

## Vorbereitung

Bitte führen Sie zur Vorbereitung folgende Schritte aus:

1. Starten Sie RStudio.
2. Löschen Sie den Workspace.
3. Setzen Sie das Arbeitsverzeichnis dort, wo Sie Ihre Daten abgelegt haben: `Session` » `Set Working Directory` » `Choose Directory`.
4. Öffnen Sie ein R-Skript.
5. Nachdem Sie die Aufgaben bearbeitet haben, speichern Sie das Skript unter einem geeigneten Namen ab.

## Aufgabe 1 - Verteilung eines Merkmals

- Untersuchen Sie die Form der Verteilung der Prokrastination (`prok`) anhand von Modalwert, Median und Mittelwert.
- Stellen Sie die Verteilung auch grafisch mithilfe eines Histogramms dar. Zeichnen Sie in diesem Histogramm den Mittelwert ein. Färben Sie die Balken lila. Benennen Sie das Diagramm mit dem Titel "Histogramm von Prokrastination". Benennen Sie die Y-Achse mit "Häufigkeit" und löschen Sie den Titel der X-Achse.

### Lösung

```
summary(erstis$prok)
```

| Min.  | 1st Qu. | Median | Mean  | 3rd Qu. | Max.  | NA's |
|-------|---------|--------|-------|---------|-------|------|
| 1.000 | 2.075   | 2.500  | 2.459 | 2.900   | 3.700 | 15   |

```
library(DescTools)
```

```
Mode(erstis$prok, na.rm = T)
```

```
[1] 2.3
```

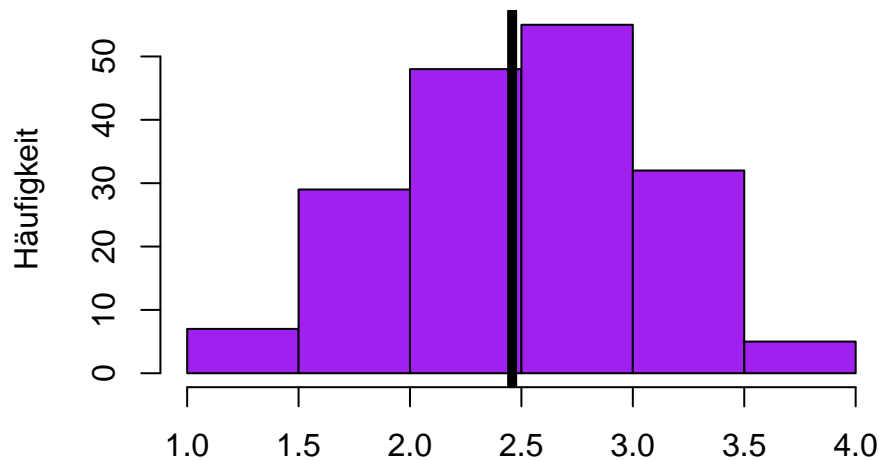
```
attr("freq")
```

```
[1] 15
```

- Verteilung von Prokrastination
  - Mittelwert und Median sind fast identisch, d. h. einigermaßen symmetrisch verteilt
  - der Mittelwert ist etwas geringer als der Median
  - Q1 und Q3 sind auch etwa gleich weit entfernt vom Median -> spricht auch für Symmetrie
  - Der Modalwert, d.h. die häufigste Punktzahl, beträgt 2.3 und ist etwas geringer als der Median und Mittelwert.

```
hist(erstis$prok, right = FALSE, col = "purple",
     main = "Histogramm von Prokrastination",
     ylab = "Häufigkeit",
     xlab = "")
abline(v = mean(erstis$prok, na.rm=T),
       lwd = 5)
```

## Histogramm von Prokrastination



### Aufgabe 2 - z-Transformation

Transformieren Sie die Variable `lz.2` (entsprechend der z-Transformation). Interpretieren Sie den z-transformierten Wert der 11. Person.

#### Lösung

```
lz.2_z <- scale(erstis$lz.2)
lz.2_z[11]
```

```
[1] -3.217855
```

Die 11. Person hat einen z-standardisierten Wert von  $-3.22$  in Lebenszufriedenheit (2. Messzeitpunkt). Ihre Lebenszufriedenheit liegt also über drei Standardabweichungen unter dem Durchschnitt.

### Aufgabe 3 - Streuungsmaße

- (i) Berechnen Sie, mit welchem Alter die Personen ihr Abitur gemacht haben. Berechnen Sie dann die Standardabweichung für dieses Alter.

```
erstis$Alter_bei_Abitur <- erstis$abi - erstis$gebjahr
sd(erstis$Alter_bei_Abitur, na.rm = T)
```

```
[1] 4.527957
```

- (ii) Zentrieren Sie die neu gebildete Variable und bestimmen Sie zur Überprüfung Mittelwert und Standardabweichung, jeweils auf zwei Nachkommastellen gerundet.

```
erstis$Alter_bei_Abitur_c <- scale(erstis$Alter_bei_Abitur, scale = F)
round(mean(erstis$Alter_bei_Abitur_c, na.rm = T), digits = 2)
```

```
[1] 0
```

```
round(sd(erstis$Alter_bei_Abitur_c, na.rm = T), digits = 2)
```

```
[1] 4.53
```

- (iii) Bestimmen Sie, mit welchem Alter die meisten Teilnehmenden ihr Abitur gemacht haben mithilfe eines Befehls.

```
Mode(erstis$Alter_bei_Abitur, na.rm = T)
```

```
[1] 20
```

```
attr("freq")
```

```
[1] 60
```