

Klausurreport Computergestützte Datenanalyse SoSe 2025

Die Klausur zum Modul C: Computergestützte Datenanalyse im Sommersemester 2025 fand am 25.07.2025 von 10.00 - 11.00 Uhr in den Computer-Pools des Universitätsrechenzentrums (URZ) mit 58 Teilnehmer:innen als elektronische Klausur in Präsenz statt. Die Klausur bestand aus 30 Multiple Choice Aufgaben mit jeweils vier Antwortmöglichkeiten und jeweils genau einer richtigen Antwort. Sie war in 2 Teile mit jeweils 15 Fragen untergliedert. Teil 1 war closed-book und Teil 2 open-book. Für den open-book Teil konnten Prüfungsteilnehmende das Internet und R am Prüfungs-PC nutzen. Die Klausur ist diesem Bericht beigelegt, richtige Antworten sind auf der letzten Seite angegeben.

Bewertungsschema

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.

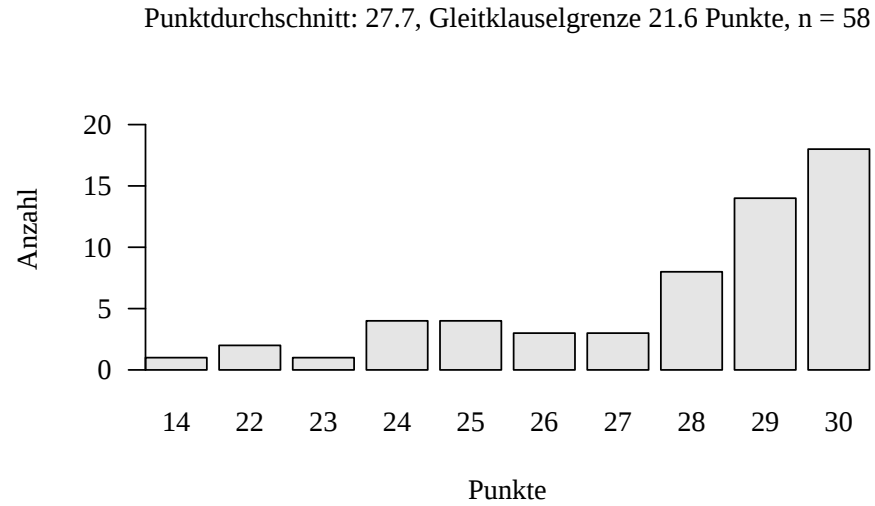
$\leq$	$\geq$	Note
100	95	1,0
94	90	1,3
89	85	1,7
84	80	2,0
79	75	2,3
74	70	2,7
69	65	3,0
64	60	3,3
59	55	3,7
54	50	4,0
49	0	5,0

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

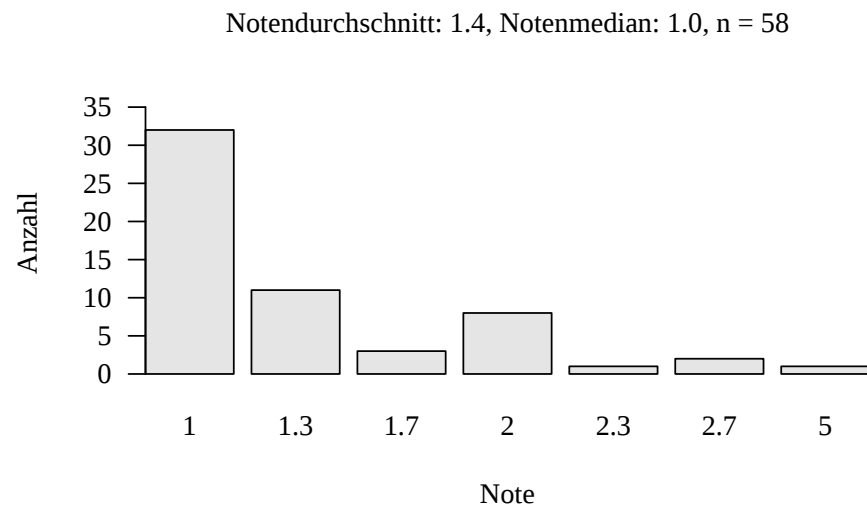
Punkte	Prozent	Note
30	100.00	1,0
29	96.67	1,0
28	93.33	1,3
27	90.00	1,3
26	86.67	1,7
25	83.33	2,0
24	80.00	2,0
23	76.67	2,3
22	73.33	2,7
21	70.00	2,7
20	66.67	3,0
19	63.33	3,3
18	60.00	3,3
17	56.67	3,7
16	53.33	4,0
15	50.00	4,0

Ergebnisse

Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erzielten Punkte.

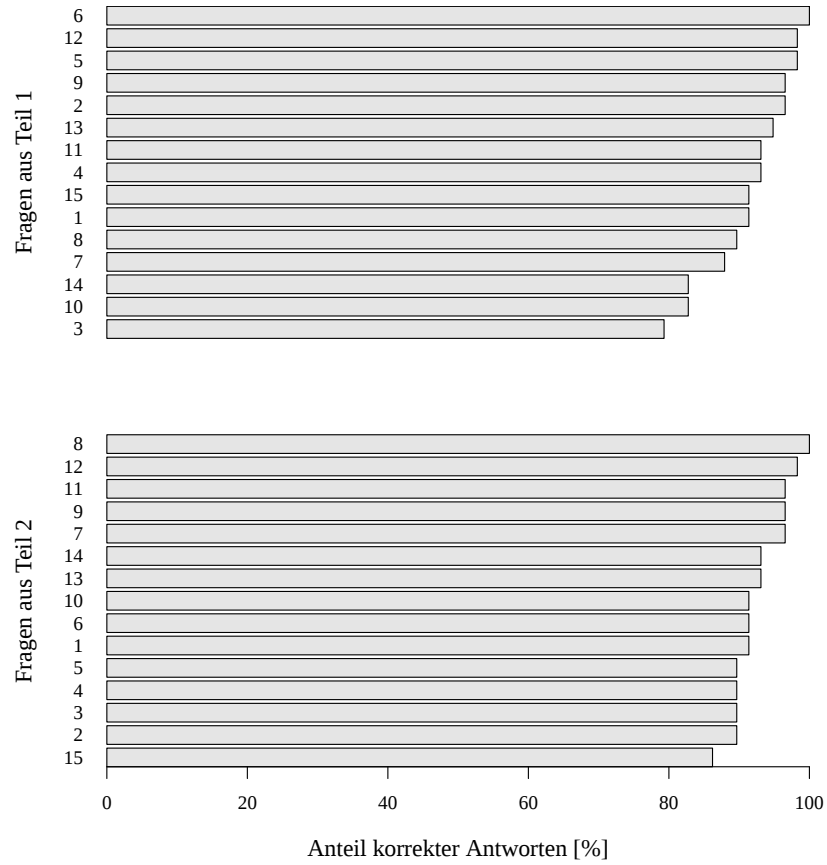


Die nachfolgende Abbildung zeigt die absolute Häufigkeitsverteilung der erreichten Noten.



Die zwei nachfolgenden Abbildungen zeigen jeweils die Reihenfolge der Klausurfragen, sortiert nach Anteil korrekter Antworten über alle Teilnehmenden für beide Teile der Klausur.

Klausurfragen sortiert nach Anzahl korrekter Antworten



OTTO-VON-GUERICKE-UNIVERSITÄT MAGDEBURG

Institut für Psychologie

Abteilung Methodenlehre I: Methoden der experimentellen und neurowissenschaftlichen Psychologie

Belinda Fleischmann und Prof. Dr. Dirk Ostwald

Klausur Modul C Einführung in empirisch-wissenschaftliches Arbeiten

Termin: 25.07.2025

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 die zutreffende Antwort an.
- Für jede richtig gelöste Aufgabe erhalten Sie einen Punkt.
- Die Kommunikation mit anderen Klausurteilnehmer:innen ist nicht gestattet.

Viel Erfolg!

Teil 1 - Inhalte aus Einheiten (1) bis (3)

Alle Fragen zu Ethik und ethischen Formalitäten beziehen sich auf die „Deklaration von Helsinki“ der *World Medical Association (WMA)*, die „Ethical Principles of Psychologists and Code of Conduct (EPPCC)“ der *American Psychological Association (APA)* und den Leitfaden „Ethisches Handeln in der psychologischen Forschung“ der *Deutsche Gesellschaft für Psychologie (DGPs)*.

1. Welche Aussage bezüglich rechtlicher Normen für ethische Fragestellungen in der Forschung ist **nicht** zutreffend?
 - a) In Deutschland gibt es einen klaren gesetzlichen Rahmen zur Forschung an und mit Menschen.
 - b) Artikel 1 und 2 des Grundgesetzes dienen als Grundlage für ethische Überlegungen und Richtlinien.
 - c) Das Arzneimittelgesetz ist im Bereich humanmedizinischer Forschung relevant.
 - d) Das Medizinprodukte-Durchführungsgesetz ist im Bereich humanmedizinischer Forschung relevant.
2. Ein Universitätsprofessor plant eine Studie in seiner Vorlesung über den Einfluss von Stress auf Leistung. Die Rekrutierung der Studierenden erfolgt während der Vorlesung, und die Teilnahmebedingungen umfassen Fragebögen und Leistungsnachweise, die ohnehin Pflichtbestandteil des Kurses sind. Welche ethischen Bedenken könnten hierbei aufkommen?
 - a) Die Teilnahme an der Studie sollte verpflichtend sein, um eine repräsentative Stichprobe zu gewährleisten.
 - b) Es besteht das Risiko, dass Studierende an der Studie teilnehmen, um den Erwartungen des Professors gerecht zu werden, was die Freiwilligkeit der Teilnahme beeinträchtigen könnte.
 - c) Die Rekrutierung während der Vorlesung ist unproblematisch, da alle Studierenden ohnehin an der Vorlesung teilnehmen müssen.
 - d) Die ethische Unbedenklichkeit der Studie ist gesichert, wenn vollständige Anonymität der Teilnahme sowie der erhobenen Daten gewährleistet ist.
3. Welche Aussage bezüglich Täuschung in einer Studie ist **nicht** zutreffend?
 - a) Wenn möglich, keine Täuschung und wenn nötig, so wenig Täuschung wie möglich.
 - b) Täuschung in Studien bedeutet, dass spezifische Informationen vorenthalten werden oder sogar falsche Informationen gegeben werden.
 - c) Das „Debriefing“ sollte im besten Fall am Ende der Teilnahme und spätestens am Ende der Datenerhebung stattfinden.
 - d) Täuschung ist grundsätzlich unvereinbar mit dem Prinzip der informierten Einwilligung und ist deshalb ethisch nicht vertretbar.
4. Welches der folgenden Kriterien ist **keine** zentrale Voraussetzung für eine gültige Einwilligungserklärung in einer Studie?
 - a) Informiertheit
 - b) Freiwilligkeit
 - c) Kompetenz und Entscheidungsfähigkeit
 - d) Aufwandsentschädigung

5. Welche der folgenden Aussagen zu den Risikotypen im Rahmen psychologischer Studien ist **nicht** zutreffend?
- a) Beeinträchtigungen des Selbstwerts sind ein Beispiel für ein psychologisches Risiko.
 - b) Kompromittierung des persönlichen sozialen Status ist ein Beispiel für ein soziales Risiko.
 - c) Ablehnung von privaten Krankenversicherungen ist ein Beispiel für ein ökonomisches Risiko.
 - d) Ein Risiko gilt als minimal, wenn es ausschließlich mentale, aber keine körperlichen Konsequenzen mit sich bringt.
6. Was ist der Unterschied zwischen anonymisierten und pseudonymisierten Daten?
- a) Bei Anonymisierung ist eine indirekte fehlerfreie Zuordnung von Daten zu einer Person noch möglich, bei Pseudonymisierung nicht.
 - b) Beide Varianten enthalten Pseudonyme, aber nur anonymisierte Daten enthalten auch Anonyme.
 - c) Pseudonymisierte Daten enthalten auch Dummy-Variablen, anonymisierte Daten nicht.
 - d) Anonymisierte Daten können nicht zu einer Person zugeordnet werden, pseudonymisierte Daten können prinzipiell einer Person zugeordnet werden.
7. Welches der folgenden Beispiele trägt zur Wahrung von Gerechtigkeit in psychologischen Studien bei?
- a) Die Bevorzugung von Minderheiten und vulnerablen Gruppen bei der Rekrutierung.
 - b) Der Ausschluss von Minderheiten und vulnerablen Gruppen bei der Rekrutierung.
 - c) Die verständliche Aufklärung über das Randomisierungsprinzip.
 - d) Eine Aufwandsentschädigung, die unabhängig vom Aufwand der Teilnahme festgelegt wird.
8. Welche der folgenden ist **kein** Mindestbestandteil eines Ethikantrags?
- a) Darstellung des Studienvorhabens
 - b) Teilnehmendeninformation
 - c) Einwilligungserklärung
 - d) Information über Bild- oder Tonaufnahmen
9. Welcher der folgenden ist **kein** Bestandteil des „Goldenen Wegs“ wissenschaftlichen Publizierens?
- a) Die Zweitveröffentlichung des Artikels in einem Online-Repository.
 - b) Der Artikel wird für alle frei zugänglich gemacht.
 - c) Die Autor:innen oder ihre Institution bezahlt eine Gebühr an den Verlag, um Kosten für Redaktionsprozess, technische Struktur und Marketing zu decken.
 - d) Die Editor:innen des Verlags schicken den Artikel an andere Wissenschaftler:innen zur Überprüfung.

10. Welcher der folgenden Aspekte trifft **nicht** auf den Journal Impact Factor zu?
- a) Der Journal Impact Factor wird von Unternehmen veröffentlicht.
 - b) Der Journal Impact Factor ist von der Größe eines Forschungsfeldes abhängig.
 - c) Der Journal Impact Factor ist ein quantitatives Maß für die wissenschaftliche Qualität einer Zeitschrift.
 - d) Zur Berechnung des Journal Impact Factors wird die Anzahl an Zitationen von Artikeln einer Fachzeitschrift der letzten zwei Jahre ins Verhältnis zu der Anzahl an Publikationen von Artikeln der Fachzeitschrift der letzten zwei Jahre gesetzt.
11. Welche drei zentralen Inhalte gehören in die Einleitung einer empirisch-quantitativen Arbeit nach den „Journal Article Reporting Standards (JARS)“ der APA?
- a) Eine Beschreibung des Studiendesigns, die wichtigsten Ergebnisse und deren Limitationen.
 - b) Den theoretischen Hintergrund, die Forschungsfrage/-n und die Hypothese/-n.
 - c) Eine detaillierte Darstellung der Stichprobe, Materialien und Verfahren.
 - d) Eine Interpretation der Ergebnisse im Kontext bestehender Literatur.
12. Welcher der folgenden Aspekte wird laut den „Journal Article Reporting Standards (JARS)“ der APA typischerweise in der Diskussion und nicht in den Ergebnissen behandelt?
- a) Die Ergebnisse aller Datenanalysen
 - b) Die methodischen Details zu Datenerhebung und Analyse
 - c) Die Bedeutung der Ergebnisse im Kontext bestehender Literatur
 - d) Die Beschreibung von Abbildungen und Tabellen
13. Was ist **kein** Unterschied zwischen Replizierbarkeit und Reproduzierbarkeit?
- a) Replizierbarkeit liegt vor, wenn andere Wissenschaftler:innen andere (neue) Daten und möglicherweise andere Methoden nutzen und zu den gleichen Schlussfolgerungen kommen wie eine Originalstudie, während Reproduzierbarkeit dann vorliegt, wenn andere Wissenschaftler:innen die gleichen Daten und Methoden verwenden und die gleichen Resultate erhalten.
 - b) Replizierbarkeit ist abhängig von den unterstellten wahren, aber unbekannten, Parametern. Reproduzierbarkeit ist abhängig von der Sorgfältigkeit der wissenschaftlichen Arbeit.
 - c) Für die Überprüfung der Reproduzierbarkeit benötigen Wissenschaftler:innen Zugriff auf die Daten der Originalstudie, für die Überprüfung der Replizierbarkeit nicht.
 - d) Für Replizierbarkeit ist größtmögliche Methoden- und Datentransparenz wichtig, für Reproduzierbarkeit nicht.
14. Welcher der folgenden Punkte ist ein ethisches Argument für Methoden- und Datentransparenz?
- a) Die Beseitigung jeglicher Forschungsrisiken
 - b) Die Sicherstellung der Replizierbarkeit einer Studie
 - c) Der Respekt im Umgang mit öffentlichen Geldern
 - d) Die Maximierung der Zahl an veröffentlichten Studien

15. Welche der folgenden ist ein Datenstandard, der in der psychologischen Forschung verwendet wird?

- a) Open Science Framework
- b) Brain Imaging Data Structure (BIDS)
- c) GitHub
- d) DSGVO

Teil 2 - Praktische Anwendung ALM

Einfaktorielle Varianzanalyse

In den folgenden Aufgaben werden zentrale Aspekte einer einfaktoriellen Varianzanalyse (EVA) behandelt. Der untenstehende **R** Code definiert den **R** Dataframe D, der einen Datensatz mit vier Gruppen (A, B, C und D) und 56 Beobachtungen umfasst, wobei jede Gruppe genau 14 Beobachtungen enthält. Die erste Spalte gibt die Gruppenzugehörigkeit an, die zweite enthält die jeweiligen Werte. Alle Referenzierungen des Datensatzes D im Folgenden beziehen sich auf den hier definierten Dataframe.

```
D <- data.frame(
  Gruppe = c(
    rep("A", 14),
    rep("B", 14),
    rep("C", 14),
    rep("D", 14)
  ),
  Werte = c(
    14.6, 16.2, 20.2, 17.0, 21.6, 16.5, 15.9, 17.7, 17.8, 17.5, 21.4, 19.1, 21.3, 16.5,
    18.6, 21.3, 21.6, 21.0, 20.4, 19.0, 21.0, 20.7, 21.5, 23.9, 22.6, 23.3, 22.5, 23.4,
    15.0, 16.8, 11.8, 16.0, 19.4, 18.8, 13.9, 19.6, 16.7, 17.8, 16.1, 17.5, 12.0, 19.1,
    19.2, 20.3, 22.7, 21.7, 20.8, 24.4, 20.6, 22.2, 21.2, 20.9, 19.0, 23.8, 21.3, 19.4
  )
)
```

1. Welche der folgenden **R** Befehle bestimmt **nicht** die Anzahl **n** der Beobachtungen?

- a) `nrow(D)`
- b) `length(D$Werte)`
- c) `length(D$Gruppe)`
- d) `length(D)`

2. Welche der folgenden **R** Befehle bestimmt die Anzahl der Gruppen?

- a) `nrow(D$Gruppe)`
- b) `length(unique(D$Gruppe))`
- c) `ncol(D)`
- d) `length(D$Gruppe)`

3. n sei die Anzahl an Beobachtungen, p die Anzahl an Gruppen und n_i ein **R** Vektor mit vier Einträgen, der die Gruppengrößen des durch D repräsentierten Datensatzes enthält. Welcher der folgenden **R** Befehle erstellt eine Designmatrix für ein EVA-Modell in Effektdarstellung mit Referenzgruppe?

- a) `matrix(c( rep(1, n_i[1]), rep(1, n_i[2]), rep(1, n_i[3]), rep(1, n_i[4]), rep(0, n_i[1]), rep(1, n_i[2]), rep(0, n_i[3]), rep(0, n_i[4]), rep(0, n_i[1]), rep(0, n_i[2]), rep(1, n_i[3]), rep(0, n_i[4]), rep(0, n_i[1]), rep(0, n_i[2]), rep(0, n_i[3]), rep(1, n_i[4]) ), nrow = n)`
- b) `matrix(c( rep(1, n_i[1]), rep(0, n_i[2]), rep(0, n_i[3]), rep(0, n_i[4]), rep(1, n_i[1]), rep(1, n_i[2]), rep(0, n_i[3]), rep(0, n_i[4]), rep(1, n_i[1]), rep(0, n_i[2]), rep(1, n_i[3]), rep(0, n_i[4]), rep(1, n_i[1]), rep(0, n_i[2]), rep(0, n_i[3]), rep(1, n_i[4]) ), nrow = n)`
- c) `kronecker(diag(p), matrix(1, nrow = n, ncol = 1))`
- d) `kronecker(diag(p), matrix(1, nrow = n, ncol = 1))[, -2]`

4. X sei die **R** Repräsentation der Designmatrix eines vollständigen EVA-Modells mit vier Betaparametern, n die Anzahl an Beobachtungen im Datensatz D . Welche der folgenden **R** Codes erzeugt die Designmatrix des reduzierten EVA-Modells mit einem Betaparameter?

- a) `matrix(rep(1, n), nrow = n)`
- b) `matrix(c(rep(1, n), rep(0, n)), nrow = n)`
- c) `matrix(c(D$Werte), nrow = n)`
- d) `X[, -1]`

5. X sei die **R** Repräsentation der Designmatrix eines vollständigen EVA-Modells in Effektdarstellung mit Referenzgruppe. Welche der folgenden **R** Codes bestimmt die Anzahl der Betaparameter?

- a) `ncol(X) - 1`
- b) `ncol(X)`
- c) `ncol(X) + 1`
- d) `length(X[, 1])`

6. **X** sei die **R** Repräsentation der Designmatrix des vollständigen EVA-Modells. Welcher der folgenden Befehle berechnet den Betaparameterschätzer, gegeben die Werte des Datensatzes **D**?
- a) `solve(t(X) %*% X) %*% t(X) %*% D$Werte`
 - b) `solve(t(X) %*% X) %*% t(X) %*% y`
 - c) `coef(lm(Werte ~ Gruppe, data = D$Werte))`
 - d) `mean(D$Werte)`
7. **X** sei die **R** Repräsentation der Designmatrix des vollständigen EVA-Modells und **beta_hat** der Betaparameterschätzer des vollständigen EVA-Modells. Welcher der folgenden Befehle berechnet den Residuenvektor des vollständigen EVA-Modells basierend auf den Werten des Datensatzes **D**?
- a) `D$Werte - X %*% beta_hat`
 - b) `D$Werte - beta_hat %*% t(X)`
 - c) `D$Werte - beta_hat %*% X`
 - d) `D$Werte - mean(D$Werte) %*% beta_hat`
8. **X_0** sei die **R** Repräsentation der Designmatrix des reduzierten EVA-Modells und **beta_hat_0** der Betaparameterschätzer des reduzierten EVA-Modells. Welcher der folgenden Befehle berechnet den Residuenvektor des reduzierten EVA-Modells basierend auf den Werten des Datensatzes **D**?
- a) `D$Werte - X %*% beta_hat`
 - b) `residuals(lm(Werte ~ Gruppe, data = D))`
 - c) `D$Werte - predict(lm(Werte ~ Gruppe, data = D))`
 - d) `D$Werte - X_0 %*% beta_hat_0`
9. **eps_hat_0** sei die **R** Repräsentation des Residuenvektors des reduzierten Modells einer EVA. Welcher der folgenden Befehle berechnet die total sum of squares (SQT)?
- a) `t(eps_hat_1) %*% eps_hat_1`
 - b) `t(eps_hat_0) %*% eps_hat_0`
 - c) `t(eps_hat) %*% eps_hat`
 - d) `t(eps_hat_0) %*% eps_hat_0 - t(eps_hat) %*% eps_hat`

10. p sei die Anzahl an Gruppen und n die Anzahl an Beobachtungen im Datensatz D . Welcher der folgenden **R** Befehle bestimmt die Zähler-Freiheitsgrade einer F-Statistik der EVA?
- a) $n - 1$
 - b) $n - p$
 - c) $p - 1$
 - d) $n - 2$
11. p sei die Anzahl an Gruppen und n die Anzahl an Beobachtungen im Datensatz D . Weiterhin seien **SQB** die between sum of squares, **SQW** die within sum of squares und **SQT** die total sum of squares einer EVA. Welcher der folgenden **R** Befehle berechnet die mean between sum of squares (MSB)?
- a) $SQW / (n - p)$
 - b) $SQT / (n - 1)$
 - c) $SQB / (n - 1)$
 - d) $SQB / (p - 1)$
12. p sei die Anzahl an Gruppen und n die Anzahl an Beobachtungen im Datensatz D . Weiterhin seien **SQB** die between sum of squares, **SQW** die within sum of squares, **MSB** die mean between sum of squares und **MSW** die mean within sum of squares einer EVA. Welcher der folgenden **R** Befehle berechnet die F-Statistik?
- a) MSW / MSB
 - b) MSB / MSW
 - c) SQB / SQW
 - d) $(SQB / (n - p)) / (SQW / (p - 1))$
13. p sei die Anzahl an Gruppen und n die Anzahl an Beobachtungen im Datensatz D und **F_stat** die F-Statistik einer EVA. Welcher der folgenden **R** Befehle berechnet den p-Wert für den F-Test?
- a) $1 - pf(F_stat, p - 1, n - p)$
 - b) $pf(F_stat, p - 1, n - p)$
 - c) $1 - pf(F_stat, n - p, p - 1)$
 - d) $pf(F_stat, n - p, p - 1)$

14. Welcher der folgenden Befehle führt eine EVA mit dem Datensatz D in **R** durch?

- a) `summary(res.aov)`
- b) `aov(Werte ~ Gruppe, data = D)`
- c) `eva(Werte ~ Gruppe, data = D)`
- d) `summary(res)`

15. Gegeben sei untenstehender **R** Output der Zusammenfassung einer mit `aov()` erfolgten EVA. Welche Aussage ist **nicht** zutreffend?

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gruppe	3	253.1	84.36	20.22	8.09e-09 ***
Residuals	52	216.9	4.17		

- a) Die within sum of squares beträgt 216.9.
- b) Die F-Statistik beträgt 20.22.
- c) Die between sum of squares beträgt 216.9.
- d) Der p-Wert ist kleiner als 0.001.

Lösungen

Teil 1

1. a)
2. b)
3. d)
4. d)
5. d)
6. d)
7. c)
8. d)
9. a)
10. c)
11. b)
12. c)
13. d)
14. c)
15. b)

Teil 2

1. d)
2. b)
3. a)
4. a)
5. b)
6. a)
7. a)
8. d)
9. b)
10. c)
11. d)
12. b)
13. a)
14. b)
15. c)