Masterprüfung WiSe 2020/21 - MUSTERLÖSUNG

Fach: Ökonometrie

Prüferin: Prof. Regina T. Riphahn, Ph.D.

Vorbemerkungen:

Anzahl der Aufgaben:	Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen. **Es wird nur der Lösungsbogen eingesammelt.**
Bewertung:	Es können maximal 90 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
Erlaubte Hilfsmittel:	• Formelsammlung (ist der Klausur beigefügt) • Tabellen der statistischen Verteilungen (sind der Klausur beigefügt) • Taschenrechner • Fremdwörterbuch
Wichtige Hinweise:	• Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert. • Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1: **[11,5 Punkte]**

Sie schätzen eine Regression mit der täglichen Schlafdauer in Stunden in Abhängigkeit von erklärenden Merkmalen. Für die Beobachtungen $i = 1, ..., N$ stellen Sie folgendes Modell auf:

$$sleep_i = \beta_1 + \beta_2 \cdot workhours_i + \beta_3 \cdot f\_workhours_i + \beta_4 \cdot female_i + \beta_5 \cdot city_i + \beta_6 \cdot youngkid_i + \varepsilon_i$$

In den folgenden Tabellen finden Sie sowohl deskriptive Statistiken als auch die Regressionsergebnisse einer KQ-Schätzung.

Variable	**Mittelwert**	**Std. Abw.**	**Min.**	**Max.**	**Beschreibung**
sleep	7.78	1.06	1.80	11.18	Tägliche Schlafdauer in Stunden
workhours	5.05	2.26	0	15.27	Tägliche Arbeitszeit in Stunden
f_workhours	3.28	3.25	0	15.27	Interaktionsterm: $female \cdot workhours$
female	0.46	0.50	0	1	=1, falls Frau, =0, falls Mann.
city	0.40	0.49	0	1	=1, falls in Großstadt wohnen, 0 sonst.
youngkid	0.13	0.34	0	1	=1, falls Person Kinder jünger als 3 Jahre hat, 0 sonst.

```
      Source |       SS       df       MS              Number of obs =     860
-------------+------------------------------           F(  5,   700) =   19.42
       Model |  81.70115       5    16.34023           Prob > F      =  0.0000
    Residual | 598.29569136   864   .6924718           R-squared     =  0.1218
-------------+------------------------------           Adj R-squared =  0.1155
       Total | 679.99684136   860  0.7906940           Root MSE      =  0.8321

------------------------------------------------------------------------------
       sleep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   workhours |  -.1420164   .0261594    -5.42   0.000    -.1933768   -.0906561
 f_workhours |  -.0539565   .0124899    -4.32   0.000    -.0784367   -.0294763
      female |   .3654238   .1233765     2.96   0.011     .1235059    .6072417
        city |  -.1839444   .0767387      ?      ?          ?           ?
    youngkid |  -.0739644   .1127516    -0.66   0.512    -.2953362    .1474074
       _cons |   8.479748   .1257961    67.41   0.000     8.232765    8.726731
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die <u>*dritte*</u> *Nachkommastelle.*

1.1 Interpretieren Sie den geschätzten Koeffizienten der Variable *youngkid* statistisch und inhaltlich. [2 Punkte]

- Statistisch: der geschätzte Koeffizient von *youngkid* ist auf dem 5%-Niveau statistisch nicht signifikant von Null verschieden. [1 Punkte]
- Inhaltlich: Personen, die ein Kind unter drei Jahren haben, schlafen c.p. im Durchschnitt 0,07 Stunden pro Tag weniger als Personen, die kein Kind unter drei Jahren haben. [1 Punkt]

1.2 Berechnen und interpretieren Sie das 95%-Konfidenzintervall für den geschätzten Koeffizienten der Variable *city*. Gehen Sie darauf ein, ob der Koeffizient statistisch signifikant von Null verschieden ist. [3 Punkte]

- t-Wert in Tabelle ablesen: 1,96 (df=854, α /2 = 0,975). [0,5 Punkte]
- Obere Grenze: $-0,184 + 1,96 * 0,077 = -0,033$. [0,5 Punkte]
- Untere Grenze: $-0,184 - 1,96 * 0,077 = -0,335$. [0,5 Punkte]
- (Das 95%-Konfidenzintervall des Koeffizienten von *city* lautet: [-0,335; -0,033])

- Interpretation: Mit wiederholten Stichproben liegt das wahre β_{city} in 95% der Fälle im auf diese Weise berechneten Konfidenzintervall. [0,5 Punkte]
- Da der Wert 0 nicht im 95%-Konfidenzintervall enthalten ist, ist der Koeffizient auf dem 5%-Niveau statistisch signifikant von Null verschieden. [1 Punkte]

1.3 Wie würden Sie testen, ob die Schlafdauer von in einer Großstadt wohnenden und nicht in einer Großstadt wohnenden Frauen ceteris paribus gleich ist? Erläutern Sie knapp verbal ihr Vorgehen. [2 Punkte]

- Die Testung könnte durch die Einführung eines Interaktionsterms zwischen den Variablen *female* und *city* in das Modell erfolgen. [1P]
- Bei Signifikanz des Koeffizienten des Interaktionsterms kann auf einen Unterschied in der Schlafdauer zwischen in der Großstadt und nicht in der Großstadt lebenden Frauen geschlossen werden. [1P]

1.4 Testen Sie mit einer Irrtumswahrscheinlichkeit von 1% anhand eines RESET-Tests mit Polynomen 2., 3., 4. Grades, ob das Modell fehlspezifiziert ist. Beschreiben Sie Ihr Vorgehen, geben Sie dabei Hilfsregression, Hypothesen, Teststatistik, Entscheidungsregel und Testentscheidung an. *Hinweis* : Verwenden Sie $S_1 = 542,223$ als Fehlerquadratsumme des unrestringierten Modells. [4,5 Punkte]

- Die Teststatistik basiert auf folgender Hilfsregression: [1P]

$$y_i = \mathbf{x_i'}\beta + \alpha_2 \hat{y}_i^2 + \alpha_3 \hat{y}_i^3 + \alpha_4 \hat{y}_i^4 + \nu_i$$

- Hypothesen: $H_0 : \alpha_2 = \alpha_3 = \alpha_4 = 0$, H_1 : mindestens ein $\alpha_j \neq 0$ mit $j = 2,3,4$ [0,5P]
- Teststatistik:

$$F^{emp} = \frac{(S_0 - S_1)/J}{S_1/(N-K)}$$

- Entscheidungsregel: H_0 verwerfen, falls $F^{emp} > F^{krit}_{J,N-K,0,05} = F^{krit}_{3;860-6-3;0,01} = 3,80$ [1P] (Angabe des kritischen Wertes an dieser Stelle nicht notwendig).

- Berechnung:

$$F^{emp} = \frac{(598,296 - 542,223)/3}{542,223/851} = 29,33 \quad [1P]$$

- Testentscheidung: Da $29,33 = F^{emp} > F^{krit}_{3,851;0,01} = 3,80$ muss die Nullhypothese auf dem 1%-Niveau verworfen werden. Der Test liefert Evidenz für eine Fehlspezifikation des Modells. [1P]

Aufgabe 2: **[17,5 Punkte]**

Der Zusammenhang zwischen Job-Zufriedenheit und gewünschter Arbeitszeit wird mit einem binären Logit-Modell und einer unabhängig gezogenen Stichprobe von N = 1230 Individuen geschätzt. Die nachfolgende Tabelle enthält die deskriptiven Statistiken und eine Beschreibung der Variablen. In der darauf folgenden Tabelle sind die Regressionsergebnisse ausgewiesen.

Variable	Mittelwert	Std. Abw.	Min.	Max.	Beschreibung
jobsat	0.07	0.25	0	1	=1, falls Person mit ihrem Job zufrieden ist, =0 sonst.
age	40.87	6.15	21	57	Alter in Jahren
mismatch	0.68	0.47	0	1	=1, falls gewünschte von tatsächlicher Arbeitszeit abweicht, =0 sonst.

```
Logistic regression                                  Number of obs = 1230
                                                           LR chi2(2) = ?
                                                     Prob > chi2 = 0.0005
Log likelihood = -155.376716                          Pseudo R2 = 0.0421

------------------------------------------------------------------------------
      jobsat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0576807   .0311626    -1.83   0.067    -0.118759   0.0033979
    mismatch |  -1.024564   .2998061    -3.42   0.001    -1.612173   -.4369546
       _cons |  -.2725406   .9748076    -0.28   0.780    -2.183128   1.6380823
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

2.1 Welche zwei Gründe sprechen dagegen, dieses Modell mit dem KQ-Verfahren zu schätzen? [2 Punkte]

- Ein Problem ist, dass die Varianz des KQ-Schätzers im Falle einer binären abhängigen Variable nicht konstant ist, es liegen daher heteroskedastische Störterme vor. [1P]

- Ein weiteres Problem ist, dass im linearen Modell die prognostizierten Wahrscheinlichkeiten außerhalb des Intervalls [0,1] liegen können. [1P]

2.2 Beschreiben Sie knapp das Prinzip des Maximum-Likelihood Verfahrens zur Schätzung der Parameter. [1,5 Punkte]

- Bei Maximum-Likelihood-Schätzverfahren wird der Vektor der unbekannten Parameter β so bestimmt, dass die Wahrscheinlichkeit, dass genau die vorliegenden Daten generiert werden, maximiert wird. [1,5 Punkte]

2.3 Erläutern Sie jeweils knapp die Vorgehensweise der Berechnung eines marginalen Effekts am Mittel der Daten und eines mittleren marginalen Effekts. [2 Punkte]

- Der marginale Effekt am Mittel der Daten wird anhand der Stichprobenmittelwerte aller erklärender Variablen berechnet. [1P]
- Für den mittleren marginalen Effekt wird für jede Beobachtung der marginale Effekt berechnet und aus diesen das arithmetische Mittel gebildet. [1P]

2.4 Bestimmen und interpretieren Sie den marginalen Effekt des Alters am Mittel der Daten. [5 Punkte]

- Marginaler Effekt $\frac{\partial E[y|X]}{\partial age_i} = \frac{e^{x'\beta}}{(1+e^{x'\beta})^2}\beta_2$ am Mittel der Daten:
- $\bar{x}'b = -0,273 - 0,058 \cdot 40,87 - 1,025 \cdot 0,68 = -3,34$ [2P]

- $\frac{\partial E[y|X]}{\partial age_i} = \frac{e^{-3,34}}{(1+e^{(-3,34)})^2} \cdot (-0,058) = -0,002$ [2P]
- Mit jedem weiteren Lebensjahr sinkt die Wahrscheinlichkeit, dass ein Individuum mit seinem Job zufrieden ist, im Mittel ceteris paribus um 0,2 Prozentpunkte. [1P]

2.5 Wie hoch ist die Wahrscheinlichkeit der Zufriedenheit mit dem Job für 32-jährige Personen ($age = 32$), deren tatsächliche Arbeitszeit nicht von der gewünschten Arbeitszeit abweicht ($mismatch = 0$)? Zeigen Sie Ihren Rechenweg und formulieren Sie einen Antwortsatz. [3 Punkte]

- Berechnung der Wahrscheinlichkeit der Zufriedenheit mit dem Job für 32-jährige Personen, deren Arbeitszeit nicht von der gewünschten Arbeitszeit abweicht:

$$P(jobsat = 1|age = 32, mismatch = 0) = \Lambda(-0,273 - 0,058 \cdot 32 - 1,025 \cdot 0) =$$

$$\Lambda(-2,129) = \frac{1}{1 + \exp(-(-2,129))} = 0,106 \text{ [2,5P]}$$

- Die vorhergesagte Wahrscheinlichkeit der Zufriedenheit mit dem Job für 32-jährige Personen, die mit ihrer Arbeitszeit zufrieden sind beträgt 10,6%. [0,5P]

2.6 Beurteilen Sie die Signifikanz des Modells anhand eines Likelihood-Ratio-Tests am 5% Niveau. Geben Sie Null- und Alternativhypothesen, Entscheidungsregel mit kritischem Wert und Testergebnis an. Hinweis: Ein Modell, das nur mit einer Konstanten geschätzt wurde, liefert einen Wert der Log-Likelihood-Funktion von -185,827. [4 Punkte]

- Hypothesen: H_0: alle Steigungsparameter = 0 gegen H_1: nicht H_0 [0,5P]
- Teststatistik: $\xi_{LR} = -2[ln(\tilde{\theta}) - ln(\widehat{\theta})] \sim \chi^2_J$ [1P]
- Entscheidungsregel : H_0 verwerfen, falls $\xi_{LR} > \chi^2_{J=2;\alpha=0,05} = 5,99$ [0,5P]
- Berechnung: $\xi_{LR} = -2[-185,827 - (-155,377)] = 60,9$ [1P]
- Entscheidung: H_0 wird verworfen. Die beiden Variablen *age* und *mismatch* führen zu einer signifikanten Verbesserung des Log-Likelihood-Wertes. [1P]

Aufgabe 3: [14 Punkte]

Sie interessieren sich für die Determinanten des Bruttostundenlohns in den USA. Dazu liegt Ihnen ein Datensatz von 3010 Personen mit folgenden Variablen vor:

$lnwage_i$	Bruttostundenlohn von Person i in US-Dollar ($)
$educ_i$	Dauer der Ausbildung von Person i in Jahren
exp_i	Arbeitsmarkterfahrung von Person i in Jahren
$exp2_i$	quadrierte Arbeitsmarkterfahrung von Person i in Jahren
$ethn_i$	Dummy-Variable,=1 wenn Person i Afroamerikaner ist, =0 sonst
$city_i$	Dummy-Variable,=1 wenn Person i in einer Stadt wohnt, =0 sonst
$south_i$	Dummy-Variable,=1 wenn Person i im Süden der USA wohnt, =0 sonst
$nearcollege_i$	Entfernung des Wohnortes, in dem Person i aufgewachsen ist, zu dem nächstgelegenen College in Meilen

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$lnwage_i = \beta_1 + \beta_2 educ_i + \beta_3 exp_i + \beta_4 exp2_i + \beta_5 ethn_i + \beta_6 city_i + \beta_7 south_i + \varepsilon_i$$

```
Source       |       SS           df       MS      Number of obs   =     3,010
-------------+----------------------------------   F(6, 3003)      =    204.93
Model        |  172.165628         6  28.6942714   Prob > F        =    0.0000
Residual     |  420.476016     3,003  .140018653   R-squared       =    0.2905
-------------+----------------------------------   Adj R-squared   =    0.2891
Total        |  592.641645     3,009  .196956346   Root MSE        =    .37419

------------------------------------------------------------------------------
lnwage       |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
educ         |    .074009   .0035054    21.11   0.000     .0671357    .0808823
exp          |   .0835958   .0066478    12.57   0.000     .0705612    .0966305
exp2         |  -.0022409   .0003178    -7.05   0.000    -.0028641   -.0016177
ethn         |  -.1896315   .0176266   -10.76   0.000    -.2241929   -.1550702
city         |    .161423   .0155733    10.37   0.000     .1308876    .1919583
south        |  -.1248615   .0151182    -8.26   0.000    -.1545046   -.0952184
_cons        |   4.886734   .0785544    62.21   0.000     4.732709     5.04076
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die <u>dritte</u> Nachkommastelle.

3.1 Bei welchem Wert der Variable *exp* ist der erwartete Lohn am höchsten? [2 Punkte]

- $\frac{\partial E[lnwage|\mathbf{x}_i]}{\partial exp} = \beta_3 + 2 \cdot \beta_4 \cdot exp \stackrel{!}{=} 0$ [1P]
- $exp = -\frac{b_3}{2b_4} \Rightarrow -\frac{0,084}{2(-0,002)} = 21$ [1P]
- Bei dem Wert $exp \approx 21$ ist der erwartete Lohn am höchsten.

3.2 Wann liegt ein Problem ausgelassener Variablen vor? Nennen Sie allgemein die nötigen Bedingungen. Diskutieren Sie am vorliegenden Beispiel, ob bei der Schätzung des Effekts von *educ* ein Problem ausgelassener Variablen vorliegt. [2 Punkte]

- Ein Problem ausgelassener Variablen liegt vor, wenn (i) eine erklärende Variable mit einer ausgelassenen Variable korreliert und (ii) die ausgelassene Variable mit der abhängigen Variable korreliert. [je 0,5P]
- Beispielsweise könnte die Bildung mit der Bildung der Eltern korrelieren, die hier nicht berücksichtigt wird. Die Bildung der Eltern könnte auch mit dem Lohn der Kinder zusammenhängen. Somit liegt vermutlich ein Problem ausgelassener Variablen vor. [1P]
  (Andere Antworten sind möglich)

3.3 Ein Kommilitone möchte die Variable *nearcollege* als Instrumentvariable für *educ* nutzen. Diskutieren Sie, ob die Variable *nearcollege* als Instrumentvariable für die Bildung geeignet ist. Nennen und erklären Sie kurz die Bedingungen, die hierfür erfüllt sein müssen und beurteilen Sie, ob diese im vorliegenden Fall erfüllt sind. [4 Punkte]

- Relevanz: Das Instrument muss mit der endogenen Variable partiell korreliert sein. [1P] Diese Bedingung ist vermutlich erfüllt, da die geografische Entfernung negativ mit dem Besuch der Hochschule korrelieren könnte. [1P]

- Exogenität: Das Instrument darf nicht mit unbeobachteten Determinanten der abhängigen Variable korreliert sein. [1P] Diese Bedingung ist hier vermutlich erfüllt, da es unwahrscheinlich ist, dass sich die unbeobachteten Determinanten des Bruttostundenlohns (z. B. Fähigkeiten) mit der Entfernung zum nächstmöglichen College unterscheiden (auch andere Antwort zulässig). [1P]

3.4 Sie entscheiden sich dafür, die Dummyvariable *nearcollege* als Instrument in einer Two Stage Least Squares (2SLS)-Schätzung zu verwenden. Erläutern Sie kurz verbal die Vorgehensweise des 2SLS-Schätzers und stellen Sie die beiden benötigten Modellgleichungen auf. [4 Punkte]

- Der 2SLS-Schätzer ist ein zweistufiger KQ-Schätzer.
- 1. Stufe: Die endogene Variable *educ* wird auf das Instrument *nearcollege* und alle exogenen Regressoren regressiert. [1P]
- 2.Stufe: Die Bruttostundenlohnregression wird geschätzt, wobei für *educ* die vorhergesagten Werte aus Stufe 1 eingesetzt werden. [1P]

1.Stufe:

$$educ_i = \theta_1 + \theta_2 nearcollege_i + \theta_3 exp_i + \theta_4 exp2_i + \theta_5 ethn_i + \theta_6 city_i + \theta_7 south_i + \upsilon_i \quad \text{[1P]}$$

2.Stufe:

$$lnwage_i = \beta_1^{IV} + \beta_2^{IV} \widehat{educ_i} + \beta_3^{IV} exp_i + \beta_4^{IV} exp2_i + \beta_5^{IV} ethn_i + \beta_6^{IV} city_i + \beta_7^{IV} south_i + \varepsilon_i \quad \text{[1P]}$$

3.5 Sie schätzen die 2SLS-Schätzung mittels Stata und erhalten folgenden Output:

```
Instrumental variables (2SLS) regression          Number of obs   =      3,010
                                                  Wald chi2(6)    =     726.69
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2252
                                                  Root MSE        =     .39058

------------------------------------------------------------------------------
lnwage       |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
educ         |   .1322888    .049176     2.69   0.007     .0359057    .2286719
exp          |    .107498   .0212758     5.05   0.000     .0657981    .1491978
exp2         |  -.0022841   .0003337    -6.84   0.000    -.0029382   -.0016299
ethn         |  -.1308019   .0528108    -2.48   0.013    -.2343091   -.0272946
city         |   .1313237   .0300948     4.36   0.000      .072339    .1903083
south        |  -.1049005   .0230463    -4.55   0.000    -.1500704   -.0597307
_cons        |    3.85716    .870208     4.43   0.000     2.151584    5.562737
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exp c.exp#c.exp ethn city south nearcollege
```

Sie sehen sich zusätzlich Teststatistiken der ersten Stufe dieser 2SLS-Schätzung an, die mit dem *estat first*-Befehl in Stata erzeugt wurden:

```
First-stage regression summary statistics
--------------------------------------------------------------------------
             |            Adjusted      Partial
Variable     |   R-sq.       R-sq.        R-sq.     F(1,3003)   Prob > F
-------------+------------------------------------------------------------
educ         |  0.4745      0.4734       0.0055       16.7176    0.0000
--------------------------------------------------------------------------
```

Erläutern Sie den Begriff des schwachen Instruments. Handelt es sich in dem vorliegenden Fall um ein schwaches Instrument? [2 Punkte]

- Ein schwaches Instrument liegt vor, wenn der Zusammenhang zwischen der Instrumentvariable und der endogenen Variable nur relativ gering ist. [1P]
- Man spricht bei einem F-Wert des Signifikanztests des Instruments in der Regression der ersten Stufe kleiner 10 von einem schwachen Instrument.

=> Bei dem vorliegenden Instrument handelt es sich um ein starkes Instrument, da der F-Wert des Signifikanztests des Instruments größer als 10 ist. [1P]

Aufgabe 4: **[17 Punkte]**

Sie interessieren sich für die Determinanten der Anzahl von Straftaten in einer fränkischen Großstadt. Ihr Datensatz enthält folgende Informationen für 12 Zeitpunkte:

$Straftaten_t$	Anzahl an Straftaten pro 100.000 EinwohnerInnen in Jahr t
alr_t	Lokale Arbeitslosenrate in % in Jahr t (kodiert 0-100)
$antf_t$	Bevölkerungsanteil von Frauen in % in Jahr t (kodiert 0-100)
$jahr_t$	Jahr t (2008-2019)

Sie stellen folgendes lineares Regressionsmodell auf und schätzen dieses anschließend mit Stata:

$$straftaten_t = \beta_1 + \beta_2 alr_t + \beta_3 antf_t + \beta_4 jahr_t + \varepsilon_t$$

```
Source       |      SS           df       MS          Number of obs   =        12
-------------+----------------------------------      F(3, 8)         =    ??????
Model        |  3728.36           3  1242.787         Prob > F        =    ??????
Residual     |  1108.60           8  138.5743         R-squared       =    ??????
-------------+----------------------------------      Adj R-squared   =    ??????
Total        |  4836.96          11  439.7232         Root MSE        =    11.772

------------------------------------------------------------------------------------
straftaten        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
alr               |   826.7158   170.1262    4.86   0.001     434.4041    1219.028
antf              |  -174.7433    92.3454   -1.89   0.095    -387.6922    38.20557
jahr              |   137.2902   50.74524    2.71   0.027     20.27147    254.3089
_cons             |    -264687   98526.12   -2.69   0.028    -491888.7   -37485.39
------------------------------------------------------------------------------------
```

4.1 Interpretieren Sie inhaltlich und statistisch den geschätzten Koeffizienten b_3. [2 Punkte]

- b_3: Erhöht sich der Bevölkerungsanteil von Frauen um 1 Prozentpunkt, so fällt c.p. im Durchschnitt die Anzahl an Straftaten pro Jahr um 174,74 pro 100.000 EinwohnerInnen. [1P]
- Der Koeffizient ist am 10% Signifikanzniveau signifikant. [1P]

4.2 Berechnen und interpretieren Sie das R^2 der Regression. [2 Punkte]

- $R^2 = MSS/TSS = 3728,36/4836,96 = 0,771$ [1P]
- Das Modell erklärt 77,1% der Variation der Straftaten pro 100.000 EinwohnerInnen. [1P]

4.3 Erläutern Sie knapp verbal, was unter Autokorrelation zu verstehen ist und nennen Sie zwei Konsequenzen von Autokorrelation für die Eigenschaften des KQ-Schätzers. [2 Punkte]

- Autokorrelation ist eine Situation, in der zeitlich aufeinanderfolgende Störterme korrelieren. [1P]
- Autokorrelation führt zur Ineffizienz der KQ-Parameterschätzer (Standardfehler sind falsch berechnet) [0,5P], Schätzer jedoch nach wie vor erwartungstreu. [0,5P] (Andere Antworten sind möglich)

4.4 Sie vermuten Autokorrelation und führen einen Breusch-Godfrey-Test auf Autokorrelation 1. bis 4. Ordnung auf dem 5%-Signifikanzniveau durch. Sie erhalten folgenden Stata-Output:

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
lags(p)      |          chi2               df              Prob > chi2
-------------+-------------------------------------------------------------
1            |          7.820               1                ??????
2            |          7.221               2                0.0270
3            |          6.864               3                0.0763
4            |          8.000               4                0.0916
---------------------------------------------------------------------------
H0: no serial correlation
```

Beschreiben Sie knapp die Komponenten der Teststatistik $LM = (T-1) \cdot R^2$ für den Test 1. Ordnung. Geben Sie eine Hilfsregression, die Null- und Alternativhypothese, kritischen Wert und Testergebnis für diesen Fall an. [5 Punkte]

- Hilfsregression: $e_t = \beta_1 + \rho e_{t-1} + \beta_2 alr_t + \beta_3 antf_t + \beta_4 jahr_t + \nu_t$; alternativ $e_t = \beta_1 + \rho e_{t-1} + \nu_t$. [1P]
- Komponenten der Teststatistik: $T-1$ entspricht der Anzahl an Messzeitpunkten - 1 (hier $12-1$) und R^2 ist das Bestimmtheitsmaß aus der Hilfsregression. [1P]
- Hypothesen: $H_0: \rho = 0$ (keine Autokorrelation 1. Ordnung); $H_1: \rho \neq 0$ (Autokorrelation 1. Ordnung). [1P]
- Kritischer Wert: $\chi^2_{1;5\%} = 3,842$. [1P]
- Testergebnis: Da $\chi^2_{empirisch} = 7,820 > 3,842 = \chi^2_{kritisch}$ kann die Nullhypothese auf dem 5% Signifikanzniveau verworfen werden. Es gibt Evidenz für Autokorrelation 1. Ordnung. [1P]

4.5 Nutzen Sie die Tabelle aus der Aufgabe 4.4. Korrelation welcher Ordnung vermuten Sie basierend auf den Tests am 5%-Signifikanzniveau? Begründen Sie Ihre Antwort (*Falls Sie das Ergebnis der Aufgabe 4.4 nicht haben, unterstellen Sie für den Test 1. Ordnung einen p-Wert von 0.01*). [2 Punkte]

- Korrelation 2. Ordnung [1P]
- Die Nullhypothese des Tests auf Autokorrelation 3. Ordnung am 5% Signifikanzniveau kann nicht abgelehnt werden. [1P]

4.6 Erläutern Sie die Idee des Cochrane-Orcutt Schätzers. Setzen Sie voraus, dass $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$ mit $\nu_t \overset{iid}{\sim} (0, \sigma^2_\nu)$ gilt und ρ bekannt ist. Benennen Sie knapp, wodurch sich dieses Verfahren vom Prais-Winsten Schätzer unterscheidet. [2 Punkte]

- Idee: Transformation des Modells: $y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + \varepsilon_t - \rho\varepsilon_{t-1} = y_t^* = (x_t^*)'\beta + \nu_t$
  Die Koeffizienten β werden nicht mit den originalwerten y_t und x_t, sondern mit den transformierten Werten $y_t^* = y_t - \rho y_{t-1}$ und $x_t^* = x_t - \rho x_{t-1}$ geschätzt. Da $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$ generiert eine Transformation von ε_t in $\nu_t = \varepsilon_t - \rho\varepsilon_{t-1}$ nicht-autokorrelierte Störterme. [1P]

- Cochrane-Orcutt kann die erste Beobachtung nicht nutzen, der Prais-Winsten Schätzer macht dies möglich. [1P]

4.7 Eine Alternative zum Breusch-Godfrey-Test ist der Durbin Watson Test. Erläutern Sie zwei Gründe warum Sie sich gegen bzw. für die Anwendung von dem Durbin Watson Test in diesem Fall entscheiden würden [2 Punkte].

Jeweils ein Punkt (insgesamt max. 2 Punkte) für folgende Antworten:

- Er eignet sich nur für den Test der Autokorrelation 1. Ordnung, deswegen ist er hier nicht anwendbar. [1P]
- Das Modell muss eine Regressionskonstante enthalten, daher ist die Verwendung hier möglich. [1P]
- Wegen des Unschärfebereichs bei den kritischen Werten sind manchmal keine Aussagen bezüglich der Autokorrelation möglich. [1P]
- Alternative Antworten möglich.

Aufgabe 5 – MC Fragen **[30 Punkte]**

Bitte geben Sie die zutreffende Antwort **auf Ihrem Multiple-Choice-Lösungsblatt** an. Zu jeder Frage gibt es genau eine richtige Antwort. Für jede korrekt beantwortete Frage erhalten Sie einen Punkt. Falsche Antworten führen nicht zu Punktabzug. Bei mehr oder weniger als einer markierten Antwort auf eine Frage gilt diese als nicht beantwortet. **Angaben auf dem Aufgabenblatt werden nicht gewertet.**

1.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$. Die aufgestellten Hypothesen lauten $H_0 : \beta_3 \leq 3$ vs. $H_1 : \beta_3 > 3$. Auf einem Signifikanzniveau von 1% führt eine t-Teststatistik von $2,68$
a	bei $n = 8$ zur Ablehnung von H_0.
b	bei $n = 10$ zur Ablehnung von H_0.
c	bei $n = 12$ zur Ablehnung von H_0.
d	bei $n = 14$ zur Ablehnung von H_0. **X**

2.	Wann liegt ein Typ II-Fehler vor ? Wenn...
a	eine wahre Nullhypothese nicht abgelehnt wird.
b	eine falsche Nullhypothese nicht abgelehnt wird. **X**
c	eine wahre Nullhypothese abgelehnt wird.
d	eine falsche Nullhypothese abgelehnt wird.

3.	Wozu führt die Aufnahme einer irrelevanten Variable x_3 in ein lineares Regressionsmodell $y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$:
a	bei kleinen Stichproben zu inkonsistenten Schätzern.
b	bei konstanter Anzahl an Beobachtungen c.p. zu einer Erhöhung der Varianz für den Parameterschätzer β_2, wenn $cov(x_2; x_3) \neq 0$. **X**.
c	bei Multikollinearität zu verzerrten Schätzern.
d	einer reduzierten Varianz des Störterms.

4.	Wenn in den erklärenden Variablen Polynome enthalten sind, dann:
a	kann ein gemeinsam signifikanter Erklärungsgehalt der Polynome anhand eines F-Tests überprüft werden. **X**
b	können lineare Zusammenhänge zwischen x und y damit nicht abgebildet werden.
c	ergibt sich zwingend ein Problem der Multikollinearität.
d	ergibt sich der marginale Effekt der Variablen x als Ableitung des auf y bedingten Erwartungswerts von x.

5.	Welche Aussage bezüglich marginaler Effekte ist richtig?
a	Der marginale Effekt am Mittel der Daten aus Logit-Regressionen variiert nicht zwischen Individuen. **X**
b	Marginale Effekte können nur für binäre Regressoren berechnet werden.
c	Die Schätzkoeffizienten von Logit-Regressionen können als marginale Effekte interpretiert werden.
d	Marginale Effekte in Probit-Regressionen können nur bezüglich der Signifikanz und des Vorzeichens interpretiert werden.

6.	Wozu führt das Vorliegen von nicht korrigierter Autokorrelation?
a	zu systematischem Messfehler in der abhängigen Variable.
b	zu falsch berechneten Schätzkoeffizienten.
c	zu überschätzten R^2-Werten.
d	zu falsch berechneten Konfidenzintervallen. **X**

7.	Welche Aussage zum Bestimmtheitsmaß im multiplen Regressionsmodell mit endlicher Stichprobe ist richtig?
a	es entspricht dem Verhältnis der Störtermvariation zur Gesamtvariation.
b	ist immer kleiner als das korrigierte Bestimmtheitsmaß $\bar{R}^2$.
c	ist immer größer als das korrigierte Bestimmtheitsmaß $\bar{R}^2$. **X**
d	es entspricht dem Verhältnis erklärter Variation zur Störtermvariation.

8.	Welche Aussage ist richtig, wenn der F-Test für die Signifikanz einer Instrumentenvariable auf der ersten Stufe einer 2SLS-Schätzung einen F-Wert von 23,09 liefert.
a	Die Exogenitätsbedingung ist erfüllt.
b	Die Relevanzbedingung ist erfüllt. **X**
c	Es kann keine Aussage bezüglich der Relevanz des Instruments getroffen werden.
d	Das Instrument ist exogen.

9.	Worauf prüft ein Strukturbruchtest:
a	ob Autokorrelation bei Zeitreihendaten vorliegt.
b	ob die Kreuzproduktmatrix $\mathbf{X}'\mathbf{X}$ invertierbar ist.
c	ob über die gesamte Stichprobe einheitliche Zusammenhänge zwischen der abhängigen und den unabhängigen Variablen vorherrschen. **X**
d	ob zusätzliche erklärende Variablen die Varianz des Störterms negativ werden lassen.

10.	Logit- und Probit-Schätzer
a	unterliegen identischen Annahmen bezüglich der Fehlertermverteilung.
b	führen zu identischen marginalen Effekten.
c	werden nicht verwendet, wenn die abhängige Variable stetig ist. **X**
d	führen zu identischen Koeffizienten.

11.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{i2} + u_i$. b_2 beträgt -5 und $\widehat{Var}(b_2) = 0,25$. Die aufgestellten Hypothesen lauten $H_0 : \beta_2 \leq -4$ vs. $H_1 : \beta_2 > -4$. Die dazugehörige t-Teststatistik beträgt
a	$t^{empirisch} = 20$.
b	$t^{empirisch} = -4$.
c	$t^{empirisch} = -2$. **X**
d	$t^{empirisch} = 2$

12.	Schätzgleichungen, die quadratische unabhängige Variablen enthalten,
a	können anhand des KQ-Verfahrens berechnet werden. **X**
b	führen zu einem niedrigeren R^2 als Schätzungen ohne quadratische Terme.
c	können nur im Rahmen nicht linearer Regressionen geschätzt werden.
d	können für logarithmierte abhängige Variablen nicht berechnet werden.

13.	Es wird das Modell $ln(Stundenlohn_i) = \beta_1 + \beta_2 frau_i + \beta_3 educ_i + \beta_4 (frau_i \cdot educ_i) + \varepsilon_i$ mittels einer KQ-Schätzung geschätzt ($educ_i$ misst Bildung in Jahren, $frau_i = 1$ falls Frau). Welche Aussage ist richtig?
a	b_1 gibt den durchschnittlichen Stundenlohnen für Frauen mit 0 Bildungsjahren an.
b	b_3+b_4 gibt die geschätzte Bildungsrendite für Frauen an. **X**
c	b_4 gibt die geschätzte Bildungsrendite für Frauen an.
d	b_2 ist der Lohnunterschied zwischen Frauen und der restlichen Stichprobe in %.

14.	Wozu führt unkorrigierte Heteroskedastie im linearen Regressionsmodel?
a	zu falschen Werten der t-Statistik. **X**
b	zu Verzerrung des KQ-Schätzers.
c	zu Effizienz des KQ-Schätzers.
d	zu korrekten Standardfehlern des KQ-Schätzers.

15.	Sie testen auf positive Autokorrelation 1. Ordnung für ein lineares Modell mit einer Konstante und 5 unabhängigen Variablen mit 80 beobachteten Perioden auf einem Signifikanzniveau von 5%. Die empirische Durbin-Watson Statistik beträgt 1,89.
a	Es besteht keine Evidenz für Autokorrelation **X**
b	Es besteht Evidenz für Autokorrelation.
c	Eine Aussage über die Autokorrelation ist nicht möglich.
d	Es besteht Evidenz für Autokorrelation 2. Ordnung .

16.	Sie schätzen das Modell $inlf = \beta_1 + \beta_2 mann_i + \beta_3 kids_i + \varepsilon_i$ mittels einer KQ-Schätzung, wobei $inlf$ eine Dummy-Variable ist und den Wert 1 annimmt, wenn die Person erwerbstätig ist. *mann* ist eine Dummyvariable für Männer und *kinder* die Anzahl Kinder in der Familie ist. Der geschätzte Koeffizient zu β_3 ist $-0,02$. Welche Interpretation ist richtig?
a	Steigt die Anzahl Kinder um eins, so fällt die Wahrscheinlichkeit der Erwerbstätigkeit c.p. im Mittel um etwa 20%-Punkte.
b	Steigt die Anzahl Kinder um eins, so fällt die Wahrscheinlichkeit der Erwerbstätigkeit c.p. im Mittel um etwa 0,2%.
c	Steigt die Anzahl Kinder um eins, so fällt die Wahrscheinlichkeit der Erwerbstätigkeit c.p. im Mittel um etwa 20%.
d	Steigt die Anzahl Kinder um eins, so fällt die Wahrscheinlichkeit der Erwerbstätigkeit c.p. im Mittel um etwa 2%-Punkte. **X**

17.	In der zweiten Stufe einer 2SLS-Schätzung
a	liegt zwangsläufig Endogenität vor.
b	wird die endogene Variable aus der ersten Stufe auf das Instrument, aber nicht die exogenen Variablen regressiert.
c	wird die endogene Variable aus der ersten Stufe auf das Instrument und die exogenen Variablen regressiert.
d	Keine der Antworten ist richtig. **X**

18.	Die geschätzten Koeffizienten von Logit- und Probitmodellen
a	lassen sich als marginale Effekte interpretieren.
b	lassen sich nur bezüglich der Signifikanz und des Vorzeichens interpretieren. **X**
c	sind für einzelne Beobachtungen nicht konstant.
d	Keine der Aussagen ist richtig.

19.	Die Inverse der matrix $\Psi = \begin{pmatrix} ausb_1 & 0 & 0 \\ 0 & ausb_2 & 0 \\ 0 & 0 & ausb_3 \end{pmatrix}$ ist
a	$\begin{pmatrix} 0 & 0 & ausb_1 \\ 0 & ausb_2 & 0 \\ ausb_3 & 0 & 0 \end{pmatrix}$
b	$\begin{pmatrix} ausb_1 & 0 & 0 \\ 0 & ausb_2 & 0 \\ 0 & 0 & ausb_3 \end{pmatrix}$
c	$\begin{pmatrix} \frac{1}{ausb_1} & 0 & 0 \\ 0 & \frac{1}{ausb_2} & 0 \\ 0 & 0 & \frac{1}{ausb_3} \end{pmatrix}$ **X**
d	$\begin{pmatrix} \sqrt{ausb_1} & 0 & 0 \\ 0 & \sqrt{ausb_2} & 0 \\ 0 & 0 & \sqrt{ausb_3} \end{pmatrix}$

20.	Die Varianz-Kovarianzmatrix der Störterme eines Regressionsmodells sei $\sigma^2\Psi$, wobei $\Psi = \begin{bmatrix} 3 & 0 & 4 \\ 0 & 3 & 0 \\ 2 & 0 & 3 \end{bmatrix}$. Hier gilt:
a	Homoskedastie und positive Autokorrelation. **X**
b	Heteroskedastie und positive Autokorrelation.
c	Homoskedastie und keine Autokorrelation.
d	Heteroskedastie und keine Autokorrelation.

21.	Der KQ-Schätzer b lässt sich nicht berechnen, wenn
a	die Matrix $X'X$ nicht singulär ist.
b	die Matrix $X'X$ invertierbar ist.
c	die Matrix $X'X$ vollen Spaltenrang aufweist.
d	perfekte Multikollinearität besteht. **X**

22.	Was ergibt sich aus dem Produkt der Matrizen $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ und $B = \begin{pmatrix} 3 & 1 \\ 7 & 2 \end{pmatrix}$?
a	$AB = \begin{pmatrix} 3 & 7 \\ 1 & 2 \end{pmatrix}$.
b	$AB = \begin{pmatrix} 2 & 1 \\ 7 & 3 \end{pmatrix}$.
c	$AB = \begin{pmatrix} 3 & 1 \\ 7 & 2 \end{pmatrix}$. **X**
d	Die Matrix AB ist nicht definiert.

23.	In einer 2SLS-Schätzung werden die Werte der endogenen Variable in der zweiten Stufe ersetzt durch
a	die Residuen der ersten Stufe Schätzung auf der zweiten Stufe.
b	die quadrierten Residuen der ersten Stufe.
c	die quadrierten Werte der Instrumentvariable.
d	die vorhergesagten Werte der abhängigen Variable aus der ersten Stufe. **X**

24.	Für das Modell $income_i = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 exper_i^2 + u_i$ ergibt eine KQ-Schätzung $b_1 = 100$, $b_2 = 50$, $b_3 = 20$ und $b_4 = -0,5$. Bei welchem Wert von $exper$ ist der marginale Effekt von Berufserfahrung -1 ?
a	21. **X**
b	18.
c	-21.
d	40.

25.	Sie schätzen das Modell $log(einkommen_i) = \beta_0 + \beta_1 female_i + \beta_2 arbeitszeit_i + u_i$. Welcher Koeffizient gibt eine Elastizität an?
a	$\hat{\beta}_0$.
b	$\hat{\beta}_1$.
c	$\hat{\beta}_2$.
d	Keiner der Koeffizienten. **X**

26.	Sie möchten überprüfen, ob es im Modell $y_i = \beta_0 + \beta_1 Einkommen_i + \beta_2 Alter_i + \beta_3 Mann_i + \varepsilon_i$ signifikante Unterschiede in den Steigungsparametern zwischen Männern und Frauen gibt. Wie viele Parameter müssen Sie insgesamt im Rahmen eines vollständig interagierten Modells schätzen?
a	4.
b	6. **X**
c	7.
d	8.

27.	Der Durbin-Watson Test
a	ist bei positiver Autokorrelation nicht durchführbar.
b	ist auch bei Schätzung ohne Konstante gültig.
c	eignet sich zum Testen auf Heteroskedastie.
d	ist auch in kleinen Stichproben gültig. **X**

28.	Die Nullhypothese im Breusch-Pagan Test mit N Beobachtungen
a	wird abgelehnt, wenn $R^2 \cdot N \leq \chi^2_{kritisch}$.
b	besagt, dass Autokorrelation vorliegt.
c	besagt, dass Homoskedastie vorliegt.**X**
d	besagt, dass Autokorrelation und Heteroskedastie vorliegen.

29.	Autokorrelation im Störterm kann behoben werden durch
a	die Aufnahme von irrelevanten erklärenden Variablen.
b	eine ML-Transformation.
c	eine Vergrößerung der Stichprobe.
d	eine FGLS Schätzung. **X**

30.	Wenn $\Psi = \mathbf{I}$ gilt, dann impliziert $Var(\varepsilon) = \sigma^2\Psi$,
a	dass Heteroskedastie vorliegt.
b	dass $Var(\varepsilon) = 1$.
c	dass $cov(\varepsilon_i, \varepsilon_j) \neq 0$ für alle $i \neq j$.
d	dass $cov(\varepsilon_i, \varepsilon_j) = 0$ für alle $i \neq j$. **X**

Tabelle 2: Perzentile der *t*-Verteilung

Zelleneintrag: x, sodass Prob$\left[t_n \leq x\right] = P$, mit n Freiheitsgraden

n \ P	0.75	0.9	0.95	0.975	0.99	0.995
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.817	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.500
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.813	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.696	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.320	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
35	0.682	1.306	1.690	2.030	2.438	2.724
40	0.681	1.303	1.684	2.021	2.423	2.705
45	0.680	1.301	1.679	2.014	2.412	2.690
50	0.679	1.299	1.676	2.009	2.403	2.678
60	0.679	1.296	1.671	2.000	2.390	2.660
70	0.678	1.294	1.667	1.994	2.381	2.648
80	0.678	1.292	1.664	1.990	2.374	2.639
90	0.677	1.291	1.662	1.987	2.368	2.632
100	0.677	1.290	1.660	1.984	2.364	2.626
∞	0.674	1.282	1.645	1.960	2.326	2.576

Quelle: In R generiert

Tabelle 3: Perzentile der χ^2-Verteilung

Zelleneintrag: c, sodass Prob $\left[\chi_n^2 \leq c\right] = P$, mit n Freiheitsgraden

n \ P	0.005	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	0.995
1	0.00004	0.0002	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.842	5.024	6.635	7.879
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
35	17.19	18.51	20.57	22.47	24.80	29.05	34.34	40.22	46.06	49.80	53.20	57.34	60.27
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
45	24.31	25.90	28.37	30.61	33.35	38.29	44.34	50.98	57.51	61.66	65.41	69.96	73.17
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49

Quelle: In R generiert

Tabelle 4b: 99% Perzentile der *F*-Verteilung

Zelleneintrag: f, sodass Prob$\left[F_{n1,n2} \leq f\right] = 0.99$

n2	**n1 = Freiheitsgrade des Zählers** 1	2	3	4	5	6	7	8	9
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
∞	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43

n2	**n1 = Freiheitsgrade des Zählers** 10	12	15	20	30	40	50	60	∞
1	6055.85	6106.32	6157.28	6208.73	6260.65	6286.78	6302.52	6313.03	6362.68
2	99.40	99.42	99.43	99.45	99.47	99.47	99.48	99.48	99.50
3	27.23	27.05	26.87	26.69	26.50	26.41	26.35	26.32	26.14
4	14.55	14.37	14.20	14.02	13.84	13.75	13.69	13.65	13.47
5	10.05	9.89	9.72	9.55	9.38	9.29	9.24	9.20	9.03
6	7.87	7.72	7.56	7.40	7.23	7.14	7.09	7.06	6.89
7	6.62	6.47	6.31	6.16	5.99	5.91	5.86	5.82	5.66
8	5.81	5.67	5.52	5.36	5.20	5.12	5.07	5.03	4.87
9	5.26	5.11	4.96	4.81	4.65	4.57	4.52	4.48	4.32
10	4.85	4.71	4.56	4.41	4.25	4.17	4.12	4.08	3.92
15	3.80	3.67	3.52	3.37	3.21	3.13	3.08	3.05	2.88
20	3.37	3.23	3.09	2.94	2.78	2.69	2.64	2.61	2.43
25	3.13	2.99	2.85	2.70	2.54	2.45	2.40	2.36	2.18
30	2.98	2.84	2.70	2.55	2.39	2.30	2.25	2.21	2.02
40	2.80	2.66	2.52	2.37	2.20	2.11	2.06	2.02	1.82
50	2.70	2.56	2.42	2.27	2.10	2.01	1.95	1.91	1.70
70	2.59	2.45	2.31	2.15	1.98	1.89	1.83	1.78	1.56
100	2.50	2.37	2.22	2.07	1.89	1.80	1.74	1.69	1.45
∞	2.34	2.20	2.06	1.90	1.72	1.61	1.54	1.50	1.16

Quelle: In R generiert

Tabelle 5: Durbin-Watson Teststatistik

d_L und d_U am 5% Signifikanzniveau

	k=2		**k=3**		**k=4**		**k=5**		**k=6**		**k=11**		**k=16**	
n	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21				
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15	0.16	3.30		
17	1.13	1.38	1.02	1.53	0.90	1.71	0.78	1.90	0.67	2.21	0.20	3.18		
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.15	0.24	3.07		
19	1.18	1.40	1.08	1.54	0.97	1.68	0.86	1.85	0.75	2.10	0.29	2.97		
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	2.06	0.34	2.89	0.06	3.68
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	2.02	0.38	2.81	0.09	3.58
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.99	0.42	2.73	0.12	3.55
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.96	0.47	2.67	0.15	3.41
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.64	0.51	2.61	0.19	3.33
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.92	0.54	2.57	0.22	3.25
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.90	0.58	2.51	0.26	3.18
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.89	0.62	2.47	0.29	3.11
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.88	0.65	2.43	0.33	3.05
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.86	0.68	2.40	0.36	2.99
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.85	0.71	2.36	0.39	2.94
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.84	0.74	2.33	0.43	2.99
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.83	0.77	2.31	0.46	2.84
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.83	0.80	2.28	0.49	2.80
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.82	0.82	2.26	0.52	2.75
35	1.40	1.52	1.34	1.53	1.28	1.65	1.22	1.73	1.16	1.81	0.85	2.24	0.55	2.72
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.81	0.87	2.22	0.58	2.68
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80	0.89	2.20	0.60	2.65
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.80	0.91	2.18	0.63	2.61
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.80	0.93	2.16	0.65	2.59
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	0.95	2.15	0.68	2.56
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.79	1.04	2.09	0.79	2.44
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.79	1.11	2.04	0.88	2.35
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.78	1.17	2.01	0.96	2.28
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.22	1.98	1.03	2.23
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77	1.27	1.96	1.09	2.18
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77	1.30	1.95	1.14	2.15
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77	1.34	1.94	1.18	2.12
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.37	1.93	1.22	2.09
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77	1.40	1.92	1.26	2.07
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.42	1.91	1.29	2.06
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78	1.44	1.90	1.32	2.04
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.46	1.90	1.35	2.03
150	1.72	1.75	1.72	1.76	1.71	1.76	1.69	1.77	1.68	1.79	1.61	1.86	1.54	1.52
200	1.76	1.78	1.76	1.78	1.75	1.79	1.74	1.80	1.73	1.81	1.68	1.86	1.62	1.61

Anmerkung: k ist die Anzahl der Regressoren mit Konstante.

Quelle: Johnston, J. and J. DiNardo, 1997, *Econometric Methods*, 4^{th} ed., New York: McGraw-Hill

Formeln Ökonometrie

I. Mathematische Grundlagen

i. Algebra

$(AB)' = B'A'$

$(A')' = A$

$AA^{-1} = I$ und $A^{-1}A = I$

$(A^{-1})' = (A')^{-1}$

$(AB)^{-1} = B^{-1}A^{-1}$

Ableitung von Matrizen

Für die Matrix A und die Vektoren x und c gilt bei passender Ordnung:

$$\frac{\partial c'x}{\partial x} = c$$

$$\frac{\partial Ax}{\partial x} = A'$$

Wenn A <u>symmetrisch</u> ist: $\frac{\partial x'Ax}{\partial x} = 2Ax$

ii. Varianz, Kovarianz und Korrelationskoeffizient

Varianz:

$$\sigma^2 = V\{Y\} = E\left\{(Y - E(Y))^2\right\}$$
$$= E\{Y^2\} - E\{Y\}^2$$

Kovarianz:

$$\sigma_{YX} = \operatorname{cov}\{Y, X\} = E\left\{(Y - E(Y))(X - E(X))\right\}$$
$$= E\{XY\} - E\{X\}E\{Y\}$$

Korrelationskoeffizient:

$$\rho_{YX} = \frac{\operatorname{cov}\{Y, X\}}{\sqrt{V\{X\} \cdot V\{Y\}}} = \frac{\sigma_{YX}}{\sigma_Y \sigma_X}, \quad -1 \leq \rho_{YX} \leq 1$$

X,Y sind *unkorreliert*, wenn cov{Y,X} = 0

***Rechenregeln*:**

Wenn a, b, c, d Skalare und X, Y Zufallsvariablen sind:

$$V\{aY + b\} = a^2\, V\{Y\}$$

$$V\{aY + bX\} = a^2 V\{Y\} + b^2 V\{X\} + 2ab \operatorname{cov}\{Y, X\}$$

II. Annahmen im linearen Modell

A 1 $E\{\varepsilon_i\} = 0$ i = 1,2,..., N

A 2 $\{x_1, \ldots, x_N\}$ und $\{\varepsilon_1, \ldots, \varepsilon_N\}$ sind unabhängig

A 3 $V\{\varepsilon_i\} = \sigma^2$ i = 1,2,…, N

A 4 $\operatorname{cov}\{\varepsilon_i, \varepsilon_j\} = 0$ i,j = 1,…, N, $i \neq j$

A 5 $\varepsilon \sim N(0, \sigma^2 I_N)$

A 5' $\varepsilon_i \sim NID(0, \sigma^2)$

A 6 $\frac{1}{N}\sum_{i=1}^{N} x_i x_i'$ konvergiert gegen eine positiv definite nichtsinguläre Matrix Σ_{xx}.

A 7 $E\{x_i \varepsilon_i\} = 0$

A 8 x_t und ε_t sind für gegebenes t statistisch unabhängig

A 9 $V\{\varepsilon \mid X\} = \sigma^2 \operatorname{Diag}\{h_i^2\} = \sigma^2 \Psi$

A 10 $E\{\varepsilon \mid X\} = 0$

A 11 $\varepsilon_t \sim IID(0, \sigma^2)$

A 12 ε_t ist über die Zeit unkorreliert, mit Erwartungswert 0.

III. Das Lineare Regressionsmodell

Lösung für β:

$$b = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i \quad \text{wenn } \sum_{i=1}^{N} x_i x_i' \text{ invertierbar}$$

ist bzw.

$b = (X'X)^{-1} X'y$ wenn $X'X$ invertierbar ist

Lösung für b_2 wenn $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$:

$$b_2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Varianz des KQ Schätzers:

$$V\{b \mid X\} = \sigma^2 (X'X)^{-1} = \sigma^2 \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$$

Unverzerrter Schätzer für σ^2:

$$s^2 = \frac{1}{N-K}\sum_{i=1}^{N} e_i^2$$

IV. Maximum Likelihood

Likelihood Funktion im Modell mit einer binären abhängigen Variable:

$$L(\beta) = \prod_{i=1}^{N} P\{y_i = 1 | x_i; \beta\}^{y_i} \; P\{y_i = 0 | x_i; \beta\}^{1-y_i}$$

Log-Likelihood Funktion im Modell mit einer binären abhängigen Variable:

$$\log L(\beta) = \sum_{i=1}^{N} y_i \log F(x_i'\beta) + \sum_{i=1}^{N} (1-y_i)\log(1-F(x_i'\beta))$$

Marginale Effekte im Probit und Logit Modell:

Probit:

$$\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\cdot\beta_k$$

Logit:

$$\frac{\partial \Lambda(x_i'\beta)}{\partial x_{ik}} = \frac{\exp(x_i'\beta)}{\left(1+\exp(x_i'\beta)\right)^2}\cdot\beta_k$$

V. Gütemaße

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

Angepasstes R^2

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^{N} e_i^2 \;/(N-K)}{\sum_{i=1}^{N}(y_i-\bar{y})^2 \;/(N-1)}$$

AIC

$$AIC = \log\frac{1}{N}\sum_{i=1}^{N} e_i^2 + \frac{2K}{N}$$

BIC

$$BIC = \log\frac{1}{N}\sum_{i=1}^{N} e_i^2 + \frac{K}{N}\log N$$

pseudo R^2

$$\text{pseudo}R^2 = 1 - \frac{1}{1+2(\log L_1 - \log L_0)/N}$$

Mc Fadden R^2

$$\text{McFadden } R^2 = 1 - (\log L_1 / \log L_0)$$

VI. Tests

Kritischer Wert bei einem einseitigen Test

$$P\{t_k > t_{N-K;\alpha}\} = \alpha$$

Konfidenzintervall zum Niveau 1 - α

$$b_k - t_{N-K;\frac{\alpha}{2}}\cdot se(b_k) < \beta_k < b_k + t_{N-K;\frac{\alpha}{2}}\cdot se(b_k)$$

Teststatistik für einen F-Test auf gemeinsame Signifikanz

$$F = \frac{(S_0 - S_1)/J}{S_1/(N-K)} \sim F_{J,N-K}$$

$$F = \frac{\left(R_1^2 - R_0^2\right)/J}{\left(1-R_1^2\right)/(N-K)}$$

Teststatistik Goldfeld-Quandt-Test

$$\lambda = \frac{s_A^2}{s_B^2} \sim F_{N_A-K,N_B-K}$$

Teststatistik Durbin-Watson-Test

$$dw = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Teststatistik Wald-Test

$$\xi_W = (R\hat{\theta} - q)'[R\hat{V}R']^{-1}(R\hat{\theta} - q) \sim \chi_J^2$$

Teststatistik Likelihood-Ratio-Test

$$\xi_{LR} = -2\left[\log L(\tilde{\theta}) - \log L(\hat{\theta})\right] \sim \chi_J^2$$