universität
innsbruck

Wiederholung: Arithmetisches Mittel & Varianz

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

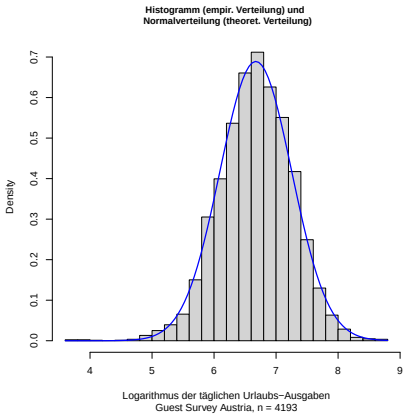
www.hsto.info/econometrics

Beispiel: Gästebefragung Österreich

age	gender	income	**expenditure**	country
57	female	35200	434	Germany
29	male	40000	1025	Austria
35	female	49280	202	Germany
30	female	38205	603	other
48	female	39050	596	Italy
70	male	30000	1489	Austria
⋮	⋮	⋮	⋮	⋮

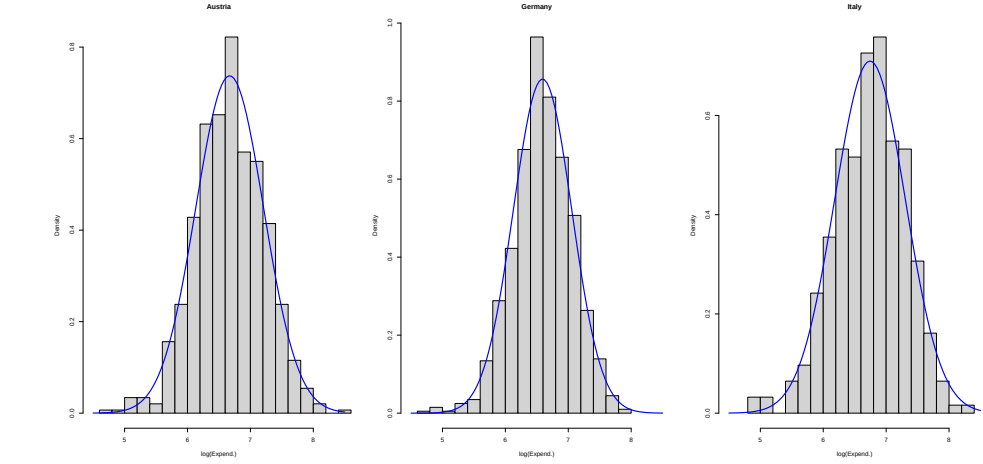
Beispiel: Gästebefragung Österreich

age	gender	income	**expenditure**	country
57	female	35200	434	Germany
29	male	40000	1025	Austria
35	female	49280	202	Germany
30	female	38205	603	other
48	female	39050	596	Italy
70	male	30000	1489	Austria
⋮	⋮	⋮	⋮	⋮



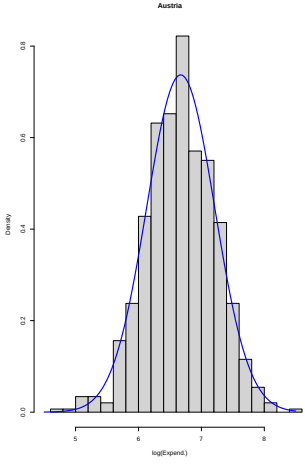
Gästebefragung Österreich

Tägliche Urlaubsausgaben (*expenditure*): Wie vergleichen?

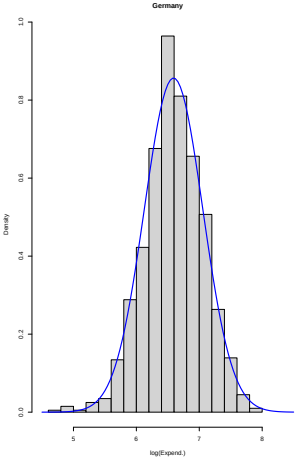


Gästebefragung Österreich

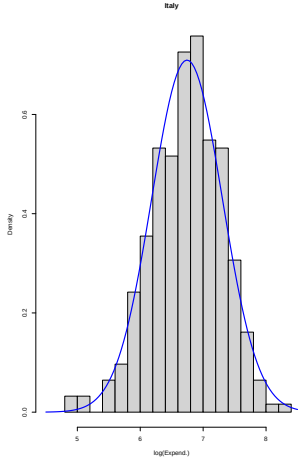
Tägliche Urlaubsausgaben (*expenditure*): Wie vergleichen?



mean = 911, sd = 522



mean = 810, sd = 384



mean = 985, sd = 547

Mittelwerte: Arithmetisches Mittel

- Mittelwert: Überbegriff für verschiedene Lagemaße.
- Arithmetisches Mittel: am häufigsten verwendet

Definition Arithmetisches Mittel

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

- nur für metrisch skalierte Variablen sinnvoll!
- empfindlich gegenüber Extremwerten.

Arithmetisches Mittel

Arithmetisches Mittel: besitzen einige der n Beobachtungen den gleichen numerischen Wert können diese zusammengefasst werden

$$\bar{x} = \frac{1}{n}\left(\underbrace{x_1 + \cdots + x_1}_{n_1\text{-mal}} + \underbrace{x_2 + \cdots + x_2}_{n_2\text{-mal}} + \cdots + \underbrace{x_k + \cdots + x_k}_{n_k\text{-mal}}\right)$$

mit Häufigkeiten $n_1, n_2, \ldots, n_k$

$$\bar{x} = \frac{1}{n}\left(x_1 n_1 + \cdots + x_k n_k\right) = \frac{1}{n}\sum_{j=1}^{k} x_j n_j = \sum_{j=1}^{k} x_j \frac{n_j}{n}$$

mit $\sum_{j=1}^{k} n_j = n$

Arithmetisches Mittel

Arithmetisches Mittel:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{k} x_j n_j = \sum_{j=1}^{k} x_j \frac{n_j}{n} = \quad \text{mit } j = 1, \ldots, k$$

bzw. mit $f_j := \frac{n_j}{n}$ (*relative* Häufigkeiten)

$$\boxed{\bar{x} = \sum_{j=1}^{k} x_j f_j}$$

$\Rightarrow$ **gewogenes arithmetisches Mittel**

Arithmetisches Mittel

Arithmetisches Mittel: 4 Eigenschaften

1. **Schwerpunkteigenschaft** Die Summe der Abweichungen der Einzelwerte vom arithm. Mittel $\bar{x}$ sind Null:

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

weil aus $\bar{x} = \frac{1}{n}\sum_i x_i$ folgt $\sum_i x_i = n\bar{x}$, und $\sum_i \bar{x} = n\bar{x}$

$\rightarrow$ Schwerpunkt einer Verteilung.

Arithmetisches Mittel: 4 Eigenschaften

2. Die Summe der quadrierten Abweichungen von $\bar{x}$ ist kleiner als von jedem beliebigen anderen Wert z

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - z)^2 \quad \text{für } \bar{x} \neq z$$

Arithmetisches Mittel

Arithmetisches Mittel: 4 Eigenschaften

2. Die Summe der quadrierten Abweichungen von $\bar{x}$ ist kleiner als von jedem beliebigen anderen Wert z

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - z)^2 \quad \text{für } \bar{x} \neq z$$

- Warum?

$$\begin{aligned}
\sum_i (x_i - z)^2 &= \sum_i (x_i - \bar{x} + \bar{x} - z)^2 = \sum_i \left[(x_i - \bar{x}) + (\bar{x} - z)\right]^2 \\
&= \sum_i (x_i - \bar{x})^2 + 2(\bar{x} - z) \underbrace{\sum_i (x_i - \bar{x})}_{=0} + \sum_i (\bar{x} - z)^2 \\
&= \sum_i (x_i - \bar{x})^2 + \sum_i (\bar{x} - z)^2 \quad \text{mit } \sum_i (\bar{x} - z)^2 > 0
\end{aligned}$$

Arithmetisches Mittel: 4 Eigenschaften

2. Die Summe der quadrierten Abweichungen von $\bar{x}$ ist kleiner als von jedem beliebigen anderen Wert z

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - z)^2 \quad \text{für } \bar{x} \neq z$$

- Warum?

$$\begin{aligned}
\sum_i (x_i - z)^2 &= \sum_i (x_i - \bar{x} + \bar{x} - z)^2 = \sum_i \left[(x_i - \bar{x}) + (\bar{x} - z)\right]^2 \\
&= \sum_i (x_i - \bar{x})^2 + 2(\bar{x} - z)\underbrace{\sum_i (x_i - \bar{x})}_{=0} + \sum_i (\bar{x} - z)^2 \\
&= \sum_i (x_i - \bar{x})^2 + \sum_i (\bar{x} - z)^2 \quad \text{mit } \sum_i (\bar{x} - z)^2 > 0
\end{aligned}$$

- $\Rightarrow$ arithm. Mittel $\bar{x}$ erzeugt kleinstmögliche Streuung (Varianz)!

Arithmetisches Mittel: 4 Eigenschaften

3. **Translationsäquivariant:** werden die Einzelwerte linear transformiert $x_i^* = b_1 + b_2 x_i$, dann gilt

$$\bar{x}^* = b_1 + b_2 \bar{x}$$

Arithmetisches Mittel: 4 Eigenschaften

3. **Translationsäquivariant:** werden die Einzelwerte linear transformiert $x_i^* = b_1 + b_2 x_i$, dann gilt

$$\bar{x}^* = b_1 + b_2 \bar{x}$$

- Warum?

$$\begin{aligned} \bar{x}^* &= \frac{1}{n} \sum_i (b_1 + b_2 x_i) \\ &= \frac{1}{n} \left(n b_1 + b_2 \sum_i x_i \right) \\ &= b_1 + b_2 \frac{1}{n} \sum_i x_i = b_1 + b_2 \bar{x} \end{aligned}$$

Arithmetisches Mittel

Arithmetisches Mittel: 4 Eigenschaften

4. **gewichtetes arithm. Mittel:** Zwei (oder mehrere) Teilgesamtheiten, deren Umfang und arithm. Mittel bekannt sind ($n_1, \bar{x}_1, n_2, \bar{x}_2$ mit $n_1 + n_2 = n$)

$$\begin{aligned} \bar{x} &= \frac{1}{n}\left(\frac{x_1}{x_1}\sum_{i=1}^{n_1} x_{1i} + \frac{n_2}{n_2}\sum_{i=1}^{n_2} x_{2i}\right) \\ &= \frac{n_1}{n}\left(\frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}\right) + \frac{n_2}{n}\left(\frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i}\right) = \frac{n_1}{n}\,\bar{x}_1 + \frac{n_2}{n}\,\bar{x}_2 \end{aligned}$$

Arithmetisches Mittel

Arithmetisches Mittel: 4 Eigenschaften

4. **gewichtetes arithm. Mittel:** Zwei (oder mehrere) Teilgesamtheiten, deren Umfang und arithm. Mittel bekannt sind ($n_1, \bar{x}_1, n_2, \bar{x}_2$ mit $n_1 + n_2 = n$)

$$\bar{x} = \frac{1}{n}\left(\frac{x_1}{x_1}\sum_{i=1}^{n_1} x_{1i} + \frac{n_2}{n_2}\sum_{i=1}^{n_2} x_{2i}\right)$$

$$= \frac{n_1}{n}\left(\frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}\right) + \frac{n_2}{n}\left(\frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i}\right) = \frac{n_1}{n}\,\bar{x}_1 + \frac{n_2}{n}\,\bar{x}_2$$

- Warum?
  Aus der Definition des arithm. Mittels

$$\sum_{i=1}^{n_1} x_{1i} = n_1\bar{x}_1 \text{ und } \sum_{i=1}^{n_2} x_{2i} = n_2\bar{x}_2$$

Varianz

Definition Varianz

Varianz s^2: mittlere quadratische Abweichung vom arithmetischen Mittel $\bar{x}$.

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

nur für metrisch skalierte Variablen.

Varianz

Warum?

$$
\begin{aligned}
\mathrm{var}(x) &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\left(\sum_i x_i^2 - 2\bar{x}\sum_i x_i + \sum_i \bar{x}^2\right) \\
&= \frac{1}{n}\left(\sum_i x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2 := \overline{x^2} - \bar{x}^2
\end{aligned}
$$

weil $\sum_i x_i = n\bar{x}$ und $\sum_i \bar{x}^2 = n\bar{x}^2$

Varianz

Linear Transformierte Daten:

Sei $x_i^* = b_1 + b_2 x_i$ für $i = 1, \ldots, n$

$$\text{var}(x^*) = \frac{1}{n} \sum_{i=1}^{n} (x_i^* - \bar{x}^*)^2$$

Wir haben bereits früher gezeigt, dass $\bar{x}^* = b_1 + b_2 \bar{x}$.

$$\begin{aligned} \text{var}(x^*) &= \frac{1}{n} \sum_i (b_1 + b_2 x_i - b_1 - b_2 \bar{x})^2 \\ &= \frac{1}{n} \sum_i (b_2 [x_i - \bar{x}])^2 \\ &= b_2^2 \frac{1}{n} \sum_i (x_i - \bar{x})^2 = b_2^2 \, \text{var}(x_i) \end{aligned}$$

$\Rightarrow$ Addition oder Subtraktion einer Konstante. hat keinen Einfluss auf die Varianz!

Varianz

Zwei Arten der Varianz

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{versus} \quad s_s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- die meisten Programme berechnen die Varianz nach der 2. Formel, d.h. sie verwenden den Vorfaktor $1/(n-1)$
- die Anwendung des Vorfaktor $1/(n-1)$ statt $1/n$ ist angebracht, wenn die Varianz aus einer Stichprobe berechnet wird und als *Schätzung* für die Varianz der Grundgesamtheit dient.
- der Grund dafür liegt im Konzept der später diskutierten *Erwartungstreue*.

Standardabweichung

- Die Varianz ist manchmal schwierig zu interpretieren, wenn z.B. x in Euro gemessen wird, hat die Varianz die Dimension Euro^2.
- Die *Standardabweichung* hat gegenüber der Varianz den Vorteil, dass sie in der gleichen Einheit wie die Beobachtungswerte gemessen wird.

Definition Standardabweichung

$$s = \sqrt{\mathrm{var}(x)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Metrisch skalierte Merkmale: Zusammenhangsmaß

Das wichtigste Zusammenhangsmaß für **metrisch skalierte Merkmale** ist die **empirische Kovarianz**.

Definition Kovarianz

Die Kovarianz ist eine (nicht standardisierte) Maßzahl für den Zusammenhang zwischen zwei metrisch skalierten statistischen Merkmalen x und y.

$$\operatorname{cov}(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

mit $\bar{x} := \frac{1}{n} \sum_i x_i$ und $\bar{y} := \frac{1}{n} \sum_i y_i$

Kovarianz

Beispiel:

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
	2	1	-3	-2	6
	3	4	-2	1	-2
	4	1	-1	-2	2
	6	4	1	1	1
	7	2	2	-1	-2
	8	6	3	3	9
$\sum$	30	18	0	0	14

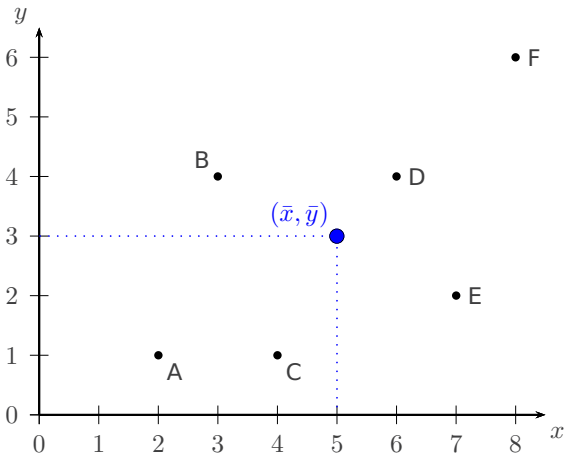
Kovarianz

Beispiel:

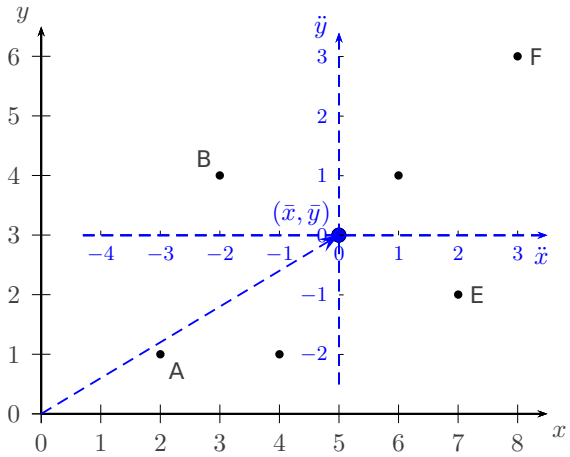
	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
	2	1	-3	-2	6
	3	4	-2	1	-2
	4	1	-1	-2	2
	6	4	1	1	1
	7	2	2	-1	-2
	8	6	3	3	9
$\sum$	30	18	0	0	14

$$\mathrm{cov}(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{14}{6} = 2.33$$

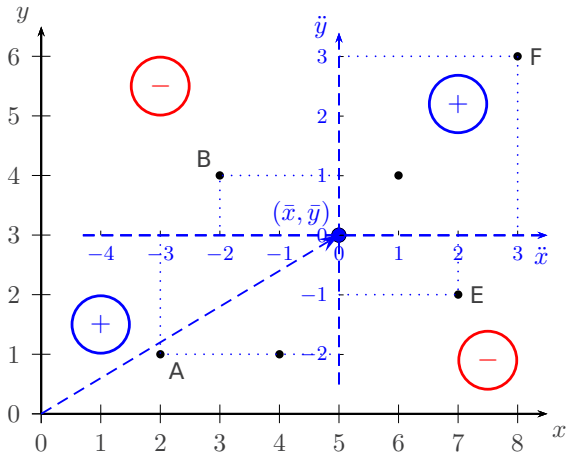
Kovarianz: Mittelwerttransformation



Kovarianz: Mittelwerttransformation



Vorzeichen der Kovarianz



Kovarianz

Kovarianz:

- Die Kovarianz ist positiv, wenn x und y tendenziell einen gleichgerichteten linearen Zusammenhang aufweisen, d.h. hohe Werte von x gehen mit hohen Werten von y einher und niedrige mit niedrigen.
- Die Kovarianz ist negativ, wenn x und y einen gegengerichteten linearen Zusammenhang aufweisen.
- Ist die Kovarianz Null, so besteht kein *linearer Zusammenhang* (es kann aber trotzdem oder ein nicht-linearer Zusammenhang bestehen, z.B. U-förmig).

Rechenregeln für empirische Kovarianzen

1) Symmetrie:

$$\boxed{\operatorname{cov}(x, y) = \operatorname{cov}(y, x)}$$

Rechenregeln für empirische Kovarianzen

1) Symmetrie:

$$\boxed{\mathrm{cov}(x,y) = \mathrm{cov}(y,x)}$$

Warum?

$$\begin{aligned}\mathrm{cov}(x,y) &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) = \mathrm{cov}(y,x)\end{aligned}$$

Rechenregeln für empirische Kovarianzen

2) Konstante Faktoren können ausgeklammert werden: für $x, y \in \mathbb{R}^n$ und Zahlen $a, b \in \mathbb{R}$

$$\boxed{\operatorname{cov}(ax, by) = ab \operatorname{cov}(x, y)}$$

Rechenregeln für empirische Kovarianzen

2) Konstante Faktoren können ausgeklammert werden: für $x, y \in \mathbb{R}^n$ und Zahlen $a, b \in \mathbb{R}$

$$\boxed{\operatorname{cov}(ax, by) = ab \operatorname{cov}(x, y)}$$

Warum?

$$\begin{aligned}
\operatorname{cov}(ax, by) &= \frac{1}{n} \sum_{i=1}^{n} (ax_i - a\bar{x})(by_i - b\bar{y}) \\
&= \frac{1}{n} \sum_{i=1}^{n} a(x_i - \bar{x}) b(y_i - \bar{y}) \\
&= ab \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \\
&= ab \operatorname{cov}(x, y)
\end{aligned}$$

Rechenregeln für empirische Kovarianzen

3) Additivität: für $x, y, z \in \mathbb{R}^n$

$$\boxed{\operatorname{cov}[x, (y + z)] = \operatorname{cov}(x, y) + \operatorname{cov}(x, z)}$$

Rechenregeln für empirische Kovarianzen

3) Additivität: für $x, y, z \in \mathbb{R}^n$

$$\boxed{\text{cov}[x, (y+z)] = \text{cov}(x, y) + \text{cov}(x, z)}$$

Warum?

$$\begin{aligned}
\text{cov}[x, (y+z)] &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})[(y_i + z_i) - (\overline{y+z})] \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})[(y_i + z_i) - (\bar{y} + \bar{z})] \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})[(y_i - \bar{y}) + (z_i - \bar{z})] \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z}) \\
&= \text{cov}(x, y) + \text{cov}(x, z)
\end{aligned}$$

Rechenregeln für empirische Kovarianzen

4) Zusammenhang mit empirischer Varianz: für $x \in \mathbb{R}^n$

$$\boxed{\operatorname{cov}(x, x) = \operatorname{var}(x)}$$

Rechenregeln für empirische Kovarianzen

4) Zusammenhang mit empirischer Varianz: für $x \in \mathbb{R}^n$

$$\boxed{\operatorname{cov}(x, x) = \operatorname{var}(x)}$$

Warum?

$$\begin{aligned} \operatorname{cov}(x, x) &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x}) \\ &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\ &= \operatorname{var}(x) \end{aligned}$$

Rechenregeln für empirische Kovarianzen

5) Empirische Varianz einer Summe: für $x, y \in \mathbb{R}^n$

$$\boxed{\operatorname{var}(x+y) = \operatorname{var}(x) + \operatorname{var}(y) + 2\operatorname{cov}(x,y)}$$

Rechenregeln für empirische Kovarianzen

5) Empirische Varianz einer Summe: für $x, y \in \mathbb{R}^n$

$$\boxed{\operatorname{var}(x+y) = \operatorname{var}(x) + \operatorname{var}(y) + 2\operatorname{cov}(x,y)}$$

Warum?

$$\begin{aligned}
\operatorname{var}(x+y) &= \frac{1}{n}\sum_{i=1}^{n}[(x_i+y_i)-(\bar{x}+\bar{y})]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}[(x_i-\bar{x})+(y_i-\bar{y})]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 + \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 + \\
&\quad +2\,\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\
&= \operatorname{var}(x) + \operatorname{var}(y) + 2\operatorname{cov}(x,y)
\end{aligned}$$

Kovarianz

Zwei Arten der Kovarianz

$$\mathrm{cov}(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

versus

$$\mathrm{cov}_s(x, y) = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

- die meisten Programme berechnen die Kovarianz nach der 2. Formel, d.h. sie verwenden den Vorfaktor $1/(n-1)$
- die Anwendung des Vorfaktor $1/(n-1)$ statt $1/n$ ist angebracht, wenn die Kovarianz aus einer Stichprobe berechnet wird und als *Schätzung* für die Kovarianz der Grundgesamtheit dient.

Korrelation

Korrelationen:

- *Kovarianzen* hängen von Maßeinheiten ab! Um einen Zusammenhang vergleichbar zu machen, muss die Kovarianz normiert werden ⇒ Korrelationskoeffizienten
- Korrelationen sind eine Gruppe von statistischen Kennwerten, die den *“Zusammenhang”* zwischen zwei Variablen messen sollen.
- Bewegen sich die Variablen in die selbe Richtung? Wie stark hängen sie zusammen?

Korrelationskoeffizient nach Bravais-Pearson

Definition *Korrelationskoeffizient nach Bravais-Pearson*

Der Korrelationskoeffizient r ist ein dimensionsloses Maß für den Grad des linearen Zusammenhangs zwischen zwei *mindestens intervallskalierten* Merkmalen.

$$\operatorname{corr}(x, y) := r_{xy} = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)}\sqrt{\operatorname{var}(y)}}$$

- erfordert mindestens *metrisches* Skalenniveau!

Korrelationskoeffizient nach Bravais-Pearson

Eigenschaften des Korrelationskoeffizient nach Bravais-Pearson: für Datenvektoren $x, y \in \mathbb{R}^n$ und Zahlen $a, b, c, d \in \mathbb{R}$ gilt

1. $r_{x,y}$ kann nur Werte zwischen -1 und $+1$ annehmen

$$-1 \leq \operatorname{corr}(x, y) \leq +1$$

2. $r_{x,y}$ ändert sich nicht bei einer linearen Transformation

$$\operatorname{corr}(ax + b, cy + d) = \operatorname{corr}(x, y)$$

3. Wenn der $\operatorname{corr}(x, y) = 0$ sind die beiden Merkmale linear unabhängig (sie können aber trotzdem nicht-linear abhängig sein); wenn $|\operatorname{corr}(x, y)| = 1$ sind die Merkmale exakt linear abhängig
   - $\operatorname{corr}(x, y) = +1$ wenn $y = a + bx$
   - $\operatorname{corr}(x, y) = -1$ wenn $y = a - bx$

Korrelationskoeffizient nach Bravais-Pearson

Beispiel: mit $\ddot{x} := x - \bar{x}$, $\ddot{y} := y - \bar{y}$

	x	y	$\ddot{x}$	$\ddot{x}^2$	$\ddot{y}$	$\ddot{y}^2$	$\ddot{x}\ddot{y}$
	2	1	-3	9	-2	4	6
	3	4	-2	4	1	1	-2
	4	1	-1	1	-2	4	2
	6	4	1	1	1	1	1
	7	2	2	4	-1	1	-2
	8	6	3	9	3	9	9
$\sum$	30	18	0	28	0	20	14

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}\sqrt{\operatorname{var}(y)}} = \frac{14}{\sqrt{28 \cdot 20}} = 0.591608$$

Korrelationskoeffizient nach Bravais-Pearson

Übung: Zeigen Sie, dass der Korrelationskoeffizient

$$\begin{aligned} r &= \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}} \\ &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \end{aligned}$$

alternativ berechnet werden kann als

$$r = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_i x_i^2 - n\bar{x})(\sum_i y_i^2 - n\bar{y}^2)}}$$