Zwischenprüfung am 13. November

WS 2020/2021, 11:00 – 13:00

Sie sollten für diese Prüfung höchstens 105 Minuten brauchen. Bitte planen Sie ausreichend viel Zeit ein, um Ihre Lösung vor 13:00 auf Learn hochzuladen! Bitte laden Sie genau eine pdf-Datei hoch.
Viel Erfolg!

1 Fallstudie 1 – Einfache lineare Regression

In dieser Fallstudie untersuchen wir den Einfluss des Pro-Kopf-Einkommens auf die Pro-Kopf-Ausgaben für öffentliche Schulen in allen Staaten der USA. Dazu untersuchen wir einen Datensatz mit 50 Beobachtungen (Wisconsin fehlt, jedoch gibt es eine Beobachtung für Washington DC) aus dem Jahre 1979 mit den Spalten `Expenditure`, welche die Pro-Kopf-Ausgaben für öffentliche Schulen enthält, und `Income`, welche das Pro-Kopf-Einkommen enthält.

1.1

Wir betrachten das folgende lineare Modell:

$$Y = \beta_0 + X\beta_1 + u, \quad \mathbb{E}(u|X) = 0. \tag{1}$$

Die abhängige Variable Y entspricht den Pro-Kopf-Ausgaben für öffentliche Schulen (in Dollar pro Kopf), die erklärende Variable X entspricht dem Pro-Kopf-Einkommen (in Dollar pro Kopf). Die Ergebnisse dieser Regression und möglicherweise relevante Plots können Sie auf Seite 8 finden.

Was ist die Kleinste-Quadrate-Schätzung (OLS estimate) des Regressionskoeffizienten, welcher zur erklärenden Variable `Income` gehört? Berichten Sie den Wert der Schätzung und die Interpretation des Parameters. Die Interpretation sollte insbesondere die richtigen Einheiten erwähnen.

1.2

Welches Vorzeichen hat die Korrelation zwischen den Variablen X `(Income)` und Y `(Expenditure)`? Rechtfertigen Sie Ihre Antwort.

1.3

Scheinen die Annahmen $\mathbb{E}(u|X) = 0$ und $\mathbb{V}(u|X) = \sigma^2$ (Homoskedaszität) im Modell 1 erfüllt zu sein? Rechtfertigen Sie Ihre Antwort.

1.4

Was ist die erwartete Veränderung der Pro-Kopf-Ausgaben für öffentliche Schulen bei
a) einem Anstieg der Pro-Kopf-Einkommens um 1000 Dollar und
b) einer Abnahme des Pro-Kopf-Einkommens um 2300 Dollar?

1.5

Nun betrachten wir das folgende log-log Modell:

$$\log(Y) = \xi_0 + \xi_1 \log(X) + \tilde{u}, \quad \mathbb{E}(\tilde{u}|X) = 0, \tag{2}$$

wobei Y und X dieselben Größen wie in Frage 1.1 sind. Sie können ebenso die Ergebnisse dieser Regression und möglicherweise relevante Plots auf Seite 9 finden.

Geben Sie das damit implizierte Modell für Y an, also eine Gleichung der Form $Y = \ldots$.

1.6

Was ist die Kleinste-Quadrate-Schätzung (OLS estimate) des Regressionskoeffizienten, welcher zur erklärenden Variable `income` gehört im log-log-Modell (2)? Berichten Sie den Wert der Schätzung und die Interpretation des Parameters. Die Interpretation sollte insbesondere die richtigen Einheiten erwähnen.

1.7

Scheinen die Annahmen $\mathbb{E}(\tilde{u}|X) = 0$ und $\mathbb{V}(\tilde{u}|X) = \sigma^2$ (Homoskedaszität) im Modell 2 erfüllt zu sein? Rechtfertigen Sie Ihre Antwort.

1.8

Welches der beiden Modelle (1) oder (2) würden Sie bevorzugen und warum?

1.9

Welche Informationen liefert Ihnen diese Regressionsanalyse in sozialpolitischer Hinsicht? Nennen Sie mögliche Maßnahmen, um die aufgeworfenen Punkte anzugehen.

Zwischenprüfung am 13. November
WS 2020/2021, 11:00 – 13:00

---

2 Fallstudie 2 – Mutliple Regression

In der folgenden Fallstudie untersuchen wir, welche Faktoren zum Gehalt eines CEOs beitragen. Der Datensatz enthält Informationen von 177 CEOs aus dem Jahre 1990. Die abhängige Variable `salary` beschreibt das Gehalt in 1000 US Dollar Schritten. Weiters sind die erklärenden Variablen:

- `mktval`: Marktwert des Unternehmens (1 Einheit = 1 Million USD)
- `sales`: Umsatz des Unternehmens (1 Einheit = 1 Million USD)
- `profits`: Gewinn des Unternehmens (1 Einheit = 1 Million USD)
- `ceoten`: Amtszeit des CEO im Unternehmen (1 Einheit = 1 Jahr)

Wir werden uns in den betrachteten Modellen auf diese Variablen mit Y_{salary} und $X_{\text{mktval}}, X_{\text{sales}}, X_{\text{profits}}$ und X_{ceoten} beziehen, wobei `salary` die abhängige Variable ist und alle anderen mögliche erklärende Variablen sind, die man in die Modelle miteinbeziehen kann. Weiters wird gegebenenfalls eine Menge von erklärenden Variablen gemeinsam mit X bezeichnet.

Beantworten Sie die folgenden Fragen unter Zuhilfenahme der Tabellen auf Seite 10.

2.1

Wir schätzen das multiple Regressionsmodell

$$\log(Y_{\text{salary}}) = \beta_0 + \beta_1 \log(X_{\text{mktval}}) + \beta_2 \log(X_{\text{sales}}) + u, \quad \mathbb{E}(u|X) = 0$$

wobei $X = (X_{\text{mktval}}, X_{\text{sales}})$, und bezeichnen dieses als Modell 1.

Was ist die Kleinste-Quadrate-Schätzung (OLS estimate) des Regressionskoeffizienten, welcher zur erklärenden Variable `sales` gehört? Geben Sie den Wert der Schätzung und die Interpretation des Parameters. Die Interpretation sollte insbesondere die richtigen Einheiten erwähnen.

2.2

Was ist die Kleinste-Quadrate-Schätzung des Schnittpunkts (intercept)? Interpretieren Sie diese im vorliegenden Kontext!

Zwischenprüfung am 13. November

WS 2020/2021, 11:00 – 13:00

2.3

i) Wie würden Sie die Variable `profits` ins Modell 1 miteinbeziehen? Ergibt es Sinn, sie als logarithmierte Variable miteinzubeziehen?

ii) Nun schätzen wir ein neues Modell:

$$\log(Y_{\text{salary}}) = \beta_0 + \beta_1 \log(X_{\text{mktval}}) + \beta_2 \log(X_{\text{sales}}) + \beta_3 X_{\text{profits}} + u, \quad \mathbb{E}(u|X) = 0,$$

wobei $X = (X_{\text{mktval}}, X_{\text{sales}}, X_{\text{profits}})$, und wir bezeichnen dieses als Modell 2.

Was ist die Kleinste-Quadrate-Schätzung des zur Variable `profits` gehörenden Regressionskoeffizienten? Berichten Sie den Wert der Schätzung und die Interpretation des Parameters. Die Interpretation sollte insbesondere die richtigen Einheiten erwähnen.

2.4

Meinen Sie, dass es sinnvoll ist, die Variable `profits` dem Modell 1 hinzuzufügen? Um diese Frage zu beantworten, betrachten Sie zuerst folgende zwei Fragen und geben Sie Ihre Antworten dazu:

- Vergleichen Sie die Werte für R-Quadrat der Modelle 1 und 2, die wir zuvor betrachtet haben.
- Besteht eine hohe Multikollinearität zwischen den Variablen `mktval` und `profits`? (Bitte prüfen Sie die gegebenen Korrelationswerte in der Tabelle auf Seite 10.) Falls diese besteht, wie würde sich das Einbeziehen der Variable `profits` ins Modell 1 auf den Standardfehler des Kleinste-Quadrate-Schätzers der Variable `mktval` auswirken?

Geben Sie letztendlich eine Antwort auf die ursprüngliche Frage, ob Sie das Hinzufügen von `profits` zum Modell 1 für sinnvoll erachten.

2.5

Nun schätzen wir ein neues Modell:

$$\log(Y_{\text{salary}}) = \beta_0 + \beta_1 \log(X_{\text{mktval}}) + \beta_2 \log(X_{\text{sales}}) + \beta_3 X_{\text{ceoten}} + u, \quad \mathbb{E}(u|X) = 0,$$

wobei $X = (X_{\text{mktval}}, X_{\text{sales}}, X_{\text{ceoten}})$, und wir bezeichnen dieses als Modell 3. Vergleichen Sie die Werte von R-Quadrat von Modell 1 und Modell 3. Welches Modell würden Sie gemäß diesem Kriterium wählen? Geben Sie Ihre Wahl an und rechtfertigen Sie Ihre Antwort.

2.6

Bestimmen Sie unter Verwendung von Modell 3 die durchschnittliche Differenz des logarithmierten Gehalts eines CEO, der schon 20 Jahre im Unternehmen ist, und einem, der seit 15 Jahren im Unternehmen ist.

2.7

Berechnen Sie unter Verwendung von Modell 3 das geschätzte bedingte Durchschnittsgehalt eines CEO in einem Unternehmen mit einem Marktwert von 100 Millionen USD, einem Umsatz von 200 Millionen USD, welcher seit 5 Jahren im Unternehmen ist.

2.8

Berechnen Sie eine (erwartungstreue/unverzerrte) Schätzung der Fehlervarianz, $(\hat{\sigma}^2)$, im Modell 3. Verwenden Sie dafür das SSR, welches in der Tabelle auf Seite 10 steht.

2.9

Wenn man die Einflussfaktoren auf das Gehalt eines CEO betrachtet, kann man erwarten, dass nicht nur dessen Amtszeit als CEO von Bedeutung ist, sondern dass auch die Anzahl von Jahren in anderen Positionen im selben Unternehmen eine Rolle spielt.

Daher schlägt Ihr(e) Dozent(in) Ihnen vor, diese Variable sowie die Gesamtanzahl an Jahren im selben Unternehmen dem Modell 3 hinzuzufügen. Würden Sie den Ratschlag Ihre(s/r) Dozent(in) befolgen und diese beiden Prädiktoren mit einbeziehen? Beantworten Sie diese Frage unter Angabe Ihrer Argumente.

3 Richtig oder falsch?

Richtig oder falsch? Geben Sie an, ob eine Aussage richtig oder falsch ist. Falls es sich um eine richtige Aussage handelt, geben Sie eine kurze Begründung dafür. Falls es sich um eine falsche Aussage handelt, geben Sie eine korrekte Aussage an oder eine Begründung, warum diese Aussage falsch ist.

3.1

Betrachten Sie das folgende einfache lineare Regressionsmodell

$$Y = \beta_0 + \beta_1 X + u,$$

wobei $\mathbb{E}(u|X) = 0$. Nehmen Sie weiters an, dass $\beta_0 = -1$, $\beta_1 = -3$ und $\mathbb{E}(X) = 2$.

1. Wenn X um 2 Einheiten stegt, dann steigt Y um genau 6 Einheiten.

2. Der Erwartungswert von Y ist -1.

3.2

Betrachten Sie das einfache lineare Regressionsmodell

$$Y = \beta_0 + \beta_1 X + u,$$

wobei $\mathbb{E}(u|X) = 0$, $\mathbb{V}(u|X) = \sigma^2 > 0$. Es sei (y_i, x_i), $i = 1, \ldots, N$, eine zufällige Stichprobe aus der Grundgesamtheit. Seien $\hat{\beta}_0$ und $\hat{\beta}_1$ die OLS-Schätzer von β_0 und β_1.

1. Je größer die Stichprobenvariation $s_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$ des Regressors ist, desto größer ist die Varianz des Schätzers $\hat{\beta}_1$.

2. Die OLS-Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ sind immer unkorreliert.

3.3

Betrachten Sie das multiple lineare Regressionsmodell

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u,$$

wobei $\mathbb{E}(u|X_1, X_2, X_3) = 0$, $\mathbb{V}(u|X_1, X_2, X_3) = \sigma^2 > 0$. Sei $(y_i, x_{1,i}, x_{2,i}, x_{3,i})$, $i = 1, \ldots, N$ eine zufällige Stichprobe.
Weiters sei $\hat{\boldsymbol{\beta}}$ der OLS-Schätzer für $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$. Wir definieren die Residuen $\hat{u}_i = y_i - \hat{y}_i$, $i = 1, \ldots, N$, wobei $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\beta}_3 x_{3,i}$.

Zwischenprüfung am 13. November

WS 2020/2021, 11:00 – 13:00

---

1. Es gilt, dass $y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\beta}_3 x_{3,i}) = 0$ für alle $i = 1, \ldots, N$.

2. Es gilt, dass $\hat{\beta}_j = \beta_j$ für alle $j = 0, 1, 2, 3$.

3. Stellen Sie sich vor, Sie hätten Daten zu Wahlen aus verschiedenen Städten. Zur Wahl standen drei verschiedene Parteien: A, B und C. Nehmen Sie an Y entspricht der Anzahl an abgegebenen Stimmen, X_1 gibt den Anteil an Stimmen für Partei A, X_2 gibt den Anteil an Stimmen für Partei B und X_3 gibt den Anteil an Stimmen für Partei C an.
   Dann kann der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ berechnet werden als
   $$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \qquad \text{wobei} \quad \boldsymbol{y} = (y_1, y_2, \ldots, y_N)'.$$

3.4

Betrachten Sie das Regressionsmodell

$$Y = cX_1^{\beta_1} X_2^{\beta_2} e^u,$$

wobei $Y > 0,\ c > 0,\ \mathbb{E}(u|X_1, X_2) = 0,\ \mathbb{V}(u|X_1, X_2) = \sigma^2 > 0$.

1. Die bedingte Wahrscheinlichkeit $\mathbb{E}(\log(Y)|X_1, X_2)$ ist eine lineare Funktion in Abhängigkeit von β_1 und β_2.

2. Wenn X_1 konstant gehalten wird und X_2 um 1% steigt, erwarten wir einen Anstieg von Y um β_2 Einheiten.

3.5

Nehmen Sie an, Sie schätzen ein einfaches lineares Regressionsmodell, bei dem Sie als abhängige Variable das Gewicht einer Person (Y) und als erklärende Variable die Körpergröße einer Person (X) verwenden. Nehmen Sie weiters an, dass die Stichprobe $N = 1000$ Erwachsenen bestand, deren Körpergrößen zwischen 150cm und 200cm betrug. Das geschätzte Modell sieht wie folgt aus

$$Y = X - 100 + u,$$

wobei u Mittelwert 0 hat gegeben X.
Dieses Modell kann von Kinderärzten verwendet werden um einzuschätzen, ob ihre Patienten übergewichtig sind.

Zwischenprüfung am 13. November

WS 2020/2021, 11:00 – 13:00

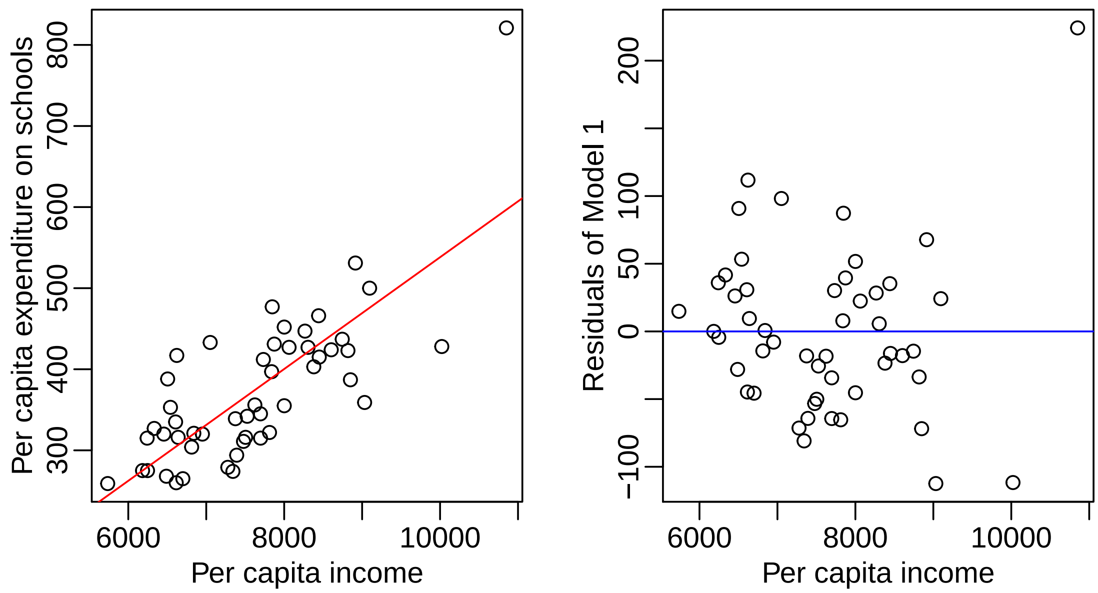

Abbildung 1: Plots belonging to Model 1 from Case Study 1

	Model 1
(Intercept)	−151.27 (64.12)
Income	0.07 (0.01)
SSR	181015.217(df = 48)
R^2	0.59
Adj. R^2	0.58
Num. obs.	50

. Standard errors reported in parentheses

Tabelle 1: Estimation results for Model 1

Zwischenprüfung am 13. November

WS 2020/2021, 11:00 – 13:00

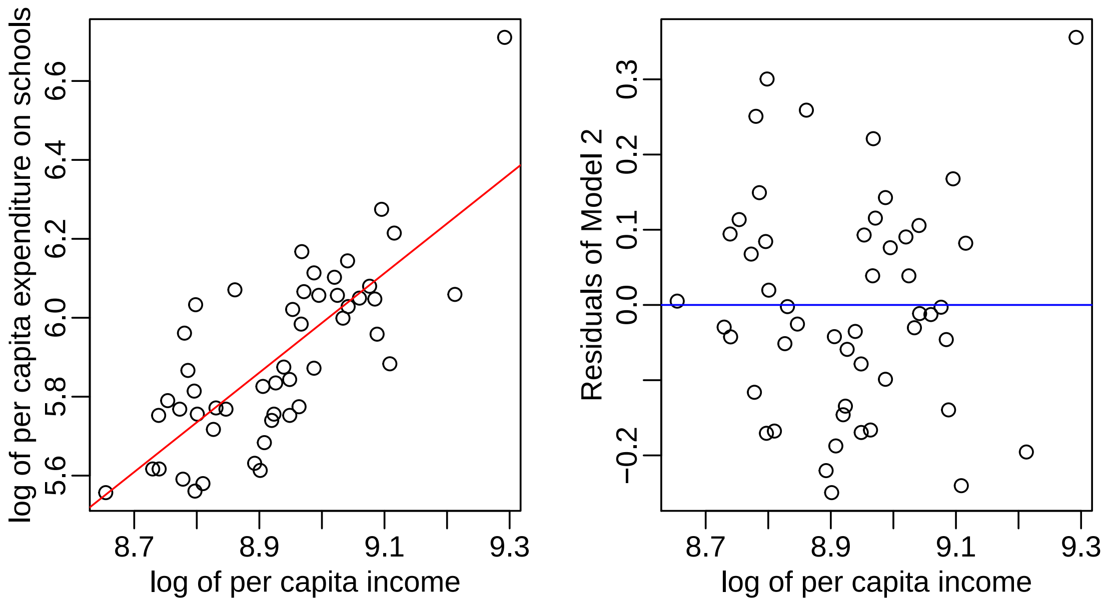

Abbildung 2: Plots belonging to Model 2 from Case Study 1

	Model 2
(Intercept)	-5.35 (1.38)
log(Income)	1.26 (0.15)
SSR	1.017(df = 48)
R^2	0.58
Adj. R^2	0.57
Num. obs.	50

. Standard errors reported in parentheses

Tabelle 2: Estimation results for Model 2

	Model 1	Model 2	Model 3
(Intercept)	4.62092 (0.25441)	4.68692 (0.37973)	4.50379 (0.25723)
log(mktval)	0.10671 (0.05012)	0.09753 (0.06369)	0.10924 (0.04959)
log(sales)	0.16213 (0.03967)	0.16137 (0.03991)	0.16285 (0.03924)
profits		0.00004 (0.00015)	
ceoten			0.01171 (0.00533)
SSR	45.31(df = 174)	45.295(df = 173)	44.079(df = 173)
R^2	0.29911	0.29934	0.31815
Adj. R^2	0.29106	0.28719	0.30633
Num. obs.	177	177	177

. Standard errors reported in parentheses

Tabelle 3: Statistical models

Tabelle 4: Correlation matrix of coefficient estimates

	salary	mktval	sales	profits	ceoten
salary	1.0000000	0.4063071	0.3802239	0.3939276	0.1429477
mktval	0.4063071	1.0000000	0.7546616	0.9181280	0.0066094
sales	0.3802239	0.7546616	1.0000000	0.7982872	-0.0677147
profits	0.3939276	0.9181280	0.7982872	1.0000000	-0.0216068
ceoten	0.1429477	0.0066094	-0.0677147	-0.0216068	1.0000000