

Lösung zu Blatt 5 (RStudio)

Hausaufgabe 13

Im ersten Schritt lesen wir die Textdatei `tips.txt` ein. Dafür verwenden wir den Befehl `read.table()`:

```
tips = read.table("/Users/paulbergold/Hausaufgaben Rstudio/tips.txt")
```

Bemerkung:

Mit `head(DATA, n)` lassen sich die ersten n Datenreihen des Datensatzes `DATA` anzeigen. Im Gegensatz dazu liefert `tail(DATA, n)` die letzten n Datenreihen. Der Standardwert von n ist 6. Zum Beispiel erhalten wir für den Datensatz `tips`:

```
head(tips)
```

```
##   tip
## 1 1.2
## 2 0.4
## 3 1.2
## 4 1.7
## 5 2.0
## 6 0.6
```

Mit dem Befehl `as.matrix()` erzeugen wir nun die Matrix `tips_matrix`, deren Einträge die Werte der Variablen `tip` sind.

```
tips_matrix = as.matrix(tips)
dim(tips_matrix)
```

```
## [1] 255  1
```

Wir sehen, dass die Matrix `tips_matrix` aus 255 Zeilen und einer Spalte besteht. Dabei stehen die 255 Werte der Variablen `tip` gerade in der ersten Spalte dieser Matrix.

Wir erstellen nun den Vektor `tip`, der diese Werte enthält:

```
tip = tips_matrix[,1]
```

Nun berechnen wir das arithmetische Mittel und die Stichprobenvarianz:

```
x_bar = mean(tip)
x_bar
```

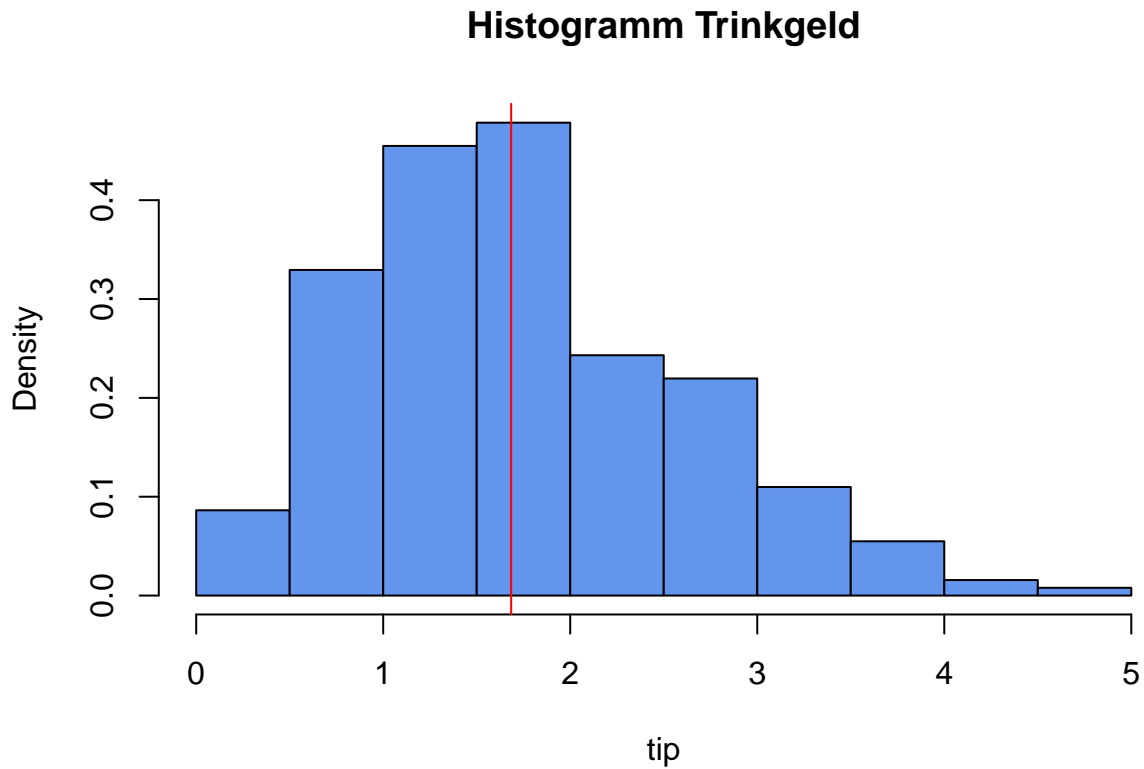
```
## [1] 1.684314
```

```
s_X = sqrt(var(tip))
s_X
```

```
## [1] 0.8676088
```

Im nächsten Schritt erstellen wir das Histogramm:

```
titel = "Histogramm Trinkgeld"  
color = "cornflowerblue"  
right = FALSE #Intervalle rechts offen  
prob = TRUE #Normierung  
hist(tip, main = titel, prob = prob, right = right, col = color)  
abline(v = x_bar, col = "red")
```



Standardisierung:

Im Fall $s_X > 0$ war die Standardisierung des Merkmals X definiert als

$$Z = \frac{X - \bar{x}}{s_X}.$$

Dementsprechend erhalten wir:

```
tip_standard = (tip - x_bar)/s_X
```

Nach Voraussetzung gilt für die Standardisierung Z :

$$\bar{z} = 0 \quad \text{und} \quad s_Z = 1.$$

Für `tip_standard` erhalten wir:

```
mean(tip_standard)
```

```
## [1] -2.967313e-18
```

```
sqrt(var(tip_standard))
```

```
## [1] 1
```

Zuletzt erstellen wir noch das Histogramm für die Standardisierung:

```
titel = "Histogramm Trinkgeld (standardisiert)"  
color = "darkseagreen"  
right = FALSE #Intervalle rechts offen  
prob = TRUE #Normierung  
hist(tip_standard, main = titel, prob = prob, right = right, col = color)
```

