Datenanalyse

Sitzung 8: Korrelation

Institut für Kommunikationswissenschaft und Medienforschung
Ludwig-Maximilians-Universität München

Ablauf der Sitzung

1. Kurze Wiederholung: Korrelation
2. Übung 1: Korrelation verstehen
3. Übung 2: Korrelation manuell berechnen
4. Korrelation in R berechnen
5. Übung 3: Korrelation in R berechnen

1. Kurze Wiederholung:
Korrelation

Was ist eine Korrelation?

Eine Korrelation bezeichnet die **Wechselbeziehung zwischen zwei metrischen Variablen**.

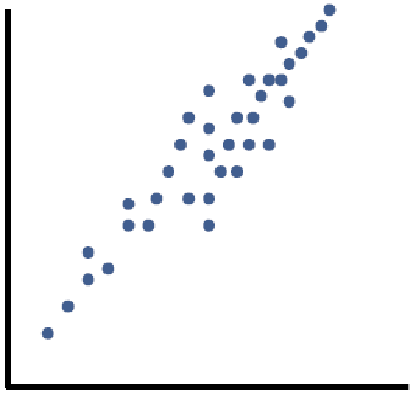
Für die Interpretation dieser Beziehung ist die **Stärke** und **Richtung** des Zusammenhangs von Bedeutung.

Arten von Korrelationen

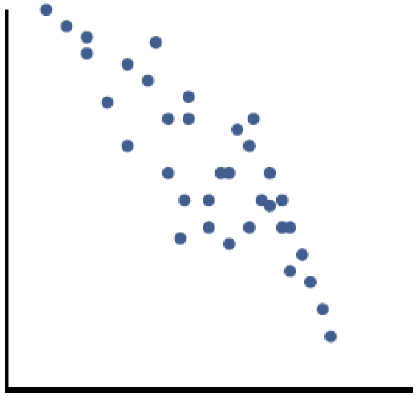
- positiver Zusammenhang ($r_{XY} > 0$):
  je größer X, desto größer Y bzw. je kleiner X, desto kleiner Y
- kein Zusammenhang ($r_{XY} \approx 0$)
- negativer Zusammenhang ($r_{XY} < 0$):
  je größer X, desto kleiner Y bzw. je kleiner X, desto größer Y

Grafische Darstellung: Streudiagramm

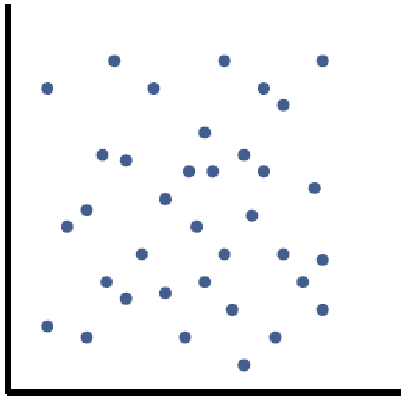
- Englisch: *scatter plot*
- Jedes Wertepaar eines Falls wird in einem Koordinatensystem als Punkt mit den Koordinaten $(x; y)$ darstellt
- Formen von Zusammenhängen:



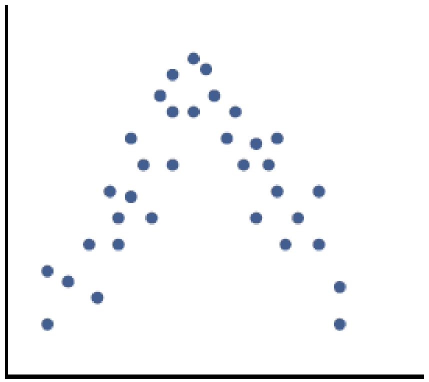
positiv linear



negativ linear



kein Zusammenhang



nicht linearer Zusammenhang (z.B. umgekehrt U-förmig)

Unstandardisiertes Maß: Kovarianz s_{XY}

$$s_{XY} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

Zur Erinnerung:

Die Kovarianz ist ein Maß für den linearen Zusammenhang zweier Variablen. Sie ist eng verwandt mit der Korrelation (aber: nicht standardisiert, daher nicht gut vergleichbar).

Standardisiertes Maß: Korrelationskoeffizient $\boldsymbol{r_{XY}}$

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}$$

Zur Erinnerung:

- Wertebereich von −1 bis +1
- Vorzeichen von *r* gibt die Richtung, Betrag von *r* die Stärke des Zusammenhangs an
- Berechnung der Standardabweichung: $s_X = \sqrt{s_X^2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$

Sonderform: partieller Korrelationskoeffizient

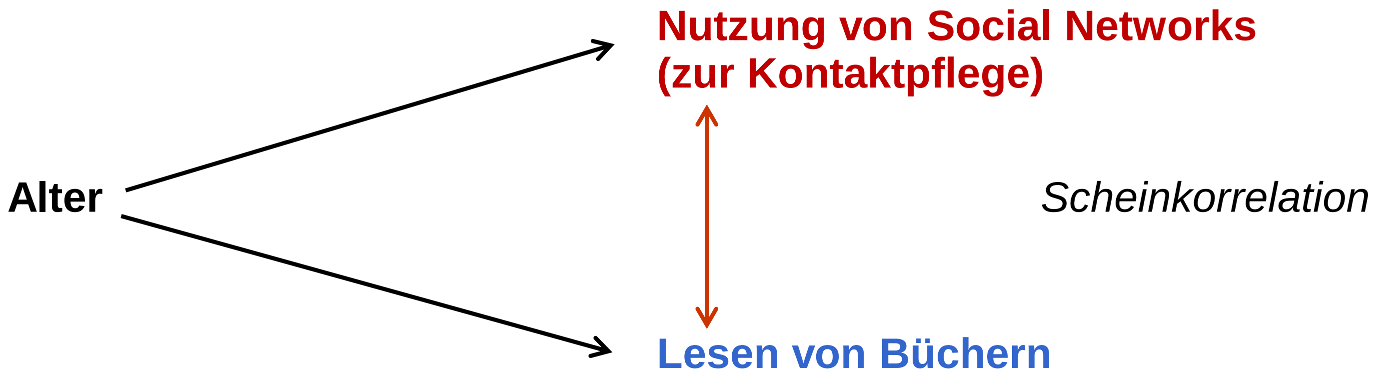
$$r_{XY|Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{1 - r^2{}_{XZ}} \cdot \sqrt{1 - r_{YZ}^2}}$$

Zur Erinnerung:

Beschreibt den Zusammenhang zwischen den Variablen X und Y, nachdem der Einfluss einer Drittvariablen Z kontrolliert wird (d.h. der Einfluss der Drittvariable wird „herausgerechnet“)

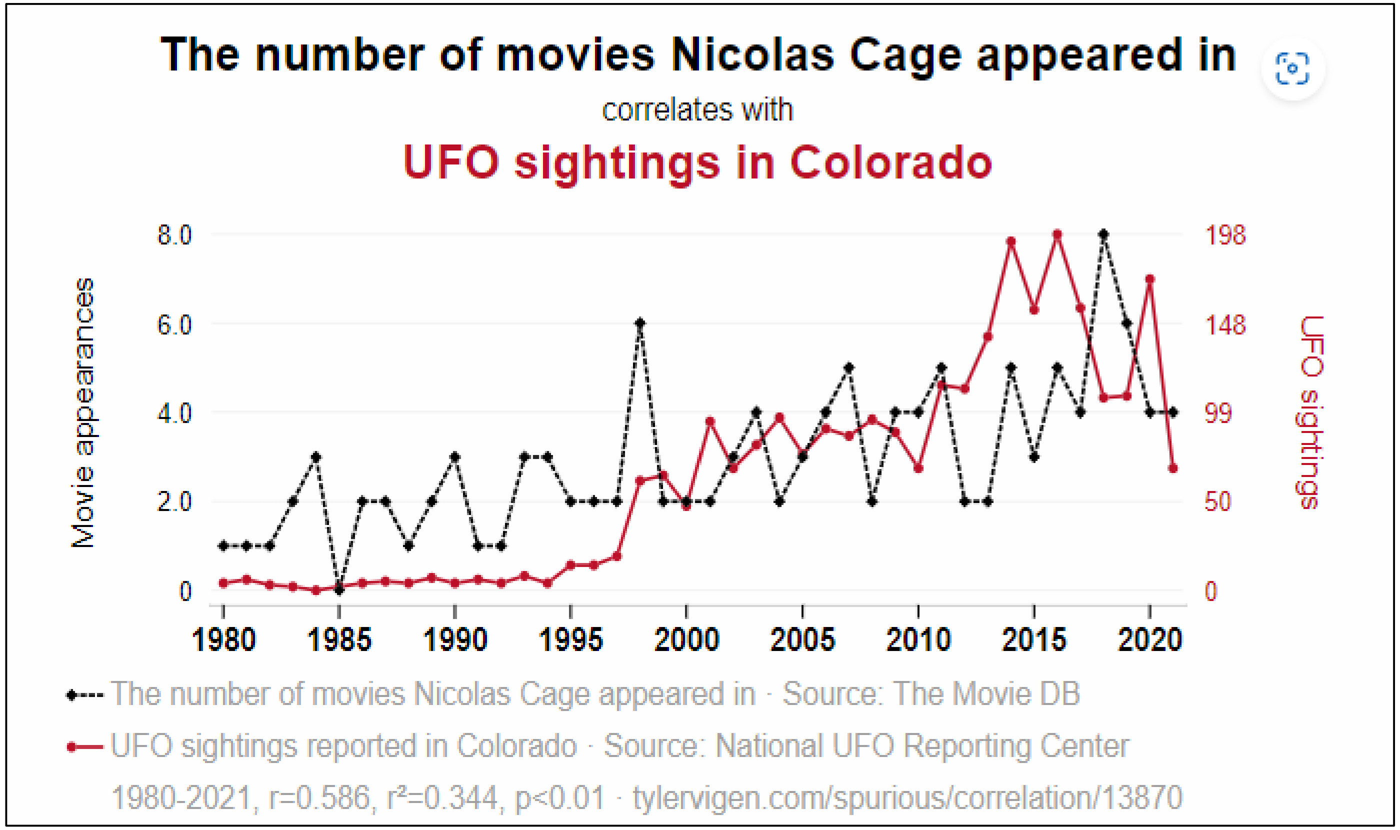
Korrelation und Kausalität

- Zusammenhangsmaße geben <u>keine</u> Auskunft über die Kausalrichtung eines Zusammenhangs (X → Y, Y → X, X ↔ Y)
- Drittvariablenproblem: Zusammenhang der beiden Variablen könnte auch durch eine dritte Variable als deren „gemeinsame Ursache“ begründet sein, die ihrerseits mit *beiden* Variablen korreliert (→ „**Scheinkorrelation**“)
- **Beispiel**: Wer häufig Social Networks nutzt, liest weniger Bücher.



→ Kontrolle von Drittvariablen durch Partialkorrelation

Achtung: Korrelation ≠ Kausalzusammenhang



Quelle: https://tylervigen.com/spurious/correlation/13870

Wichtige Take-Aways

- **Kovarianz**: wechselseitige Varianz zwischen zwei Variablen, d.h. nichtstandardisierte Maßzahl für deren (linearen) Zusammenhang. Abhängig vom Wertebereich der Variablen.
- **Korrelation**: standardisierte Maßzahl für die (lineare) Wechselbeziehung zwischen zwei metrischen Variablen

2. Übung 1:

Korrelation verstehen

Übung 1: Korrelation verstehen

Frage 1: Was sind Voraussetzungen, um eine Pearson-Korrelation zwischen zwei Variablen berechnen zu können?

A) Beide Variablen müssen metrisch skaliert sein.

B) Die Variablen müssen normalverteilt sein.

C) Die Variablen müssen kausal zusammenhängen.

D) Die Variablen müssen einen linearen Zusammenhang haben.

Lösung für Übung 1: Korrelation verstehen

Frage 1: Was sind Voraussetzungen, um eine Pearson-Korrelation zwischen zwei Variablen berechnen zu können?

A) Beide Variablen müssen metrisch skaliert sein.

B) Die Variablen müssen normalverteilt sein.

D) Die Variablen müssen einen linearen Zusammenhang haben.

Übung 1: Korrelation verstehen

Frage 2: Welche der folgenden Annahmen ließe sich mit einer Korrelation auf Basis von Befragungsdaten belegen?

A) Je häufiger die Teilnahme am R-Tutorium, desto besser die Statistik-Note.

B) Die Teilnahme am R-Tutorium hat einen positiven Effekt auf den eigenen Spaß an der Statistik.

C) Je häufiger die Teilnahme am R-Tutorium, desto höher der Spaß an der Statistik bzw. je höher der Spaß an der Statistik, desto häufiger die Teilnahme am R-Tutorium.

Lösung für Übung 1: Korrelation verstehen

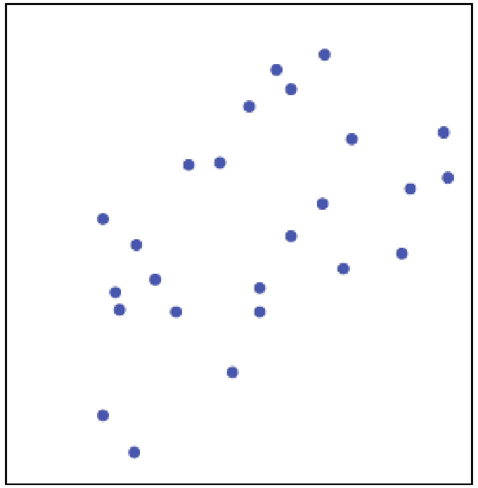
Frage 2: Welche der folgenden Hypothesen ließe sich mit einer Korrelation auf Basis von Befragungsdaten beantworten?

C) Je häufiger die Teilnahme am R-Tutorium, desto höher der Spaß an der Statistik bzw. je höher der Spaß an der Statistik, desto häufiger die Teilnahme am R-Tutorium.

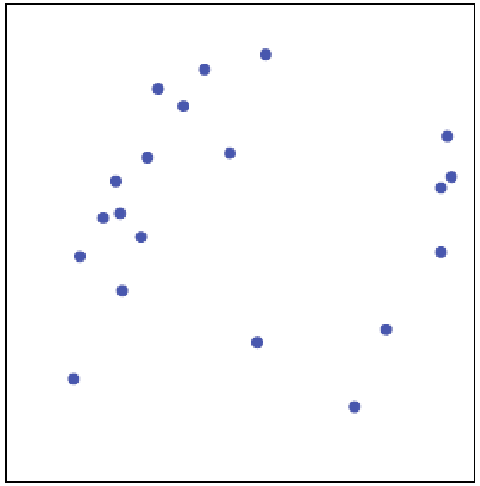
Übung 1: Korrelation verstehen

Frage 3: Welche der folgenden Grafiken illustriert einen linearen, positiven Zusammenhang zwischen zwei Variablen – und woran sieht man dies?

A



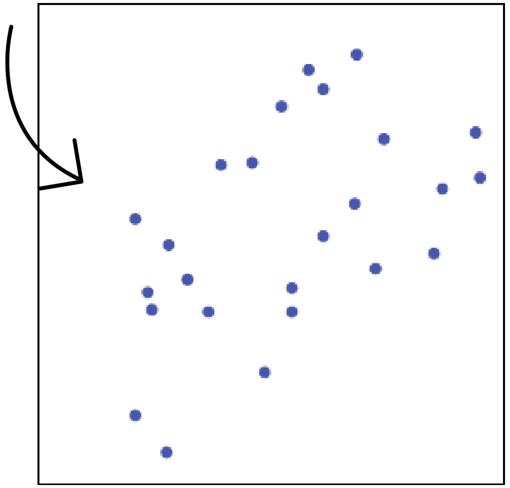
B



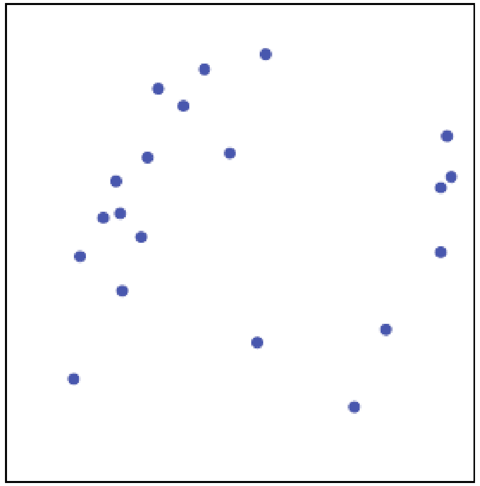
Lösung für Übung 1: Korrelation verstehen

Frage 3: Welche der folgenden Grafiken illustriert einen linearen, positiven Zusammenhang zwischen zwei Variablen – und woran sieht man dies?

A



B



3. Übung 2:

Korrelation manuell berechnen

Übung 2 (Hausaufgabe)

Pierre erhebt bei drei zufällig ausgewählten Mitgliedern seiner Familie (N = 3) jeweils die tägliche Fernsehnutzungsdauer (in Stunden; $\bar{x}$ = 2,0; s^2 = 1,0) und ihr Vertrauen in andere Menschen (Skala 1-10; $\bar{x}$ = 5,0; s^2 = 1,0). Er nimmt entsprechend der Kultivierungshypothese an, dass TV-Nutzungsdauer und interpersonales Vertrauen negativ zusammenhängen. Er erhebt folgende Daten:

ID	TV-Nutzung	Vertrauen
1	1	5
2	2	6
3	3	4

Übung 2 (Hausaufgabe)

a) Pierre setzt voraus, dass die Variablen linear zusammenhängen. Berechnen Sie ein geeignetes Maß, das einen möglichen Zusammenhang der Variablen in der Stichprobe ausdrückt. Was lässt sich anhand des Resultats über den Zusammenhang in der Stichprobe sagen?

b) Kann Pierre auch für die gesamte Familie einen Zusammenhang annehmen? (Signifikanzniveau 5%)

c) Pierre hat außerdem festgestellt, dass das Alter der drei Personen positiv mit der TV-Nutzung (r = 0,50) und negativ mit dem Vertrauen korreliert (r = –0,50). Muss er befürchten, dass der in a) gefundene Zusammenhang (in der Stichprobe) nur eine Scheinkorrelation ist?

Übung 2a

a) Pierre setzt voraus, dass die Variablen linear zusammenhängen. Berechnen Sie ein geeignetes Maß, das einen möglichen Zusammenhang der Variablen in der Stichprobe ausdrückt. Was lässt sich anhand des Resultats über den Zusammenhang in der Stichprobe sagen?

→ geeignetes Maß: **(Produkt-Moment-)Korrelation** → **Pearson's *r***

- drückt **Stärke** (durch Höhe) und **Richtung** (pos./neg. Vorzeichen) des Zusammenhangs zweier Variablen aus
- geeignet für **lineare** Zusammenhänge **metrischer** Variablen

  → hier beides gegeben

1. Schritt: Kovarianz s_{XY} berechnen

wobei $s_{XY} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

Zur Erinnerung:
Die Kovarianz ist ein Maß für den linearen Zusammenhang zweier Variablen. Sie ist eng verwandt mit der Korrelation (aber: nicht standardisiert, daher nicht gut vergleichbar).

1. Schritt: Kovarianz s_{XY} berechnen

wobei $s_{XY} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

ID	TV-Nutzung (x)	Vertrauen (y)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1	5	– 1,0	0,0	0,0
2	2	6	0,0	1,0	0,0
3	3	4	1,0	– 1,0	– 1,0
	$\bar{x}$ = 2	$\bar{y}$ = 5			

1. Schritt: Kovarianz s_{XY} berechnen

wobei $s_{XY} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

ID	TV-Nutzung (x)	Vertrauen (y)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1	5	– 1,0	0,0	0,0
2	2	6	0,0	1,0	0,0
3	3	4	1,0	– 1,0	– 1,0
	$\bar{x}$ = 2	$\bar{y}$ = 5			

$\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$= – 1

1. Schritt: Kovarianz s_{XY} berechnen

wobei $s_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$

ID	TV-Nutzung (x)	Vertrauen (y)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1	5	– 1,0	0,0	0,0
2	2	6	0,0	1,0	0,0
3	3	4	1,0	– 1,0	– 1,0
	$\bar{x}$ = 2	$\bar{y}$ = 5			

$\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = -1$

$$\boldsymbol{s_{XY}} = \frac{1}{3-1} \cdot -1 = \mathbf{-0,50}$$

2. Schritt: Korrelationskoeffizient r_{XY} berechnen

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}$$

Zwischenschritt: Standardabweichungen s_x und s_y berechnen

$$s_X = \sqrt{s_X^2} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$s_X = \sqrt{s_X^2} = \sqrt{\frac{1}{3-1}((1-2)^2 + (2-2)^2 + (3-2)^2)} = 1$$

Entsprechend ebenso…

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{1}{3-1}((5-5)^2 + (6-5)^2 + (4-5)^2)} = 1$$

ID	TV-Nutzung (x)	Vertrauen (y)
1	1	5
2	2	6
3	3	4
	$\bar{x}$ = 2	$\bar{y}$ = 5

2. Schritt: Korrelationskoeffizient r_{XY} berechnen

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}$$

$$r_{XY} = \frac{-0{,}50}{1 \cdot 1}$$

$$\boldsymbol{r_{XY} = -0{,}50}$$

3. Schritt: Korrelationskoeffizient r_{XY} interpretieren

- Interpretation:
  - 1 = perfekt positiver linearer Zusammenhang
  - 0 = kein linearer Zusammenhang
  - -1 = perfekt negativer linearer Zusammenhang
- Konventionen:

r_{XY}	**Stärke des Zusammenhangs**
$0{,}10 \leq \lvert r \rvert < 0{,}30$	schwacher Zusammenhang
$0{,}30 \leq \lvert r \rvert < 0{,}50$	mittlerer Zusammenhang
$0{,}50 \leq \lvert r \rvert < 0{,}70$	starker Zusammenhang
$\lvert r \rvert \geq 0{,}70$	sehr starker Zusammenhang

Aufgabe 2a

a) Pierre setzt voraus, dass die Variablen linear zusammenhängen. Berechnen Sie ein geeignetes Maß, das einen möglichen Zusammenhang der Variablen in der Stichprobe ausdrückt.
Was lässt sich anhand des Resultats über den Zusammenhang in der Stichprobe sagen?

$$\boldsymbol{r = -0,50}$$

- **Vorzeichen** negativ → negativer Zusammenhang: Je höher der TV-Konsum, desto geringer das interpersonale Vertrauen bzw. je höher das interpersonale Vertrauen, desto geringer der TV-Konsum. Für die Stichprobe (!) stützen die Daten also Pierres Vermutung.
- **Höhe von *r***: per Konvention (knapp) starker Zusammenhang

Aufgabe 2b

b) Kann Pierre auch für die gesamte Familie einen Zusammenhang annehmen? (Signifikanzniveau 5%)

→ Frage nach **Verallgemeinerbarkeit auf Grundgesamtheit**

→ **Signifikanztest** für *r* notwendig

Signifikanztest für r_{XY}

- *Fragestellung:* Besteht für Stichprobe ermittelter Zusammenhang zwischen zwei Variablen auch in der Grundgesamtheit?
- *Formulierung der Hypothesen:*
  - **Zweiseitiges Problem:** Besteht Zusammenhang (egal welcher Richtung) oder sind beide Variablen unabhängig voneinander?

$$H_0: \rho_{XY} = 0$$

$$H_1: \rho_{XY} \neq 0$$

- *Prüfgröße:*

$$t = \frac{r_{XY} \cdot \sqrt{N-2}}{\sqrt{1-r_{XY}^2}} \quad \text{mit} \quad df = N - 2$$

Testentscheidung:
H_0 wird abgelehnt, wenn
$|t| > t_{krit}$

Aufgabe 2b

Signifikanztest für Korrelationskoeffizient

	einseitig		zweiseitig	
df	α = 5%	α = 1%	α = 5%	α = 1%
1	6,31	31,82	12,71	63,66

- Berechnung Prüfgröße

$$t = \frac{r_{XY} \cdot \sqrt{N-2}}{\sqrt{1-r_{XY}^2}} = \frac{-0{,}50 \cdot \sqrt{3-2}}{\sqrt{1-0{,}25}} = \frac{-0{,}50}{\sqrt{0{,}75}} = -0{,}58$$

$$df = N - 2 = 1$$

- kritischer t-Wert bei $\alpha = 0{,}05$ (zweiseitig) aus der Tabelle:
  - für $df = 1$: $t_{\text{krit}} = 12{,}71$
- Testentscheidung:
  - $|t| = 0{,}58 < t_{\text{krit}}$ → ***H_0* kann nicht abgelehnt werden**

Aufgabe 2b

b) Kann Pierre auch für die gesamte Familie einen Zusammenhang annehmen? (Signifikanzniveau 5%)

→ Nein, der festgestellte negative Zusammenhang kann **nicht** auf die Grundgesamtheit (Familie) verallgemeinert werden.

Aufgabe 2c

c) Pierre hat außerdem festgestellt, dass das Alter der drei Personen positiv mit der TV-Nutzung (r = 0,50) und negativ mit dem Vertrauen korreliert (r = – 0,50). Muss er befürchten, dass der in a) gefundene Zusammenhang (in der Stichprobe) nur eine Scheinkorrelation ist?

→ Prüfung auf Scheinkorrelation bzw. Kontrolle von Drittvariablen („gemeinsame Ursache“) mittels **Partialkorrelation**.

Aufgabe 2c

Berechnung partieller Korrelationskoeffizient

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{1 - {r^2}_{XZ}} \cdot \sqrt{1 - r_{YZ}^2}}$$

→ Beschreibt den Zusammenhang zwischen den Variablen X und Y, nachdem der Einfluss einer Drittvariablen Z kontrolliert wird (d.h. der Einfluss der Drittvariable wird „herausgerechnet")

Aufgabe 2c

Berechnung partieller Korrelationskoeffizient

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{1 - {r^2}_{XZ}} \cdot \sqrt{1 - r_{YZ}^2}}$$

$$r_{XY} = r_{TV \cdot Vertr} = -0{,}50$$
$$r_{XZ} = r_{TV \cdot Alter} = 0{,}50$$
$$r_{YZ} = r_{Vertr \cdot Alter} = -0{,}50$$

$$r_{TV \cdot Vertr|Alter} = \frac{-0{,}5 + 0{,}5 \cdot 0{,}5}{\sqrt{1 - 0{,}25} \cdot \sqrt{1 - 0{,}25}} = \frac{-0{,}25}{0{,}75} = -0{,}33$$

Aufgabe 2c

→Prüfung auf Scheinkorrelation bzw. Kontrolle von Drittvariablen („gemeinsame Ursache“) mittels **Partialkorrelation**.

→Ergebnis: $r_{TV \cdot Vertr|Alter} = -0{,}33$

→Durch Kontrolle des Einflusses des Alters **reduziert** sich der Zusammenhang zwischen TV-Nutzungsdauer und Vertrauen zwar auf einen knapp mittleren Zusammenhang, verschwindet aber **nicht** → <u>keine</u> (reine) Scheinkorrelation

4. Korrelation in R berechnen

Korrelationen berechnen mittels `tidycomm`

Sie können Korrelationen mit dem `tidycomm` Package berechnen. Dafür können Sie u. a. die folgenden Funktionen nutzen:

1. `correlate` [berechnet Korrelationen]
2. `to_correlation_matrix` [gibt die Ergebnisse zusätzlich als Korrelationsmatrix aus]

Struktur:

```
data %>%
  correlate(variable)


data %>%
  correlate(variable) %>%
  to_correlation_matrix(variable)
```

Übung 1 in R umsetzen

Als erstes legen wir den Datensatz in R an:

```
# Anlegen des Datensatzes
tv_nutzung <- c(1, 2, 3)
vertrauen <- c(5, 6, 4)
befragten_id <- c(1:3)
data <- cbind(befragten_id, tv_nutzung, vertrauen)
data <- tidyr::as_tibble(data)
```

Übung 1 in R umsetzen

Kurze R-Auffrischung

Als erstes legen wir den Datensatz in R an:

```r
# Anlegen des Datensatzes
tv_nutzung <- c(1, 2, 3)
vertrauen <- c(5, 6, 4)
befragten_id <- c(1:3)
data <- cbind(befragten_id, tv_nutzung, vertrauen)
data <- tidyr::as_tibble(data)
```

Funktionen sind vorgegeben und erhalten immer Klammern dahinter. Achten Sie darauf, alle geöffneten Klammern auch wieder zu schließen.

Parameter spezifizieren Funktionsaufrufe, stehen in deren Klammern und werden durch Kommata voneinander getrennt.

Pfeile weisen **Objekte** zu. Die links vom Pfeil stehenden Objektnamen sind essentiell, denn darunter ist fortan das Objekt verwendbar.

Übung 1 in R umsetzen

Anschliessend berechnen wir die Korrelationen & lassen uns die Korrelationsmatrix ausgeben:

```
#Berechnung von Korrelationen in R
data %>%
  correlate(tv_nutzung, vertrauen)
```

```
# A tibble: 1 × 6
  x          y              r    df     p     n
* <chr>      <chr>      <dbl> <int> <dbl> <int>
1 tv_nutzung vertrauen   -0.5     1 0.667     3
```

```
#Ausgabe als Korrelationsmatrix
data %>%
  correlate(tv_nutzung, vertrauen) %>%
  to_correlation_matrix()
```

```
# A tibble: 2 × 3
  r          tv_nutzung vertrauen
* <chr>           <dbl>     <dbl>
1 tv_nutzung        1        -0.5
2 vertrauen        -0.5       1
```

Sonderfrage: Warum ist die Diagonale „1“?

Übung 1 in R umsetzen

Kurze R-Auffrischung

Anschliessend berechnen wir die Korrelationen & lassen uns die Korrelationsmatrix ausgeben:

```
#Berechnung von Korrelationen in R
data %>%
  correlate(tv_nutzung, vertrauen)
```

```
# A tibble: 1 × 6
  x          y              r    df     p     n
  <chr>      <chr>      <dbl> <int> <dbl> <int>
1 tv_nutzung vertrauen   -0.5     1 0.667     3
```

Eine **Ausgabe** erhalten Sie nur dann, wenn Sie das Ergebnis einer Funktion *nicht* einem Objekt zuweisen (also kein <-).

```
#Ausgabe als Korrelationsmatrix
data %>%
  correlate(tv_nutzung, vertrauen) %>%
  to_correlation_matrix()
```

```
# A tibble: 2 × 3
  r          tv_nutzung vertrauen
  <chr>           <dbl>     <dbl>
1 tv_nutzung        1        -0.5
2 vertrauen        -0.5       1
```

Das **Pipe-Symbol** nimmt das Objekt/Ergebnis links davon und macht damit das Folgende (rechts davon bzw. nächste Zeile). Achten Sie darauf, eine Pipe-Befehlskette abzuschließen (also nicht mit einem Pipe-Symbol zu enden).

Wichtige Take-Aways

- **correlate()**: berechnet die Korrelation zwischen zwei Variablen
- **to_correlation_matrix()**: gibt zusätzlich die Korrelationsmatrix heraus

5. Übung 3

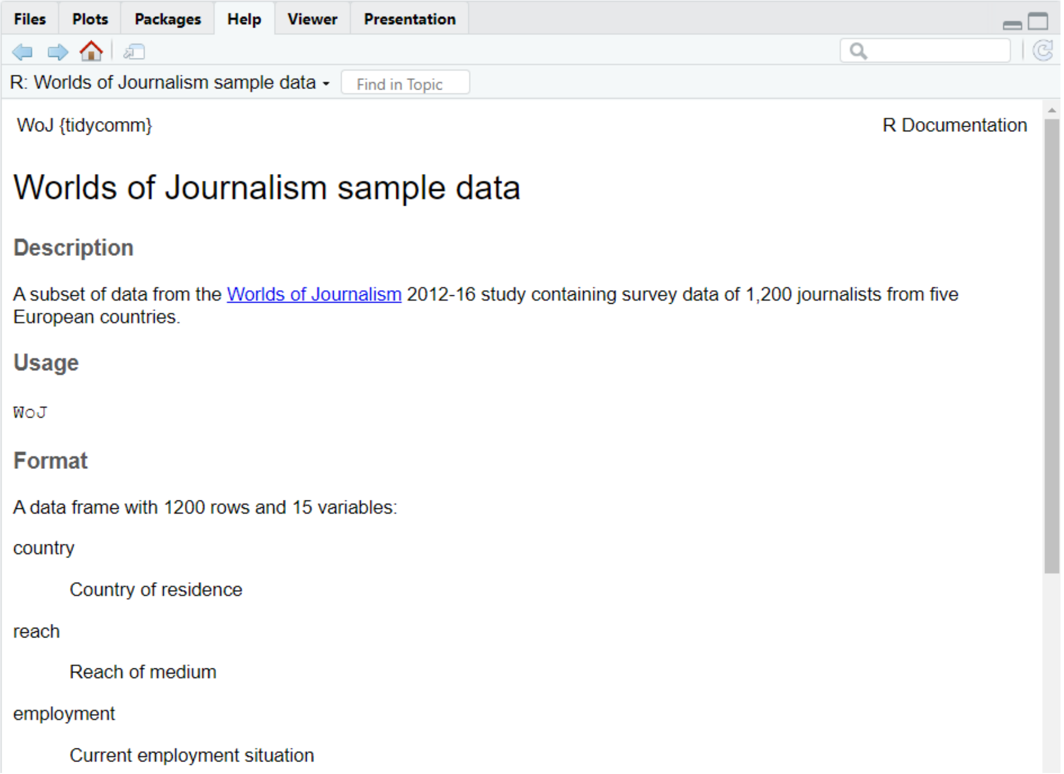
Korrelationen in R berechnen

Übung 3

Laden Sie nun erneut den Datensatz „WoJ“ vom tidycomm-Package in Ihr RStudio-Environment.

```
#Abspeichern des World of Journalism Datensatzes
data <- WoJ
```

Erinnerung: Hilfeseite zum WoJ-Datensatz (`?WoJ`)



Übung 3

Journalist*innen im WoJ-Datensatz wurden gefragt, …

- wie lange sie bereits in der Branche arbeiten («work_experience»),
- inwiefern sie denken, dass Journalist*innen immer ethischen Richtlinien folgen sollten («ethics_1») sowie
- wie sehr sie der Regierung trauen («trust_government»).

Beantworten Sie mittels `tidycomm`:

a) Korrelieren journalistische Arbeitserfahrung, Zustimmung zu ethischen Richtlinien und das Vertrauen in die Regierung.

b) Können Sie diese Korrelationen mittels eines Streudiagramms visualisieren?*

c) Wie berechnen Sie die partielle Korrelation zwischen Zustimmung zu ethischen Richtlinien und Vertrauen in die Regierung, wenn Sie für Arbeitserfahrung kontrollieren wollen?*

*für b) und c) müssen Sie ggf. via help() oder dem R-Kompendium die Lösung recherchieren

Lösung für Übung 3 in R umsetzen mittels `tidycomm`

Aufgabe 3a): Ausgeben der Korrelationen

Sie können `correlate()` aus dem `tidycomm`-Package anwenden:

Der Code:

```
#Abspeichern des World of Journalism Datensatzes
WoJ <- WoJ

#Berechnung von Korrelationen in R
data %>%
  correlate(ethics_1, trust_government, work_experience)
```

Das Ergebnis:

```
# A tibble: 3 × 6
  x                y                      r    df        p     n
* <chr>            <chr>              <dbl> <int>    <dbl> <int>
1 ethics_1         trust_government -0.102   1198 0.000421  1200
2 ethics_1         work_experience  -0.103   1185 0.000387  1187
3 trust_government work_experience  -0.0708  1185 0.0146    1187
```

Lösung für Übung 3 in R umsetzen mittels `tidycomm`

Aufgabe 3a): Ausgeben der Korrelationen

Sie können `correlate()` aus dem `tidycomm`-Package anwenden:

Der Code:

```
#Abspeichern des World of Journalism Datensatzes
WoJ <- WoJ

#Berechnung von Korrelationen in R
data %>%
  correlate(ethics_1, trust_government, work_experience)
```

Das Ergebnis:

```
# A tibble: 3 x 6
  x                y                      r    df        p     n
* <chr>            <chr>              <dbl> <int>    <dbl> <int>
1 ethics_1         trust_government -0.102   1198 0.000421  1200
2 ethics_1         work_experience  -0.103   1185 0.000387  1187
3 trust_government work_experience  -0.0708  1185 0.0146    1187
```

<u>Beispiel</u>: Es besteht eine signifikante Korrelation zwischen der Zustimmung zu ethischen Richtlinien und dem Vertrauen in die Regierung (r = -.102; p < .001).

Lösung für Übung 3 in R umsetzen mittels `tidycomm`

Aufgabe 3b): Visualisierung der Korrelationen

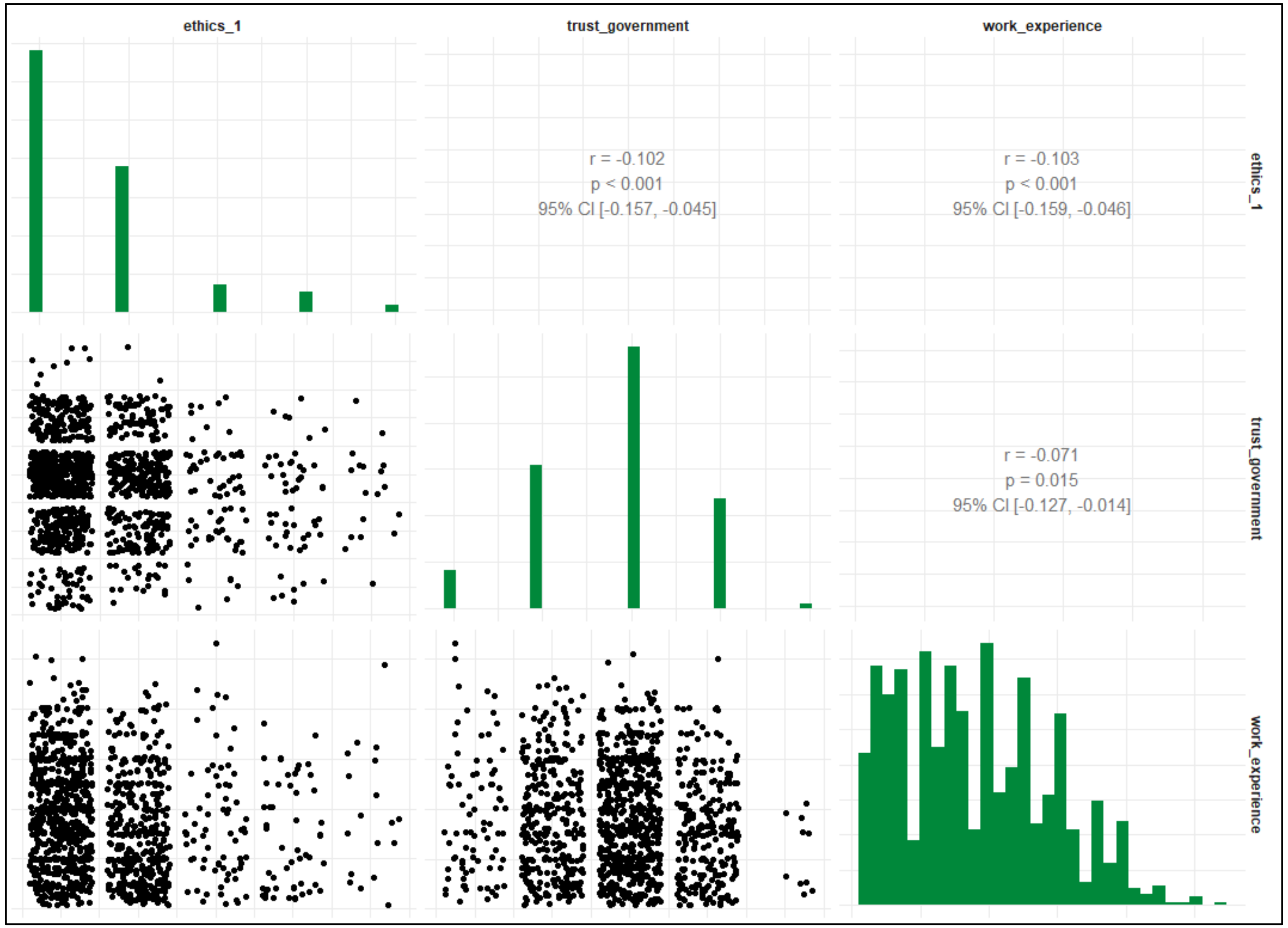
Sie können `correlate()` sowie `visualize()` aus dem `tidycomm`-Package anwenden:

Der Code:

```
#Visualisierung von Korrelationen
WoJ %>%
  correlate(ethics_1,  trust_government, work_experience) %>%
  visualize()
```

Das Ergebnis: ….

Lösung für Übung 3 in R umsetzen mittels `tidycomm`



Lösung für Übung 3 in R umsetzen mittels tidycomm

Aufgabe 3c): Partialkorrelation

Sie können `correlate()` aus dem `tidycomm`-Package anwenden und das argument `partial` nutzen:

Der Code:

```
#Partieller Korrelationskoeffizient- Kontrolle für "Work Experience"
WoJ %>%
  correlate(ethics_1, trust_government, partial = work_experience)
```

Das Ergebnis:

```
  x        y                z                     r    df        p     n
* <chr>    <chr>            <chr>             <dbl> <dbl>    <dbl> <int>
1 ethics_1 trust_government work_experience -0.111  1184 0.000124  1187
```

Interpretation: Der Zusammenhang ist weiterhin signifikant und bleibt ähnlich (in Richtung wie Größe).

Danke für Ihre Aufmerksamkeit!