Prof. Dr. Mitja Stadje
Björn Kriesche, M.Sc.

SS 2015
5. 5. 2015

Ökonometrie - Übungsblatt 2

Abgabe am 12. 5. **vor** Beginn der Übung

Aufgabe 1 (2+2+2 Punkte)

Es sei $X_1, \dots, X_n$ eine Stichprobe aus unabhängigen und identisch verteilten Zufallsvariablen.

(a) Es sei $X_1 \sim \mathrm{N}(\mu, \sigma^2)$ mit bekannter Varianz $\sigma^2 > 0$. Konstruiere ein exaktes symmetrisches Konfidenzintervall für μ zum Niveau γ.

(b) Es sei $X_1 \sim \mathrm{N}(\mu, \sigma^2)$ mit unbekannter Varianz $\sigma^2 > 0$. Konstruiere ein exaktes symmetrisches Konfidenzintervall für μ zum Niveau γ.

(c) Es sei $X_1 \sim \mathrm{Exp}(\lambda)$. Konstruiere ein asymptotisches symmetrisches Konfidenzintervall für λ zum Niveau γ. Verwende den zentralen Grenzwertsatz.

Berechne in den Teilaufgaben (a) und (b) die Länge des Konfidenzintervalls, falls $\gamma = 0,95$ und $n = 20$. Zur Bestimmung der Quantile nutze **R** oder die Tabellen auf der Homepage.
Hinweis: In der Vorlesung wurde das (1-α)-Quantil durch z_α symbolisiert, während es in anderen Quellen (z.B. in den Tabellen) oft durch $z_{1-\alpha}$ symbolisiert wird.

Aufgabe 2 (3+2+2,5+1,5+1+2 Punkte)

Diese Aufgabe soll ohne die Verwendung von **R** bearbeitet werden. Die folgende Tabelle enthält Stichprobendaten für die Anzahl von Stunden, die 8 Studenten zum Lernen für eine Ökonometrieprüfung aufgewendet haben, sowie die Ergebnisse, die sie in der Prüfung erzielt haben.

Lernzeit in Stunden (x)	20	16	34	23	27	32	18	22
Punkte in der Prüfung (y)	64	61	84	70	88	92	72	77

Es wird angenommen, dass sich die erzielten Punkte als lineare Funktion der Lernzeit darstellen lassen.

(a) Bestimme und interpretiere die Regressionsparameter β_0 und β_1 und zeichne die Regressionsgerade zusammen mit den Daten in ein Streudiagramm (scatter plot).

(b) Berechne die Residuen für alle 8 Studenten und überprüfe, ob sich diese zu null aufsummieren.

(c) Beurteile, ob die Annahme der Exogenität (d.h. $\mathbb{E}(u|x) = 0$) grundsätzlich als gegeben angenommen werden kann. Finde und erläutere eine weitere externe Größe, die

  (i) zu einer Überschätzung von β_1 führen könnte,

  (ii) zu einer Unterschätzung von β_1 führen könnte,

  (iii) vermutlich keinen Einfluss auf die Schätzung haben sollte.

(d) Wie ändert sich das Prüfungsergebnis eines Studenten (laut unserem Modell), der sich entschließt 9 Stunden mehr zu lernen als ursprünglich geplant? Prognostiziere das Prüfungsergebnis eines Studenten, der 30 Stunden für die Prüfung gelernt hat bzw. der überhaupt nicht für die Prüfung gelernt hat.

(e) Ein Student möchte die Klausur ohne Abzug, d.h. mit vollen 100 Punkten bestehen. Wie viele Stunden müsste er dafür lernen, wenn man die Störterme außer acht lässt?

(f) Bestimme und interpretiere das Bestimmtheitsmaß R^2.

Aufgabe 3 (1 Punkt)
Betrachte das lineare Regressionsmodell $y = \beta_0 + \beta_1 x + u$ wobei gelte, dass $\alpha_0 = \mathbb{E}(u) \neq 0$. Erläutere und beweise wie das Modell umgeschrieben werden kann, um ein lineares Regressionsmodell zu erhalten, dessen Störterme Erwartungswert 0 haben.

Aufgabe 4 (2+2+1 Punkte)
Betrachte das einfache lineare Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ für $i = 1, \ldots, n$, sowie die in der Vorlesung eingeführten Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für β_0 und β_1.

(a) Betrachte das modifizierte Regressionsmodell $c_1 y_i = \beta_0 + \beta_1 c_2 x_i + u_i$ mit $c_1 \in \mathbb{R}$ und $c_2 \neq 0$. Zeige, dass die Schätzer $\tilde{\beta}_0$ und $\tilde{\beta}_1$ für β_0 und β_1 durch $\tilde{\beta}_0 = c_1 \hat{\beta}_0$ und $\tilde{\beta}_1 = \frac{c_1}{c_2} \hat{\beta}_1$ gegeben sind.

(b) Betrachte das modifizierte Regressionsmodell $c_1 + y_i = \beta_0 + \beta_1 (c_2 + x_i) + u_i$ mit $c_1, c_2 \in \mathbb{R}$. Zeige, dass die Schätzer $\tilde{\beta}_0$ und $\tilde{\beta}_1$ für β_0 und β_1 durch $\tilde{\beta}_0 = \hat{\beta}_0 + c_1 - c_2 \hat{\beta}_1$ und $\tilde{\beta}_1 = \hat{\beta}_1$ gegeben sind.

Gegeben sei nun das quasi-lineare Regressionsmodell $\log(y_i) = \beta_0 + \beta_1 x_i + u_i$ mit $y_i > 0$ für $i = 1, \ldots, n$, sowie die zugehörigen Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für β_0 und β_1.

(c) Betrachte das modifizierte Regressionsmodell $\log(c_1 y_i) = \beta_0 + \beta_1 x_i + u_i$ mit $c_1 > 0$. Zeige, dass die Schätzer $\tilde{\beta}_0$ und $\tilde{\beta}_1$ für β_0 und β_1 durch $\tilde{\beta}_0 = \log(c_1) + \hat{\beta}_0$ und $\tilde{\beta}_1 = \hat{\beta}_1$ gegeben sind.

Aufgabe 5 (6 Punkte)
Betrachte den Datensatz mit den 2053 Mietdaten aus Blatt 1- Aufgabe 4 (Datei `miete03.txt`). Wir wollen untersuchen, ob sich die Nettomiete linear in Abhängigkeit von der Wohnfläche darstellen lässt.

(a) Berechne in **R** Schätzer für die Regressionsparameter β_0 und β_1 und zeichne die Regressionsgerade zusammen mit den Daten in eine Abbildung. Berechne das Bestimmtheitsmaß R^2.

(b) Möglicherweise besteht ein anderer funktionaler Zusammenhang zwischen der Nettomiete und der Wohnfläche. Wir betrachten die folgenden 3 Möglichkeiten:

  (i) der Logarithmus der Nettomiete hängt linear von der Wohnfläche ab,

  (ii) die Nettomiete hängt linear vom Logarithmus der Wohnfläche ab,

(iii) der Logarithmus der Nettomiete hängt linear vom Logarithmus der Wohnfläche ab.

Berechne für jeden der 3 Fälle die Regressionsparameter und zeichne die Regressionsgerade zusammen mit den Daten in eine Abbildung. Welches Modell passt deiner Meinung nach am besten? Entscheide anhand des Bestimmtheitsmaßes R^2, welche Abhängigkeit am besten linear modelliert werden kann.

Die Regressionsparameter und das Bestimmtheitsmaß sollen anhand der Formeln berechnet werden. Die **R**-Funktion `lm` soll nicht verwendet werden.

Hinweis: Das Bestimmheitsmaß R^2 eines linearen Regressionsmodells $y_i = \beta_0 + \beta_1 x_i + u_i$ für $i = 1, \ldots, n$ mit Schätzern $\hat{\beta}_0$ und $\hat{\beta}_1$ für β_0 und β_1 ist wie folgt definiert:

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_n)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_n)^2},$$

wobei $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ für $i = 1, \ldots, n$.