Einfache lineare Regression: Signifikanztests

statistik-online.ch

Bei der **einfachen linearen Regression** kann man testen, ob die Regressionsparameter statistisch bedeutsam von 0 abweichen. Besonders wichtig sind Tests für die Steigung b_1, den Achsenabschnitt b_0 und das Bestimmtheitsmass R^2.

1 Grundidee

Signifikanztests prüfen, ob ein beobachteter Zusammenhang stark genug ist, um H_0 zu verwerfen.

Die Regressionsgleichung lautet:

$$\hat{y} = b_0 + b_1 x$$

- b_0: Achsenabschnitt.
- b_1: Steigung.
- R^2: Anteil der Streuung von y, der durch das Modell erklärt wird.

Wichtig: In diesem Dokument wird ausschliesslich mit kritischen Werten entschieden. P-Werte werden nicht verwendet.

2 Fehlerstreuung und Standardfehler

Signifikanztests in der Regression beruhen auf den Residuen.

$$e_i = y_i - \hat{y}_i$$

Die Summe der quadrierten Residuen heisst SSE:

$$SSE = \sum e_i^2$$

Der Residualstandardfehler ist:

$$s_e = \sqrt{\frac{SSE}{n-2}}$$

Symbol	Bedeutung
n	Stichprobengrösse
$n-2$	Freiheitsgrade der einfachen linearen Regression
SSE	Summe der quadrierten Residuen
s_e	typische Abweichung der Punkte von der Regressionsgerade

3 t-Test für die Steigung b_1

Der wichtigste Test in der einfachen linearen Regression prüft, ob die Steigung von 0 verschieden ist.

Hypothesen

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Inhaltlich:

- H_0: Es gibt keinen linearen Zusammenhang zwischen x und y.
- H_1: Es gibt einen linearen Zusammenhang zwischen x und y.

Teststatistik

$$t = \frac{b_1 - 0}{SE(b_1)}$$

$$SE(b_1) = \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}$$

$$df = n - 2$$

Entscheidungsregel

Bei einem zweiseitigen Test:

$$H_0 \text{ verwerfen, wenn } |t| > t_{krit}$$

Der kritische Wert t_{krit} wird mit $df = n - 2$ aus der t-Tabelle gelesen.

4 t-Test für den Achsenabschnitt b_0

Der Achsenabschnitt kann ebenfalls getestet werden. Dieser Test ist aber nur dann inhaltlich interessant, wenn $x = 0$ sinnvoll interpretierbar ist.

Hypothesen

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

Teststatistik

$$t = \frac{b_0 - 0}{SE(b_0)}$$

$$SE(b_0) = s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$$

$$df = n - 2$$

Entscheidungsregel

$$H_0 \text{ verwerfen, wenn } |t| > t_{krit}$$

5 F-Test für R^2

Der F-Test prüft, ob das Regressionsmodell insgesamt erklärende Kraft hat. Bei der einfachen linearen Regression ist dieser Test eng mit dem t-Test für b_1 verbunden.

Hypothesen

$$H_0 : R^2_{\text{Population}} = 0$$

$$H_1 : R^2_{\text{Population}} > 0$$

Inhaltlich ist das bei der einfachen linearen Regression gleichbedeutend mit:

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

Teststatistik

Allgemein gilt bei k Prädiktoren:

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$

Bei der einfachen linearen Regression gibt es nur einen Prädiktor, also $k = 1$:

$$F = \frac{R^2}{1 - R^2} \cdot (n - 2)$$

Die Freiheitsgrade sind:

$$df_1 = 1 \qquad df_2 = n - 2$$

Entscheidungsregel

$$H_0 \text{ verwerfen, wenn } F > F_{krit}$$

Der kritische Wert F_{krit} wird mit $df_1 = 1$ und $df_2 = n - 2$ aus der F-Tabelle gelesen.
Merksatz: Bei der einfachen linearen Regression gilt:

$$F = t_{b_1}^2$$

6 Beispiel

Eine Lehrperson untersucht, ob Lernzeit x Prüfungspunkte y vorhersagen kann.

Person	x	y	$\hat{y}$	$e = y - \hat{y}$
1	1	52	52.6	-0.6
2	2	56	56.3	-0.3
3	3	61	60.0	1.0
4	4	65	63.7	1.3
5	5	66	67.4	-1.4

Aus der Regression wurden berechnet:

$$\hat{y} = 48.9 + 3.7x$$

$$n = 5 \qquad df = n - 2 = 3 \qquad \sum(x_i - \bar{x})^2 = 10$$

$$SSE = (-0.6)^2 + (-0.3)^2 + 1.0^2 + 1.3^2 + (-1.4)^2 = 5.10$$

$$s_e = \sqrt{\frac{5.10}{3}} = 1.30$$

Test für b_1

$$SE(b_1) = \frac{1.30}{\sqrt{10}} = 0.41$$

$$t = \frac{3.7}{0.41} = 8.98$$

Bei $\alpha = 0.05$, zweiseitig, und $df = 3$ gilt:

$$t_{krit} = 3.182$$

$$|8.98| > 3.182$$

Entscheidung: H_0 wird verworfen. Die Steigung ist statistisch signifikant von 0 verschieden.

Test für b_0

$$SE(b_0) = 1.30\sqrt{\frac{1}{5} + \frac{3^2}{10}} = 1.37$$

$$t = \frac{48.9}{1.37} = 35.69$$

$$|35.69| > 3.182$$

Entscheidung: H_0 wird verworfen. Der Achsenabschnitt ist statistisch signifikant von 0 verschieden.
Hinweis: Inhaltlich ist dieser Test nur sinnvoll, wenn $x = 0$ eine sinnvolle Bedeutung hat.

F-Test für R^2

Aus den Daten ergibt sich:

$$R^2 = 0.964$$

$$F = \frac{0.964}{1 - 0.964} \cdot 3 = 80.33$$

Bei $\alpha = 0.05$, $df_1 = 1$ und $df_2 = 3$ gilt:

$$F_{krit} = 10.13$$

$$80.33 > 10.13$$

Entscheidung: H_0 wird verworfen. Das Modell erklärt statistisch signifikant Streuung in y.

7 Zusammenfassung

Test	H_0	**Statistik**	**Entscheidung**
Steigung	$\beta_1 = 0$	$t = b_1/SE(b_1)$	$\lvert t\rvert > t_{krit}$
Achsenabschnitt	$\beta_0 = 0$	$t = b_0/SE(b_0)$	$\lvert t\rvert > t_{krit}$
Modell / R^2	$R^2 = 0$	$F = \frac{R^2}{1-R^2}(n-2)$	$F > F_{krit}$

Aufgaben: Signifikanztests in der Regression

1 – t-Test für b_1

Eine Regression mit $n = 30$ ergibt:

$$b_1 = 1.80 \qquad SE(b_1) = 0.45$$

Es gilt $df = 28$, $\alpha = 0.05$, zweiseitig und $t_{krit} = 2.048$.

a) Formulieren Sie H_0 und H_1.

b) Berechnen Sie die Teststatistik.

c) Entscheiden Sie mit t_{krit}.

2 – t-Test für b_0

Eine Regression ergibt:

$$b_0 = 12 \qquad SE(b_0) = 4$$

Es gilt $df = 28$, $\alpha = 0.05$, zweiseitig und $t_{krit} = 2.048$.

a) Berechnen Sie die Teststatistik.

b) Entscheiden Sie mit t_{krit}.

c) Wann ist dieser Test inhaltlich sinnvoll?

3 – F-Test für R^2

Eine einfache lineare Regression mit $n = 30$ ergibt:

$$R^2 = 0.36$$

Es gilt $\alpha = 0.05$, $df_1 = 1$, $df_2 = 28$ und $F_{krit} = 4.20$.

a) Berechnen Sie F.

b) Entscheiden Sie mit F_{krit}.

c) Interpretieren Sie das Ergebnis.

4 – Multiple Choice

Welche Aussagen sind korrekt? (Mehrfachantworten möglich)

- ❑ Der t-Test für b_1 prüft, ob die Steigung von 0 verschieden ist.
- ❑ Der F-Test für R^2 wird mit $F > F_{krit}$ entschieden.
- ❑ Ein signifikanter Achsenabschnitt ist immer inhaltlich wichtig.
- ❑ Bei einfacher linearer Regression gilt $F = t_{b_1}^2$.
- ❑ Wenn $R^2 = 0.36$, erklärt das Modell 36% der Streuung von y.

Lösungen: Signifikanztests in der Regression

Lösung zu Aufgabe 1

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

$$t = \frac{1.80}{0.45} = 4.00$$

$$|4.00| > 2.048$$

Entscheidung: H_0 wird verworfen. Die Steigung ist statistisch signifikant von 0 verschieden.

Lösung zu Aufgabe 2

$$t = \frac{12}{4} = 3.00$$

$$|3.00| > 2.048$$

Entscheidung: H_0 wird verworfen. Der Achsenabschnitt ist statistisch signifikant von 0 verschieden.
Inhaltlich sinnvoll: Der Test ist vor allem dann sinnvoll, wenn $x = 0$ eine reale und interpretierbare Bedeutung hat.

Lösung zu Aufgabe 3

$$F = \frac{0.36}{1 - 0.36} \cdot (30 - 2) = \frac{0.36}{0.64} \cdot 28 = 15.75$$

$$15.75 > 4.20$$

Entscheidung: H_0 wird verworfen. Das Modell erklärt statistisch signifikant Streuung in y.

Lösung zu Aufgabe 4

✓ Richtig – Der t-Test für b_1 prüft die Steigung.

✓ Richtig – Beim F-Test wird rechtsseitig entschieden.

✗ Falsch – Der Achsenabschnitt ist nur sinnvoll interpretierbar, wenn $x = 0$ sinnvoll ist.

✓ Richtig – In der einfachen linearen Regression gilt $F = t_{b_1}^2$.

✓ Richtig – $R^2 = 0.36$ entspricht 36% erklärter Streuung.