Programmierung und Deskriptive Statistik

BSc Psychologie WiSe 2023/24

Belinda Fleischmann

Datum	Einheit	Thema
11.10.23	Einführung	(1) Einführung
18.10.23	R Grundlagen	(2) R und Visual Studio Code
25.10.23	R Grundlagen	(2) R und Visual Studio Code
01.11.23	R Grundlagen	(3) Vektoren
08.11.23	R Grundlagen	(4) Matrizen
15.11.23	R Grundlagen	(5) Listen und Dataframes
22.11.23	R Grundlagen	(6) Datenmanagement
29.11.23	Deskriptive Statistik	(7) Häufigkeitsverteilungen
06.12.23	Deskriptive Statistik	(8) Verteilungsfunktionen und Quantile
13.12.23	Deskriptive Statistik	(9) Maße der zentralen Tendenz
20.12.23	*Leistungsnachweis Teil 1*	
20.12.23	**Deskriptive Statistik**	**(10) Maße der Datenvariabilität**
	Weihnachtspause	
10.01.24	Deskriptive Statistik	(11) Anwendungsbeispiel (Deskriptive Statistik)
17.01.24	Inferenzstatistik	(12) Anwendungsbeispiel (Parameterschätzung, Konfidenzintervalle)
24.01.24	Inferenzstatistik	(13) Anwendungsbeispiel (Hypothesentest)
25.01.24	*Leistungsnachweis Teil 2*	

(10) Maße der Variabilität

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Übungen und Selbstkontrollfragen

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Übungen und Selbstkontrollfragen

Definition (Spannbreite)

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

$$sb := \max(x_1, ..., x_n) - \min(x_1, ..., x_n). \tag{1}$$

Bestimmen der Spannbreite in R

Die Spannbreite kann mit `range()` berechnet werden

```
# Einlesen des Beispieldatensatzes
fpath   <- file.path(data_path, "psychotherapie_datensatz.csv")
D       <- read.table(fpath, sep = ",", header = T)

# 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
```

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Übungen und Selbstkontrollfragen

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

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

und die *empirische Stichprobenvarianz* von x ist definiert als

$$\tilde{s}^2 := \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{3}$$

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

$$\tilde{s}^2 = \frac{n-1}{n} s^2, s^2 = \frac{n}{n-1} \tilde{s}^2 \text{ und } 0 \leq \tilde{s}^2 < s^2. \tag{4}$$

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

$$s_y^2 = a^2 s_x^2. \tag{5}$$

Beweis

$$\begin{aligned}
s_y^2 := \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i + b - (a\bar{x} + b))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i + b - a\bar{x} - b)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(a(x_i - \bar{x}))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}a^2(x_i - \bar{x})^2 \\
&= a^2\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= a^2 s_x^2
\end{aligned} \tag{6}$$

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
```

Verschiebungssatz zur empirischen Stichprobenvarianz

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

$$\tilde{s}^2 = \overline{x^2} - \bar{x}^2 \tag{7}$$

Beweis

$$\begin{aligned}
\tilde{s}^2 &:= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 \\
&= \overline{x^2} - 2\bar{x}\bar{x} + \frac{1}{n}n\bar{x}^2 \\
&= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 \\
&= \overline{x^2} - \bar{x}^2
\end{aligned} \tag{8}$$

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
```

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Übungen und Selbstkontrollfragen

Stichprobenstandardabweichung

Definition (Stichprobenstandardabweichung, empirische)

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

$$s := \sqrt{s^2} \tag{9}$$

und die *empirische Stichprobenstandardabweichung* von x ist definiert als

$$\tilde{s} := \sqrt{\tilde{s}^2}. \tag{10}$$

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

$$\tilde{s} = \sqrt{(n-1)/n}s. \tag{11}$$

Berechnung der Stichprobenstandardabweichung in R

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

```
# 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
```

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

```
[1] 1.740167
```

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

```
[1] 1.731444
```

```
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

$$s_y = |a| s_x. \tag{12}$$

Beweis

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

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

Stichprobenstandardabweichung bei linear-affinen Transformationen

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
```

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Übungen und Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition der Spannbreite eines Datensatzes wieder.
2. Berechnen Sie die Spannbreite der Post.BDI Daten.
3. Geben Sie die Definition der Stichprobenvarianz und der empirischen Stichprobenvarianz wieder.
4. Berechnen Sie die Stichprobenvarianz und die empirische Stichprobenvarianz der Post.BDI Daten.
5. Geben Sie das Theorem zur Stichprobenvarianz bei linear-affinen Transformationen wieder.
6. Geben Sie den Verschiebungssatz zur empirischen Stichprobenvarianz wieder.
7. Geben Sie die Definition der Stichprobenstandardabweichung und der empirischen Stichprobenstandardabweichung wieder.
8. Berechnen Sie die Stichprobenstandardabweichung und die empirische Stichprobenstandardabweichung der Post.BDI Daten.
9. Geben Sie das Theorem zur Stichprobenstandardabweichung bei linear-affinen Transformationen wieder.