Lernzettel

Zusammenhangsmessung: Korrelationen, Streuung und einfache Regressionsanalyse

Universität: Technische Universität Berlin
Kurs/Modul: Statistik I für Wirtschaftswissenschaften
Erstellungsdatum: September 19, 2025

Lernzettel: Statistik I für Wirtschaftswissenschaften – Zusammenhangsmessung: Korrelationen, Streuung und einfache Regressionsanalyse

(1) Streuung und zentrale Maße. Streuung beschreibt, wie stark die Daten um ihren Mittelpunkt streuen. Wichtige Größen sind Varianz und Standardabweichung.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2, \qquad \sigma = \sqrt{\sigma^2}.$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2, \qquad s = \sqrt{s^2}.$$

(2) Zusammenhangsmessung: Korrelationen. Der Pearson-Korrelationskoeffizient misst den linearen Zusammenhang zwischen zwei Variablen X und Y.

$$r = \frac{\mathrm{cov}(X,Y)}{s_X s_Y}.$$

$$r = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2 \ \sum_{i=1}^{n}(y_i-\overline{y})^2}}.$$

Interpretation:

$$-1 \le r \le 1, \quad r > 0 \text{ positiver Zusammenhang, } r < 0 \text{ negativer Zusammenhang.}$$

Je näher |r| an 1 liegt, desto stärker ist der lineare Zusammenhang; je näher an 0, desto schwächer.

Zusatz: Spearman-Rangkorrelation (falls Rangordnung statt Linearität relevant):

$$\rho = \text{Korrelation der Ränge von } (X_i), (Y_i).$$

(3) Einfache Regressionsanalyse. Ziel: Vorhersage einer abhängigen Variable Y aus der unabhängigen Variable X.

Zwei-Gleichungs-Modell:

$$Y_i = a + bX_i + \varepsilon_i,$$

mit Residuen ε_i.

Die Parameter werden so gewählt, dass die quadratischen Abweichungen minimiert werden (Kleinste-Quadrat-Methode).

Berechnung der Koeffizienten:

$$b = \frac{S_{xy}}{S_{xx}}, \qquad a = \overline{Y} - b\overline{X},$$

mit

$$S_{xy} = \sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}), \qquad S_{xx} = \sum_{i=1}^{n}(X_i-\overline{X})^2.$$

Vorhersagen und Residuen:

$$\hat{Y}_i \;=\; a \;+\; bX_i, \qquad e_i \;=\; Y_i \;-\; \hat{Y}_i.$$

Güte der Anpassung (Bestimmtheitsmaß):

$$SST \;=\; \sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$

$$SSR \;=\; \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2, \qquad SSE \;=\; \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2,$$

$$R^2 \;=\; \frac{SSR}{SST} \;=\; 1 - \frac{SSE}{SST}.$$

(4) Voraussetzungen und Interpretation. Für eine sinnvolle einfache Regression gelten grob: - Linearität der Beziehung zwischen X und Y. - Unabhängige Beobachtungen. - Zufallsfehler (ε_i) ohne systematische Muster. - Homoskedastizität: Varianz der Residuen konstant über X. - Normalverteilung der Residuen (für Tests).

Interpretation der Ergebnisse: - Die Regressionsgerade beschreibt die mittlere Veränderung von Y bei einer Änderung von X. - R^2 gibt an, wie viel der Varianz von Y durch X erklärt wird. - Korrelation zeigt Richtung und Stärke des Zusammenhangs, aber keine Kausalität.

(5) Vorgehen in einer typischen Analysepraxis. 1) Deskriptive Voranalyse: Streuung, Mittelwerte, Scatterplot von X gegen Y. 2) Berechne Pearson-Korrelation r und prüfe Richtung/Stärke. 3) Schätze die einfache Regressionsgerade: Bestimme a und b. 4) Prüfe Güte der Anpassung (R^2)*undResiduenmuster.*5)*BerichteInterpretationenimwirtschaftswissenschaftlichenKontext*

Zusammenfassung: - Streuungsmmaße geben Auskunft über die Verteilung der Daten. - Korrelationen messen lineare Zusammenhänge zwischen zwei Größen. - Die einfache Regression modelliert eine lineare Abhängigkeit und ermöglicht Vorhersagen sowie die Quantifizierung des Anteils der Varianz, der erklärt wird.