**Hausübung Einführung in die Statistische Software – WS2016/17**

Nachname: Oberhauser

Vorname: Cornelia

Matrikelnummer: 11111111111

Email-Adresse: [Cornelia.Oberhauser@stat.uni-muenchen.de](mailto:Cornelia.Oberhauser@stat.uni-muenchen.de)

Studieren Sie Bachelor Statistik? - Ja

Möchten Sie eine ausführliche Rückmeldung zu Ihrer Abgabe? - Nein

Aufgabe 0)

**R-Syntax, mit der die Daten erzeugt und gespeichert wurden:**

# zu Beginn der Analyse immer Arbeitsverzeichnis (working directory) festlegen, z.B.

setwd("Z:\\Einf\_Software\_WS2016\\Hausübung")

#------------------------------------------------------------

# Vorbereitung

# Daten einlesen

student\_performance <- read.table(file="student\_performance.csv", sep=";", header=TRUE)

# Stichprobe mit Matrikelnummer als Startwert (seed) ziehen

matrikelnummer <- 1111111111 # hier bitte eigene Matrikelnummer hinschreiben

# Startwert festlegen

set.seed(matrikelnummer)

# Zufallsstichprobe vom Umfang 300 ohne Zurücklegen aus den Daten ziehen

lines\_sample <- sample(x=1:nrow(student\_performance), size=300, replace = FALSE, prob = NULL)

student\_performance\_sample <- student\_performance[lines\_sample,]

# Datensatz im .dat-Format speichern

# mit verschiedenen Trennzeichen für Spalten (sep) und Dezimaltrennzeichen (dec)

write.table(x=student\_performance\_sample,

file="student\_performance\_sample\_semicolon\_dot.dat", sep = ";", dec = ".",

row.names = FALSE, col.names = TRUE)

write.table(x=student\_performance\_sample,

file="student\_performance\_sample\_semicolon\_comma.dat", sep = ";", dec = ",",

row.names = FALSE, col.names = TRUE)

write.table(x=student\_performance\_sample,

file="student\_performance\_sample\_tab\_dot.dat", sep = "\t", dec = ".",

row.names = FALSE, col.names = TRUE)

write.table(x=student\_performance\_sample,

file="student\_performance\_sample\_tab\_comma.dat", sep = "\t", dec = ",",

row.names = FALSE, col.names = TRUE)

# diese Datensätze dann in R, SPSS und SAS einlesen und

# für jedes Softwarepaket einen Datensatz, der fehlerfrei eingelesen wurde,

# für die weiteren Analysen verwenden**Aufgabe 1) Syntax zum Daten einlesen** (Die Syntax zum Einlesen jeweils eines Datensatzes pro Softwarepaket ist ausreichend. Der für das jeweilige Softwarepaket ausgewählte Datensatz sollte fehlerfrei in das Programm eingelesen werden.)

**R:**

**SPSS:**

**SAS:**

Beispiel:

**Aufgabe 2) Mittelwert von Alter**

**R:**

mean(data$age)

[1] 16.65667

**SPSS:**

DESCRIPTIVES VARIABLES=age

/STATISTICS=MEAN.

|  |  |  |
| --- | --- | --- |
| **Deskriptive Statistik** | | |
|  | N | Mittelwert |
| age | 300 | 16,66 |
| Gültige Werte (Listenweise) | 300 |  |

**SAS:**

**PROC** **MEANS** DATA=sample mean;

VAR age;

**RUN**;

|  |
| --- |
| Das SAS System |

Die Prozedur MEANS

| **Analysevariable : age** |
| --- |
| **Mittelwert** |
| 16.6566667 |

**Interpretation:**

Der Mittelwert von Alter beträgt 16.66.Beispiel:

**Aufgabe 3) Boxplots zur Beschreibung des Zusammenhangs von Geschlecht und Alter**

**R:**

boxplot(data$age~data$sex)



**SPSS:**

\* Diagrammerstellung.

GGRAPH

/GRAPHDATASET NAME="graphdataset" VARIABLES=sex age MISSING=LISTWISE REPORTMISSING=NO

/GRAPHSPEC SOURCE=INLINE.

BEGIN GPL

SOURCE: s=userSource(id("graphdataset"))

DATA: sex=col(source(s), name("sex"), unit.category())

DATA: age=col(source(s), name("age"))

DATA: id=col(source(s), name("$CASENUM"), unit.category())

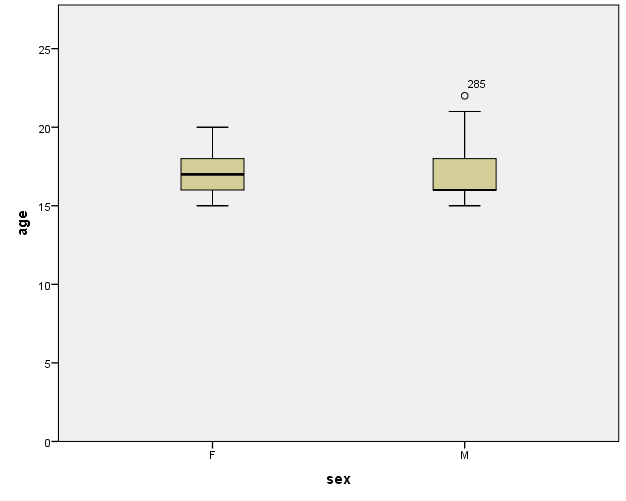
GUIDE: axis(dim(1), label("sex"))

GUIDE: axis(dim(2), label("age"))

SCALE: linear(dim(2), include(0))

ELEMENT: schema(position(bin.quantile.letter(sex\*age)), label(id))

END GPL.

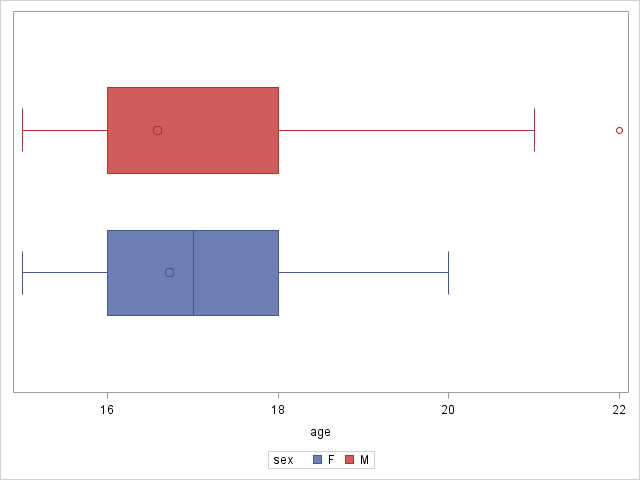


**SAS:**

**PROC** **SGPLOT** DATA=sample;

HBOX age / GROUP=sex;

**RUN**;



**Interpretation:**

Der Median bei den Männern ist niedriger als bei den Frauen. Bei den Männern gibt es extremere Werte nach oben.