Vorlesung Statistik A
Explorative Datenanalyse

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Wintersemester 2026/2027

Lineare Transformationen

Um die numerische Vergleichbarkeit verschiedener Messwertreihen zu gewährleisten (z. B. wenn mit unterschiedlichen metrischen Skalen gemessen wurde), kann man diese ineinander umrechnen

Eine hierbei häufig auftretende Klasse von Umrechnungen sind die als *lineare Transformationen* bezeichneten (korrekterweise eigentlich *affine Transformationen*) der Form

$$y_i = p \cdot x_i + q \text{ für alle } i = 1, \ldots, n$$

mit Konstanten $p \neq 0$ (*positiv linear*, wenn $p > 0$) und q

Beispiel

- Umrechnung der Temperaturmessung von Grad Celsius in Grad Fahrenheit

Lineare Transformation: Mittelwert

Es gilt

$$\bar{y} = p \cdot \bar{x} + q$$

Beweis

$$\begin{aligned}
\bar{y} &= \frac{1}{n} \sum_{i=1}^{n} y_i \\
&= \frac{1}{n} \sum_{i=1}^{n} (p \cdot x_i + q) \\
&= \frac{1}{n} \sum_{i=1}^{n} p \cdot x_i + \frac{1}{n} \sum_{i=1}^{n} q \\
&= p \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + \frac{1}{n} \cdot n \cdot q \\
&= p \cdot \bar{x} + q
\end{aligned}$$

Lineare Transformation: Varianz

Es gilt

$$\tilde{s}_y^2 = p^2 \cdot \tilde{s}_x^2$$

Beweis

$$\begin{aligned}
\tilde{s}_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}[p \cdot x_i + q - (p \cdot \bar{x} + q)]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}[p \cdot x_i - p \cdot \bar{x}]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}[p^2 \cdot (x_i - \bar{x})^2] \\
&= p^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = p^2 \cdot \tilde{s}_x^2
\end{aligned}$$

Lineare Transformationen

Für die Standardabweichung gilt dann entsprechend

$$\tilde{s}_y = |p| \cdot \tilde{s}_x$$

Zentrierung

- Die spezielle Transformation

  $$y_i = x_i - \bar{x}$$

  wird als *Zentrierung* bezeichnet
- Offensichtlich gilt nach Zentrierung

  $$\bar{y} = 0$$

z-Transformation

Eine in der Statistik sehr häufige Transformation ist die *z-Transformation* (auch *Standardisierung*)

$$z_i = \frac{x_i - \bar{x}}{\tilde{s}_x} \text{ für alle } i = 1, \ldots, n$$

Die z-Transformation ist ein Spezialfall einer linearen Transformation mit

$$p = \frac{1}{\tilde{s}_x}, \qquad q = -\frac{\bar{x}}{\tilde{s}_x}$$

z-Transformation

Beispiel (s. o.)

- Mit $\bar{x} = 10$ und $\tilde{s}_x = \sqrt{17.2} \approx 4.15$ berechnet man

i	x_i	z_i	z_i^2
1	16	1.45	2.09
2	4	−1.45	2.09
3	13	0.72	0.52
4	9	−0.24	0.06
5	8	−0.48	0.23
$\sum$	50	0	5

- Damit ergibt sich

$$\bar{z} = 0, \qquad \tilde{s}_z^2 = \frac{1}{5} \cdot 5 - 0^2 = 1$$

z-Transformation

Es gilt allgemein

$$\bar{z} = 0, \qquad \tilde{s}_z^2 = 1$$

Beweise

$$\bar{z} = \frac{1}{\tilde{s}_x} \cdot \bar{x} - \frac{\bar{x}}{\tilde{s}_x}$$

$$\tilde{s}_z^2 = \frac{1}{\tilde{s}_x^2} \cdot \tilde{s}_x^2 = 1$$

Schiefe einer Verteilung

Schiefe (skewness) als Kennwert einer Verteilung charakterisiert die Abweichung von der Symmetrie

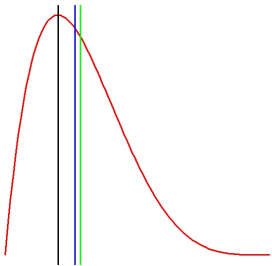
Einen ersten Anhaltspunkt zum Ausmaß der Schiefe liefert die Differenz zwischen dem arithmetischen Mittel $\bar{x}$ und dem Median $\tilde{x}$ der Verteilung

Es gilt

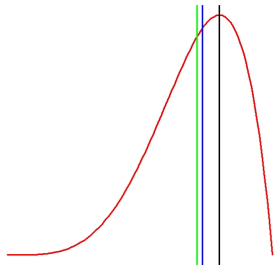
- $\bar{x} - \tilde{x} = 0$: Verteilung symmetrisch (Schiefe gleich Null)
- $\bar{x} - \tilde{x} > 0$: Verteilung tendenziell rechtsschief bzw. linkssteil (positive Schiefe)
- $\bar{x} - \tilde{x} < 0$: Verteilung tendenziell linksschief bzw. rechtssteil (negative Schiefe)

Schiefe

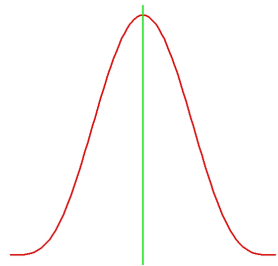
Positive Schiefe
$\bar{x} - \tilde{x} > 0$



Negative Schiefe
$\bar{x} - \tilde{x} < 0$



Keine Schiefe
$\bar{x} - \tilde{x} = 0$



(x_{mod} —, $\tilde{x}$ —, $\bar{x}$ —)

Maßzahlen der Schiefe

Quantilskoeffizient der Schiefe

$$g_p = \frac{(\tilde{x}_{1-p} - \tilde{x}) - (\tilde{x} - \tilde{x}_p)}{\tilde{x}_{1-p} - \tilde{x}_p}$$

Für $p = 0.25$ erhält man den *Quartilskoeffizienten*

Es gilt

- $g_p = 0$ für symmetrische Verteilungen (Schiefe gleich Null)
- $g_p > 0$ für rechtsschiefe bzw. linkssteile Verteilungen (positive Schiefe)
- $g_p < 0$ für linksschiefe bzw. rechtssteile Verteilungen (negative Schiefe)

Maßzahlen der Schiefe

Momentkoeffizient der Schiefe

$$g_m = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\tilde{s}_x}\right)^3 = \frac{1}{n}\sum_{i=1}^{n} z_i^3$$

Es gilt

- $g_m = 0$ für symmetrische Verteilungen (Schiefe gleich Null)
- $g_m > 0$ für rechtsschiefe bzw. linkssteile Verteilungen (positive Schiefe)
- $g_m < 0$ für linksschiefe bzw. rechtssteile Verteilungen (negative Schiefe)

Der Momentkoeffizient liefert eine präzisere, aber nicht so robuste Kennzeichnung der Schiefe, wie der Quantilskoeffizient

Wölbung einer Verteilung

Die Wölbung (Exzess, engl. kurtosis) als Kennwert einer Verteilung charakterisiert deren Schmal- bzw. Breitgipfligkeit

Es gilt

- Je größer die Wölbung einer Verteilung, desto schmalgipfliger ist ihr Verlauf ("peakedness")

Beachte: Die Wölbung einer Verteilung sollte nur bei unimodalen (eingipfligen) Verteilungen berechnet werden

Eine Abschätzung der Wölbung kann über folgenden Momentkoeffizienten vorgenommen werden

Maßzahl der Wölbung

Wölbungsmaß von Fisher

$$\gamma = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\tilde{s}_x}\right)^4 - 3 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3$$

- $\gamma = 0$ normalgipflig (mesokurtisch, Normalverteilung als Referenz)
- $\gamma > 0$ schmalgipflige Verteilung (leptokurtisch)
- $\gamma < 0$ breitgipflige Verteilung (platykurtisch)

Die Momentkoeffizienten g_m für die Schiefe und γ für die Wölbung werden häufig unter Verwendung der Stichprobenvarianz s^2 anstelle der empirischen Varianz $\tilde{s}^2$ definiert

Schiefe und Wölbung einer Verteilung

Beispiel (s. o.): $n = 5$, $\bar{x} = 10$, $\tilde{x} = 9$ ($x_i = 16, 4, 13, 9, 8$)

Schiefe

- Es gilt $\bar{x} - \tilde{x} = 10 - 9 = +1$ (tendenziell rechtsschief/linkssteil)
- Quartilskoeffizient ($p = 0.25$)

p	$n \cdot p$	k	$\tilde{x}_p$
0.25	1.25	2	8
0.75	3.75	4	13

- Damit ergibt sich

$$g_p = \frac{(\tilde{x}_{1-p} - \tilde{x}) - (\tilde{x} - \tilde{x}_p)}{\tilde{x}_{1-p} - \tilde{x}_p} = \frac{(13 - 9) - (9 - 8)}{13 - 8} = \frac{3}{5} = 0.6$$

Schiefe und Wölbung einer Verteilung

Beispiel (s. o.): $n = 5$, $\bar{x} = 10$, $\tilde{s}_x = 4.147$

i	x_i	z_i	z_i^3	z_i^4
1	16	1.447	3.028	4.381
2	4	−1.447	−3.028	4.381
3	13	0.723	0.379	0.274
4	9	−0.241	−0.014	0.003
5	8	−0.482	−0.112	0.054
$\sum$	50	0.000	0.252	9.093

Momentkoeffizient der Schiefe

$$g_m = \frac{1}{n}\sum_{i=1}^{n} z_i^3$$
$$= \frac{1}{5} \cdot 0.252 = 0.050$$

Wölbungsmaß von Fisher

$$\gamma = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3$$
$$= \frac{1}{5} \cdot 9.093 - 3 = -1.181$$

Schiefe und Wölbung: Berechnung mit R

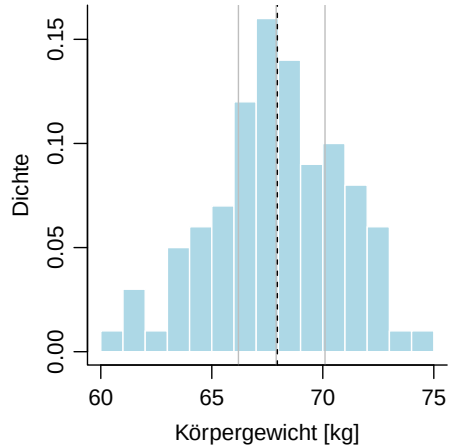
Beispiel: Körpergewicht von 100 männlichen Studenten

```
## Urliste
x <- c(61.1,62.0,62.2,61.8,60.2,63.4,63.5,64.1,65.0,64.2,
       67.5,65.4,64.7,64.4,65.1,65.2,68.2,63.8,63.6,65.3,
       66.9,63.4,65.0,67.2,67.2,67.2,68.4,66.6,67.4,67.6,
       66.0,68.7,66.8,66.1,66.3,66.5,68.9,67.9,69.4,67.7,
       66.4,68.4,67.7,67.2,66.7,66.0,67.8,67.8,67.7,68.2,
       68.8,66.5,68.8,66.8,66.1,67.1,68.7,68.8,68.9,66.8,
       68.3,68.3,67.9,67.3,66.0,70.9,70.5,69.0,69.1,69.7,
       69.6,70.9,70.1,70.1,69.3,71.1,70.2,71.7,69.3,71.9,
       71.1,70.9,71.6,71.0,70.6,69.5,69.1,71.3,70.2,71.1,
       69.5,71.7,73.8,72.3,72.2,73.0,72.4,74.9,72.5,72.3)
```

Schiefe und Wölbung: Berechnung mit R

Beispiel: Körpergewicht von 100 männlichen Studenten

```
hist(x, breaks = 60:75,
     freq = FALSE, right = TRUE,
     col = "lightblue",
     xlab = "Koerpergewicht [kg]",
     ylab = "Dichte")
```



Schiefe und Wölbung: Berechnung mit R

Beispiel: Körpergewicht von 100 männlichen Studenten

```
x.q  <- mean(x)
x.50 <- median(x)
x.q - x.50
[1] 0.053

## Quartilskoeffizient
p    <- 0.25
x.25 <- quantile(x, probs = p,     type = 2)
x.75 <- quantile(x, probs = 1 - p, type = 2)
gp   <- ((x.75 - x.50) - (x.50 - x.25)) / (x.75 - x.25)
[1] 0.128
```

Schiefe und Wölbung: Berechnung mit R

Beispiel: Körpergewicht von 100 männlichen Studenten

```
## z-Transformation
s <- sqrt(mean((x - x.q)^2))  # empir. Standardabweichung
z <- (x - x.q)/s

## Momentkoeffizient der Schiefe
gm <- mean(z^3)
[1] -0.238

## Fishers Woelbungsmass
gamma <- mean(z^4) - 3
[1] -0.144
```