

Einführung in die Empirische Wirtschaftsforschung

Muster-Prüfung

Anweisungen:

- Die Prüfung besteht aus fünfzehn Multiple Choice Aufgaben.
- Jede Multiple Choice Aufgabe hat genau eine richtige Lösung. Wenn Sie diese ankreuzen, erhalten Sie einen Punkt. Wenn Sie eine falsche Lösung ankreuzen, werden Ihnen 0.33 Punkte abgezogen. Wenn Sie mehr als eine Lösung ankreuzen oder es nicht 100%ig klar ist, welche Lösung Sie angekreuzt haben, erhalten Sie null Punkte.
- Kopien der notwendigen Verteilungs-Tabellen sind am Ende der Prüfung zu finden.

1. Welche Aussage bezüglich der R^2 -Statistik ist FALSCH?

- (a) Wenn wir eine Variable im Modell (Zielvariable oder Regressor) linear transformieren, ändert sich R^2 nicht.
- (b) Eine starke, jedoch nicht-lineare Beziehung kann in einem kleinen R^2 resultieren.
- (c) Wenn wir dem Modell einen Regressor hinzufügen, der mit der Zielvariablen rein gar nichts zu tun hat, dann kann R^2 auf keinen Fall grösser werden.
- (d) Mit Hilfe von R^2 und s_Y^2 (d.h., der Stichprobenvarianz der Zielvariablen) lässt sich s^2 berechnen (d.h., der Schätzer für die Varianz der Fehler-Terme), so lange wir auch n und k kennen (d.h., den Stichprobenumfang und die Anzahl der ‘interessanten’ Regressoren).

2. Man schätzt ein Modell “Gemeinsame Steigung” wie folgt: $\hat{y} = 1.2 + 0.5d + 2.0x$, wobei d eine Dummy-Variable darstellt. Die gleichen Daten ergäben das folgende geschätzte Modell:

- (a) $\hat{y} = 1.2(1 - d) + 0.5d + 2.0x$
- (b) $\hat{y} = 1.2(1 - d) + 1.7d + 2.0x$
- (c) $\hat{y} = 0.2(1 - d) + 0.5d + 2.0x$
- (d) $\hat{y} = 0.2(1 - d) + 1.7d + 2.0x$

3. Man will ein Konfidenz-Intervall für $\Delta = \mu_1 - \mu_2$ berechnen und hat die Wahl zwischen zwei Stichproben-Designs: (i) zwei unabhängige Stichproben (mit $n_1 = n_2 = n$) oder (ii) paarweise Daten (mit n Datenpaaren). Welche Aussage ist FALSCH?

- (a) Wenn eine Korrelation innerhalb eines (beliebigen) Datenpaares besteht, ist das Design (ii) besser.
- (b) In beiden Designs sind die Schätzer ($\bar{X} - \bar{Y}$ bzw. $\bar{D}$) unverzerrt (oder unbiased).
- (c) Obwohl eine starke Abhängigkeit innerhalb eines (beliebigen) Datenpaares besteht, kann es sein, dass die beiden Designs gleich gut sind.
- (d) Design (ii) liefert gültige Inferenz, auch wenn keinerlei Abhängigkeit innerhalb der Datenpaare besteht.

4. Man möchte sich, aus theoretischen Gründen, für eines der beiden folgenden multiplen Regressions-Modelle entscheiden:

- (1) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$
- (2) $y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + u$

- (a) Man kann eine Entscheidung basierend auf einem F -Test treffen.
- (b) Man sollte sich für das Modell mit dem grösseren R^2 entscheiden.
- (c) Man sollte sich für das Modell entscheiden, in dem $\hat{\beta}_1$ einen kleineren p -Wert hat.
- (d) Nichts von alledem.

5. Ein multiples Regressions-Modell mit drei ‘interessanten’ Variablen ($k = 3$) wurde geschätzt, basierend auf einem Stichprobenumfang von $n = 200$. Der Schätzer des beta-Vektors lautet:

$$\hat{\beta} = (1.9, 5, 4, -3.5)'$$

Die geschätzte Kovarianz-Matrix lautet:

$$\widehat{Var}(\hat{\beta}) = \begin{pmatrix} 3 & -1 & 0.5 & 0 \\ -1 & 4 & 1.5 & 2 \\ 0.5 & 1.5 & 6 & -2 \\ 0 & 2 & -2 & 5 \end{pmatrix}$$

Weiterhin ist die Summe der quadrierten Residuen gegeben durch $SSR = 5'880$. Man beobachtet $x_{neu} = (1, -2, 0, 1)'$. Unter der Annahme, dass die Fehler-Terme eine Normal-Verteilung mit konstanter Standard-Abweichung haben, welches ist das 90% Vorhersage-Intervall für y_{neu} ?

- (a) $[-21.8, -1.4]$
- (b) $[-20.3, -2.9]$
- (c) $[-24.9, 1.7]$
- (d) $[-23.2, 0.0]$

6. Das folgende ist der Software-Output für eine einfache Regression. Leider sind einige Einträge verloren gegangen. Was können wir über $\hat{\beta}_1$ aussagen?

Call:

```
lm(formula = y ~ x)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.1061	0.1966	-0.540	0.5935
x	????	0.1764	????	????

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.048 on 28 degrees of freedom

Multiple R-Squared: 0.09829, Adjusted R-squared: 0.06608

F-statistic: 3.052 on 1 and 28 DF, p-value: 0.0916

- (a) $|\hat{\beta}_1| \leq 0.2$
- (b) $0.2 < |\hat{\beta}_1| \leq 0.3$
- (c) $|\hat{\beta}_1| > 0.3$
- (d) Gar nichts

7. In einem einfachen Regressionsmodell hat die erste Beobachtung eine aussergewöhnlich grosse Cook's Distance.

- (a) Die erste Beobachtung ist ein Ausreisser.
- (b) Der Abstand $|x_1 - \bar{x}|$ ist aussergewöhnlich gross.
- (c) Mindestens eins von beiden.
- (d) Beides.

8. Prof. Werwolf unterrichtet Statistik II an der Universität von Transsylvanien. 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 SD = 25; Final: Mittelwert = 150 und SD = 20; Eine einfache Regression mit Final als Zielvariable und Midterm als Regressor ergibt einen geschätzten Abschnitt von 21.2 und einen Wert von $s = 5.2$ (as Schätzer der SA der Fehler-Terme u_i in der Regression). Was ist die Stichprobenkorrelation zwischen Midterm und Final.

- (a) $r = 0.8$
- (b) $r = 0.7$
- (c) $r = 0.6$
- (d) Die Stichprobenkorrelation kann mit den angegebenen Informationen nicht ermittelt werden.

9. Was ist der Standardfehler von $\hat{\beta}_0 = 21.2$ (d.h. dem geschätzten Abschnitt) in der Problemstellung von Aufgabe 8?

- (a) 3.84
- (b) 4.84
- (c) 5.84
- (d) Der Standardfehler kann mit der angegebenen Information nicht berechnet werden.

10. Welchen Befehl benützt man in R, um ein Regressions-Modell mit den zwei ‘interessanten’ Regressoren X_1 und $X_2 - X_3$ Zu schätzen?

- (a) `> fit = lm(y ~ x1 + (x2 - x3))`
- (b) `> fit = lm(y ~ x1 + I(x2 - x3))`
- (c) `> fit = lm(x1 + (x2 -x3) ~ y)`
- (d) `> fit= lm(x1 + I(x2 -x3) ~ y)`

11. Der folgende Software Output basiert auf Daten der Firma FedEx im Zeitraum 1974–1976. Es geht darum, wie viele Einheiten (*volume*) eines bestimmten Produkts (nämlich des “Courier Paks”) verkauft wurden in Abhängigkeit des Preises (*price*) und des Monats (*trend*). Hierbei ist *trend* ein linearer Zeittrend mit *trend* = 1 für 01/1974, *trend* = 2 für 02/1974, ... und *trend* = 29 für 05/1976 (der Monat der letzten Beobachtung im Datensatz).

```
> fedex.fr
  volume price trend
1    140   5.0     1
2    195   5.0     2
3    268   5.0     3
.      .    .     .
.      .    .     .
.      .    .     .
28   1370  12.5    28
29   1304  12.5    29

> summary(lm(volume ~ price + trend))
Call:
lm(formula = volume ~ price + trend)

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  319.264      77.472   4.121 0.000341 ***
price        -31.942      14.435  -2.213 0.035891 *
trend         44.781       4.087  10.956 3.06e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Versetzen Sie sich in die Schuhe von FedEx und zurück nach Ende 05/1976. Der momentane Preis ist $price = 12.5$. Welchen der vier unten angeführten Preise sollten Sie für 06/1976 wählen, wenn das Ziel die Maximierung des Umsatzes in diesem Monat ist?

- (a) $price = 7.5$
- (b) $price = 10.0$
- (c) $price = 12.5$
- (d) $price = 15.0$

12. Eine Forscherin möchte herausfinden, wie der Ernteertrag einer Pflanze (Y) auf die Menge des verabreichten Düngers (X) reagiert. Sie hat dazu $n = 10$ Parzellen (identischen) Bodens zur Verfügung. Auf der i -ten Parzelle wird sie die Menge x_i an Dünger verabreichen, wobei jeweils $x_i \in [0, 100]$ sein muss (in einer angemessenen Einheit). Die Forscherin geht davon aus, dass es eine lineare Beziehung gibt der Art

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Sie wählt $x_1, \dots, x_{10}$ so aus, um die Varianz des Kleinste-Quadrate Schätzers $\hat{\beta}_1$ zu minimieren. Die Analyse der daraufhin erhaltenen Daten ergibt einen Standardfehler von 0.0797 für $\hat{\beta}_1$. Welches ist ein unverzerrter Schätzer für σ^2 (d.h. für die konstante Varianz der Fehlerterme u_i)?

- (a) 17.6
- (b) 158.8
- (c) 1'429.2
- (d) Die Frage kann mit Hilfe der angegebenen Informationen nicht beantwortet werden.

13. Man betrachtet ein Modell "Gemeinsame Steigung": $y = \beta_0 + \delta d + \beta_1 x + e$, wobei d eine Dummy-Variable darstellt. Die Schätzung, basierend auf $n = 500$ Beobachtungen, ergibt: $\hat{y} = 0.5 + 1.2d + 2.0x$. Der Standardfehler für $\hat{\beta}_0$ ist 0.1, der Standardfehler für $\hat{\delta}$ ist 0.2 und der Standardfehler für $\hat{\beta}_1$ ist 0.5. Ein 95% Konfidenzintervall für den Abschnitt in der Gruppe mit $d = 1$ ist gegeben als

- (a) 1.7 ± 0.588
- (b) 1.7 ± 0.438
- (c) 1.7 ± 1.073
- (d) Das Konfidenzintervall kann mit Hilfe der angegebenen Informationen nicht berechnet werden.

14. Stichwort Heteroskedastie: Welches ist der HC3-Schätzer für σ_i (d.h. für die Standardabweichung des Fehlerterms u_i), wobei $\hat{u}_i$ das OLS-Residuum für die i -te Beobachtung ist? Hinweis hierzu:

$$\text{Stand-Resid}_i = \frac{\hat{u}_i}{s\sqrt{1 - r_{ii}}}$$

- (a) $\hat{\sigma}_i = \hat{u}_i$
- (b) $\hat{\sigma}_i = \hat{u}_i / \sqrt{1 - r_{ii}}$
- (c) $\hat{\sigma}_i = \hat{u}_i / (1 - r_{ii})$
- (d) Nichts davon

15. Das folgende ist der Software Output für eine einfache Regression, leider gingen jedoch ein paar Zahlen verloren. Was ist der Wert der R^2 -Statistik?

Call:

```
lm(formula = y ~ x)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.79998	-0.33930	-0.05944	0.13508	1.17749

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.7170	*****	2.055	0.0701 .
x	0.1655	0.2406	*****	0.5087

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6069 on 9 degrees of freedom

Multiple R-Squared: *****, Adjusted R-squared: *****

F-statistic: 0.4735 on 1 and 9 DF, p-value: 0.5087

- (a) 0.05
- (b) 0.10
- (c) 0.20
- (d) Die Frage kann mit Hilfe der gegebenen Informationen nicht beantwortet werden.