9. Vorlesung Statistik II
Effektgrößen und Stichprobenumfangsplanung in der ELR und MLR

Wdh.: Beschreibung linearer Zusammenhänge

- In Statistik I (Vorlesung 3) hatten wir uns im Rahmen der Korrelation bereits mit der Beschreibung von linearen Zusammenhängen beschäftigt. Dieselben Aspekte spielen bei der Wahl sinnvoller Effektgrößen in der ELR eine wichtige Rolle.
- Wichtige Aspekte linearer Zusammenhänge:
  - **Richtung** des Zusammenhangs.
  - **Stärke** des Zusammenhangs.
  - **Einheitsunabhängigkeit** des Zusammenhangs.
- Wichtige Aspekte hinsichtlich der Stärke eines linearen Zusammenhangs:
  - **Steigung** der Regressionsgeraden.
    -> abgebildet durch den Parameter β
  - **Streuung** der AV-Werte um die Regressionsgerade.
    -> abgebildet durch den Parameter σ^2

Vorbemerkung: Effektgrößen bei der ELR

- Im Folgenden werden wir zwei Effektgrößen kennenlernen, die in Regressionsmodellen verwendet werden:
  - Die Effektgröße ρ^2
  - Die Effektgröße β_z

- Beide Effektgrößen zeichnen sich dadurch aus, dass sie beide Aspekte der Stärke von linearen Zusammenhängen berücksichtigen und von der Maßeinheit der Rohdaten unabhängig sind.
- Die Richtung des linearen Zusammenhangs wird hingegen nur von der Effektgröße β_z berücksichtigt.

Notation

Wir unterscheiden wie immer zwischen dem Parameter, der Schätzfunktion und dem Schätzwert:

- Parameter: ρ^2 bzw. β_z
- Schätzfunktion: $\hat{\rho}^2 = R^2$ bzw. $\hat{\beta}_z = B_z$
- Schätzwert: r^2 bzw. b_z

Definition der Effektgröße ρ^2 und deren Schätzfunktion und Konfidenzintervall

- Der Quotient aus der Varianz der bedingten Erwartungswerte μ_i der AV und der Gesamtvarianz σ^2_{ges} der AV definiert die Effektgröße ρ^2:

$$\rho^2 = \frac{\sigma^2_{\mu_i}}{\sigma^2_{ges}}$$

- Interpretation: Anteil der Gesamtvarianz der AV in der Population, der durch die UV erklärt werden kann.

- Eine häufig verwendete Schätzfunktion für ρ^2 ist

$$\hat{\rho}^2 = R^2 = \frac{\hat{\sigma}^2_{\mu_i}}{\hat{\sigma}^2_{ges}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- Für ρ^2 kann ein Konfidenzintervall konstruiert werden.
  Die Herleitung des Intervalls ist kompliziert, weswegen wir uns auf die Berechnung in R (z.B. Funktion ci.R2 im MBESS Paket) beschränken.

Schätzwert r^2

- Die Realisation r^2 von R^2 ist unser Schätzwert für ρ^2:

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- Bemerkung I: Der Schätzwert r^2 ist im Rahmen der ELR identisch mit der quadrierten Pearson-Korrelation.
- Bemerkung II: Man bezeichnet r^2 auch als Determinationskoeffizienten.

Beispiel 1 in R

- Erstes Beispiel aus der letzten Vorlesung: Vorhersage von Depression durch negative Selbstbewertung:

```
Residual standard error: 7.879 on 189 degrees of freedom
Multiple R-squared:  0.5119, Adjusted R-squared:  0.5093
F-statistic: 198.2 on 1 and 189 DF,  p-value: < 0.00000000000000022

> ci.R2(R2 = 0.5119, p = 1, N = 191)
$Lower.Conf.Limit.R2
[1] 0.4062197

$Prob.Less.Lower
[1] 0.025

$Upper.Conf.Limit.R2
[1] 0.604553

$Prob.Greater.Upper
[1] 0.025
```

- Wir gehen davon aus, dass 40.62 bis 60.46 % der Gesamtvarianz der Depression in der Population durch die negative Selbstbewertung erklärt werden können.

Beispiel 2 in R

- Zweites Beispiel aus der letzten Vorlesung: Vorhersage von Depression durch Abhängigkeitskognitionen:

```
Residual standard error: 10.48 on 189 degrees of freedom
Multiple R-squared:  0.1362, Adjusted R-squared:  0.1316
F-statistic: 29.79 on 1 and 189 DF,  p-value: 0.0000001496

> ci.R2(R2 = 0.1362, p = 1, N = 191)
$Lower.Conf.Limit.R2
[1] 0.05711453

$Prob.Less.Lower
[1] 0.025

$Upper.Conf.Limit.R2
[1] 0.2348333

$Prob.Greater.Upper
[1] 0.025
```

- Wir gehen davon aus, dass 5.71 % bis 23.48 % der Gesamtvarianz der Depression in der Population durch die Abhängigkeitskognitionen erklärt werden können.

Effektgröße β_z

- Die **Effektgröße $\boldsymbol{\beta_z}$** entspricht dem „normalen“ β, falls sowohl AV und UV **z-standardisiert** werden.

➢ Interpretation 1: Wenn die unabhängige Variable sich um eine **Standardabweichung** erhöht, dann erhöht sich die abhängige Variable im Mittel um β_z **Standardabweichungen**.

- Man kann zeigen, dass **in der ELR** die β_z der Korrelation von AV und UV in der Population entspricht.

➢ Interpretation 2: β_z ist die **Korrelation von AV und UV** in der Population.

- Zusammenhang zwischen β_z und ρ^2:

$$\rho^2 = \beta_z^2$$

- Vorteil von β_z gegenüber ρ^2: β_z liefert Informationen über die **Richtung** des linearen Zusammenhangs.

Schätzfunktion B_z und Schätzwert b_z, Konfidenzintervall für β_z

- Die Schätzfunktion B_z entspricht der „normalen“ Schätzfunktion für den Steigungsparameter nach z-Standardisierung der Daten.
- Die Realisation b_z ist der Schätzwert für β_z.
- Der Schätzwert b_z ist im Rahmen der ELR identisch mit der Pearson-Korrelation in der Stichprobe (siehe auch VL 3 in Statistik I).

- Für β_z kann ein Konfidenzintervall konstruiert werden.
- Dieses KI entspricht dem in der letzten Vorlesung besprochenen KI für den Steigungsparameter nach z-Standardisierung der Daten.

Beispiel 1 in R

- Erstes Beispiel aus der letzten Vorlesung: Vorhersage von Depression durch negative Selbstbewertung:

```
> fit_nsb_z <- lm(scale(bdi_ges) ~ scale(fie_nsb), data = daten)
> summary(fit_nsb_z)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.043e-17  5.068e-02    0.00        1
scale(fie_nsb) 7.155e-01  5.082e-02   14.08   <2e-16 ***

> confint(fit_nsb_z)
                     2.5 %     97.5 %
(Intercept)    -0.09997963 0.09997963
scale(fie_nsb)  0.61524526 0.81573004
```

- Wir gehen davon aus, dass die Korrelation zwischen Depression und negativer Selbstbewertung in der Population zwischen 0.62 und 0.82 liegt.

Beispiel 2 in R

- Zweites Beispiel aus der letzten Vorlesung: Vorhersage von Depression durch Abhängigkeitskognitionen:

```
> fit_abk_z <- lm(scale(bdi_ges) ~ scale(fie_abk), data = daten)
> summary(fit_abk_z)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.606e-16  6.743e-02   0.000        1
scale(fie_abk) 3.690e-01  6.761e-02   5.458  1.5e-07 ***

> confint(fit_abk_z)
                    2.5 %    97.5 %
(Intercept)    -0.1330088 0.1330088
scale(fie_abk)  0.2356583 0.5023750
```

- Wir gehen davon aus, dass die Korrelation zwischen Depression und Abhängigkeitskognitionen in der Population zwischen 0.24 und 0.50 liegt.

Zwischengliederung

- Bislang:
  - Effektstärken ρ^2 und β_z in der ELR
- Jetzt:
  - Stichprobenumfangsplanung in der ELR

Stichprobenumfangsplanung bei der ELR I

- Es soll mithilfe von R der Mindeststichprobenumfang für den Hypothesentest berechnet werden, mit den in der ELR äquivalenten Hypothesen:
  - $H_0: \beta = 0, \;\; H_1: \beta \neq 0$
  - $H_0: \beta_z = 0, \;\; H_1: \beta_z \neq 0$
  - $H_0: \rho^2 = 0, \;\; H_1: \rho^2 \neq 0$

  (Wenn $\beta = 0$ gilt, dann gilt auch $\beta_z = 0$, dann gilt auch $\rho^2 = 0$)

- Der statistische Test soll dabei folgende Eigenschaften haben:

  $\alpha = 0.005$ und $1 - \beta = 0.8$

- Als Mindesteffekt geben wir $\beta_z = 0.2$ bzw. $\rho^2 = 0.2^2 = 0.04$ vor.
- Für die Berechnung in R müssen wir β_z bzw. ρ^2 in eine alternative Effektgröße f^2 umrechnen:

$$f^2 = \frac{\beta_z^2}{1 - \beta_z^2} = \frac{\rho^2}{1 - \rho^2} \approx 0.04$$

Stichprobenumfangsplanung bei der ELR II


```
> pwr.f2.test(u = 1, f2 = 0.04, sig.level = 0.005, power = 0.8)

     Multiple regression power calculation

              u = 1
              v = 334.7723
             f2 = 0.04
      sig.level = 0.005
          power = 0.8
```

- u gibt in diesem Fall die Anzahl der Prädiktoren an.
  Bei der ELR ist dies immer 1.
- ν ist der Parameter der t-Verteilung der Teststatistik unter der H_0.
- Wegen $\nu = n - 2$ ist die erforderliche Stichprobengröße $\nu + 2$.
- Wir brauchen also mindestens 337 Personen.

Zwischengliederung

- Bislang:
  - Effektstärken ρ^2 und β_z in der ELR
  - Stichprobenumfangsplanung in der ELR
- Jetzt:
  - Effektstärken in der MLR

Wiederholung: Multiple lineare Regression (MLR)

- Zur Erinnerung: Wenn das statistische Modell mehr als einen Prädiktor enthält, spricht man von einer multiplen Regression.

- In der letzten Vorlesung haben wir ein Modell mit zwei Prädiktoren betrachtet. Dieses lässt sich sehr einfach auf mehr als zwei Prädiktoren verallgemeinern.

- Allgemeine Modellgleichung für *k* Prädiktoren:

$$Y_i = \alpha + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \ ... + \beta_k \cdot X_{ik} + \varepsilon_i, \text{wobei } \varepsilon_i \sim N(0, \sigma^2)$$

- Wir werden die Themen der restlichen Vorlesung im Rahmen dieses allgemeinen Modells besprechen.

Effektgrößen in der MLR

- Im Rahmen der MLR können zwei Kategorien von Effektgrößen unterschieden werden:
    - Effektgrößen, die die Stärke des Zusammenhang **aller Prädiktoren gemeinsam** mit dem Kriterium quantifizieren.
    - Effektgrößen, die die Stärke des Zusammenhang eines **einzelnen Prädiktors** mit dem Kriterium quantifizieren.

- Aus der ersten Kategorie werden wir die Effektgröße ρ^2 besprechen.
- Aus der zweiten Kategorie werden wir die Effektgrößen β_{z_j} und ρ_j^2 besprechen.

Definition ρ^2

- Das ρ^2 in der MLR ist eine direkte Verallgemeinerung des ρ^2 aus der ELR.
- Es ist wie dort definiert als der Quotient aus der Varianz der bedingten Erwartungswerte μ_i der AV und der Gesamtvarianz σ_{ges}^2 der AV:

$$\rho^2 = \frac{\sigma_{\mu_i}^2}{\sigma_{ges}^2}$$

- Interpretation: Anteil der Gesamtvarianz der AV in der Population, der durch alle UVs gemeinsam erklärt werden kann.
- Bemerkung: Die Wurzel aus ρ^2 wird auch als multiple Korrelation bezeichnet.

Schätzfunktion R^2 und Schätzwert r^2

- Eine häufig verwendete Schätzfunktion für ρ^2 ist die gleiche wie in der ELR (Folie 5)

$$\hat{\rho}^2 = R^2 = \frac{\hat{\sigma}^2_{\mu_i}}{\hat{\sigma}^2_{ges}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)2}{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Mit dem Schätzwert

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- Bemerkung: Die Wurzel aus r^2 entspricht der Korrelation zwischen den vorhergesagten AV-Werten $\hat{y}_i$ und den tatsächlichen AV-Werten y_i in der Stichprobe.

Konfidenzintervall für ρ^2

- Für ρ^2 kann ein Konfidenzintervall konstruiert werden.

- Die Herleitung des Intervalls ist kompliziert, weswegen wir uns auf die Berechnung in R (z.B. Funktion ci.R2 im MBESS Paket) beschränken.

Beispiel in R

- Beispiel aus der letzten Vorlesung. Vorhersage von Depression durch negative Selbstbewertung und Abhängigkeitskognitionen:

```
Residual standard error: 7.854 on 188 degrees of freedom
Multiple R-squared:  0.5176, Adjusted R-squared:  0.5125
F-statistic: 100.9 on 2 and 188 DF,  p-value: < 0.00000000000000022

> ci.R2(R2 = 0.5176, p = 2, N = 191)
$Lower.Conf.Limit.R2
[1] 0.4092714

$Prob.Less.Lower
[1] 0.025

$Upper.Conf.Limit.R2
[1] 0.607498

$Prob.Greater.Upper
[1] 0.025
```

- Wir gehen davon aus, dass 40.93 bis 60.75 % der Gesamtvarianz der Depression in der Population durch die negative Selbstbewertung und die Abhängigkeitskognitionen zusammen erklärt werden können.

Definition β_{z_j}

- β_{z_j} in der MLR ist eine direkte Verallgemeinerung von β_z aus der ELR.
- Analog entsprechen die β_{z_j} den „normalen“ β_j, falls **die AV und alle UVs** z-standardisiert werden.
- Interpretation: Wenn sich die unabhängige Variable j um eine Standardabweichung erhöht, dann erhöht sich die abhängige Variable im Mittel um β_{z_j} Standardabweichungen, falls alle anderen UVs konstant bleiben.
- Es gibt also für jede UV j eine eigene Effektstärke β_{z_j}, die angibt, wie stark die jeweilige UV bei Konstanthaltung aller anderen UVs mit der AV zusammenhängt.
- Bemerkung: β_{z_j} entspricht in der MLR **nicht** der Korrelation von AV und UV j in der Population.

Schätzfunktion B_{z_j}, Schätzwert b_{z_j} und Konfidenzintervall für β_{z_j}

- Die Schätzfunktion B_{z_j} für β_{z_j} entspricht der „normalen“ Schätzfunktion für den jeweiligen Steigungsparameter nach z-Standardisierung der Daten.

- Die Realisation b_{z_j} ist der Schätzwert für β_{z_j}
- Bemerkung: Der Schätzwert b_{z_j} ist im Rahmen der MLR **nicht** identisch mit der Pearson-Korrelation zwischen der UV j und der AV.

- Für β_{z_j} kann ein Konfidenzintervall konstruiert werden.
- Dieses KI entspricht dem „normalen“ KI für den jeweiligen Steigungsparameter nach z-Standardisierung der Daten.

Beispiel in R I

- Erstes Beispiel aus der letzten Vorlesung. Vorhersage von Depression durch negative Selbstbewertung und Abhängigkeitskognitionen:

```
> fit_z <- lm(scale(bdi_ges) ~ scale(fie_nsb) + scale(fie_abk), data =
daten)
> summary(fit_z)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.421e-17  5.052e-02   0.000    1.000
scale(fie_nsb) 6.806e-01  5.582e-02  12.192   <2e-16 ***
scale(fie_abk) 8.301e-02  5.582e-02   1.487    0.139

> confint(fit_z)
                     2.5 %     97.5 %
(Intercept)    -0.09966423 0.09966423
scale(fie_nsb)  0.57048377 0.79072610
scale(fie_abk) -0.02711148 0.19313084
```

Beispiel in R II

- Wir gehen davon aus, dass sich die Depression im Mittel um 0.57 bis 0.79 Standardabweichungen erhöht, falls sich die negative Selbstbewertung um eine Standardabweichung erhöht und die Abhängigkeitskognitionen konstant bleiben.

- Wir gehen davon aus, dass sich die Depression im Mittel um -0.03 bis 0.19 Standardabweichungen erhöht, falls sich die Abhängigkeitskognitionen um eine Standardabweichung erhöht und die negative Selbstbewertung konstant bleibt.

Definition ρ_j^2

- Eine Alternative zu β_{z_j} in der MLR stellt die Effektstärke ρ_j^2 dar.
- Definition:

$$\rho_j^2 = \rho^2 - \rho_{-j}^2$$

- wobei ρ_{-j}^2 demjenigen ρ^2 entspricht, das man erhält, wenn man den Prädiktor j aus dem MLR-Modell entfernt.

- Es gibt also für jeden Prädiktor ein eigenes ρ_j^2.
- Interpretation: Anteil der Gesamtvarianz der AV in der Population, der über die anderen Prädiktoren hinaus nur durch die UV j erklärt werden kann.
- Bemerkung I: ρ_j^2 wird auch als quadrierte Semipartialkorrelation bezeichnet.
- Bemerkung II: Die Summe der einzelnen ρ_j^2 muss nicht ρ^2 ergeben.

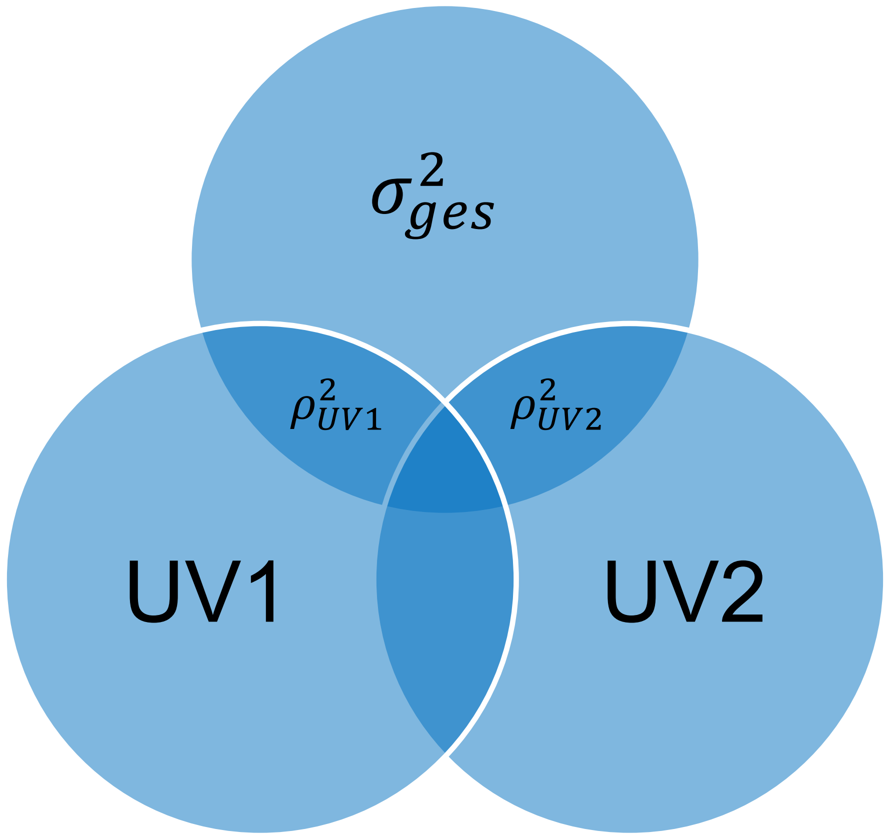
Beispiel ρ_j^2 in der Population

- Wir nehmen an, dass im Fall eines MLR-Modells mit der AV Depression und den UVs negative Selbstbewertung und Abhängigkeitskognitionen $\rho^2 = 0.5$ ist. Negative Selbstbewertung um Abhängigkeitskognitionen können also zusammen 50% der Varianz der Depression erklären.

- ρ^2_{-nsb} wäre in diesem Fall das ρ^2 aus dem Modell ohne die negative Selbstbewertung. In diesem Fall bliebe nur ein Prädiktor übrig: Die Abhängigkeitskognitionen.
- ρ^2_{-nsb} ist hier also das ρ^2 eines ELR-Modells mit Depression als AV und Abhängigkeitskognitionen als UV.
- Nehmen wir an, dass dieses ρ^2_{-nsb} gleich 0.3 wäre.
- Dann wäre der Anteil der Gesamtvarianz der Depression, der über die Abhängigkeitskognitionen hinaus nur durch die negative Selbstbewertung erklärt werden kann

$$\rho^2_{nsb} = \rho^2 - \rho^2_{-nsb} = 0.5 - 0.3 = 0.2$$

Graphische Veranschaulichung ρ_j^2



Beispiel Schätzwerte für ρ_j^2 aus der Stichprobe I

- Berechnung der Schätzwerte für ρ_j^2 in unserem Beispiel mit Depression als AV, negativer Selbstbewertung als UV1 und Abhängigkeitskognitionen als UV2:

- Der Schätzwert für ρ^2 entspricht dem r^2 aus dem MLR-Modell mit beiden Prädiktoren (siehe oben):

```
> fit = lm(bdi_ges ~ fie_nsb + fie_abk, data = daten)
> summary(fit)

Residual standard error: 7.854 on 188 degrees of freedom
Multiple R-squared:  0.5176, Adjusted R-squared:  0.5125
F-statistic: 100.9 on 2 and 188 DF,  p-value: < 0.00000000000000022
```

Beispiel Schätzwerte für ρ_j^2 aus der Stichprobe II

- Der Schätzwert für ρ^2_{-nsb} entspricht dem r^2 aus dem Modell ohne den Prädiktor negative Selbstbewertung (also nur mit dem Prädiktor Abhängigkeitskognitionen):

```
> fit = lm(bdi_ges ~ fie_abk, data = daten)
> summary(fit)

Residual standard error: 10.48 on 189 degrees of freedom
Multiple R-squared:  0.1362, Adjusted R-squared:  0.1316
F-statistic: 29.79 on 1 and 189 DF,  p-value: 0.0000001496
```

- Ein Schätzwert für $\rho^2_{nsb} = \rho^2 - \rho^2_{-nsb}$ wäre also die Differenz der Schätzwerte für ρ^2 und ρ^2_{-nsb} , in diesem Fall also $0.5176 - 0.1362 = 0.3814$

Beispiel Schätzwerte für ρ_j^2 aus der Stichprobe III

- Der Schätzwert für ρ^2_{-abk} entspricht dem r^2 aus dem Modell ohne den Prädiktor Abhängigkeitskognitionen (also nur mit dem Prädiktor negative Selbstbewertung):

```
> fit = lm(bdi_ges ~ fie_nsb, data = daten)
> summary(fit)

Residual standard error: 7.879 on 189 degrees of freedom
Multiple R-squared:  0.5119, Adjusted R-squared:  0.5093
F-statistic: 198.2 on 1 and 189 DF,  p-value: < 0.00000000000000022
```

- Ein Schätzwert für $\rho^2_{abk} = \rho^2 - \rho^2_{-abk}$ wäre also die Differenz der Schätzwerte für ρ^2 und ρ^2_{-abk} , in diesem Fall also $0.5176 - 0.5119 = 0.0057$

Inferenzstatistische Verfahren für ρ_j^2

- KIs für ρ_j^2 werden wir im Rahmen dieser Vorlesung nicht näher besprechen.

- Wir werden ρ_j^2 jedoch im Rahmen der Stichprobenplanung verwenden.

Zwischengliederung

- Bislang:
  - Effektstärken ρ^2 und β_z in der ELR
  - Stichprobenumfangsplanung in der ELR
  - Effektgrößen im Rahmen der MLR
- Jetzt:
  - Stichprobenplanung in der MLR

Stichprobenplanung für den Hypothesentest über ein β_j

- Stichprobenumfangsplanungen für allgemeine Hypothesen über β_j sind leider recht umständlich. Für den einfachsten Fall können wir aber den Umweg über ρ_j^2 machen.
- Dabei machen wir uns zu Nutze, dass die folgenden Hypothesenpaare äquivalent sind:
  - $H_0: \beta_j = 0,\ H_1: \beta_j \neq 0$
  - $H_0: \rho_j^2 = 0,\ H_1: \rho_j^2 \neq 0$
- Um die Stichprobengröße für diesen Hypothesentest in R bestimmen zu können, müssen wir eine weitere alternative Effektgröße berechnen:

$$f_j^2 = \frac{\rho_j^2}{1 - \rho^2}$$

- Falls wir z.B. ein $\rho_j^2 = 0.02$ und ein $\rho^2 = 0.1$ vorgeben, erhalten wir

$$f_j^2 = \frac{\rho_j^2}{1 - \rho^2} = \frac{0.02}{1 - 0.1} \approx 0.02$$

Stichprobenplanung für den Hypothesentest über ein β_j

- In R:

```
> pwr.f2.test(u = 1, f2 = 0.02, sig.level = 0.005, power = 0.8)

     Multiple regression power calculation

              u = 1
              v = 667.5817
             f2 = 0.02
      sig.level = 0.005
          power = 0.8
```

- Bei der Stichprobenplanung des Hypothesentests **für** $\boldsymbol{\beta_j}$ in der MLR gilt **immer** $\mathbf{u = 1}$ (unabhängig von der Anzahl der Prädiktoren).
- Die benötigte Stichprobengröße ist $n = v + k + 1$, also in unserem Fall mit $k = 2$ Prädiktoren aufgerundet 671 Personen.

Stichprobenplanung für den Omnibustest

- Um die Stichprobengröße für den Omnibustest mit den Hypothesen

$$H_0\colon \beta_j = 0 \text{ für alle } j$$
$$H_1\colon \beta_j \neq 0 \text{ für mindestens ein } j$$

- in R bestimmen zu können, müssen wir

$$f^2 = \frac{\rho^2}{1-\rho^2}$$

berechnen.

- Falls wir z.B. $\rho^2 = 0.1$ vorgeben, erhalten wir

$$f^2 = \frac{\rho^2}{1-\rho^2} = \frac{0.1}{1-0.1} \approx 0.11$$

Stichprobenplanung für den Omnibustest

- In R:

```
> pwr.f2.test(u = 2, f2 = 0.11, sig.level = 0.005, power = 0.8)
Multiple regression power calculation
              u = 2
              v = 144.6068
             f2 = 0.11
      sig.level = 0.005
          power = 0.8
```

- Bei der Stichprobenplanung des Omnibustests in der MLR entspricht u der Anzahl der Prädiktoren.
- Die benötigte Stichprobengröße ist $n = v + k + 1$, mit $k = 2$ Prädiktoren also z.B. aufgerundet 148 Personen.

Zwischengliederung

- Bislang:
  - Effektstärken ρ^2 und β_z in der ELR
  - Stichprobenumfangsplanung in der ELR
  - Effektgrößen im Rahmen der MLR
  - Stichprobenplanung in der MLR
- Jetzt:
  - Kollinearität

Was ist Kollinearität?

- Man spricht im Rahmen der MLR von Kollinearität, wenn einer oder mehrere der Prädiktoren stark mit den jeweils anderen Prädiktoren zusammenhängen.
- Kollinearität hat einen Einfluss auf die Standardfehler der Schätzfunktionen für die Steigungsparameter β_j.
- Damit wirkt sich Kollinearität auch auf die Konfidenzintervalle und Hypothesentests für die Steigungsparameter β_j aus.

- Konfidenzintervalle für ρ^2, der Omnibustest der MLR, sowie die Schätzung von Vorhersagewerten sind von der Kollinearität jedoch nicht beeinträchtigt.
- Bemerkung: In der Literatur wird statt „Kollinearität“ häufig auch der Begriff „Multikollinearität“ verwendet.

Kollinearität und Standardfehler von B_j

- Ausgangslage: MLR Modell mit k Prädiktoren

$$Y_i = \alpha + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \ldots + \beta_k \cdot X_{ik} + \varepsilon_i$$

- Der Standardfehler der Schätzfunktion $B_j = \hat{\beta}_j$ ist:

$$SE(B_j) = \sqrt{Var(B_j)} = \sqrt{\frac{1}{1-\boldsymbol{r_j^2}} \cdot \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}}$$

- Hierbei ist r_j^2 der Determinationskoeffizient eines Regressionsmodells, in das der Prädiktor j als Kriterium und alle anderen Prädiktoren als Prädiktoren aufgenommen werden. Dieser gibt den Anteil der Gesamtvarianz des Prädiktors j an, der durch die anderen Prädiktoren erklärt werden kann und ist somit ein Maß für die Kollinearität d.h. ein Maß dafür, wie stark der Prädiktor mit den anderen Prädiktoren linear zusammenhängt.

r_j^2 für $k = 2$ Prädiktoren

- Im Falle von zwei Prädiktoren ist r_j^2 einfach die quadrierte Pearson-Korrelation zwischen Prädiktor 1 und Prädiktor 2 ($\boldsymbol{r_{x_1x_2}^2}$).
- Die Standardfehler der Schätzfunktionen sind in diesem Fall

$$SE(B_1) = \sqrt{Var(B_1)} = \sqrt{\frac{1}{1 - \boldsymbol{r_{x_1x_2}^2}} \cdot \frac{\sigma^2}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}}$$

$$SE(B_2) = \sqrt{Var(B_2)} = \sqrt{\frac{1}{1 - \boldsymbol{r_{x_1x_2}^2}} \cdot \frac{\sigma^2}{\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2}}$$

➢ Je größer r_j^2, desto größer der Standardfehler $SE(B)$.

➢ Je größer der Standardfehler $SE(B)$ desto geringer die Power der Hypothesentests über β_j und desto größer werden Konfidenzintervalle für β_j.

Umgang mit Kollinearität

- Wenn Kollinearität also Konfidenzintervalle für β_j größer und die Power des Hypothesentest für β_j kleiner macht, sollte sie dann nicht um jeden Preis verhindert werden?
- Vorsicht! Welche Prädiktoren in einem Modell sind, ist in erster Linie eine **inhaltliche** Frage und **keine statistische**.
- Es besteht inhaltlich ein ggf. wichtiger Unterschied zwischen Fragestellungen wie:
  - „Was kann ich durch die negative Selbstbewertung (UV) über die Depressivität einer Person (AV) herausfinden?“
  - „Was kann ich durch die negative Selbstbewertung (UV1) zusätzlich über die Depressivität einer Person (AV) herausfinden, wenn ich die Abhängigkeitskognitionen (UV2) der Personen bereits kenne?“
- Dem Effekt der Kollinearität auf die Standardfehler kann auch durch eine größere Stichprobe *n* entgegengewirkt werden:

$$SE(B_j) = \sqrt{Var(B_j)} = \sqrt{\frac{1}{1 - r_j^2} \cdot \frac{\sigma^2}{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}}$$

Welche Prädiktoren sollten also ins Modell und welche nicht?

- In der Literatur wird bei der Aufnahme von Drittvariablen („Kovariate“) als weitere Prädiktoren ins Regressionsmodell davon gesprochen, dass der Einfluss dieser Drittvariablen statistisch kontrolliert wird.
- Drittvariablen sind Variablen, die uns primär nicht interessieren, die aber womöglich den Zusammenhang zwischen AV und interessierender UV beeinflussen können.
- Prinzipiell kann die Kontrolle von solchen Variablen sinnvoll sein, um Fehlschlüsse zu vermeiden.
- Eine unüberlegte Hinzunahme von Variablen ins Modell kann aber echte Zusammenhänge auch verzerren und Fehlschlüsse erst herbeiführen.

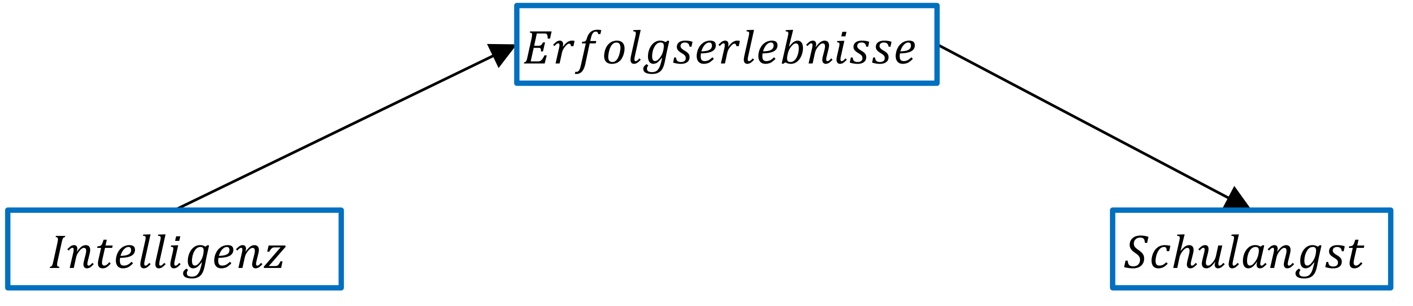
➢ Die Entscheidung, für Drittvariablen zu kontrollieren, diese also in das Modell mitaufzunehmen, sollte deshalb nie leichtfertig geschehen und immer vor dem Hintergrund theoretischer Überlegungen passieren!

Directed Acyclic Graphs (DAGs)

- Directed Acyclic Graphs (DAG) können genutzt werden, um theoretische Kausalannahmen graphisch zu veranschaulichen und potentielle Störfaktoren („Confounder“) aufzudecken. Sie helfen uns zu entscheiden, ob bzw. welche Drittvariablen in das Modell aufgenommen werden sollten und welche nicht.

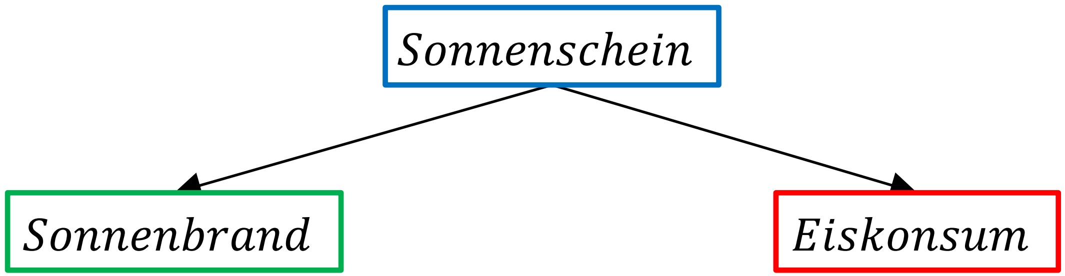
- DAGs veranschaulichen komplexe kausale Zusammenhänge durch gerichtete Pfeile. Die Richtung der Pfeile gibt dabei die kausale Wirkrichtung an.

- Ein Pfeil kann dabei sowohl gleichgerichtete als auch gegengerichtete Zusammenhänge darstellen:



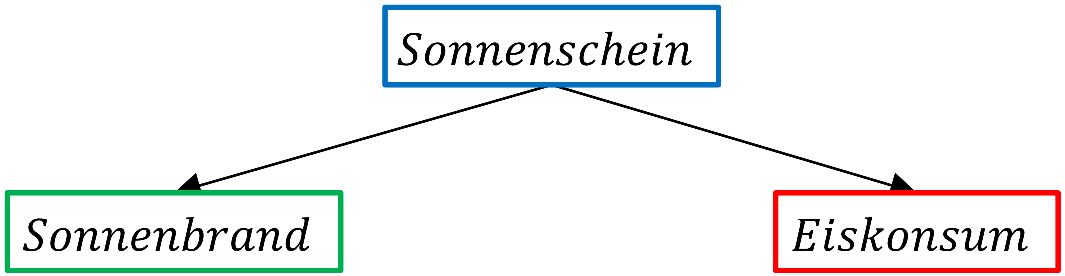
- Aus DAGs können eindeutige Regeln abgeleitet werden, ob die Aufnahme einer Drittvariablen ins Modell sinnvoll ist oder nicht, je nach Anordnung der Variablen und Pfeile (also je nach theoretischen Annahmen und Überlegungen).

Beispiel 1: Fork – the Confounder



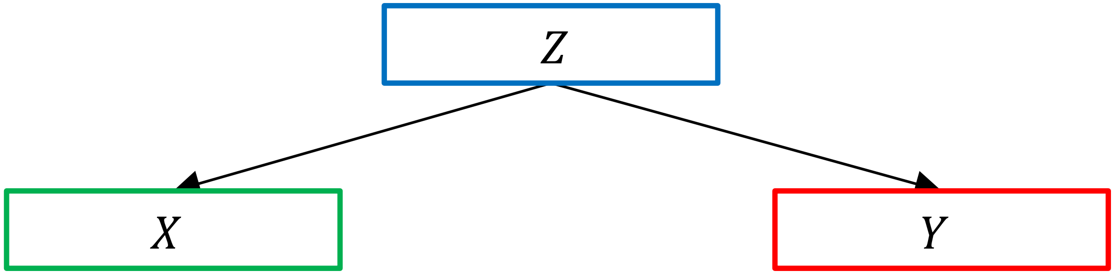
- In einer ELR mit Eiskonsum als AV und Sonnenbrand als UV wird ein Zusammenhang dieser beiden Variablen „entdeckt“, obwohl hier gar kein offensichtlicher (kausaler) Zusammenhang gilt.
- In einer MLR mit Sonnenschein als zusätzlichem Prädiktor wird dieser Effekt verschwinden und wir lernen, dass der Sonnenschein hier eigentlich entscheidend sein könnte.

Beispiel 1: Fork – the Confounder



- Anders ausgedrückt: betrachten wir den Zusammenhang zwischen Sonnenbrand und Eiskonsum „alleine“, finden wir einen (positiven) Zusammenhang. Kontrollieren wir für Sonnenschein, betrachten den Zusammenhang also für eine spezifische Ausprägung der Variable Sonnenschein, verschwindet der Zusammenhang zwischen Sonnenbrand und Eiskonsum.
- (Zur Erinnerung: Bei einer MLR ist der Zusammenhang zwischen UV1 und AV über alle Ausprägung von UV2 hinweg gleich)
- Wir könnten also sagen: Sonnenschein erklärt den Zusammenhang zwischen Sonnenbrand und Eiskonsum.

Beispiel 1: Fork




```
> n <- 1000
> z <- rnorm(n)
> x <- rnorm(n, mean = 1*z)
> y <- rnorm(n, mean = 1*z)
```

Verzerrter Effekt

```
Call: lm(y ~ x)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03823    0.03932   0.972    0.331
x            0.50619    0.02887  17.533   <2e-16 ***
```

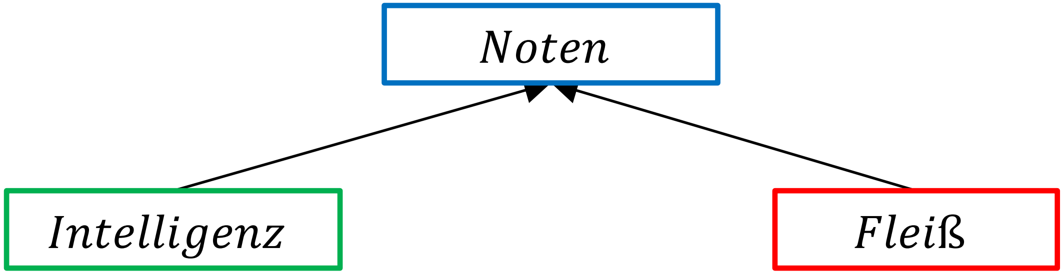
Korrekter Effekt

```
Call: lm(y ~ x + z)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02378    0.03174   0.749    0.454
x           -0.02706    0.03279  -0.825    0.409
z            1.01662    0.04396  23.124   <2e-16 ***
```

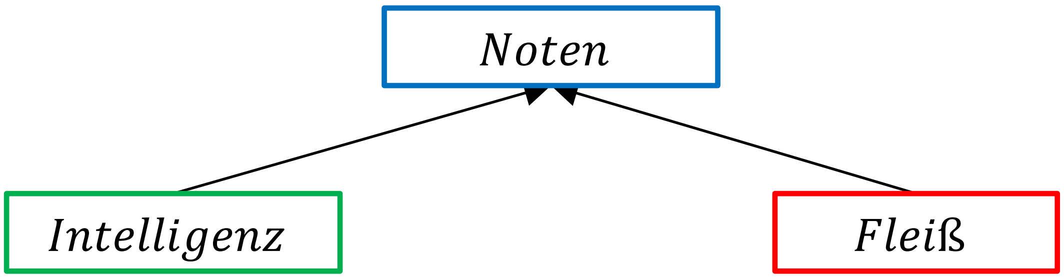
- Ist eine Variable ein Confounder (wie hier die Variable Z / Sonnenschein), dann sollte sie (wenn möglich) in das Modell als UV mitaufgenommen werden, um den tatsächlichen Zusammenhang zwischen X und Y zu schätzen.

Beispiel 2: Collider



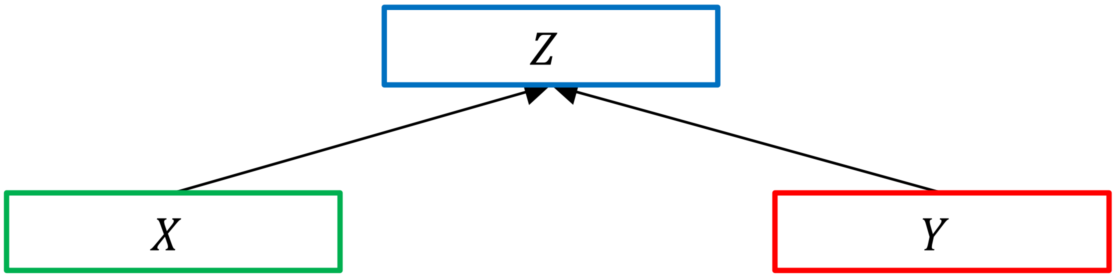
- Muss bei der Suche nach dem (in diesem theoretischen Modell nicht vorhandenen) Zusammenhang von Intelligenz (UV) und Fleiß (AV) die Note als zweite UV mit aufgenommen werden?
- In einer MLR mit beiden Prädiktoren (Intelligenz und Noten) wird plötzlich ein Effekt zwischen Intelligenz und Fleiß auftauchen, der uns in die Irre leiten kann!
- Hier sollte also die Note nicht als Prädiktor ins Modell aufgenommen und eine ELR gerechnet werden, um den Zusammenhang von Intelligenz und Fleiß auf den richtigen Wert (= 0) zu schätzen.

Beispiel 2: Collider



- Begründung: (in diesem Beispiel) hängen Intelligenz und Fleiß nicht zusammen. Kontrollieren wir aber für die Note, welche von beiden Variablen kausal verursacht wird, bedeutet das inhaltlich, dass wir uns den Zusammenhang von Intelligenz und Fleiß unter Konstanthaltung der Noten anschauen. Dadurch entsteht aber ein (nicht-kausaler) Zusammenhang zwischen diesen beiden Variablen.
- Beispiel: Wenn wir die Intelligenz von Personen betrachten und dabei für die Note kontrollieren (also z.B. die Intelligenz aller Personen mit einer guten Note), dann können wir in Abhängigkeit der Intelligenz etwas über den Fleiß dieser Personen sagen. Bei einer guten Note und niedrigen Intelligenz müsste dann ja im Mittel der Fleiß höher sein, andernfalls hätten diese Personen (nach unserer theoretischen Annahme) keine gute Note. Wir würden also einen negativen Zusammenhang zwischen Intelligenz und Fleiß finden („explain-away effect“).

Beispiel 2: Collider




```
> n <- 1000
> x <- rnorm(n)
> y <- rnorm(n)
> z <- rnorm(n, mean = 1*x + 1*y)
```

Verzerrter Effekt

```
Call: lm(y ~ x + z)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.001578   0.022284  -0.071    0.944
x           -0.531703   0.027384 -19.417   <2e-16 ***
z            0.536929   0.015840  33.897   <2e-16 ***
```

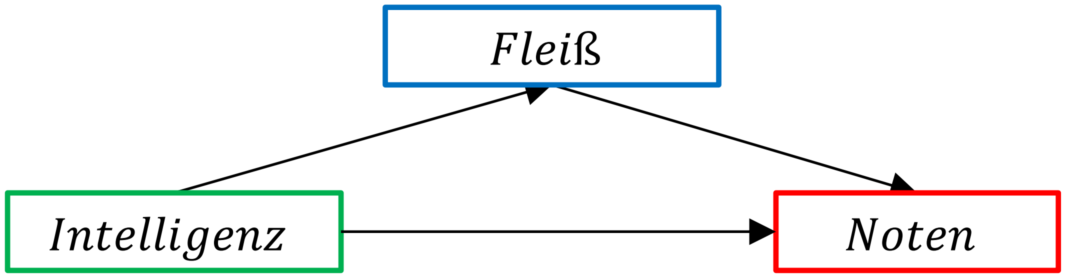
Korrekter Effekt

```
lm(y ~ x)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.001448   0.032677  -0.044    0.965
x            0.004991   0.032763   0.152    0.879
```

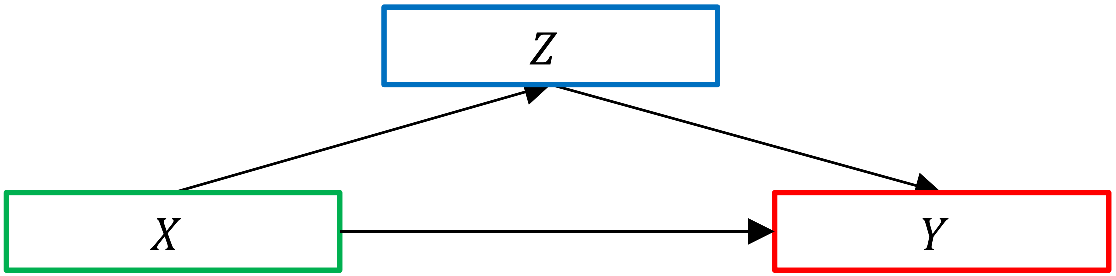
- Ist eine Variable ein Collider (wie hier die Variable Z / Noten), dann sollte sie (wenn möglich) NICHT in das Modell als UV mitaufgenommen werden.

Beispiel 3: Pipe (Mediation)



- In einem alternativen Modell gehen wir davon aus, dass Intelligenz zu besseren Noten führt.
- Gleichzeitig verhindert Intelligenz die Drittvariable Fleiß (= negativer kausaler Zusammenhang), obwohl Fleiß selbst wieder zu besseren Noten führen würde.
- In einer ELR würde nun der gesamte Einfluss von Intelligenz auf Noten geschätzt werden, also inklusive dem gegenläufigen Effekt, der über den indirekten Pfad der Variable Fleiß führt.
- In einer MLR bekommen wir eine bessere Schätzung der einzelnen Einflüsse.
- (In der Psychologie würde Fleiß hier häufig auch als Mediator bezeichnet werden)

Beispiel 3: Pipe (Mediation)




```
n <- 1000
x <- rnorm(n)
z <- rnorm(n, mean = -1*x)
y <- rnorm(n, mean = 1*x + 1*z)
```

Totaler Effekt

```
Call: lm(y ~ x)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.03143    0.04333  -0.725    0.468
x           -0.03880    0.04291  -0.904    0.366
```

Einzelne Effekte

```
Call: lm(y ~ x + z)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.03073    0.03183  -0.966    0.334
x            0.90425    0.04513  20.035   <2e-16 ***
z            0.91394    0.03130  29.195   <2e-16 ***
```

- Ist eine Variable ein Mediator (wie hier die Variable Z / Fleiß), dann hängt es von der Fragestellung ab, ob sie in das Modell als UV mitaufgenommen werden sollte.

Weitere Info zu DAGs

- Zusätzlich zu den beiden gezeigten Arten von Kausalzusammenhängen gibt es noch eine weitere, nämlichen den Descendant.
- Mit diesen vier Regeln lässt sich aus jedem noch so komplexen DAG ableiten, welche Variablen für welche Fragestellung nötig sind (siehe auch http://www.dagitty.net/dags.html).
- **Ob ein DAG zutreffend ist, ist wieder eine inhaltliche Diskussion. Ist der DAG eine falsche Repräsentation der Realität, dann sind auch die daraus abgeleiteten notwendigen Prädiktoren im Modell unter Umständen falsch.**
- Das ist allerdings kein Argument gegen die Verwendung von DAGs, sondern eher eine Kritik an ungenau formulierten psychologischen Theorien:

➢ Annahmen über Zusammenhänge zwischen Variablen und deren Rollen im Modell verschwinden nicht einfach, nur, weil wir diese Annahmen nicht offenlegen (z.B. in einem DAG).

➢ Sobald wir ein Modell in R schätzen wollen und die entsprechende Syntax eingeben (z.B. „Intelligenz ~ Fleiß + Noten“), haben wir bereits implizit Annahmen über Zusammenhänge und Rollen von Variablen getroffen!

Weitere Info zu DAGs

Weiterführende Literatur zu diesem Thema:

- Cinelli, C., Forney, A., & Pearl, J. (2022). A Crash Course in Good and Bad Controls. Sociological Methods & Research. https://doi.org/10.1177/00491241221099552
- Elwert, F. (2013). Graphical Causal Models. In S. L. Morgan (Ed.), Handbook of Causal Analysis for Social Research (pp. 245–273). Dordrecht: Springer Netherlands
- McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. 2nd Edition. Chapman and Hall/CRC. Oder in seiner Online Vorlesung auf Youtube (u.a. Lecture 6: https://youtu.be/UpP-_mBvECI)
- Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. Advances in Methods and Practices in Psychological Science, 27–42. https://doi.org/10.1177/2515245917745629

Ab nächster Woche

- Bislang:
  - ELR (stetige AV, eine stetige UV)
  - MLR (stetige AV, k stetige UVs)

- Die nächsten Vorlesungen:
  - Regressionsmodelle mit stetiger AV sowie diskreten (und stetigen) UVs
  - Regressionsmodelle mit diskreter (binärer) AV
  - Ausblick: Regressionsmodelle mit diskreter oder stetiger AV, diskreten und/oder stetigen UVs (mit oder ohne Interaktionen)