

# Statistik Vorlesung

17. März 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) Für alle Daten gilt: Der Minimalwert ist kleiner als der Mittelwert.

Richtig  Falsch

b) Der korrigierte Kontingenzkoeffizient zwischen den Merkmalen „Wolkengattung“ und „Windgeschwindigkeit“ betrage 0,85. Dies bedeutet, dass Gewitterwolken tendenziell mit höheren Windgeschwindigkeiten einhergehen.

Richtig  Falsch

c) Bei einer diskreten Zufallsvariablen  $X$ , definiert auf den ganzen Zahlen, gilt stets:

$$P(-2 < X \leq 10) = F(10) - F(-3)$$

Richtig  Falsch

d) Eine Verteilungsfunktion kann keine Funktionswerte über 1 annehmen.

Richtig  Falsch

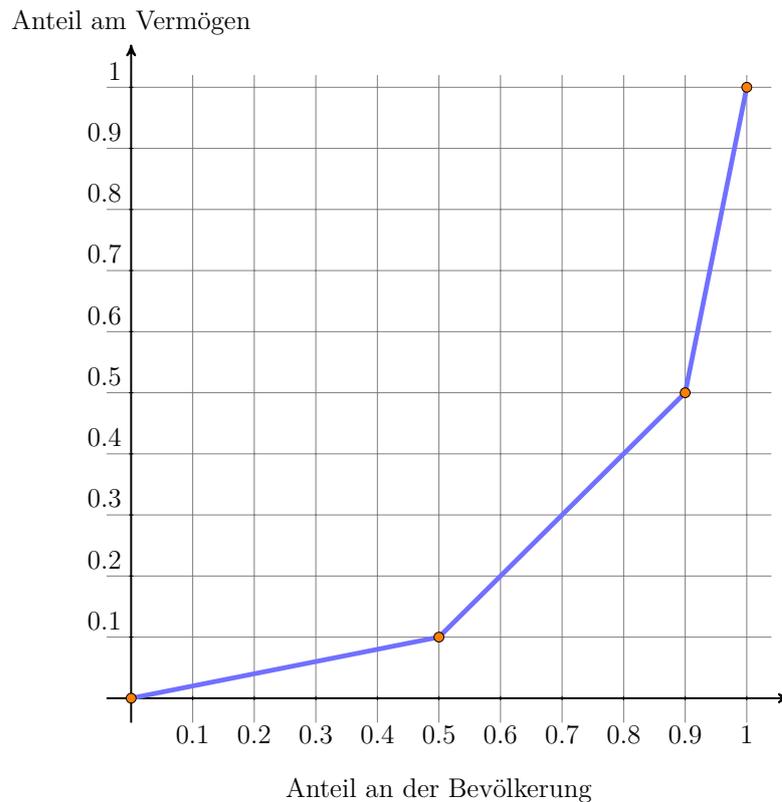
e) Die Stichprobenvarianz ist chi-Quadrat-verteilt mit  $n-1$  Freiheitsgraden.

Richtig  Falsch

f) In einem Regressionsmodell  $Ausgaben = 3 + 0,2 \cdot \log(Einkommen)$  bedeutet 0,2, dass die Ausgaben um 0,2 % steigen, wenn das Einkommen um 1 % steigt.

Richtig  Falsch

2. (10 Punkte) Durch die folgende Grafik wird die Vermögensverteilung in einem Land beschrieben. Welche Aussagen treffen zu?



- a) Die reichsten 10 % der Einwohner besitzen 50 % des gesamten Vermögens des Landes.  
 Richtig  Falsch
- b) Das gesamte Vermögen des Landes verteilt sich zu gleichen Anteilen auf drei Gruppen.  
 Richtig  Falsch
- c) Der Gini-Koeffizient entspricht hier der Fläche unterhalb der Lorenzkurve.  
 Richtig  Falsch
- d) Die Lorenzkurve ist immer monoton steigend.  
 Richtig  Falsch
- e) Die ärmsten 10 % der Einwohner des Landes besitzen 50 % des Vermögens.  
 Richtig  Falsch

3. a) (10 Punkte) Ostern naht und laut einer Studie ist zu befürchten, dass die Österreicher heuer weniger Osterhasen verschenken werden. Es stellt sich die Frage, wie viele Osterhasen voraussichtlich pro Person verschenkt werden. Dazu wurden 250 Personen befragt. Die Ergebnisse der Umfrage sind in folgender Tabelle zusammengefasst:

Anzahl der Osterhasen	0	1	2	3	4	5
Absolute Häufigkeit	10	25	25	50	80	60

1. Zeichnen Sie die geeignete Verteilungsfunktion! Bestimmen und interpretieren Sie den Wert  $F(2)$ !
2. Bestimmen Sie anhand der Daten alle drei Werte der zentralen Tendenz! Welcher dieser Parameter beschreibt die Lage der Daten am besten und warum?

Ausführung Beispiel 3a:

- b) (10 Punkte) Die Schoko-Osterhasen sind zu Ostern Pflichtprogramm! Es gibt sie in allen Farben, Größen und Preisklassen. Es wurden 6 Milkschokolade-Osterhasen zu je 100 g einem Geschmackstest unterzogen. Die folgende Tabelle gibt über den Verkaufspreis [in EURO/Osterhase] und die erzielten Punkte des Geschmackstests Auskunft:

Osterhase	A	B	C	D	E	F
Preis in EURO	4,19	2,99	4,49	0,99	2,49	1,19
Punkte	94	93	93	91	90	89

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) (12 Punkte) In einem Grazer Haubenlokal werden täglich 30 Salatköpfe bestellt. Bei der Anlieferung wird zufällig ein Salatkopf aus der Lieferung ausgewählt und kontrolliert. Ist der Salatkopf frisch und knackig, wird ein weiterer Salat zur Kontrolle entnommen. Dieser Vorgang wird fortgeführt bis (A) die ersten fünf Salatköpfe die Kontrolle bestehen oder (B) ein unzureichend schöner Salatkopf in der Auswahl gefunden wurde. In Fall (A) wird die Lieferung angenommen, in Fall (B) geht die gesamte Bestellung an den Absender zurück. Gehen Sie für die folgenden Fragestellungen davon aus, dass sich zwei „schlechte“ Salatköpfe in einer Tageslieferung befinden.
1. (3 Punkte) Stellen Sie die möglichen Ereignisse des Kontrollprozesses mit deren Wahrscheinlichkeiten in einem Wahrscheinlichkeitsbaum dar!
  2. (3 Punkte) Wie groß ist die Wahrscheinlichkeit, dass die empfangene Lieferung zurückgesandt wird?
  3. (3 Punkte) Angenommen die ersten 3 kontrollierten Salatköpfe haben die Qualitätskontrolle bereits bestanden. Wie groß ist jetzt die Wahrscheinlichkeit, dass die Lieferung nach Vollendung der Qualitätskontrolle zurückgesandt wird?
  4. (3 Punkte) Wie oft im Monat ist zu erwarten, dass die Sendung zurückgeschickt wird? (Annahme: 1 Monat = 30 Tage)

Ausführung Beispiel 4a:

- b) (8 Punkte) Das Gewicht eines Salatkopfs ist mit  $\mu = 0,42$  kg und  $\sigma = 0,05$  kg normalverteilt.
1. (4 Punkte) Wie viel Prozent der Salate wiegen maximal 0,45 kg?
  2. (4 Punkte) Innerhalb welcher um den Erwartungswert symmetrischen Toleranzgrenzen befindet sich das Gewicht von 90 % der Salate?

Ausführung Beispiel 4b:

5. a) Ein Autohersteller führt bei seinen Mittelklassewagen ein Softwareupdate durch, um das Einhalten von gesetzlich vorgeschriebenen  $\text{NO}_x$  Grenzwerten sicherzustellen. Eine Umweltorganisation misst die  $\text{NO}_x$ -Emissionen bei zufällig ausgewählten PKWs eines bestimmten Modells **jeweils** vor und nach dem Softwareupdate (Angaben in mg/km):

X = ohne Update	1090	1117	1118	1076	1080	1146	1108	1139
Y = mit Update	1077	1079	1081	1057	1014	1029	1083	1119

Die Umweltorganisation möchte wissen, ob das Softwareupdate tatsächlich zu einer Reduktion der  $\text{NO}_x$ -Emissionen im Realbetrieb geführt hat. Die Messwerte können als normalverteilt angesehen werden.

1. (2 Punkte) Wie lauten die Hypothesen der Umweltorganisation?
2. (2 Punkte) Welcher Test kommt zur Anwendung? Begründen Sie ausführlich!
3. (11 Punkte) Führen Sie den Test durch! Wie lautet die Testentscheidung und wie ist diese zu interpretieren (Signifikanzniveau 5 %)?

Ausführung Beispiel 5a:

- b) Eine weitere Umweltorganisation misst ebenso die  $\text{NO}_x$ -Emissionen von zufällig ausgewählten PKWs desselben Modells wie in Aufgabe 5.a). Allerdings werden die Messungen an acht PKWs vor und an anderen acht PKWs nach dem Softwareupdate durchgeführt. Die Messungen können hier nicht als normalverteilt angesehen werden.

X = ohne Update	1099	1077	1073	1047	1067	1046	1028	1034
Y = mit Update	920	1028	1000	970	1004	982	967	1021

1. (2 Punkte) Wie lauten die Hypothesen der Umweltorganisation hinsichtlich der  $\text{NO}_x$ -Emissionen?
2. (2 Punkte) Welcher Test kommt zur Anwendung? Begründen Sie ausführlich!
3. (11 Punkte) Führen Sie den Test durch! Wie lautet die Testentscheidung und wie ist diese zu interpretieren (Signifikanzniveau 5 %)??

Ausführung Beispiel 5b:

6. Ein Dachziegelproduzent analysiert seine Umsätze und hat dazu einige Daten aus seinen Aufzeichnungen herangezogen. Er schätzt seinen Umsatz pro Woche (in Euro) aus den Variablen **Anzahl der Kunden**, **Einkaufspreis für Zement** (in Euro pro Sack) und der **10-Tages-Wettervorhersage** mit den Ausprägungen regnerisch, wechselhaft und trocken. Das Ergebnis der von ihm durchgeführten Regressionsanalyse wird in der folgenden Tabelle zusammengefasst:

	Koeffizient	Standardfehler	t-Statistik	p-Wert
Schnittpunkt	211	87	2,425	0,06
Kundenanzahl	432	42,4	10,189	0,00
Zementpreis	-12	7,3	-1,644	0,18
regnerisch	-127	105	-1,209	0,22
wechselhaft	-68	243	-0,28	0,62

- (3 Punkte) Wie viel Umsatz erwartet er in einer Woche mit 200 Kunden bei einem Zementpreis von 4,20 Euro pro Sack, wenn trockenes Wetter vorhergesagt wird?
- (3 Punkte) Wie ändert sich, bei gleichem Zementpreis, der Umsatz im Schnitt, wenn das Wetter regnerisch angesagt ist und 15 Kunden weniger kommen?
- (2 Punkte) Welche Variablen sind auf dem 5%-Niveau signifikant? Begründen Sie!

Ausführung Beispiel 6: