

Ergebnisbericht

Der Leistungsnachweis für Modul C1 – Programmierung und Deskriptive Statistik im Wintersemester 2023/2024 fand in zwei Teilen als Online Schriftliche Ausarbeitung (Open Book Exam) statt. An Teil 1 am 20.12.2023 nahmen 64 Studierende und an Teil 2 am 25.01.2024 63 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.

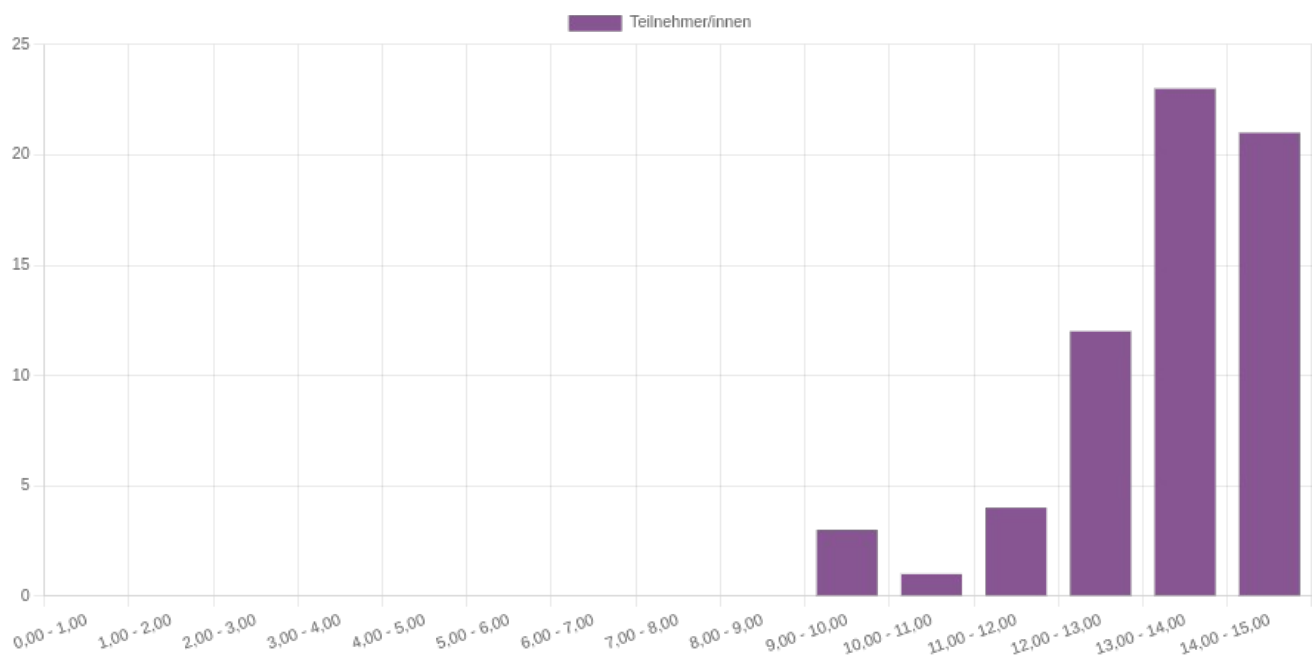
Bewertungsschema

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 *jeweils* mindestens die Hälfte der Punkte erreicht wurde.

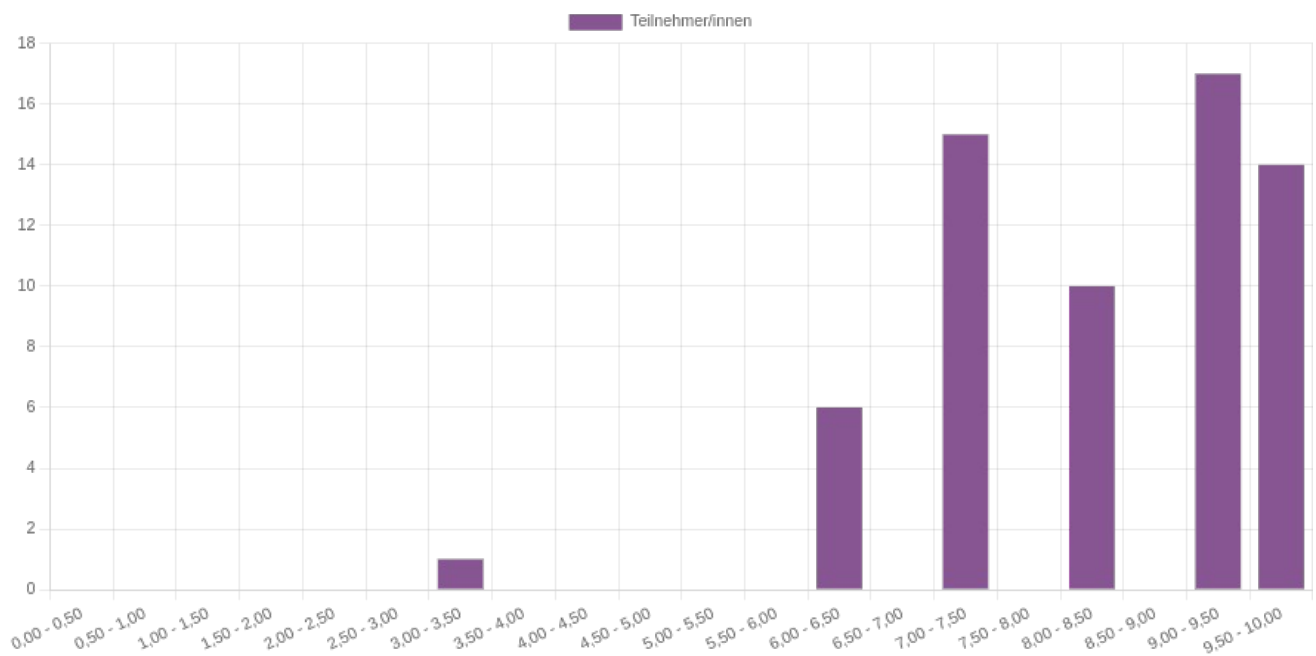
Ergebnisse

Die nachfolgenden Abbildungen zeigen die absolute Häufigkeitsverteilung der erzielten Punkte.

Teil 1 (20.12.2023)



Teil 2 (25.01.2024)



Fragen und Lösungen – Teil 1

Frage 1

Richtig

Erreichte Punkte 1,00 von 1,00

Welche Aussage bezüglich der Klassenzahl k von Histogrammen ist zutreffen?

- a. Bei einer Klassenzahl $k \leq 1$ entspricht das Erscheinungsbild eines Histogramms das einer relativen Häufigkeitsverteilung.
- b. Je höher die Klassenzahl k , desto größer ist die Breite der Rechtecke in einem Histogramm.
- c. Keine der Aussagen ist zutreffend. ✓
- d. Je geringer die Klassenzahl k , desto geringer ist die relative Häufigkeit einzelner Merkmalsausprägungen.

Die richtigen Antworten sind:

Je höher die Klassenzahl k , desto größer ist die Breite der Rechtecke in einem Histogramm. ,

Keine der Aussagen ist zutreffend.



Frage 2

Richtig

Erreichte Punkte 1,00 von 1,00

Gegeben sei der folgende Variablenzuweisungen in R:

```
x <- 3  
y = x + 1  
z <- x
```

Welche logische Operation ergibt **TRUE**?

- a. `y == x`
- b. `z <= x` ✓
- c. `z == 4`
- d. `y <= x`

Die richtige Antwort ist:

```
z <= x
```

Frage 3

Richtig

Erreichte Punkte 1,00 von 1,00

Welche Aussage bezüglich Vektorindizierung in R trifft zu?

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

Die richtige Antwort ist:

Vektoren in R können mit Vektoren positiver oder negativer Zahlen indiziert werden.

Frage 4

Richtig

Erreichte Punkte 1,00 von 1,00

Welcher R Konsolenoutput resultiert aus folgendem R Code?

```
D <- data.frame(a = c(1, 2, 3), b = c(FALSE, FALSE, TRUE), c = c(7, 8, 9))
print(D$c[D$a])
```

- a. [1] NA NA NA
- b. [1] 7 8 9 ✓
- c. Error in D\$c[D\$a] : invalid subscript type 'list'
- d. [1] 9

Die richtigen Antworten sind:

[1] 7 8 9,

Frage 5

Richtig

Erreichte Punkte 1,00 von 1,00

Welchen Output erzeugt der R-Befehl `summary()`?

- a. Zwei Quartile, das Minimum, das Maximum, den Median und den Mittelwert eines Datensatzes.
- b. Alle Aussagen sind zutreffend. ✓
- c. Das Minimum, das 0.25-Quantil, das 0.5-Quantil, das 0.75-Quantil, den Mittelwert und das Maximum eines Datensatzes.
- d. Drei Quartile, das Minimum, das Maximum und den Mittelwert eines Datensatzes.

Die richtigen Antworten sind: Drei Quartile, das Minimum, das Maximum und den Mittelwert eines Datensatzes.,
Alle Aussagen sind zutreffend.

Frage 5 ist uneindeutig und wurde rückwirkend gestrichen

Frage 6

Richtig

Erreichte Punkte 1,00 von 1,00

Welche Aussage bezüglich der R Datenstruktur "List" ist zutreffend?

- a. Der Datentyp von Listen ist `list`. ✓
- b. Listen sind ein wesentlicher Baustein atomarer Vektoren.
- c. Listenelementinhalte können immer mit einfachen eckigen Klammern [] indiziert werden.
- d. Listen sind atomar, können also Objekte verschiedenen Datentyps enthalten.

Die richtige Antwort ist:

Der Datentyp von Listen ist `list`.

Frage 7

Richtig

Erreichte Punkte 1,00 von 1,00

In welchem Fall werden in R Vektorelemente recycled?

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

Die richtige Antwort ist:

Bei Durchführung binärer arithmetischer Operationen mit Vektoren unterschiedlicher Längen.

Frage 8

Richtig

Erreichte Punkte 1,00 von 1,00

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

- a. Mit der R Funktion View() können .csv Dateien im Data Viewer in Visual Studio Code geöffnet werden.
- b. Mit der R Funktion read.table() können .csv Dateien ohne Kopfzeile importiert werden. ✓
- c. Mit der R Funktion write.table() werden Daten einer geöffneten Datei in ein Dataframe geschrieben.
- d. Mit der R Funktion write.table() werden Daten eines R Objekts in ein Tabelle im Arbeitsspeicher geschrieben.

Die richtige Antwort ist:

Mit der R Funktion read.table() können .csv Dateien ohne Kopfzeile importiert werden.

Frage 9

Richtig

Erreichte Punkte 1,00 von 1,00

Welche Aussage bezüglich der Gemeinsamkeiten und Unterschiede von Mittelwert und Median ist zutreffend?

- a. Bei Normalverteilung der Daten sind Median und Mittelwert identisch.
- b. Ausreißer im Datensatz führen zu höheren Mittelwerten und geringeren Medianwerten.
- c. Bei bimodaler Verteilung der Daten sind beide Maße größer als der Modalwert.
- d. Ausreißer im Datensatz haben meist einen größeren Effekt auf den Mittelwert, als auf den Median. ✓

Die richtige Antwort ist:

Ausreißer im Datensatz haben meist einen größeren Effekt auf den Mittelwert, als auf den Median.

Frage 10

Richtig

Erreichte Punkte 1,00 von 1,00

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

- a. Der Median und das arithmetische Mittel, wenn eine Normalverteilung vorliegt.
- b. Die relativen Häufigkeiten der in einem Datensatz vorkommenden Zahlenwerte, sofern sich diese in der Menge der reellen Zahlen befinden.
- c. Quantile. ✓
- d. Die Standardabweichung, falls die Daten standardnormalverteilt sind.

Die richtige Antwort ist: Quantile.

Frage 11

Richtig

Erreichte Punkte 1,00 von 1,00

Mit welchem R-Befehl wird **nicht** der Mittelwert eines Datensatzes x mit 15 Datenpunkten berechnet?

- a.

```
y <- 3 * x + 5
y_bar <- mean(y)
x_bar <- (y_bar - 5) * (1 / 3)
```
- b.

```
y_bar <- (1 / 15) * sum(x)
```
- c.

```
y <- x + 3
z <- x + y
z_bar <- mean(z)
z_bar <- x_bar + y_bar
```

 ✓
- d.

```
x_bar = (1 / length(x)) * sum(x)
```

Die richtige Antwort ist:

```
y <- x + 3
z <- x + y
z_bar <- mean(z)
z_bar <- x_bar + y_bar
```

Frage 12

Richtig

Erreichte Punkte 1,00 von 1,00

Mit welchem R Befehl wird die 3. Zeile der im Folgenden definierten Matrix modifiziert?

```
A <- matrix(c(4, 3, 2, 1,
             3, 2, 1, 4,
             2, 1, 4, 3),
           nrow = 3,
           byrow = TRUE)
```

- a. `a_2 <- c(1, 2, 3, 4)`
- b. `a_.2 <- matrix(1, 2, 3, 4)`
- c. `A[3,] <- c(4, 1, 2, 3)` ✓
- d. `A[, 3] <- c(4, 1, 2, 3)`

Die richtige Antwort ist:

```
A[3, ] <- c(4, 1, 2, 3)
```

Frage 13

Richtig

Erreichte Punkte 1,00 von 1,00

Was ist **kein** typischer Bestandteil eines Datenanalyseskripts?

- a. Befehle zum Einlesen und Bereinigen eines digitalen Datensatzes.
- b. Kompilierung eines Quellcodes, falls dieser als R-Skript vorliegt. ✓
- c. Befehle zur Berechnung deskriptiver Statistiken.
- d. Dokumentation der Datenvisualisierung.

Die richtige Antwort ist:

Kompilierung eines Quellcodes, falls dieser als R-Skript vorliegt.

Frage **14**

Richtig

Erreichte Punkte 1,00 von 1,00

Was haben binäre und textuelle Dateiformate gemeinsam?

- a. Keine der Aussagen ist zutreffend. ✓
- b. Beide sind generell offene Dateiformate und lassen sich mit einfachen allgemeinen Texteditoren öffnen und bearbeiten.
- c. Beide Formate sind oft proprietär.
- d. Die Metadaten ist in beiden Formaten automatisch in der Kopfzeile enthalten.

Die richtige Antwort ist:

Keine der Aussagen ist zutreffend.

Frage **15**

Richtig

Erreichte Punkte 1,00 von 1,00

Welcher R Befehl erzeugt **nicht** die gleiche Sequenz wie folgender Befehl mit dem Colonoperator : ?

`1:10`

- a. `seq(1, 11)` ✓
- b. `seq(1, 10, len = 10)`
- c. `seq(1, 10)`
- d. `seq(1, 10, by = 1)`

Die richtige Antwort ist:

`seq(1, 11)`

Fragen und Lösungen – Teil 2

Frage 1

Vollständig

Erreichte Punkte 1,00 von 1,00

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

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

Frage 2

Vollständig

Erreichte Punkte 1,00 von 1,00

Gegeben sei der Datensatz

```
> x <- c(-3, 1, 2, 3, 4, 4, 4, 5, 5, 5, 5, 5, 5, 12)
```

Welche Aussage ist zutreffend?

- a. Das Maximum der relativen Häufigkeiten beträgt 6.
- b. Der Modalwert beträgt 5.
- c. Der Modalwert beträgt 6.
- d. Das Maximum der absoluten Häufigkeiten beträgt 5.

Frage 3

Vollständig

Erreichte Punkte 1,00 von 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
```

Welche Aussage bezüglich des folgenden R Befehls ist zutreffend?

```
boxplot(
  D$Delta.BDI ~ D$Bedingung,
  ylim = c(0, 12),
  col = "gray90",
  ylab = "",
  xlab = "Bedingung",
  main = TeX("$\\Delta$ BDI$")
)
```

- a. Wenn die Spalte `Data_frame\$Delta.BDI` existieren würde, dann würde dieser Befehl zwei Boxplots aus Werten dieser Spalte erstellen.
- b. Keine der Aussagen ist zutreffend.
- 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, und das entsprechende Boxplot erstellt.
- d. Der Befehl erzeugt eine Fehlermeldung, weil tilde (~) keine zulässige arithmetische Operation ist.

Frage 4

Vollständig

Erreichte Punkte 1,00 von 1,00

Welche ist **keine** Modellannahme für Parameterschätzung und Konfidenzintervalle im Rahmen der Frequentistischen Inferenz, wenn Erwartungswert- und Varianzparameter für ein Datensatz $x := (x_1, \dots, x_n)$ mit $x_i \in \mathbb{R}$ geschätzt werden?

- a. Es wird das Normalverteilungsmodell zugrunde gelegt.
- b. Es wird angenommen, dass der Erwartungswertparameter der Fehlervariable 0 beträgt.
- c. Es wird angenommen, dass die Datenwerte Realisierungen von unabhängig verteilten Zufallsvariablen sind.
- d. Es wird angenommen, dass bei unabhängig und identisch verteilten Zufallsvariablen, auch alle Realisierungen dieser Variable identisch sein müssten.

Frage 5

Vollständig

Erreichte Punkte 1,00 von 1,00

Welche Aussage bezüglich eines Zweistichproben-T-Tests kann **nicht** zutreffen?

- a. Der Zweistichproben-T-Test vergleicht eine T-Teststatistik mit einem kritischen Wert.
- b. Die Nullhypothese besagt, dass die Differenz der Erwartungswertparameter zweier Stichproben gleich 0 ist.
- c. Die Alternativhypothese besagt, dass die Erwartungswertparameter zweier Stichproben ungleich sind.
- d. In die Berechnung der T-Teststatistik fließen keine Werte eines Datensatzes (Realisierungen einer Zufallsvariable) ein.

Frage 6

Vollständig

Erreichte Punkte 1,00 von 1,00

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

```
n      = length(data)
delta  = 0.90
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. Das 0.90-Konfidenzintervall für den Varianzparameter.
- b. Ein Punktschätzer für den Erwartungswertparameter, der mit einer Wahrscheinlichkeit von 0.95 von dem Konfidenz-Intervall $[G_u, G_o]$ abgedeckt ist.
- c. Die Stichprobenstandardabweichung des vorliegenden Datensatzes.
- d. Das 0.95-Konfidenzintervall für den Erwartungswertparameter.

Frage 7

Vollständig

Erreichte Punkte 1,00 von 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 zutreffend?

- a. Der R Befehl `sum(Data_frame$Bedingung[Data_frame$Bedingung == "Klassisch"])` ergibt eine gruppenspezifische Stichprobengröße.
- b. Keine der Aussagen ist zutreffend.
- c. Mit dem R Befehl `var(Data_frame$Post.BDI)` wird die empirische Stichprobenvarianz für die Werte der Spalte `Data_frame$Post.BDI` bestimmt.
- d. Mit dem R Befehl `sqrt(variance(Data_frame$Pre.BDI))` wird die Stichprobenstandardabweichung für die Werte der Spalte `Data_frame$Pre.BDI` berechnet.

Frage 8

Vollständig

Erreichte Punkte 1,00 von 1,00

Welche Aussage bezüglich Boxplots ist zutreffend?

- a. In einem Boxplot werden immer Extremwerte (Minimum und Maximum) abgebildet.
- b. In einem Boxplot werden typischerweise Mittelwert, Varianz und Standardabweichung abgebildet.
- c. In einem Boxplot werden nie Quantile abgebildet.
- d. Boxplots können mit dem R Befehl `boxplot()` erstellt werden.

Frage 9

Vollständig

Erreichte Punkte 1,00 von 1,00

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

- a. Spannweite.
- b. Standardabweichung.
- c. Die Summe aller Abweichungen vom Mittelwert.
- d. Die Wurzel aus der Varianz.

Frage 10

Vollständig

Erreichte Punkte 1,00 von 1,00

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

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