Programmierung und Deskriptive Statistik

BSc Psychologie WiSe 2024/25

Belinda Fleischmann

Datum	Einheit	Thema	Form
15.10.24	R Grundlagen	(1) Einführung	Seminar
22.10.24	R Grundlagen	(2) R und Visual Studio Code	Seminar
29.10.24	R Grundlagen	(2) R und Visual Studio Code	Übung
05.11.24	R Grundlagen	(3) Vektoren, (4) Matrizen	Seminar
12.11.24	R Grundlagen	(5) Listen und Dataframes	Seminar
	Leistungsnachweis 1		
19.11.24	R Grundlagen	(6) Datenmanagement	Seminar
26.11.24	R Grundlagen	(2)-(6) R Grundlagen	Übung
03.12.24	Deskriptive Statistik	(7) Häufigkeitsverteilungen	Seminar
10.12.24	Deskriptive Statistik	(8) Verteilungsfunktionen und Quantile	Seminar
	Leistungsnachweis 2		
17.12.24	**Deskriptive Statistik**	(9) **Maße der zentralen Tendenz und Datenvariabilität**	**Seminar**
	Weihnachtspause		
07.01.25	R Grundlagen	(10) Strukturiertes Programmieren: Kontrollfluss, Debugging	Seminar
14.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel	Übung
	Leistungsnachweis 3		
21.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel	Seminar
28.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel, Q&A	Seminar

(9) Maße der zentralen Tendenz und Datenvariabilität

Maße der zentralen Tendenz

Maße der Datenvariabilität

Programmierübungen und Selbstkontrollfragen

Maße der zentralen Tendenz

Maße der Datenvariabilität

Programmierübungen und Selbstkontrollfragen

Definition (Mittelwert)

$x = (x_1, ..., x_n)$ sei ein Datensatz. Dann heißt

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

der *Mittelwert* von x.

Bemerkung

- Im Kontext der Inferenzstatistik heißt der Mittelwert *Stichprobenmittel*.
- Die Inferenzstatistik gibt der Mittelwertsbildung ihren Sinn.

Berechnung des Mittelwerts in R

"Manuelle" Berechnung des Mittelwerts

```
# Einlesen des Beispieldatensatzes und Abbildungsverzeichnisdefinition
psy_data_path <- file.path(data_dir_path, "psychotherapie_datensatz.csv")
D             <- read.table(psy_data_path, sep = ",", header = TRUE)

# Mittelwertberechnung
x             <- D$Pre.BDI                  # double Vektor der Pre-BDI Werte
n             <- length(x)                  # Anzahl der Werte
x_bar         <- (1 / n) * sum(x)           # Mittelwertsberechnung
print(x_bar)                                # Ausgabe
```

```
[1] 18.61
```

`mean()` zur Berechnung des Mittelwerts

```
x_bar <- mean(x)                  # Mittelwertsberechnung
print(x_bar)                      # Ausgabe
```

```
[1] 18.61
```

Theorem (Eigenschaften des Mittelwerts)

$x = (x_1, ..., x_n)$ und sei ein Datensatz und $\bar{x}$ sei der Mittelwert von x. Dann gelten

(1) Die Summe der Abweichungen vom Mittelwert ist Null,

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0. \tag{2}$$

(2) Die absoluten Summen negativer und positiver Abweichungen vom Mittelwert sind gleich, d.h. wenn $j = 1, ..., n_j$ die Datenpunktindizes mit $(x_j - \bar{x}) < 0$ und $k = 1, ..., n_k$ die Datenpunktindizes mit $(x_k - \bar{x}) \geq 0$ bezeichnen, dann gilt mit $n_j + n_k$

$$|\sum_{j=1}^{n_j}(x_j - \bar{x})| = |\sum_{k=1}^{n_k}(x_k - \bar{x})|. \tag{3}$$

(3) Der Mittelwert der Summe zweier gleich großer Datensätze entspricht der Summe ihrer Mittelwerte, d.h. für einen weiteren Datensatz $y = (y_1, ..., y_n)$ mit Mittelwert $\bar{y}$ gilt

$$\overline{x + y} = \bar{x} + \bar{y} \tag{4}$$

(4) Eine linear-affine Transformation eines Datensatz transformiert den Mittelwert des Datensatzes linear-affin, d.h für $a, b \in \mathbb{R}$ gilt

$$\overline{ax + b} = a\bar{x} + b$$

Eigenschaften des Mittelwerts

Beweis

(1) Es gilt

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = \sum_{i=1}^n x_i - n\bar{x} = \sum_{i=1}^n x_i - \frac{n}{n}\sum_{i=1}^n x_i = \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0.$$

(2) Seien $j = 1, ..., n_j$ die Indizes mit $(x_j - \bar{x}) < 0$ und $k = 1, ..., n_k$ die Indizes mit $(x_k - \bar{x}) \geq 0$, so dass $n = n_j + n_k$. Dann gilt

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \Leftrightarrow \sum_{j=1}^{n_j} (x_j - \bar{x}) + \sum_{k=1}^{n_k} (x_k - \bar{x}) = 0 \Leftrightarrow \sum_{k=1}^{n_k} (x_k - \bar{x}) = -\sum_{j=1}^{n_j} (x_j - \bar{x})$$

$$\Leftrightarrow |\sum_{j=1}^{n_j} (x_j - \bar{x})| = |\sum_{k=1}^{n_k} (x_k - \bar{x})|.$$

Eigenschaften des Mittelwerts

Beweis

(3) Es gilt

$$\overline{x+y} := \frac{1}{n}\sum_{i=1}^{n}(x_i + y_i) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} y_i =: \bar{x} + \bar{y}$$

(4) Es gilt

$$\begin{aligned}
\overline{ax+b} &:= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) \\
&= \sum_{i=1}^{n}\left(\frac{1}{n}ax_i + \frac{1}{n}b\right) \\
&= \sum_{i=1}^{n}\left(\frac{1}{n}ax_i\right) + \sum_{i=1}^{n}\left(\frac{1}{n}b\right) \\
&= a\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} b \\
&= a\bar{x} + b
\end{aligned}$$

Eigenschaften des Mittelwerts

Summe der Abweichungen

```
x <- D$Pre.BDI                          # double Vektor der Werte
s <- sum(x - mean(x))                   # Summe der Abweichungen vom Mittelwert
print(s)                                # Ausgabe (Ergebnis hat Rundungsfehler)
```

```
[1] 5.684342e-14
```

Beträge der positiven und negativen Abweichungen

```
x   <- D$Pre.BDI                        # double Vektor der  Werte
s_1 <- sum(x[x <= mean(x)] - mean(x))   # Summe aller negativer Abweichungen
s_2 <- sum(x[x >  mean(x)] - mean(x))   # Summe aller positiver Abweichungen
print(abs(s_1))                         # Ausgabe des Betrags
```

```
[1] 71.28
```

```
print(abs(s_2))                         # Ausgabe des Betrags
```

```
[1] 71.28
```

Eigenschaften des Mittelwerts

Summation von Datensätzen

```
x       <- D$Pre.BDI                 # double Vektor der Pre.BDI-Werte
x_bar   <- mean(x)                   # Mittelwert der Pre.BDI-Werte
y       <- D$Post.BDI                # double Vektor der Post.BDI Werte
y_bar   <- mean(y)                   # Mittelwert der Post.BDI Werte
z       <- x + y                     # double Vektor der Summe der Pre und Post Werte
z_bar   <- mean(z)                   # Mittelwert der Summe der Werte
print(z_bar)                         # Ausgabe
```

```
[1] 31.68
```

```
xy_bar  <- x_bar + y_bar             # Summe der Mittelwerte der Pre- und Post.BDI Werte
print(xy_bar)                        # Ausgabe
```

```
[1] 31.68
```

Linear-affine Transformation

```
x        <- D$Pre.BDI                # double Vektor der Pre.BDI Werte
x_bar    <- mean(x)                  # Mittelwert der Pre.BDI Werte
a        <- 2                        # Multiplikationskonstante
b        <- 5                        # Additionskonstante
y        <- a * x + b                # linear-affine Transformation der Pre.BDI Werte
y_bar    <- mean(y)                  # Mittelwert der transfomierten Pre.BDI Werte
print(y_bar)                         # Ausgabe
```

```
[1] 42.22
```

```
ax_bar_b <-a * x_bar + b             # Transformation des Mittelwerts
print(ax_bar_b)                      # Ausgabe
```

```
[1] 42.22
```

Median

Definition (Median)

$x = (x_1, ..., x_n)$ sei ein Datensatz und $x_s = (x_{(1)}, ..., x_{(n)})$ der zugehörige aufsteigend sortierte Datensatz. Dann ist der Median von x definiert als

$$\tilde{x} := \begin{cases} x_{((n+1)/2)} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right) & \text{falls } n \text{ gerade} \end{cases} \tag{5}$$

Bemerkungen

- Der Median ist identisch mit dem 0.5-Quantil.
- Mindestens 50% aller x_i sind kleiner oder gleich $\tilde{x}$
- Mindestens 50% aller x_i sind größer oder gleich $\tilde{x}$.
- Anstelle eines Beweises verweisen wir auf untenstehende Abbildungen.

Median - Beispiele

Beispiele

Beispiel für n ungerade

$$n := 5 \Rightarrow \left(\frac{5+1}{2}\right) = (3) \Rightarrow \tilde{x} := x_{(3)}$$

$$x_{(1)}, x_{(2)}, x_{(3)} \leq \tilde{x} \leq x_{(3)}, x_{(4)}, x_{(5)}$$

Beispiel für n gerade

$$n := 6 \Rightarrow \left(\frac{6}{2}\right) = (3), \left(\frac{6}{2}+1\right) = (4) \Rightarrow \tilde{x} := \frac{1}{2}\left(x_{(3)} + x_{(4)}\right)$$

$$x_{(1)}, x_{(2)}, x_{(3)} < \tilde{x} < x_{(4)}, x_{(5)}, x_{(6)}$$

Berechnung des Medians in R

Manuelle Bestimmung des Medians

```
x   <- D$Pre.BDI                                     # double Vektor der Pre.BDI Werte
n   <- length(x)                                     # Anzahl der Werte
x_s <- sort(x)                                       # aufsteigend sortierter Vektor
if (n %% 2 == 1) {                                   # n ungerade, n mod 2 == 1
   x_tilde <- x_s[(n + 1) / 2]
} else {                                             # n gerade, n mod 2 == 0
   x_tilde <- (x_s[n / 2] + x_s[n / 2 + 1]) / 2
}
print(x_tilde)
```

```
[1] 19
```

Berechnung des Medians mit `median()`

```
x_tilde <- median(x)                                 # Berechnung des Medians
print(x_tilde)                                       # Ausgabe
```

```
[1] 19
```

Median vs. Mittelwert

Der Median ist weniger anfällig für Ausreißer als der Mittelwert

```
x       <- D$Pre.BDI                      # double Vektor der Pre.BDI Werte
x_bar   <- mean(x)                        # Mittelwert der Pre.BDI Werte
x_tilde <- median(x)                      # Median der Pre.BDI Werte
print(x_bar)                              # Ausgabe
```

```
[1] 18.61
```

```
print(x_tilde)                            # Ausgabe
```

```
[1] 19
```

```
y       <- x                              # neuer Datensatz...
y[1]    <- 10000                          # ... mit einem Extremwert
y_bar   <- mean(y)                        # Mittelwert des neuen Datensatzes
print(y_bar)                              # Ausgabe
```

```
[1] 118.44
```

```
y_tilde <- median(y)                      # Median des neuen Datensatzes
print(y_tilde)                            # Ausgabe
```

```
[1] 19
```

Definition (Modalwert)

$x := (x_1, ..., x_n)$ mit $x_i \in \mathbb{R}$ sei ein Datensatz, $A := \{a_1, ..., a_k\}$ mit $k \leq n$ seien die im Datensatz vorkommenden verschiedenen Zahlenwerte und $h : A \to \mathbb{N}$ sei die absolute Häufigkeitsverteilung der Zahlwerte von x. Dann ist der *Modalwert (oder Modus)* von x definiert als

$$\text{argmax}_{a \in A} h(a), \tag{6}$$

also der am häufigsten im Datensatz vorkommende Wert.

Bemerkungen

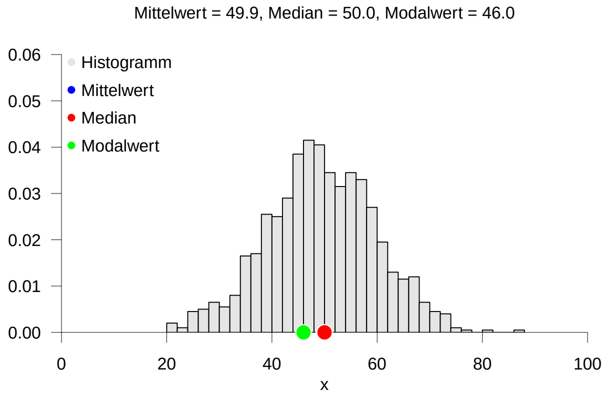
- Modalwerte sind nur bei Datensätzen mit Datenpunktwiederholungen sinnvoll.

Bestimmung des Modalwertes in R

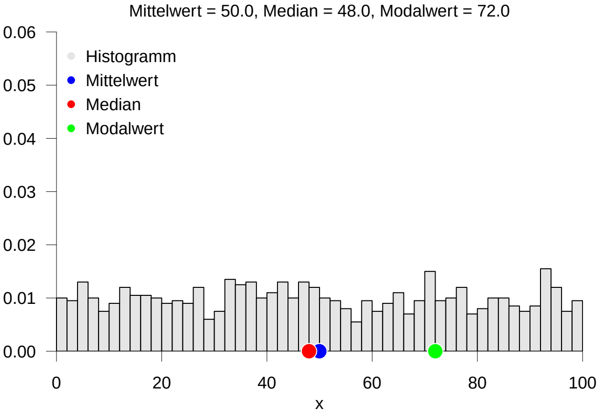
```
x        <- D$Pre.BDI                        # double Vektor der Pre.BDI Werte
n        <- length(x)                        # Anzahl der Datenwerte (100)
H        <- as.data.frame(table(x))          # absolute Haeufigkeitsverteilung (dataframe)
names(H) <- c("a", "h")                      # Konsistente Benennung
mod      <- H$a[which.max(H$h)]              # Modalwert
print(as.numeric(as.vector(mod)))            # Ausgabe als numeric vector, nicht factor
```

```
[1] 18
```

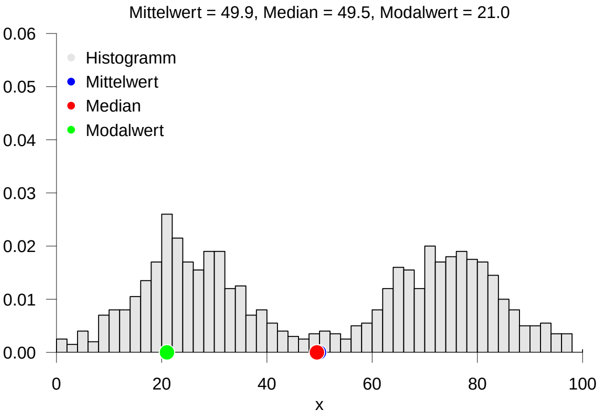
Visuelle Intuition zu Maßen zentraler Tendenz bei **Normalverteilung**



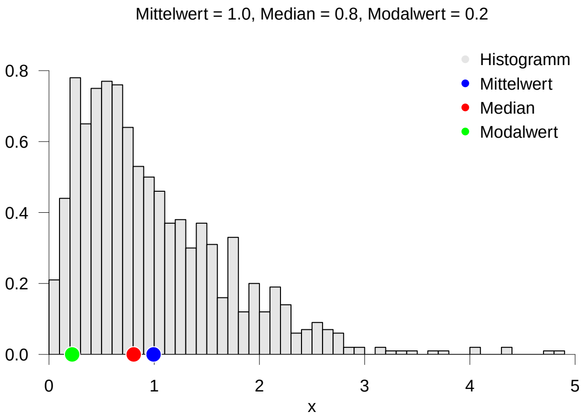
Visuelle Intuition zu Maßen zentraler Tendenz bei **Gleichverteilung**



Visuelle Intuition zu Maßen zentraler Tendenz bei **bimodalen Verteilungen**



Visuelle Intuition [...] bei **nicht-symmetrischen Verteilungen**



Maße der zentralen Tendenz

Maße der Datenvariabilität

Programmierübungen und Selbstkontrollfragen

Definition (Spannbreite)

$x = (x_1, ..., x_n)$ sei ein Datensatz. Dann ist die *Spannbreite* von $x_1, ..., x_n$ definiert als

$$sb := \max(x_1, ..., x_n) - \min(x_1, ..., x_n). \tag{7}$$

Bestimmen der Spannbreite in R

Die Spannbreite kann mit `range()` berechnet werden

```
# Manuelle Spannbreitenberechnung
x       <- D$Pre.BDI                          # double Vektor der Pre-BDI Werte Werte
x_max   <- max(x)                             # Maximum der Pre-BDI Werte
x_min   <- min(x)                             # Mininum der Pre-BDI Werte
sb      <- x_max - x_min                      # Spannbreite
print(sb)
```

```
[1] 9
```

```
# Automatische Spannbreitenberechnung
MinMax  <- range(x)                           # "Automatische" Berechnung von min(x), max(x)
sb      <- MinMax[2] - MinMax[1]              # Spannbreite
print(sb)
```

```
[1] 9
```

```
is.vector(MinMax)
```

```
[1] TRUE
```

Stichprobenvarianz

Definition (Stichprobenvarianz, empirische Stichprobenvarianz)

$x = (x_1, ..., x_n)$ sei ein Datensatz und $\bar{x}$ das Stichprobenmittel. Die *Stichprobenvarianz* von x ist definiert als

$$s^2 := \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 \tag{8}$$

und die *empirische Stichprobenvarianz* von x ist definiert als

$$\tilde{s}^2 := \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2. \tag{9}$$

Bemerkungen

- s^2 ist ein unverzerrter Schätzer von $\mathbb{V}(\xi)$, $\tilde{s}^2$ ist ein verzerrter Schätzer $\mathbb{V}(\xi)$.
- Für $n \to \infty$ gilt $\frac{1}{n} \approx \frac{1}{n-1}$, $\tilde{s}^2$ ist ein asymptotisch unverzerrter Schätzer von $\mathbb{V}(\xi)$.
- $\tilde{s}^2$ ist der ML Schätzer, s^2 ist der ReML Schätzer von σ^2 bei $\xi_1, ..., \xi_n \sim N(\mu, \sigma^2)$.
- Es gelten

$$\tilde{s}^2 = \frac{n-1}{n}s^2, s^2 = \frac{n}{n-1}\tilde{s}^2 \text{ und } 0 \le \tilde{s}^2 < s^2. \tag{10}$$

Berechnen der Stichprobenvarianz in R

Die Sitchprobenvarianz kann mit `var()` berechnet werden

```
x         <- D$Pre.BDI                               # double Vektor der Pre-BDI Werte Werte
n         <- length(x)                               # Anzahl der Werte
s2        <- (1 / (n - 1)) * sum((x - mean(x))^2)    # Stichprobenvarianz
print(s2)
```

```
[1] 3.028182
```

```
s2        <- var(x)                                  # "automatische" Stichprobenvarianz
print(s2)
```

```
[1] 3.028182
```

```
s2_tilde  <- (1 / n) * sum((x - mean(x))^2)          # Empirische Stichprobenvarianz
print(s2_tilde)
```

```
[1] 2.9979
```

```
s2_tilde  <- ((n - 1) / n) * var(x)                  # "automatische" empirische Stichprobenvarianz
print(s2_tilde)
```

```
[1] 2.9979
```

Theorem (Stichprobenvarianz bei linear-affinen Transformationen)

$x = (x_1, ..., x_n)$ sei ein Datensatz mit Stichprobenvarianz s_x^2 und $y = (ax_1 + b, ..., ax_n + b)$ sei der mit $a, b \in \mathbb{R}$ linear-affin transformierte Datensatz mit Stichprobenvarianz s_y^2. Dann gilt

$$s_y^2 = a^2 s_x^2. \tag{11}$$

Beweis

$$\begin{aligned}
s_y^2 := \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2 &= \frac{1}{n-1}\sum_{i=1}^n (ax_i + b - (a\bar{x} + b))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n (ax_i + b - a\bar{x} - b)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n (a(x_i - \bar{x}))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n a^2(x_i - \bar{x})^2 \\
&= a^2 \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 \\
&= a^2 s_x^2
\end{aligned} \tag{12}$$

Stichprobenvarianz bei linear-affinen Transformationen

Beispiel: Stichprobenvarianz bei linear-affinen Transformationen

```
# Stichprobenvarianz nach Transformation
x    <- D$Pre.BDI                       # double Vektor der Pre-BDI Werte
s2x  <- var(x)                          # Stichprobenvarianz von x_1,...,x_n
a    <- 2                               # Multiplikationskonstante
b    <- 5                               # Additionskonstante
y    <- a * x + b                       # y_i <-  ax_i + b
s2y  <- var(y)                          # Stichprobenvarianz y_1,...,y_n
print(s2y)
```

```
[1] 12.11273
```

```
# Stichprobenvarianz nach Theorem
s2y  <- a^2 * s2x                       # Stichprobenvarianz y_1,...,y_n
print(s2y)
```

```
[1] 12.11273
```

Theorem (Verschiebungssatz zur empirischen Stichprobenvarianz)

$x = (x_1, ..., x_n)$ sei ein Datensatz, $x^2 := (x_1^2, ..., x_n^2)$ sei sein elementweises Quadrat und $\bar{x}$ und $\overline{x^2}$ seien die respektiven Mittelwerte. Dann gilt

$$\tilde{s}^2 = \overline{x^2} - \bar{x}^2 \tag{13}$$

Beweis

$$\begin{aligned}
\tilde{s}^2 &:= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 \\
&= \overline{x^2} - 2\bar{x}\bar{x} + \frac{1}{n}n\bar{x}^2 \\
&= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 \\
&= \overline{x^2} - \bar{x}^2
\end{aligned} \tag{14}$$

Verschiebungssatz zur empirischen Stichprobenvarianz

Beispiel: Verschiebungssatz zur empirischen Stichprobenvarianz

```
# Direkte Berechnung der empirischen Stichprobenvarianz
x        <- D$Pre.BDI                            # double Vektor der Pre-BDI Werte Werte
n        <- length(x)                            # Anzahl Datenpunkte
x_bar    <- mean(x)                              # Stichprobenmittel
s2_tilde <- ((n - 1) / n) * var(x)               # empirische Stichprobenvarianz
print(s2_tilde)
```

```
[1] 2.9979
```

```
# Berechnung der empirischen Stichprobenvarianz mit Theorem
s2_tilde <- mean(x^2) - (mean(x))^2              # \bar{x^2} - \bar{x}^2
print(s2_tilde)
```

```
[1] 2.9979
```

```
# Das Theorem gilt nicht für die Stichprobenvarianz
s2       <- var(x)                               # s^2 \neq \bar{x^2} - \bar{x}^2
print(s2)
```

```
[1] 3.028182
```

Definition (Stichprobenstandardabweichung, empirische)

$x = (x_1, ..., x_n)$ sei ein Datensatz. Die *Stichprobenstandardabweichung* von x ist definiert als

$$s := \sqrt{s^2} \tag{15}$$

und die *empirische Stichprobenstandardabweichung* von x ist definiert als

$$\tilde{s} := \sqrt{\tilde{s}^2}. \tag{16}$$

Bemerkungen

- s ist ein verzerrter Schätzer von $\mathbb{S}(\xi)$.
- s^2 misst Variabilität in quadrierten Einheiten, zum Beispiel Quadratmeter (m^2).
- s misst Variabilität in unquadrierten Einheiten, zum Beispiel Meter (m).
- Es gilt

$$\tilde{s} = \sqrt{(n-1)/n}s. \tag{17}$$

Berechnung der Stichprobenstandardabweichung in R

Die Stichprobenstandardabweichung kann mit `sd()` berechnet werden.

```r
# Manuelle Berechnung der Stichprobenstandardabweichung
x       <- D$Pre.BDI                                    # double Vektor der Pre-BDI Werte Werte
n       <- length(x)                                    # Anzahl der Werte
s       <- sqrt((1 / (n - 1)) * sum((x - mean(x))^2))  # Standardabweichung
print(s)
```

```
[1] 1.740167
```

```r
# Automatische Berechnung der Stichprobenstandardabweichung
s       <- sd(x)                                        # "automatische" Berechnung
print(s)
```

```
[1] 1.740167
```

```r
# Empirische Standardabweichung
s_tilde <- sqrt((1 / (n)) * sum((x - mean(x))^2))      # empirische Standardabweichung
print(s_tilde)
```

```
[1] 1.731444
```

```r
s_tilde <- sqrt((n - 1) / n) * sd(x)                    # empirische Standardabweichung
print(s_tilde)
```

```
[1] 1.731444
```

Stichprobenstandardabweichung bei linear-affinen Transformationen

Theorem (Stichprobenstandardabweichung bei linear-affinen Transformationen)

$x = (x_1, ..., x_n)$ sei ein Datensatz mit Stichprobenstandardabweichung s_x und $y = (ax_1 + b, ..., ax_n + b)$ sei der mit $a, b \in \mathbb{R}$ linear-affin transformierte Datensatz mit Stichprobenstandardabweichung s_y. Dann gilt

$$s_y = |a| s_x. \tag{18}$$

<u>Beweis</u>

$$\begin{aligned}
s_y := \left(\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2\right)^{1/2} & = \left(\frac{1}{n-1}\sum_{i=1}^n \left(ax_i + b - (a\bar{x} + b)\right)^2\right)^{1/2} \\
& = \left(\frac{1}{n-1}\sum_{i=1}^n \left(a(x_i - \bar{x})\right)^2\right)^{1/2} \\
& = \left(\frac{1}{n-1}\sum_{i=1}^n a^2 (x_i - \bar{x})^2\right)^{1/2} \\
& = \left(a^2\right)^{1/2} \left(\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2\right)^{1/2}
\end{aligned}$$

Also gilt $s_y = a s_x$, wenn $a \geq 0$ und $s_y = -a s_x$, wenn $a < 0$. Dies aber entspricht $s_y = |a| s_x$.

Beispiel: Stichprobenstandardabweichung bei linear-affinen Transformationen

```
# a >= 0
x   <- D$Pre.BDI                          # double Vektor der Pre-BDI Werte Werte
s_x <- sd(x)                              # Stichprobenstandardabweichung von x
a   <- 2                                  # Multiplikationskonstante
b   <- 5                                  # Additionskonstante
y   <- a*x + b                            # y_i =  ax_i + b
s_y <- sd(y)                              # Stichprobenstandardabweichung von y
print(s_y)
```

```
[1] 3.480334
```

```
s_y <- a*s_x                              # Stichprobenstandardabweichung von y
print(s_y)
```

```
[1] 3.480334
```

```
# a < 0
x   <- D$Pre.BDI                          # double Vektor der Pre-BDI Werte Werte
s_x <- sd(x)                              # Stichprobenstandardabweichung von x
a   <- -3                                 # Multiplikationskonstante
b   <- 10                                 # Additionskonstante
y   <- a*x + b                            # y_i =  ax_i + b
s_y <- sd(y)                              # Stichprobenstandardabweichung von y
print(s_y)
```

```
[1] 5.220502
```

```
s_y <- (-a)*s_x                           # Stichprobenstandardabweichung von y
print(s_y)
```

```
[1] 5.220502
```

Maße der zentralen Tendenz

Maße der Datenvariabilität

Programmierübungen und Selbstkontrollfragen

Programmierübungen

1. Berechne den Mittelwert der Post.BDI Daten.
2. Berechne den Median der Post.BDI Daten.
3. Berechne den Modalwert des Post.BDI Datensatzes.
4. Visualisiere die Häufigkeitsverteilung des Post.BDI Datensatzes und diskutiere die berechneten Werte von Mittelwert, Median und Modalwert vor dem Hintergrund dieser Häufigkeitsverteilung.
5. Berechne die Spannbreite der Post.BDI Daten.
6. Berechne die Stichprobenvarianz und die empirische Stichprobenvarianz der Post.BDI Daten.
7. Berechne die Stichprobenstandardabweichung und die empirische Stichprobenstandardabweichung der Post.BDI Daten.

Selbskontrollfragen

1. Gebe die Definition des Mittelwertes eines Datensatzes wieder.
2. Gebe das Theorem zu den Eigenschaften des Mittelwerts wieder.
3. Gebe die Definition des Median eines Datensatzes wieder.
4. Wie verhalten sich Mittelwert und Median in Bezug auf Datenausreißer?
5. Gebe die Definition des Modalwertes eines Datensatzes wieder.
6. Gebe die Definition der Spannbreite eines Datensatzes wieder.
7. Gebe die Definition der Stichprobenvarianz und der empirischen Stichprobenvarianz wieder.
8. Gebe das Theorem zur Stichprobenvarianz bei linear-affinen Transformationen wieder.
9. Gebe den Verschiebungssatz zur empirischen Stichprobenvarianz wieder.
10. Gebe die Definition der Stichprobenstandardabweichung und der empirischen Stichprobenstandardabweichung wieder.
11. Gebe das Theorem zur Stichprobenstandardabweichung bei linear-affinen Transformationen wieder.