

Statistik Vorlesung

29. November 2024

Dauer der Prüfung: 100 Minuten

ZUNAME:			
VORNAME:		MATR.NR.:	

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

VERBOTEN: **alle sonstigen Unterlagen, Handys**

Bei den Single-Choice-Fragen 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) Die Standardabweichung der Grundgesamtheit lässt sich auch dann berechnen, wenn diese in Gruppen aufgeteilt ist und nur der Stichprobenumfang, der Mittelwert und die Varianz jeder Gruppe bekannt sind.

☐

Richtig

☐

Falsch

- b) Wenn das χ^2 zwischen 0 und 1 liegt, braucht der Kontingenzkoeffizient nicht mehr korrigiert zu werden.

☐

Richtig

☐

Falsch

- c) Die Ereignisse C und D sind dann stochastisch unabhängig, wenn die Wahrscheinlichkeit für das Eintreten des Ereignisses D nicht vom Eintreten des Ereignisses C abhängt.

☐

Richtig

☐

Falsch

- d) Die Funktionswerte jeder Dichte- und jeder Verteilungsfunktion sind niemals negativ.

☐

Richtig

☐

Falsch

- e) Die Varianz des arithmetischen Mittels ist bei stochastisch unabhängigen Zufallsvariablen gleich groß wie das arithmetische Mittel der Varianzen.

☐

Richtig

☐

Falsch

- f) Die Quadratsumme der Residuen ist bei der linearen Regression typischerweise 0.

☐

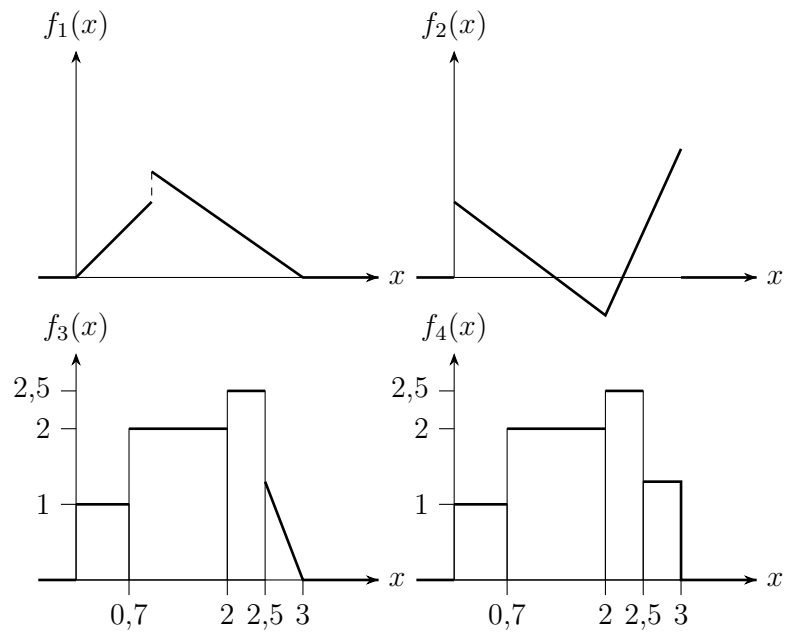
Richtig

☐

Falsch

Lösung: R F R R F F

2. (10 Punkte)



a) Bei $f_1(x)$ kann es sich um eine Dichtefunktion handeln.

☐

Richtig

☐

Falsch

b) Bei $f_2(x)$ kann es sich um eine Dichtefunktion handeln.

☐

Richtig

☐

Falsch

c) Bei $f_3(x)$ kann es sich um eine Dichtefunktion handeln.

☐

Richtig

☐

Falsch

d) Bei $f_4(x)$ kann es sich um eine Dichtefunktion handeln.

☐

Richtig

☐

Falsch

e) Bei $f_4(x)$ kann es sich um ein Histogramm handeln.

☐

Richtig

☐

Falsch

Lösung: R F F F F

3. a) Der Leiter eines Tenniskurses möchte untersuchen, wie regelmäßig die angemeldeten Personen im Semester tatsächlich an den Kurseinheiten teilnehmen. Zu diesem Zweck hat er im letzten Semester wöchentlich Aufzeichnungen über die Anwesenheit der Kursteilnehmerinnen und Kursteilnehmer geführt. Die erfassten Daten geben einen Überblick über die Anzahl insgesamt besuchter Kurseinheiten von jedem der 12 Teilnehmerinnen und Teilnehmer.

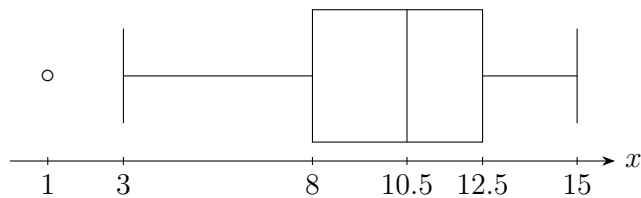
Anzahl besuchter Einheiten:	11	8	13	15	12	10	9	1	3	8	15	12
------------------------------------	----	---	----	----	----	----	---	---	---	---	----	----

- (8 Punkte) Bestimmen Sie die für einen Boxplot benötigten Größen und zeichnen Sie diesen!
- (1 Punkt) Berechnen Sie das arithmetische Mittel!
- (1 Punkt) Interpretieren Sie den Boxplot hinsichtlich der Schiefe der Daten!

Ausführung Beispiel 3a:

Lösung:

i.



- $\bar{x} = 9,75$
- linksschief

- b) Der Betreiber eines Fitnessstudio-Netzwerks möchte analysieren, ob Studios mit höheren Mitgliedsbeiträgen von ihren Kundinnen und Kunden auch als qualitativ besser wahrgenommen werden und dies in einer höheren Zufriedenheit resultiert. Dazu wurden sechs Mitglieder verschiedener Studios befragt. Für jedes Mitglied wurden der monatliche Mitgliedsbeitrag (in Euro) sowie die Zufriedenheit mit dem Angebot erhoben. Die Zufriedenheit wurde auf einer Skala von 1 (= sehr zufrieden) bis 5 (= sehr unzufrieden) bewertet.

Mitglied	1	2	3	4	5	6
Mitgliedsbeitrag in Euro	29	79	45	99	65	39
Zufriedenheit	4	2	1	1	2	3

- (2 Punkte) Mit welcher Kennzahl kann ein Zusammenhang zwischen den beiden Merkmalen am besten dargestellt werden? Begründen Sie Ihre Wahl!
- (6 Punkte) Berechnen Sie diese Kennzahl!
- (2 Punkte) Interpretieren Sie Ihr Ergebnis!

Ausführung Beispiel 3b:

Lösung:

- Rangkorrelationskoeffizient
- 0,743
- starker Zusammenhang

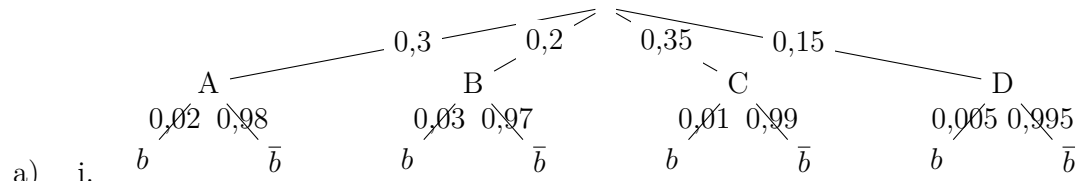
Kunden, die in Studios mit höherem Mitgliedsbeitrag trainieren, sind tendenziell zufriedener mit dem Angebot ihres Studios.

4. a) In einer Bäckerei sind 4 Bäckerinnen mit der Teigherstellung für Vanillekipferl beauftragt, 30 % des Teiges stammen von Bäckerin A, 20 % von Bäckerin B, 35 % von Bäckerin C und 15 % von Bäckerin D. Nach der Kühlung des Teiges werden die Kipferl geformt, wobei erfahrungsgemäß der Anteil der Kipferl, die beim Formen brechen, je Teig unterschiedlich ist: 2 % von Bäckerin A, 3 % von Bäckerin B, 1 % von Bäckerin C und 0,5 % von Bäckerin D.
- (2 Punkte) Stellen Sie den Sachverhalt in einem Baumdiagramm dar!
 - (2 Punkte) Wie groß ist die Wahrscheinlichkeit, dass ein Kipferl beim Formen brechen wird?
 - (1 Punkt) Mit welcher Wahrscheinlichkeit stammt der Teig von Bäckerin A und das Kipferl wird beim Formen brechen?
 - (3 Punkte) Ein nicht gebrochenes Kipferl wurde geformt. Wie groß ist die Wahrscheinlichkeit, dass der Teig von Bäckerin D stammt?
- b) Die Vanillekipferl werden in Dosen abgefüllt. Das Füllgewicht ist normalverteilt mit einem Erwartungswert von 180 g und einer Standardabweichung von 25 g.
- (4 Punkte) Bestimmen Sie den Wert der Verteilungsfunktion an der Stelle 185 und interpretieren Sie diesen!
 - (4 Punkte) Bestimmen Sie die Zahl d derart, dass $P(X > d) = 0,9$! Skizzieren Sie die **Verteilungsfunktion** samt Skalen auf beiden Achsen und kennzeichnen Sie darin die Lösung!
 - (4 Punkte) Zehn Keksdosen werden in einem Karton verpackt. Das Gewicht des Kartons ist normalverteilt mit einem Erwartungswert von 300 g und einer Standardabweichung von 10 g. Mit welcher Wahrscheinlichkeit wiegt der volle Karton mehr als 2 kg?

Ausführung Beispiel 4:

Ausführung Beispiel 4:

Lösung:



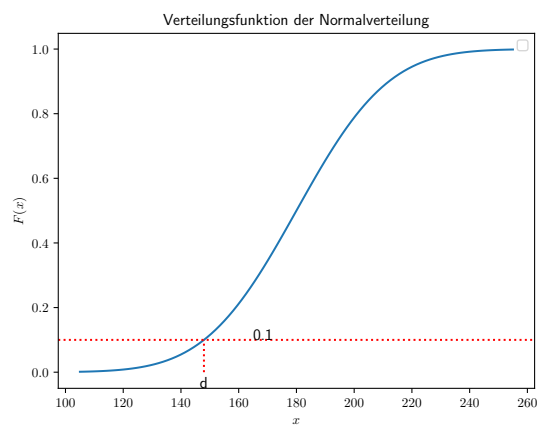
ii. 0,01625

iii. 0,006

iv. 0,152

b) i. 0,579; 57,9 % der Füllgewichte betragen höchstens 185 g.

ii. 147,961



iii. 0,895

5. a) Eine PR-Firma sucht kurzfristig Studierende, die an den Adventsamstagen in einem Grazer Einkaufszentrum als Elfen verkleidet auftreten und tanzen. Es gibt sechs Bewerbungen. Da die Tanzleistungen der Studierenden teilweise sehr zu wünschen übrig lassen, organisiert die PR-Firma einen Crash-Tanzkurs für alle. Die Tanzleistungen der Studierenden vor und nach dem Tanzkurs werden wie folgt beurteilt (höhere Werte bedeuten bessere Tanzleistungen):

Studierender	A	B	C	D	E	F
Vor Tanzkurs (X)	15	30	28	31	19	22
Nach Tanzkurs (Y)	32	39	28	27	20	29

Hat der Tanzkurs zu besseren Tanzleistungen geführt?

- i. (3 Punkte) Welche Tests können für die Überprüfung der Fragestellung angewendet werden? Welche Voraussetzungen müssen für die verschiedenen Tests gegeben sein? Wie werden diese Voraussetzungen überprüft?
 - ii. (15 Punkte) Gehen Sie nun davon aus, dass die Tanzleistungen nicht normalverteilt sind! Führen Sie den Test zum Niveau $\alpha = 5\%$ durch und interpretieren Sie Ihre Entscheidung!
- b) Bei einem Stand des Christkindlmarkts wird Glühwein in Bechern mit einer Füllmenge von 200 ml verkauft. Fünf Freunde, die in diesem Semester die Statistik-Vorlesung besuchen, wollen wissen, ob die versprochene Menge an Glühwein auch eingehalten wird. Sie messen nach und erhalten folgende Ergebnisse:

Studierender	A	B	C	D	E
Glühwein in ml	205	195	198	208	199

- i. (11 Punkte) Bestimmen Sie ein zweiseitiges 0,95-Konfidenzintervall für die mittlere Glühweinmenge und interpretieren Sie dieses!
- ii. (1 Punkt) Drei der fünf Freunde haben weniger als die versprochenen 200 ml Glühwein in ihren Bechern und sind enttäuscht. Mit welchem Test kann überprüft werden, ob die ausgeschenkte Glühweinmenge tatsächlich im Schnitt unter der versprochenen Sollmenge von 200 ml liegt?

Ausführung Beispiel 5:

Ausführung Beispiel 5:

Lösung:

- a)
 - i. t -Test für Differenzen, wenn normalverteilt
Wilcoxon-Vorzeichen-Rang-Test, wenn nicht normalverteilt
Jarque-Bera-Test, Prüfung auf Normalverteilung
 - ii. $H_1 : d_{0,5} > 0$; $H_0 : d_{0,5} \leq 0$; $t_0 = 1,483$; $K =]1,645; \infty[$
Es konnte nicht gezeigt werden, dass der Tanzkurs zu besseren Tanzleistungen führt.
- b)
 - i. $[194,372; 207,628]$
Der wahre Erwartungswert liegt mit einer Wahrscheinlichkeit von 95 % in diesem Intervall.
 - ii. Einstichproben- t -Test

6. Untenstehend finden Sie eine Regressionsanalyse basierend auf Daten von 50 Startup-Unternehmen aus den US-Bundesstaaten New York, Kalifornien und Florida. Das Modell soll den Unternehmensgewinn (in Mio. \$) basierend auf den Investitionen in Forschung und Entwicklung (F & E), und Marketing (jeweils in Mio. \$), sowie den Unternehmensstandort, schätzen.

Regressions-Statistik

Bestimmtheitsmaß	−0,29
Standardfehler	4,29
Beobachtungen	50

Variable	Freiheitsgrade (df)	Quadratsummen (SS)	Prüfgröße (F)	F kritisch
Regression (SSR)	4	300,00	4,22	0,01
Residuen (SSE)	45	800,00		
Gesamt (SST)	49	100,00		

Variable	Koeffizienten	Standardfehler	t-Statistik	P-Wert	Untere 95%	Obere 95%
Schnittpunkt	10,00	2,68	3,73	0,00	4,60	15,40
F & E	0,30	0,14	2,14	0,04	0,02	0,58
Marketing	0,10	0,06	1,67	0,10	−0,02	0,22
Kalifornien	−2,00	1,18	−1,69	0,10	−4,38	0,38
Florida	−3,00	1,66	−1,81	−0,08	−6,34	0,34

- (3 Punkte) In den drei Tabellen oben hat sich jeweils ein Fehler eingeschlichen. Markieren Sie diese und begründen Sie Ihre Wahl!
- (2 Punkte) Schätzen Sie den Gewinn eines Unternehmens, das jeweils 10 Mio. \$ in F & E bzw. Marketing investiert und den Sitz in New York hat.
- (2 Punkte) Wie ändert sich Ihre Schätzung, wenn das Unternehmen alle 20 Mio. \$ in Marketing investiert und den Sitz in Florida hat?
- (1 Punkt) Welche der Modellparameter Marketing und F & E sind mit Niveau $\alpha = 0.05$ signifikant?

Ausführung Beispiel 6:

Ausführung Beispiel 6:

Lösung:

a) 1. $0 \leq R^2 \leq 1$, 2. $SST = SSR + SSE$, 3. $0 \leq \text{p-Value} \leq 1$

b) $G = 10 + 0.3 \cdot 10 + 0.1 \cdot 10 = 14$

c) $G' = 10 + 0.3 \cdot 0 + 0.1 \cdot 20 - 3 = 9$, $\Delta = G' - G = -5$ oder $\Delta = -0.3 \cdot 10 + 0.1 \cdot 10 - 3 = -5$

d) F & E