Klausurreport »Allgemeines Lineares Modell«, SoSe 2025

Die Klausur zur Vorlesung “Allgemeines Lineares Modell” (Modul B2: Inferenzstatistik) im Sommersemester 2025 wurde am 18.07.2025 von 10:00 bis 11:00 Uhr in den Computer-Pools des Universitätsrechenzentrums (URZ) als elektronische Klausur (E-Klausur) mit 53 Teilnehmenden durchgeführt. Sie bestand aus 30 Multiple-Choice-Aufgaben mit jeweils vier Antwortmöglichkeiten und jeweils genau einer richtigen Antwort. Die Klausur ist diesem Bericht beigefügt, richtige Antworten sind auf der letzten Seite angegeben.

Bewertungschema

Die Aufteilung der zugelassenen Noten auf die erreichten Prozentpunkte wurde anhand untenstehender Tabelle vorgenommen. Diese trifft folgende Zuordnung der erreichten Prozentpunkte zu den zugelassenen Noten anhand von geschlossenen Intervallen gerundeter Prozentpunkte.

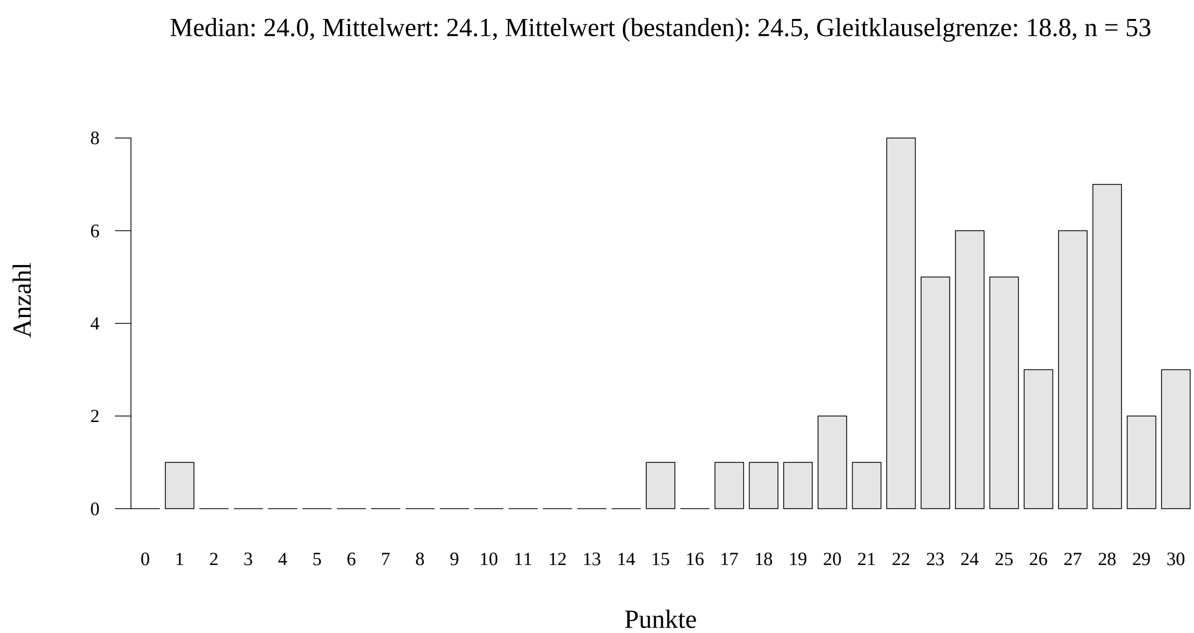
$\leq$	$\geq$	Note
100	95	1,0
94	90	1,3
89	85	1,7
84	80	2,0
79	75	2,3
74	70	2,7
69	65	3,0
64	60	3,3
59	55	3,7
54	50	4,0
49	0	5,0

Es ergibt sich folgendes Punktenotenschema, wobei < 15 Punkte mit 5,0 bewertet wurden.

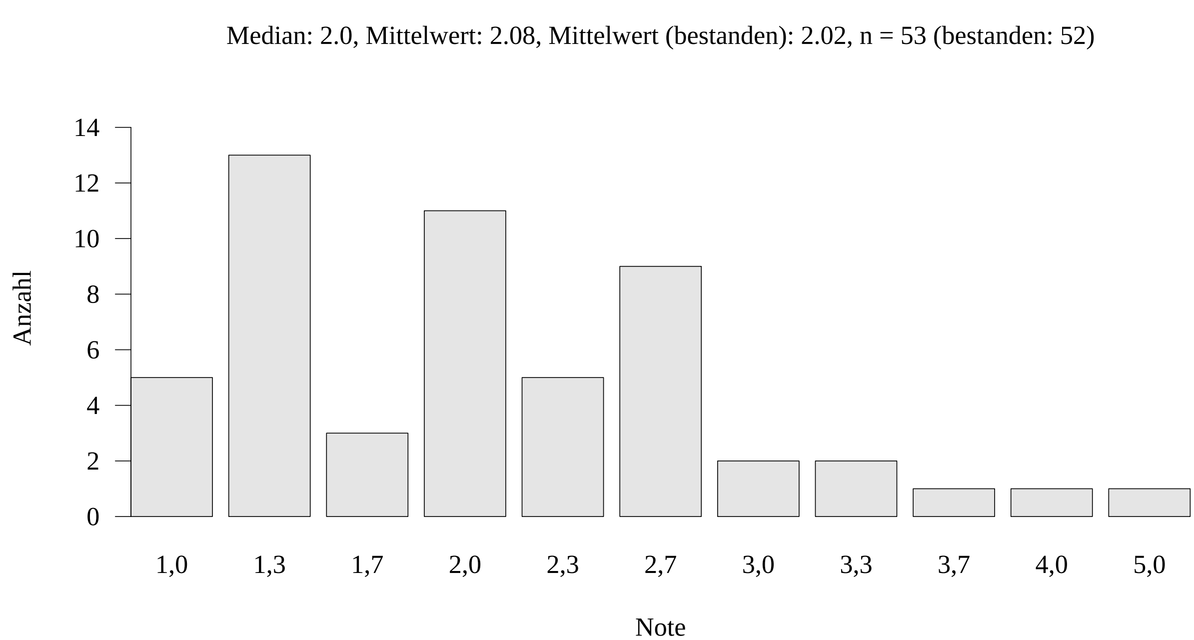
Punkte	Prozent	Note
30	100,0	1,0
29	96,7	1,0
28	93,3	1,3
27	90,0	1,3
26	86,7	1,7
25	83,3	2,0
24	80,0	2,0
23	76,7	2,3
22	73,3	2,7
21	70,0	2,7
20	66,7	3,0
19	63,3	3,3
18	60,0	3,3
17	56,7	3,7
16	53,3	4,0
15	50,0	4,0

Ergebnisse

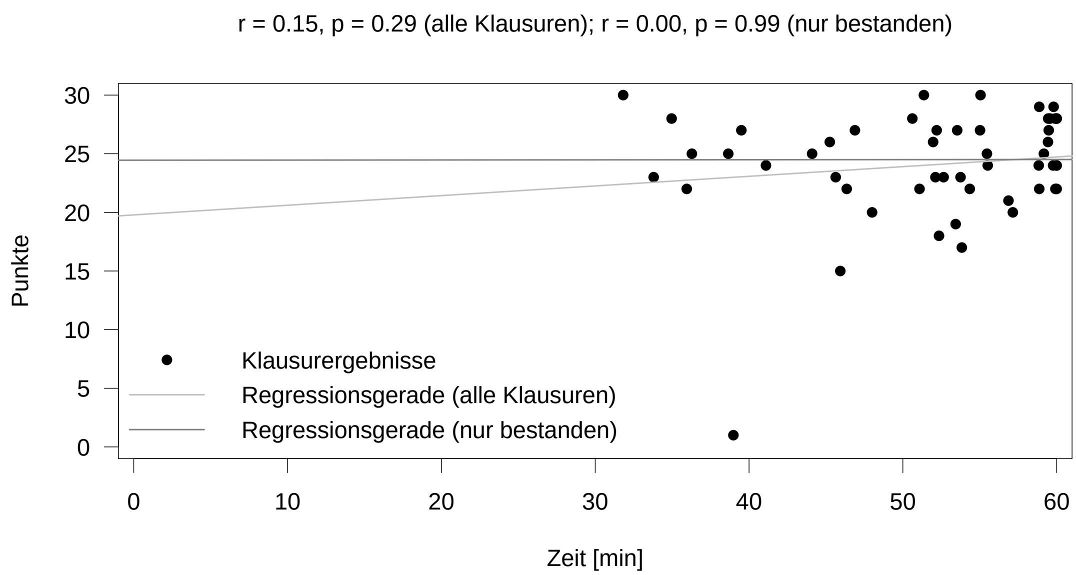
Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erzielten Punkte.



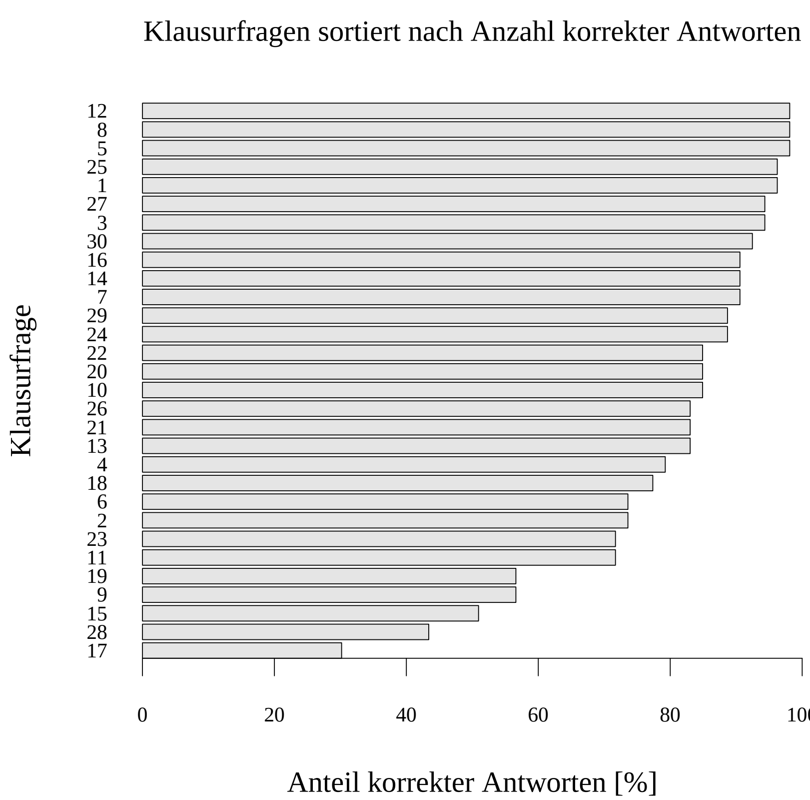
Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erreichten Noten.



Die nachfolgende Abbildung zeigt, dass verbrauchte Zeit (= Dauer von Start der Klausur bis zu elektronischer Abgabe) nicht mit erreichter Punktzahl (= Anzahl richtig beantworteter Fragen) korreliert, wenn nicht-bestandene Klausuren aus dem Datensatz entfernt werden.



Die nachfolgende Abbildung zeigt die Reihenfolge der Klausurfragen, wenn man sie nach Anteil korrekter Antworten über alle Teilnehmenden hinweg sortiert. Dieser Anteil kann annäherungsweise als umgekehrter Schweregrad einer Frage interpretiert werden.



Otto-Von-Guericke-Universität Magdeburg
Fakultät für Naturwissenschaften
Institut für Psychologie
Lehrstuhl Methodenlehre I
Dr. rer. nat. Joram Soch

Klausur "Allgemeines Lineares Modell"
(Modul B2: Inferenzstatistik)
Termin: 18.07.2025

Name, Vorname: ______________________

Matrikelnummer: ______________________

Bearbeitungshinweise:

- Die Klausur besteht aus **30 Aufgaben**.
- Sie haben zur Bearbeitung **60 Minuten** Zeit.
- Bei jeder Aufgabe sind jeweils **vier Antwortmöglichkeiten** vorgegeben.
- Es trifft **immer genau eine** Antwort zu.
- Bitte kreuzen Sie bei jeder Aufgabe nur die **zutreffende Antwort** an.
- Für jede **richtig gelöste Aufgabe** erhalten Sie einen Punkt.
- Die Klausur ist bestanden, wenn Sie mindestens **15 Punkte** erreichen.

Viel Erfolg!

1. Welche Aussage über univariate, reellwertige linear-affine Funktionen der Form $f(x) = \beta_0 + \beta_1 x$ stimmt **nicht**?
   a) Eine linear-affine Funktion ist durch zwei skalare Parameter gekennzeichnet.
   b) Jede linear-affine Funktion geht durch den Koordinatenursprung, d.h. den Punkt $(0, 0)$.
   c) Der Steigungsparameter einer linear-affinen Funktion gibt an, um wieviele Einheiten der Funktionswert steigt, wenn man das Funktionsargument um eine Einheit erhöht.
   d) Der Offset-Parameter einer linear-affinen Funktion gibt an, an welcher Stelle die zu dieser Funktion gehörende Gerade die y-Achse schneidet.

2. Wieviele Spalten hat die Designmatrix im Modell einfachen linearen Regression?
   a) 1
   b) 2
   c) 3
   d) 4

3. Welche Aussage über Stichprobenkorrelationen ist **nicht** korrekt?
   a) Häufige richtungsgleiche Abweichung der x_i und y_i von ihren Mittelwerten führt im Allgemeinen zu einer positiven Korrelation.
   b) Häufige richtungsungleiche Abweichung der x_i und y_i von ihren Mittelwerten führt im Allgemeinen zu einer negativen Korrelation.
   c) Halten sich richtungsgleichen und richtungsungleichen Abweichungen die Waage, führt das im Allgemeinen zu geringer Korrelation.
   d) Die Richtungsgleichheit oder -ungleichheit der Abweichung der x_i und y_i von ihren Mittelwerten spielt für die Stichprobenkorrelation keine Rolle.

4. Welche der folgenden Aussagen über die Extremwerte des Bestimmtheitsmaßes R^2 stimmt **nicht**?
   a) Wenn die erklärte Quadratsumme Null ist, ist $\mathrm{R}^2 = 0$.
   b) Wenn die residuelle Quadratsumme Null ist, ist $\mathrm{R}^2 = 1$.
   c) Wenn die erklärte gleich der totalen Quadratsumme ist, ist $\mathrm{R}^2 = 1$.
   d) Wenn die totale Quadratsumme Null ist, ist $\mathrm{R}^2 = 0$.

5. Es seien $A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$, $x = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ und $b = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$. Berechnen Sie $Ax + b$!
   a) $Ax + b = \begin{pmatrix} 6 \\ 7 \end{pmatrix}$
   b) $Ax + b = \begin{pmatrix} 8 \\ 5 \end{pmatrix}$
   c) $Ax + b = \begin{pmatrix} 4 \\ 9 \end{pmatrix}$
   d) $Ax + b = \begin{pmatrix} 11 \\ 6 \end{pmatrix}$

6. Welche Aussage über die n-dimensionale Einheitsmatrix I_n stimmt **nicht**?
   a) Die Transponierte von I_n ist I_n.
   b) Die Inverse von I_n ist I_n.
   c) Die Determinante von I_n ist 1.
   d) Die Spur von I_n ist 1.

7. Gegeben sei die Matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$. Welche Aussage trifft dann zu?
   a) A ist eine symmetrische Matrix.
   b) A ist eine Diagonalmatrix.
   c) Die Determinante von A ist 5.
   d) Die Spur von A ist 6.

8. $\xi \in \mathbb{R}^n$ sei ein multivariat normalverteilter Zufallsvektor mit $\xi \sim N(\mu, \Sigma)$. Was ist die Kovarianzmatrix $\mathbb{C}(\xi)$?
   a) n
   b) μ
   c) Σ
   d) $\mathbb{R}$

9. $\xi \sim N(\mu, \Sigma)$ sei ein multivariat normalverteilter Zufallsvektor mit den Parametern $\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ und $\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$. Welche Aussage ist **nicht** korrekt?
   a) Die Anzahl der Komponenten des Zufallsvektors ξ ist zwei.
   b) Die Erwartungswerte der Komponenten des Zufallsvektors ξ sind 1 und 2.
   c) Die Kovarianz der Komponenten des Zufallsvektors ξ ist 3.
   d) Der Zufallsvektor ξ folgt einer sphärischen Normalverteilung.

10. Was gehört **nicht** zu den Standardproblemen Frequentistischer Inferenz?
    a) Parameterschätzung
    b) Konfidenzintervalle
    c) Hypothesentests
    d) Modellformulierung

11. Sind die Komponenten des Datenvektors im Allgemeinen Linearen Modell unabhängig und identisch verteilt?
    a) nein, niemals
    b) im Allgemeinen nicht, in Spezialfällen (z.B. Einstichproben-T-Test) ja
    c) im Allgemeinen ja, in Spezialfällen (z.B. multiple lineare Regression) nicht
    d) ja, immer

12. Wie lautet die Formel für den Betaparameterschätzer im Allgemeinen Linearen Modell?
    a) $\hat{\beta} = X\beta + \varepsilon$
    b) $\hat{\beta} = (I_n - X\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}})y$
    c) $\hat{\beta} = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}y$
    d) $\hat{\beta} = X\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}y$

13. Wie lautet die Formel für den (unverzerrten) Varianzparameterschätzer im Allgemeinen Linearen Modell?
    a) $\hat{\sigma}^2 = (y - \bar{y})^{\mathrm{T}}(y - \bar{y})$
    b) $\hat{\sigma}^2 = (y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$
    c) $\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$
    d) $\hat{\sigma}^2 = \frac{1}{n-p}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$

14. Was ist eine (zentrale) T-Zufallsvariable?
    a) Eine T-Zufallsvariable ist ein Quotient $Z/\sqrt{U/n}$, wobei Z normalverteilt mit Erwartungswertparameter ungleich 0 ist und U eine Chi-Quadrat-Zufallsvariable mit Freiheitsgradparameter n ist.
    b) Eine T-Zufallsvariable ist ein Quotient $Z/\sqrt{U/n}$, wobei Z standardnormalverteilt ist und U eine Chi-Quadrat-Zufallsvariable mit Freiheitsgradparameter n ist.
    c) Eine T-Zufallsvariable ist ein Quotient $\frac{U_1/n_1}{U_2/n_2}$, wobei U_1 nicht-zentral χ^2-verteilt mit Freiheitsgradparameter n_1 und Nichtzentralitätsparameter $d \neq 0$ und U_2 χ^2-verteilt mit Freiheitsgradparameter n_2 ist.
    d) Eine T-Zufallsvariable ist ein Quotient $\frac{U_1/n_1}{U_2/n_2}$, wobei U_1 und U_2 Chi-Quadrat-Zufallsvariablen mit den Freiheitsgradparametern n_1 und n_2 sind.

15. Gegeben sei ein Allgemeines Lineares Modell mit der Designmatrix $X \in \mathbb{R}^{n \times p}$. Welcher Verteilung folgt die T-Statistik unter der Nullhypothese, dass $c^{\mathrm{T}}\beta = c^{\mathrm{T}}\beta_0$ ist?
    a) $T \sim t(n)$
    b) $T \sim t(p)$
    c) $T \sim t(n-p)$
    d) $T \sim t(\delta, n-p)$ mit $\delta \neq 0$

16. Was ist eine (zentrale) F-Zufallsvariable?
    a) Eine F-Zufallsvariable ist ein Quotient $Z/\sqrt{U/n}$, wobei Z normalverteilt mit Erwartungswertparameter ungleich 0 ist und U eine Chi-Quadrat-Zufallsvariable mit Freiheitsgradparameter n ist.
    b) Eine F-Zufallsvariable ist ein Quotient $Z/\sqrt{U/n}$, wobei Z standardnormalverteilt ist und U eine Chi-Quadrat-Zufallsvariable mit Freiheitsgradparameter n ist.
    c) Eine F-Zufallsvariable ist ein Quotient $\frac{U_1/n_1}{U_2/n_2}$, wobei U_1 nicht-zentral χ^2-verteilt mit Freiheitsgradparameter n_1 und Nichtzentralitätsparameter $d \neq 0$ und U_2 χ^2-verteilt mit Freiheitsgradparameter n_2 ist.
    d) Eine F-Zufallsvariable ist ein Quotient $\frac{U_1/n_1}{U_2/n_2}$, wobei U_1 und U_2 Chi-Quadrat-Zufallsvariablen mit den Freiheitsgradparametern n_1 und n_2 sind.

17. Gegeben sei ein vollständiges ALM mit der Designmatrix $X \in \mathbb{R}^{n \times p}$ und ein reduziertes ALM mit der Designmatrix $X \in \mathbb{R}^{n \times p_0}$, wobei $p_1 = p - p_0$. Welcher Verteilung folgt die F-Statistik unter der Nullhypothese, dass das reduzierte Modell zutrifft?

    a) $F \sim f(p_0, n - p_0)$
    b) $F \sim f(p_0, n - p)$
    c) $F \sim f(p_1, n - p)$
    d) $F \sim f(\delta, p_1, n - p)$ mit $\delta \neq 0$

18. Wieviele Spalten hat die Designmatrix im Einstichproben-T-Test-Modell?

    a) 1
    b) 2
    c) 3
    d) 4

19. Von welcher der folgenden Größen hängt die Powerfunktion eines zweiseitigen Einstichproben-Tests **nicht** ab?

    a) von den wahren, aber unbekannten Parametern μ und σ^2
    b) von den Schätzern der Parameter, $\bar{y}$ und s_y^2
    c) vom gewählten kritischen Wert k
    d) vom Stichprobenumfang n

20. Gegeben sei das Szenario der einfaktoriellen Varianzanalyse mit p Gruppen randomisierter experimenteller Einheiten, wobei $n_1, \ldots, n_p$ die Anzahlen der Einheiten pro Gruppe seien. Wieviele Spalten hat die Designmatrix des Modells der einfaktoriellen Varianzanalyse in Effektdarstellung?

    a) p
    b) $p + 1$
    c) $\sum_i^p n_i$
    d) $\frac{1}{p} \sum_i^p n_i$

21. Wie stehen die Quadratsummenzerlegung im Modell der einfaktoriellen Varianzanalyse und die Quadratsummenzerlegung im Modell der (einfachen) linearen Regression miteinander in Beziehung?

    a) Die *between sum of squares* entspricht der erklärten Quadratsumme, die *within sum of squares* entspricht der residuellen Quadratsumme.
    b) Die *between sum of squares* entspricht der residuellen Quadratsumme, die *within sum of squares* entspricht der erklärten Quadratsumme.
    c) Die *between sum of squares* entspricht der totalen Quadratsumme, die *within sum of squares* entspricht der erklärten Quadratsumme.
    d) Die *between sum of squares* entspricht der residuellen Quadratsumme, die *within sum of squares* entspricht der totalen Quadratsumme.

22. Gegeben sei das Szenario der zweifaktoriellen Varianzanalyse mit zwei Faktoren mit jeweils zwei Faktorleveln. Wieviele Spalten hat die Designmatrix des Modells der zweifaktoriellen Varianzanalyse mit Interaktion in Effektdarstellung?
    a) 2
    b) 3
    c) 4
    d) 8

23. Im Modell der 2×2 Varianzanalyse mit Interaktion und Referenzgruppe seien $\mu_0 = 2$, $\alpha_2 = 1$, $\beta_2 = -1$ und $\gamma_{22} = 3$. Bestimmen Sie μ_{21}, d.h. μ_{ij} für $i = 2$ und $j = 1$.
    a) $\mu_{21} = 3$
    b) $\mu_{21} = 2$
    c) $\mu_{21} = 1$
    d) $\mu_{21} = 5$

24. Wie lautet die Definition für den p-Wert?
    a) Der p-Wert entspricht der Wahrscheinlichkeit, dass die Nullhypothese wahr ist.
    b) Der p-Wert ist die Wahrscheinlichkeit, dass die beobachteten Effekte zufällig zustande gekommen sind.
    c) Der p-Wert ist das kleinste Signifikanzlevel, bei welchem man die Nullhypothese basierend auf einem vorliegenden Wert der Teststatistik ablehnen würde.
    d) Der p-Wert ist das größte Signifikanzlevel, bei welchem man die Nullhypothese unter Gebrauch des tatsächlichen Stichprobenumfangs mit 80%-iger Wahrscheinlichkeit ablehnen würde.

25. Worin besteht die Motivation, bedingte bzw. partielle Korrelationen zu berechnen?
    a) Bedingte und partielle Korrelation können Schwächen in der Datenerhebung ausgleichen, wie z.B. niedrige Datenqualität, schlechte Durchführung des Experiments oder geringes Signal-zu-Rauschen-Verhältnis.
    b) Bei der Betrachtung von bedingten und partiellen Korrelationen gilt ein anderes Signifikanzniveau als für "normale" Stichprobenkorrelationen.
    c) Bedingte und partielle Korrelationen sind besser in der Lage, auch bei niedrigem Stichprobenumfang signifikante Ergebnisse zu erzeugen.
    d) Bei der Betrachtung von zwei Zufallsvariablen ist es möglich, dass diese scheinbar hoch korreliert sind, aber weniger stark miteinander zusammenhängen, wenn man Einflüsse einer dritten Zufallsvariable mit in Betracht zieht.

26. Gegeben sei eine Menge von Realisierungen (x_i, y_i, z_i) mit $i = 1, \ldots, n$ von drei Zufallsvariablen x, y, z. Wie ist die partielle Stichprobenkorrelation $r_{x,y\setminus z}$ definiert?
    a) Die partielle Stichprobenkorrelation ist die Stichprobenkorrelation der x_i und y_i, nachdem der Einfluss der z_i im Sinne eines Modells der einfachen linearen Regression aus ihnen herausgerechnet wurde.
    b) Die partielle Stichprobenkorrelation ist der Durchschnitt der Stichprobenkorrelationen $r_{x,z}$ und $r_{y,z}$.
    c) Die partielle Stichprobenkorrelation ist die Stichprobenkorrelation der $\hat{z}_i^{(x)}$ und $\hat{z}_i^{(y)}$, wobei einmal der Einfluss der x_i und einmal der Einfluss der y_i im Sinne eines Modells der einfachen linearen Regression aus z_i herausgerechnet wurde.
    d) Die partielle Stichprobenkorrelation ist immer identisch mit der bedingten Korrelation $\rho(x, y|z)$.

27. Gegeben sei ein multiples Regressionsmodell mit einer abhängigen Variable, n Datenpunkten, m unabhängigen Variablen und einem zusätzlichen Interzeptparameter. Wieviele Spalten hat die Designmatrix dieses Modells?
    a) m
    b) $m+1$
    c) n
    d) $n+m$

28. $X \in \mathbb{R}^{n\times 4}$ und $\beta = \begin{pmatrix}\beta_1 & \beta_2 & \beta_3 & \beta_4\end{pmatrix}^{\mathrm{T}}$ seien die Designmatrix und der Betaparametervektor eines multiplen Regressionsmodells. Welche Nullhypothese wird durch die F-Statistik unter der Partitionierung des Modells mit $p_0 = 2$ getestet?
    a) H_0: $\beta_1 = 0 \ \wedge \ \beta_2 = 0$
    b) H_0: $\beta_3 = 0 \ \wedge \ \beta_4 = 0$
    c) H_0: $\beta_3 = 2 \ \wedge \ \beta_4 = 2$
    d) H_0: $\beta_3 = 0 \ \vee \ \beta_4 = 0$

29. Welches der folgenden ist kein parametrisches oder faktoriell-parametrisches ALM-Design?
    a) einfache lineare Regression
    b) multiple lineare Regression
    c) Zweistichproben-T-Test
    d) Kovarianzanalyse

30. Gegeben sei das Modell der einfaktoriellen Kovarianzanalyse mit Interaktion, wobei I die Anzahl der Faktorlevel und $n_1, \ldots, n_I$ die Anzahlen der Einheiten pro Faktorlevel seien. Wieviele Spalten hat die Designmatrix dieses Modells?
    a) I
    b) $I+1$
    c) $2I$
    d) $2\sum_i^I n_i$

Lösungen:

1. b)
2. b)
3. d)
4. d)
5. b)
6. d)
7. a)
8. c)
9. c)
10. d)
11. b)
12. c)
13. d)
14. b)
15. c)
16. d)
17. c)
18. a)
19. b)
20. a)
21. a)
22. c)
23. a)
24. c)
25. d)
26. a)
27. b)
28. b)
29. c)
30. c)