er kann unterschiedliche Formen annehmen, solange die Funktion nicht negativ ist und die Fläche unter dem Graphen 1 beträgt.

      ◯ Er muss an der Stelle 0 den Funktionswert 1 besitzen.

   f) Welchen statistischen Fehler begeht man, wenn man sich fälschlicherweise für die H_0 entscheidet? ____________________

   Lösung: R F R F 3 Fehler 2. Art

2. (10 Punkte)

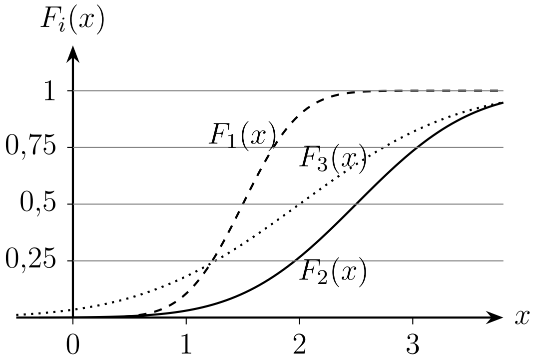


Abbildung 1

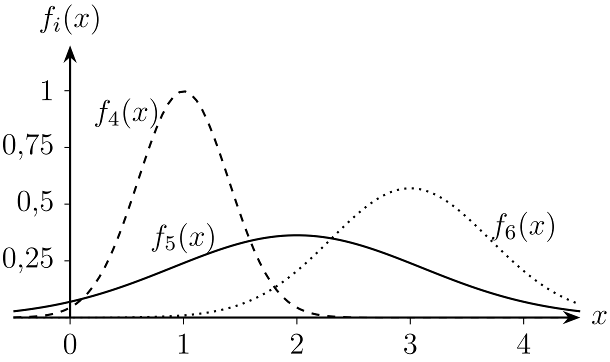


Abbildung 2

In Abbildung 1 sind für drei normalverteilte Zufallsvariablen X_1, X_2 und X_3 die drei entsprechenden Verteilungsfunktionen $F_1(x)$ (strichliert), $F_2(x)$ (durchgehend) und $F_3(x)$ (gepunktet) gezeichnet. In Abbildung 2 sind für drei weitere normalverteilte Zufallsvariablen X_4, X_5 und X_6 die drei entsprechenden Dichtefunktionen $f_4(x)$ (strichliert), $f_5(x)$ (durchgehend) und $f_6(x)$ (gepunktet) gezeichnet.

a) Welche Zufallsvariable hat den größten Erwartungswert?

☐ X_1 ☐ X_2 ☐ X_3

Welche Zufallsvariable hat den kleinsten Erwartungswert?

☐ X_1 ☐ X_2 ☐ X_3

b) Welche Zufallsvariable hat die größte Varianz?

☐ X_1 ☐ X_2 ☐ X_3

Welche Zufallsvariable hat die kleinste Varianz?

☐ X_1 ☐ X_2 ☐ X_3

c) Welche Zufallsvariable hat das größte 5-Prozent-Quantil?

☐ X_1 ☐ X_2 ☐ X_3

Welche Zufallsvariable hat das kleinste 5-Prozent-Quantil?

☐ X_1 ☐ X_2 ☐ X_3

d) Welche Zufallsvariable hat die größte Varianz?

☐ X_4 ☐ X_5 ☐ X_6

Welche Zufallsvariable hat die kleinste Varianz?

☐ X_4 ☐ X_5 ☐ X_6

e) Füllen Sie die Lücken im folgenden Satz so, dass eine richtige Aussage entsteht:

Die Verteilungsfunktion __________ und die Dichtefunktion __________ beschreiben die gleiche Zufallsvariable.

Lösung: $X_2 X_1$ $X_3 X_1$ $X_2 X_3$ $X_5 X_4$ $F_3 f_5$

3.A In der folgenden Tabelle sind die Noten und die erreichten Punkte von Studierenden bei einer Klausur angegeben:

Note	5	4	3	2	1
Erreichte Punkte	[0; 50]	]50; 60]	]60; 70]	]70; 80]	]80; 100]
Absolute Häufigkeit	11	7	12	6	4

a) (9 Punkte) Berechnen Sie den Modalwert und näherungsweise den Median und das arithmetische Mittel der erreichten Punkte! Welcher der berechneten Werte beschreibt die zentrale Tendenz der Punkteverteilung am besten? Begründen Sie Ihre Antwort!

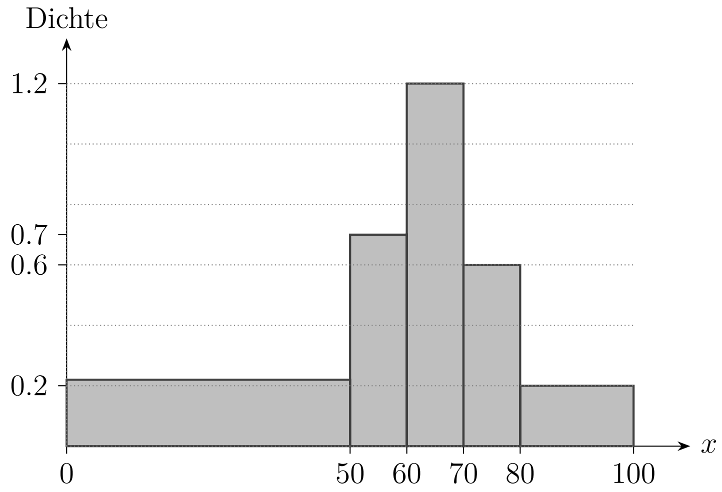
b) (6 Punkte) Stellen Sie die Daten in einem geeigneten Diagramm graphisch dar!

c) (1 Punkt) Interpretieren Sie das Diagramm aus b) hinsichtlich der Schiefe der Daten!

Ausführung Beispiel 3.A: Lösung:

a) $x_{mod} = 65$, $x_{0,5} = 61{,}667$, $\bar{x} = 56{,}25$, $x_{0,5}$ oder $\bar{x}$, beide genähert; $x_{0,5}$, da nicht durch Extremwerte verzerrt oder $\bar{x}$, da metrisch

b)



c) linksschief/rechtssteil

3.B Die LV-Leiterin möchte wissen, ob Studierende, die mehr Zeit für die Klausurvorbereitung verwendet haben, bessere Noten haben als Studierende, die weniger lange für die Klausur gelernt haben. Deshalb befragt sie die Studierenden nach ihrem Lernaufwand (in Stunden). Es ergeben sich die folgenden Informationen:

Note	5	4	3	2	1
durchschnittlicher Lernaufwand (in Stunden)	7,5	10,3	15,7	20,1	18,9

a) (2 Punkte) Welche Kennzahl ist für die Berechnung des Zusammenhangs zwischen der Note und dem Lernaufwand geeignet? Begründen Sie! Sie müssen diese Kennzahl nicht berechnen.

b) (2 Punkte) Angenommen, die Kennzahl aus a) beträgt $-0{,}938$. Interpretieren Sie diesen Wert im gegebenen Kontext!

Ausführung Beispiel 3.B:

Lösung:

a) Rangkorrelationskoeffizient

b) Sehr starke negative Korrelation: Studierende, die viel Zeit für Lernen verwendet haben, bekommen tendenziell bessere Noten.

4.A In einer Stadt wird eine Umfrage unter Berufstätigen zur Nutzung von Verkehrsmitteln und zur Pünktlichkeit bei der Ankunft am Arbeitsplatz durchgeführt. Jede befragte Person gibt an, ob sie mit dem Auto, mit dem Fahrrad oder mit öffentlichen Verkehrsmitteln zur Arbeit fährt und ob sie dabei meist pünktlich oder häufig verspätet ankommt. Es liegen folgende Ergebnisse vor:

45 % der Befragten fahren mit dem Auto zur Arbeit, 20% mit dem Fahrrad. Die übrigen nutzen öffentliche Verkehrsmittel. Insgesamt geben 82 % aller Befragten an, pünktlich am Arbeitsplatz anzukommen. Von den Fahrradfahrenden geben sogar 95% an, pünktlich anzukommen. 11% der Befragten fahren mit dem Auto und kommen häufig verspätet zur Arbeit.

a) (5 Punkte) Vervollständigen Sie die folgende Wahrscheinlichkeitstabelle mithilfe der vorliegenden Ergebnisse!

	Pünktlich	**Verspätet**	**Summe**
Auto			0,45
Fahrrad			
Öffentlich			
Summe			**1**

b) (2 Punkte) Wie groß ist die Wahrscheinlichkeit, dass eine unter den Befragten zufällig ausgewählte Person mit dem Auto oder mit dem Fahrrad zur Arbeit fährt?

c) (2 Punkte) Mit welcher Wahrscheinlichkeit ist eine zufällig ausgewählte Person pünktlich, wenn man weiß, dass sie mit dem Auto zur Arbeit fährt?

d) (2 Punkte) Zeigen Sie rechnerisch, dass die Ereignisse $A =$ „Pünktliche Ankunft“ und $B =$ „AutofahrerIn“ stochastisch abhängig sind.

Ausführung Beispiel 4.A:

Lösung:

a)

	Pünktlich	**Verspätet**	**Summe**
Auto	0,34	0,11	0,45
Fahrrad	0,19	0,01	0,20
Öffentlich	0,29	0,06	0,35
Summe	0,82	0,18	1

b) 0,65

c) 0,756

d) $0{,}82 \neq 0{,}756$

4.B An einem Flughafen kommen im Durchschnitt 15 Taxis pro Stunde an, die Fahrgäste aufnehmen möchten. Die Ankünfte erfolgen zufällig und unabhängig voneinander. Es wird angenommen, dass sich die Anzahl der Taxis, die innerhalb einer Stunde eintreffen, durch eine Poisson-Verteilung modellieren lässt.

a) (1 Punkt) Geben Sie den Parameter dieser Verteilung an.

b) (2 Punkte) Mit welcher Wahrscheinlichkeit treffen genau 10 Taxis innerhalb einer Stunde am Flughafen ein?

c) (3 Punkte) Ein Fahrgast wartet 10 Minuten in der Taxizone. Wie groß ist die Wahrscheinlichkeit, dass in diesen 10 Minuten kein Taxi eintrifft?

d) (3 Punkte) Wie groß ist die Wahrscheinlichkeit, dass ein Fahrgast, der gerade am Flughafen ankommt, länger als 5 Minuten warten muss, bis ein Taxi eintrifft? (Hinweis: Exponentialverteilung!)

Ausführung Beispiel 4.B:

Lösung:

a) $\lambda = 15$

b) 0,049

c) 0,082

d) 0,287

5.A (8 Punkte) In einer Umfrage unter Adria-Urlaubern gaben 86 Prozent der Befragten an, Quallen in den Gewässern als äußerst störend zu empfinden. Wie viele Personen wurden befragt, wenn mit einer Sicherheit von 90 Prozent mindestens 81,55 Prozent der Urlauber angegeben haben, die Quallen als äußerst störend empfunden zu haben?

Ausführung Beispiel 5.A:

Lösung: 100

5.B a) (14 Punkte) In einer Umfrage wurden Urlauber befragt, wie viel sie für Essen und Souvenirs pro Urlaubstag ausgegeben haben. Dazu wurden 3 Frauen und 4 Männer befragt und die Ergebnisse für die Ausgaben in der folgenden Tabelle zusammengefasst:

Frauen (X)	48	80	44	
Männer (Y)	86	17	47	119

Kann mit den vorliegenden Daten gezeigt werden ($\alpha = 0{,}1$), dass Urlauber täglich mehr als 50 Euro ausgeben? Es wird angenommen, dass die Daten normalverteilt sind.

b) (3 Punkte) Es wird vermutet, dass die Streuung der Ausgaben bei Männern höher ist als bei Frauen. Welchen Test ist durchzuführen, um das zu zeigen? Formulieren Sie die Hypothesen! Der Test selbst braucht nicht durchgeführt zu werden.

Ausführung Beispiel 5.B:

Lösung:

a) Einstichproben-t-Test; $H_1 : \mu > 50$, $H_0 : \mu \leq 50$; $t_0 = 1{,}013$; $K =]1{,}44; \infty[$; $H_!$ nicht bestätigt; Nein, das konnte nicht gezeigt werden.

b) F-Test; $H_1 : \sigma_x^2/\sigma_y^2 < 1$; $H_0 : \sigma_x^2/\sigma_y^2 \geq 1$

6. Der Besitzer einer mobilen Eisdiele möchte herausfinden, bei welcher Außentemperatur sich ein Ausflug mit seiner Eisdiele lohnt. In den letzten Tagen hat er folgende Verkaufszahlen pro Stunde und die vorherrschende Temperatur notiert:

Temperatur [in °C]	32	25	31	22
Anzahl der verkauften Eistüten pro Stunde	40	27	38	12

Die Regressionsgerade wurde aus dem Modell wie folgt geschätzt: $\hat{Y} = 42{,}290 + 2{,}601x$

Der Standardfehler der Regression wurde ebenfalls geschätzt: $\hat{\sigma} = 3{,}728$

a) (5 Punkte) Berechnen Sie den Standardfehler des Koeffizienten $\hat{\beta}_1$ ($\hat{\sigma}_{\hat{\beta}_1}$) und interpretieren Sie diesen!

b) (8 Punkte) In der Regressionsanalyse geht es auch darum, erklärende Merkmale zu finden. Kann die Außentemperatur als ein für die Verkaufszahlen unverzichtbares erklärendes Merkmal angesehen werden? Führen Sie den entsprechenden Test zum Niveau $\alpha = 5$ % durch und beantworten Sie die Frage!

Ausführung Beispiel 6:

Lösung:

a) 0,449; Schwankung des geschätzten Koeffizienten β_1

b) $H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$
$t_0 = 5{,}796$; $K =]-\infty; -4{,}303[\cup]4{,}303; \infty[$ $t_0 \in K$; H_1 bestätigt.
Ja, die Temperatur kann als unverzichtbar angesehen werden.