

# Statistik Vorlesung

13. Mai 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	20	
4	20	
5	30	
6	8	
Summe	100	
Note		

1. (12 Punkte)

a) Der Modalwert kann kleiner als das arithmetische Mittel sein.

Richtig  Falsch

b) Der korrigierte Kontingenzkoeffizient kann negativ sein.

Richtig  Falsch

c) Die Dichte einer stetigen Zufallsvariable darf keine Sprungstellen enthalten.

Richtig  Falsch

d) Der Bildbereich der Verteilungsfunktion einer Zufallsvariablen liegt im Intervall  $[0,1]$ .

Richtig  Falsch

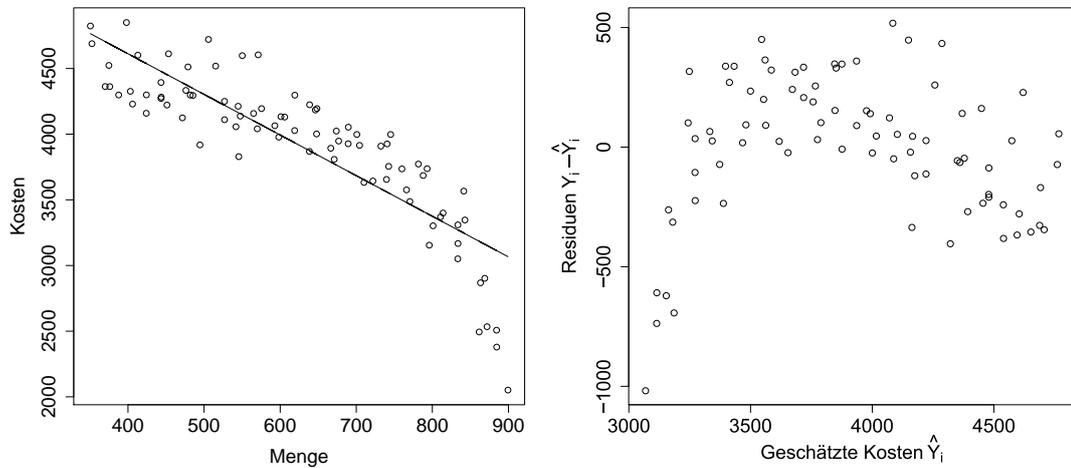
e) Je größer der Stichprobenumfang bei gegebenem  $\alpha$  ist, umso kleiner ist der  $\beta$ -Fehler.

Richtig  Falsch

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

Richtig  Falsch

2. (10 Punkte) Othilie schätzt mit Hilfe einer Regressionsanalyse die Transportkosten für die Lieferung von Faserplatten nach Nordeuropa in Abhängigkeit der gelieferten Menge. Die Graphik zeigt das Streudiagramm und die Residuen. Welche Aussagen treffen zu?



- a) Die Kovarianz zwischen Transportkosten und Liefermenge ist positiv.  
 Richtig  Falsch
- b) Sowohl sehr kleine als auch sehr große Mengen hätten vor der Regressionsanalyse entfernt werden müssen.  
 Richtig  Falsch
- c) Aus den Residuen ist ersichtlich, dass eigentlich ein Polynom 2. Grades statt der Geraden geschätzt hätte werden müssen.  
 Richtig  Falsch
- d) Die Residuen zeigen, dass für kleine Mengen die Kosten unterschätzt werden.  
 Richtig  Falsch
- e) Eigentlich hätte Othilie eine Regressionsgerade für kleine Liefermengen und eine für große Liefermengen bestimmen müssen.  
 Richtig  Falsch

3. a) (10 Punkte) Ein Großbetrieb hat erhoben, wie viele Betriebsunfälle/Jahr jeweils in den letzten Jahren angefallen sind.

18	2	36	51	25	16	22	20	22	18
----	---	----	----	----	----	----	----	----	----

1. Erstellen Sie einen Boxplot!
2. Berechnen Sie Mittelwert, Modalwert und Spannweite!
3. Wie ist das Merkmal skaliert? Ist es diskret oder stetig?

Ausführung Beispiel 3a:

- b) (10 Punkte) Durchschnittlich waren in den letzten Jahren 3.460 Frauen und 4.155 Männer beschäftigt. Von den oben angeführten 230 Unfällen passierten 150 den Männern. Sie möchten untersuchen, ob es einen Zusammenhang zwischen Unfallhäufigkeit und Geschlecht gibt. Berechnen Sie eine geeignete Kennzahl und interpretieren Sie diese!

Ausführung Beispiel 3b:

4. a) (10 Punkte) Ein Spargelanbaubetrieb setzt zur Ernte seiner Spargelfelder drei Spargelerntemaschinen ein, die vollautomatisch durch Kameraerkennung die Spargelstangen ernten. Die Hälfte der Felder wird mit Maschine *A*, 20 Prozent der Felder mit Maschine *B* und die restlichen Felder mit Maschine *C* abgeerntet. Der Anteil der krummen Spargelstangen beträgt bei Maschine *A* 0,5 Prozent, bei Maschine *B* 1,5 Prozent und bei Maschine *C* 4 %.
1. Welcher Anteil an der gesamten Spargelernte ist nicht krumm?
  2. Mit welcher Wahrscheinlichkeit stammt eine krumme Spargelstange von Maschine *B*?
  3. Wie wahrscheinlich ist es, bei einer zufälligen Überprüfung einer Spargelstange eine von Maschine *C* geerntete, nicht krumme Spargelstange zu finden?
  4. Wie wahrscheinlich ist es, eine Spargelstange zu entnehmen, die weder von *B* geerntet noch krumm ist?

Ausführung Beispiel 4a:

- b) (10 Punkte) Erdbeeren werden in 250 g Becher abgefüllt. Das Füllgewicht ist normalverteilt mit einem Erwartungswert von 265 g und einer Standardabweichung von 20 g.
1. Bestimmen Sie den Wert der Verteilungsfunktion an der Stelle 250!
  2. Mit welcher Wahrscheinlichkeit wiegt ein Becher genau 250 g?
  3. Bestimmen Sie das 0,3-Quantil dieser Verteilung!
  4. Bestimmen Sie die Zahl  $d$  derart, dass  $P(X > d) = 0,1$ ! Skizzieren Sie die Verteilungsfunktion samt Skalen auf beiden Achsen und kennzeichnen Sie darin die Lösung!

Ausführung Beispiel 4b:

5. a) (15 Punkte) Der Hersteller eines Elektroautos wirbt mit einem überraschend geringem Durchschnittsverbrauch von 10 kWh/100 km. Da Ihnen dieser Wert sehr gering vorkommt und Sie die Korrektheit dieser Aussage interessiert, leihen Sie sich das besagte E-Auto eine Woche lang aus und unternehmen jeden Tag eine Testfahrt, bei der Sie den durchschnittlichen Verbrauch auf 100 km aufzeichnen. Ihre Stichprobe liefert die folgenden Werte:

	Mo	Di	Mi	Do	Fr	Sa	So
kWh/100 km	12,1	15,9	9,1	17,7	8,9	13,1	18,9

Der Verbrauch eines Elektroautos kann als normalverteilte Größe betrachtet werden.

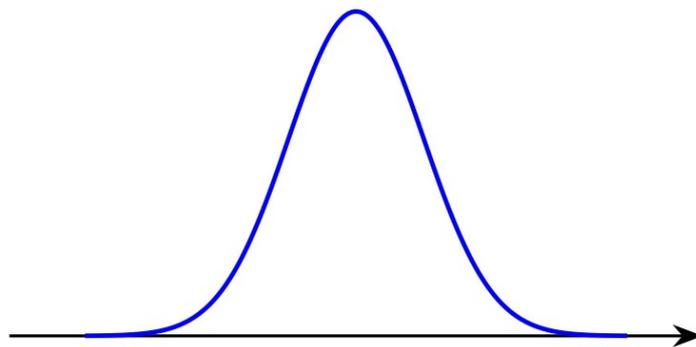
1. Geben Sie den Schätzwert für den tatsächlichen durchschnittlichen Verbrauch an!
2. Bestimmen Sie ein zweiseitiges 90-%-Konfidenzintervall für den durchschnittlichen Verbrauch!
3. Bei welchem Konfidenzniveau wäre der Verbrauch von 10 kWh/100 km gerade noch im Konfidenzintervall enthalten (d. h. bildete die Grenze des Konfidenzintervalls)?

Ausführung Beispiel 5a:

b) (15 Punkte) Lässt sich anhand Ihrer Daten mit einem Signifikanzniveau von 0,05 zeigen, dass der zu erwartende Verbrauch über 10 kWh/100 km liegt? Beantworten Sie dazu folgende Aufgaben!

1. Welcher Test ist anzuwenden?
2. Wie sind die Hypothesen zu formulieren?
3. Geben Sie den Testwert an!
4. Bestimmen Sie den kritischen Bereich!
5. Wie entscheiden Sie?
6. Wie ist das Ergebnis zu interpretieren?

Veranschaulichen Sie die Grenzen des kritischen Bereichs, das Signifikanzniveau und den berechneten Testwert in der nachstehenden Grafik!



Ausführung Beispiel 5b:

6. (8 Punkte) Siegfried hat seine Leidenschaft für die Forstwirtschaft entdeckt und hilft privaten Waldbesitzern bei Aufforstungsarbeiten. Um seinen Auftraggebern keine unrealistischen Kostenvoranschläge zu geben, analysiert er mittels Regressionsanalyse, wie lange er für das Pflanzen von Fichten bestimmter Höhe benötigt. Seine Aufzeichnungen sind der folgenden Tabelle zu entnehmen.

Baumhöhe	62	32	73	75	59
Setzdauer	3,6	2,1	3,9	4	3,2

Siegfried hat auch bereits die Koeffizienten der Regressionsgerade berechnet: Achsenabschnitt  $a = 0,686$  und Steigung  $b = 0,044$ .

- Berechnen Sie die Residuen!
- Berechnen Sie die Standardabweichung des Störterms!

Ausführung Beispiel 6:

Ausführung Beispiel 6: