Einführung in die Empirische Wirtschaftsforschung

Übungsaufgaben 5

Konfidenzintervall - Hypothesentest

1. Betrachten Sie folgendes lineares Modell

$$Y_i = 0.2 + 1.2x_i + \epsilon_i \quad \forall\, i = 1, \ldots, N,$$

wobei

$$\epsilon_i \overset{i.i.d.}{\sim} Normal(0, \sigma^2).$$

a) Beschreiben Sie in eigenen Worten die Begriffe Konfidenzintervall, Hypothesentest, p-Wert und Güte in dem Kontext des obigen linearen Modelles.

b) Sie sind an dem Hypothesentest $H0: \ \beta_1 = 1$ gegen $H1: \beta_1 \neq 1$ auf dem $\alpha = 0.05$ Niveau interessiert. Was ist die Güte des Tests, wenn $N = 150$, $\sigma^2 = 3$ und

$$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = 0.33?$$

c) Welche Annahmen müssten Sie ändern, um eine höhere Güte zu erhalten?

d) (In R) Simulieren Sie $S = 1000$ Realisationen von b_1, wenn $N = 150$ und $\sigma^2 = 3$ (wie in b)). Generieren Sie die $x-$Werte einmalig mittels einer $Uniform(-1, 1)-$Verteilung. Setzten Sie dabei den Seed auf 747. Stellen Sie die Verteilung von b_1 in einem Histogramm dar. Was fällt Ihnen auf?

e) In der Simulation von d) zählen Sie wie oft von den $S = 1000$ Simulationen der Hypothesentest aus b) die $H0$ verwirft. Vergleichen Sie den erhaltenen Wert mit der berechneten Güte aus b).

2. Sei folgende geschätzte lineare Einfachregression gegeben mit den Standardfehlern der geschätzten Koeffizienten in Klammern

$$\hat{Y} = \underset{(10.2)}{43.2} + \underset{(7.4)}{61.5}X.$$

Nehmen Sie an, dass alle Annahmen des linearen Modelles erfüllt sind und N gross ist.

a) Konstruieren Sie ein 95% Konfidenzintervall für β_0.

b) Testen Sie die Nullhypothese $H_0 : \beta_1 = 55$ vs. $H_1 : \beta_1 \neq 55$ auf dem 5% Signifikanzniveau.

c) Testen Sie die Nullhypothese $H_0 : \beta_1 = 55$ vs. $H_1 : \beta_1 > 55$ auf dem 5% Signifikanzniveau.

3. Sie benutzen ein Datenset von 50 Klassen und regressieren den durchschnittlichen "Testscore" (Y) auf "Class Size" (X). Sie erhalten folgende Regression mit den Standardfehlern der geschätzten Koeffizienten in Klammern

$$\widehat{TestScore} = \underset{(23.5)}{640.3} - \underset{(2.02)}{4.93} \times ClassSize.$$

a) Konstruieren Sie ein 95% Konfidenzintervall für β_1.

b) Berechnen Sie den p-Wert für den zweiseitigen Hypothesentest mit der Nullhypothese, dass $\beta_1 = 0$. Verwerfen Sie die Nullhypothese auf dem 5% Signifikanzlevel? Wie sieht es aus bei 1%?

c) Berechnen Sie den p-Wert für den zweiseitigen Hypothesentest mit der Nullhypothese, dass $\beta_1 = -5$. Ohne zusätzliche Berechnungen: Ist -5 enthalten im 95% Konfidenzintervall für β_1.

d) Konstruieren Sie ein 90% Konfidenzintervall für β_0.

4. Prof. Rechter-Winkel unterrichtet Vektorgeometrie II an der Universität von Orthogonalia. Im Kurs gibt es insgesamt 100 StudentInnen. Alle haben eine Midterm-Prüfung (in der Mitte des Semesters) und eine Final-Prüfung (am Ende des Semesters) geschrieben. Die Daten ergeben die folgenden Kennzahlen. Midterm: Mittelwert = 230 und SA = 25; Final: Mittelwert = 150 und SA = 20. Eine lineare Einfachregression mit Final als Zielvariable und Midterm als Regressor ergibt einen geschätzten Abschnitt von 21.2 und eine Stichproben-Varianz der Residuen von 206.08.

a) Finden Sie b_1.

b) Finden Sie ein 95% Konfidenz-Intervall für β_0 (d.h. für den Abschnitt).

c) Finden Sie ein 95% Konfidenz-Intervall für β_1 (d.h. für die Steigung).

1. Betrachten Sie folgendes lineares Modell

$$Y_i = 0.2 + 1.2x_i + \epsilon_i \quad \forall\, i = 1, \ldots, N,$$

wobei

$$\epsilon_i \overset{i.i.d.}{\sim} Normal(0, \sigma^2).$$

a) Beschreiben Sie in eigenen Worten die Begriffe Konfidenzintervall, Hypothesentest, p-Wert und Güte in dem Kontext des obigen linearen Modelles.

b) Sie sind an dem Hypothesentest $H0: \ \beta_1 = 1$ gegen $H1: \beta_1 \neq 1$ auf dem $\alpha = 0.05$ Niveau interessiert. Was ist die Güte des Tests, wenn $N = 150$, $\sigma^2 = 3$ und

$$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = 0.33?$$

c) Welche Annahmen müssten Sie ändern, um eine höhere Güte zu erhalten?

d) (In R) Simulieren Sie $S = 1000$ Realisationen von b_1, wenn $N = 150$ und $\sigma^2 = 3$ (wie in b)). Generieren Sie die $x-$Werte einmalig mittels einer $Uniform(-1, 1)-$Verteilung. Setzten Sie dabei den Seed auf 747. Stellen Sie die Verteilung von b_1 in einem Histogramm dar. Was fällt Ihnen auf?

e) In der Simulation von d) zählen Sie wie oft von den $S = 1000$ Simulationen der Hypothesentest aus b) die $H0$ verwirft. Vergleichen Sie den erhaltenen Wert mit der berechneten Güte aus b).

Modell: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ mit $\varepsilon_i \sim N(0, \sigma^2)$ und $\beta_0 = 0.2, \beta_1 = 1.2$

a)

1.) Konfidenzintervall für β_1: Intervall, dass mit einer Konfidenz von $(1-\alpha)$-% den Steigungsparameter β_1 enthält. (Argument mit 100-facher Repetition)

2.) Hypothesentest für β_1: Wir testen, ob wir die Nullhypothese $H_0: \beta_1 = \beta_{1,0}$ verwerfen können (Meist $\beta_{1,0} = 0$)

3.) p-Wert: Wahrscheinlichkeit ein solches oder sogar noch extremeres Ergebnis zu erhalten, gegeben dass die Nullhypothese stimmt. (Vergleich mit α).

4.) Güte: Wahrscheinlichkeit H_0 zu verwerfen, gegeben dass H_0 nicht stimmt.

Wahrheit

Entscheidung	H_0 wahr	H_0 falsch
H_0 nicht verworfen	✓ $P(\cdot) = 1-\alpha$	Fehler II Art $P(\cdot) = \beta$
H_0 verworfen	Fehler I Art mit $P(\cdot) = \alpha$	✓ $P(\cdot) = 1-\beta :=$ Güte

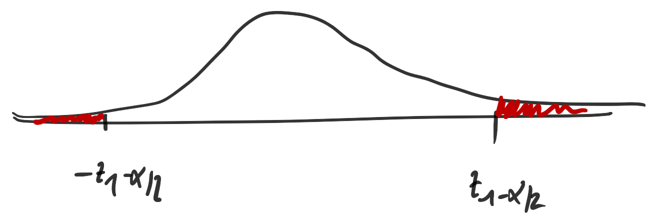
b) $H_0: \beta_1 = 1$

$H_1: \beta_1 \neq 1$

Berechne die Güte des Tests:

1. Für welche Werte $\hat{\beta}$ wird H_0 verworfen?

$|z| > z_{1-\alpha/2}$ $(p < \alpha)$



$$\frac{|\hat{\beta}_1 - \beta_{1,0}|}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}} > z_{1-\alpha/2}$$

$$\frac{|\hat{\beta}_1 - 1|}{\sqrt{\frac{3}{149 \times 0.33}}} > 1.96$$

$$\frac{\hat{\beta}_1 - 1}{\sqrt{0.06}} < -1.96 \quad \text{oder} \quad \frac{\hat{\beta}_1 - 1}{\sqrt{0.06}} > 1.96$$

$\hat{\beta}_1 < 0.52$ oder $\hat{\beta}_1 > 1.48$

2. Was sind die Wahrscheinlichkeiten, dass $\hat{\beta}_1 < 0.52$ oder $\hat{\beta}_1 > 1.48$ eintreffen?

$$P(\hat{\beta}_1 < 0.52) = P\left(\underbrace{\frac{\hat{\beta}_1 - \beta_1}{\frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}}_{:= Z \sim N(0,1)} < \frac{0.52 - \beta_1}{\frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}\right)$$

$$= P(Z < -2.75) = 0.0030$$

$$P(\hat{\beta}_1 > 1.48) = P\left(Z > \frac{1.48 - 1.2}{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}}\right)$$

$$= P(Z > 1.13) = 0.1292 ,$$

sodass Güte $= 0.0030 + 0.1292 = 0.1322$

c) Güte höher:

- N grösser
- σ^2 kleiner
- σ_x^2 grösser
- $|\beta_1 - \beta_{1,0}|$ grösser

2. Sei folgende geschätzte lineare Einfachregression gegeben mit den Standardfehlern der geschätzten Koeffizienten in Klammern

$$\hat{Y} = \underset{(10.2)}{43.2} + \underset{(7.4)}{61.5}X.$$

Nehmen Sie an, dass alle Annahmen des linearen Modelles erfüllt sind und N gross ist.

a) Konstruieren Sie ein 95% Konfidenzintervall für β_0.

b) Testen Sie die Nullhypothese $H_0 : \beta_1 = 55$ vs. $H_1 : \beta_1 \neq 55$ auf dem 5% Signifikanzniveau.

c) Testen Sie die Nullhypothese $H_0 : \beta_1 = 55$ vs. $H_1 : \beta_1 > 55$ auf dem 5% Signifikanzniveau.

a.) $b_0 \pm z_{1-\alpha/2} \cdot \sqrt{\hat{Var}(b_0)}$ oder $b_1 \pm z_{1-\alpha/2} \sqrt{\hat{Var}(b_1)}$

$= 43.2 \pm 1.96 \cdot 10.6 = [23.2, 63.2]$

b.) 1) Hypothese festlegen

$H_0: \beta_1 = 55$

$H_1: \beta_1 \neq 55$

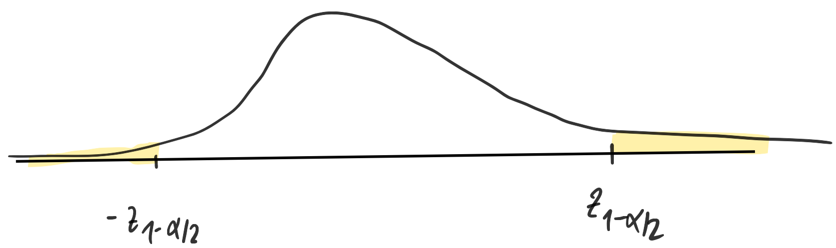
2) Teststatistik unter H_0 berechnen

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}\right) \rightarrow \frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}}} \sim N(0,1)$$

Hier: $z = \frac{61.5 - 55}{7.4} = 0.88$

3) Entscheidung treffen (zwei äquiv. Möglichkeiten):

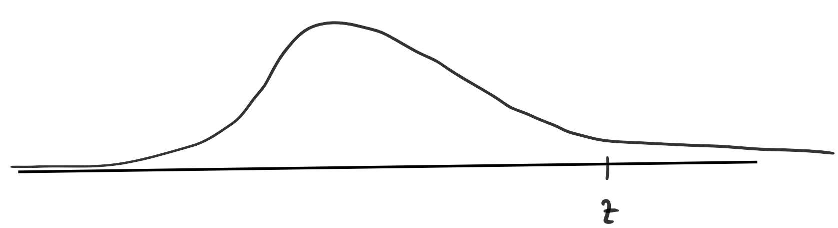
i) Ablehnungsregion:



- H_0 verwerfen, falls $|z| > z_{1-\alpha/2}$ (zweiseitiger Test)
- H_0 verwerfen, falls $z > z_{1-\alpha}$ (rechtsseitiger Test)
- H_0 verwerfen, falls $z < -z_{1-\alpha}$ (linksseitiger Test)

Hier: $z = 0.88 < 1.96$. Wir können nicht verwerfen.

(ii) p-Wert:



- H_0 verwerfen, falls $2P(Z \geq |z|) < \alpha$ (zweiseitig)
- H_0 verwerfen, falls $P(Z > z) < \alpha$ (rechtsseitig)
- H_0 verwerfen, falls $P(Z < z) < \alpha$ (linksseitig)

Hier: $2P(Z > |0.88|) = 2P(Z > 0.88) = 2(1 - P(Z < 0.88)) = 0.37$.

Wir können H_0 nicht verwerfen.

c) Rechtsseitiger Test:

$$z = \frac{61.55 - 55}{7.4} = 0.88 < z_{1-\alpha} = 1.645$$

Wir können nicht verwerfen.

3. Sie benutzen ein Datenset von 50 Klassen und regressieren den durchschnittlichen "Testscore" (Y) auf "Class Size" (X). Sie erhalten folgende Regression mit den Standardfehlern der geschätzten Koeffizienten in Klammern

$$\widehat{TestScore} = \underset{(23.5)}{640.3} - \underset{(2.02)}{4.93} \times ClassSize.$$

a) Konstruieren Sie ein 95% Konfidenzintervall für β_1.

b) Berechnen Sie den p-Wert für den zweiseitigen Hypothesentest mit der Nullhypothese, dass $\beta_1 = 0$. Verwerfen Sie die Nullhypothese auf dem 5% Signifikanzlevel? Wie sieht es aus bei 1%?

c) Berechnen Sie den p-Wert für den zweiseitigen Hypothesentest mit der Nullhypothese, dass $\beta_1 = -5$. Ohne zusätzliche Berechnungen: Ist -5 enthalten im 95% Konfidenzintervall für β_1.

d) Konstruieren Sie ein 90% Konfidenzintervall für β_0.

a) $-4.93 \pm 1.96 \cdot 2.02 = [-8.89, -0.97]$

b) $z = \dfrac{-4.93 - 0}{2.02} = -2.44$

$p\text{-Wert} = 2P(Z > |z|) = 2P(Z > 2.44)$

$= 2(1 - P(Z < 2.44))$

$= 0.0146$ → in R zeigen

→ Wir können also für $\alpha = 0.05$ verwerfen, nicht aber für $\alpha = 0.01$.

c) $z = \frac{-4.93 - (-5)}{2.02} = 0.03$

$2P(z > |z|) = 2P(z > 0.03) = 0.9760$

Wir können Ho nicht verwerfen.

Dualität Hypothesentest und Konfidenzintervalle:

"Falls ein Wert im Hypothesentest zum Level α nicht verworfen wurde, so liegt es im $(1-\alpha)\%$ Konfidenzintervall"

d) $640.3 \pm 1.645 \cdot 23.5 = [601.64, 678.95]$

4. Prof. Rechter-Winkel unterrichtet Vektorgeometrie II an der Universität von Orthogonalia. Im Kurs gibt es insgesamt 100 StudentInnen. Alle haben eine Midterm-Prüfung (in der Mitte des Semesters) und eine Final-Prüfung (am Ende des Semesters) geschrieben. Die Daten ergeben die folgenden Kennzahlen. Midterm: Mittelwert = 230 und SA = 25; Final: Mittelwert = 150 und SA = 20. Eine lineare Einfachregression mit Final als Zielvariable und Midterm als Regressor ergibt einen geschätzten Abschnitt von 21.2 und eine Stichproben-Varianz der Residuen von 206.08.

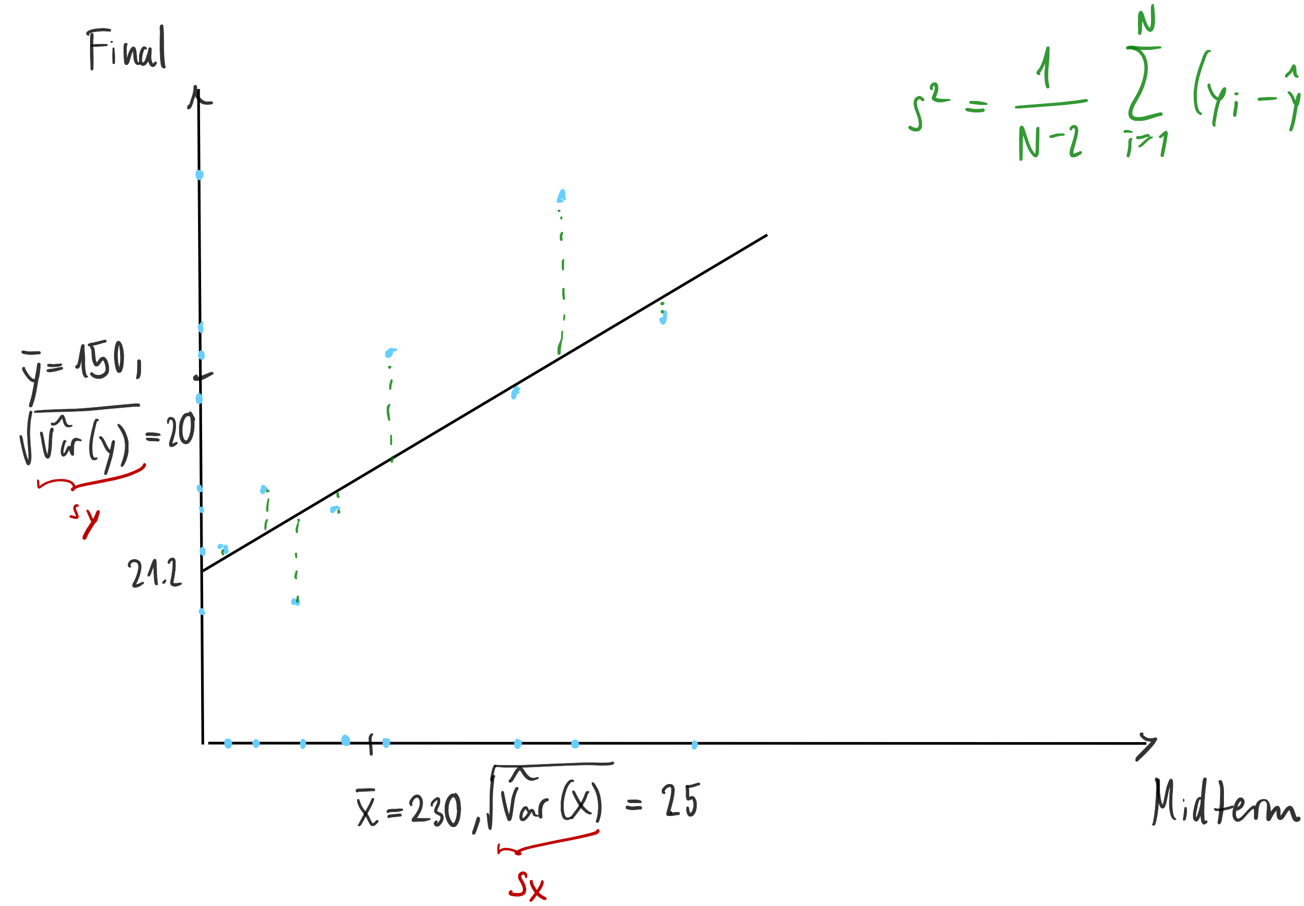
a) Finden Sie b_1.

b) Finden Sie ein 95% Konfidenz-Intervall für β_0 (d.h. für den Abschnitt).

c) Finden Sie ein 95% Konfidenz-Intervall für β_1 (d.h. für die Steigung).

Angaben: $N = 100,\ z_{1-\alpha/2} = 1.96,\ \bar{x} = 230,\ \bar{y} = 150,\ s_x = 25,$
$s_y = 20,\ b_0 = 21.2,\ s^2 = 206.08$

$$s^2 = \frac{1}{N-2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$



a) b_1 herleiten:

$$b_0 = \bar{y} - b_1 \cdot \bar{x} \iff b_1 = \frac{\bar{y} - b_0}{\bar{x}} = 0.56$$

b) 95% -KI

$$= \left[b_0 - z_{1-\alpha/2} \sqrt{s^2 \left(\frac{1}{N} + \frac{\bar{x}}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right)}, b_0 + z_{1-\alpha/2} \sqrt{s^2 \left(\frac{1}{N} + \frac{\bar{x}}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right)} \right]$$

Wir kennen nur $\sum_{i=1}^{N} (x_i - \bar{x})^2$ nicht, aber:

$$s_x = \sqrt{\hat{Var}(X)} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2},$$

sodass $(N-1) s_x^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 = 61'875$

Somit: $95\%\text{-KI} = \left[21.2 \pm 1.96 \cdot \sqrt{206.08 \left(\frac{1}{100} + \frac{230^2}{61'875} \right)} \right]$

$$= [-4.97, 47.37]$$

c) 95% -KI

$$= \left[b_1 \pm 1.96 \cdot \sqrt{\frac{s^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}} \right] = [0.45, 0.67]$$