Klausurreport »Allgemeines Lineares Modell«, SoSe 2024

Die Klausur zur Vorlesung "Allgemeines Lineares Modell" (Modul B2: Inferenzstatistik) im Sommersemester 2024 wurde am 19.07.2024 von 10:00 bis 11:00 Uhr in den Computer-Pools des Universitätsrechenzentrums (URZ) als elektronische Klausur (E-Klausur) mit 56 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.

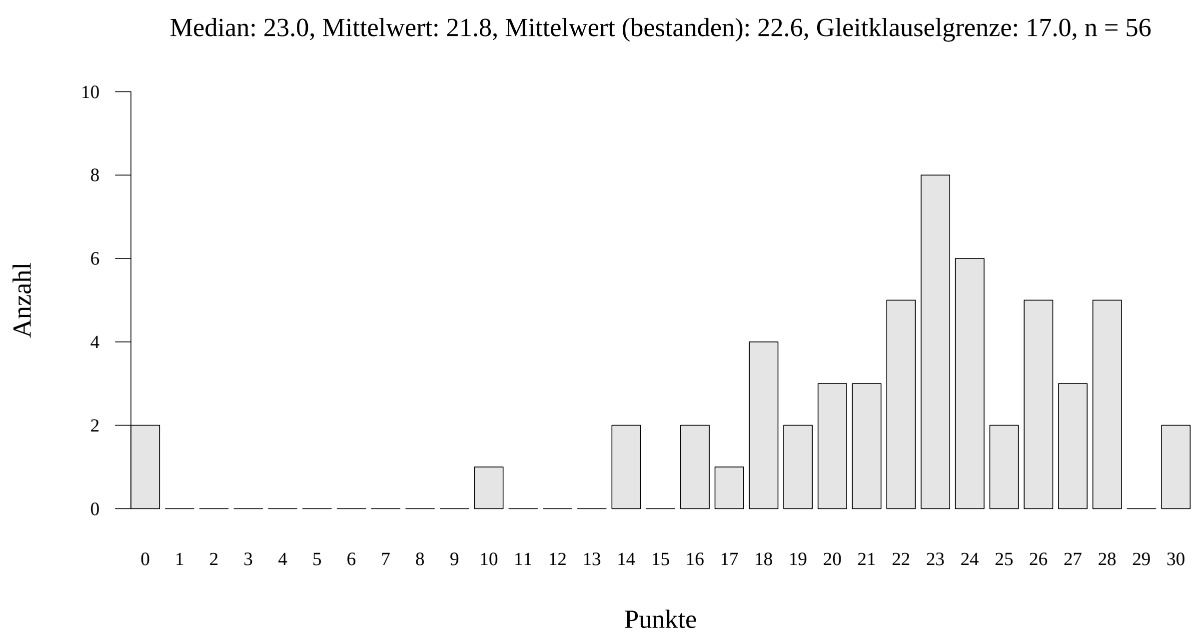
$\leq$	$\geq$	Note
100	95	1,0
94	90	1,3
89	85	1,7
84	80	2,0
79	75	2,3
74	70	2,7
69	65	3,0
64	60	3,3
59	55	3,7
54	50	4,0
49	0	5,0

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

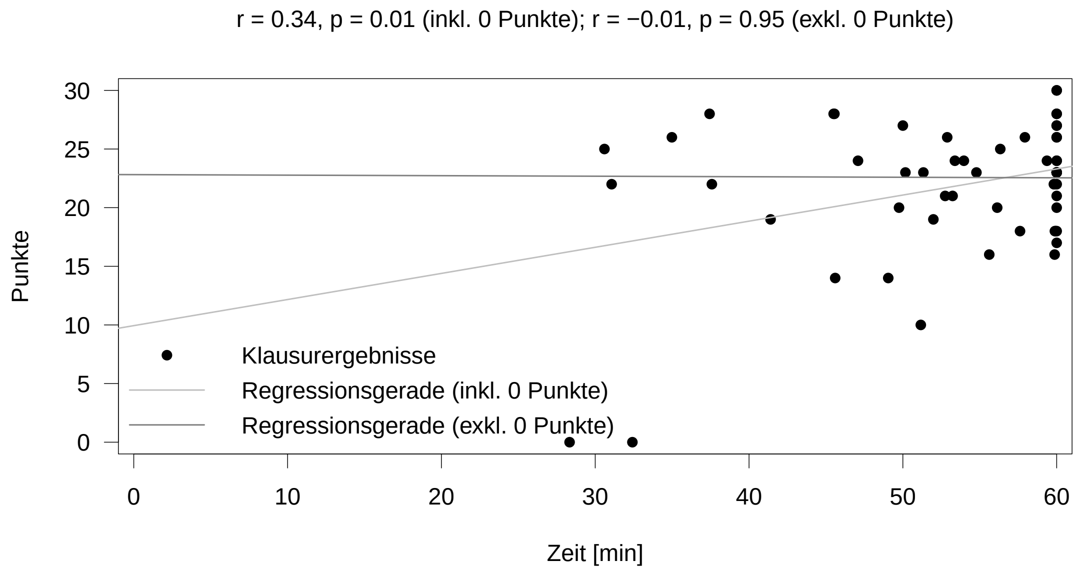
Ergebnisse

Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erzielten Punkte.

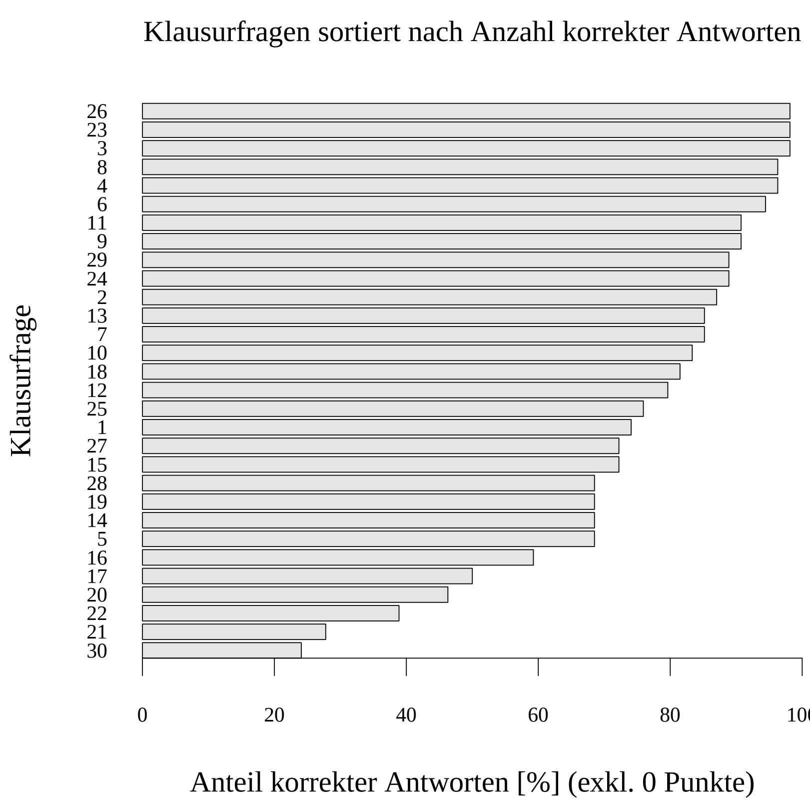


Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erreichten Noten.

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 Klausuren mit der Punktzahl 0 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: 19.07.2024

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 die Parameter einer Ausgleichsgerade ist korrekt?
   a) Die Parameter β einer Ausgleichsgerade minimieren die Summe der quadrierten vertikalen Abweichungen der gemessenen Datenpunkte y_i von den vorhergesagten Werten $f_\beta(x_i)$.
   b) Die Parameter β einer Ausgleichsgerade sind diejenigen Parameterwerte, für die die gemessenen Datenpunkte y_i gleich den vorhergesagten Werten $f_\beta(x_i)$ sind.
   c) Die Parameter einer Ausgleichsgerade entsprechen den Parametern des Ausgleichspolynoms mit Polynomgrad $k = 2$.
   d) Die Parameter einer Ausgleichsgerade unterscheiden sich von den Maximum-Likelihood-Schätzern der Parameter β_0 und β_1 im Modell der einfachen linearen Regression.

2. Welche Gleichung gibt das Modell der einfachen linearen Regression wieder?
   a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, 1),\ i = 1, \ldots, n$
   b) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(1, \sigma^2),\ i = 1, \ldots, n$
   c) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma_i^2),\ i = 1, \ldots, n$
   d) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2),\ i = 1, \ldots, n$

3. Wie ist die Korrelation zweier Zufallsvariablen definiert?
   a) Die Korrelation zweier Zufallsvariablen ist definiert als die Kovarianz dieser Zufallsvariablen, geteilt durch das Produkt ihrer Standardabweichungen.
   b) Die Korrelation zweier Zufallsvariablen entspricht der Differenz ihrer Erwartungswerte.
   c) Die Korrelation zweier Zufallsvariablen entspricht dem Steigungsparameter der Ausgleichsgerade.
   d) Die Korrelation zweier Zufallsvariablen ist immer Null, da Zufallsvariablen zufällig sind und demzufolge nicht korreliert sein können.

4. Wie ist das Bestimmtheitsmaß R^2 definiert?
   a) R^2 entspricht der residuellen Quadratsumme, geteilt durch die totale Quadratsumme.
   b) R^2 entspricht der erklärten Quadratsumme, geteilt durch die totale Quadratsumme.
   c) R^2 entspricht der erklärten Quadratsumme, minus der residuellen Quadratsumme.
   d) R^2 entspricht dem Quadrat des Steigungsparameters der Ausgleichsgerade.

5. Welche dieser Matrixoperationen ist **nicht** elementweise definiert?
   a) Matrixaddition
   b) Matrixsubtraktion
   c) Skalarmultiplikation
   d) Matrixmultiplikation

6. Es seien $A = \begin{pmatrix} -1 & 0 & 3 \\ 4 & 2 & -2 \end{pmatrix}$ und $B = \begin{pmatrix} 2 & 2 & 4 \\ 3 & 0 & -3 \end{pmatrix}$. Berechnen Sie $C = A + B$!

   a) $C = \begin{pmatrix} -3 & 2 & 7 \\ 7 & 2 & -5 \end{pmatrix}$
   b) $C = \begin{pmatrix} 1 & 2 & 7 \\ 7 & 2 & 5 \end{pmatrix}$
   c) $C = \begin{pmatrix} 1 & 2 & 7 \\ 7 & 2 & -5 \end{pmatrix}$
   d) $C = \begin{pmatrix} 1 & 0 & 7 \\ 7 & 0 & -5 \end{pmatrix}$

7. Es seien $A = \begin{pmatrix} 0 & 2 & 4 \\ -1 & -5 & 3 \end{pmatrix}$ und $B = \begin{pmatrix} 1 & 4 \\ -1 & 2 \\ 0 & 3 \end{pmatrix}$. Berechnen Sie $C = AB$!

   a) $C = \begin{pmatrix} -2 & 16 \\ 4 & 5 \end{pmatrix}$
   b) $C = \begin{pmatrix} -2 & 16 \\ 4 & -5 \end{pmatrix}$
   c) $C = \begin{pmatrix} 2 & 16 \\ 4 & -5 \end{pmatrix}$
   d) $C = \begin{pmatrix} 2 & 20 \\ 4 & -5 \end{pmatrix}$

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. $\xi \in \mathbb{R}^n$ sei ein multivariat normalverteilter Zufallsvektor mit $\xi \sim N(\mu, \Sigma)$. Unter welcher Bedingung ist ξ sphärisch normalverteilt?

   a) wenn $n = 1$ ist
   b) wenn μ dem Nullvektor 0_n entspricht
   c) wenn Σ die Form $\sigma^2 I_n$ mit $\sigma^2 \in \mathbb{R}$ annimmt
   d) wenn es sich um eine bivariate Normalverteilung handelt

10. Wieviele skalare Parameter hat das Allgemeine Lineare Modell mit der Designmatrix $X \in \mathbb{R}^{n \times p}$ und der Kovarianzmatrix $\sigma^2 I_n$?

    a) 1
    b) p
    c) $p + 1$
    d) $n \cdot p$

11. Welche Aussage macht das Theorem zur Datenverteilung im Allgemeinen Linearen Modell über den Datenvektor?
    a) $y = X\beta + \varepsilon$
    b) $y \sim N(X\beta, \sigma^2 I_n)$
    c) $\varepsilon \sim N(0_n, \sigma^2 I_n)$
    d) $y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i$

12. Warum sind die Komponenten des Zufallsfehlers im Allgemeinen Linearen Modell unabhängig und identisch verteilt?
    a) weil $\varepsilon_i \sim N(0, \sigma^2)$ für alle $i = 1, \ldots, n$ gilt
    b) weil die Designmatrix vollen Spaltenrang $\text{rg}(X) = p$ hat
    c) weil der Datenvektor aus der Addition des Zufallsfehlers ε zu dem konstanten Vektor $X\beta$ resultiert
    d) weil Zufallsfehler in probabilistischen Modellen immer unabhängig und identisch normalverteilt sind

13. 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) $\hat{\beta} = X\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}y$

14. Was ist eine (zentrale) Chi-Quadrat-Zufallsvariable?
    a) Eine Chi-Quadrat-Zufallsvariable ist eine Summe von n quadrierten Zufallsvariablen Z_i, wobei alle Z_i mit $i = 1, \ldots, n$ standardnormalverteilt sind.
    b) Eine Chi-Quadrat-Zufallsvariable ist eine Summe von n quadrierten Zufallsvariablen Z_i, wobei alle Z_i mit $i = 1, \ldots, n$ normalverteilt mit Erwartungswertparameter ungleich 0 sind.
    c) Eine Chi-Quadrat-Zufallsvariable ist eine Zufallsvariable mit dem Ergebnisraum $\mathbb{R}$, sodass alle Werte der Zufallsvariable die gleiche Wahrscheinlichkeit haben aufzutreten.
    d) Eine Chi-Quadrat-Zufallsvariable ist eine Zufallsvariable ξ mit der Wahrscheinlichkeitsdichtefunktion $f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$ für alle $x \in \mathbb{R}$.

15. Gegeben sei ein Allgemeines Lineares Modell mit $p = 4$ Spalten in der Designmatrix. Welches der folgenden ist für dieses Modell ein Kontrastgewichtsvektor, der die Differenz zwischen erster und zweiter Komponente des Betaparametervektors berechnet?
    a) $c = \begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathrm{T}}$
    b) $c = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$
    c) $c = \begin{pmatrix} 1 & -1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$
    d) $c = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$

16. Welche Aussage über die Rolle des Nullparameters β_0 für den Einsatz von T-Statistiken stimmt **nicht**?

    a) Wählt man β_0 als den p-dimensionalen Nullvektor, erhält man eine Deskriptivstatistik, die den Effekt von $c^{\mathrm{T}}\hat{\beta}$ im Sinne eines Signal-zu-Rauschen-Verhältnisses quantifiziert.

    b) Setzt man für β_0 den wahren, aber unbekannten Wert des Betaparametervektors ein, erlaubt die T-Statistik die Bestimmung von Konfidenzintervallen für einzelne Komponenten des Betaparametervektors.

    c) Wählt man β_0 beliebig, so impliziert der Nullparameter zusammen mit dem Kontrastgewichtsvektor c für einen Hypothesentest die Nullhypothese $H_0 : c^{\mathrm{T}}\beta = c^{\mathrm{T}}\beta_0$.

    d) Wählt man $\beta_0 = I_n$, induziert dies eine sphärische Normalverteilung für den Betaparameterschätzer.

17. Gegeben sei das Modell der einfachen linearen Regression mit dem Betaparametervektor $\beta = \begin{pmatrix}\beta_0 & \beta_1\end{pmatrix}^{\mathrm{T}}$, wobei β_0 dem Offsetparameter und β_1 dem Steigungsparameter entspricht. Wie muss p_0 gewählt werden, um dieses Modell so zu partitionieren, dass das reduzierte Modell nur den Effekt des Offsetparameters beinhaltet?

    a) $p_0 = 0$

    b) $p_0 = 1$

    c) $p_0 = 2$

    d) $p_0 = 3$

18. Welche Aussage über die F-Statistik stimmt **nicht**?

    a) Der Zähler der F-Statistik misst die Reduktion der residuellen Quadratsumme von reduziertem zu vollständigem Modell, im Verhältnis zur Anzahl der zusätzlichen Regressoren im vollständigen Modell.

    b) Der Zähler der F-Statistik entspricht dem Varianzparameterschätzer des reduzierten Modells.

    c) Der Nenner der F-Statistik misst die residuelle Quadratsumme des vollständigen Modells, im Verhältnis zu $n - p$, wobei n und p die Dimensionen der Designmatrix im vollständigen Modell sind.

    d) Der Nenner der F-Statistik entspricht dem Varianzparameterschätzer des vollständigen Modells.

19. Was versteht man unter Testumfangkontrolle?

    a) Testumfangkontrolle bedeutet, dass man das für einen statistischen Test erforderliche probabilistische Modell angibt.

    b) Testumfangkontrolle bedeutet, dass man die Übereinstimmung der Parameterschätzer mit den wahren, aber unbekannten Werten der Parameter kontrolliert.

    c) Testumfangkontrolle bedeutet, dass man den optimalen Stichprobenumfang anhand von Signifikanzniveau α_0, minimaler detektierbarer Effektstärke δ^* und gewünschter statistischer Power b berechnet.

    d) Testumfangkontrolle bedeutet, dass man den kritischen Wert k eines statistischen Tests so wählt, dass dieser Test eine akzeptable maximale Falschpositivrate α_0 einhält.

20. Wieviele Spalten hat die Designmatrix im Zweistichproben-T-Test-Modell?

    a) 1

    b) 2

    c) 3

    d) 4

21. 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) $\frac{1}{p} \sum_i^p n_i$

22. Wie lautet die Nullhypothese im Szenario der einfaktoriellen Varianzanalyse?

    a) Der gruppenübergreifende Erwartungswertparameter μ_0 ist gleich 0.
    b) Der gruppenübergreifende Varianzparameter σ^2 ist gleich 0.
    c) Alle gruppenspezifischen Erwartungswertparameter μ_i sind gleich 0 für $i = 1, \ldots, p$.
    d) Alle gruppenspezifischen Effektparameter α_i sind gleich 0 für $i = 2, \ldots, p$.

23. Gegeben sei das Szenario einer 3×4 zweifaktoriellen Varianzanalyse mit 10 Datenpunkten pro Zelle. Aus wievielen Datenpunkten besteht der zu diesem Szenario gehörende Datensatz?

    a) 12
    b) 40
    c) 120
    d) 144

24. Welche Rolle spielt der Parameter γ_{22} im Modell der 2×2 Varianzanalyse mit Interaktion und Referenzgruppe?

    a) γ_{22} beschreibt den Erwartungswertparameter der Referenzgruppe A1B1.
    b) γ_{22} beschreibt die Differenz beim Übergang von Level 1 zu Level 2 von Faktor A.
    c) γ_{22} beschreibt die Differenz beim Übergang von Level 1 zu Level 2 von Faktor B.
    d) γ_{22} beschreibt den Unterschied der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B, und zwar zwischen Level 1 und Level 2 von Faktor A.

25. Welche Aussage ist **nicht** korrekt?

    a) Die bedingte Kovarianz zweier Zufallsvariablen ist die Kovarianz dieser Zufallsvariablen in einer auf eine dritte Zufallsvariable bedingten Verteilung.
    b) Die bedingte Korrelation zweier Zufallsvariablen ist die Korrelation dieser Zufallsvariablen in einer auf eine dritte Zufallsvariable bedingten Verteilung.
    c) Um bedingte Kovarianz und bedingte Korrelation zu bestimmen, sind mindestens drei Zufallsvariablen erforderlich.
    d) Bedingte Kovarianz und bedingte Korrelation sind unter multivariater Normalverteilung immer identisch.

26. Jemand konfrontiert Sie mit einer signifikanten Korrelation zwischen dem Ausmaß des Konsums von Speiseeis und der Häufigkeit von Sonnenbränden in einem Gebiet, berechnet über eine Menge von Gebieten ($n = 67$, $r = 0.46$, $p < 0.001$). Wie reagieren Sie, wenn Sie diesen Befund im Sinne des Begriffs der partiellen Korrelation kritisieren?

    a) Ich bezweifle die Zuverlässigkeit, mit der die empirischen Daten erhoben wurden.
    b) Ich bezweifle die Sinnhaftigkeit wissenschaftlicher Studien, in der solche Paare von Variablen miteinander in Beziehung gesetzt werden.
    c) Ich schlage vor, eine zweite, deutlich größere Stichprobe zu erheben, um den Befund einer signifikanten Korrelation zu erhärten.
    d) Ich schlage vor, eine dritte, bisher nicht berücksichtigte Variable in Betracht zu ziehen, die sich mutmaßlich kausal auf sowohl Eiskonsum als auch Sonnenbrandinzidenz auswirkt.

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. Wie lässt sich die Formel für den Betaparameterschätzer des Allgemeinen Linearen Modells im Kontext des Modells der multiplen linearen Regression interpretieren?

    a) $\hat{\beta} \approx$ Regressorkovariabilität$^{-1}\cdot$ Regressordatenkovariabilität
    b) $\hat{\beta} \approx$ Regressordatenkovariabilität$^{-1}\cdot$ Regressorkovariabilität
    c) $\hat{\beta} \approx$ Stichprobenumfänge$^{-1}\cdot$ Datenpunktsummen
    d) $\hat{\beta} \approx$ Datenpunktsummen$^{-1}\cdot$ Stichprobenumfänge

29. Welches der folgenden ist kein faktorielles ALM-Design?

    a) Einstichproben-T-Test
    b) einfache lineare Regression
    c) einfaktorielle Varianzanalyse
    d) zweifaktorielle Varianzanalyse

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