FORSCHEN.
FÖRDERN.
VERNETZEN.
DEUTSCHE GESELLSCHAFT FÜR
FREMDSPRACHENFORSCHUNG

DGFF-Kolleg: Regressionsanalyse

17. Mai 2024

Bastian Fuchs

bastian.fuchs@rub.de

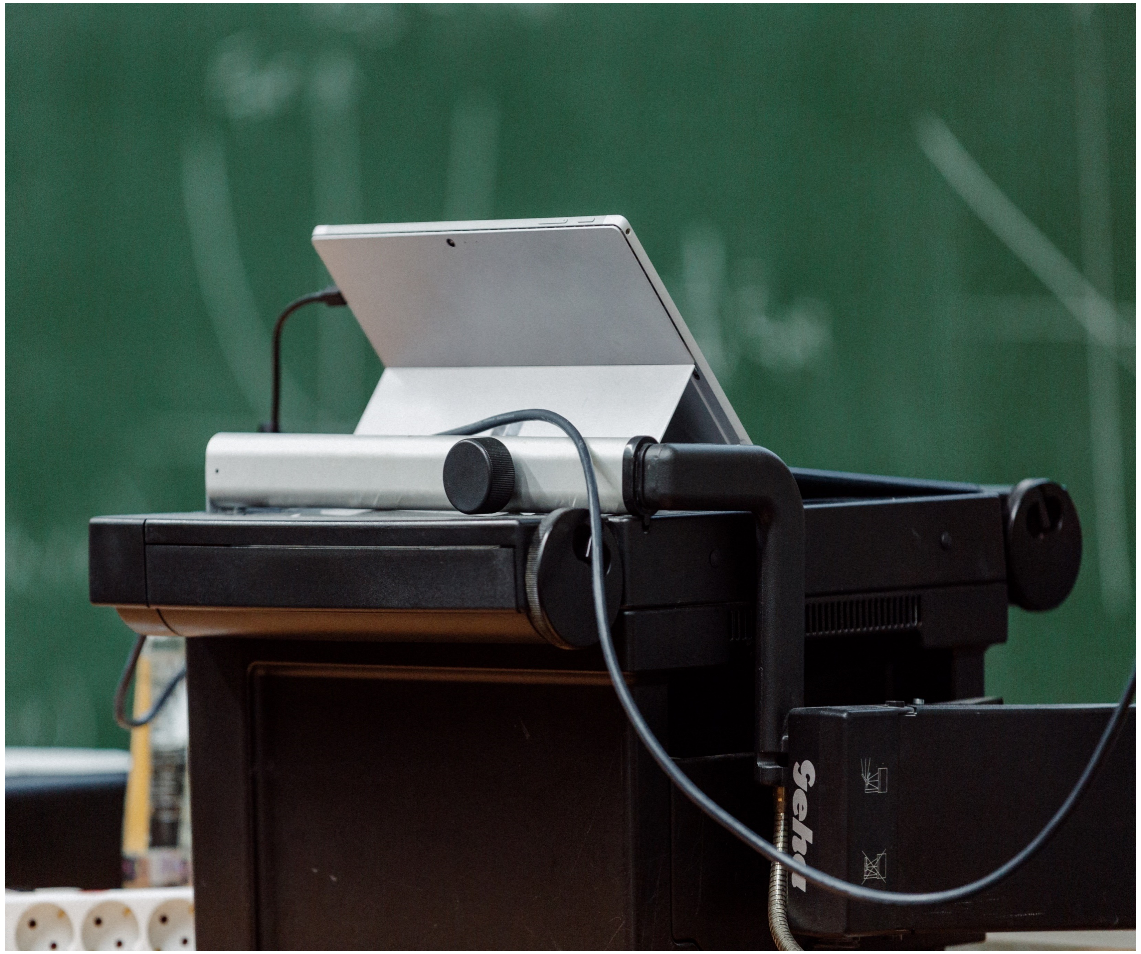


Foto: © Universität Bielefeld

Übersicht

- Einfache lineare Regression
- Regressionsgleichung
- Methode der kleinsten Quadrate
- Determinationskoeffizient R^2
- Standardisiertes Regressionsgewicht *ß*
- Ausblick: Multiple Regression

Vorbereitung

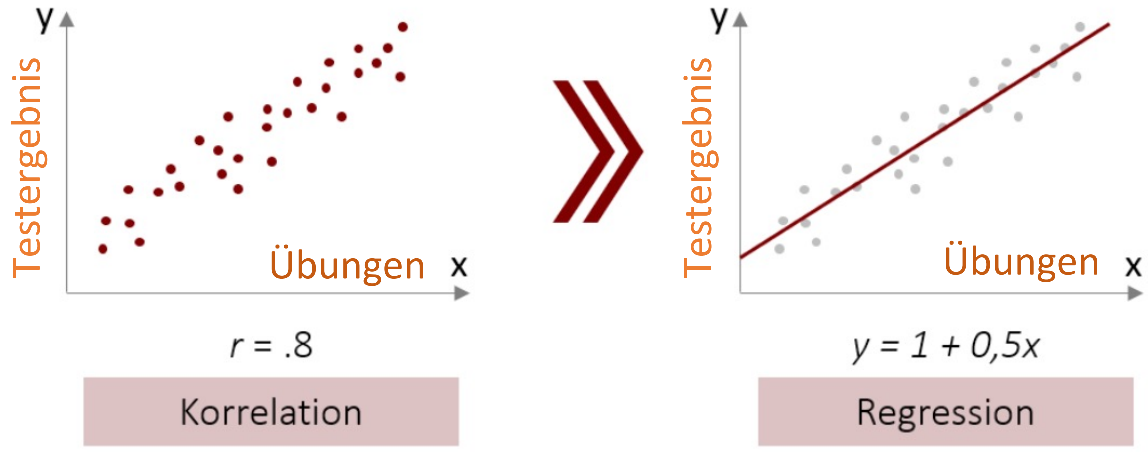
- Kapitel 13 (S. 209-214) in Hanna & Dempster (2017)
- 4.3. Einfache lineare Regression (S. 105-118) in Rasch et al. (2021)

- Video zur Inferenzstatistik (DGFF-Kolleg am 23.06.2023, Prof. Dr. Dominik Rumlich)

→ *Inferenzstatistik: Zusammenhänge (Regression)* (ab 23:24 Min. / Folie 33)

Einfache lineare Regression

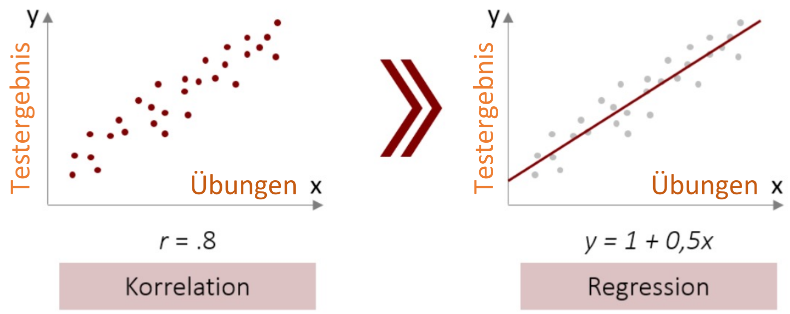
- **Korrelation** (Zusammenhang) → **Regression** (Vorhersage)



Bildquelle 1

- **Korrelation:** Es besteht ein starker positiver Zusammenhang zwischen der Anzahl der Übungen und den Testergebnissen der SuS.
- **Regression:** Schüler A hat 5 Übungen gemacht und wird im Test 70 Punkte bekommen (Vorhersage).

Einfache lineare Regression

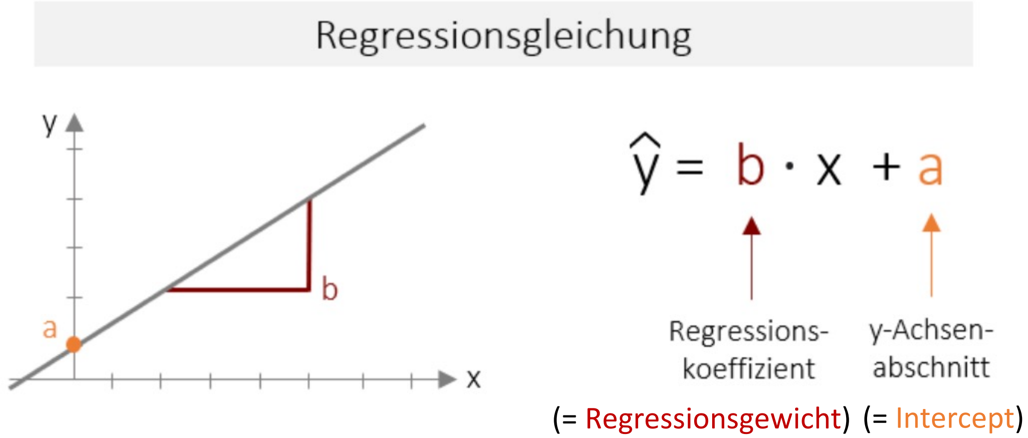


Bildquelle 1

- **Ziel:** “auf der Grundlage von Korrelationen die bestmögliche Vorhersage für eine Variable zu bestimmen“ (Sedlmeier & Renkewitz, 2018, S. 246)
- Eine Variable (*Kriterium*) wird durch eine andere Variable (*Prädiktor*) vorhergesagt
- Abhängigkeit einer Variable von der anderen bzw. Einfluss einer Variable auf die andere Variable
- Warum *einfache* Regression? → nur 1 Prädiktor
- Warum *lineare* Regression? → linearer Zusammenfang zw. Prädiktor und Kriterium (Rasch et al., 2021, S. 105)

Regressionsgleichung

- **Ziel:** eine Variable (*Kriterium*) durch eine andere Variable (*Prädiktor*) vorhersagen
- Diese Vorhersage basiert auf einer **Regressionsgleichung**



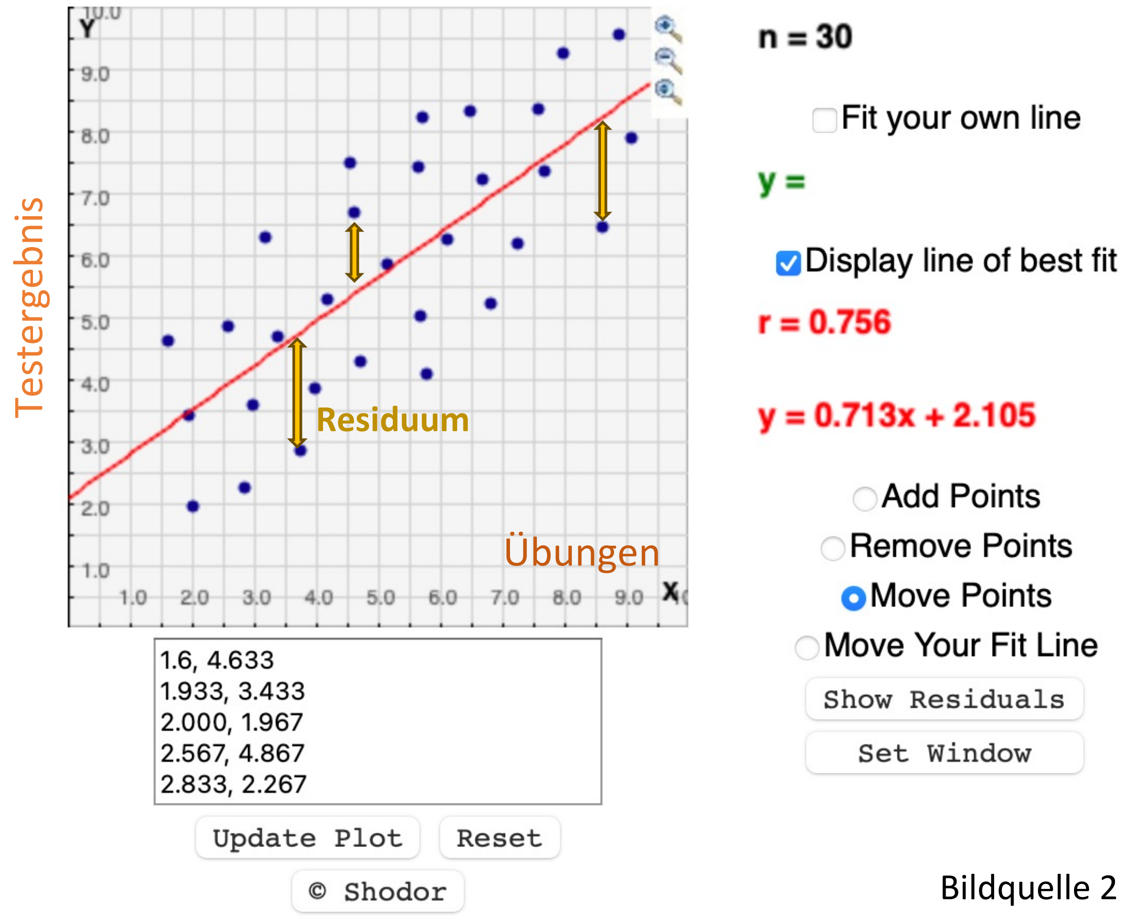
Bildquelle 1

„Das Ziel dieses Verfahrens ist es, den stochastischen Zusammenhang zwischen zwei Variablen durch eine **lineare Funktion** wiederzugeben. Anschaulich bedeutet dies, dass die Punktewolke durch eine **einzige, möglichst repräsentative Gerade** ersetzt wird. Dies gelingt natürlich umso besser, je enger die Punktewolke ist, also je höher die beiden Merkmale tatsächlich miteinander korrelieren."

Rasch et al., 2021, S. 106

- Testergebnis $_{\text{Schüler A}}$ = b * Anzahl der Übungen $_{\text{Schüler A}}$ + a

Methode der kleinsten Quadrate



Bildquelle 2

- Wie entsteht die Regressionsgleichung?
- Residuen müssen möglichst klein sein
- **Methode der kleinsten Quadrate**:

Minimierung der Summe der (quadrierten) Residuen

tatsächlicher Messwert — vorhergesagter Wert

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

Bildquelle 1

Determinationskoeffizient R^2

Determinations-koeffizient → $$\mathrm{R}^2 = r^2_{(x,y)} = \frac{s^2_{\hat{y}}}{s^2_y}$$

Regressionsvarianz „erklärte Varianz“

Gesamtvarianz

quadrierte Korrelation der Variablen x und y

Bildquelle 1

- Ein Gütemaß für die Vorhersage: Wie viel Varianz (Streuung) erklärt der Prädiktor?
- Prozentualle Varianzerklärung einer Variable durch die andere
- Wird in empirischen Studien berichtet

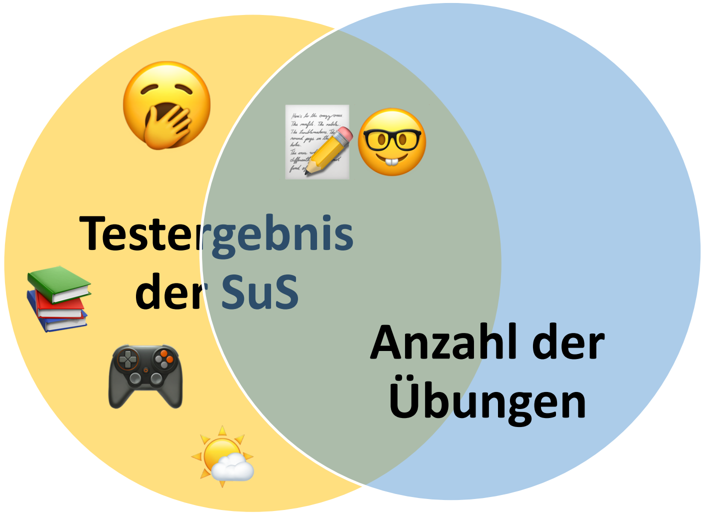
Nicht-erklärte Varianz

R^2 = 0.85 → 85% der Varianz in den Testergebnissen geht auf die Anzahl der Übungen zurück

Gesamtvarianz

Erklärte Varianz durch den Prädiktor

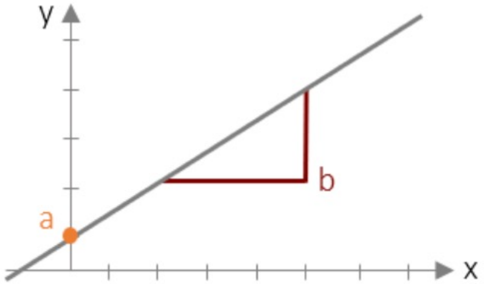
R^2 = 0.55 → 55% der Varianz in den Testergebnissen geht auf die Anzahl der Übungen zurück



Standardisiertes Regressionsgewicht ß

- Zur Erinnerung:

Regressionsgleichung



$$\hat{y} = b \cdot x + a$$

Regressions-koeffizient

y-Achsen-abschnitt

Problem: abhängig von der Skala

Bildquelle 1

- Lösung: Standardisierung

standardisierter Regressionskoeffizient (β-Gewicht)

$$\beta_{yx} = b \cdot \frac{s_x}{s_y} = \frac{\mathrm{cov}(xy)}{s_x^2} \cdot \frac{s_x}{s_y} = \frac{cov_{(xy)}}{s_x \cdot s_y}$$

Regressions-koeffizient

Kovarianz

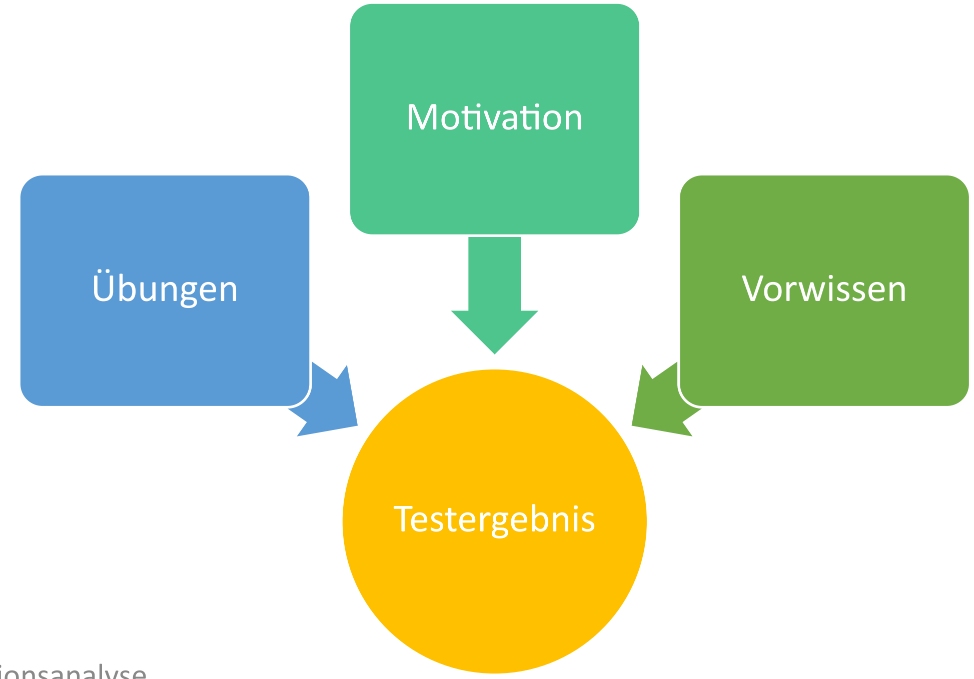
Standardabweichung von y

Standardabweichung von x

Bildquelle 1

Ausblick: Multiple Regression

- Bisher: *einfache* lineare Regression → nur 1 Prädiktor
- Erweiterung: *multiple* lineare Regression → mehrere Prädiktoren
- Testergebnis $_{\text{Schüler A}}$ = b_1 * Anzahl der Übungen $_{\text{Schüler A}}$ + b_2 * Motivation $_{\text{Schüler A}}$ + b_3 * Vorwissen $_{\text{Schüler A}}$ + a



Literaturangaben

- Hanna, D., & Dempster, M. (2017). *Statistik für Psychologen*. WILEY.
- Planing, P. *Regression*. https://statistikgrundlagen.de/ebook/chapter/regression/
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2021). *Quantitative Methoden 1. Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Auflage). Springer.
- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler*. Pearson.

Bildquellen:

1. Planing, P. *Regression*. https://statistikgrundlagen.de/ebook/chapter/regression/
2. http://www.shodor.org/interactivate/activities/Regression/