

Statistik Vorlesung

27. September 2022

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	22	
4	16	
5	32	
6	8	
Summe	100	
Note		

1. (12 Punkte)

a) Bei einem Quartilsabstand von 0 stimmt der Median mit dem ersten Quartil überein.

Richtig Falsch

b) Für alle Daten gilt: $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

Richtig Falsch

c) Die Zufallsgröße „Augenzahl eines fairen Würfels“ ist stetig gleichverteilt auf dem Intervall $[1,6]$

Richtig Falsch

d) Die Wahrscheinlichkeit $P(X \leq x)$ kann bei einer Dichtefunktion an der y -Achse abgelesen werden.

Richtig Falsch

e) Ist beim Gauß-Test die Testgröße groß (z.B. $t_0 > z_\gamma$), ist signifikant bestätigt, dass die Daten einer Normalverteilung unterliegen.

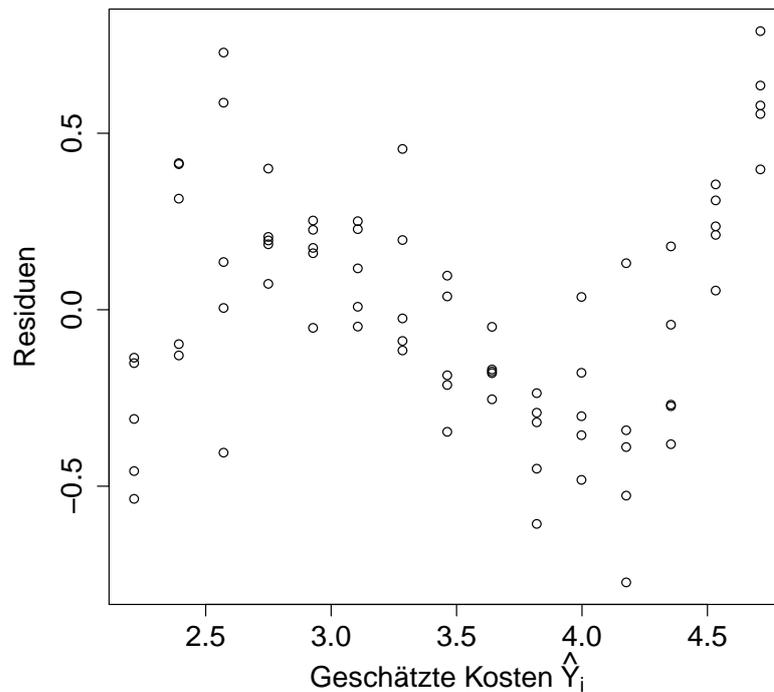
Richtig Falsch

f) Wenn die überwiegende Mehrheit der Residuen $(Y_i - \hat{Y}_i)$ positiv ist, ist die Steigung der Regressionsgeraden jedenfalls auch positiv.

Richtig Falsch

2. (10 Punkte)

Ferdinand baut Beeren in Weißrussland an, verpackt sie dort tiefgekühlt in Schachteln und importiert sie in Österreich. Für den außerhalb der EU liegenden Standort seiner Produktion schätzt er die Produktionskosten (in Euro) in Abhängigkeit der Verpackungsgröße (in Dekagramm) mittels einer Regressionsgerade aus den Aufzeichnungen seines Verwalters. Zur Modellvalidierung verwendet Ferdinand den Residuenplot. Welche Aussagen treffen zu?



- a) Da die Residuen einen polynomialen Verlauf aufweisen, hätte Ferdinand nicht eine Gerade schätzen dürfen, sondern ein Polynom 3. Grades.
 Richtig Falsch
- b) In jenen Fällen, in denen das Regressionsmodell die Kosten auf vier Euro schätzt ($\hat{Y} = 4$), waren die wahren Kosten Y tendenziell niedriger als die Schätzwerte \hat{Y} . (Hinweis: Residuen $\hat{r}_i = Y_i - \hat{Y}_i$.)
 Richtig Falsch
- c) In Ferdinands Residuenplot ist Multikollinearität der Daten deutlich ersichtlich.
 Richtig Falsch
- d) Das Muster der Residuen zeigt eindeutig, dass die Störtermvarianz σ^2 nicht konstant ist.
 Richtig Falsch
- e) Ferdinand hat richtigerweise eine lineare Regression geschätzt, aber er darf keinen Residuenplot zeichnen.
 Richtig Falsch

3. a) Wie viel kg haben Sie für Ihre „Sommerfigur 2022“ abgenommen? Diese Frage wurde 600 Personen gestellt und folgende Daten liegen dazu vor:

Gewichtsabnahme [in kg]	[0; 2[[2; 5[[5; 10[[10; 15[[15; 20]
Absolute Häufigkeit	120	150	204	96	30

1. (4 Punkte) Stellen Sie die Daten in einem entsprechenden Diagramm grafisch dar!
2. (8 Punkte) Zeichnen Sie die geeignete Verteilungsfunktion, markieren Sie darin das 0,9-Quantil und interpretieren Sie dieses!

Ausführung Beispiel 3a:

- b) (10 Punkte) Der Sommerurlaub 2022 war ein teurer Spaß, aber wie zufrieden waren die Urlauber mit dem Preis-Leistungs-Verhältnis? Dazu wurden von 6 Urlaubern jeweils die Ausgaben für den Urlaub (in EURO) sowie die Zufriedenheit erhoben. Für die Zufriedenheit wurden die Bewertung zwischen 1 (= Sehr zufrieden) und 5 (= Sehr unzufrieden) vergeben. Die folgende Tabelle gibt Auskunft über die erhobenen Daten:

Urlauber	1	2	3	4	5	6
Ausgaben in Euro	1.575	987	1.440	1.081	1.823	2.050
Zufriedenheit	4	2	3	1	4	5

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. Bei der Herstellung von Arbeitsplatten für Küchen wird auf eine Spanplatte eine Kunststoffbeschichtung aufgebracht. Die Stärke der Spanplatte X ist normalverteilt mit $\mu = 31,7$ mm und $\sigma = 0,4$ mm, die Stärke der Kunststoffschicht Y ist normalverteilt mit $\mu = 1,5$ mm und $\sigma = 0,3$ mm. Die Gesamtstärke der Arbeitsplatte $Z = X + Y$ soll zwischen 32 mm und 34 mm liegen.
- (3 Punkte) Wie ist die Gesamtstärke der Arbeitsplatte Z verteilt? Geben Sie die Parameter dieser Verteilung explizit an!
 - (4 Punkte) Wie groß ist die Wahrscheinlichkeit, dass die Gesamtstärke der Arbeitsplatte Z innerhalb der Toleranz liegt?
 - (3 Punkte) Welche Gesamtstärke wird nur von 5 % aller Platten überschritten?
 - (Unabhängig von a)–c)) Erfahrungsgemäß sind $p = 10$ % der Arbeitsplatten aus dieser Produktion fehlerhaft. Wie groß ist die Wahrscheinlichkeit, dass in einer Lieferung von $n = 30$ Stück dieser Platten
 - (3 Punkte) genau 2 defekte Platten sind,
 - (2 Punkte) mindestens eine Platte defekt ist?
 - (1 Punkt) Wie viele fehlerhafte Platten kann man im Durchschnitt in einer Lieferung von 30 Stück erwarten?

Ausführung Beispiel 4:

Ausführung Beispiel 4:

5. a) (7 Punkte) Die Meinungsforscherin Sabrina hat sich im Urlaub sehr über die Quallen-Plage im Mittelmeer geärgert und eine Umfrage gestartet. Bei einer Umfrage unter 200 Urlaubern gaben 86 % an, dass sie die Quallen als äußerst lästig empfunden haben. Bestimmen Sie das noch unten abgeschlossene 95%-Konfidenzintervall und interpretieren Sie dieses!

Ausführung Beispiel 5:

- b) (11 Punkte) Ein Kartenspiel besteht aus den vier Farben Herz, Kreuz, Karo und Pik. An lauen Abenden wurde in den Ferienunterkünften oftmals Karten gespielt. Moritz war beim Kartenspiel recht glücklos und behauptet nun, dass die Karten regelmäßig so schlecht gemischt worden seien, dass er nicht gleich viele Karten von jeder Farbe erhalten habe. Er hat genau mitgeschrieben, wie viele Karten der vier Farben er im Laufe eines Abends erhalten hat.

Farbe	Herz	Kreuz	Karo	Pik
Anzahl	64	91	75	90

1. Mit welchem Test können Sie feststellen, ob das Kartenspiel bzw. der Mischvorgang fair war, also die Wahrscheinlichkeit für jede der vier Farben gleich war?
2. Formulieren Sie dazu die Hypothesen!
Hinweis: H_0 in Form einer Wahrscheinlichkeitstabelle.
3. Führen Sie den Test zum Niveau 5 Prozent durch und nehmen Sie zu Moritz' Behauptung Stellung!

Ausführung Beispiel 5:

- c) (14 Punkte) Um Probleme mit Fehlern einer Spezialsoftware im Konzern lösen zu können, hat ein Mitarbeiter der EDV-Abteilung gleich nach dem Urlaub auf einigen zufällig ausgewählten Computern eine neue Version dieser Software installiert. Für die ausgewählten Computer hat er am Tag vor und am Tag nach dem Update notiert, wie viele Fehler jeweils aufgetreten sind.

Vorher (X) 4 3 5 2

Nachher (Y) 0 1 1 1

Können Sie zum Niveau $\alpha = 0,01$ bestätigen, dass die Fehleranzahl durch das Update geringer geworden ist, wenn bekannt ist, dass die Anzahl der Fehler als normalverteilt betrachtet werden darf?

Ausführung Beispiel 5:

6. (8 Punkte) Die Studienrichtungsvertretung möchte herausfinden, wie die Zeit für die Vorbereitung auf die Statistik-VO-Prüfung am sinnvollsten eingesetzt werden kann und erhebt unter 400 BWL-Studierenden die für die Prüfungsvorbereitung aufgewendete Zeit und die erreichte Punkteanzahl. Es wird erfasst, wie viele Stunden für das Rechnen der Beispiele auf den UE-Blättern (UEBUNGEN) und für das Rechnen von Beispielen alter Klausuren (KLAUSUREN) aufgewendet wurde. Für die Studienrichtungsvertretung ist auch interessant, um den wievielten Prüfungsantritt es sich handelt. Das Ergebnis der von ihr durchgeführten Regressionsanalyse wird in der folgenden Tabelle zusammengefasst:

	Koeffizient	Standardfehler	t-Statistik	P-Wert
Schnittpunkt	-16,81	5,24	-3,21	0
2. Antritt	-7,34	1,7	-4,33	0
3. Antritt	1,53	1,48	1,03	0,3
4. Antritt	3,47	2,16	1,61	0,11
UEBUNGEN	0,35	0,04	8,34	0
KLAUSUREN	0,58	0,15	3,86	0

- a) Studierende verwenden bei ihrem 3. Antritt zum Rechnen alter Klausuren um 20 Stunden mehr als beim 1. Antritt. Die Zeit für die Beispiele aus der Übung bleibt unverändert. Um wie viele Punkte erhalten sie beim 3. Antritt im Mittel mehr als beim 1. Antritt?
- b) Wie viele Punkte erhält eine Kandidatin beim ersten Antritt im Mittel, wenn sie 80 Stunden für das Rechnen der Übungsbeispiele und 30 Stunden für das Rechnen alter Klausuren aufwendet?

Ausführung Beispiel 6: