Exercise 1: Multiple Regression

Henrik Singmann

January 2019 (updated: 2019-01-14)

Violent Crime Rates by US State

The data set `USArrests` contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

Research question: *Which variables predict murder rates and to what degree?*

Design

The data comes with `R`, so we can load it right away. But at first, we load the `tidyverse` package, because these are the functions we want to use throughout this exercise.

```
library("tidyverse")
theme_set(theme_bw(base_size = 17))
data(USArrests)
glimpse(USArrests)
```

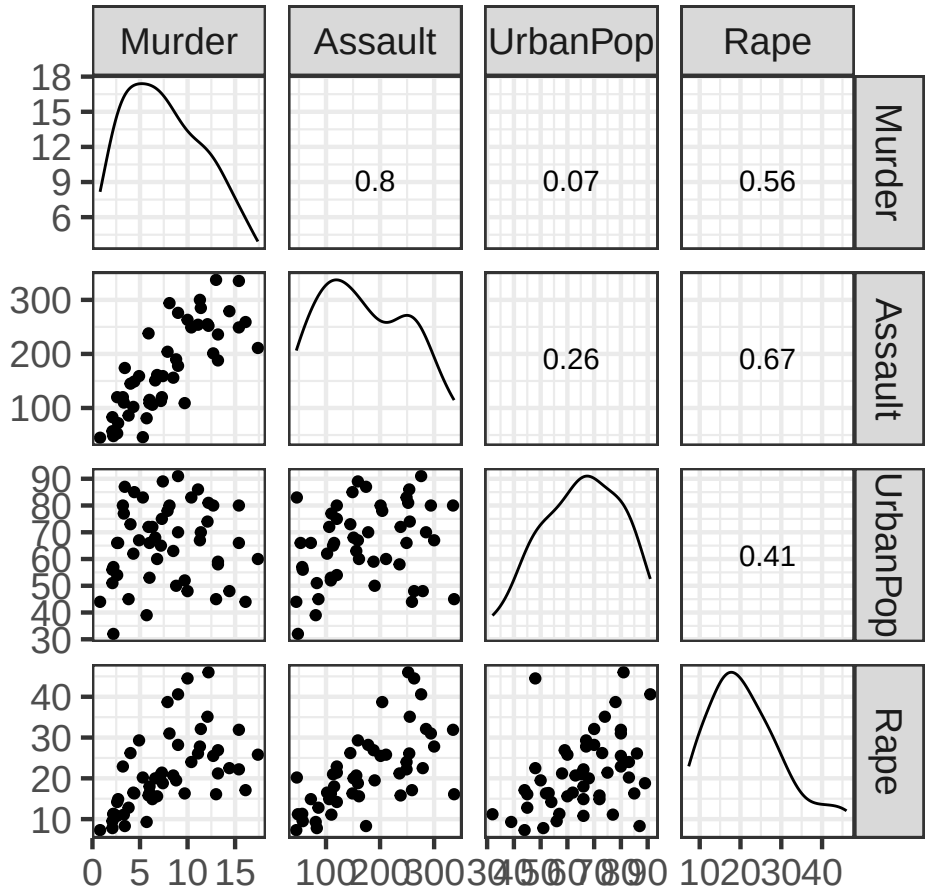
```
## Observations: 50
## Variables: 4
## $ Murder   <dbl> 13.2, 10.0, 8.1, 8.8, 9.0, 7.9, 3.3, 5.9, 15.4, 17.4,...
## $ Assault  <int> 236, 263, 294, 190, 276, 204, 110, 238, 335, 211, 46,...
## $ UrbanPop <int> 58, 48, 80, 50, 91, 78, 77, 72, 80, 60, 83, 54, 83, 6...
## $ Rape     <dbl> 21.2, 44.5, 31.0, 19.5, 40.6, 38.7, 11.1, 15.8, 31.9,...
```

The `data.frame` has 50 observations (one for each state) on 4 variables. The columns in the data are:

- `Murder`: Murder arrests (per 100,000)
- `Assault`: Assault arrests (per 100,000)
- `UrbanPop`: Percent urban population
- `Rape`: Rape arrests (per 100,000)

Several of the variables show correlations among each other.

```
GGally::ggscatmat(USArrests)
```



When looking at the zero-order correlation displayed in the upper-triangle of the pairs plot, we see that both, `Assault` and `Rape` appear to predict murder rates.

Exercise 1: Simple Linear Models

Create one simple linear regression model with murder rate as dependent variable and each of the other variables as independent variable.

```
# simple linear models:
```

Which of the three variables shows a significant effect on murder rates?

Exercise 2: Multiple Regression

Create one overall model with murder rate as dependent variable and the three other variables as independent variables (without interactions).

```
## multiple regression model
```

Which of the three variables now shows a significant effect on murder rates? If there are differences, how can these differences be explained?