

Statistik Vorlesung

02. Juli 2026

Arbeitszeit: 100 Minuten

VORNAME:		MATR.NR.:	
NACHNAME:			

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

VERBOTEN: **alle sonstigen Unterlagen, Handys**

Aufgaben 8 und 9: Für jede vollständig richtig gelöste Frage gibt es 2 Punkte. Es gibt keine Teilpunkte und keine Minuspunkte.

Lösungswege müssen nachvollziehbar aufgeschrieben sein.

Aufgabe	max. Punkte	erreichte Punkte	Teilsumme
1	11		
2	08		
3	12		
4	13		
5	09		
6	16		
7	11		
8	10		
9	10		
Summe	100		
Note			

1. Ein Obstbetrieb möchte überprüfen, wie stark das Gewicht einzelner Marillen schwankt. Eine Stichprobe von 12 Marillen liefert folgende Werte (in Gramm):

20	22	23	25	27	29	29	31	33	35	48	56
----	----	----	----	----	----	----	----	----	----	----	----

- a) (3 Punkte) Bestimmen Sie Modalwert, Median und arithmetisches Mittel und beurteilen Sie anhand dieser Lageparameter die Symmetrie der Daten.
- b) (5 Punkte) Ermitteln Sie die für einen Boxplot relevanten Größen und zeichnen Sie diesen.
- c) (3 Punkte) Beschreiben Sie verbal, wie sich die grafische Darstellung des Boxplots verändert, wenn alle Werte um 5 Gramm erhöht werden. Erklären Sie, welche Auswirkungen diese Änderung auf den Modalwert, das arithmetische Mittel und die Standardabweichung der Daten hat.

Lösung:

- a) $x_{\text{mod}} = x_{0,5} = 29$; $\bar{x} = 31,5$; nicht symmetrisch, rechtsschief

b)



- c) Der Boxplot liegt um 5 Werte weiter rechts. Die Form ist unverändert. Die Lagemaße sind jeweils um 5 größer. Die Standardabweichung bleibt gleich.

2. An 8 Sommertagen hat Professor Paul die Temperatur in °C sowie seine Konzentrationsleistung erhoben. Die Konzentration wurde mit einer Skala von 1 bis 5 erfasst, wobei 1 = sehr gut, 2 = gut, 3 = mittel, 4 = eher schlecht und 5 = sehr schlecht bedeuten. Die Daten sind in folgender Tabelle zusammengefasst:

Tag	1	2	3	4	5	6	7	8
Temperatur (in °C)	26	34	32	28	35	37	36	22
Konzentration	2	4	3	2	4	5	5	1

- a) (2 Punkte) Welche Kennzahl ist am besten geeignet, um einen Zusammenhang zwischen den Merkmalen festzustellen? Begründen Sie Ihre Wahl.
- b) (6 Punkte) Berechnen Sie diese und interpretieren Sie Ihr Ergebnis.

Lösung:

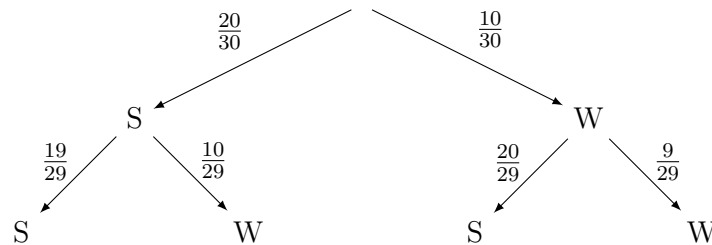
- a) Spearman, da Temperatur kardinal, aber Konzentration nur ordinal ist.
- b) 0,982; Starker monotoner Zusammenhang. Je höher die Temperatur, desto tendentiell niedriger die Konzentration.

3. In einer Urne befinden sich 30 Kugeln: 20 schwarze und 10 weiße.

- a) Es werden zwei Kugeln nacheinander ohne Zurücklegen gezogen.
- (2 Punkte) Zeichnen Sie das Baumdiagramm (inkl. der zugehörigen Wahrscheinlichkeiten).
 - (2 Punkte) Geben Sie mit Hilfe des Baumdiagramms die Wahrscheinlichkeit an, dass beide Kugeln die gleiche Farbe haben.
 - (2 Punkte) Geben Sie mit Hilfe des Baumdiagramms die Wahrscheinlichkeit an, dass die zweite gezogene Kugel schwarz ist, wenn man weiß, dass die erste gezogene Kugel weiß war.
- b) Nun werden 8 Kugeln aus der Urne gezogen (wiederum ohne Zurücklegen). Berechnen Sie unter Verwendung einer geeigneten Formel:
- (4 Punkte) die Wahrscheinlichkeit, dass mehr als eine Kugel weiß ist und
 - (2 Punkte) die Anzahl der zu erwartenden weißen Kugeln.

Lösung:

- a) i.



ii. $\frac{47}{87} \approx 0,54$

iii. $\frac{20}{29} \approx 0,69$

b) i. 0,846

ii. $\frac{8}{3}$

4. Ein Unternehmen produziert elektronische Bauteile in zwei unabhängigen Produktionsschritten. Die Dauer des ersten Produktionsschrittes X (in Minuten) ist normalverteilt mit $X \sim \mathcal{N}(20; 4)$. Die Dauer des zweiten Produktionsschrittes Y (in Minuten) ist normalverteilt mit $Y \sim \mathcal{N}(\mu_Y; 3)$. Die gesamte Produktionsdauer eines Bauteils wird durch die Zufallsvariable $T = X + Y$ beschrieben.
- a) (3 Punkte) Berechnen Sie die Wahrscheinlichkeit, dass der erste Produktionsschritt höchstens 25 Minuten dauert.
 - b) (3 Punkte) Die Bauteile mit den kürzesten 10 % der Produktionsdauern gelten als besonders schnell produziert. Wie viele Minuten darf der erste Produktionsschritt höchstens dauern, damit er noch zu den schnellsten 10 % gehört?
 - c) (3 Punkte) Geben Sie die Verteilung von T mitsamt ihren Parametern in Abhängigkeit von μ_Y an.
 - d) (4 Punkte) Bestimmen Sie den Wert von μ_Y , sodass die Wahrscheinlichkeit, dass die gesamte Produktionszeit höchstens 40 Minuten beträgt, gleich 0,90 ist.

Lösung:

- a) 0,894
- b) 14,872
- c) $T \sim \mathcal{N}(20 + \mu_Y; 5)$
- d) 13,59

5. Ein Unternehmen möchte wissen, wie hoch der durchschnittliche Tagesumsatz in einer Filiale ist. In einer Stichprobe von 25 Tagen wurde ein durchschnittlicher Tagesumsatz von 8.200 Euro mit einer geschätzten Standardabweichung von 500 Euro ermittelt. Aus Erfahrung weiß man, dass die Tagesumsätze normalverteilt sind.
- a) (5 Punkte) Berechnen Sie das zweiseitige 95-%-Konfidenzintervall für den erwarteten durchschnittlichen Tagesumsatz. Interpretieren Sie das Konfidenzintervall verbal im Kontext der Aufgabe.
 - b) (2 Punkte) Wie würde sich das Konfidenzintervall ändern, wenn statt eines 95-%-Konfidenzniveaus ein 99-%-Konfidenzniveau benutzt wird? Begründen Sie Ihre Antwort ohne Rechnung.
 - c) (2 Punkte) Angenommen, die geschätzte Standardabweichung wäre größer. Wie würde sich das auf das Konfidenzintervall bei gleichem Konfidenzniveau auswirken? Begründen Sie Ihre Antwort ohne Rechnung.

Lösung:

- a) Konf [7.993,61; 8.406,39]
- b) Konfidenzintervall wäre größer, da man mehr Sicherheit mit größeren Intervallen erhält.
- c) Konfidenzintervall wäre größer, da s im Zähler steht.

6. Ein Casino behauptet, dass seine Münzen „fair“ sind, d. h. beim Werfen ist die Wahrscheinlichkeit für „Kopf“ und „Zahl“ jeweils $p = 0,5$. Ein Casinobesucher wirft die Münze 20 Mal und erhält 14 Mal „Kopf“. Er legt Beschwerde beim Casino-Manager ein, weil er der Meinung ist, dass die Münze nicht fair sei, also dass die tatsächliche Wahrscheinlichkeit für „Kopf“ von 0,5 abweiche.

Die Tabelle der Verteilungsfunktion für das Merkmal „Kopf“ bei 20 Würfeln mit einer fairen Münze finden Sie unterhalb der Aufgabe d).

- a) (3 Punkte) Welcher Test ist für die Überprüfung der Beschwerde anzuwenden? Formulieren Sie die Hypothesen.
- b) (7 Punkte)
 - i. Geben Sie die Testgröße an.
 - ii. Bestimmen Sie unter Zuhilfenahme der angegebenen Tabelle zu einem Signifikanzniveau von $\alpha = 0,1$ den kritischen Bereich.
 - iii. Entscheiden Sie, ob das Ergebnis statistisch signifikant ist und interpretieren Sie Ihre Entscheidung im Kontext der Aufgabe.
- c) (4 Punkte) Angenommen, der Casinobesucher vermutet, dass „Kopf“ häufiger auftritt als bei einer fairen Münze. Formulieren Sie die entsprechenden Hypothesen und prüfen Sie erneut. Ist die Beschwerde auf dem Signifikanzniveau von 10 % statistisch begründet?
- d) (2 Punkte) Angenommen, die Münze ist in Wahrheit nicht fair und die Wahrscheinlichkeit für „Kopf“ beträgt tatsächlich $p = 0,7$. Welche Art von Fehlentscheidung wurde dann in Aufgabenteil b) getroffen? Beschreiben Sie diesen Fehler kurz in eigenen Worten.

k	$F(k)$	k	$F(k)$	k	$F(k)$	k	$F(k)$
0	0,000						
1	0,000	6	0,058	11	0,748	16	0,999
2	0,000	7	0,132	12	0,868	17	1,000
3	0,001	8	0,252	13	0,942	18	1,000
4	0,006	9	0,412	14	0,979	19	1,000
5	0,021	10	0,588	15	0,994	20	1,000

Lösung:

- a) Binomialtest; $H_0 : p = 0,5$; $H_1 : p \neq 0,5$
- b) $t_0 = 14$; $K = \{0, 1, 2, 3, 4, 5\} \cup \{15, 16, 17, 18, 19, 20\}$
 $t_0 \notin K$; H_0 beibehalten. Es konnte nicht gezeigt werden, dass die Münze unfair ist.
- c) $H_0 : p \leq 0,5$; $H_1 : p > 0,5$, $K = \{14, 15, 16, 17, 18, 19, 20\}$
 $t_0 \in K$; H_1 bestätigt. Es konnte gezeigt werden, dass die Münze unfair und die Beschwerde begründet ist.
- d) Fehler 2. Art; Akzeptanz einer falschen H_0 .

7. Mit den Verkaufsdaten eines mobilen Eisstandes soll der tägliche Umsatz [Euro] mittels Regressionsrechnung vorhergesagt werden. Folgende Variablen wurden zur Erklärung verwendet: Temperatur [Grad Celsius]; Luftfeuchtigkeit [Prozent]; Windgeschwindigkeit [km/h]; Standort (Ausprägungen: Park, Strand, Stadtzentrum); Wochenende (1 = Wochenende, 0 = Werktag).

Die Analyse erbrachte folgendes Ergebnis (Referenzkategorie Standort: Stadtzentrum):

	Koeffizient	Standardfehler	t-Statistik	p-Wert
Schnittpunkt	850,32	412,78	2,060	0,039
Temperatur	12,45	0,38	32,763	0,000
Luftfeuchtigkeit	-3,18	1,413	-2,251	0,024
Windgeschwindigkeit	-5,67	1,676	-3,383	0,001
Standort[Park]	124,80	8,93	13,975	0,000
Standort[Strand]	287,50	10,62	27,072	0,000
Wochenende	95,40	6,75	14,133	0,000

- a) i. (3 Punkte) Schätzen Sie mit Hilfe des Modells den erwarteten Tagesumsatz bei einer Temperatur von 28°C, einer Luftfeuchtigkeit von 45 %, einer Windgeschwindigkeit von 8 km/h, Standort Park und einem Werktag.
- ii. (2 Punkte) Betrachten Sie das Vorzeichen des Koeffizienten für die Variable „Windgeschwindigkeit“. Interpretieren Sie dieses Vorzeichen im Kontext der Aufgabe.
- iii. (2 Punkte) Ein Analyst vermutet, dass die Variable „Luftfeuchtigkeit“ keinen statistisch signifikanten Einfluss auf den Eisumsatz hat. Unterstützen oder widerlegen Sie diese Vermutung mithilfe eines passenden Wertes aus der Tabelle auf einem Signifikanzniveau von $\alpha = 0,05$ und begründen Sie Ihre Entscheidung.
- b) Für eine gezielte Standortanalyse wurde ein Teilmodell gerechnet, das ausschließlich den Standort und das Wochenende betrachtet. Folgende relevante Koeffizienten wurden ermittelt (Referenzkategorie bleibt Stadtzentrum):

Variable	Koeffizient
Standort [Park]	120,00
Standort [Strand]	290,50
Wochenende	96,00

- i. (2 Punkte) Wie ändert sich der Tagesumsatz im Durchschnitt, wenn der Eisstand im Park anstatt am Strand aufgestellt wird und alle anderen Variablen gleich bleiben?
- ii. (2 Punkte) Wie ändert sich der erwartete Tagesumsatz im Durchschnitt, wenn der Eisstand vom Stadtzentrum an den Strand verlegt wird und statt an einem Werktag an einem Wochenende betrieben wird, während alle anderen Variablen unverändert bleiben?

Lösung:

- a) i. 1.135,26
- ii. Je höher die Windgeschwindigkeit, desto tendenziell niedriger der Umsatz.
- iii. Analysten-Vermutung widerlegt; $p\text{-Wert} = 0,024 < 0,05$; Luftfeuchtigkeit ist daher signifikant.
- b) i. -170,5; sinkt
- ii. 386,5

8. (10 Punkte)

a) Welche Aussagen sind korrekt? Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen:

- ☐ Differenzen sind bei ordinalskalierten Daten interpretierbar.
- ☐ Verhältnisskalen besitzen einen natürlichen Nullpunkt.
- ☐ Metrische Daten sind immer stetig.
- ☐ Mittelwerte sind bei nominalen Daten sinnvoll interpretierbar.
- ☐ Bei intervallskalierten Daten sind Differenzen interpretierbar.

b) Welche Aussagen zur Kovarianz $\text{Cov}(X, Y)$ sind korrekt? Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen:

- ☐ Die Kovarianz besitzt immer dasselbe Vorzeichen wie der Pearsonsche Korrelationskoeffizient.
- ☐ Eine Kovarianz von 0 bedeutet, dass kein Zusammenhang zwischen X und Y besteht.
- ☐ Die Kovarianz ist normiert zwischen -1 und 1 .
- ☐ Eine positive Kovarianz bedeutet, dass große Werte von X tendenziell mit großen Werten von Y einhergehen.
- ☐ Die Kovarianz ist unabhängig von den Einheiten der Variablen.

c) Ergänzen Sie die Textlücken so, dass jedenfalls eine wahre Aussage entsteht.

Ist der p -Wert kleiner als das Signifikanzniveau α , so wird die -hypothese

.

d) Ergänzen Sie die Textlücken im nachstehenden Satz durch Ankreuzen des jeweils zutreffenden Ausdrucks aus den nachstehenden Tabellen so, dass jedenfalls eine wahre Aussage entsteht.

Eine diskrete Zufallsvariable X beschreibt die Anzahl der defekten Teile in einer Stichprobe, F ist die zugehörige Verteilungsfunktion. Für die Wahrscheinlichkeit, dass mehr als 2, aber höchstens 5 Teile defekt sind, gilt allgemein:

$$P(2 < X \leq 5) = \textcircled{1} - \textcircled{2}$$

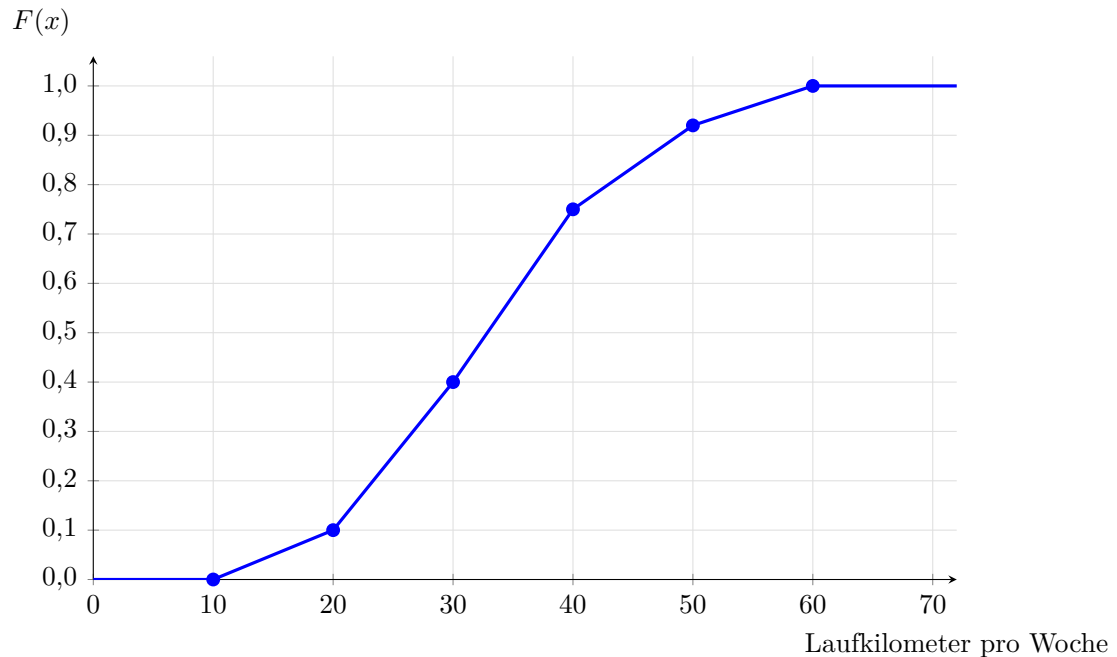
①	②
☐ $F(4)$	☐ $F(1)$
☐ $F(5)$	☐ $F(2)$
☐ $1 - F(5)$	☐ $F(3)$

e) Welche zwei Aussagen beschreiben die Eigenschaften einer Binomialverteilung mathematisch korrekt? Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.

- ☐ Sie beschreibt die Anzahl der Versuche bis zum ersten Erfolg.
- ☐ Sie beschreibt die Anzahl der Erfolge in einer festen Anzahl von n Versuchen.
- ☐ Die Wahrscheinlichkeit für einen Erfolg ändert sich von Versuch zu Versuch.
- ☐ Die Erfolgswahrscheinlichkeit p muss bei jedem der n Versuche konstant bleiben.
- ☐ Die einzelnen Versuche müssen voneinander abhängig sein.

Lösung: (2, 5); (1, 4); (Alternativ-, bestätigt ODER Null-, verworfen); (2, 2); (2, 4);

9. (10 Punkte) Um herauszufinden, wie viele Laufkilometer pro Woche zurückgelegt werden, wurden 800 Personen befragt. Die Ergebnisse der Umfrage sind in folgender approximierender Verteilungsfunktion $F(x)$ grafisch dargestellt:



- a) Welches Skalenniveau hat das erhobene Merkmal in dieser Umfrage?
- b) Kreuzen Sie die zutreffende Aussage an. Die Verteilungsfunktion in der Grafik gehört zu
- ☐ gewichteten Daten
 - ☐ unklassierten Daten
 - ☐ klassierten Daten
 - ☐ gruppierten Daten
- c) Welche maximale Laufleistung wurde von 40 % der Befragten angegeben?
- d) Wie viele Personen laufen zwischen 20 km und 30 km pro Woche?
- e) Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.
- ☐ Für diese Verteilungsfunktion wurde unterstellt, dass die Laufkilometer pro Woche in den einzelnen Klassen jeweils gleichverteilt sind.
 - ☐ Eine Person ist genau 5 km pro Woche gelaufen.
 - ☐ Der Median der Laufkilometer pro Woche liegt im Intervall $[30; 40[$
 - ☐ Die Spannweite der Verteilung beträgt 70.
 - ☐ Drei Viertel der Befragten gaben an mehr als 40 km pro Woche zu laufen.

Lösung: metrisch, 3, 30, 240, (1, 3)