Statistik und Datenanalyse: Aufbau

GLM – Regression

Benjamin Fretwurst

▶ PDF-Version der Folien

Inhalt

- 1 Regression – bivariat
  - 1.1 Die Regressionsidee
  - 1.2 Der Regressionskoeffizient b
  - 1.3 Der standardisierte Regressionskoeffizient
  - 1.4 Das Bestimmtheitsmass R^2
- 2 Regression multivariat
  - 2.1 Das Modell mit 2 UV’s
  - 2.2 Das Modell mit 3 UV’s
- 3 Übung 1
  - 3.1 Installieren Sie R und R-Studio (neu)
  - 3.2 Laden Sie die Daten
  - 3.3 Rechnen Sie
- LEF 2
- Take Home
- Ausblick

Orga

- nutzen Sie das Forum
- Studienteilnahmepunkte
- Klausurformat (kein «open book»!)
- Ihre Fragen

Lernziele

- bivariate Regression
- Kennwerte der Regressionsanalysen
- multivariate Regression
- Übung 1a

1 Regression – bivariat

Anruf Biewald Berliner Kurier

- NB: Stimmt es, dass die Geburtenrate mit der Storchenpopulation zusammenhängt? Und haben Sie dazu Daten?
- Ich: Stimmt, das ist aber nur ein didaktischer Kalauer um Scheinkorrelationen zu erklären
- NB: Könnte ich ein Foto von Ihnen haben?
- Ich: Moment, was wollen Sie denn da berichten; ich habe dazu genausowenig gemacht wie jeder andere
- NB: Nein, wir wollen nur das Sommerloch füllen und da ist uns die Sache mit den Störchen eingefallen und Sie brauchen wir als Experten, der das Phänomen erklärt.


Streik angekündigt
Heute droht Bahn-Chaos!
SEITEN 2–3
BERLINER KURIER
ZEITUNG FÜR BERLIN-BRANDENBURG
1:1 in Magdeburg
Gogia rettet Union-Serie
SEITEN 22–23
Sie versprechen Berlin-Ausflüge, aber dann ...
Die miesen Tricks der Rentner-Abzocker
Statt Besichtigung gibt es nur stundenlange Verkaufsveranstaltungen. Angebliche Wunder-Matratzen für mehr als 2000 Euro angepriesen. KURIER und Berliner Rundfunk 91.4 enthüllen die Methoden einer fragwürdigen Branche SEITEN 6–7
30 JAHRE – 30 PROZENT!
Großer Jubiläumssonderverkauf mit bis zu 30 Prozent Rabatt nur noch bis Sonntag, den 16.12.!
Die ersten 100 Kunden bekommen einen Premium-Weihnachtsbaum!
DIE GÜNSTIGSTEN AUTOS IN BERLIN UND BRANDENBURG!

Berliner Forscher beweisen: Wo es Störche gibt, gibt es wirklich mehr Babys

Gelangweilte Statistiker haben beide Kurven übereinander gelegt. Das Ergebnis ist verblüffend

NB

Berlin – Die alte Geschichte vom Storch, der die Babys bringt. Ist da vielleicht doch etwas dran? Eine Untersuchung an der Freien Universität (FU) ergab jedenfalls: Wo es viele Störche gibt, kommen tatsächlich auch viele Babys auf die Welt.
Benjamin Fretwurst (30) von der FU völlig baff nach der Auswertung von zwei entsprechenden Statistiken: "Die Zahl der Störche korreliert eindeutig mit der Anzahl der Geburten in einer bestimmten Region."
Wahrscheinlich liegt das einfach nur daran, dass es die meisten Störche in ländlichen Regionen gibt und dort die Geburtenraten schon immer höher sind als in der Stadt.
Aber jetzt kommt's: Die Storchengleichung gilt sogar für Berlin. Um das zu beweisen, mussten die Statistiker ihre Zahlen aber ganz genau und von allen Seiten betrachten. Aber jetzt passt alles.
1993 kamen in der Hauptstadt 28 724 Kinder auf die Welt – bei drei Storchenpaaren. Zehn Jahre später waren es 28 723 Kinder (also genau ein Baby weniger) und nur noch zwei Storchenpaare. Ein Blick nach Brandenburg: Bei 1212 Storchenpaaren gab es 1993 12 238 Babys. 2003 stieg die Zahl der Gefiederten auf 1318 und die Zahl der Kinder auf 17 970. Wissenschaft ist doch etwas Herrliches!

Die Idee vom Modell

Modellidee

Für das Ergebnis der Datenerhebung wird ein Modell entworfen, das Zusammenhänge einfach darstellt. Da das Modell nie zu 100% das Ergebnis treffen wird, bleibt ein Rest, den wir Modellfehler oder einfach Fehler nennen.

Grundmodell

Ergebnis = (Modell) + Fehler

Beispiel

Mittelwert von $x_i = \overline{x} + Fehler_i$ (die Abweichungen vom Mittelwert)

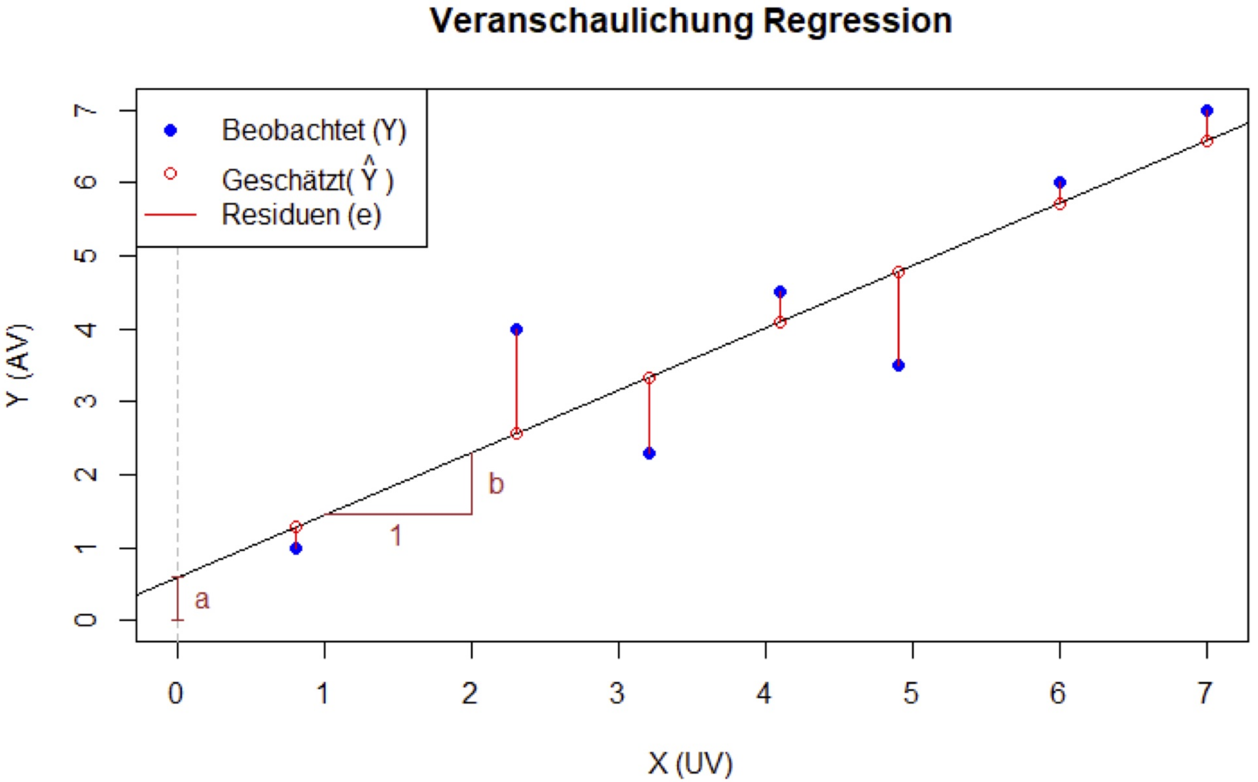
1.1 Die Regressionsidee

Was ist Regression?

Regression ist ein Weg um die Werte einer Variablen (AV) mit Hilfe einer (UV) oder mehrerer anderer Variablen (UVs) vorherzusagen.

- Es handelt sich um ein hypothetisches Model über die Beziehungen zwischen zwei und mehr Variablen.
- Das Modell nimmt lineare Beziehungen an.
    - Daher wird die Beziehung als gerade Linie dargestellt.
    - Darum spricht man auch vom «Linearen Modell».

8.2 Grundprinzip der Regressionsanalyse



Allgemeine Geradenformel: **Y = a + b * X** mit minimalen Abweichungen der beobachteten Werte von der berechneten Gerade.

Geschätzte Y-Werte: $\boldsymbol{\hat{Y}} = \boldsymbol{a} + \boldsymbol{bX}$; Beobachtete Werte: $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{bX} + \boldsymbol{e}$

Einführung in Statistik

Modul 8: Zusammenhang metrischer Variablen

8.1 Grundbegriffe und Streudiagramme

8.2 Grundprinzip der Regressionsanalyse

8.3 Fehleranalyse bei der Regressionsanalyse

8.4 Regression mit R

8.5 Teststatistik

8.6 Varianzaufklärung

8.7 Kovarianz

8.8 Korrelation

8.9 Der standardisierte Regressionskoeffizient

8.10 Ergänzungen

Notation der (multivariaten) Regression

Wir ändern die Notation etwas

$$Y = a + bX + e$$

➪ \class$fragment = b_1 + b_2X_2 + e$

Table 4

Regression results using mpg as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	*sr²*	*sr²* 95% CI [LL, UL]	*r*	Fit
(Intercept)	29.60**	[27.09, 32.11]						
disp	-0.04**	[-0.05, -0.03]	-0.85	[-1.05, -0.65]	.72	[.51, .81]	-.85**	
								R^2 = .718** 95% CI[.51,.81]

Note. * indicates $p < .05$; ** indicates $p < .01$. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights; *beta* indicates the standardized regression weights; sr^2 represents the semi-partial correlation squared; *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

Warum?

- In Tabellen (auch in R) steht die Konstante (a) in der Spalte der b's (estimates).
- Und weil wir im Multivariaten mehrere X und zugehörige b's haben, nummerieren wir sie durch, wobei wir in der ersten Zeile mit b_1 für die Konstante anfangen.
- Uuuuuund: In der Matrixschreibweise würde man B als Vector für die Regressionskoeffizienten nehmen, wobei die Konstante in der ersten Zeile steht.

8.2 Grundprinzip der Regressionsanalyse

- Die optimale Anpassung wird erreicht, indem man versucht, die Residuen zu minimieren.
- Das Minimierungsziel richtet sich auf die Summe der Quadratabweichungen (ordinary least square)

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \rightarrow minimal!$$

- Diese Bedingung ist erfüllt für:

$$Steigungskoeffizient\ (b, slope) = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}$$

$$Achsenabschnitt\ (a, intercept) = \overline{Y} - b\overline{X}$$

Einführung in Statistik
Modul 8: Zusammenhang metrischer Variablen

8.1 Grundbegriffe und Streudiagramme
8.2 Grundprinzip der Regressionsanalyse
8.3 Fehleranalyse bei der Regressionsanalyse
8.4 Regression mit R
8.5 Teststatistik
8.6 Varianzaufklärung
8.7 Kovarianz
8.8 Korrelation
8.9 Der standardisierte Regressionskoeffizient
8.10 Ergänzungen

1.2 Der Regressionskoeffizient b

Regressionskoeffizient b (aka Steigungskoeffizient) und r

$$b = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2} \qquad (1)$$

$$= \frac{\frac{\sum (X - \overline{X})(Y - \overline{Y})}{n}}{\frac{\sum (X - \overline{X})^2}{n}} \qquad (2)$$

$$= \frac{COV_{YX}}{VAR_X} \qquad (3)$$

$$= \frac{r_{YX} \cdot \cancel{s_X} \cdot s_Y}{s_X \cdot \cancel{s_X}} \qquad (4)$$

$$= r_{YX} \cdot \frac{s_y}{s_X} \qquad (5)$$

$$r_{YX} = b \cdot \frac{s_X}{s_Y} \qquad (6)$$

1.3 Der standardisierte Regressionskoeffizient

Der standardisierte Regressionskoeffizient BETA aka b*

$$BETA = b \cdot \frac{s_X}{s_Y} = r_{YX}$$

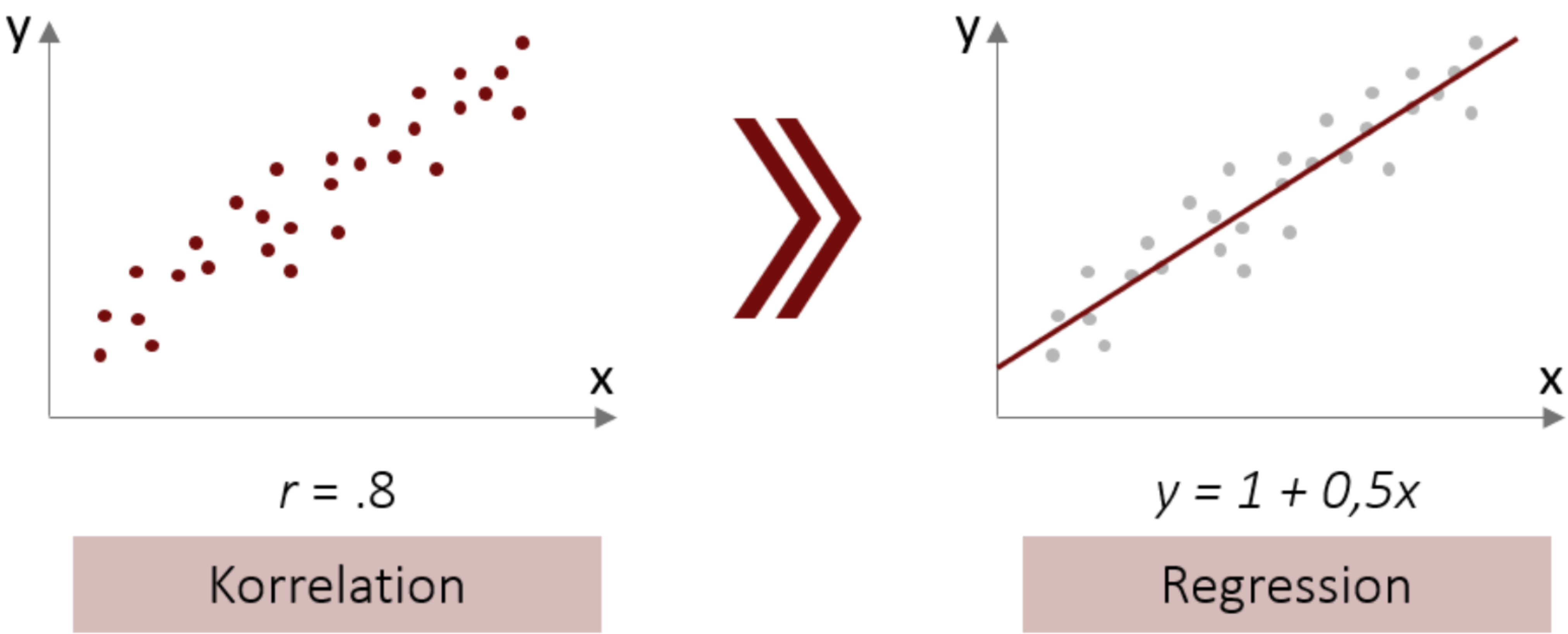
Beschreibung

Die standardisierten Regressionskoffizienten geben einen Zusammenhang in Standardabweichungen an: Wenn x um eine Standardabweichung grösser ist, um wie viele Standardabweichungen ist dann y grösser (kann negativ sein)?

Wie Korrelationen bzw. partielle Korrelationen

Die BETAs sind den Korrelationen sehr ähnlich: +1 ist ein perfekter positiver Zusammenhang, 0 kein Zusammenhang und -1 ein perfekter negativer Zusammenhang. Interpretieren würde ich ab 0.1, wenn sie signifikant sind.

Korrelation und Regression

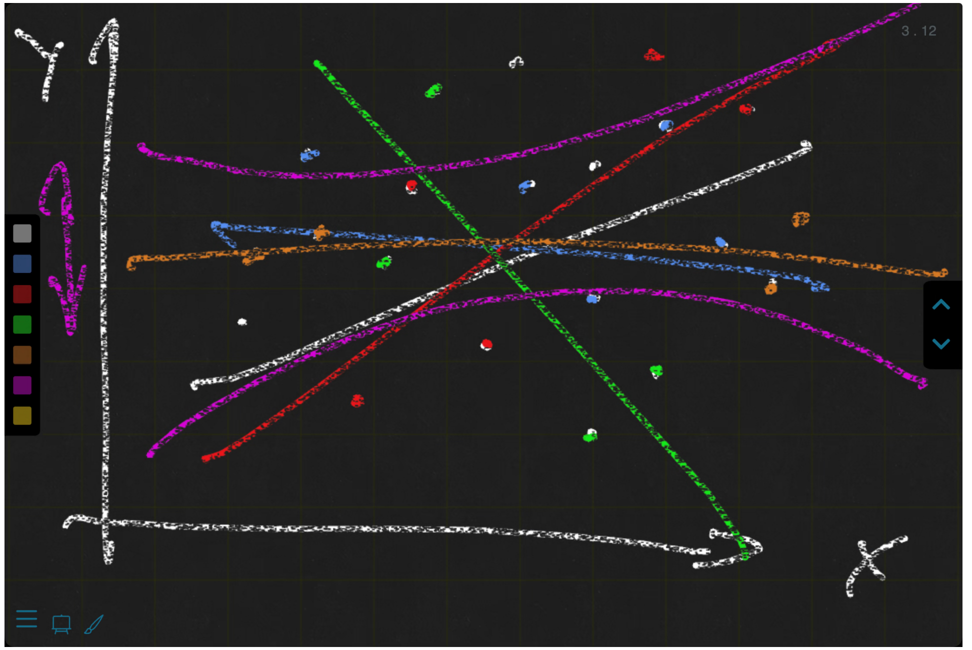


Von der Korrelation zur Regression und zurück

1.3.1 Standardfehler der b's

$$se_b^2 = \frac{s_e^2}{n \cdot s_2^2} \text{ mit } s_e^2 = \frac{1}{n-3} \sum e_i^2$$

Die Standardfehler der b's sind (bei sehr vielen Ziehungen) die «durchschnittliche» Abweichung der b's von dem wahren Wert β. Standardfehler kann man auch für die standardisierten Regressionskoeffizienten (BETA) berechnen.



Streuung von Regressionsgeraden

Varianzzerlegung der Gesamtmodellgüte

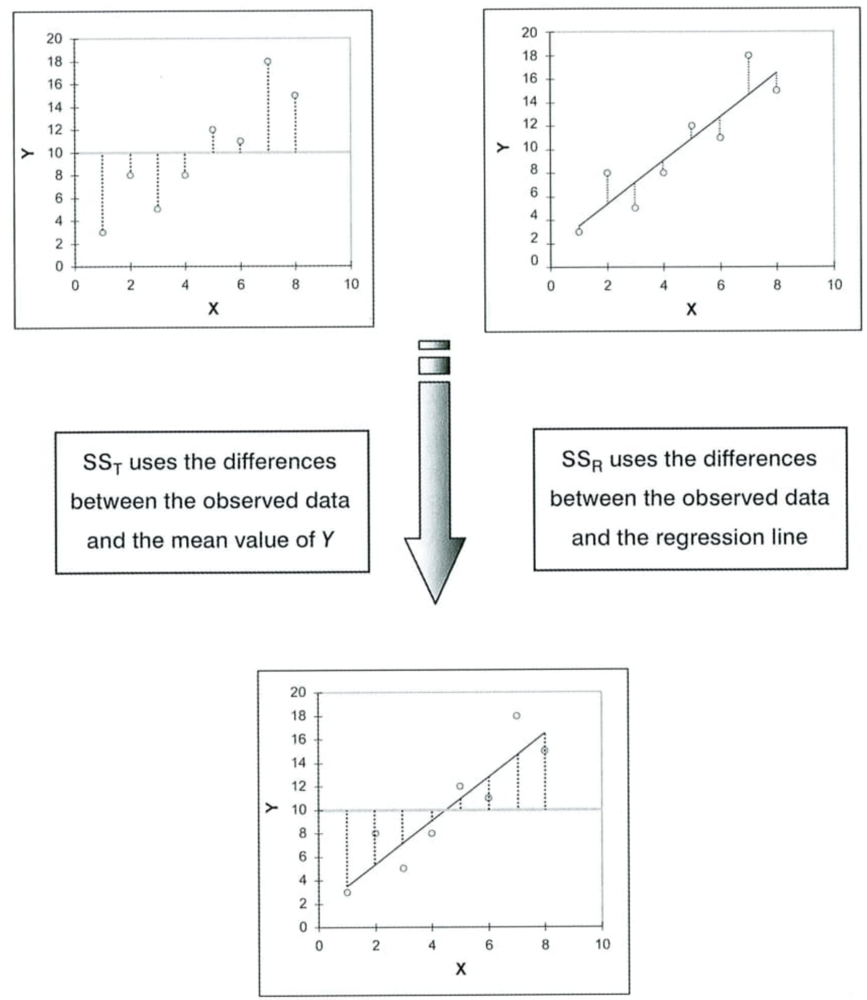
$$Y_i = \overline{Y} + e_i \quad (7)$$

$$Y_i = b_1 + b_2 X_i + e_i \quad (8)$$

$$\hat{Y}_i = b_1 + b_2 X_i \quad (9)$$

$$Y_i = \hat{Y}_i + e_i \quad (10)$$

$$SS_T = SS_R + SS_M \quad (11)$$



Summenzerlegung

1.4 Das Bestimmtheitsmass R^2

Was sagt das Bestimmtheitsmass R^2?

Das Bestimmtheitsmass R^2 gibt an, wie gut die Werte der AV durch die Werte der UV vorhergesagt werden können.

$R^2 = \frac{SS_M}{SS_T}$

- Anteil der erklärten Varianz der AV durch die UVs.
- SS_T: Summe der quadrierten Abweichungen total für die AV (Y).
- SS_M: Summe der quadrierten Abweichungen des Modells ($\hat{\text{Y}}$)
- $R^2 = \frac{\text{aufgeklärte Varianz}}{\text{Gesamtvarianz}}$

1.4.1 Bestimmtheistmass und Modellgüte

$$Y_i = \overline{Y} + e_i \tag{12}$$

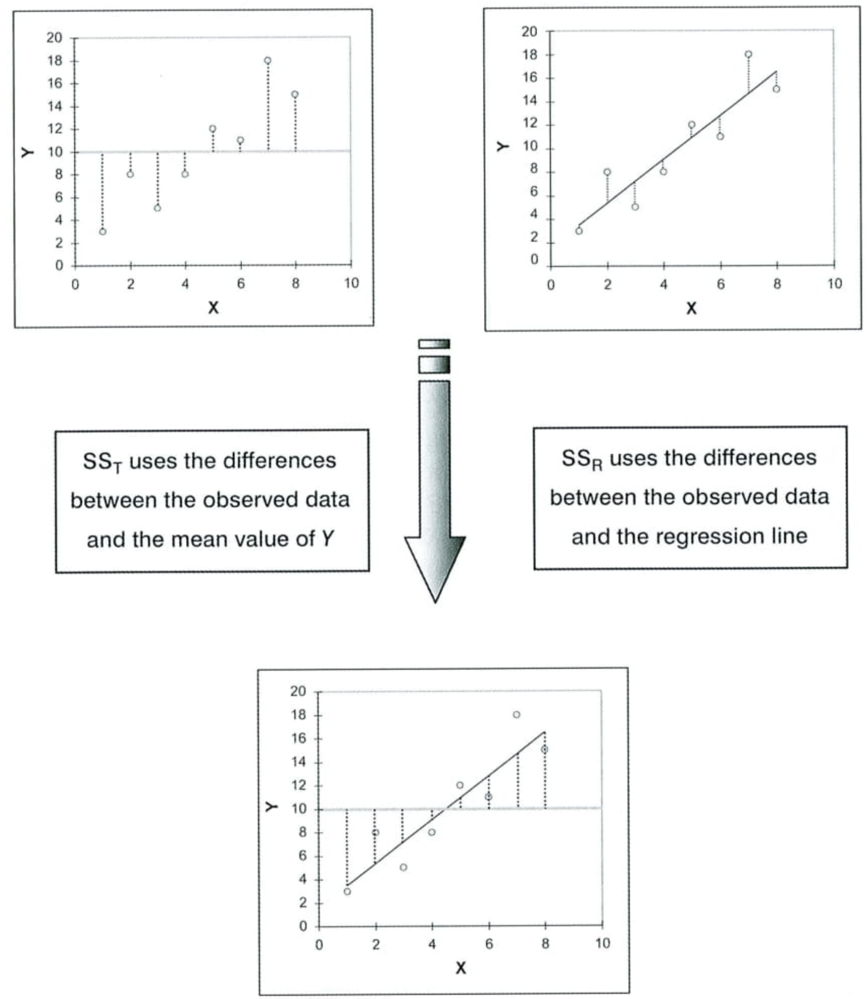
$$Y_i = b_1 + b_2 X_i + e_i \tag{13}$$

$$\hat{Y}_i = b_1 + b_2 X_i \tag{14}$$

$$Y_i = \hat{Y}_i + e_i \tag{15}$$

$$SS_T = SS_R + SS_M \tag{16}$$

$$R^2 = \frac{SS_M}{SS_T} \tag{17}$$



Summenzerlegung

1.4.2 Interpretation von R^2

R^2 ist Varianzaufklärung

Das Bestimmtheitsmass R^2 gibt an, wie viel Variant der AV durch die UV's aufgeklärt werden konnte. R^2 geht von 0 bis 1, bzw., wenn in Prozent ausgedrückt, von 0% bis 100%.

$R^2_{adj.}$

Korrigiertes $R^2 \cdot \frac{n-k-1}{n-1}$ bei kleinen Stichproben (wobei k die Anzahl UVs ist).

F-Test (R^2)

Gibt an, ob durch das Modell insgesamt überzufällig gut Varianz aufgeklärt wurde. Also, ob die Nullhypothese zurückgewiesen werden kann, dass die AV nicht durch sämtliche UVs im Modell erklärt werden kann.

1.4.3 Kennwerte von Regressionsanalysen – Signifikanz

t-Werte der b's oder standardisierten Regressionskoeffizienten (BETA)

Umrechnung der b's in t-Werte, die sich (bei gegebenem Stichprobenumfang bzw. den Degrees of Freedom) unter der Annahme der Nullhypothese ergeben. Sie sind innerhalb einer Regressionsanalyse vergleichbar. Sie sind für die b's und BETAS identisch.

p-Werte der b's bzw. BETAS (p oder sig.)

Geben die Wahrscheinlichkeit an, dass ein in einer Stichprobe gefundenes b zustandekommt, obwohl die Nullhypothese gilt. Ist auch für die b's und BETAS identisch. Bei $p < .05$ sprechen wir von einem von 0 signifikant verschiedenen b, wenn das Signifikanzniveau bei 95% liegt (5% Irrtumswahrscheinlichkeit).

Model Summary

R	0.343	RMSE	1.787
R-Squared	0.118	Coef. Var	40.845
Adj. R-Squared	0.078	MSE	3.193
Pred R-Squared	-0.050	MAE	1.333

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	9.375	1	9.375	2.936	0.1007
Residual	70.250	22	3.193		
Total	79.625	23			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper

2 Regression multivariat

2.1 Das Modell mit 2 UV's

Regressionsgleichung

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + U_i \tag{18}$$

$$Y_i = b_1 + b_2 X_{i2} + b_3 X_{i3} + \varepsilon_i \tag{19}$$

Regressionskoeffizienten b

Die Variablen sind alle im erhobenen Datensatz. Die Regressionsrechnung ermöglicht Aussagen darüber wie die Einflussgrössen wirken, welches Gewicht sie also haben. Jetzt müssen Regressionsgewichte bestimmt werden.

$$b_2 = \frac{r_{Y2} - r_{23} r_{Y3}}{(1 - R^2_{2.3})} \frac{S_y}{S_2} \tag{20}$$

2.2 Das Modell mit 3 UV's

Regressionsgleichung

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + U_i \tag{21}$$

$$Y_i = b_1 + b_2 X_{i2} + b_3 X_{i3} + b_4 X_{i4} + \varepsilon_i \tag{22}$$

3 Übung 1

3.1 Installieren Sie R und R-Studio (neu)

Eine Anleitung zur Installation finden Sie hier.

Erstellen Sie eine Quarto-Datei.qmd

1. Öffnen Sie R-Studio
2. In R-Studio ➪ File ➪ New File ➪ Quarto Document…
3. Klicken Sie unten links auf «Create Empty Document»
4. (Wählen Sie als `title` «Erste Regression»)
5. Fügen Sie einen r-Chunk hinzu mit diesem Schlater: 
6. speichern Sie an einem günstigen Ort
   (am besten in der Cloud + nicht auf Desktop)

Installieren Sie ein paar Pakete

Kopieren Sie in Ihre Datei:

```
install.packages("tidyverse", "sjmisc", "sjlabelled", "ggupbr", "corrr")

install.packages("devtools")

devtools::install_github("joon-e/soscisurvey")
```

3.2 Laden Sie die Daten

Laden Sie den Fragebogen hier runter und schauen ihn an

```r
suppressPackageStartupMessages(library(tidyverse))

Daten_import <- soscisurvey::read_sosci("https://www.soscisurvey.de/StatAufbau/?act=

DATEN <- Daten_import

DATEN  |>
  sjlabelled::get_label()
```

```
CASE

"Interview-Nummer (fortlaufend)"

QUESTNNR
                                                                                  "Fragebogen,
der im Interview verwendet wurde"

MODE

"Interview-Modus"

STARTED
```

"Zeitpunkt zu dem das Interview begonnen hat (Europe/Berlin)"

3.2.1 Schauen Sie sich zwei Variablen an

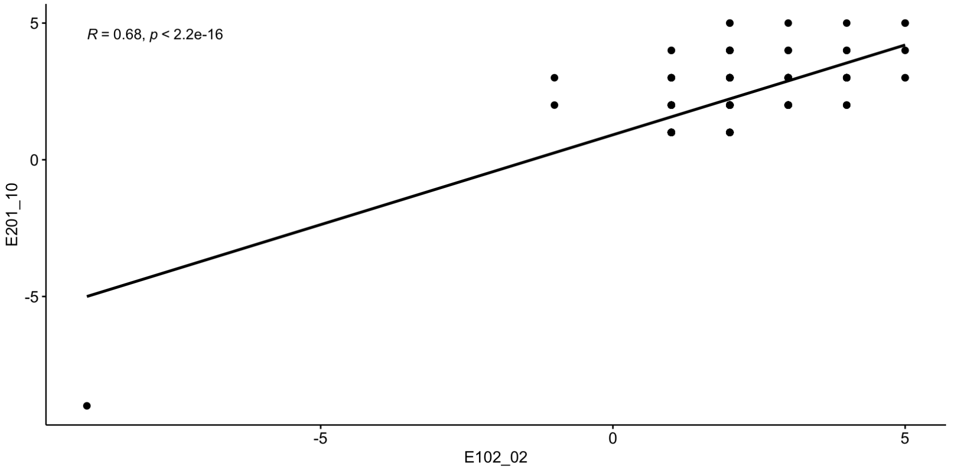
```r
DATEN   |>
  sjmisc::frq(E102_01, E201_10)
```

```
Statistik_Einführung: Statistik Einführung ist mir leichtgefallen. (E102_01) <integer>
# total N=167 valid N=167 mean=2.58 sd=1.40

Value |                      Label |  N | Raw % | Valid % | Cum. %
-------------------------------------------------------------------
   -9 |           nicht beantwortet |  1 |  0.60 |    0.60 |   0.60
   -2 |               keine Antwort |  0 |  0.00 |    0.00 |   0.60
   -1 |                 weiss nicht |  1 |  0.60 |    0.60 |   1.20
    1 | 1 trifft überhaupt nicht zu | 21 | 12.57 |   12.57 |  13.77
    2 |                           2 | 55 | 32.93 |   32.93 |  46.71
    3 |                           3 | 53 | 31.74 |   31.74 |  78.44
    4 |                           4 | 29 | 17.37 |   17.37 |  95.81
    5 |    5 trifft voll und ganz zu |  7 |  4.19 |    4.19 | 100.00
 <NA> |                        <NA> |  0 |  0.00 |    <NA> |   <NA>

Erwartungen STAT A: Ich möchte viel mehr über Statistik wissen. (E201_10) <integer>
# total N=167 valid N=167 mean=2.35 sd=1.37
```

3.2.2 Schauen Sie sich den Scatterplot an

1. Schauen Sie sich den Scatterplot an und die Correlation.
2. Was sehen Sie für ein Problem mit den Daten? Was tun?

```
DATEN  |>
  ggpubr::ggscatter(x = "E102_02", y =
    add = "reg.line", cor.coef = TRUE)
```



3.3 Rechnen Sie

```
DATEN  |>
  select(E102_02, E201_10)  |>
  cor()
```

```
          E102_02   E201_10
E102_02 1.0000000 0.6815875
E201_10 0.6815875 1.0000000
```

3.3.1 Rechnen Sie ein Regressionsmodell

```
Modell_1 <- lm(E201_10 ~ E102_02, dat
summary(Modell_1)
```

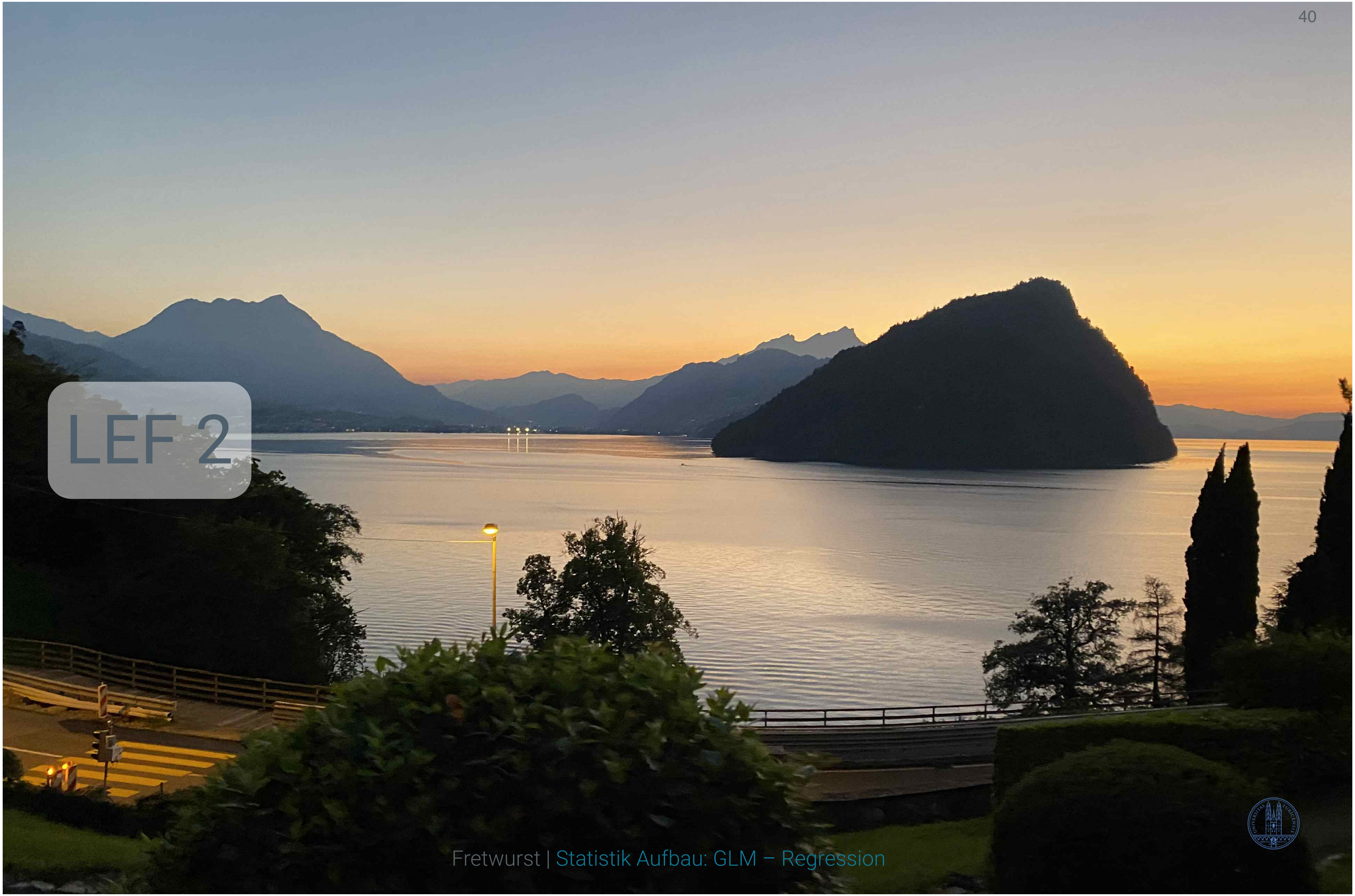
```
Call:
lm(formula = E201_10 ~ E102_02, data =
DATEN)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2726 -0.7684 -0.2641  0.7274  2.7274

Coefficients:
            Estimate Std. Error t value
Pr(>|t|)
(Intercept)  1.28105    0.16912   7.575
2.60e-12 ***
E102_02      0.49578    0.06668   7.435
5.73e-12 ***
---
```

3.3.2 Verändern Sie das Regressionsmodell

Kopieren Sie den r-Chunk der letzten Folie und setzen Sie andere Variablen ein: Nehmen Sie die Variablen für «Statistik Einführung hat mir Spass gemacht» und erklären Sie damit: «Ich freu mich auf Statistik Aufbau!».

Beantworten Wie wieder die Fragen:

1. Wie gross ist R^2?
2. Wie gross ist die bivariate Korrelation r? (selbst ausrechnen)
3. Ist der Zusammenhang positiv oder negativ?
4. Ist der Zusammenhang signifikant?


LEF 2

Essayfragen 2

E2.1 Wie ist die Korrelation definiert?

E2.2 Was ist das Analyseziel einer Regression?

E2.3 Wie sind die Regressionskoeffizienten gekennzeichnet? (Welcher Buchstabe)

E2.4 Was ist der Unterschied zwischen BETA und β?

E2.5 Was drückt der Standardfehler der Regressionskoeffizienten b aus?

E2.6 Mit welchen Kennwerten kann die Modellgüte insgesamt bewertet werden?

E2.7 Was ist a) $R_a^2dj.$ und b) wann würde man es verwenden?

E2.8 Was sagt die Signifikanz des F-Tests für ein Regressionsmodell aus?

MC-Fragen 2

MC 2.2.

MC 2.2: Sind folgende Aussagen richtig oder falsch?

richtig	falsch	Aussagen
○	○	Die Regressionskoeffizienten (b’s) sind unstandardisiert.
○	○	Die standardisierten Regressionskoeffizienten messen die unbekannten Parameter {tex`eta`}.
○	○	Die standardisierten Regressionskoeffizienten sind wie Korrelationen interpretierbar.
○	○	Die Standardfehler der b’s sind immer 1.

Punkte: 0

MC 2.2.

MC 2.2: Sind folgende Aussagen richtig oder falsch?

richtig	falsch	Aussagen
○	○	Das Bestimmtheitsmass R^2 gibt an, welcher Varianzanteil der AV durch die UV bzw. die UVs erklärt werden kann.
○	○	R^2_adj. wird nur bei sehr grossen Stichproben gebraucht, um zufällige Signifikanzen zu vermeiden.
○	○	Der F-Test zum R testet, ob alle UVs signifikant sind.
○	○	Wenn R^2 kleiner als .05 ist, dann ist die Regression signifikant.

Punkte: 0

MC 2.3.

MC 2.3: Sind folgende Aussagen richtig oder falsch?

richtig	falsch	Aussagen
○	○	Die standardisierten Regressionskoeffizienten entsprechen bei der bivariaten Regression der Korrelation.
○	○	Die Regressionskoeffizienten b liegen immer zwischen -1 und 1.
○	○	Das Bestimmtheitsmass R^2 gibt den Prozentanstieg der Regressionsgeraden an.
○	○	Die Konstante in der Regressionsgleichung wird bei multivariaten Modellen auch als b gekennzeichnet.

Punkte: 0

MC 2.4.

MC 2.4: Sind folgende Aussagen richtig oder falsch?

richtig	falsch	Aussagen
○	○	Bei der multivariaten Regression hängt b auch von der Covarianz der UVs ab
○	○	Wenn es zwei UVs gibt, spricht man schon von «multivariat»
○	○	Wenn ein b signifikant ist, wird auch R^2 des Gesamtmodells signifikant.
○	○	R^2 gibt an, wie viel Prozent der Varianz der AV durch alle UVs zusammen erklärt werden können.

Punkte: 0

MC 2.5.

MC 2.5: Sind folgende Aussagen richtig oder falsch?

richtig	falsch	Aussagen
○	○	R-Studio ist eine Programmiersprache für statistische Analysen.
○	○	R ist eine Benutzer:innenoberfläche für die Programmiersprache R-Studio
○	○	tidyverse ist eine Sammlung von R-Paketen.) signifikant.
○	○	Wenn man Drittvariablen «herausrechnet», werden die anderen b's immer kleiner.

Punkte: 0

MC 2.6.

MC 2.6: Sind folgende Aussagen richtig oder falsch?

richtig	falsch	Aussagen
○	○	Wenn man Drittvariablen «herausrechnet», werden die anderen b's immer kleiner.
○	○	Für die standardisierten Regressionskoeffizienten können auch Konfidenzintervalle angegeben werden.
○	○	Die Standardfehler von BETAS sind immer 1.
○	○	Regressionen sind eine Form der GLM.

Punkte: 0

Für LEF 2: 0 von 12 Punkten, was 0% und etwa einer 1 entspricht.

Take Home

- Elemente der Regressionsanalyse
- Modell: R^2, $R^2_{adj.}$ F-Wert, F-Test, p-Wert
- UVs: B's, se_b, BETA's, t-Wert, p-Wert, TOL, VIF
- Mit Regressionsanalysen können die linearen Effekte von UVs auf eine AV berechnet werden, wobei die Effekte der UVs gegenseitig kontrolliert werden.
- Jedes B steht für eine Hypothese.
- Bei der multivariaten Regression werden "Scheinkorrelationen" herausgerechnet.

Ausblick

GLM – BLUE

- Voraussetzungen für OLS-Regressionen
- Wann sind die b «Best Linear Unbiased Estimators»
- Wie mit Verletzungen der Grundannahmen umgehen

Heteroskedastizität

Die Modellschätzung ist nicht für alle Fälle gleich gut.

Multikollinearität

Die UVs sind stark miteinander verbunden (aka konfundiert).