**Hausübung Einführung in die Statistische Software – WS2015/16**

Nachname: Oberhauser

Vorname: Cornelia

Matrikelnummer: 11111111111

Email-Adresse: Cornelia.Oberhauser@stat.uni-muenchen.de

Aufgabe 0)

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

# zu Beginn der Analyse immer Working Directory festlegen

setwd("Z:\\Einf\_Software WS2015\\Hausaufgabe")

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

# Vorbereitung

# Daten einlesen

miete\_all <- read.table(file="miete\_2015.txt")

# Stichprobe mit Matrikelnummer als Startwert (seed) ziehen

matrikelnummer <- 1111111111 # hier bitte eigene Matrikelnummer hinschreiben

# Startwert festlegen

set.seed(matrikelnummer)

# Zufallsstichprobe vom Umfang 2500 ohne Zurücklegen aus den Daten ziehen

lines\_sample <- sample(x=1:nrow(miete\_all), size=2500, replace = FALSE, prob = NULL)

miete <- miete\_all[lines\_sample,]

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

# Datensatz im .dat-Format speichern

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

# Strichpunkt als Spaltentrennzeichen, Punkt als Dezimaltrennzeichen

write.table(x=miete, file = "miete\_sample\_semicolon\_dot.dat", sep = ";", dec = ".",

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

# Strichpunkt als Spaltentrennzeichen, Komma als Dezimaltrennzeichen

write.table(x= miete, file = " miete\_sample\_semicolon\_comma.dat", sep = ";", dec = ",",

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

# tab als Spaltentrennzeichen, Punkt als Dezimaltrennzeichen

write.table(x= miete, file = " miete\_sample\_tab\_dot.dat", sep = "\t", dec = ".",

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

# tab als Spaltentrennzeichen, Komma als Dezimaltrennzeichen

write.table(x= miete, file = " miete\_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 Nettomiete pro Quadratmeter**

**R:**

mean(data$Nettomiete\_pro\_Quadratmeter)

[1] 10.73427

**SPSS:**

DATASET ACTIVATE DataSet1.

DESCRIPTIVES VARIABLES=Nettomiete\_pro\_Quadratmeter

/STATISTICS=MEAN.

|  |  |  |
| --- | --- | --- |
| **Deskriptive Statistik** | | |
|  | N | Mittelwert |
| Nettomiete\_pro\_Quadratmeter | 2500 | 10,7343 |
| Gültige Werte (Listenweise) | 2500 |  |

**SAS:**

**PROC** **MEANS** DATA=data MEAN;

VAR Nettomiete\_pro\_Quadratmeter;

**RUN**;

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

Die Prozedur MEANS

| **Analysevariable : Nettomiete\_pro\_Quadratmeter** |
| --- |
| **Mittelwert** |
| 10.7342720 |

**Interpretation:**

Der Mittelwert von Nettomiete pro Quadratmeter ist 10.73 Euro (in der Stichprobe mit 2500 Wohnungen, gezogen mit dem seed 1111111111).Beispiel:

**Aufgabe 3) Streudiagramm zwischen Nettomiete und Wohnflaeche**

**R:**

plot(data$Wohnflaeche, data$Nettomiete)



**SPSS:**

\* Diagrammerstellung.

GGRAPH

/GRAPHDATASET NAME="graphdataset" VARIABLES=Wohnflaeche Nettomiete MISSING=LISTWISE

REPORTMISSING=NO

/GRAPHSPEC SOURCE=INLINE.

BEGIN GPL

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

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

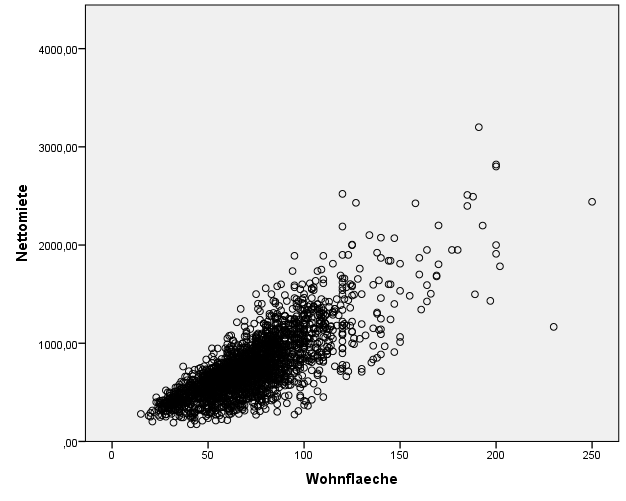
DATA: Nettomiete=col(source(s), name("Nettomiete"))

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

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

ELEMENT: point(position(Wohnflaeche\*Nettomiete))

END GPL.

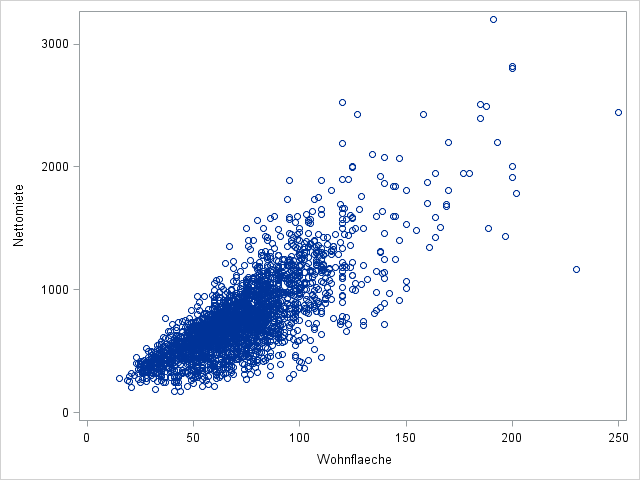


**SAS:**

**PROC** **SGPLOT** DATA=data;

SCATTER y=Nettomiete x=Wohnflaeche;

**RUN**;



**Interpretation:**

Mit steigender Wohnfläche nimmt die Nettomiete zu. Mit steigender Wohnfläche nimmt die Streuung der Nettomiete zu.