Statistik Vorlesung

04. Februar 2025

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) Das arithmetische Mittel ist kleiner als der Median, wenn sich in den Daten Ausreißer mit relativ kleinen Zahlenwerten befinden.

◯ Richtig ◯ Falsch

b) Die Fläche unter einer beobachteten Lorenz-Kurve kann bei vollkommener Ungleichverteilung auch null sein.

◯ Richtig ◯ Falsch

c) Wenn die Ankunftszeit eines Busses an einer bestimmten Haltestelle stetig gleichverteilt zwischen 8.00 Uhr und 8.10 Uhr ist, dann ist die Wahrscheinlichkeit dafür, dass der Bus zwischen 8:03 Uhr und 8:05 Uhr kommt, gleich groß wie die Wahrscheinlichkeit für eine Ankunft zwischen 8.08 Uhr und 8:10 Uhr.

◯ Richtig ◯ Falsch

d) In der Dichtefunktion einer Standardnormalverteilung beträgt der Abstand zwischen Wendestelle und der Maximalstelle genau 1.

◯ Richtig ◯ Falsch

e) Der Jarque-Bera-Anpassungstest prüft mittels einer chi-Quadrat-verteilten Teststatistik, ob die zu testende Verteilung einer Normalverteilung entspricht.

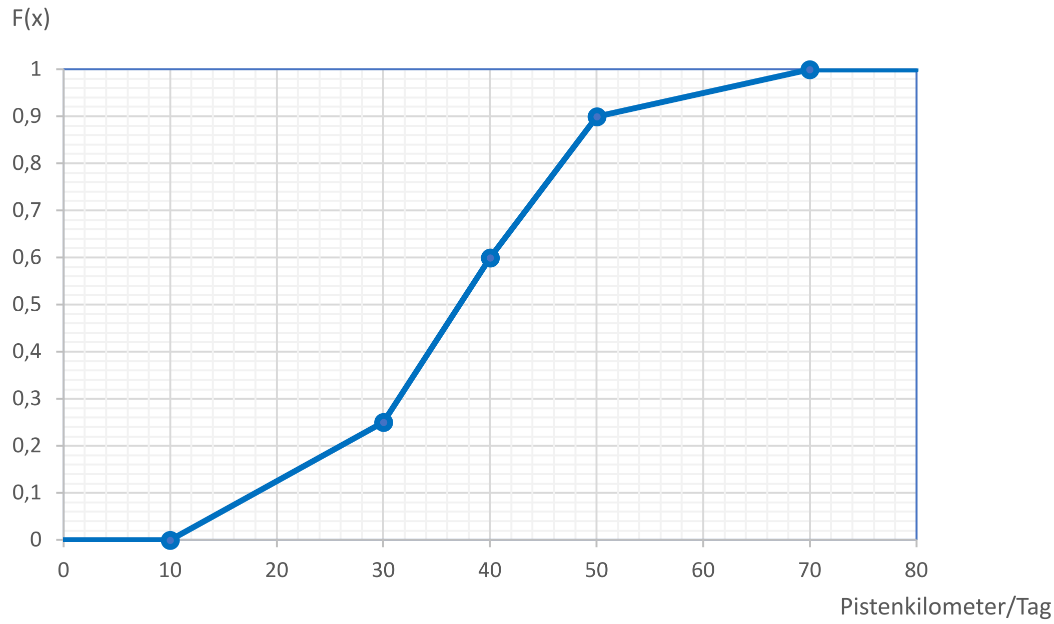
◯ Richtig ◯ Falsch

f) Die Steigung der Regressionsgerade zeigt an, wie sich die abhängige Variable ändert, wenn die unabhängige Variable um den Wert des jeweiligen Koeffizienten erhöht wird.

◯ Richtig ◯ Falsch

Lösung: F F R R R F

2. (10 Punkte)
Um herauszufinden, wie viele Pistenkilometer an einem Ski-Tag zurückgelegt werden, wurden 1.000 Personen befragt. Die Ergebnisse der Umfrage sind in folgender approximierenden Verteilungsfunktion $F(x)$ grafisch dargestellt:



Welche der folgenden Aussagen treffen zu?

a) Die Verteilungsfunktion in der Graphik gehört zu klassierten Daten.

◯ Richtig ◯ Falsch

b) Für diese Verteilungsfunktion wurde unterstellt, dass die Pistenkilometer/Tag in den einzelnen Klassen jeweils gleichverteilt sind.

◯ Richtig ◯ Falsch

c) Der Median der Umfrageergebnisse kann zumindest näherungsweise berechnet werden.

◯ Richtig ◯ Falsch

d) 10 % der befragten Personen legen mindestens 50 Pistenkilometer/Tag zurück.

◯ Richtig ◯ Falsch

e) 350 Personen legen zwischen 40 und 50 Pistenkilometer/Tag zurück.

◯ Richtig ◯ Falsch

Lösung: R R R R F

3. a) Um herauszufinden, wie viele Pausen ein Schifahrer an einem Schitag in einer Schihütte macht, wurden 200 Personen befragt. Die Ergebnisse der Umfrage sind in folgender Tabelle zusammengefasst:

Anzahl der Pause in einer Schihütte	0	1	2	3	4
Absolute Häufigkeit	8	30	64	72	26

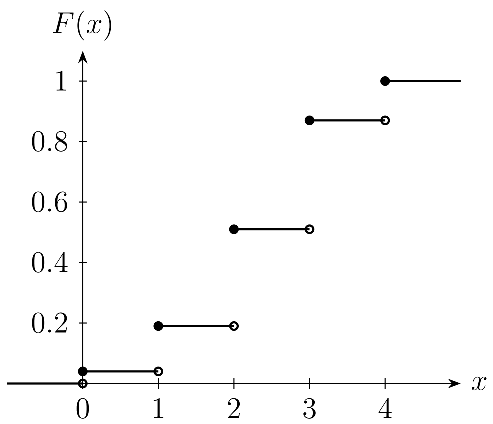
i. (5 Punkte) Zeichnen Sie die geeignete Verteilungsfunktion! Bestimmen und interpretieren Sie den Wert $F(2{,}5)$!

ii. (7 Punkte) Berechnen Sie Modalwert, Median, Mittelwert und Standardabweichung! Welcher dieser Parameter beschreibt die Lage der Daten am besten und warum?

Ausführung Beispiel 3a:

Lösung:

i.



$F(2{,}5) = 0{,}51$ 51 % der Schifahrer machen höchstens zwei Pausen.

ii. Modalwert 3; Median 2; Mittelwert 2,39; Standardabweichung 1,019; Mittelwert, da das Skalenniveau metrisch ist.

b) In einer weiteren Umfrage wurden die durchschnittlichen Ausgaben für Pausen in einer Schihütte an einem Schitag auf der Planai, dem Hauser Kaibling sowie der Reiteralm erhoben. Folgende Daten liegen dazu vor:

Schigebiet	Planai	Hauser Kaibling	Reiteralm
Anzahl der Befragten	50	70	40
Mittelwert der Ausgaben	65	54	72
Standardabweichung der Ausgaben	20	12	35

i. (2 Punkte) In welchem Schigebiet ist die Streuung der Ausgaben am kleinsten? Wählen Sie ein geeignetes Streuungsmaß!

ii. (6 Punkte) Berechnen Sie das arithmetische Mittel sowie die Standardabweichung der Ausgaben über alle drei Schigebiete!

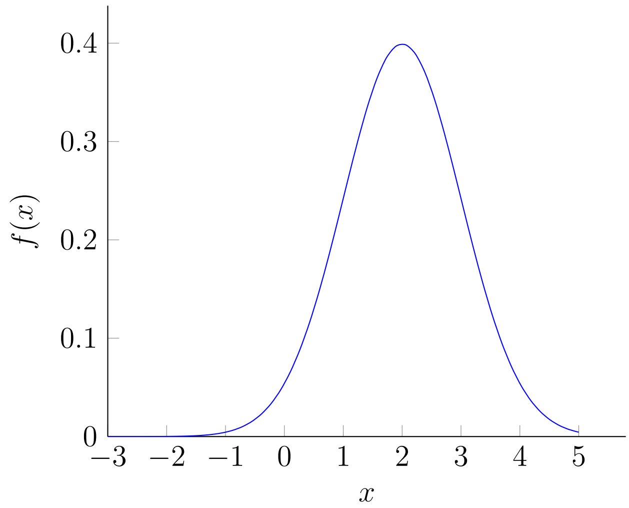
Ausführung Beispiel 3b:

Lösung:

i. Hauser Kaibling

ii. 61,9375; 23,453

4. Eine Zufallsvariable X ist normalverteilt mit $X \sim N(\mu;\sigma)$ wobei $\mu = 2$, und $\sigma = 1$. Das α-Quantil dieser Verteilung ist x_α, und die zugehörige Verteilungsfunktion $F(x)$. Die Dichtefunktion $f(x)$ dieser Verteilung finden Sie in untenstehendem Diagramm.



a) (2 Punkte) Bestimmen Sie den Wert von $F(x_{0.95})$!

b) (2 Punkte) Bestimmen Sie den Wert von $F(x_\alpha) + F(x_{1-\alpha})$ für $\alpha = 0.25$!

c) (2 Punkte) Bestimmen Sie $F(\frac{1}{3})$!

d) (2 Punkte) Bestimmen Sie die Wahrscheinlichkeit $P(X = \frac{2}{3})$!

e) (4 Punkte) Bestimmen Sie ein (beliebiges) Intervall $]a;b[$, sodass $P(a < X < b) = 0{,}5$!

f) (2 Punkte) Eine andere Zufallsvariable Y ist normalverteilt mit $Y \sim N(0;1)$. Skizzieren Sie die Dichtefunktion dieser Verteilung in das Diagramm oben!

g) (3 Punkte) Bestimmen Sie die Fläche, wo sich die beiden Dichtefunktionen überlappen!

h) (3 Punkte) Wie ist $X+Y$ verteilt, wenn X und Y voneinander unabhängig sind? Wie lauten die Parameter?

Ausführung Beispiel 4:

Ausführung Beispiel 4:

Lösung:

a) $F(x_{0.95}) = 0.95$

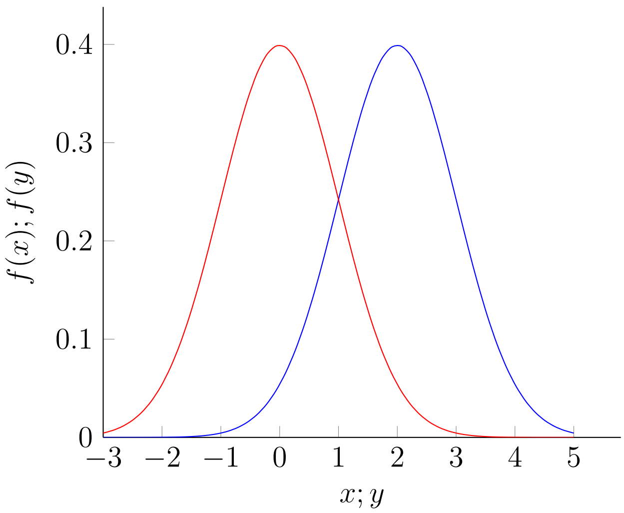
b) $F(x_\alpha) + F(x_{1-\alpha}) = 1$

c) $F(\frac{1}{3}) = 0{,}047$

d) $P(X = x) = 0$

e) z.B.: $a = -\infty$, $b = \mu$, oder $a = 1{,}326$, $b = 2{,}674$

f)



g) $A = 2 \cdot P(Y \geq 1) = 2 \cdot [1 - P(Y \leq 1)] = 1 \cdot [1 - \Phi(1)] = 2 \cdot [1 - 0{,}841] = 0.318$

h) $(X + Y) \sim N(0 + 2; \sqrt{1^2 + 1^2}) = N(2; \sqrt{2})$

5. a) Im Rahmen des alljährlichen Winterfestes zum Semesterabschluss werden von einer Schule eine Vielzahl winterlicher Aktivitäten organisiert. Ein besonderer Höhepunkt ist das Schlittschuhrennen auf der schuleigenen Eisbahn, bei dem die Laufzeiten (in Sekunden) aller teilnehmenden Kinder erfasst werden. Die benötigte Zeit, um eine vorgegebene Strecke auf dem Eis zurückzulegen, kann als normalverteilte Zufallsvariable angenommen werden.

Laufzeit (in Sekunden)	92	114	84	79	123	95	86	102	118	94

i. (8 Punkte) Bestimmen Sie, basierend auf den gegebenen Ergebnissen von 10 Kindern, ein zweiseitiges 0,95-Konfidenzintervall für die durchschnittliche Zeit, die für diese Eislaufstrecke benötigt wird!

ii. (1 Punkt) Wie verändert sich (bei ansonsten unveränderten Bedingungen) das in i. berechnete Konfidenzintervall, wenn die Laufzeiten von 100 Kindern vorliegen würden? Begründen Sie Ihre Antwort (keine Rechnung erforderlich)!

Ausführung Beispiel 5a:

Lösung:

i. $[87{,}894; 109{,}506]$

ii. Das Intervall wird kleiner, da mehr Information vorliegt, n im Nenner größer, Gesamtausdruck kleiner

b) Ein weiterer Höhepunkt des Winterfestes ist der Wettbewerb im Schneemannbauen, an dem in diesem Jahr 11 Mädchen und 9 Buben teilnehmen. Die Preisrichter vergeben Punkte, um den schönsten Schneemann zu küren. Bei den vorliegenden Ergebnissen fällt auf, dass die besten drei Schneemänner (mit den meisten Punkten) von Mädchen gebaut worden sind. Lässt sich zum Signifikanzniveau von $\alpha = 5\ \%$ auch statistisch nachweisen, dass Mädchen tendenziell schönere Schneemänner bauen als Buben? Die vergebenen Punkte können nicht als Realisation einer normalverteilten Zufallsgröße angesehen werden.

Mädchen (X)	76	92	90	80	64	52	67	89	34	73	62
Buben (Y)	36	75	68	42	55	79	83	48	56		

i. (3 Punkte) Welcher Test ist hier anzuwenden? Begründen Sie Ihre Wahl und formulieren Sie die Hypothesen! Der Test selbst ist nicht durchzuführen.

ii. (2 Punkte) Der Test wurde bereits für Sie durchgeführt und ergab einen Testwert von $t_0 = 1{,}254$, was einem p-Value von 0,105 entspricht. Wie lautet Ihre Testentscheidung? Begründen Sie Ihre Antwort!

iii. (2 Punkte) Interpretieren Sie Ihr Ergebnis inhaltlich mit Bezug zur Aufgabenstellung!

Ausführung Beispiel 5b:

Lösung:

i. Wilcoxon-Rangsummentest; $H_0 : c \geq 0$, $H_1 : c < 0$

ii. H_1 nicht bestätigt, da p-Value $> 0{,}05$

iii. Es konnte nicht gezeigt werden, dass Mädchen signifikant schönere Schneemänner bauen.

c) Eine österreichweite Auswertung über gewählte Wintersportaktivitäten bei von Schulen organisierten Wintersportwochen hat ergeben, dass 10 % aller Schülerinnen und Schüler am liebsten rodeln gehen. In einer Befragung von 10 zufällig ausgewählten Kindern aus Salzburger Schulen haben 2 angegeben, Rodeln als bevorzugte Aktivität für die bevorstehende Sportwoche gewählt zu haben. Kann mit einem Signifikanzniveau von $\alpha = 0{,}1$ nachgewiesen werden, dass Salzburger Kinder eine höhere Präferenz für das Rodeln haben als der österreichische Durchschnitt?

   i. (11 Punkte) Formulieren Sie die Hypothesen und führen Sie den entsprechenden Test durch!

   ii. (3 Punkte) Wie lautet Ihre Testentscheidung? Interpretieren Sie Ihr Ergebnis auch inhaltlich mit Bezug zur Fragestellung laut Angabe!

   Ausführung Beispiel 5c:

   Lösung:

   i. Binomialtest; $H_0 : p \leq 0{,}1,\ H_1 : p > 0{,}1; t_0 = 2; K = \{3, 4, \ldots, 10\}$

   ii. t_0 nicht in K; H_1 nicht bestätigt; Es konnte nicht gezeigt werden, dass Salzburger Kinder eine höhere Präferenz für das Rodeln haben.

6. Um die Abhängigkeit des Weizenertrages von der Düngemittelmenge zu schätzen, führt ein Agrarbetrieb eine lineare Regressionsanalyse durch. Für die statistische Untersuchung werden fünf gleich große Äcker verwendet. Die Untersuchung brachte folgendes Ergebnis (ha ... Hektar = 10.000 m^2):

Düngemittel/ha in kg (X)	1	3	5	9	12
Weizenertrag/ha in Tonnen (Y)	3	5	5	7	16

Die Regressionsgerade wurde aus dem Datensatz bereits ermittelt: $y = 1{,}05 + 1{,}025x$. Das Bestimmtheitsmaß beträgt 0,802.

a) (3 Punkte) Berechnen Sie die Residuen!

b) (3 Punkte) Wie hoch ist die Standardabweichung des Störterms?

c) (2 Punkte) Bestimmen Sie den Pearson-Korrelationskoeffizienten zwischen Düngemittel und Weizenertrag!

Ausführung Beispiel 6:

Lösung:

a)

Residuen	0,925	0,875	−1,175	−3,275	2,65

b) 2,630

c) 0,896