

Einführung in die Empirische Wirtschaftsforschung

Prüfung 2008

Anweisungen:

- Die Prüfung besteht aus zwanzig 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, wird Ihnen ein Drittelpunkt 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 Standardnormal-, t - und Chi-quadrat-Tabellen sind am Ende der Prüfung zu finden.

1. Man untersucht die Beziehung zweier numerischer Variablen und erhält eine Stichproben-Korrelation von $r = -0.95$. Welche Aussage ist FALSCH?

- (a) Es ist möglich, dass eine starke nichtlineare Beziehung besteht.
- (b) Es ist möglich, dass eine mittel-starke lineare Beziehung besteht, aber Ausreisser vorhanden sind.
- (c) Es ist möglich, dass eine nichtlineare Transformation einer der beiden Variablen in einer grösseren Stichproben-Korrelation resultiert.
- (d) Der Bestimmtheits-Koeffizient R^2 ist grösser als r .

2. Eine Autofirma verkauft ein bestimmtes Modell in 3 Farben. Eine Zufalls-Stichprobe von Käufern ergibt die folgenden Daten (wobei die Käufer in drei Altersklassen eingeordnet werden). Der zugehörige Wert der χ^2 -Test-Statistik ergibt den Wert $X^2 = 7.88$.

	Weiss	Rot	Schwarz
Jung	34	60	95
Mittel	50	70	94
Alt	45	62	61

Für den p -Wert PW bzgl. der Null-Hypothese, dass "Farbe" und "Altersklasse" voneinander unabhängig sind, gilt:

- (a) $PW > 0.1$
- (b) $0.05 < PW \leq 0.1$
- (c) $0.01 < PW \leq 0.05$
- (d) $PW \leq 0.01$

3. 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 für diesen Monat ein möglichst hoher Umsatz erzielt werden soll.

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

4. Wir betrachten das Regressionsmodell $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{2,i}^2 + e_i$. Welchen Befehl benützt man in R, um das Modell unter der Nullhypothese $H_0 : \beta_2 + 80\beta_3 = 1$ zu schätzen?

- (a) `fit = lm(I(y - x2) ~ x1 + I(x2^2 - 80 * x2))`
- (b) `fit = lm(I(y + x2) ~ x1 + I(x2^2 - 80 * x2))`
- (c) `fit = lm(I(y + x2) ~ x1 + I(x2^2 + 80 * x2))`
- (d) `fit = lm(I(y - x2) ~ x1 + I(x2^2 + 80 * x2))`

5. In einem einfachen Regressionsmodell hat die erste Beobachtung eine aussergewöhnlich grosse Cook's Distance. Es gilt:

- (a) Die erste Beobachtung ist ein Ausreisser.
- (b) Der Abstand $|x_1 - \bar{x}|$ ist aussergewöhnlich gross im Vergleich mit den anderen $|x_i - \bar{x}|$.
- (c) Mindestens eins von beiden, d.h. (a) oder (b).
- (d) Beides, d.h. (a) und (b).

6. Basierend auf einer Stichprobe von $n = 100$ Beobachtungen berechnet man die folgenden Ergebnisse in einem multiplen Regressions-Modell ($k = 3$):

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} -0.1 \\ 1.2 \\ 1.7 \\ -2.4 \end{pmatrix} \quad \text{und} \quad \widehat{Var}(\hat{\beta}) = \begin{pmatrix} 3 & -1 & 0.5 & 0 \\ -1 & 4 & 1.5 & 2 \\ 0.5 & 1.5 & 6 & -2 \\ 0 & 2 & -2 & 5 \end{pmatrix}$$

Welches ist die Test-Statistik für $H_0 : \beta_1 - 2\beta_2 = 0$ gegen $H_A : \beta_1 - 2\beta_2 \neq 0$?

- (a) 0.175
- (b) 1.750
- (c) -0.469
- (d) -4.690

7. Man untersucht die Beziehung zweier numerischer Variablen (X und Y) mit Hilfe eines Regressions-Modells. Sei $\hat{\beta}_{1,A}$ der Schätzer der Steigung im allgemeinen Modell $y_i = \beta_0 + \beta_{1,A}x_i + e_i$ (mit zugehörigem Bestimmtheitskoeffizienten R_A^2) und sei $\hat{\beta}_{1,V}$ der Schätzer des Abschnitts im vereinfachten Modell $y_i = \beta_{1,V}x_i + e_i$ (mit zugehörigem Bestimmtheitskoeffizienten R_V^2). Welche der folgenden Aussagen ist FALSCH?

- (a) $R_V^2 \leq R_A^2$.
- (b) $Var(\hat{\beta}_{1,V}) \leq Var(\hat{\beta}_{1,A})$.
- (c) Wenn $\beta_0 \neq 0$, dann ist $\hat{\beta}_{1,V}$ ein verzerrter Schätzer für $\beta_{1,A}$.
- (d) $\hat{\beta}_{1,V} = \sum x_i y_i / \sum x_i^2$.

8. Eine Firma verkauft einen Benzinzusatz, der angeblich den Benzinverbrauch von Autos senken soll. Die Packung behauptet, dass die durchschnittliche Senkung des Verbrauches bei 0.3 (Liter/100 km) liegt. Man wählt eine (Zufalls)Stichprobe von $n = 100$ Autos aus und misst ihren Benzinverbrauch jeweils ohne und mit dem Zusatz. Die folgenden Statistiken kommen dabei heraus:

Gruppe	n	$\bar{x}$	s
Ohne	100	8.50	1.5
Mit	100	8.35	1.2

Die Stichproben-Korrelation (zwischen "Ohne" und "Mit") beträgt $r = 0.9$. Welches ist ein 90% Konfidenz-Intervall für die durchschnittliche Senkung des Verbrauches?

- (a) 0.15 ± 0.316
- (b) 0.15 ± 0.377
- (c) 0.15 ± 0.110
- (d) 0.15 ± 0.131

9. 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

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

Sie wählt $x_1, \dots, x_{10}$ so aus, um die Varianz des Kleinst-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)?

- (a) 17.6
- (b) 158.8
- (c) 1,429.2

(d) Die Frage kann mit Hilfe der angegebenen Information nicht beantwortet werden.

10. 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 der angegebenen Information nicht berechnet werden.

11. Das folgende ist der Software Output für das Eiskrem-Beispiel aus der Vorlesung (leider gingen jedoch ein paar Zahlen verloren ...). Was ist das 90% Konfidenzintervall für β_0 (d.h., für den Abschnitt) basierend auf der angemessenen t -Verteilung?

```
Call:
lm(formula = sales ~ temp)

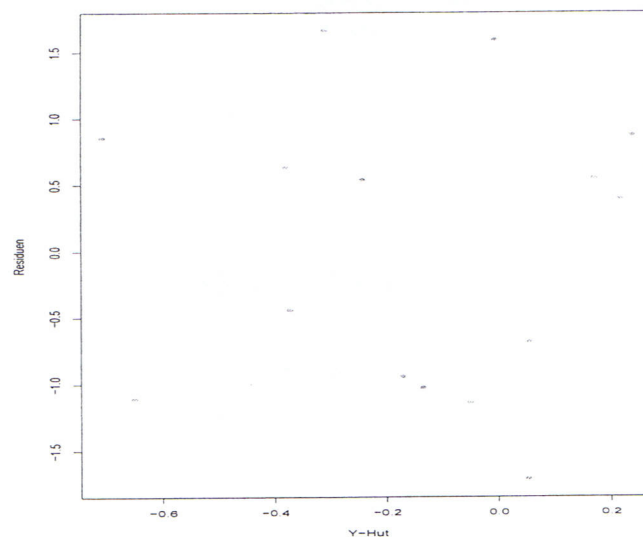
Residuals:
    Min       1Q   Median       3Q      Max
-72.386 -25.303  -1.980  35.390  59.575

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  348.8975      ????   32.411  < 2e-16 ***
temp          5.9055       ????    7.145 8.93e-08 ***

Residual standard error: 40.33 on 28 degrees of freedom
Multiple R-Squared:  0.6458, Adjusted R-squared:  0.6332
F-statistic: 51.05 on 1 and 28 DF,  p-value: 8.935e-08
```

- (a) 348.9 ± 14.1
- (b) 348.9 ± 18.3
- (c) 348.9 ± 22.0
- (d) Das Konfidenzintervall kann mit der angegebenen Information nicht berechnet werden.

12. Das folgende ist das Residuen-Diagramm für eine einfache Regression. Welche Aussage ist WAHR?



- (a) Das falsche Modell wurde geschätzt.
- (b) Es handelt sich klar um eine Zeitreihe.
- (c) Es liegt klar Heteroskedastie vor.
- (d) Die (Standard-)Inferenz in diesem Modell basierend auf der t -Verteilung kann sehr unterschiedlich sein im Vergleich zur (Standard-)Inferenz basierend auf der Normalverteilung.

13. Basierend auf einer Stichprobe von $n = 20$ BuchhalterInnen werden die folgenden drei Modelle geschätzt:

```
Call: lm(formula = salary ~ age + qs)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -137.113    127.843   -1.073  0.298483
age           5.556      1.288    4.316  0.000469 ***
qs           1.559      4.388    0.355  0.726738
```

```
Call: lm(formula = age ~ qs)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   80.052     13.847    5.781 1.77e-05 ***
qs          -1.724      0.693   -2.487  0.0229 *
```

```
Call: lm(formula = qs ~ age)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.54117     2.83365    9.366 2.42e-08 ***
age        -0.14840     0.05967   -2.487  0.0229 *
```

Welches ist die geschätzte Steigung in dem Modell, das salary auf qs regressiert? D.h., welches ist der fehlende Wert in der Spalte "Estimate" in dem folgenden geschätzten Modell:

```
Call: lm(formula = salary ~ qs)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  307.693     106.410    2.892  0.00972 **
qs           xxx         xxx      xxx  0.14952
```

- (a) -8.0
- (b) 0.8
- (c) 1.4
- (d) Der fehlende Wert kann mit der angegebenen Information nicht berechnet werden.

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

- (i) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i$
- (ii) $\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e_i$

Welche Aussage ist WAHR?

- (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 mit dem grösseren adjustierten R^2 entscheiden.
- (d) Nichts von alledem.

15. 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

16. Volker Bierbrauer lebt in Melbourne und interessiert sich für den Alkoholkonsum in Victoria (ein Bundesstaat von Australien). Volker ist ein leidenschaftlicher Biertrinker. Aus eigener Erfahrung weiss er, dass der Bierkonsum stark von der Aussentemperatur und damit der Jahreszeit abhängt. Er möchte deshalb untersuchen, ob im Herbst und im Winter (3. und 4. Quartal) mehr Alkohol konsumiert wird als im Frühling und im Sommer (1. und 2. Quartal). Er verfügt über eine Stichprobe mit Quartalsdaten, die den Zeitraum 1995.3 bis 2006.1 abdeckt (JJJJ.Q bezeichnet das Q-te Quartal des Jahres JJJJ). Für folgende drei Variablen liegen Beobachtungen vor (t ist von der Form JJJJ.Q):

q_t : reale pro-Kopf Ausgaben für Alkohol

p_t : realer Preisindex für Alkohol

y_t : verfügbares, reales pro-Kopf Einkommen

Um den Einfluss der Jahreszeit auf den Alkoholkonsum modellieren zu können, definiert Volker 4 Dummy-Variablen. Für $i \in \{1, \dots, 4\}$ sei

$$Q_{i,t} := \begin{cases} 1, & \text{jeweils im } i\text{-ten Quartal eines Jahres,} \\ 0 & \text{sonst.} \end{cases}$$

Volker modelliert die Alkohalnachfrage mit verschiedenen Regressionsmodellen:

$$\begin{aligned} \mathcal{M}_1 : \quad q_t &= \beta_0 + \beta_1 p_t + \beta_2 y_t + \gamma_1 Q_{1,t} + \gamma_2 Q_{2,t} + \gamma_3 Q_{3,t} + \gamma_4 Q_{4,t} + e_t \\ \mathcal{M}_2 : \quad q_t &= \beta_0 + \beta_1 p_t + \beta_2 y_t + \gamma_1 Q_{1,t} + \gamma_2 Q_{2,t} + \gamma_3 Q_{3,t} + e_t \\ \mathcal{M}_3 : \quad q_t &= \beta_1 p_t + \beta_2 y_t + \gamma_1 Q_{1,t} + \gamma_2 Q_{2,t} + \gamma_3 Q_{3,t} + \gamma_4 Q_{4,t} + e_t \end{aligned}$$

Welche der folgenden Aussagen ist FALSCH?

- (a) Für alle t gilt $\sum_{i=1}^4 Q_{i,t} = 1$.
- (b) Im Modell $\mathcal{M}_2$ "misst" γ_3 den Einfluss (ceteris paribus) des 3. Quartals auf die realen pro-Kopf Ausgaben für Alkohol.
- (c) Im Modell $\mathcal{M}_3$ "misst" γ_3 den Einfluss (ceteris paribus) des 3. Quartals auf die realen pro-Kopf Ausgaben für Alkohol.
- (d) Die Parameter des Modells $\mathcal{M}_1$ können mit der Methode der kleinsten Quadrate nicht (eindeutig) geschätzt werden.

17. Volker hat das Regressionsmodell $\mathcal{M}_2$ von Aufgabe 16 mit R geschätzt und den folgenden Software-Output erhalten:

```
Call:
lm(formula = q ~ y + p + Q1 + Q2 + Q3)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.5109      1.5284   5.569 2.41e-06 ***
y              2.1774      0.2740   7.945 1.62e-09 ***
p              1.1635      1.7733   0.656  0.516
Q1             0.3958      0.2411   1.642  0.109
Q2            -1.4169      0.2307  -6.142 4.03e-07 ***
Q3            -1.7323      0.2274  -7.618 4.33e-09 ***

Residual standard error: 0.4025 on 37 degrees of freedom
Multiple R-Squared:  0.9291,    Adjusted R-squared:  0.9196
F-statistic: 97.03 on 5 and 37 DF,  p-value: < 2.2e-16
```

Interpretieren sie den Wert 1.1635 in der Spalte "Estimate" des Software-Outputs:

- (a) *Ceteris paribus*, wenn der reale Preisindex für Alkohol um eine Einheit steigt, steigen die realen pro-Kopf Ausgaben für Alkohol um 1.1635 Einheiten.
- (b) *Ceteris paribus*, wenn der reale Preisindex für Alkohol um eine Einheit steigt, steigen die durchschnittlichen realen pro-Kopf Ausgaben für Alkohol um 1.1635 Einheiten.
- (c) *Ceteris paribus*, wenn der reale Preisindex für Alkohol um eine Einheit steigt, steigen die realen pro-Kopf Ausgaben für Alkohol um geschätzte 1.1635 Einheiten.
- (d) *Ceteris paribus*, wenn der reale Preisindex für Alkohol um eine Einheit steigt, steigen die durchschnittlichen realen pro-Kopf Ausgaben für Alkohol um geschätzte 1.1635 Einheiten.

18. Die matrizenweise Schreibweise des Regressionsmodells $\mathcal{M}_2$ (siehe Aufgaben 16 und 17) für alle Beobachtungen der Stichprobe sei $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Mit Hilfe von R berechnet Volker $(\mathbf{X}'\mathbf{X})^{-1}$. Die Werte auf der Hauptdiagonalen dieser Matrix sind dabei leider "verloren" gegangen (mit ? markiert). Für die übrigen Werte von $(\mathbf{X}'\mathbf{X})^{-1}$ erhält er

	(Intercept)	y	p	Q1	Q2	Q3
(Intercept)	?	-0.2875	-13.6102	-0.3358	-0.2764	-0.2804
y	-0.2875	?	-1.4498	0.2786	0.2607	0.2523
p	-13.6102	-1.4498	?	-0.7973	-0.7911	-0.7553
Q1	-0.3358	0.2786	-0.7973	?	0.2477	0.2427
Q2	-0.2764	0.2607	-0.7911	0.2477	?	0.2329
Q3	-0.2804	0.2523	-0.7553	0.2427	0.2329	?

Nun möchte Volker den folgenden Hypothesentest durchführen:

H_0 : Im 1. Quartal sind die durchschnittlichen Ausgaben für Alkohol gleich hoch wie im 2. Quartal.

H_A : Im 1. Quartal sind die durchschnittlichen Ausgaben für Alkohol höher als im 2. Quartal.

Welche Aussage über den Wert der Teststatistik t für diesen Hypothesentest ist WAHR?

(Bemerkung: verwenden Sie auch den Software-Output der Aufgabe 17.)

- (a) t ist kleiner als 5.
- (b) t liegt im Intervall $[5, 9]$.
- (c) t ist grösser als 9.
- (d) Aufgrund der fehlenden Angaben kann t nicht berechnet werden.

19. Volker vermutet, dass es in der Nachfrage nach Alkohol (siehe Aufgabe 16) im 1. Quartal 2003 einen Strukturbruch gegeben hat. Um diese Hypothese zu testen, definiert er zunächst eine Dummy-Variable D_t , die vor 2003.1 den Wert 0 und sonst den Wert 1 hat. Dann schätzt er das Regressionsmodell M_2 von Aufgabe 16 mit R und erhält den folgenden Software-Output:

```
Call:
lm(formula = q ~ y + p + Q1 + Q2 + Q3)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.5109      1.5284   5.569 2.41e-06 ***
y              2.1774      0.2740   7.945 1.62e-09 ***
p              1.1635      1.7733   0.656  0.516
Q1             0.3958      0.2411   1.642  0.109
Q2            -1.4169      0.2307  -6.142 4.03e-07 ***
Q3            -1.7323      0.2274  -7.618 4.33e-09 ***

Residual standard error: 0.4025 on 37 degrees of freedom
Multiple R-Squared:  0.9291,    Adjusted R-squared:  0.9196
F-statistic: 97.03 on 5 and 37 DF,  p-value: < 2.2e-16
```

Anschliessend schätzt er ein erweitertes Regressionsmodell, in dem die Dummy-Variable D zum Einsatz kommt. Der entsprechende Software-Output lautet wie folgt:

```
Call:
lm(formula = q ~ y + p + Q1 + Q2 + Q3 + I(D*y) + I(D*p) + I(D*Q1) + I(D*Q2) + I(D*Q3))

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.67701     2.31118   6.783 1.15e-07 ***
y             1.41043     0.28814   4.895 2.70e-05 ***
p            -3.66563     2.02133  -1.813  0.0791 .
Q1             0.05159     0.25130   0.205  0.8386
Q2            -1.75618     0.24862  -7.064 5.20e-08 ***
Q3            -1.99929     0.22367  -8.939 3.28e-10 ***
I(D * y)       1.83965     1.05941   1.736  0.0921 .
I(D * p)      -5.85511     4.11773  -1.422  0.1647
I(D * Q1)       0.42413     0.67751   0.626  0.5357
I(D * Q2)       0.44781     0.67920   0.659  0.5144
I(D * Q3)       0.54335     0.73009   0.744  0.4622

Residual standard error: 0.3249 on ? degrees of freedom
Multiple R-Squared:  ?,    Adjusted R-squared:  0.9476
F-statistic: 76.92 on 10 and 32 DF,  p-value: < 2.2e-16
```

Leider sind dabei ein paar Angaben "verloren" gegangen (mit ? markiert). Mit Hilfe dieser beiden Software-Outputs möchte Volker die Hypothese eines Strukturbruchs mittels Chow-Test testen. Welche der folgenden Aussagen ist WAHR?

- (a) Der Wert der Teststatistik F ist kleiner als 3.5.
- (b) Der Wert der Teststatistik F liegt im Intervall $[3.5, 4.5]$.
- (c) Der Wert der Teststatistik F ist grösser als 4.5.
- (d) Aufgrund der vorliegenden Informationen kann der Test auf einen Strukturbruch nicht durchgeführt werden.

20. In einem multiplen Regressions-Modell mit drei 'interessanten' Regressoren möchte man die Null-Hypothese testen $H_0: \beta_2 = 0.5 \ \& \ \beta_1 + 2\beta_3 = 0$. Finden Sie den Wert der F -Test-Statistik, basierend auf dem folgenden Software Output:

```
Call: lm(formula = y ~ x1 + x2 + x3)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.9672	0.1502	6.439	6.35e-08 ***
x1	1.8892	0.1400	13.492	< 2e-16 ***
x2	0.4475	0.1392	3.214	0.00239 **
x3	-0.4831	0.1595	-3.029	0.00401 **

Residual standard error: 1.05 on 46 degrees of freedom
Multiple R-Squared: 0.8344, Adjusted R-squared: 0.8236
F-statistic: 77.28 on 3 and 46 DF, p-value: < 2.2e-16

```
Call: lm(formula = I(y - 0.5 * x2) ~ I(x1 - 0.5 * x3))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.0081	0.1557	6.477	4.67e-08 ***
I(x1 - 0.5 * x3)	1.6937	0.1200	14.117	< 2e-16 ***

Residual standard error: 1.097 on 48 degrees of freedom
Multiple R-Squared: 0.8059, Adjusted R-squared: 0.8019
F-statistic: 199.3 on 1 and 48 DF, p-value: < 2.2e-16

```
Call: lm(formula = I(y - 0.5 * x2) ~ I(x1 + 0.5 * x3))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.9386	0.2445	3.838	0.000362 ***
I(x1 + 0.5 * x3)	1.6304	0.2228	7.318	2.40e-09 ***

Residual standard error: 1.711 on 48 degrees of freedom
Multiple R-Squared: 0.5274, Adjusted R-squared: 0.5175
F-statistic: 53.56 on 1 and 48 DF, p-value: 2.396e-09

```
Call: lm(formula = I(y + 0.5 * x2) ~ I(x1 + 0.5 * x3))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.9935	0.2933	3.387	0.00142 **
I(x1 + 0.5 * x3)	1.5471	0.2673	5.789	5.26e-07 ***

Residual standard error: 2.053 on 48 degrees of freedom

Multiple R-Squared: 0.4111, Adjusted R-squared: 0.3988

F-statistic: 33.51 on 1 and 48 DF, p-value: 5.263e-07

```
Call: lm(formula = I(y + 0.5 * x2) ~ I(x1 - 0.5 * x3))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.0504	0.2098	5.007	7.87e-06 ***
I(x1 - 0.5 * x3)	1.6901	0.1617	10.453	5.83e-14 ***

Residual standard error: 1.478 on 48 degrees of freedom

Multiple R-Squared: 0.6948, Adjusted R-squared: 0.6884

F-statistic: 109.3 on 1 and 48 DF, p-value: 5.825e-14

- (a) $F = 3.2$
- (b) $F = 40.7$
- (c) $F = 4.0$
- (d) $F = 24.6$