

# Statistik Vorlesung

3. Februar 2023

Dauer der Prüfung: 100 Minuten

ZUNAME:			
VORNAME:		MATR.NR.:	

ERLAUBT: Skriptum des Instituts, nicht-graphikfähiger Taschenrechner

VERBOTEN: alle sonstigen Unterlagen, graphikfähiger Taschenrechner, Handys

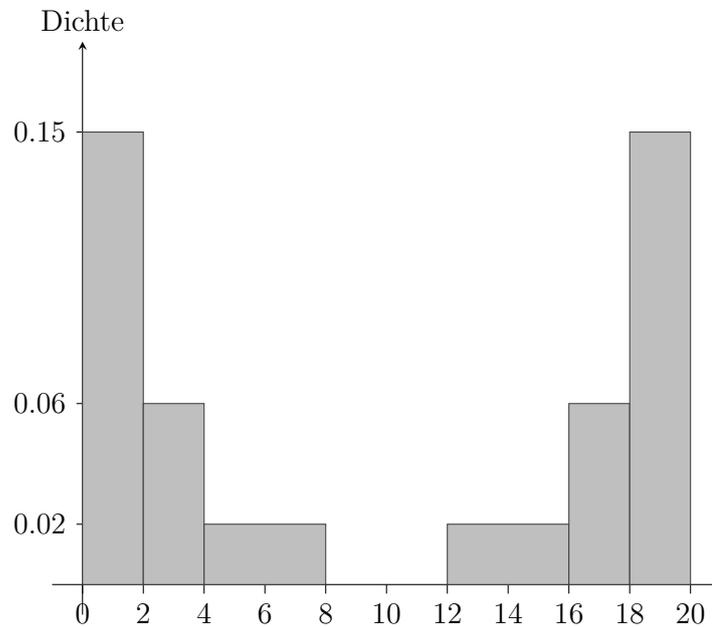
Bei den Single-Choice-Fragen bringt eine richtige Antwort 2 Punkte und eine falsche 1 Punkt Abzug.  
Es gibt keine negative Punktemitnahme in ein anderes Beispiel.

Aufgabe	max. Punkte	erreichte Punkte
1	12	
2	10	
3	20	
4	20	
5	30	
6	8	
Summe	100	
Note		

1. (12 Punkte)

- a) Wenn Daten Ausreißer enthalten, sollte als Lagemaß eher der Median als der Mittelwert verwendet werden.  
 Richtig  Falsch
- b) Für alle Daten liegt der Pearson'sche Korrelationskoeffizient im Intervall  $[-1; 1]$ .  
 Richtig  Falsch
- c) Gegeben seien die Zufallsvariablen  $X$  und  $Y$  mit den Erwartungswerten 3 und 7. Dann hat die Zufallsvariable  $Z = 3X + 7Y$  einen Erwartungswert von 58.  
 Richtig  Falsch
- d) Eine Wahrscheinlichkeit  $P(x_1 \leq X \leq x_2)$  ist bei einer Dichtefunktion als Fläche zwischen der Funktion, der x-Achse und den jeweiligen Grenzen ( $x_1$  und  $x_2$ ) gegeben.  
 Richtig  Falsch
- e) Ist der p-Value kleiner als die Irrtumswahrscheinlichkeit  $\alpha$ , so muss die  $H_0$  beibehalten werden.  
 Richtig  Falsch
- f) Die Regressionsgerade  $Statistikpunkte = 22 + 0,2 \cdot Mathematikpunkte$  wurde aus den Klausurergebnissen von 100 BWL-Studierenden ermittelt. Wenn der Mittelwert der Mathematikpunkte 30 war, musste der Mittelwert der Statistikpunkte 28 gewesen sein.  
 Richtig  Falsch

2. (10 Punkte) Nachdem der Hauptdarsteller in der Serie „Fernweh“ durch einen anderen Schauspieler ersetzt wurde, möchte der Filmproduzent anhand einer Umfrage unter Stammseherinnen über 50, die die Serie bisher regelmäßig mitverfolgt haben, herausfinden, wie die Zuseherinnen die neue Serie finden. Es wird erhoben, wie viele der ausgestrahlten 20 neuen Folgen nach dem Schauspielerwechsel von den Zuseherinnen angesehen wurden. Die folgende Graphik zeigt das aus den Daten generierte Histogramm:



- a) 30 % der Befragten haben 18 Folgen oder mehr gesehen.  
 Richtig  Falsch
- b) Die Höhe der Stäbe gibt die absolute Häufigkeit an, mit der die Befragten eine bestimmte Anzahl von Folgen gesehen haben.  
 Richtig  Falsch
- c) Das Histogramm lässt schließen, dass die Befragten 2 Gruppen bilden, von denen eine eher nach dem Schauspielerwechsel die Serie nicht mehr so häufig oder gar nicht mehr anschaut.  
 Richtig  Falsch
- d) Das Merkmal „Anzahl der Folgen“ ist ein stetiges Merkmal.  
 Richtig  Falsch
- e) Das Histogramm ist falsch gezeichnet, weil auf der x-Achse eigentlich die Merkmalsausprägung und auf der y-Achse die relative Häufigkeit stehen müsste.  
 Richtig  Falsch

3. a) (10 Punkte) Ein Stahlwerk produziert auf sogenannten Stranggussanlagen Stahlvorblöcke für die Schienen- und Fahrzeugindustrie. Der Prozessschritt der Roheisenentschwefelung wird halbautomatisch durchgeführt, d.h. es gibt einen menschlichen Benutzer der eine Maschine steuert. Für die Dauer des Prozessschrittes liegen folgende Messdaten in Sekunden vor:

13 | 15 | 10 | 14 | 10 | 16 | 22 | 9

1. Bestimmen Sie die benötigten Größen und zeichnen Sie den Boxplot!
2. Bestimmen Sie anhand der Daten alle drei Werte der zentralen Tendenz (Modalwert, Median, Mittelwert)! Welcher dieser Werte beschreibt die Lage der Daten am besten, und warum?

Ausführung Beispiel 3:

- b) (10 Punkte) Die Prozessdauer für die Roheisenentschwefelung im Stahlwerk wurde nun in zwei Kategorien (kurz und lang) erfasst, und es wurden in weiterer Folge erneut 100 Messungen durchgeführt. Diese wurden kategorisiert und dabei wurde notiert, ob die Messung während Nacht- oder Tagschicht gemacht wurde.

<i>Schicht</i>	<i>Prozessdauer</i>	
	kurz	lang
Tagschicht	16	35
Nachtschicht	31	18

1. Welche Skalenniveaus haben hier die Variablen *Prozessdauer* und *Schicht*?
2. Berechnen Sie das geeignete Zusammenhangsmaß und interpretieren Sie dieses!

Ausführung Beispiel 3:

4. Die Anzahl an Einbrüchen in einer Stadt kann als annähernd poissonverteilt angenommen werden. Im Mittel werden **pro Woche** 1.5 Einbrüche verübt.
- a) (3 Punkte) Wie groß ist die Wahrscheinlichkeit, dass in einer Woche kein Einbruch verübt wird?
  - b) (4 Punkte) Wie groß ist die Wahrscheinlichkeit von mindestens 2 Einbrüchen in einer Woche?
  - c) (4 Punkte) Wie groß ist die Wahrscheinlichkeit dafür, dass in drei zufällig ausgewählten Wochen jeweils mindestens zwei Einbrüche begangen wurden?
  - d) (4 Punkte) Wie viele Einbrüche sind in dieser Stadt pro Jahr (52 Wochen) zu erwarten? Begründen Sie ausführlich!
  - e) (5 Punkte) Die Wartezeit bis zum ersten Einbruch ist exponentialverteilt mit dem gleichen Parameter wie für die Aufgaben a) und b). Wie lange dauert es im Schnitt, bis der erste Einbruch eines Jahres begangen wird? Geben Sie das Resultat **in Tagen** an!

Ausführung Beispiel 4:

Ausführung Beispiel 4:

5. Ein Landwirt betreibt nebegewerblich einen 24-Stunden-Lebensmittel-Shop. Unter anderem befüllt er seine Automaten täglich mit selbstgebackenen frischen Brotlaiben. Im Zuge einer Qualitätskontrolle wurden 6 Brotlaibe zufällig ausgewählt und auf deren Gewicht kontrolliert. Die Messungen ergaben folgende Werte:

Brot	1	2	3	4	5	6
Gewicht [g]	1041	992	1009	1095	998	1006

Das Gewicht der Brotlaibe kann als normalverteilte Zufallsgröße angesehen werden.

- a) (10 Punkte) Bestimmen Sie ein zweiseitiges 90%-Intervall für das durchschnittliche Brotlaibgewicht und interpretieren Sie!
- b) (20 Punkte) Der Landwirt behauptet, dass die Brotlaibe seines Nachbarn signifikant leichter sind als seine. Heimlich besorgt er sich sieben Brotlaibe seines Nachbarn und wiegt diese. Er erhält dabei einen Mittelwert von 995 g bei einer Standardabweichung von 41 g. Können Sie die Behauptung des Landwirts bestätigen? Auch die Brotlaibgewichte des Nachbarn können als normalverteilt angenommen werden. Das Signifikanzniveau beträgt zehn Prozent.
1. Welcher Test ist anzuwenden? Begründen Sie ausführlich!
  2. Wie sind die Hypothesen zu formulieren?
  3. Bestimmen Sie den kritischen Bereich!
  4. Wie entscheiden Sie?
  5. Wie kann das Ergebnis interpretiert werden?

Ausführung Beispiel 5:

Ausführung Beispiel 5:

6. Der Fachverband der Holzindustrie möchte herausfinden, wie hoch der Abschlag vom aktuellen Preis ist, den Sägewerke für Drehwuchs in Abhängigkeit der Stammlänge verrechnen. Das Ergebnis der Regressionsanalyse anhand von 102 Beobachtungen ist der folgenden Tabelle zu entnehmen.

	Koeffizient	Standardfehler
Schnittpunkt	4,871	0,401
Länge [Meter]	1,245	0,474

- a) (2 Punkte) Wie hoch ist der Abschlag für Drehwuchs bei 5 Meter langen Stämmen im Mittel?
- b) (6 Punkte) Wie hoch ist der p-Wert für den Koeffizienten des Merkmals „Länge“ im Modell?

Ausführung Beispiel 6: