

# Statistik Vorlesung

01. Februar 2024

Dauer der Prüfung: 100 Minuten

ZUNAME:			
VORNAME:		MATR.NR.:	

ERLAUBT: Skriptum des Instituts, nicht-graphikfähiger Taschenrechner

VERBOTEN: alle sonstigen Unterlagen, graphikfähiger Taschenrechner, 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	28	
6	10	
Summe	100	
Note		

1. (12 Punkte)

a) Die Varianz wird jedenfalls größer, wenn die Stichprobe um eine weitere Beobachtung ergänzt wird.

Richtig

Falsch

b) Beträgt der Pearson'sche Korrelationskoeffizient zwischen zwei Merkmalen  $-0,9$ , so ist der lineare Zusammenhang zwischen diesen Merkmalen nur sehr gering.

Richtig

Falsch

c) Ist die Zufallsgröße  $X$  normalverteilt, so ist der Mittelwert einer Stichprobe von  $X$  ebenfalls normalverteilt.

Richtig

Falsch

d) Die Dichtefunktion einer stetigen Zufallsvariablen ist stets stetig.

Richtig

Falsch

e) Wenn der  $t$ -Test für Differenzen signifikant ist, bedeutet dies, dass die Erwartungswerte der beiden betrachteten Datenreihen signifikant anders sind.

Richtig

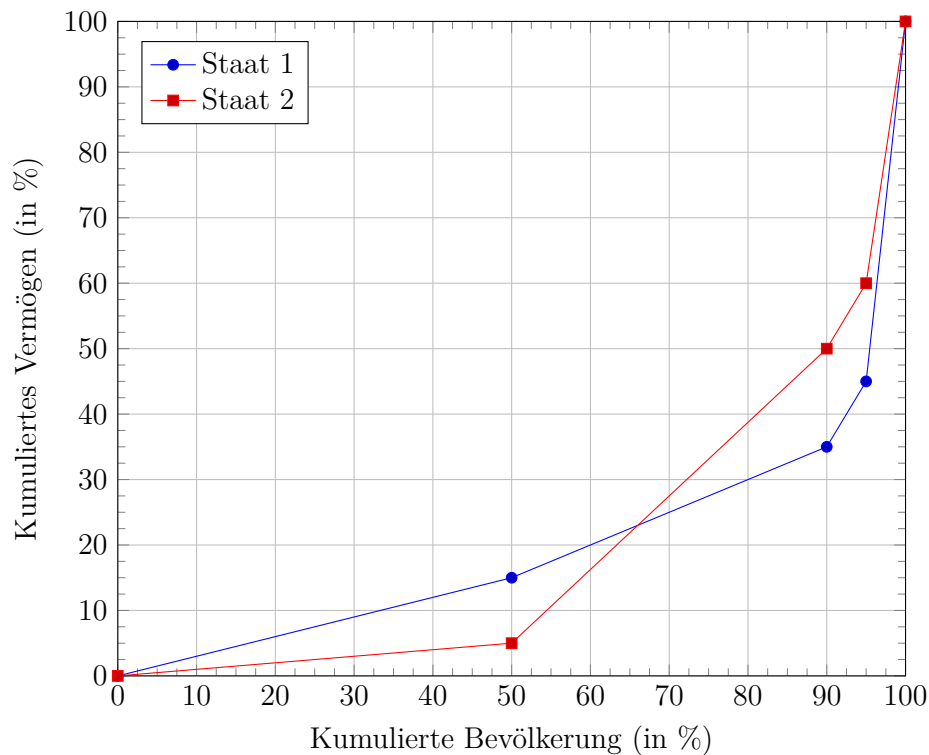
Falsch

f) In der einfachen linearen Regression gibt der Achsenabschnitt den Schätzwert für  $Y$  an, wenn die erklärende Variable gleich 0 ist.

Richtig

Falsch

2. (10 Punkte) Um die Vermögenskonzentration zweier Staaten zu vergleichen, soll die Verteilung des Vermögens auf die Bevölkerung analysiert werden. Für eine graphische Darstellung anhand von Lorenzkurven wurde die Bevölkerung in jeweils 4 Gruppen unterteilt: Gruppe I repräsentiert den Teil der Bevölkerung mit dem geringsten Vermögen, Gruppe II repräsentiert die Mittelschicht, Gruppe III die obere Mittelschicht, und in Gruppe IV befindet sich der reichste Anteil der Bevölkerung.



Welche der folgenden Aussagen sind richtig?

- a) In beiden Staaten sind 5 % der Bevölkerung der Gruppe III zugeordnet.  
 Richtig  Falsch
- b) Die reichsten 5 % der Bevölkerung besitzen in beiden Staaten jeweils mehr als 50 % des Vermögens.  
 Richtig  Falsch
- c) Der Anteil am Gesamtvermögen, der auf Gruppe II der Bevölkerung fällt, ist in beiden Ländern gleich groß.  
 Richtig  Falsch
- d) Der Anteil, den die unteren 50 % am Gesamtvermögen besitzen, ist im Staat 1 höher als im Staat 2.  
 Richtig  Falsch
- e) Der Gini-Koeffizient ist in Staat 1 eindeutig größer.  
 Richtig  Falsch

3. a) Eine Bäckerei möchte mit ihren Faschingskrapfen an der jährlichen Verkostung teilnehmen. Aufgrund Personalmangels befüllt aber der Großvater eines Bäckers händisch die Krapfen mit Marillenmarmelade. Um sich bei der Verkostung nicht zu blamieren, werden 11 Krapfen hinsichtlich der Füllmenge von Marillenmarmelade überprüft und folgende Daten erhoben:

Füllmenge [in g]	5	3	8	31	2	15	22	14	10	8	7
------------------	---	---	---	----	---	----	----	----	----	---	---

- i. (9 Punkte) Bestimmen Sie die für einen Boxplot relevanten Größen und zeichnen Sie diesen!
  - ii. (2 Punkte) Interpretieren Sie diesen Boxplot hinsichtlich der Schiefe der Daten!
- Ausführung Beispiel 3a:

- b) (9 Punkte) Ein verliebtes Pärchen nimmt an dieser Verkostung teil und bewertet die Krapfen nach der bekannten Notenskala (1 bedeutet, der Krapfen ist sehr gut, 5 bedeutet, der Krapfen ist ungenießbar). Die folgende Tabelle gibt Auskunft über die erhobenen Daten:

Bäckerei	A	B	C	D	E	F	G	H	I	J
Noten Tester <i>A</i>	1	2	5	4	3	1	2	2	3	4
Noten Testerin <i>B</i>	2	1	5	4	3	2	3	2	4	3

Haben die beiden Verliebten auch hinsichtlich der Krapfen den gleichen Geschmack? Mit welcher Kennzahl kann ein Zusammenhang zwischen den beiden Merkmalen am besten dargestellt werden? Berechnen Sie diese und interpretieren Sie Ihr Ergebnis!

Ausführung Beispiel 3b:

4. a) Das Organisationskomitee für den Frühlingsball 2024 möchte weiße und rote Tulpen als Blumenschmuck. Für die Fahrt zur Gärtnerei benützt der Angestellte das Auto. Für die Hinfahrt wählt er die längere Strecke über die Autobahn, für die Rückfahrt benützt er die Bundesstraße. Die Fahrzeiten auf diesen beiden Strecken werden als unabhängige normalverteilte Größen betrachtet, wobei:

Fahrzeit (in Stunden) für die Hinfahrt über die Autobahn:  $N(1; 0,2)$

Fahrzeit (in Stunden) für die Rückfahrt über die Bundesstraße:  $N(1; 0,5)$

- i. (2 Punkte) Wie ist die Gesamtfahrzeit (Hin- und Rückfahrt)  $X$  verteilt? Geben Sie die Parameter dieser Verteilung an!
- ii. (3 Punkte) Mit welcher Wahrscheinlichkeit ist der Angestellte insgesamt höchstens 2 Stunden aber mindestens 1,75 Stunden unterwegs?
- iii. (3 Punkte) Bestimmen Sie zu dieser Verteilung das 0,9-Quantil und interpretieren Sie dieses!

Ausführung Beispiel 4a:

- b) Endlich bei der Gärtnerei angekommen, kauft der Angestellte 110 rote und 90 weiße Tulpenzwiebel. Die roten Tulpen werden mit einer Wahrscheinlichkeit von 90 % bis zum Ball blühen, die weißen mit einer Wahrscheinlichkeit von 65 %.
- i. (2 Punkte) Stellen Sie den Sachverhalt in einem geeigneten Wahrscheinlichkeitsbaum dar!
  - ii. (2 Punkte) Mit welcher Wahrscheinlichkeit wird eine zufällig ausgewählte Tulpenzwiebel bis zum Ball blühen?
  - iii. (2 Punkte) Wie viele blühende Tulpen kann der Angestellte bei den weißen Tulpenzwiebeln erwarten?
  - iv. (3 Punkte) Wie wahrscheinlich ist es, dass genau 105 rote Tulpen blühen werden?
  - v. (3 Punkte) Mit welcher Wahrscheinlichkeit ist eine Zwiebel, die beim Ball nicht geblüht hat, die einer roten Tulpe?

Ausführung Beispiel 4b:

5. a) Ein Unternehmen füllt Getränke auf unterschiedlichen Abfüllanlagen in Flaschen ab. Um zu überprüfen, ob die Toleranzen der Füllmengen eingehalten werden, werden regelmäßig Stichproben von befüllten Flaschen entnommen und deren Inhalte gemessen. Die Toleranzen garantieren sowohl eine Mindest- als auch eine Höchstfüllmenge. Sämtliche Messergebnisse sind annähernd normalverteilt. Die Messergebnisse einer Stichprobe der Abfüllanlage  $X$  sind:

Flascheninhalte in Milliliter (Anlage $X$ )	994	1.000	1.012	987	1.007
---	-----	-------	-------	-----	-------

Der Mittelwert dieser Daten beträgt  $\bar{x} = 1.000$ .

- i. (2 Punkte) Bestimmen Sie die zugehörige *Stichprobenvarianz* und runden Sie das Ergebnis ganzzahlig und verwenden Sie diesen **gerundeten** Wert in den nachfolgenden Aufgaben!
- ii. (8 Punkte) Bestimmen Sie ein geeignetes *90%-Konfidenzintervall* (ganzzahlig gerundet), begründen Sie Ihre Wahl und interpretieren Sie das Ergebnis!
- iii. (8 Punkte) Das Unternehmen möchte die Hypothese  $H_0 : \mu = 1.000$  anhand der durchgeführten Stichprobe testen. Welcher Test kommt zur Anwendung und wie lautet der Testwert? Wählen Sie ein geeignetes Signifikanzniveau und führen Sie den Test durch! Ändert sich Ihre Testentscheidung, je nachdem, welches Signifikanzniveau  $\alpha$  Sie wählen? Begründen Sie!

Ausführung Beispiel 5a:



- b) Es wird nun eine *zweite Stichprobe* von fünf Flaschen ( $n_y = 5$ ) der Anlage  $Y$  entnommen mit  $\bar{y} = 1.010$  und  $s_y^2 = 100$ . Nehmen Sie außerdem an, dass  $s_x^2 = 100$ !
- i. (7 Punkte) Prüfen Sie mit einem geeigneten Test auf *Varianzhomogenität* der beiden Stichproben mit Signifikanzniveau 0,1. *Hinweis:*  $F_{4;4;0,95} = 6,388$ .
  - ii. (3 Punkte) Welcher Test ist durchzuführen, wenn Sie zeigen wollen, dass mehr Flascheninhalt von Anlage  $Y$  als von Anlage  $X$  zu erwarten ist? Bestimmen Sie die Hypothesen, aber führen Sie den Test nicht durch!

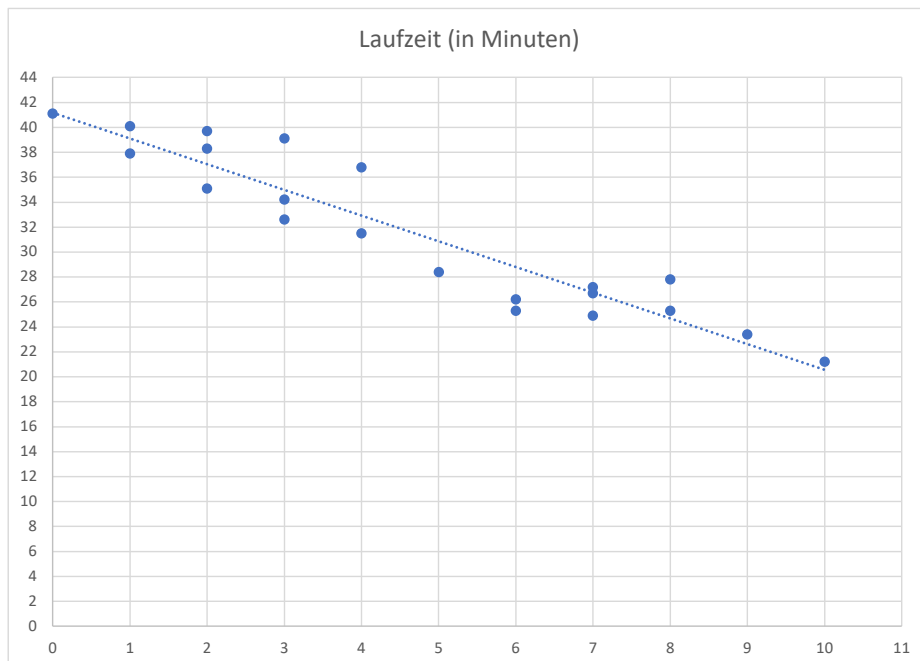
Ausführung Beispiel 5b:

6. Es soll untersucht werden, ob und wie sich die Anzahl an Stunden Lauftraining pro Woche auf die Laufzeit über 5 km auswirkt. Dazu wurden die Trainingsstunden pro Woche (Merkmal  $X$ ) und die 5-km-Laufzeiten in Minuten (Merkmal  $Y$ ) von 21 männlichen Schülern erhoben.

Die Analyse ergab folgende Kleinste-Quadrate-Schätzung:

	Koeffizienten	Standardabweichung des Koeffizienten
Achsenabschnitt	41,189	0,940
Trainingsstunden	-2,063	0,172

- (2 Punkte) Interpretieren Sie den Koeffizienten des Merkmals „Trainingsstunden“ im gegebenen Kontext!
- (2 Punkte) Welche geschätzte Laufzeit würde sich bei 5 Stunden Lauftraining pro Woche auf Basis des Modells ergeben?
- (2 Punkte) Die untenstehende Grafik zeigt die geschätzte Regressionsgerade sowie die tatsächlich gemessenen Datenpunkte. Zeichnen Sie in dieser Grafik die geschätzte Laufzeit bei 5 Stunden Lauftraining ein und kennzeichnen Sie das Residuum!
- (4 Punkte) Welche Hypothesen sind aufzustellen, um zu prüfen, ob der Einfluss des Merkmals „Trainingsstunden“ signifikant ist? Berechnen Sie den Testwert eines geeigneten Tests und treffen Sie eine Entscheidung, wenn  $K = ] - \infty; -2,093[ \cup ]2,093; \infty[$ !



Ausführung Beispiel 6:

Ausführung Beispiel 6: