



Programmierung und Deskriptive Statistik

BSc Psychologie WiSe 2022/23

Belinda Fleischmann

Inhalte basieren auf Programmierung und Deskriptive Statistik von Dirk Ostwald, lizenziert unter CC BY-NC-SA 4.0

(10) Maße der Variabilität

Datum	Einheit	Thema
12.10.22	Einführung	(1) Einführung
19.10.22	R Grundlagen	(2) R und RStudio I
26.10.22	R Grundlagen	(2) R und RStudio II
02.11.22	R Grundlagen	(3) Vektoren
09.11.22	R Grundlagen	(4) Matrizen
16.11.22	R Grundlagen	(5) Listen und Dataframes
23.11.22	R Grundlagen	(6) Datenmanagement
30.11.22	Deskriptive Statistik	(7) Häufigkeitsverteilungen I
07.12.22	Deskriptive Statistik	(7) Häufigkeitsverteilungen II
14.12.22	Deskriptive Statistik	(8) Verteilungsfunktionen und Quantile
19.12.22	<i>Leistungsnachweis Teil 1</i>	
21.12.22	Deskriptive Statistik	(9) Maße der zentralen Tendenz
	Weihnachtspause	
04.01.23	Deskriptive Statistik	(10) Maße der Datenvariabilität
11.01.23	Deskriptive Statistik	(12) Anwendungsbeispiel (Deskriptive Statistik)
18.01.23	Inferenzstatistik	(13) Anwendungsbeispiel (Parameterschätzung, Konfidenzintervalle)
25.01.23	Inferenzstatistik	(14) Anwendungsbeispiel (Hypothesentest)
27.01.23	<i>Leistungsnachweis Teil 2</i>	Online, 9:00 - 10:00 Uhr

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Selbstkontrollfragen

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Selbstkontrollfragen

Definition (Spannbreite)

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

$$S := \max(x_1, \dots, x_n) - \min(x_1, \dots, x_n). \quad (1)$$

Berechnen der Spannbreite mit `range()`

```
# Einlesen des Beispieldatensatzes
```

```
fname = file.path(data_dir, "psychotherapie_datensatz.csv")
```

```
D      = read.table(fname, sep = ",", header = T)
```

```
# Manuelle Spannbreitenberechnung
```

```
x      = D$Pre.BDI
```

```
# double Vektor der Pre-BDI Werte
```

```
x_max  = max(x)
```

```
# Maximum der TA1 Werte
```

```
x_min  = min(x)
```

```
# Minimum der TA1 Werte
```

```
S      = x_max - x_min
```

```
# Spannbreite
```

```
print(S)
```

```
> [1] 9
```

```
# automatische Spannbreitenberechnung
```

```
MinMax = range(x)
```

```
# "automatische" Berechnung von min(x), max(x)
```

```
S      = MinMax[2] - MinMax[1]
```

```
# Spannbreite
```

```
print(S)
```

```
> [1] 9
```

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Selbstkontrollfragen

Definition (Stichprobenvarianz, empirische Stichprobenvarianz)

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

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

und die *empirische Stichprobenvarianz* von x ist definiert als

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

Bemerkungen

- S^2 ist ein unverzerrter Schätzer von $\mathbb{V}(X)$, \tilde{S}^2 ist ein verzerrter Schätzer $\mathbb{V}(X)$.
- Für $n \rightarrow \infty$ gilt $\frac{1}{n} \approx \frac{1}{n-1}$, \tilde{S}^2 ist ein asymptotisch unverzerrter Schätzer von $\mathbb{V}(X)$.
- \tilde{S}^2 ist der ML Schätzer, S^2 ist der ReML Schätzer von σ^2 bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$.
- Es gelten

$$\tilde{S}^2 = \frac{n-1}{n} S^2, S^2 = \frac{n}{n-1} \tilde{S}^2 \text{ und } 0 \leq \tilde{S}^2 < S^2. \quad (4)$$

Berechnen der Stichprobenvarianz mit `var()`

```
x      = D$Pre.BDI           # double Vektor der Pre-BDI Werte
n      = length(x)           # Anzahl der Werte
S2     = (1/(n-1))*sum((x - mean(x))^2) # Stichprobenvarianz
print(S2)
```

```
> [1] 3.03
```

```
S2     = var(x)              # "automatische" Stichprobenvarianz
print(S2)
```

```
> [1] 3.03
```

```
S2_tilde = (1/n)*sum((x - mean(x))^2) # Empirische Stichprobenvarianz
print(S2_tilde)
```

```
> [1] 3
```

```
S2_tilde = ((n-1)/n)*var(x)          # "automatische" empirische Stichprobenvarianz
print(S2_tilde)
```

```
> [1] 3
```

Theorem (Stichprobenvarianz bei linear-affinen Transformationen)

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

$$S_y^2 = a^2 S_x^2. \quad (5)$$

Beweis

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

Stichprobenvarianz bei linear-affinen Transformationen

```
# Stichprobenvarianz nach Transformation
```

```
x      = D$Pre.BDI
```

```
S2x     = var(x)
```

```
a      = 2
```

```
b      = 5
```

```
y      = a*x + b
```

```
S2y     = var(y)
```

```
print(S2y)
```

```
# double Vektor der Pre-BDI Werte
```

```
# Stichprobenvarianz von  $x_1, \dots, x_n$ 
```

```
# Multiplikationskonstante
```

```
# Additionskonstante
```

```
#  $y_i = ax_i + b$ 
```

```
# Stichprobenvarianz  $y_1, \dots, y_n$ 
```

```
> [1] 12.1
```

```
# Stichprobenvarianz nach Theorem
```

```
S2y     = a^2*S2x
```

```
# Stichprobenvarianz  $y_1, \dots, y_n$ 
```

```
print(S2y)
```

```
> [1] 12.1
```

Theorem (Verschiebungssatz zur empirischen Stichprobenvarianz)

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

$$\tilde{S}^2 = \overline{x^2} - \bar{x}^2 \quad (7)$$

Beweis

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

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] 3
```

Berechnung der empirischen Stichprobenvarianz mit Theorem

```
S2_tilde   = mean(x^2) - (mean(x))^2    #  $\overline{x^2} - \bar{x}^2$ 
print(S2_tilde)
```

```
> [1] 3
```

Das Theorem gilt nicht für die Stichprobenvarianz

```
S2          = var(x)              #  $S^2 \neq \overline{x^2} - \bar{x}^2$ 
print(S2)
```

```
> [1] 3.03
```

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Selbstkontrollfragen

Definition (Stichprobenstandardabweichung, empirische)

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

$$S := \sqrt{S^2} \quad (9)$$

und die *empirische Stichprobenstandardabweichung* von x ist definiert als

$$\tilde{S} := \sqrt{\tilde{S}^2}. \quad (10)$$

Bemerkungen

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

$$\tilde{S} = \sqrt{(n-1)/n} S. \quad (11)$$

Stichprobenstandardabweichung

Berechnung der Stichprobenstandardabweichung mit sd()

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

```
> [1] 1.74
```

```
# Automatische Berechnung der Stichprobenstandardabweichung
S      = sd(x)                    # "automatische" Berechnung
print(S)
```

```
> [1] 1.74
```

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

```
> [1] 1.73
```

```
S_tilde = sqrt((n-1)/n)*sd(x)          # empirische Standardabweichung
print(S_tilde)
```

```
> [1] 1.73
```

Theorem (Stichprobenvarianz bei linear-affinen Transformationen)

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

$$S_y = |a|S_x. \quad (12)$$

Beweis

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

Also gilt $S_y = aS_x$, wenn $a \geq 0$ und $S_y = -aS_x$, wenn $a < 0$. Dies aber entspricht $S_y = |a|S_x$.

Stichprobenstandardabweichung

Stichprobenstandardabweichung bei linear-affinen Transformationen

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

```
> [1] 3.48
Sy = a*Sx           # Stichprobenvarianz von y
print(Sy)
```

```
> [1] 3.48
# a < 0
x  = D$Pre.BDI      # double Vektor der Pre-BDI Werte
Sx = sd(x)          # Stichprobenvarianz von x
a  = -3             # Multiplikationskonstante
b  = 10             # Additionskonstante
y  = a*x + b        #  $y_i = ax_i + b$ 
Sy = sd(y)          # Stichprobenvarianz von y
print(Sy)
```

```
> [1] 5.22
Sy = (-a)*Sx        # Stichprobenvarianz von y
print(Sy)
```

```
> [1] 5.22
```

Spannbreite

Stichprobenvarianz

Stichprobenstandardabweichung

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition der Spannweite eines Datensatzes wieder.
2. Berechnen Sie die Spannweite 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.