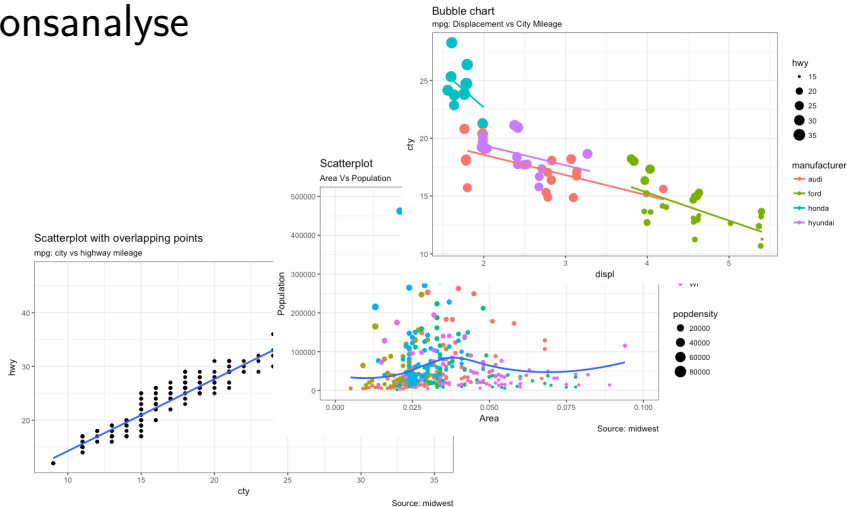


Data Science für Geistes- und Sozialwissenschaftler

Thema: Regressionsanalyse

Martin Kerntopf
25. Januar 2022



Zielsetzung:

- Wiederholung:
 - Einlesen von Datensätzen
 - Arbeiten mit R/RStudio
 - Skalenniveaus, Variablenarten, Datentypen
- Regressionen:
 - Grundlagen
 - Formen
 - Anwendung
 - Auswertung
 - Darstellung

Zielsetzung:

- Wiederholung:
 - Einlesen von Datensätzen
 - Arbeiten mit R/RStudio
 - Skalenniveaus, Variablenarten, Datentypen
- Regressionen:
 - Grundlagen
 - Formen
 - Anwendung
 - Auswertung
 - Darstellung

Struktur:

- ① Beispiel
- ② Theorie
- ③ Praxis
- ④ Take Home Message

- Daten:
→ <https://github.com/MKernt/Lehrprobe/tree/master/Lehrprobe>

Beispiel

Forschungsfrage:

- Welchen Einfluss haben Bildung und Frauenrechte auf Geburtenraten?

Forschungsfrage:

- Welchen Einfluss haben Bildung und Frauenrechte auf Geburtenraten?

Shireen J. Jejeebhoy (1995). *Women's Education, Autonomy, and Reproductive Behaviour: Experience from Developing Countries*. International Studies in Demography. Oxford, New York: Oxford University Press

Vera Sagalova et al. (2021). "Long-Term Consequences of Early Marriage and Maternity in West and Central Africa: Wealth, Education, and Fertility". In: *Journal of Global Health* 11

Goleen Samari (2019). "Education and Fertility in Egypt: Mediation by Women's Empowerment". In: *SSM - Population Health* 9

Katarina R. I. Keller (2018). "The Economic Effects of Girls' and Young Women's Education in Development". In: *SSRN*

Forschungsfrage:

- Welchen Einfluss haben Bildung und Frauenrechte auf Geburtenraten?

Shireen J. Jejeebhoy (1995). *Women's Education, Autonomy, and Reproductive Behaviour: Experience from Developing Countries*. International Studies in Demography. Oxford, New York: Oxford University Press

Vera Sagalova et al. (2021). "Long-Term Consequences of Early Marriage and Maternity in West and Central Africa: Wealth, Education, and Fertility". In: *Journal of Global Health* 11

Goleen Samari (2019). "Education and Fertility in Egypt: Mediation by Women's Empowerment". In: *SSM - Population Health* 9

Katarina R. I. Keller (2018). "The Economic Effects of Girls' and Young Women's Education in Development". In: *SSRN*

Hypothesen:

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

H2: Je weniger Frauenrechte in einem Land, desto höher die Geburtenrate

Forschungsfrage:

- Welchen Einfluss haben Bildung und Frauenrechte auf Geburtenraten?

Shireen J. Jejeebhoy (1995). *Women's Education, Autonomy, and Reproductive Behaviour: Experience from Developing Countries*. International Studies in Demography. Oxford, New York: Oxford University Press

Vera Sagalova et al. (2021). "Long-Term Consequences of Early Marriage and Maternity in West and Central Africa: Wealth, Education, and Fertility". In: *Journal of Global Health* 11

Goleen Samari (2019). "Education and Fertility in Egypt: Mediation by Women's Empowerment". In: *SSM - Population Health* 9

Katarina R. I. Keller (2018). "The Economic Effects of Girls' and Young Women's Education in Development". In: *SSRN*

Hypothesen:

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

H2: Je weniger Frauenrechte in einem Land, desto höher die Geburtenrate

→ **Korrelation:** wie eng der lineare statistische Zusammenhang zwischen zwei Größen ist

Forschungsfrage:

- Welchen Einfluss haben Bildung und Frauenrechte auf Geburtenraten?

Shireen J. Jejeebhoy (1995). *Women's Education, Autonomy, and Reproductive Behaviour: Experience from Developing Countries*. International Studies in Demography. Oxford, New York: Oxford University Press

Vera Sagalova et al. (2021). "Long-Term Consequences of Early Marriage and Maternity in West and Central Africa: Wealth, Education, and Fertility". In: *Journal of Global Health* 11

Goleen Samari (2019). "Education and Fertility in Egypt: Mediation by Women's Empowerment". In: *SSM - Population Health* 9

Katarina R. I. Keller (2018). "The Economic Effects of Girls' and Young Women's Education in Development". In: *SSRN*

Hypothesen:

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

H2: Je weniger Frauenrechte in einem Land, desto höher die Geburtenrate

Achtung!

Korrelation \neq Kausalbeziehung

Datensatz

- Basierend auf:
 - UN Daten
 - <http://data.un.org/>
 - CIRI Human Rights Dataset
 - <http://www.humanrightsdata.com/p/data-documentation.html>

Datensatz

- Basierend auf:

- UN Daten

→ <http://data.un.org/>

- CIRI Human Rights Dataset

→ [http:](http://www.humanrightsdata.com/p/data-documentation.html)

[//www.humanrightsdata.com/p/
data-documentation.html](http://www.humanrightsdata.com/p/data-documentation.html)

	Land	Geburtenrate_2012	Bildung_2012	Frauenrechte_2012
1	Afghanistan	5.141	9	2
2	Albania	1.760	9	2
3	Algeria	2.820	11	2
4	Andorra	5.979	6	3
5	Angola	2.102	11	3
6	Antigua and Barbuda	2.188	13	3

- Land
- Durchschnittliche Geburtenrate
- Durchschnittliche Bildungsdauer (in Jahren)
- Frauenrechte:
 - Kategorische Variable
 - 0: keine Rechte im Gesetz
 - 3: (alle) Rechte im Gesetz und in der Praxis

Datensatz

- Basierend auf:

- UN Daten

→ <http://data.un.org/>

- CIRI Human Rights Dataset

→ [http:](http://www.humanrightsdata.com/p/data-documentation.html)

[//www.humanrightsdata.com/p/
data-documentation.html](http://www.humanrightsdata.com/p/data-documentation.html)

	Land	Geburtenrate_2012	Bildung_2012	Frauenrechte_2012
1	Afghanistan	5.141	9	2
2	Albania	1.760	9	2
3	Algeria	2.820	11	2
4	Andorra	5.979	6	3
5	Angola	2.102	11	3
6	Antigua and Barbuda	2.188	13	3

- Land
- Durchschnittliche Geburtenrate
- Durchschnittliche Bildungsdauer (in Jahren)
- Frauenrechte:
 - Kategorische Variable
 - 0: keine Rechte im Gesetz
 - 3: (alle) Rechte im Gesetz und in der Praxis

→ **Querschnittsdaten**

Theorie

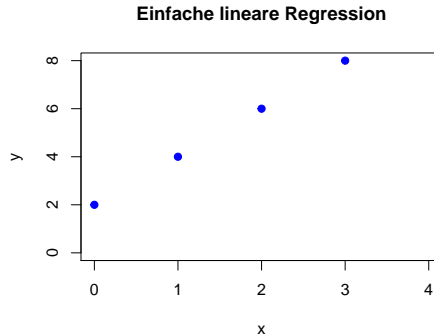
Die Regressionsanalyse

- Einfluss einer oder mehrerer unabhängiger Variablen auf (**x**)
eine abhängige Variable (**y**)
herausfinden
→ Regression von Y auf X

Die Regressionsanalyse

- Einfluss einer oder mehrerer unabhängiger Variablen (x) auf **eine** abhängige Variable (y) herausfinden

→ Regression von Y auf X

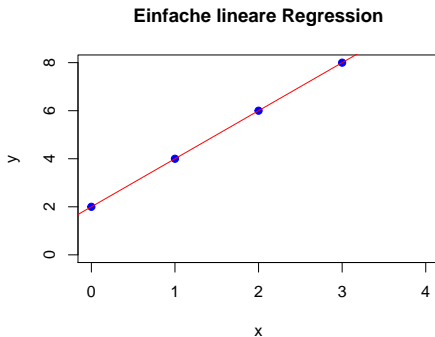


x_i	0	1	2	3
y_i	2	4	6	8

Die Regressionsanalyse

- Einfluss einer oder mehrerer unabhängiger Variablen (x) auf **eine** abhängige Variable (y) herausfinden

→ Regression von Y auf X



x_i	0	1	2	3
y_i	2	4	6	8

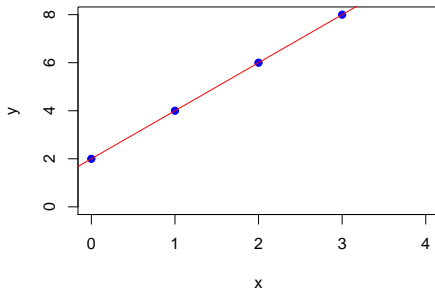
Die Regressionsanalyse

- Einfluss einer oder mehrerer unabhängiger Variablen (**x**) auf **eine** abhängige Variable (**y**) herausfinden

→ Regression von Y auf X

- Formel: $y = a + bx$
 - a = Schnittpunkt y-Achse (intercept)
 - b = Steigung (slope)

Einfache lineare Regression



x_i	0	1	2	3
y_i	2	4	6	8

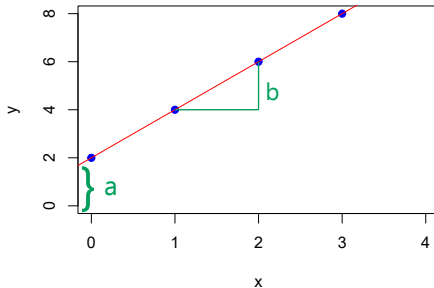
Die Regressionsanalyse

- Einfluss einer oder mehrerer unabhängiger Variablen (**x**) auf **eine** abhängige Variable (**y**) herausfinden

→ Regression von Y auf X

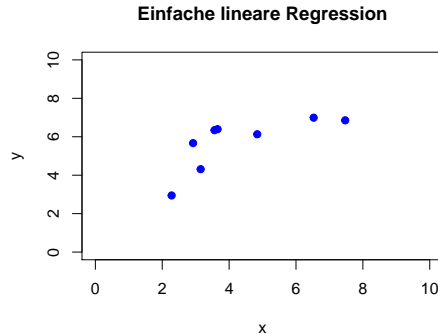
- Formel: $y = a + bx$
 - a = Schnittpunkt y-Achse (intercept)
 - b = Steigung (slope)

Einfache lineare Regression

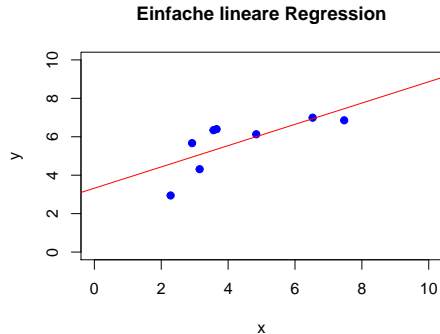


x_i	0	1	2	3
y_i	2	4	6	8

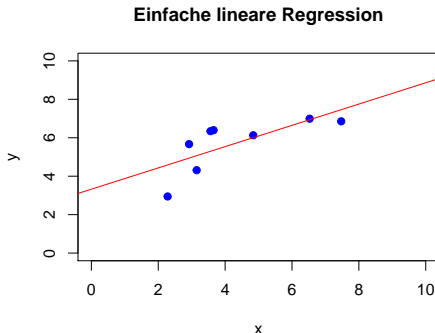
- Zusammenhang fast nie perfekt



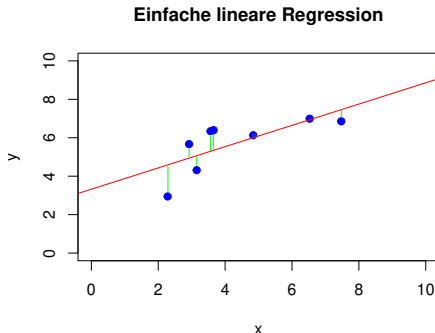
- Zusammenhang fast nie perfekt



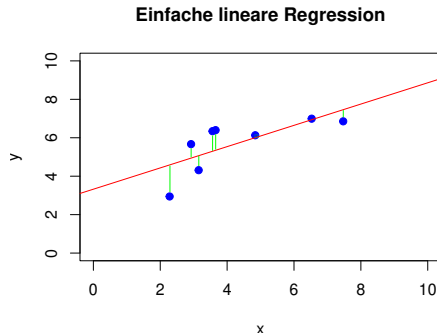
- Zusammenhang fast nie perfekt
- Hinzufügen eines Fehlerterms
→ $y = a + b * x + \epsilon$
- ϵ = 'Rauschen'
→ beinhaltet unbeobachtete Faktoren, die sich auf die abhängige Variable auswirken



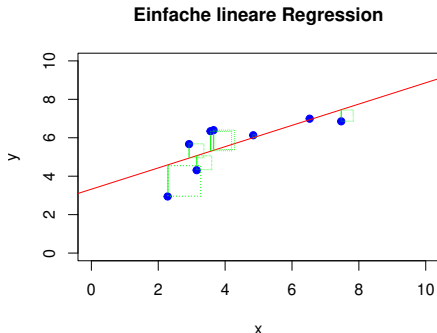
- Zusammenhang fast nie perfekt
- Hinzufügen eines Fehlerterms
→ $y = a + b * x + \epsilon$
- ϵ = 'Rauschen'
→ beinhaltet unbeobachtete Faktoren, die sich auf die abhängige Variable auswirken



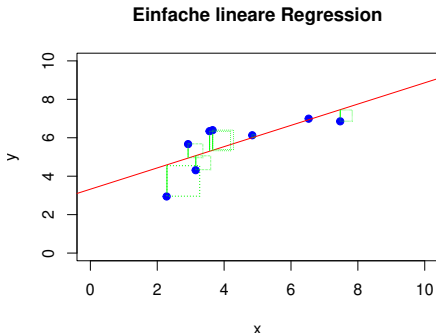
- Residuen sind die **Abweichungen** der **geschätzten y-Werte** zu den **tatsächlichen y-Werten**
- Ziel ist es die Residuen möglichst klein zu halten
→ Verfahren: Ordinary Least Squares (OLS)



- Residuen sind die **Abweichungen** der **geschätzten y-Werte** zu den **tatsächlichen y-Werten**
- Ziel ist es die Residuen möglichst klein zu halten
→ Verfahren: Ordinary Least Squares (OLS)



- Residuen sind die **Abweichungen** der **geschätzten y-Werte** zu den **tatsächlichen y-Werten**
- Ziel ist es die Residuen möglichst klein zu halten
→ Verfahren: Ordinary Least Squares (OLS)



$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$a = \bar{y} - b * \bar{x}$$

\bar{x} und \bar{y} = Mittelwerte aller x und y

- ① **Lineare Regression**
- ② Multiple lineare Regression
- ③ Logistische Regression
- ④ ...

$$y = a + b * x + \epsilon$$

Merkmale:

- Linearität zwischen x und Mittelwert von y
- Kontinuierliche abhängige Variable

Annahmen:

- Der Erwartungswert der Residuen ist Null
- Die Fehlerterme sind unkorreliert
- Es herrscht Homoskedastizität
- Die Fehlerterme sind normalverteilt

- ① Lineare Regression
- ② **Multiple lineare Regression**
- ③ Logistische Regression
- ④ ...

$$y_i = a + b_1 * x_{i1} + b_2 * x_{i2} + \dots + \epsilon_i,$$
$$i = 1, \dots, n$$

Merkmale:

- Siehe einfache lineare Regression
- Mehrere unabhängige Variablen
- Unabhängige Variablen müssen nicht kontinuierlich sein

- ① Lineare Regression
- ② **Multiple lineare Regression**
- ③ Logistische Regression
- ④ ...

$$y_i = a + b_1 * x_{i1} + b_2 * x_{i2} + \dots + \epsilon_i,$$
$$i = 1, \dots, n$$

Merkmale:

- Siehe einfache lineare Regression
- Mehrere unabhängige Variablen
- Unabhängige Variablen müssen nicht kontinuierlich sein

Achtung!

Multiple \neq Multivariate

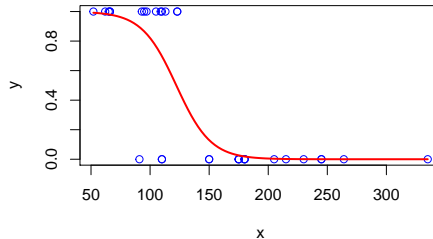
- ① Lineare Regression
- ② Multiple Regression
- ③ **Logistische Regression**
- ④ ...

$$\pi_i = \frac{1}{1+e^{x_i\beta}}$$

Merkmale:

- Dichotome abhängige Variable
→ Mann/Frau, ja/nein, etc.

Logistische Regression



- ① Lineare Regression
 - ② Multiple Regression
 - ③ Logistische Regression
 - ④ ...
- Ordinale Regression
 - AV mit mehr als zwei Kategorien (bspw. Antwortmöglichkeiten bei Umfragen)
 - Poisson Regression
 - AV mit Zählvariablen (bspw. Anzahl Anrufe bei einer Notfallnummer)
 - ...

Praxis

Vorbereitung:

- ① Datensatz herunterladen
- ② Jupyter Hub oder lokales RStudio öffnen

→ <https://jupyterhub.wolke.uni-greifswald.de/>

- ③ Datensatz einlesen
 - via GUI, oder
 - direkt im Script

```
world_data <- read.csv("~/Pfad/world_data.csv")
```


Vorbereitung:

- ① Datensatz herunterladen
- ② Jupyter Hub oder lokales RStudio öffnen

→ <https://jupyterhub.wolke.uni-greifswald.de/>

- ③ Datensatz einlesen
 - via GUI, oder
 - direkt im Script

```
world_data <- read.csv("~/Pfad/world_data.csv")
```

Vorgehen:

- Erstes Sichten der Daten

Vorbereitung:

- 1 Datensatz herunterladen
- 2 Jupyter Hub oder lokales RStudio öffnen

→ <https://jupyterhub.wolke.uni-greifswald.de/>

- 3 Datensatz einlesen
 - via GUI, oder
 - direkt im Script

```
world_data <- read.csv("~/Pfad/world_data.csv")
```

Vorgehen:

- Erstes Sichten der Daten

```
head(world_data)  
str(world_data)  
summary(world_data)
```

Vorbereitung:

- 1 Datensatz herunterladen
- 2 Jupyter Hub oder lokales RStudio öffnen

→ <https://jupyterhub.wolke.uni-greifswald.de/>

- 3 Datensatz einlesen
 - via GUI, oder
 - direkt im Script

```
world_data <- read.csv("~/Pfad/world_data.csv")
```

Vorgehen:

- Erstes Sichten der Daten

```
head(world_data)
str(world_data)
summary(world_data)
```

- Daten formatieren

Vorbereitung:

- 1 Datensatz herunterladen
- 2 Jupyter Hub oder lokales RStudio öffnen

→ <https://jupyterhub.wolke.uni-greifswald.de/>

- 3 Datensatz einlesen
 - via GUI, oder
 - direkt im Script

```
world_data <- read.csv("~/Pfad/world_data.csv")
```

Vorgehen:

- Erstes Sichten der Daten

```
head(world_data)
str(world_data)
summary(world_data)
```

- Daten formatieren

```
world_data$Frauenrechte_2012 <- as.factor
(world_data$Frauenrechte_2012)
```

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
model_H1 <- lm(Geburtenrate_2012 ~ Bildung_2012,  
  data=world_data)  
summary(model_H1)
```

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
model_H1 <- lm(Geburtenrate_2012 ~ Bildung_2012,
  data=world_data)
summary(model_H1)
```

Output:

```
Call:
lm(formula = Geburtenrate_2012 ~ Bildung_2012, data = world_data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8819 -1.1009 -0.5064  1.1004  4.6031

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.34393    0.26128  12.798  <2e-16 ***
Bildung_2012 -0.05329    0.02866  -1.859   0.0646 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 180 degrees of freedom
Multiple R-squared:  0.01884,    Adjusted R-squared:  0.01339
F-statistic: 3.457 on 1 and 180 DF,  p-value: 0.06461
```

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
model_H1 <- lm(Geburtenrate_2012 ~ Bildung_2012,
  data=world_data)
summary(model_H1)
```

Output:

```
Call:
lm(formula = Geburtenrate_2012 ~ Bildung_2012, data = world_data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8819 -1.1009 -0.5064  1.1004  4.6031

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.34393    0.26128  12.798  <2e-16 ***
Bildung_2012 -0.05329    0.02866  -1.859   0.0646 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 180 degrees of freedom
Multiple R-squared:  0.01884,    Adjusted R-squared:  0.01339
F-statistic: 3.457 on 1 and 180 DF,  p-value: 0.06461
```

Interpretation:

- Intercept = 3.34 → bei 0 Jahren Schulzeit durchschnittlich 3.34 Kinder

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
model_H1 <- lm(Geburtenrate_2012 ~ Bildung_2012,
  data=world_data)
summary(model_H1)
```

Output:

```
Call:
lm(formula = Geburtenrate_2012 ~ Bildung_2012, data = world_data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8819 -1.1009 -0.5064  1.1004  4.6031

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.34393    0.26128  12.798  <2e-16 ***
Bildung_2012 -0.05329    0.02866  -1.859   0.0646 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 180 degrees of freedom
Multiple R-squared:  0.01884,    Adjusted R-squared:  0.01339
F-statistic: 3.457 on 1 and 180 DF,  p-value: 0.06461
```

Interpretation:

- Intercept = 3.34 → bei 0 Jahren Schulzeit durchschnittlich 3.34 Kinder
- Bildung 2012 = -0.05 → pro Einheit Schulbildung nimmt durchschnittlich Anzahl an Kindern um 0.05 ab (p-Wert > 0.05 → nicht signifikant)

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
model_H1 <- lm(Geburtenrate_2012 ~ Bildung_2012,
  data=world_data)
summary(model_H1)
```

Output:

```
Call:
lm(formula = Geburtenrate_2012 ~ Bildung_2012, data = world_data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8819 -1.1009 -0.5064  1.1004  4.6031

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.34393    0.26128   12.798  <2e-16 ***
Bildung_2012  -0.05329    0.02866   -1.859   0.0646 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 180 degrees of freedom
Multiple R-squared:  0.01884,    Adjusted R-squared:  0.01339
F-statistic: 3.457 on 1 and 180 DF,  p-value: 0.06461
```

Interpretation:

- Intercept = 3.34 → bei 0 Jahren Schulzeit durchschnittlich 3.34 Kinder
- Bildung 2012 = -0.05 → pro Einheit Schulbildung nimmt durchschnittlich Anzahl an Kindern um 0.05 ab (p-Wert > 0.05 → nicht signifikant)
- $R^2 = 0.018$ → lediglich 1.8% der Varianz der Daten werden durch das Modell erklärt

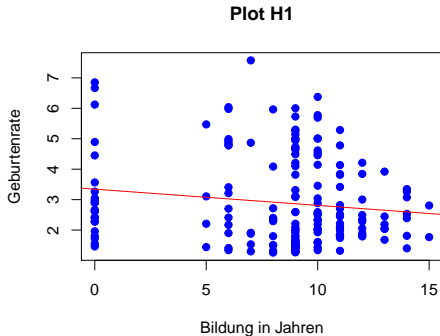
H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
plot(y = world_data$Geburtenrate_2012,  
     x = world_data$Bildung_2012,  
     pch = 19,  
     col = "blue",  
     main= "Plot H1",  
     xlab = "Bildung in Jahren",  
     ylab = "Geburtenrate")  
  
abline(lm(world_data$Geburtenrate_2012 ~  
          world_data$Bildung_2012),  
       col = "red")
```

H1: Je länger die obligatorische Bildung in einem Land dauert, desto niedriger ist die Geburtenrate

```
plot(y = world_data$Geburtenrate_2012,  
     x = world_data$Bildung_2012,  
     pch = 19,  
     col = "blue",  
     main= "Plot H1",  
     xlab = "Bildung in Jahren",  
     ylab = "Geburtenrate")  
  
abline(lm(world_data$Geburtenrate_2012 ~  
          world_data$Bildung_2012),  
       col = "red")
```

Output:



Zeit: 10 Minuten

H2: Je weniger Frauenrechte in einem Land, desto höher die Geburtenrate

- Führen Sie folgende Schritte aus:
 - Erstellung Ergebnistabelle
 - Interpretation der Ergebnisse
 - Plotten der Regression

H2: Je weniger Frauenrechte in einem Land, desto höher die Geburtenrate

```
model_H2 = lm(Geburtenrate_2012 ~
  Frauenrechte_2012,
  data=world_data)
summary(model_H2)

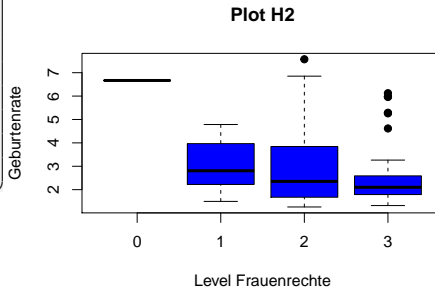
plot(y = world_data$Geburtenrate_2012,
  x = world_data$Frauenrechte_2012,
  pch = 19,
  col = "blue",
  main= "Plot H1",
  xlab = "Level Frauenrechte",
  ylab = "Geburtenrate")
```

```
Call:
lm(formula = Geburtenrate_2012 ~ Frauenrechte_2012, data = world_data)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.6186 -1.0932 -0.4893  0.8874  4.6954
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.667      1.439   4.634 6.92e-06 ***
Frauenrechte_20121 -3.557      1.476  -2.410  0.01699 *
Frauenrechte_20122 -3.788      1.444  -2.623  0.00947 **
Frauenrechte_20123 -3.930      1.463  -2.685  0.00793 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.439 on 178 degrees of freedom
Multiple R-squared:  0.04126,    Adjusted R-squared:  0.0251
F-statistic: 2.553 on 3 and 178 DF,  p-value: 0.05703
```



Was, wenn beide unabhängige
Variablen in das Modell aufgenommen
werden?

Was, wenn beide unabhängigen
Variablen in das Modell
aufgenommen werden?

```
model_multi <- lm(Geburtenrate_2012 ~  
  Bildung_2012 +  
  Frauenrechte_2012,  
  data = world_data)
```

Was, wenn beide unabhängigen Variablen in das Modell aufgenommen werden?

```
model_multi <- lm(Geburtenrate_2012 ~  
  Bildung_2012 +  
  Frauenrechte_2012,  
  data = world_data)
```

```
Call:  
lm(formula = Geburtenrate_2012 ~ Bildung_2012 + Frauenrechte_2012,  
    data = world_data)
```

```
Residuals:  
    Min       1Q   Median       3Q      Max   
-1.7675 -1.0892 -0.4935  0.9359  4.6375
```

```
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)        
(Intercept)   6.66700    1.43437   4.648 6.53e-06 ***  
Bildung_2012   -0.04186    0.02880  -1.454  0.1478      
Frauenrechte_20121 -3.21351    1.49051  -2.156  0.0324 *     
Frauenrechte_20122 -3.43749    1.45986  -2.355  0.0196 *     
Frauenrechte_20123 -3.58075    1.47855  -2.422  0.0165 *     
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.434 on 177 degrees of freedom  
Multiple R-squared:  0.05257,    Adjusted R-squared:  0.03116   
F-statistic: 2.455 on 4 and 177 DF,  p-value: 0.0475
```


Was, wenn beide unabhängigen Variablen in das Modell aufgenommen werden?

```
library(relaimpo) # (relative importance)
calc.relimp(model_multi, type="lmg", rela=T, rank=T)
```

```
model_multi <- lm(Geburtenrate_2012 ~
  Bildung_2012 +
  Frauenrechte_2012,
  data = world_data)
```

```
Call:
lm(formula = Geburtenrate_2012 ~ Bildung_2012 + Frauenrechte_2012,
    data = world_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.7675	-1.0892	-0.4935	0.9359	4.6375

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.66700	1.43437	4.648	6.53e-06 ***
Bildung_2012	-0.04186	0.02880	-1.454	0.1478
Frauenrechte_2012	-3.21351	1.49051	-2.156	0.0324 *
Frauenrechte_2013	-3.43749	1.45986	-2.355	0.0196 *
Frauenrechte_2014	-3.58075	1.47855	-2.422	0.0165 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.434 on 177 degrees of freedom
Multiple R-squared: 0.05257, Adjusted R-squared: 0.03116
F-statistic: 2.455 on 4 and 177 DF, p-value: 0.0475

Was, wenn beide unabhängigen Variablen in das Modell aufgenommen werden?

```
model_multi <- lm(Geburtenrate_2012 ~
  Bildung_2012 +
  Frauenrechte_2012,
  data = world_data)
```

```
Call:
lm(formula = Geburtenrate_2012 ~ Bildung_2012 + Frauenrechte_2012,
    data = world_data)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.7675 -1.0892 -0.4935  0.9359  4.6375
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.66700    1.43437   4.648 6.53e-06 ***
Bildung_2012   -0.04186    0.02880  -1.454  0.1478
Frauenrechte_20121 -3.21351    1.49051  -2.156  0.0324 *
Frauenrechte_20122 -3.43749    1.45986  -2.355  0.0196 *
Frauenrechte_20123 -3.58075    1.47855  -2.422  0.0165 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.434 on 177 degrees of freedom
Multiple R-squared:  0.05257,    Adjusted R-squared:  0.03116
F-statistic: 2.455 on 4 and 177 DF,  p-value: 0.0475
```

```
library(relaimpo) # (relative importance)
calc.relimp(model_multi, type="lmg", rela=T, rank=T)
```

Proportion of variance explained by model: 5.26%
Metrics are normalized to sum to 100% (rela=TRUE).

Relative importance metrics:

```

                                lmg
Frauenrechte_2012 0.7131772
Bildung_2012      0.2868228
```

Take Home Message

Zusammenfassung:

- Regressionen als Mittel um Zusammenhänge zwischen einer oder mehrerer Variablen zu überprüfen
- Wahl der Regressionsart abhängig von:
 - Art der Variablen
 - Verteilung der Standardfehler
 - Heteroskedastizität
 - Multikollinearität

Achtung!

Korrelation \neq Kausalbeziehung

Weiterführende Literatur:

- Jürgen Hedderich and Lothar Sachs (2016). *Angewandte Statistik: Methodensammlung mit R*. Heidelberg: Springer
- Christof Wolf and Henning Best, eds. (2010). *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften
- Giuseppe Ciaburro (2018). *Regression Analysis with R*. Birmingham: Packt Publishing
- Iain Pardoe (2020). *Applied Regression Modeling*. John Wiley & Sons