Klausurreport Inferenzstatistik SoSe 2023

Die Klausur zum Modul B2 Inferenzstatistik im Sommersemester 2023 fand am 20.07.2023 von 13.00 - 14.00 Uhr in Hörsaal 6, Gebäude 44 der OVGU mit 78 Teilnehmer:innen statt. Die Klausur 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 grün markiert.

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 Prozentpunktintervallen.

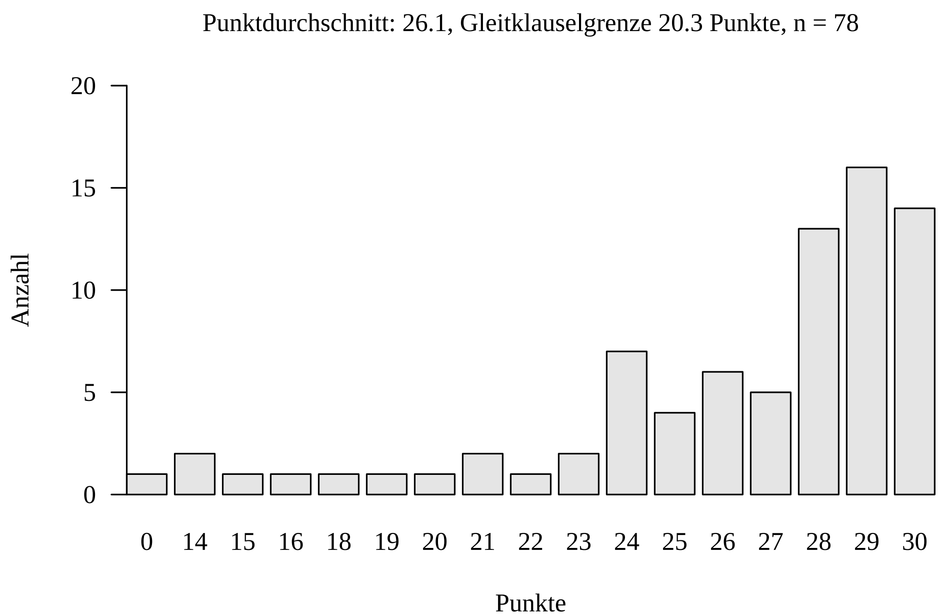
$\leq$	$\geq$	Note
100	95	1,0
94	90	1,3
89	85	1,7
84	80	2,0
79	75	2,3
74	70	2,7
69	65	3,0
64	60	3,3
59	55	3,7
54	50	4,0
49	0	5,0

Es ergibt sich folgendes Punktenotenschema, wobei < 15 Punkte mit 5.0 bewertet wurden.

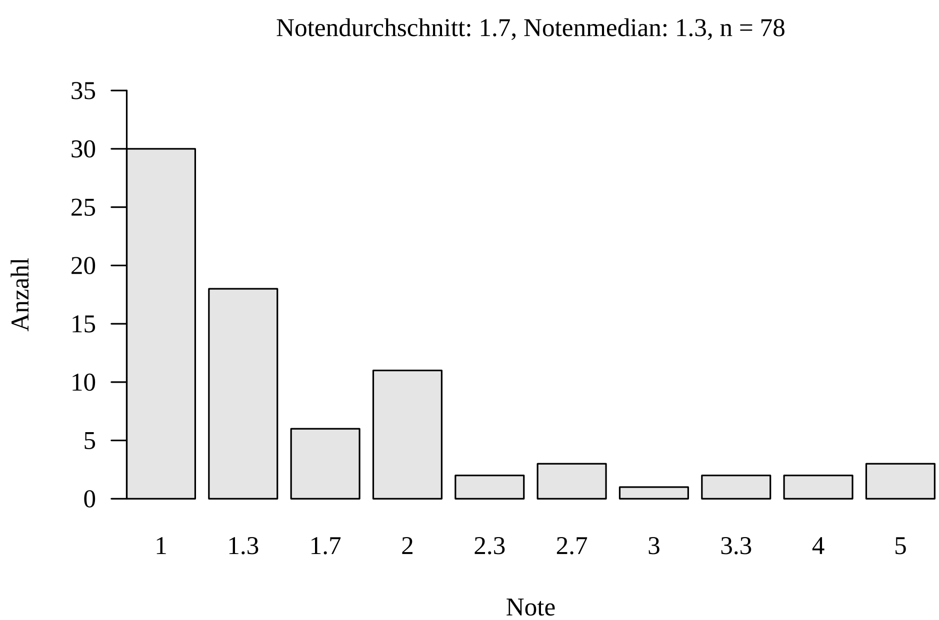
Punkte	Prozent	Note
30	100,0	1,0
29	96,7	1,0
28	93,3	1,3
27	90,0	1,3
26	86,7	1,7
25	83,3	2,0
24	80,0	2,0
23	76,7	2,3
22	73,3	2,7
21	70,0	2,7
20	66,7	3,0
19	63,3	3,3
18	60,0	3,3
17	56,7	3,7
16	53,3	4,0
15	50,0	4,0

Ergebnisse

Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erzielten Punkte.



Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erreichten Noten.



Otto-Von-Guericke-Universität Magdeburg
Institut für Psychologie
Abteilung Methodenlehre I: Methoden der experimentellen und neurowissenschaftlichen Psychologie
Prof. Dr. Dirk Ostwald

Klausur Modul B2 Inferenzstatistik

Termin: 20.07.2023

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 die zutreffende Antwort an.
- Für jede richtig gelöste Aufgabe erhalten Sie einen Punkt.

Viel Erfolg!

1. Welche Aussage zur Definition einer Ausgleichsgerade f_β trifft **nicht** zu?
   a) f_β ist eine linear-affine Funktion.
   b) Es gilt $f_\beta(x) := \beta_0 + \beta_1 x^2$.
   c) Die Ausgleichsgerade ist mit Bezug zu einem Datensatz definiert.
   d) Die Ausgleichsgerade minimiert eine Funktion von quadrierten vertikalen Abweichungen.

2. Welche Aussage zum Modell der einfachen linearen Regression

$$\upsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

   für $i = 1, ..., n$ trifft **nicht** zu?
   a) Die Zufallsvariablen υ_i modellieren Werte einer abhängigen Variable.
   b) Die x_i modellieren Werte einer unabhängigen Variable.
   c) Es gilt $\varepsilon_i \sim N(0, \sigma^2)$.
   d) Für $x_i = 0$ gilt $\upsilon_i = \beta_1$ für $i = 1, ..., n$.

3. Welche Aussage zur Stichprobenkorrelation eines Datensatzes $\{(x_i, y_i)\}_{i=1,...,n}$ bei linear-affinen Transformationen trifft zu?
   a) Multiplikation aller x_i Werte mit $a_x := 2$ verdoppelt die Stichprobenkorrelation.
   b) Multiplikation aller y_i Werte mit $a_y := 2$ verdoppelt die Stichprobenkorrelation.
   c) Addition von $b_x := 2$ zu allen x_i Werten verdoppelt die Stichprobenkorrelation.
   d) Der Betrag der Stichprobenkorrelation ändert sich bei linear-affinen Transformationen nicht.

4. Welche Aussage zum Bestimmtheitsmaß R^2 bei Ausgleichsgerade und Korrelation trifft **nicht** zu?
   a) R^2 ist die quadrierte Stichprobenkorrelation eines Datensatzes.
   b) Es gilt $0 \le \mathrm{R}^2 \le 1$.
   c) Für $\mathrm{R}^2 = 1$ ist die erklärte Streuung der Daten durch die Ausgleichsgerade gleich Null.
   d) R^2 ergibt sich aus dem Verhältnis von Explained Sum of Squares und Total Sum of Squares.

5. Es seien

$$A := \begin{pmatrix} 1 & 5 \\ 3 & 2 \end{pmatrix}, x := \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ und } b := \begin{pmatrix} 4 \\ 0 \end{pmatrix}. \tag{2}$$

   Welche Aussage trifft dann zu?
   a) $Ax + b = (12, 4)^T$.
   b) $Ax + b = (15, 2)^T$.
   c) $Ax + b = (11, 8)^T$.
   d) $Ax + b = (19, 1)^T$.

6. Es sei $x := (x_1, ..., x_n)^T \in \mathbb{R}^n$. Welche Aussage trifft dann zu?
   a) $x^T x = \sum_{i=1}^n x_i$
   b) $x^T x = \sum_{i=1}^n |x_i|$
   c) $x^T x = \sum_{i=1}^n x_i^2$
   d) $x^T x = \sum_{i=1}^n \sqrt{x_i}$

7. Welche Aussage zur Inversen einer Matrix trifft zu?
   a) Die Inverse einer Matrix ist nur für quadratische Matrizen definiert.
   b) Es gibt zu allen quadratischen Matrizen eine Inverse.
   c) Multiplikation einer Matrix mit ihrer Inversen ergibt eine Nullmatrix.
   d) Die Inverse zu einer Matrix A wird mit A^T bezeichnet.

8. Es sei $X = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Welche Aussage trifft dann zu?
   a) X ist eine symmetrische Matrix.
   b) X ist eine Diagonalmatrix.
   c) X ist eine Nullmatrix.
   d) X ist eine Einheitsmatrix.

9. Wie lautet die funktionale Form der WDF einer multivariaten Normalverteilung mit Parametern μ und Σ?
   a) $p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$.
   b) $p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^{-1} \Sigma^{T}(x-\mu)\right)$.
   c) $p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{T}(x-\mu)\right)$.
   d) $p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\Sigma)^T \mu^{-1}(x-\Sigma)\right)$.

10. Welche Aussage zum Theorem zu sphärischen (multivariaten) Normalverteilungen trifft **nicht** zu?
    a) Kovarianzmatrixparameter der Form $\sigma^2 I_n$ werden sphärisch genannt.
    b) Das Theorem stellt eine Beziehung zwischen der Wahrscheinlichkeitsdichtefunktion einer n-variaten Normalverteilung mit sphärischem Kovarianzmatrixparameter und der Wahrscheinlichkeitsdichtefunktion der gemeinsamen Verteilung von n univariaten normalverteilten Zufallsvariablen her.
    c) Die unabhängigen univariaten normalverteilten Zufallsvariablen des Theorems haben nach Voraussetzung den gleichen Erwartungswert.
    d) Die unabhängigen univariaten normalverteilten Zufallsvariablen des Theorems haben nach Voraussetzung die gleiche Varianz.

11. Welche Aussage zur Datenverteilung des Allgemeinen Linearen Modells trifft **nicht** zu?
    a) Die Daten werden als n-dimensionaler normalverteilter Zufallsvektor modelliert.
    b) Der Erwartungswertparameter der Daten ergibt sich aus Designmatrix und Betaparameter.
    c) Der Kovarianzmatrixparameter der Daten hat die Form einer Diagonalmatrix.
    d) Die Komponenten des Datenvektors sind immer unabhängig und identisch verteilt.

12. Welche Aussage zum Szenario von n unabhängig und identisch normalverteilten Zufallsvariablen in ALM Matrixschreibweise trifft **nicht** zu?
    a) Das Modell kann als $\upsilon = X\beta + \varepsilon$ mit $\varepsilon \sim N(0_n, \sigma^2 I_n)$ geschrieben werden.
    b) Das Modell kann als $\upsilon \sim N(X\beta, \sigma^2 I_n)$ geschrieben werden.
    c) Die Designmatrix des Modells ist die Einheitsmatrix, $X := I_n$.
    d) Für den Varianzparameter gilt $\sigma^2 > 0$.

13. Welche Aussage zum Szenario der einfachen linearen Regression in ALM Matrixschreibweise trifft **nicht** zu?
    a) Für den Betaparametervektor gilt $\beta \in \mathbb{R}^2$.
    b) Die Designmatrix hat mindestens drei Spalten.
    c) Eine Spalte der Designmatrix besteht nur aus Einsen.
    d) Eine Spalte der Designmatrix entspricht den Werten der unabhängigen Variable.

14. Welche Aussage zum Betaparameterschätzer $\hat{\beta}$ im ALM $\upsilon = X\beta + \varepsilon$ mit $\varepsilon \sim N(0_n, \sigma^2 I_n)$ trifft zu?
    a) Der Betaparameterschätzer hat die Form $(X^{-1}X)^T X^{-1}\upsilon$.
    b) Der Betaparameterschätzer ist ein verzerrter Schätzer von β.
    c) Der Betaparameterschätzer ist ein Maximum-Likelihood Schätzer von β.
    d) Die Frequentistische Verteilung des Betaparameterschätzers hängt nicht von der Designmatrix ab.

15. Es seien β und σ^2 die wahren, aber unbekannten Parameter des ALMs, $\hat{\beta}$ und $\hat{\sigma}^2$ ihre Schätzer, c ein Kontrastgewichtsvektor und β_0 ein Parameter. Welche Aussage zur Definition der T-Statistik trifft dann zu?
    a) Der Zähler der T-Statistik hat die Form $c^T\beta - c^T\beta_0$.
    b) Der Zähler der T-Statistik hat die Form $c^T\hat{\beta} - c^T\beta_0$.
    c) Der Zähler der T-Statistik hat die Form $\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}$.
    d) Der Zähler der T-Statistik hat die Form $\sqrt{\sigma^2 c^T (X^T X)^{-1} c}$.

16. Welche Aussage zu einer Likelihood-Quotienten-Statistik (LQS) trifft **nicht** zu?
    a) Eine LQS beruht auf zwei statistischen Modellen.
    b) Eine LQS setzt Wahrscheinlichkeitsmassen oder -dichten eines Datensatzes ins Verhältnis.
    c) Zur Bestimmung einer LQS müssen häufig zunächst Modellparameter optimiert werden.
    d) Likelihood-Funktionen sind für eine LQS irrelevant.

17. Welche Aussage zur F-Statistik trifft **nicht** zu?
    a) Die F-Statistik basiert auf einem vollständigen und einem reduzierten Modell.
    b) Der Zähler der F-Statistik misst eine Reduktion einer Residualquadratsumme.
    c) Der Nenner der F-Statistik entspricht einem Varianzparameterschätzer.
    d) Die Anzahl der zusätzlichen Betaparameter im vollständigen gegenüber dem reduzierten Modell sind für die F-Statistik ohne Bedeutung.

18. Welche Aussage zum Theorem zur Parameterschätzung im Einstichproben-T-Test Modell trifft zu?
    a) Der Betaparameterschätzer und das Stichprobenmittel der Daten sind identisch.
    b) Der Betaparameterschätzer ergibt sich zu $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \upsilon_i^2$.
    c) Der Varianzparameterschätzer und die Stichprobenstandardabweichung der Daten sind identisch.
    d) Der Varianzparameterschätzer ergibt sich zu $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} |\upsilon_i - \bar{\upsilon}|$.

19. Welche Aussage zum Theorem zur T-Teststatistik des Einstichproben-T-Tests trifft zu?
    a) Der Kontrastgewichtsvektor hat die Form $c = (1, -1)^T$.
    b) Die T-Teststatistik hängt von der Stichprobengröße ab.
    c) Die T-Teststatistik ist im Allgemeinen nichtzentral χ^2-verteilt.
    d) Der Nichtzentralitätsparameter der Verteilung der T-Teststatistik ist gegeben durch $\delta := \mu - \mu_0$.

20. Welche Aussage zum Anwendungsszenario eines Zweistichproben-T-Tests trifft **nicht** zu?
    a) Man geht von zwei Gruppen (Stichproben) randomisierter experimenteller Einheiten aus.
    b) Man setzt die Varianzparameter beider Gruppen als bekannt voraus.
    c) Man setzt die Varianzparameter beider Gruppen als identisch voraus.
    d) Man beabsichtigt, die Unsicherheit beim Vergleich der Gruppenerwartungswertparameter zu quantifizieren.

21. Welche Aussage zum Anwendungsszenario einer einfaktoriellen Varianzanalyse (EVA) trifft **nicht** zu?
    a) Man geht von zwei oder mehr Gruppen (Stichproben) randomisierter experimenteller Einheiten aus.
    b) Man nimmt an, dass die Zufallsfehlervariablen unabhängig und identisch normalverteilt sind.
    c) Man setzt die Varianzparameter aller Gruppen als identisch voraus.
    d) Man ist an Haupteffekten und Interaktionen der beiden Faktoren interessiert.
22. Welche Aussage zur Betaparameterschätzung im EVA Modell in Effektdarstellung mit Referenzgruppe trifft zu?
    a) Der Referenzgruppeneffektparameter wird durch das Stichprobenmittel aller Datenpunkte geschätzt.
    b) Der Referenzgruppeneffektparameter wird durch die Stichprobenvarianz aller Datenpunkte geschätzt.
    c) Die Effektparameter der Nicht-Referenzgruppen werden durch die Stichprobenmittel der entsprechenden Gruppen geschätzt.
    d) Die Effektparameter der Nicht-Referenzgruppen werden durch die Differenzen zwischen ihren Stichprobenmitteln und dem Stichprobenmittel der Referenzgruppe geschätzt.
23. In einem EVA Szenario seien die Between Sum of Squares, Within Sum of Squares und Total Sum of Squares mit SQB, SQW und SQT, respektive, bezeichnet. Welche Aussage zum Effektstärkemaß η^2 trifft dann zu?
    a) $\eta^2 = \text{SQB}/\text{SQT}$.
    b) $\eta^2 = \text{SQW}/\text{SQT}$.
    c) $\eta^2 = \text{SQW}/\text{SQB}$.
    d) $\eta^2 = \text{SQB}/\text{SQW}$.
24. Welche Aussage zur intuitiven Bedeutung von Haupteffekten und Interaktionen im Kontext einer 2×2 zweifaktoriellen Varianzanalyse mit Faktoren A und B trifft **nicht** zu?
    a) Intuitiv spricht man vom Vorliegen eines Haupteffekts von Faktor A, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor A, jeweils gemittelt über die zwei Level von Faktor B, unterscheiden.
    b) Intuitiv spricht man vom Vorliegen eines Haupteffekts von Faktor B, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor B, jeweils gemittelt über die zwei Level von Faktor A, unterscheiden.
    c) Intuitiv spricht man vom Vorliegen einer Interaktion der Faktoren A und B, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor B unterscheiden.
    d) Intuitiv beziehen sich Haupteffekte auf Unterschiede (Differenzen), während sich Interaktionen auf Unterschiede von Unterschieden (Differenzen von Differenzen) beziehen.
25. Es seien

$$X_1 := \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad X_2 := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad X_3 := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad X_4 := \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{3}$$

    Welche dieser Matrizen ist eine mögliche Designmatrix des Modells einer balancierten 2×2 zweifaktoriellen Varianzanalyse mit Interaktion und Referenzgruppe und 2 Datenpunkten pro Designzelle?

    a) X_1 b) X_2 c) X_3 d) X_4.

26. x, y und z seien drei gemeinsam multivariat normalverteilte Zufallsvariablen. Welche Aussage zur bedingten Korrelation $\rho(x, y|z)$ von x und y gegeben z trifft dann **nicht** zu?

   a) $\rho(x, y|z)$ ist die Korrelation von x und y in der bedingten Verteilung von x und y gegeben z.

   b) Zur Definition von $\rho(x, y|z)$ benötigt man unter anderem die bedingte Kovarianz $\mathbb{C}(x, y|z)$.

   c) $\rho(x, y|z)$ kann aus den paarweisen Korrelationen $\rho(x, y)$, $\rho(x, z)$ und $\rho(y, z)$ berechnet werden.

   d) $\rho(x, y|z)$ und die Korrelation $\rho(x, y)$ von x und y sind immer gleich.

27. x, y und z seien drei gemeinsam multivariat normalverteilte Zufallsvariablen. Welche Aussage zur partiellen Korrelation $\rho(x, y\backslash z)$ von x und y gegeben z trifft dann **nicht** zu?

   a) Die Definition von $\rho(x, y\backslash z)$ basiert auf linear-affinen Abhängigkeiten zwischen x und z sowie y und z.

   b) $\rho(x, y\backslash z)$ ist als eine Korrelation von Residualvariablen definiert.

   c) Intuitiv ist $\rho(x, y\backslash z)$ die Korrelation von x und y, aus denen der Einfluss von z herausgerechnet wurde.

   d) $\rho(x, y\backslash z)$ und die bedingte Korrelation $\rho(x, y|z)$ von x und y gegeben z sind immer verschieden.

28. Was ist **kein** typisches Ziel einer multiplen Regressionsanalyse?

   a) Die Quantifizierung des Erklärungspotentials der Variation unabhängiger Variablen für die Variation einer abhängigen Variablen.

   b) Die Quantifizierung des Einflusses einzelner unabhängiger Variablen auf die abhängige Variable im Kontext anderer unabhängiger Variablen.

   c) Die Quantifzierung von Unsicherheit beim inferentiellen Vergleich zweier Gruppenerwartungswertparameter.

   d) Die Prädiktion von Werten einer abhängigen Variable aus Werten von unabhängigen Variablen nach Parameterschätzung.

29. Welche Aussage zu den Betaparameterschätzern im Rahmen der multiplen Regression und partiellen Stichprobenkorrelationen trifft zu?

   a) Die Betaparameterschätzer und partiellen Stichprobenkorrelationen hängen nicht zusammen.

   b) Die Betaparameterschätzer können sich von partiellen Stichprobenkorrelationen unterscheiden.

   c) Die Betaparameterschätzer sind immer mit partiellen Stichprobenkorrelationen identisch.

   d) Es gilt immer "Betaparameterschätzer $=$ Anzahl Datenpunkte $\cdot$ Partielle Stichprobenkorrelation^{-1}".

30. Welcher Kontrastgewichtsvektor kann zum Testen der Nullhypothese zweier gleicher Regressoreffekte im Rahmen eines T-Tests bei einer multiplen Regressionsanalyse mit drei Regressoren benutzt werden?

   a) $c = (1, 0, 0)$.

   b) $c = (1, 0, 1)$.

   c) $c = (1, -1, 0)$.

   d) $c = (1, -1, -1)$.