

Aufgabe Tag 3 - Aufgabe von Tag 2 in R

Fabian Fiedler

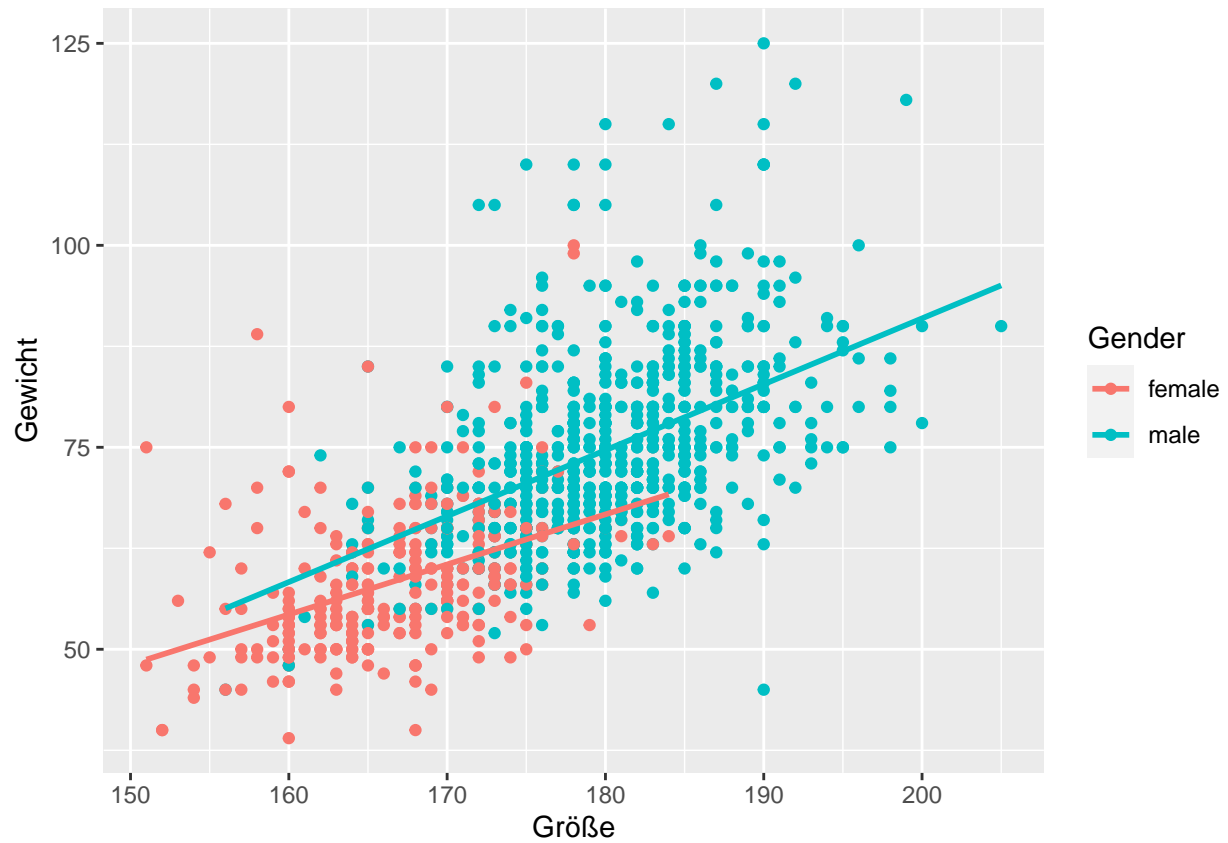
18.01.2021

```
library(tidyverse)
library(magrittr)
library(readxl)
library(knitr)
```

```
studdat <- read_excel(file.path("Daten", "StudierendenDaten",
                                "StudDat.xlsx"))
```

Aufgabe 1: Zusammenhang Körpergröße & Gewicht

```
studdat %>%
  select(Gender, Size, Weight) %>%
  drop_na() %>%
  ggplot(aes(x=Size, y=Weight, color=Gender)) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE) +
  scale_x_continuous("Größe") +
  scale_y_continuous("Gewicht")
```



Wie man anhand der Grafik sehen kann, ist die Steigung der Geraden relativ gleich und man kann somit davon ausgehen, dass der Zusammenhang zwischen Gewicht und Körpergröße für beide Geschlechter gleich ist.

(Alle Daten bei denen entweder die Größe oder das Gewicht gefehlt haben, wurden aus der Analyse ausgenommen.)

Aufgabe 2: Regressionsanalyse - Zusammenhang Körpergröße Eltern & Studierende

```
studdat %>%
  select(Gender, Size, SizeMother, SizeFather)%>%
  drop_na()%>%
  summarize(Gender, Size, SizeParents=(SizeMother+SizeFather)/2) %>%
# filter(Gender=="male") %>%
  lm(Size ~ SizeParents, data=.)
```

Call:

```
lm(formula = Size ~ SizeParents, data = .)
```

Coefficients:

(Intercept)	SizeParents
55.4992	0.7035

Anhand des Regressionskoeffizienten kann man sehen, dass für jeden cm den die Eltern durchschnittlich groß sind, das Kind durchschnittlich ca 0,7 cm groß ist.

(Alle Daten bei denen eine der Größen gefehlt hat, wurden aus der Analyse ausgenommen.)

Aufgabe 3: Zusammenhang Rauchverhalten zwischen Eltern und Studierenden

Vater

```
studdat %>%  
  select(Gender, Smoker, SmokeFather, SmokeMother) %>%  
  drop_na() %>%  
  group_by(Smoker, SmokeFather) %>%  
  tally() -> table_father
```

```
table_father %>%  
  select(SmokeFather, Smoker, n) %>%  
  pivot_wider(names_from = SmokeFather, values_from=n) %>%  
  rename(Smoker_yes=yes, Smoker_no=no) %>%  
  mutate(Smoking=Smoker_yes+Smoker_no) %>%  
  mutate(Smoker_no=Smoker_no/Smoking, Smoker_yes=Smoker_yes/Smoking) %>%  
  select(-Smoking) %>%  
  kable(digits=3)
```

Smoker	Smoker_no	Smoker_yes
no	0.768	0.232
yes	0.698	0.302

Mutter

```
studdat %>%  
  select(Gender, Smoker, SmokeFather, SmokeMother) %>%  
  drop_na() %>%  
  group_by(Smoker, SmokeMother) %>%  
  tally() -> table_mother
```

```
table_mother %>%  
  select(SmokeMother, Smoker, n) %>%  
  pivot_wider(names_from = SmokeMother, values_from=n) %>%  
  rename(Smoker_yes=yes, Smoker_no=no) %>%  
  mutate(Smoking=Smoker_yes+Smoker_no) %>%  
  mutate(Smoker_no=Smoker_no/Smoking, Smoker_yes=Smoker_yes/Smoking) %>%  
  select(-Smoking) %>%  
  kable(digits=3)
```

Smoker	Smoker_no	Smoker_yes
no	0.841	0.159
yes	0.801	0.199

Anhand der Analyse kann man sehen, dass das Rauchverhalten des Vaters mehr Auswirkungen auf das der Studierenden hat, aber generell gesagt kann man keinen definitiven Einfluss des Rauchverhaltens der Eltern auf das der Kinder sehen.

(Alle Daten bei denen eines der Rauchverhalten gefehlt hat, wurden aus der Analyse ausgenommen. Da hier das ignorieren Fehlender Daten für Mutter und Vater leichter ist, werden die Berechnungen zu denen die ich in Excel gemacht habe abweichen, da in Excel nur Daten spezifisch für jede Einzelberechnung ausgenommen wurden und nicht gemeinsam für beide Berechnungen wie hier.(z.B.: Wenn SmokeMother fehlt wurde es in Excel bei der Berechnung der Mutter ausgenommen, aber nicht beim Vater, wenn in SmokeVater ein Wert vorhanden war.))