

Statistik Vorlesung

25. September 2025

Arbeitszeit: 100 Minuten

VORNAME:		MATR.NR.:	
NACHNAME:			

ERLAUBT: **Skriptum des Instituts, Taschenrechner gemäß Liste**

VERBOTEN: **alle sonstigen Unterlagen, Handys**

Aufgabe 1 und 2: Für jede vollständig richtig gelöste Frage gibt es 2 Punkte. Es gibt keine Teilpunkte und keine Minuspunkte.

Lösungswege müssen nachvollziehbar aufgeschrieben sein.

Aufgabe	max. Punkte	erreichte Punkte	Teilsumme
1	10		
2	10		
9	11		
3	11		
4	9		
5	10		
6	15		
7	12		
8	12		
Summe	100		
Note			

1. (10 Punkte)

- a) Kreuzen Sie die beiden statistischen Kennzahlen an, die aus einem Boxplot im Allgemeinen abgelesen werden können.

- ☐ arithmetische Mittel
☐ Modus
☒ Median
☐ Variationskoeffizient
☒ Spannweite

- b) Ergänzen Sie die Textlücke im nachstehenden Satz so, dass jedenfalls eine wahre Aussage entsteht.

Ist der p-Wert kleiner als die Irrtumswahrscheinlichkeit α , so ist die H_1

bestätigt

- c) Markieren Sie, ob die folgende Aussage zutrifft oder nicht, und kennzeichnen Sie die jeweils passende Begründung.

Der Modalwert teilt eine sortierte Liste von Einzeldaten in zwei Hälften.

☐	Wahr
☒	Falsch

Begründung:

☐	Der Modalwert entspricht dem 50 %-Quantil.
☐	Der Modalwert ist definiert als durchschnittliche Merkmalsausprägung in der Stichprobe.
☒	Der Modalwert ist bei Einzeldaten die am häufigsten vorkommende Ausprägung.

- d) Die Zufallsvariable X ist normalverteilt mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 . Man erhebt eine Stichprobe, um μ zu schätzen. Welche Behauptungen über Konfidenzintervalle der Form $\left[\bar{x} - \frac{s}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}} ; \bar{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}} \right]$ treffen zu? Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.

- ☐ Je größer die Standardabweichung, desto schmaler das Konfidenzintervall.
☐ Die Breite des Konfidenzintervalls nimmt mit steigendem Stichprobenumfang im Allgemeinen zu.
☒ Die t-Verteilung wird verwendet, weil die Varianz σ^2 unbekannt ist.
☒ Je höher das Konfidenzniveau gewählt wird, desto breiter ist das Konfidenzintervall.
☐ Für verschiedene Stichproben erhält man immer exakt dasselbe realisierte Konfidenzintervall, wenn man immer dasselbe Konfidenzniveau vorgibt.

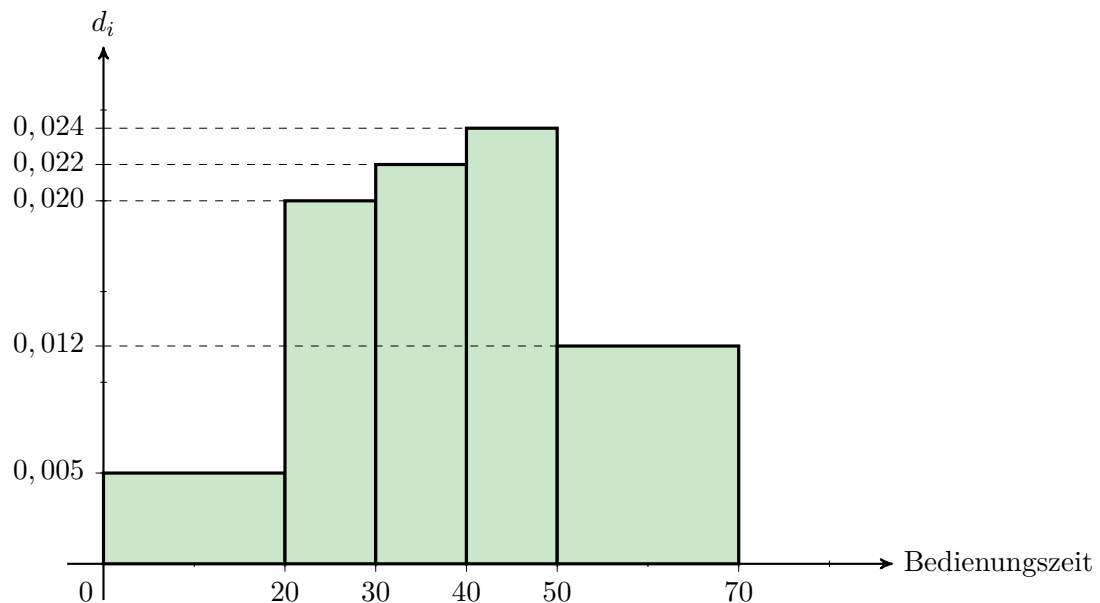
- e) Ergänzen Sie die Textlücke im nachstehenden Satz so, dass jedenfalls eine wahre Aussage entsteht.

Der Zusammenhang zwischen Einkommen (x in Euro) und Mietausgaben (y in Euro) von Haushalten wurde durch folgende Regressionsgerade geschätzt: $\hat{y} = 500 + 0,3x$. Jede Erhöhung des Einkommens um 100 Euro führt zu einem mittleren Anstieg der geschätzten Miete um

30

Euro.

2. (10 Punkte) In der nachstehenden Graphik sind die Bedienungszeiten (in Sekunden) einer Kasse eines Supermarktes für 200 aufeinanderfolgende Kunden dargestellt:



- a) Welches Merkmal wurde hier erhoben und welches Skalenniveau hat es in dieser Umfrage?

Bedienungszeit; metrisch

- b) Was stellen die einzelnen Rechtecksflächen im obigen Histogramm dar?

die relativen Klassenhäufigkeiten

- c) Wie viele Kunden hatten eine Bedienungszeit von mindestens 40 Sekunden?

96

- d) Welche Bedienungszeit hatten höchstens 10 % der Kunden?

20

- e) Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.



20 Kunden hatten eine Bedienungszeit unter 20 Sekunden.



Genau $\frac{2}{5}$ der untersuchten Kunden hatten eine Bedienungszeit von mehr als 40 Sekunden..



Der Median der Verteilung liegt im Intervall $[30; 40[$.

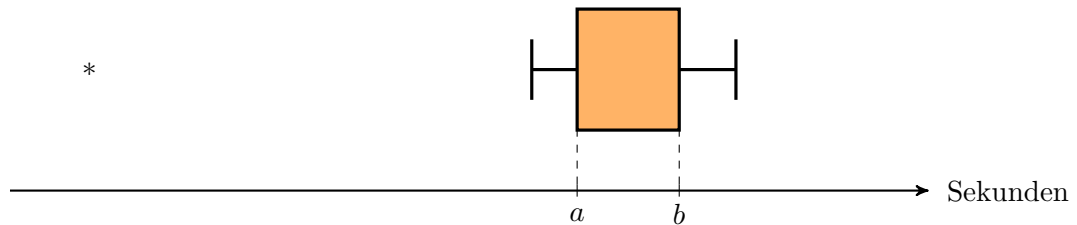


4 Kunden hatten eine Bedienungszeit von 25 Sekunden.

3. Ein Autohersteller überprüft regelmäßig die Effizienz seiner Fertigungsanlagen, insbesondere die Arbeitsabläufe an den Fließbändern. Bei einer solchen Untersuchung wurden die Montagezeiten eines bestimmten Werkstücks für 10 Arbeiter in Sekunden gemessen.

92	105	53	107	98	110	96	100	102	105
----	-----	----	-----	----	-----	----	-----	-----	-----

Die erhobenen Daten sind im folgenden Boxplot dargestellt:



- (4 Punkte) Berechnen Sie das arithmetische Mittel und den Median der Montagezeiten.
- (2 Punkte) Warum ist für diese Daten der Mittelwert kleiner als der Median? Begründen Sie!
- (2 Punkte) Berechnen Sie die Werte a und b des obigen Boxplots.
- (3 Punkte) Wie müsste man den Datenpunkt 92 des obigen Datensatzes abändern, damit er zu einem potentiellen Ausreißer nach unten wird? Geben Sie das Intervall an, in dem der geänderte Wert liegen muss.

Lösung:

- Metrisch; arithmetisches Mittel: $\bar{x} = 96,8$; Standardabweichung: 15,471
- wird tendenziell größer als das arithmetische Mittel sein, da der Wert 53 einen Ausreißer nach unten darstellt, der das arithmetische Mittel nach unten „zieht“.
- $a = 96$, $b = 105$
- $[69; 82,5]$

4. Im Rahmen eines Beurteilungsverfahrens wurden sieben Kandidaten anhand zweier verschiedener Kriterien bewertet:

Kriterium 1: Benötigte Zeit (in Minuten) zur Lösung von zehn vorgegebenen Logikaufgaben

Kriterium 2: Platzierung in einem Englisch-Sprachkompetenztest

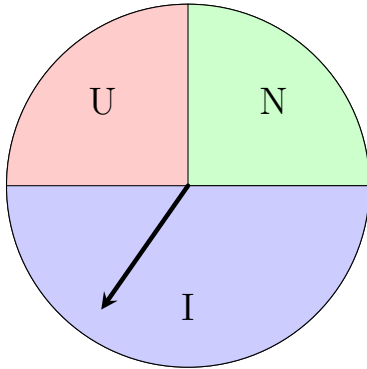
Kandidat	1	2	3	4	5	6	7
Zeit	25	27	19	21	20	19	24
Platzierung	7	6	3	4	1	2	5

- a) (2 Punkte) Ist es sinnvoll, in diesem Fall eine Regressionsgerade zu berechnen? Begründen Sie Ihre Antwort!
- b) (7 Punkte) Bestimmen Sie eine geeignete Kennzahl, um den Zusammenhang zwischen den beiden Merkmalen zu beschreiben, und interpretieren Sie den erhaltenen Wert im gegebenen Kontext.

Lösung:

- a) Da die Platzierung im Englisch-Test eine ordinal skalierte Variable ist, ist es nicht sinnvoll, eine Regressionsgerade zu berechnen. Eine Regressionsanalyse setzt voraus, dass beide Merkmale mindestens intervallskaliert sind.
- b) Der Rangkorrelationskoeffizient von Spearman beträgt $r_s \approx 0,848$. Dies zeigt einen starken positiven Zusammenhang zwischen der benötigten Zeit und der Platzierung. Das bedeutet: Kandidaten, die mehr Zeit für die Logikaufgaben benötigen, erzielen tendenziell schlechtere (höhere) Platzierungen im Englisch-Test.

5. (10 Punkte) Im Rahmen einer Semestereröffnungsparty wird folgendes Glücksspiel angeboten:
- Man darf das nachstehende Glücksrad dreimal drehen. Es wird angenommen, dass der Zeiger des Glücksrades mit derselben Wahrscheinlichkeit an jeder beliebigen Stelle des Kreises stehen bleiben kann. Erzielt man dabei das Wort „UNI“ (in genau dieser Reihenfolge) so gewinnt man € 10.- Dreht man dreimal denselben Buchstaben hintereinander, so gewinnt man € 5.-. In allen anderen Fällen gibt es keinen Gewinn. Berechnen Sie den Erwartungswert und die Varianz des Gewinns pro Spiel.

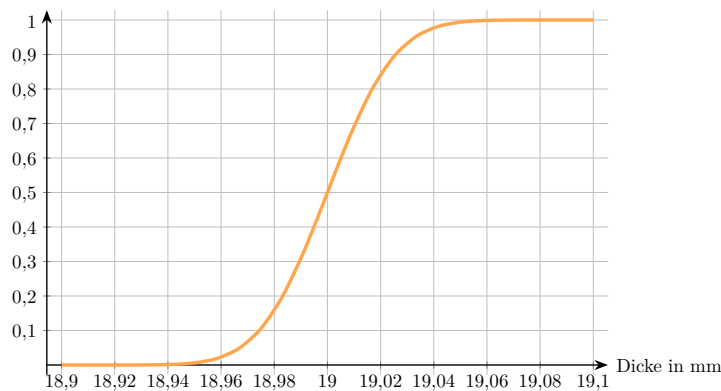


Lösung:

- a) $E(G) = 1,094$
- b) $Var(G) = 5,835$

6. Auf einer Maschine werden Platten hergestellt, deren Dicke X (bei korrekter Einstellung der Maschine) normalverteilt ist mit dem Erwartungswert $\mu = 19$ mm und der Soll-Standardabweichung $\sigma = 0,02$ mm. Platten mit einer Dicke zwischen $18,98 \leq X \leq 19,02$ mm sind Qualität A, alle anderen Qualität B.

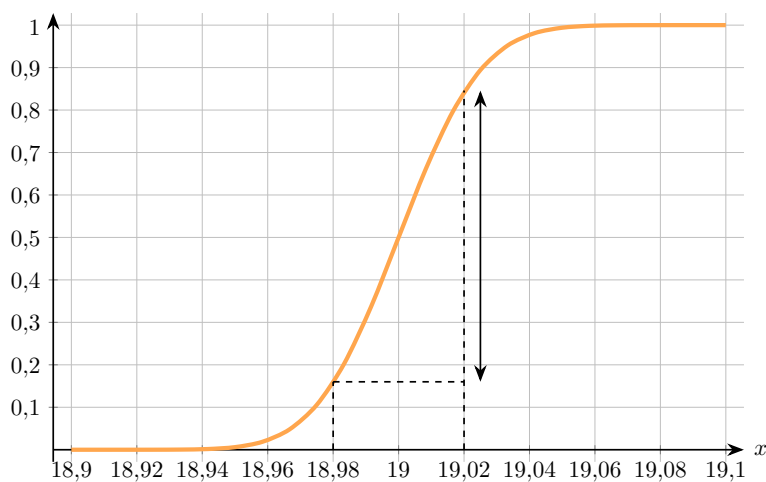
- (4 Punkte) Wie viel Prozent der Produktion sind Qualität A?
- (2 Punkte) In der nachstehenden Abbildung ist der Graph der Verteilungsfunktion dieser Normalverteilung dargestellt. Kennzeichnen Sie in dieser Abbildung die Wahrscheinlichkeit, dass eine zufällig ausgewählte Platte von Qualität A ist.
- (4 Punkte) Wie müssten die Toleranzgrenzen symmetrisch zu μ festgelegt werden, um nur 5 % Platten der Qualität B zu erhalten?
- (5 Punkte) Eine neu angeschaffte Maschine kann 98 % der Platten in Qualität A herstellen. Wie groß ist die Wahrscheinlichkeit, in einer nicht auf Qualität kontrollierten Lieferung von 10 Platten mehr als eine Platte der Güte B anzutreffen?



Lösung:

a) 0,683

b)



c) [18,961; 19,039]

d) 0,016

7. Die Reichweite X (in Kilometer) eines Elektroautos mit Semi-Festkörperbatterie ist $N(\mu, 18)$ -verteilt.
- a) (6 Punkte) Testen Sie zum Niveau $\alpha = 0,01$ die Hypothese $H_0 : \mu \geq 500$ aufgrund einer Stichprobe vom Umfang $n = 225$, die einen Mittelwert von 496,8 ergab und interpretieren Sie Ihr Ergebnis.
 - b) (3 Punkte) Bestimmen Sie den p-Wert.
 - c) (3 Punkte) Unterhalb welcher Reichweite würden Sie die Nullhypothese verwerfen?

Lösung:

- a) H_0 verworfen Es konnte gezeigt werden, dass die Autos signifikant weniger als 500 km zurücklegen.
- b) 0,004
- c) 497,21

8. Ein Unternehmen hat zwei parallele Fertigungsstraßen. Die Anzahl der bearbeiteten Produkte wurde an sechs zufällig ausgewählten Tagen beobachtet.

Fertigung A	120	113	118	124	128	116
Fertigung B	120	125	111	114	140	113

- a) (9 Punkte) Können Sie mit einem Signifikanzniveau von $\alpha = 0,1$ bestätigen, dass die Standardabweichung auf Fertigungsstraße B signifikant größer als auf Straße A ist? Gehen Sie dabei von normalverteilten Daten aus.

Hinweis: Die empirische Standardabweichung auf Fertigungsstrasse A beträgt: $s_A = 5,456$.

- b) (3 Punkte) Können Sie aus dem Ergebnis der Aufgabe a) schließen, dass die Varianzen der beiden Fertigungsstraßen unterschiedlich sind? Begründen Sie Ihre Antwort ausführlich!

Lösung:

a) $H_1 = \frac{\sigma_{x^2}}{\sigma_{y^2}} < 1; t_0 = 0,252 \quad K = [0; 0,29]; \text{ Ja.}$

- b) Wenn der einseitige Test signifikant ist, bedeutet das nur, dass die Varianz von Fertigungsstraße B größer ist. Es schließt jedoch nicht aus, dass die Varianzen auch bei einem zweiseitigen Test als gleich betrachtet werden könnten, da der zweiseitige Test zwar dieselbe Testgröße verwendet, aber die kritischen Bereiche anders berechnet werden.

9. Im Folgenden wird eine Regressionsanalyse dargestellt, die den Zusammenhang zwischen den Arbeitslosenzahlen und der Anzahl offener Stellen in Österreich für die Jahre 2016 bis 2024 untersucht. Das Modell soll die Abhängigkeit der Arbeitslosenzahl (in Tausend) von der Anzahl offener Stellen (in Tausend) schätzen.

Die Analyse ergab folgenden Output:

Regressions-Statistik	
Bestimmtheitsmaß	0,015
Standardfehler	57,485
Beobachtungen	9

	Koeffizient	Standardfehler	t-Statistik	p-Wert	Untere 95%	Obere 95%
Schnittpunkt	318,396	87,750	3,628	0,008	110,900	525,892
offene Stellen	-0,405	1,231			-3,316	2,506

- (1 Punkt) Formulieren Sie die Regressionsgleichung des Modells auf Grundlage der vorliegenden Regressionsanalyse. Geben Sie dabei die Gleichung der geschätzten Regressionsgeraden in Abhängigkeit von der unabhängigen Variablen an.
- (2 Punkte) Welche Arbeitslosenzahl prognostiziert das Modell für das Jahr 2025, in dem 78.000 offene Stellen erwartet werden?
- (2 Punkte) Interpretieren Sie das Bestimmtheitsmaß in Bezug auf die Güte des Modells.
- (6 Punkte) Hat die Anzahl der offenen Stellen einen signifikanten Einfluss auf die Arbeitslosenzahlen? Führen Sie den entsprechenden Test zum Niveau $\alpha = 5\%$ durch und beantworten Sie die Frage.

Lösung:

- $\hat{Y} = 318,396 - 0,405 \cdot x$
- 286.802
- $R^2 = 0,015$. Das Bestimmtheitsmaß von 0,015 bedeutet, dass 1,5% der Varianz der Arbeitslosenzahlen durch die offenen Stellen erklärt werden können. Das Bestimmtheitsmaß von $R^2 = 0,015$ zeigt, dass das Modell die Arbeitslosenzahlen nur sehr schlecht erklärt. Es ist daher kein gutes Modell, um den Zusammenhang zwischen den Arbeitslosenzahlen und der Anzahl offener Stellen zu beschreiben.
- $t_0 = -0,329$; $K =]-\infty; -2,365[\cup]2,365; \infty[$; $t_0 \notin K \Rightarrow H_0$ beibehalten (kein signifikanter Einfluss).