Ökonometrieprüfung WS 2014/2015 - MUSTERLÖSUNG

Fach: Ökonometrie

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

Vorbemerkungen:

Bitte beachten Sie:	**Es wird nur der Lösungsbogen eingesammelt und bewertet.** **Angaben auf dem Aufgabenbogen werden nicht gewertet.**
Anzahl der Aufgaben:	Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen.
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: **[17,5 Punkte]**

1.1 Im Modell $y = X\beta + \varepsilon$ sei y ein [Nx1]-Vektor und X eine [NxK]-Matrix. Beschreiben Sie formal und knapp verbal das Minimierungsproblem zur Bestimmung des KQ-Schätzers b_{KQ}. [2 Punkte]

- Minimierungsproblem: Minimiere $Q(b) = [y - Xb]'[y - Xb] = \sum_{i=1}^{N} (y_i - x_i'b)^2$ über die Wahl von b.
- Gesucht wird ein [Kx1]-Vektor b, sodass die Residuenquadratsumme $Q(b) > 0$ minimiert wird. Dieses b ist der KQ-Schätzer.

1.2 In einer Umfrage sammeln Sie für vier Personen folgende Informationen zu Geschlecht und Stundenlohn: Person 1 ist weiblich und verdient 24 €/h, Person 2 ist männlich und verdient 26 €/h, Person 3 ist männlich und verdient 27 €/h und Person 4 ist weiblich und verdient 20 €/h.

Sie regressieren die Variable $Stundenlohn_i$ (in Euro pro Stunde) auf eine Konstante und die binäre Variable $Weiblich_i$, welche den Wert 1 annimmt, wenn Person i weiblich ist (sonst = 0):

$$Stundenlohn_i = \beta_1 + \beta_2 Weiblich_i + \varepsilon_i$$

1.2.1 Berechnen Sie mithilfe der KQ-Methode die geschätzten Koeffizienten b_1 und b_2. Verwenden Sie $(X'X)^{-1} = \begin{pmatrix} 0,5 & -0,5 \\ -0,5 & 1 \end{pmatrix}$ als die Inverse der Matrix $X'X$. [7 Punkte]

- $y = \begin{pmatrix} 24 \\ 26 \\ 27 \\ 20 \end{pmatrix}$ und $X = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$ bzw. $X' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$
- $b = (X'X)^{-1}X'y = \begin{pmatrix} 0,5 & -0,5 \\ -0,5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 24 \\ 26 \\ 27 \\ 20 \end{pmatrix} = \begin{pmatrix} 0 & 0,5 & 0,5 & 0 \\ 0,5 & -0,5 & -0,5 & 0,5 \end{pmatrix} \begin{pmatrix} 24 \\ 26 \\ 27 \\ 20 \end{pmatrix} = \begin{pmatrix} 26,5 \\ -4,5 \end{pmatrix}$
  $= \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

1.2.2 Interpretieren Sie b_1 inhaltlich. Sollten Sie keine Lösung haben, verwenden Sie $b_1 = 20$. [1,5 Punkte]

- Der durchschnittliche Stundenlohn der männlichen Befragten beträgt 26,5 €/h.

1.2.3 Stellen Sie für diesen Fall die Varianz-Kovarianz-Matrix der Störterme $Var(\varepsilon)$ formal dar und unterstellen Sie, dass weder Autokorrelation noch Heteroskedastie vorliegen.
Hinweis: Keine Berechnung notwendig. [3 Punkte]

- $Var(\varepsilon) = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \sigma_\varepsilon^2 \cdot I$

1.3 Ein Kommilitone hat ebenfalls die Determinanten des Stundenlohns untersucht und dazu dieselben Personen wie Sie befragt. Zusätzlich zum Geschlecht hat er die Variable *Bildung* abgefragt und in das Modell aufgenommen. Sein lineares Regressionsmodell lautet demnach:

$$Stundenlohn_i = \widetilde{\beta}_1 + \widetilde{\beta}_2 Weiblich_i + \widetilde{\beta}_3 Bildung_i + \varepsilon_i$$

1.3.1 Diskutieren Sie allgemein die Konsequenzen der Nichtberücksichtigung der Variablen *Bildung* in Ihrem Modell für Ihren geschätzten Koeffizienten für *Weiblich*. [2 Punkte]

- Der geschätzte Koeffizient für *Weiblich* im Modell ohne *Bildung* ist verzerrt, wenn *Weiblich* und *Bildung* korreliert sind und der Koeffizient für *Bildung* ungleich Null ist.

1.3.2 Wie wird sich der geschätzte Koeffizient für *Weiblich* im Modell Ihres Kommilitonen mit *Bildung* ($\widetilde{b}_2$) von Ihrem geschätzten Koeffizienten b_2 unterscheiden, wenn Sie annehmen, dass *Weiblich* und *Bildung* positiv korreliert sind und der Koeffizient für *Bildung* positiv ist? Erläutern Sie Ihre Antwort. [2 Punkte]

- Der Koeffizient $\widetilde{b}_2$ im Modell mit *Bildung* ist kleiner als b_2 im Modell ohne *Bildung*. Durch die pos. Korrelation und den pos. Koeffizienten für *Bildung* liegt im Modell ohne *Bildung* eine positive Verzerrung vor, der Effekt von *Weiblich* wird somit überschätzt.

Aufgabe 2: **[14,5 Punkte]**

Sie interessieren sich für die Determinanten der Anzahl von Morden in einer Stadt. Ihr Datensatz enthält folgende Informationen für 16 Zeitpunkte:

$morde_t$	Anzahl an Morden pro 10.000 BewohnerInnen in Jahr t
alr_t	Lokale Arbeitslosenrate in % in Jahr t (kodiert 0-100)
$jmann_t$	Bevölkerungsanteil von Männern zwischen 18 und 25 Jahren in % in Jahr t (kodiert 0-100)
$jahr_t$	Jahr t (1990-2005)

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

$$morde_t = \beta_1 + \beta_2 alr_t + \beta_3 jmann_t + \beta_4 jahr_t + \varepsilon_t$$

```
      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  3,    12) =    5.90
       Model |  .000011774     3  3.9246e-06           Prob > F      =  0.0103
    Residual |  7.9762e-06    12  6.6468e-07           R-squared     =  0.5961
-------------+------------------------------           Adj R-squared =  0.4952
       Total |   .00001975    15  1.3167e-06           Root MSE      =  .00082

------------------------------------------------------------------------------
       morde |      Coef.   Std. Err.      t    P>|t|     [95\% Conf. Interval]
-------------+----------------------------------------------------------------
         alr |   .4055569   .1414595     2.87   0.014     .0973432    .7137707
       jmann |   .1003417   .0299902     3.35   0.006     .0349986    .1656848
        jahr |   .0007577   .0003453     2.19   0.049     5.47e-06      .00151
       _cons |  -1.493862   .6976248    -2.14   0.053    -3.013856    .0261318
------------------------------------------------------------------------------
```

2.1 Interpretieren Sie inhaltlich die geschätzten Koeffizienten b_3 und b_4. [2 Punkte]

- b_3: Erhöht sich der Bevölkerungsanteil von Männern zwischen 18 und 25 Jahren um 1 Prozentpunkt, so steigt c.p. im Durchschnitt die Anzahl an Morden pro Jahr um 0,1 pro 10.000 BewohnerInnen.

- b_4: Die Anzahl an Morden steigt jedes Jahr c.p. im Durchschnitt um 0,00076 pro 10.000 BewohnerInnen.

2.2 Erläutern Sie knapp verbal, was unter Autokorrelation zu verstehen ist und nennen Sie zwei Konsequenzen von Autokorrelation für die Eigenschaften des KQ-Schätzers. [3 Punkte]

- Autokorrelation ist eine Situation, in der zeitlich aufeinanderfolgende Störterme korrelieren.
- Autokorrelation führt zur Ineffizienz der KQ-Parameterschätzer, Schätzer jedoch nach wie vor erwartungstreu (Standardfehler sind falsch berechnet).

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

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          7.35               1                     ????
---------------------------------------------------------------------------
```

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

- Hilfsregression: $e_t = \beta_1 + \rho e_{t-1} + \beta_2 alr_t + \beta_3 jmann_t + \beta_4 jahr_t + \nu_t$; alternativ $e_t = \beta_1 + \rho e_{t-1} + \nu_t$.
- Komponenten der Teststatistik: $T - 1$ entspricht der Anzahl an Messzeitpunkten - 1 (hier 15) und R^2 ist das Bestimmtheitsmaß aus der Hilfsregression.
- Hypothesen: $H_0 : \rho = 0$ (keine Autokorrelation 1. Ordnung); $H_1 : \rho \neq 0$ (Autokorrelation 1. Ordnung).
- Kritischer Wert: $\chi^2_{1;5\%} = 3,842$.
- Testergebnis: Da $\chi^2_{empirisch} = 7,35 > 3,842 = \chi^2_{kritisch}$ kann die Nullhypothese auf dem 5% Signifikanzniveau verworfen werden. Es liegt wohl Autokorrelation 1. Ordnung vor.

2.4 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. Begründen Sie kurz, welches Verfahren Sie im vorliegenden Fall bevorzugen. [3 Punkte]

- Idee: 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.
- Cochrane-Orcutt kann die erste Beobachtung nicht nutzen, der Prais-Winsten Schätzer macht dies möglich.
- Im vorliegenden Fall, bei nur 16 Beobachtungen, ist der Prais-Winsten Schätzer aufgrund der zusätzlichen ersten Beobachtung zu bevorzugen.

2.5 In den meisten Fällen ist ρ unbekannt. Können Sie trotzdem die Verfahren aus 2.4 anwenden? Begründen Sie Ihre Antwort. [1,5 Punkte]

- Ja, bei unbekanntem ρ kann dieses in einem ersten Schritt aus einer Regression anhand der vorhergesagten KQ-Residuen konsistent geschätzt werden: $e_t = \rho e_{t-1} + \nu_t$.

Aufgabe 3: **[11 Punkte]**

Sie interessieren sich für die Determinanten von Kindersterblichkeit. Ihnen steht ein Querschnittsdatensatz für 77 Länder mit folgenden Informationen zur Verfügung:

$Kindersterb_i$	Anzahl von Kindern im Land i, die früh sterben, bezogen auf 1000 Lebendgeburten
$Armut_i$	Anteil der Personen im Land i, die unter der Armutsgrenze leben, in % der Gesamtbevölkerung (kodiert 0-100)
$Alphabet_i$	Alphabetisierungsrate im Land i in % der Gesamtbevölkerung (kodiert 0-100)
$Gesundheitsausg_i$	Ausgaben der Regierung im Land i für Gesundheit in % der Gesamtausgaben (kodiert 0-100)

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

$$Kindersterb_i = \beta_1 + \beta_2 Armut_i + \beta_3 Alphabet_i + \beta_4 Gesundheitsausg_i + \varepsilon_i$$

```
      Source |       SS       df       MS              Number of obs =      77
-------------+------------------------------           F(  3,    73) =   82.18
       Model |  153877.565     3  51292.5215           Prob > F      =  0.0000
    Residual |  45565.4225    73   624.18387           R-squared     =  0.7715
-------------+------------------------------           Adj R-squared =  0.7621
       Total |  199442.987    76  2624.24983           Root MSE      =  24.984

---------------------------------------------------------------------------------
    Kindersterb |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
          Armut |   .8718229   .1467527     5.94   0.000     .5793452    1.164301
       Alphabet |  -1.214231    .210099    -5.78   0.000    -1.632958   -.7955044
Gesundheitsausg |  -.2470155   .1744259    -1.42   0.161    -.5946458    .1006149
          _cons |   145.3079   18.89028     7.69   0.000     107.6596    182.9562
---------------------------------------------------------------------------------
```

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

3.1 Interpretieren Sie inhaltlich und statistisch den Zusammenhang zwischen Alphabetisierungsrate und Kindersterblichkeit. [2 Punkte]

- Inhaltlich: $b_3 = -1,214$: Wenn die Alphabetisierungsrate um 1 Prozentpunkt steigt, sinkt die Kindersterblichkeit c.p. im Mittel um 1,214 pro 1000 Lebendgeburten.
- Statistisch: $p = 0,000 < 0,01$: Der Koeffizient ist statistisch signifikant auf dem 1%-Niveau.

3.2 Testen Sie mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$, ob das Maß der Kindersterblichkeit um mehr als 0,2 abnimmt, wenn die Variable $Gesundheitsausg_i$ um einen Prozentpunkt steigt. Geben Sie Null- und Alternativhypothese, Teststatistik, kritischen Wert und Testentscheidung an. [4 Punkte]

- Hypothesen: $H_0 : \beta_4 \geq -0,2; H_1 : \beta_4 < -0,2$
- Teststatistik: $t_{empirisch} = \frac{-0,247-(-0,2)}{0,174} = -0,27$
- Kritischer Wert: $t_{kritisch} = t_{\alpha;N-K} = t_{0,05;73}$ nicht tabelliert. Nächster Wert bei geringeren Freiheitsgraden $t_{kritisch} = t_{0,05;70} = -1,667$
- Testentscheidung: Auf einem Signifikanzniveau von 5% kann die Nullhypothese nicht verworfen werden, da $t_{empirisch} = -0,27 > -1,667 = t_{kritisch}$. Der Effekt ist somit wohl nicht kleiner als -0,2.

3.3 Sie glauben, dass der Effekt der Gesundheitsausgaben auf die Kindersterblichkeit systematisch vom Niveau der Armutsrate abhängt. Sie nehmen daher einen Interaktionsterm in Ihr Modell auf und schätzen dieses erneut. Ihr Regressionsmodell und die Schätzergebnisse lauten wie folgt:

$$Kindersterb_i = \beta_1 + \beta_2 Armut_i + \beta_3 Alphabet_i + \beta_4 Gesundheitsausg_i + \beta_5 GesundheitsausgArmut_i + \varepsilon_i$$

wobei $GesundheitsausgArmut_i = Gesundheitsausg_i \cdot Armut_i$ gilt.

```
      Source |       SS       df       MS              Number of obs =      77
-------------+------------------------------           F(  4,    72) =   60.79
       Model |   153878.94     4   38469.735           Prob > F      =  0.0000
    Residual |  45564.0472    72  632.833989           R-squared     =  0.7715
-------------+------------------------------           Adj R-squared =  0.7589
       Total |  199442.987    76  2624.24983           Root MSE      =  25.156

---------------------------------------------------------------------------------------
          Kindersterb |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------+----------------------------------------------------------------
                Armut |   .8888328   .3936623     2.26   0.027     .1040814    1.673584
             Alphabet |  -1.214039   .2115899    -5.74   0.000    -1.635836   -.7922422
       Gesundheitsausg |  -.2365791   .2845403    -0.83   0.408    -.8037998    .3306416
GesundheitsausgArmut |  -.0003093   .0066347    -0.05   0.963    -.0135353    .0129167
                _cons |   144.6889    23.1971     6.24   0.000     98.44633    190.9315
---------------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die vierte Nachkommastelle.

3.3.1 Leiten Sie den marginalen Effekt der Variablen *Gesundheitsausg* allgemein her. [1 Punkt]

- Marginaler Effekt von *Gesundheitsausg*:

$$\frac{\partial \widehat{Kindersterb_i}}{\partial Gesundheitsausg_i} = b_4 + b_5 \cdot Armut_i$$

3.3.2 Berechnen und interpretieren Sie den marginalen Effekt für Länder mit einer Armutsrate von 10% und Länder mit einer Armutsrate von 80%. [4 Punkte]

- Marginaler Effekt bei $Armut_i = 10$:

$$\left.\frac{\partial \widehat{Kindersterb_i}}{\partial Gesundheitsausg_i}\right|_{Armut_i=10} = b_4 + b_5 \cdot 10 = -0,2366 - 0,0003 \cdot 10 = -0,2396$$

- Interpretation: Ein Anstieg der Variablen *Gesundheitsausg* um einen Prozentpunkt reduziert c.p. im Mittel die Kindersterblichkeit um 0,2396 pro 1000 Lebendgeburten in Ländern mit einer Armutsrate von 10%.
- Marginaler Effekt bei $Armut_i = 80$:

$$\left.\frac{\partial \widehat{Kindersterb_i}}{\partial Gesundheitsausg_i}\right|_{Armut_i=80} = b_4 + b_5 \cdot 80 = -0,2366 - 0,0003 \cdot 80 = -0,2606$$

- Interpretation: Ein Anstieg der Variablen *Gesundheitsausg* um einen Prozentpunkt reduziert c.p. im Mittel die Kindersterblichkeit um 0,2606 pro 1000 Lebendgeburten in Ländern mit einer Armutsrate von 80%.

Aufgabe 4 **[17 Punkte]**

Die Wahrscheinlichkeit, in einer Partei politisch aktiv zu sein, wurde anhand eines binären Modells analysiert.

$$P(polaktiv_i = 1|\mathbf{x}_i) = F(\beta_1 + \beta_2 ln\_einkommen_i + \beta_3 alter_i + \beta_4 mann_i + \beta_5 bildung_i)$$

Die folgende Tabelle beschreibt die verwendeten Variablen:

Variable	Mittelwert	Std.abw.	Min	Max	Beschreibung
polaktiv	.373	.484	0	1	Politische Aktivität der Person (wenn politisch aktiv=1, sonst=0)
ln_einkommen	9.823	1.083	3.829	13.708	Log. Einkommen der Person in Euro
einkommen	29615.52	42442.69	46	897756	Einkommen der Person in Euro
alter	43.712	12.131	18	87	Alter der Person in Jahren
mann	.511	.5	0	1	Geschlecht der Person (wenn männlich=1, sonst=0)
bildung	12.21	2.731	8.7	18	Jahre der Ausbildung der Person

4.1 Nennen Sie ein Problem, das bei der KQ-Schätzung von Modellen mit binären abhängigen Variablen besteht. [1 Punkt]

- Der KQ-Schätzer weist fehlerhafte Standardfehler aus, da er die Heteroskedastie nicht berücksichtigt.
- alternativ:
- Beim linearen Modell können die prognostizierten Wahrscheinlichkeiten außerhalb des Intervalls $[0,1]$ liegen.

4.2 Nehmen Sie nun an, dass die Wahrscheinlichkeitsfunktion einer logistischen Verteilung folgt. Die Ergebnisse einer Logit-Schätzung des Modells sind in der folgenden Tabelle enthalten:

```
Logistic regression                               Number of obs   =       3216
                                                  LR chi2(4)      =     301.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -1973.3689                       Pseudo R2       =     0.0709
```

polaktiv	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_einkommen	.132518	.0401888	3.30	0.001	.0537493	.2112867
alter	.0226214	.0032779	6.90	0.000	.0161969	.0290459
mann	.3048543	.0809412	3.77	0.000	.1462124	.4634961
bildung	.1828778	.0145248	12.59	0.000	.1544098	.2113459
_cons	-5.239909	.3995143	-13.12	0.000	-6.022943	-4.456876

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

4.2.1 Leiten Sie den marginalen Effekt einer Erhöhung des Einkommens für die Person i allgemein her. [3 Punkte]

- Bestimmung des marginalen Effekts im Logit-Modell für $einkommen_i$:

$$\frac{\partial P(polaktiv_i = 1|\mathbf{x}_i)}{\partial einkommen_i} = \frac{\partial F(\mathbf{x}_i'\beta)}{\partial einkommen_i} = \frac{\partial F(\mathbf{x}_i'\beta)}{\partial \mathbf{x}_i'\beta}\frac{\partial \mathbf{x}_i'\beta}{\partial einkommen_i} = \frac{\exp(\mathbf{x}_i'\beta)}{(1+\exp(\mathbf{x}_i'\beta))^2}\frac{\beta_2}{einkommen_i}$$

4.2.2 Berechnen Sie für 18- und 60-jährige Männer die Wahrscheinlichkeiten, politisch aktiv zu sein. Benennen und interpretieren sie deren Differenz. Nehmen Sie für die übrigen Variablen Stichprobenmittelwerte an. [7 Punkte]

- Berechnung der Wahrscheinlichkeit für 18-jährige Männer:

$$P(polaktiv = 1|alter = 18, mann = 1, \bar{\mathbf{x}}) = F(-5,24 + 0,133 \cdot 9,823 + 0,023 \cdot 18 + 0,305 \cdot 1 + 0,183 \cdot 12,21)$$

$$= F(-0,98) = \frac{1}{1+\exp(-(-0,98))} = 0,273$$

- Berechnung der Wahrscheinlichkeit für 60-jährige Männer:

$$P(polaktiv = 1|alter = 60, mann = 1, \bar{\mathbf{x}}) = F(-5,24 + 0,133 \cdot 9,823 + 0,023 \cdot 60 + 0,305 \cdot 1 + 0,183 \cdot 12,21)$$

$$= F(-0,014) = \frac{1}{1+\exp(-(-0,014))} = 0,496$$

- Differenz: Die Wahrscheinlichkeit, politisch aktiv zu sein, unterscheidet sich zwischen 18-jährigen und 60-jährigen Männern am Stichprobenmittelwert um 49,6-27,3=22,3 Prozentpunkte.

4.2.3 In ihrem Datensatz wird die Lebenszufriedenheit mit der Variablen *zufriedenheit* gemessen. Prüfen Sie, ob die Lebenszufriedenheit einen signifikanten Effekt auf die Wahrscheinlichkeit, politisch aktiv zu sein, hat. Führen Sie dazu einen Likelihood-Ratio-Test am 5% Signifikanzniveau durch. Geben Sie zuerst das restringierte Modell und das unrestringierte Modell an. Benennen Sie dann Hypothesen und kritischen Wert. Treffen Sie eine Testentscheidung.
Hinweis: Der Log-Likelihood-Wert des Modells inkl. $zufriedenheit_i$ beträgt -1963,774. [6 Punkte]

- restringiertes Modell: $P(polaktiv_i = 1|\mathbf{x}_i) = F(\beta_1 + \beta_2 ln\_einkommen_i + \beta_3 alter_i + \beta_4 mann_i + \beta_5 bildung_i)$
- unrestringiertes Modell: $P(polaktiv_i = 1|\mathbf{x}_i, zufriedenheit_i) = F(\beta_1 + \beta_2 ln\_einkommen_i + \beta_3 alter_i + \beta_4 mann_i + \beta_5 bildung_i + \beta_6 zufriedenheit_i)$
- Nullhypothese: $H_0 : \beta_6 = 0$
- Alternativhypothese: $H_1 : \beta_6 \neq 0$
- kritischer Wert: $\chi^2_{1,0,05} = 3,842$
- Testentscheidung: $LR_{empirisch} = -2(-1973,369 - (-1963,774)) = -2(-9,595) = 19,19 > 3,842$. Die H0 kann auf dem 5% Signifikanzniveau verworfen werden.

Aufgabe 5 - MC Fragen **[30 Punkte]**

Bitte geben Sie die zutreffende Antwort **auf Ihrem Multiple-Choice-Lösungsblatt** an. **Angaben auf dem Aufgabenblatt werden nicht gewertet.** 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 auf dem Lösungsblatt gilt diese als nicht beantwortet.

1.	Wenn a ein $[1xN]-$Zeilen- und b ein $[Nx1]-$Spaltenvektor ist, dann ist ab
a	das innere Produkt der Vektoren a und b.**X**
b	das äußere Produkt der Vektoren a und b.
c	das orthogonale Produkt der Vektoren a und b.
d	nicht definiert.

2.	Berechnen Sie die Determinante der Matrix $A = \begin{pmatrix} 4 & 2 \\ 3 & 7 \end{pmatrix}$
a	-28.
b	-4.
c	16.
d	22. **X**

3.	Wenn zwei Zufallsvariablen in einem nicht-linearen Zusammenhang stehen, dann
a	ist die Kovarianz stets 0.
b	kann dieser Zusammenhang mit der Kovarianz abgebildet werden.
c	kann dieser Zusammenhang mit dem Korrelationskoeffizienten abgebildet werden.
d	kann die Kovarianz positiv sein.**X**

4.	Ein Typ I Fehler bezeichnet die Wahrscheinlichkeit,
a	eine zutreffende Nullhypothese abzulehnen.**X**
b	eine nicht zutreffende Nullhypothese abzulehnen.
c	eine zutreffende Nullhypothese nicht abzulehnen.
d	eine nicht zutreffende Nullhypothese nicht abzulehnen.

5.	Wenn für zwei Zufallsvariablen X und Y gilt, dass $E(Y\|X) = E(Y) = 0$, dann
a	sind die Variablen X und Y statistisch unabhängig.
b	ist die Kovarianz von X und Y gleich 0. **X**
c	ist $E(X\|Y) = E(X) = 0$.
d	ist $Var(Y\|X) = Var(Y)$.

6.	Der KQ-Schätzer minimiert
a	die Wahrscheinlichkeit, einen Typ I Fehler zu begehen.
b	die Summe der quadrierten Residuen.**X**
c	die Summe der absoluten Residuen.
d	die quadrierte Summe der Residuen.

7.	Der KQ-Schätzer lässt sich nur berechnen, wenn
a	$X'X$ ein Skalar ist.
b	$X'y$ symmetrisch ist.
c	$X'y$ invertierbar ist.
d	$X'X$ vollen Spaltenrang besitzt. **X**

8.	Im einfachen linearen Regressionsmodell ist das Bestimmtheitsmaß R^2
a	größer als die Anzahl der Freiheitsgrade.
b	genauso groß wie das korrigierte Bestimmtheitsmaß.
c	das Verhältnis von unerklärter Variation zu Gesamtvariation der abhängigen Variablen.
d	zwischen 0 und 1.**X**

9.	Bei Vorliegen von perfekter Multikollinearität
a	hat die Kreuzproduktmatrix $X'X$ vollen Rang.
b	nimmt das Bestimmtheitsmaß R^2 den Wert 1 an.
c	werden die Parameter mittels KQ unpräzise geschätzt.
d	ist mindestends eine erklärende Variable eine Linearkombination der anderen erklärenden Variablen.**X**

10.	Die Annahmen $E[\varepsilon_i] = 0$ und $Var(\varepsilon) = \sigma^2$ für alle $i = 1, ..., n$ implizieren, dass
a	die Störterme autokorreliert sein können. **X**
b	die Störterme heteroskedastisch sein können.
c	Heteroskedastie und Autokorrelation ausgeschlossen werden können.
d	die Störterme heteroskedastisch und autokorreliert sein können.

11.	Der marginale Effekt in einem Logit-Modell
a	entspricht dem Wert des Koeffizienten.
b	wird immer als Elastizität interpretiert.
c	hat nicht das gleiche Vorzeichen wie der geschätzte Koeffizient.
d	hängt von den Aussprägungen der erklärenden Variablen ab.**X**

12.	Das Maximum-Likelihood-Verfahren
a	schätzt für ein lineares Modell dieselben Koeffizientenwerte wie das KQ-Verfahren.**X**
b	hat die BLUE-Eigenschaft bei Gültigkeit der Gauß-Markov-Annahmen.
c	bestimmt die Parameter durch Maximierung des R^2.
d	schätzt die Varianz des Störterms unverzerrt.

13.	Welche Aussage ist richtig?
a	Der Lagrange-Multiplier Test wird nach einer restringierten Schätzung angewendet.**X**
b	Der Likelihood-Ratio-Test hat N Freiheitsgrade.
c	Die Teststatistik des t-Tests folgt asymptotisch der χ^2-Verteilung.
d	Zur Durchführung eines Wald-Tests ist die zweimalige Schätzung eines Modells (mit und ohne Restriktion) notwendig.

14.	Bei korrekter Spezifizierung der Likelihoodfunktion ist der ML-Schätzer
a	unverzerrt.
b	konsistent. **X**
c	asymptotisch χ^2-verteilt.
d	effizient.

15.	Bei unabhängigen Beobachtungen entspricht die logarithmierte Likelihoodfunktion
a	dem Produkt der logarithmierten Dichten der einzelnen Beobachtungen der abhängigen Variable.
b	der Summe der logarithmierten Dichten der einzelnen Beobachtungen der abhängigen Variable. **X**
c	dem Logarithmus der Dichten der einzelnen Beobachtungen der abhängigen Variable.
d	der Summe der quadrierten logarithmierten Dichten der einzelnen Beobachtungen der abhängigen Variable.

16.	Für das binäre Modell $raucher_i = \beta_1 + \beta_2 alter_i + \beta_3 alter_i^2 + \varepsilon_i$ ergibt eine KQ-Schätzung b_2 = - 6,8 und b_3=0,125. In welchem Alter wird die Wahrscheinlichkeit, ein Raucher zu sein, minimiert?
a	17
b	27 **X**
c	37
d	47

17.	Ein Strukturbruchtest überprüft,
a	ob ein Problem der Modellspezifikation vorliegt **X**.
b	ob bei Zeitreihendaten Autokorrelation vorliegt.
c	ob die Kreuzproduktmatrix $\mathbf{X'X}$ invertierbar ist.
d	ob zusätzliche erklärende Variablen die Varianz des Störterms negativ werden lässt.

18.	Für ein Modell mit Konstante, 4 Regressoren und 105 Beobachtungen schätzen Sie eine Fehlerquadratsumme von 42350. Welchen Wert hat die geschätzte Varianz des Störterms?
a	100,8
b	423,5 **X**
c	10587,5
d	21302,2

19.	Sie schätzen folgendes Modell: $y_i = \beta_1^* + \beta_2^* x_i + \varepsilon_i^*$, obwohl das wahre Modell $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ lautet mit $\beta_2 > 0$ und $\beta_3 < 0$. Sei $cov(x,z) < 0$. Dann ist
a	$E[b_2^*] = \beta_2$.
b	$E[b_2^*] > \beta_2$. **X**
c	$E[b_2^*] < \beta_2$.
d	$E[b_2^*] = \beta_2 + \beta_3$.

20.	Im linearen Modell $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$ mit $\beta_2 < 0$ und $\beta_3 > 0$ ist y_i eine
a	konkave Funktion von x_i.
b	konvexe Funktion von x_i. **X**
c	lineare Funktion von x_i.
d	konstante Funktion von x_i.

21.	Sie schätzen das Modell $y_i = \beta_1 + \beta_2 ln(x_i) + \varepsilon_i$ mit KQ und erhalten $b_2 = 3$. Ceteris paribus und im Mittel gilt: wenn x um...
a	eine Einheit steigt, so steigt y um 3%.
b	eine Einheit steigt, so steigt y um 3 Einheiten.
c	1% steigt, so steigt y um 3%.
d	1% steigt, so steigt y um 0,03 Einheiten. **X**

22.	Wie lautet der marginale Effekt des Alters im folgenden linearen Wahrscheinlichkeitsmodell: $raucher_i = \beta_1 + \beta_2 alter_i + \beta_3 alter_i^2 + \beta_4 mann_i + \beta_5 abi_i + \beta_6 abi_i \cdot alter_i + \beta_7 mann_i \cdot alter_i + \varepsilon_i$
a	$\beta_2 + 2\beta_3 alter_i$
b	$\beta_2 + 2\beta_3 alter_i + \beta_6 abi_i + \beta_7 mann_i$ **X**
c	$\beta_2 + 2\beta_3 alter_i + \beta_6 + \beta_7$
d	$\beta_2 + 2\beta_3 alter_i + \beta_4 + \beta_5 + \beta_6 + \beta_7$

23.	Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Tests
a	prüfen jeweils, ob die Bedingungen erster Ordnung der Likelihoodfunktion erfüllt sind.
b	werden jeweils nach einer unrestringierten Schätzung angewendet.
c	haben asymptotisch die gleiche Verteilung. **X**
d	können bei mit KQ-geschätzten linearen Modellen angewendet werden.

24.	Das Bestimmtheitsmaß im multiplen Regressionsmodell und endlicher Stichprobe
a	ist immer kleiner als das korrigierte Bestimmtheitsmaß R^2.
b	ist immer größer als das korrigierte Bestimmtheitsmaß R^2.**X**
c	entspricht dem Verhältnis der Störtermvariation zur Gesamtvariation.
d	entspricht dem Verhältnis erklärter Variation zur Störtermvariation.

25.	Das Informationskriterium BIC
a	wird unter Verwendung des R^2 berechnet.
b	bewertet die Effizienz einer Schätzung.
c	kann zum Vergleich nicht-genesteter Modelle verwendet werden. **X**
d	berücksichtigt bei der Berechnung ausschließlich die Anzahl der Regressoren.

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

27.	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.

28.	Der FGLS-Schätzer beim Prais-Winsten-Verfahren
a	ist konsistent, aber nicht unverzerrt. **X**
b	ist nicht effizienter als der Cochrane-Orcutt Schätzer.
c	gehört zur Klasse der BLUE Schätzer (best-linear-unbiased estimator).
d	benutzt die erste Beobachtung nicht.

29.	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**

30.	Bei Heteroskedastie im linearen Modell
a	variiert die Varianz der Störterme über die Beobachtungen. **X**
b	ist der KQ-Schätzer BLUE.
c	sind die Parameter des KQ-Schätzers verzerrt geschätzt.
d	sind die Elemente der Hauptdiagonalen der Varianz-Kovarianz-Matrix identisch.