

Ergebnisbericht

Der Leistungsnachweis für Modul C1 – Programmierung und Deskriptive Statistik im Wintersemester 2022/2023 fand in zwei Teilen als Online Schriftliche Ausarbeitung (Open Book Exam) statt. An Teil 1 am 19.12.2022 nahmen 68 Studierende und an Teil 2 am 27.01.2023 66 Studierende teil. Der erste Teil umfasste 15 und der zweite 10 Multiple Choice Aufgaben mit jeweils vier Antwortmöglichkeiten und genau einer richtigen Antwort. Die Lösungen sind diesem Bericht beigelegt.

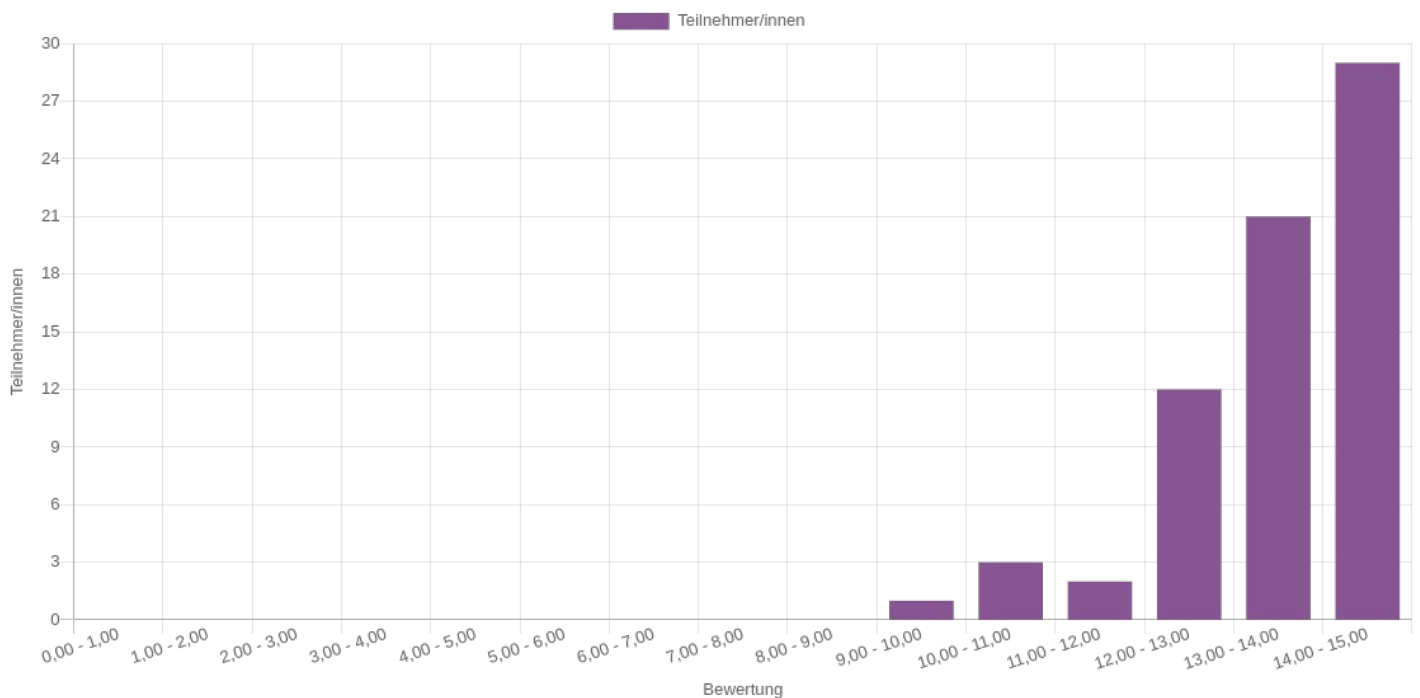
Bewertungsschema

Für jede richtige Antwort gab es einen Punkt. Somit konnten im ersten Teil 15 und im zweiten 10 Punkte erreicht werden. Der Leistungsnachweis war unbenotet und galt als bestanden, wenn in beiden Teilen mindestens die Hälfte der Punkte erreicht wurde.

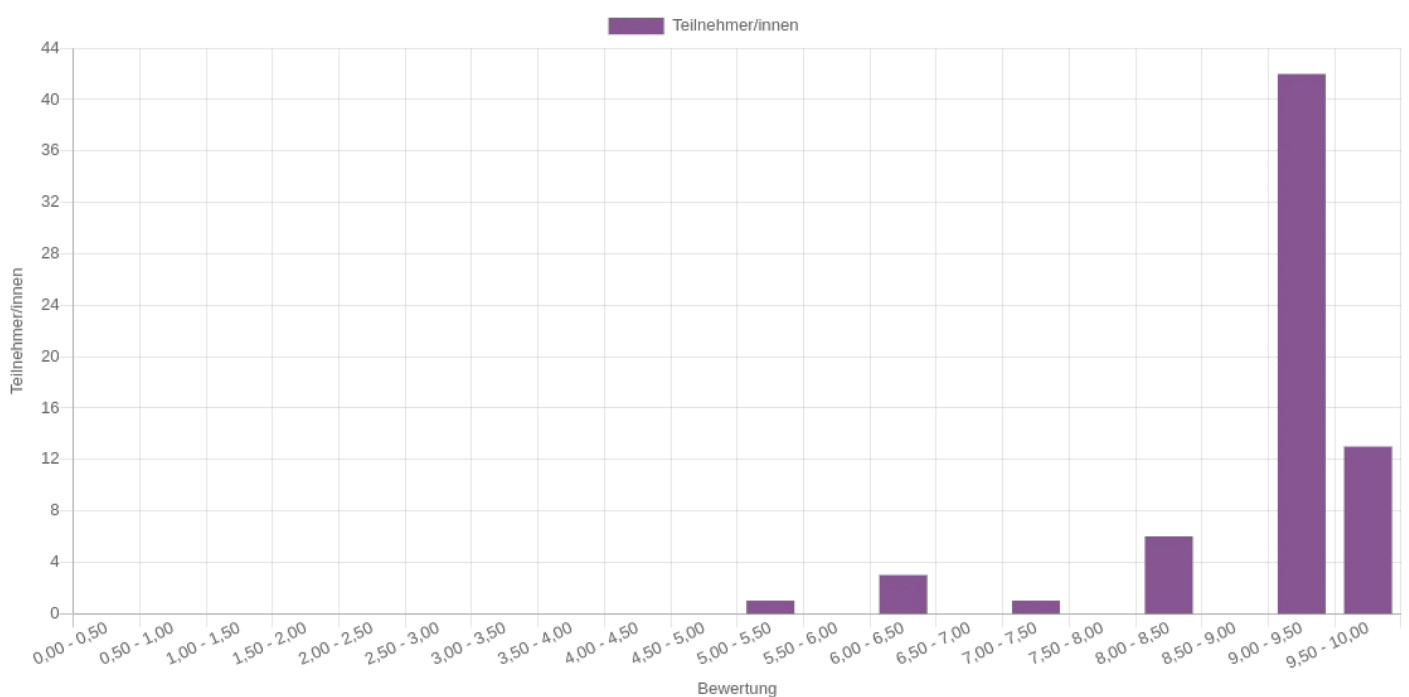
Ergebnisse

Die nachfolgenden Abbildungen zeigen die absolute Häufigkeitsverteilung der erzielten Punkte.

Teil 1 (19.12.2022)



Teil 2 (27.01.2023)



Fragen und Lösungen – Teil 1

Frage 1

Erreichbare Punkte: 1,00

Welche R Datenstruktur ist atomar?

- a. Matrix
- b. Dataframe
- c. Liste
- d. Eine Liste von Listen

Die richtige Antwort ist:

Matrix

Frage 2

Erreichbare Punkte: 1,00

Gegeben der in R vordefinierten Operatorpräzedenz, welcher R Konsolenoutput resultiert aus folgendem R Code?

```
1+(-3)^2
```

- a. `> [1] -10`
- b. `> [1] 8`
- c. `> [1] 10`
- d. `> [1] -8`

Die richtige Antwort ist:

```
> [1] 10
```



Frage 3

Erreichbare Punkte: 1,00

Welche Aussage bezüglich Histogrammen ist zutreffend?

- a. In Histogrammen werden die relativen Häufigkeiten aller Zahlenwerte eines Datensatzes visualisiert.
- b. In R können Histogramme mit dem Befehl **histplot()** erzeugt werden.
- c. Bei gegebenem k können für die Bestimmung der Intervallgrenzen $[b_{j-1}, b_j]$ für $j = 1, \dots, k$ die Werte für b_0 und b_k frei gewählt werden.
- d. Keine der Aussagen ist zutreffend.

Die richtige Antwort ist:

Keine der Aussagen ist zutreffend.

Frage 4

Erreichbare Punkte: 1,00

Welche Aussage bezüglich Vektorindizierung in R trifft **nicht** zu?

- a. Nichtganzzahlige Indizes werden abgerundet.
- b. Vektoren in R können mit logischen Vektoren indiziert werden.
- c. Indizierung mit einem character Vektor verursacht einen Fehler.
- d. Vektoren in R können mit Vektoren positiver oder negativer Zahlen indiziert werden.

Die richtige Antwort ist:

Indizierung mit einem character Vektor verursacht einen Fehler.

Frage 5

Erreichbare Punkte: 1,00

Welche Aussage bezüglich Datenimport/-export in R ist zutreffend?

- a. Mit der R Funktion **read.table()** können nur Daten von .tsv Dateien eingelesen werden.
- b. Mit der R Funktion **read.csv()** können .csv Dateien importiert werden.
- c. Mit der R Funktion **write.table()** werden Daten einer geöffneten Datei in ein Dataframe geschrieben.
- d. .csv Dateien können nur mit der R Funktion **read.table()** eingelesen werden.

Die richtige Antwort ist:

Mit der R Funktion **read.csv()** können .csv Dateien importiert werden.

Frage 6

Erreichbare Punkte: 1,00

Welcher R Konsolenoutput resultiert aus folgendem R Code?

```
l = list(a = 3,  
        b = TRUE,  
        c = sqrt)  
print(typeof(l[[2]]))
```

- a. `> [1] "list"`
- b. `> [1] TRUE`
- c. `> [1] "logical"`
- d. Keine der Aussagen ist zutreffend.

Die richtige Antwort ist:

```
> [1] "logical"
```

Frage 7

Erreichbare Punkte: 1,00

Welche Aussage bezüglich RStudio Projekten trifft **nicht** zu?

- a. Eine Projektdati speichert automatisch alle Konsolenoutputs und zeigt diese bei Wiederöffnung in RStudio an.
- b. Wenn eine neue RStudio Sitzung über die Projektdatei gestartet wird, werden zuvor geöffnete Programmdateien im Editor geöffnet.
- c. Mit einem RStudio Projekt können Projektspezifische Metadaten gespeichert werden.
- d. Bei Öffnung eines Projektes über die Projektdatei wird das Projektverzeichnis zum Working Directory.

Die richtige Antwort ist:

Eine Projektdati speichert automatisch alle Konsolenoutputs und zeigt diese bei Wiederöffnung in RStudio an.

Frage 8

Erreichbare Punkte: 1,00

Was wird mit nachstehendem Befehl erzeugt?

```
x = c(2,5,1,2,3,4,5,12,3,4,5,1,2,3,42)
H = as.data.frame(table(x))
```

- a. Keine der Aussagen ist zutreffend.
- b. Eine Tabelle mit kumulativen absoluten Häufigkeiten der Zahlenwerte von x.
- c. Ein Dataframe mit zwei Spalten, in welchen die in x vorkommenden Zahlenwerte beziehungsweise deren absoluten Häufigkeiten stehen.
- d. Ein Dataframe mit zwei Spalten, in welchen die absoluten beziehungsweise relativen Häufigkeiten der Zahlenwerte von x stehen.

Die richtige Antwort ist:

Ein Dataframe mit zwei Spalten, in welchen die in x vorkommenden Zahlenwerte beziehungsweise deren absoluten Häufigkeiten stehen.

Frage 9

Erreichbare Punkte: 1,00

Welche(s) Maß(e) der deskriptiven Statistik können in der Visualisierung einer empirischen Verteilungsfunktion abgelesen (visuell bestimmt) werden?

- a. Der Median und Quartile.
- b. Die Standardabweichung, falls die Daten standardnormalverteilt sind.
- c. Die relativen Häufigkeiten aller Zahlenwerte der Menge der reellen Zahlen.
- d. Das arithmetische Mittel.

Die richtige Antwort ist:

Der Median und Quartile.

Frage 10

Erreichbare Punkte: 1,00

In welchem Fall werden in R Vektorelemente recycled?

- a. Bei Durchführung binärer arithmetischer Operationen mit Vektoren unterschiedlicher Längen.
- b. Bei Konkatenation von Vektoren unterschiedlichen Datentyps (Coercion).
- c. Bei Kopie einzelner Vektorelemente mithilfe positiver oder negativer Indizes.
- d. Bei Durchführung arithmetischer Operationen mit NA Werten.

Die richtige Antwort ist:

Bei Durchführung binärer arithmetischer Operationen mit Vektoren unterschiedlicher Längen.

Frage 11

Erreichbare Punkte: 1,00

Mit welchem R Befehl kann immer das aktuelle Working Directory ausgegeben werden?

- a. `basename()` oder `dirname()`
- b. `getwd()`
- c. `set.wd()`
- d. `file.path(work_dir)`

Die richtige Antwort ist:

`getwd()`

Frage 12

Erreichbare Punkte: 1,00

Welcher R Konsolenoutput resultiert aus folgendem R Code?

```
D = data.frame(a = c(4,5,6), b = c(T,T,F))
print(D$a[D$b])
```

- a. `[1] 4 5 6`
- b. `[1] TRUE TRUE`
- c. `Error in print(D$a[b]) : object 'b' not found`
- d. `[1] 4 5`

Die richtige Antwort ist:

`[1] 4 5`

Frage **13**

Erreichbare Punkte: 1,00

Welche Aussage bezüglich Datenformaten trifft zu?

- a. .csv Dateien können mit dem Befehl **view()** in R geöffnet und inspiziert werden.
- b. .pdf und .xlsx, .jpg, .mp4 sind binäre Dateiformate.
- c. Im .json long Format stehen alle Variablen einer Einheit in einer geschweiften Klammer.
- d. Textuelle Dateiformate lassen sich nur mit proprietärer Software bearbeiten.

Die richtige Antwort ist:

.pdf und .xlsx, .jpg, .mp4 sind binäre Dateiformate.

Frage **14**

Erreichbare Punkte: 1,00

Mit welchem R Befehl können die Namen der Spalten eines Dataframes ausgegeben werden?

- a. `columnnames()`
- b. `rownames()`
- c. Mit keinem der genannten Befehle.
- d. `names()`

Die richtige Antwort ist:

`names()`

Frage **15**

Erreichbare Punkte: 1,00

Welcher R Code resultiert **nicht** in folgendem R Konsolenoutput?

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

- a. `seq(1, 20, by=20)`
- b. `1:20`
- c. `seq(1, 20, by=1)`
- d. `seq(1, 20, len=20)`

Die richtige Antwort ist:

```
seq(1, 20, by=20)
```


Fragen und Lösungen – Teil 2

Frage 1

Erreichbare Punkte: 1,00

Welche Aussage bezüglich Boxplots ist zutreffend?

- a. Boxplots können mit dem R Befehl `boxplot()` erstellt werden.
- b. In einem Boxplot werden typischerweise Mittelwert, Varianz und Standardabweichung abgebildet.
- c. Boxplots eignen sich zur Visualisierung der empirischen Stichprobenvarianz.
- d. In einem Boxplot werden immer Extremwerte (Minimum und Maximum) abgebildet.

Die richtige Antwort ist:

Boxplots können mit dem R Befehl `boxplot()` erstellt werden.

Frage 2

Erreichbare Punkte: 1,00

Welche Aussage bezüglich Maßen der zentralen Tendenz ist zutreffend?

- a. Bei einer bimodalen Häufigkeitsverteilung kann es auch zwei Modalwerte geben.
- b. Der Median ist identisch mit dem 0.5-Quantil.
- c. Verglichen mit dem Mittelwert, ist der Median anfälliger für Ausreißer.
- d. Die Summe der Abweichungen aller Datenpunkte vom Mittelwert entspricht der Standardabweichung.

Die richtige Antwort ist:

Der Median ist identisch mit dem 0.5-Quantil.



Frage **3**

Erreichbare Punkte: 1,00

Welches der folgenden Kenngrößen eignet sich **nicht** als Maß der Variabilität der Daten?

- a. Stichprobengröße minus 1.
- b. Die Summe der quadrierten Abweichungen aller Datenpunkte vom Mittelwert geteilt durch die Gesamtanzahl an Datenpunkten.
- c. Standardabweichung.
- d. Spannweite.

Die richtige Antwort ist:
Stichprobengröße minus 1.

Frage 4

Erreichbare Punkte: 1,00

Data_frame sei ein R dataframe mit folgendem Inhalt:

```
> Data_frame
  Bedingung Pre.BDI Post.BDI
1 Klassisch    17     9
2 Klassisch    20    14
3 Klassisch    16    13
4 Klassisch    18    12
5 Klassisch    21    12
6 Klassisch    17    14
7 Klassisch    17    12
8 Klassisch    17     9
9 Klassisch    18    11
10 Klassisch   18    14
11 Online     22    16
12 Online     19    15
13 Online     21    13
14 Online     18    15
15 Online     19    13
16 Online     17    16
17 Online     20    13
18 Online     19    16
19 Online     19    13
20 Online     19    14
```

Was passiert bei Ausführen des folgenden R Befehls?

```
> Data_frame$Delta.BDI = -(Data_frame$Post.BDI - Data_frame$Pre.BDI)
```

- a. Der Befehl erzeugt eine Fehlermeldung, weil die Spalte `Data_frame\$Delta.BDI` nicht existiert.
- b. Der Befehl erzeugt eine Fehlermeldung weil die arithmetische Operation in negativen Werten resultiert.
- c. Dem Dataframe wird eine neue Spalte hinzugefügt, in dem die Differenzen zwischen den Werten in `Data_frame\$Post.BDI` und `Data_frame\$Pre.BDI` stehen.
- d. Dem Dataframe werden zwei neue Spalten hinzugefügt, in denen die positiven und negativen Differenzen, respektive, zwischen den Werten in `Data_frame\$Post.BDI` und `Data_frame\$Pre.BDI` stehen.

Die richtige Antwort ist:

Dem Dataframe wird eine neue Spalte hinzugefügt, in dem die Differenzen zwischen den Werten in `Data_frame\$Post.BDI` und `Data_frame\$Pre.BDI` stehen.

Frage 5

Nicht beantwortet

Erreichbare Punkte: 1,00

Gegeben sei das R Dataframe

```
> Data_frame
  Bedingung Pre.BDI Post.BDI
1  Klassisch    17     9
2  Klassisch    20    14
3  Klassisch    16    13
4  Klassisch    18    12
5  Klassisch    21    12
6  Klassisch    17    14
7  Klassisch    17    12
8  Klassisch    17     9
9  Klassisch    18    11
10 Klassisch    18    14
11  Online     22    16
12  Online     19    15
13  Online     21    13
14  Online     18    15
15  Online     19    13
16  Online     17    16
17  Online     20    13
18  Online     19    16
19  Online     19    13
20  Online     19    14
```

Welche Aussage bezüglich der Auswertung deskriptiver Statistiken mit R ist **nicht** zutreffend?

- a. Der R Befehl `length(Data_frame$Bedingung[Data_frame$Bedingung == "Klassisch"])` ergibt eine gruppenspezifische Stichprobengröße.
- b. Der R Befehl `mean(Data_frame$Bedingung[Data_frame$Bedingung == "Klassisch"])` ergibt ein gruppenspezifisches Stichprobenmittel.
- c. Mit dem R Befehl `var(Data_frame$Post.BDI)` wird die Stichprobenvarianz für die Werte der Spalte `Data_frame$Post.BDI` bestimmt.
- d. Mit dem R Befehl `sd(Data_frame$Pre.BDI)` wird die Stichprobenstandardabweichung für die Werte der Spalte `Data_frame$Pre.BDI` berechnet.

Die richtige Antwort ist:

Der R Befehl `mean(Data_frame$Bedingung[Data_frame$Bedingung == "Klassisch"])` ergibt ein gruppenspezifisches Stichprobenmittel.

Frage 6

Erreichbare Punkte: 1,00

data sei ein Datensatz, der in R als Datenvektor vom Typ `double` vorliegt. Was wird mit folgendem R Code **nicht** ermittelt?

```
n      = length(data)
x_bar  = mean(data)
delta  = 0.95
psi_inv = qt((1+delta)/2,n-1)
sigsqu = sd(data)
G_u    = x_bar - (sigsqu/sqrt(n))*psi_inv
mu_hat = x_bar
G_o    = x_bar + (sigsqu/sqrt(n))*psi_inv
```

- a. Die Stichprobenvarianz.
- b. Ein 0.95-Konfidenzintervall.
- c. Das Stichprobenmittel.
- d. Ein Parameterpunktschätzer für den Erwartungswert im Normalverteilungsmodell.

Die richtige Antwort ist:
Die Stichprobenvarianz.

Frage 7

Erreichbare Punkte: 1,00

Welches der nachfolgenden R Befehle eignet sich, um Parameterschätzwerte nach der Maximum-Likelihood Schätzmethode zu bestimmen?

- a. `sd()`
- b. `var()`
- c. `median()`
- d. `mean()`

Die richtige Antwort ist:
`mean()`

Frage 8

Erreichbare Punkte: 1,00

Welche der nachfolgenden R Funktionen erzeugt **keine** Abbildung?

- a. `par()`
- b. `barplot()`
- c. `hist()`
- d. `boxplot()`

Die richtige Antwort ist:

`par()`

Frage 9

Erreichbare Punkte: 1,00

data sei Datensatz, der in R als Datenvektor vom Typ double vorliegt. Welche Kenngröße(n) wird mit folgendem R Code ermittelt?

```
n      = length(data)
delta  = 0.95
psi_inv = qt((1+delta)/2, n-1)
x_bar  = mean(data)
S      = sd(data)
G_u    = x_bar - (S/sqrt(n))*psi_inv
mu_hat = x_bar
G_o    = x_bar + (S/sqrt(n))*psi_inv
```

- a. Gruppenspezifische Mittelwerte.
- b. Der wahre, aber unbekannte Erwartungswertparameter.
- c. Das 0.95-Konfidenzintervall für den Erwartungswertparameter.
- d. Das 0.95-Konfidenzintervall für den Varianzparameter.

Die richtige Antwort ist:

Das 0.95-Konfidenzintervall für den Erwartungswertparameter.

Frage **10**

Erreichbare Punkte: 1,00

Gegeben sei der Datensatz

$x = c(1, 2, 2, 3, 12)$

Welche Aussage ist zutreffend?

- a. Das Stichprobenmittel beträgt 2.
- b. Der Modalwert ist 12.
- c. Die Spannweite beträgt 11.
- d. Der Median ist größer als das Stichprobenmittel.

Die richtige Antwort ist:

Die Spannweite beträgt 11.