Masterprüfung SS 2014 - MUSTERLÖSUNG

Fach: Ökonometrie

Prüfer: 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: **[17 Punkte]**

Sie interessieren sich für die Determinanten von Urlaubsausgaben. Nach einer Befragung von 4500 Personen steht Ihnen ein Datensatz mit folgenden Informationen zur Verfügung:

$Urlaub_i$	Jährliche Urlaubsausgaben von Person i in €
$Einkommen_i$	Monatliches Einkommen von Person i in 1000 €
$Alter_i$	Alter von Person i in Jahren
$Kinder_i$	Anzahl Kinder von Person i

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

$$log(Urlaub_i) = \beta_1 + \beta_2 Einkommen_i + \beta_3 Alter_i + \beta_4 Kinder_i + \varepsilon_i$$

```
      Source |       SS       df       MS              Number of obs =    4500
-------------+------------------------------           F(  3,  4496) = 1071.71
       Model |  ??????????     3  ??????????           Prob > F      =  0.0000
    Residual |  ??????????  4496  ??????????           R-squared     =  ??????
-------------+------------------------------           Adj R-squared =  ??????
       Total |  3337.54158  4499   .74184076           Root MSE      =  ??????

------------------------------------------------------------------------------
  log_Urlaub |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   Einkommen |   .0087434   9.76e-04     8.95   0.000      .0000682    .0001065
       Alter |   .0388558   .0015017    25.88   0.000      .0359118    .0417998
      Kinder |   .0909318   .0137208     6.63   0.000      .1178314    .0640323
       _cons |   5.702747   .0348647   163.57   0.000      5.634395    5.771099
------------------------------------------------------------------------------
```

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

1.1 Berechnen und interpretieren Sie, welchen Effekt eine Erhöhung des monatlichen Einkommens um 500€ auf die jährlichen Urlaubsausgaben hat. [2 Punkte]

- Berechnung der Änderung: $\Delta log(\widehat{Urlaub}) = b_2 \cdot 0{,}5 = 0{,}004$. [1P]
- Interpretation: Erhöht sich das monatliche Einkommen um 500€ , so führt dies c.p. im Mittel zu einer Erhöhung der monatlichen Urlaubsausgaben von 0,4%. [1P]

1.2 Testen Sie mit einer Irrtumswahrscheinlichkein von $\alpha = 0,05$, ob der Koeffizient der Variable $Kinder_i$ kleiner als 0,1 ist. Geben Sie Null- und Alternativhypothese, Teststatistik, kritischen Wert und Testentscheidung an. [4 Punkte]

- Hypothesen: $H_0 : \beta_4 \geq 0,1; H_1 : \beta_4 < 0,1$ [1P]
- Teststatistik: $t^{empirisch} = \frac{0{,}091-0{,}1}{0{,}014} = -0{,}643$ [1P]
- Kritischer Wert: $t^{kritisch} = t_{\alpha;N-K} = t_{0,05;4496} = -1,645$ [1P]
- Testentscheidung: Auf einem Signifikanzniveau von 5% kann die Nullhypothese nicht verworfen werden, da $t^{empirisch} = -0{,}643 > -1{,}645 = t^{kritisch}$; der Effekt ist somit wohl nicht kleiner als 0,1. [1P]

1.3 Stata berechnet das adjustiertes Bestimmtheitsmaß mit $\overline{R^2} = 0,417$ und weist $SST = \sum_{i=1}^{4500}(y_i - \bar{y}) = 3337,542$ aus. Wie hoch ist die Summe der quadrierten Residuen über alle 4500 Beobachtungen ($\sum_{i=1}^{4500} e_i^2$)? Zeigen Sie Ihren Rechenweg. [3 Punkte]

- $\overline{R^2} = 1 - \frac{SSR/(n-k)}{SST/(n-1)}$
- Mit $\overline{R^2} = 0,417$, $SST = 3337,542$, $n = 4500$ und $k = 4$ [2P]
- $SSR = \frac{(-\overline{R^2}+1)\cdot SST\cdot(n-k)}{n-1} = \frac{(-0,417+1)\cdot 3337,542\cdot(4500-4)}{4500-1} = 1944,49$ [1P]
- Die Summe der quadrierten Residuen über alle Beobachtungen beträgt 1944,49.

1.4 Testen Sie, ob Ihr Modell fehlspezifiziert ist. Verwenden Sie dazu folgendes Regressionsmodell:

$$log(Urlaub_i) = \beta_1 + \beta_2 Einkommen_i + \beta_3 Alter_i + \beta_4 Kinder_i + \alpha_2[log(\widehat{Urlaub_i})]^2 + \alpha_3[log(\widehat{Urlaub_i})]^3 + \nu_i,$$

wobei $log(\widehat{Urlaub_i})$ den vorhergesagten Wert von $log(Urlaub_i)$ basierend auf den obigen Schätzergebnissen darstellt. Sie erhalten $\hat{\alpha}_2 = -1,191$ und $\hat{\alpha}_3 = -3,874$.
Für $\boldsymbol{\alpha} = (\alpha_2, \alpha_3)'$ lautet die geschätzte Varianz-Kovarianz-Matrix $\widehat{\mathbf{V}(\hat{\alpha})}$ bzw. ihre Inverse $\widehat{\mathbf{V}(\hat{\alpha})}^{-1}$:

$$\widehat{\mathbf{V}(\hat{\alpha})} = \begin{pmatrix} 0,047 & 0,371 \\ 0,371 & 0,935 \end{pmatrix}, \quad \widehat{\mathbf{V}(\hat{\alpha})}^{-1} = \begin{pmatrix} -9,979 & 3,960 \\ 3,960 & -0,502 \end{pmatrix}$$

Führen Sie einen Wald-Test auf dem 10%-Signifikanzniveau durch und geben Sie Hypothesen, Teststatistik, kritischen Wert und Testentscheidung an. [8 Punkte]

- Hypothesen: $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, H_1: nicht H_0 [1P]
- Teststatistik: $\xi_{Wald} = (\mathbf{R}\theta)'(\mathbf{RVR'})^{-1}(\mathbf{R}\theta)$, wobei $\mathbf{V}$ die geschätzte Varianz-Kovarianz-Matrix bezeichnet.
- Berechnung: Mit $\mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \theta = \boldsymbol{\alpha}$ und $(\mathbf{RVR'})^{-1} = \widehat{\mathbf{V}(\hat{\alpha})}^{-1}$ folgt:

  $$\xi = (-1,191 - 3,874)\begin{pmatrix} -9,979 & 3,960 \\ 3,960 & -0,502 \end{pmatrix}\begin{pmatrix} -1,191 \\ -3,874 \end{pmatrix} = (-3,456 - 2,772)\begin{pmatrix} -1,191 \\ -3,874 \end{pmatrix} = 14,853 \text{ [4P]}$$
- Kritischer Wert: $\chi^2_{2;0,1} = 4,61$ [1P]
- Testentscheidung: H_0 kann auf einem Signifikanzniveau von 10% verworfen werden, da $14,853 > 4,61$. Das Modell ist wohl fehlspezifiziert. [2P]

Aufgabe 2: **[13 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen den Mieten für gewerbliche Immobilien in verschiedenen Wohngebieten einer Stadt und der Entfernung zum Stadtzentrum. Ihr Datensatz enthält folgende Variablen:

$Miete_i$	Miete einer Immobilie i in € pro m^2
$Entfernung_i$	Entfernung zum Stadtzentrum einer Immobilie i in km

Die Regression der Variable $Miete_i$ auf die Variable $Entfernung_i$ und eine Konstante ergibt folgenden Output:

```
      Source |       SS       df       MS              Number of obs =    1200
-------------+------------------------------           F(  1,  1198) = 4178.26
       Model |  2100.10324     1  2100.10324           Prob > F      =  0.0000
    Residual |  602.146286  1198  .502626282           R-squared     =  0.7772
-------------+------------------------------           Adj R-squared =  0.7770
       Total |  2702.24952  1199  2.25375273           Root MSE      =  .70896

------------------------------------------------------------------------------
       Miete |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  Entfernung |  -1.073535    .??????   ??????   ?????    -1.106119   -1.040951
       _cons |   17.39303    .048161   361.14   0.000     17.29854    17.48752
------------------------------------------------------------------------------
```

2.1 Interpretieren Sie inhaltlich und statistisch den Zusammenhang zwischen Mietpreis und Entfernung zum Stadtzentrum. [2 Punkte]

- Erhöht sich die Entfernung zum Stadtzentrum einer Wohnung um 1km, so sinkt der Mietpreis c.p. im Mittel um 1,07 € pro m^2. [1P]
- Da das 95%-Konfidenzintervall den Wert 0 nicht enthält, ist der Koeffizient statistisch signifikant von Null verschieden auf dem 5%-Niveau. [1P]

2.2 Erläutern Sie den Begriff der Heteroskedastie verbal. [1 Punkt]

- Unter Heteroskedastie versteht man eine Situation, in der die Varianz der Störterme im linearen Regressionsmodell nicht konstant ist, sondern über die Beobachtungen variiert. [1P]

2.3 Erläutern Sie verbal, warum im vorliegenden Fall Heteroskedastie vorliegen könnte. [2 Punkte]

- Es ist zu vermuten, dass die Varianz des Störterms mit steigender Entfernung zum Stadtzentrum zunimmt. Wohnungen direkt im Stadtzentrum zeigen eine sehr geringe Variation in den Mietpreisen, in Randbezirken oder auf dem Dorf ist von einer hohen Variation in den Mietpreisen auszugehen. [2P] (andere Antwort möglich)

2.4 Der folgende Stata-Output enthält die Ergebnisse eines White-Tests auf Heteroskedastie. Führen Sie den Test auf einem Signifikanzniveau von $\alpha = 0,01$ durch. Geben Sie Hypothesen, Teststatistik (inkl. Hilfsregression), kritischen Wert und Testentscheidung an. [5 Punkte]

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(?)      =    236.28
         Prob > chi2  =    ??????
```

- Hypothesen: $H_0 : V(\varepsilon_i) = \sigma^2$ für alle i (Homoskedastie); H_1 : nicht H_0 [1P]
- Hilfsregression: $e_i^2 = \alpha_1 + \alpha_2 Entfernung + \alpha_3 Entfernung^2 + \nu_i$ [1P]

- Teststatistik: $\chi^2_{emp} = N \cdot R^2$, wobei N die Anzahl der Beobachtungen bezeichnet; R^2 ist das Bestimmtheitsmaß aus der Hilfsregression. [1P]
- Kritischer Wert: $\chi^2_J = \chi^2_2 = 9,21$ [1P]
- Da $\chi^2_{empirisch} = 236,28 > 9,21 = \chi^2_{kritisch}$ wird die Nullhypothese auf dem 1%-Niveau verworfen. Der Test weist auf Heteroskedastie in dem vorliegenden Modell hin. [1P]

2.5 Es sei $Var(\varepsilon_i) = \sigma^2 \cdot Entfernung_i^4$. Zeigen Sie formal die GLS-Transformation, die zu konstanter Störtermvarianz führt und leiten Sie die Varianz des Störterms im transformierten Modell her. [3 Punkte]

- GLS-Transformation: $\frac{Miete_i}{Entfernung_i^2} = \frac{\beta_1}{Entfernung_i^2} + \frac{\beta_2 Entfernung_i}{Entfernung_i^2} + \frac{\varepsilon_i}{Entfernung_i^2}$
- $Var(\frac{\varepsilon_i}{Entfernung_i^2}) = \frac{1}{Entfernung_i^4} \cdot Var(\varepsilon_i) = \frac{1}{Entfernung_i^4} \cdot \sigma^2 \cdot Entferung_i^4 = \sigma^2$

Aufgabe 3 [15 Punkte]

3.1 Erläutern Sie knapp verbal, was unter Autokorrelation zu verstehen ist und welche Konsequenzen Autokorrelation für die Eigenschaften des KQ-Schätzers hat. [3 Punkte]

- Autokorrelation ist eine Situation, in der zeitlich aufeinanderfolgende Störterme korrelieren. [1P]
- Autokorrelation führt zur Ineffizienz der KQ-Parameterschätzer, Schätzer jedoch nach wie vor erwartungstreu. (Standardfehler sind falsch berechnet.) [2P]

3.2 Die Anzahl der Verkehrsunfälle (Variable *unfall*) wird mit tagesgenauen Zeitreihendaten analysiert. Erklärende Variablen sind Temperatur (Variable *temp*, in Grad Celsius gemessen) und eine Indikatorvariable für das Wochenende (Variable *wochenende*= 1, falls Wochentag Samstag oder Sonntag;= 0 sonst). Die Temperatur wird zusätzlich in quadrierter Form in das Modell aufgenommen (Variable *tempsq*). Das Modell lautet:

$$\log(unfall_t) = \beta_1 + \beta_2 temp_t + \beta_3 tempsq_t + \beta_4 wochenende_t + \varepsilon_t$$

Interpretieren Sie die Effekte der Temperatur und des Wochenendes im folgenden Stata-Output inhaltlich. *Hinweis*: Interpretieren Sie die Ergebnisse für eine bestehende Temperatur von zehn Grad Celsius. [4 Punkte]

```
      Source |       SS       df       MS              Number of obs =     365
-------------+------------------------------           F(  3,   361) =   15.63
       Model |  128.921938     3  42.9739794           Prob > F      =  0.0000
    Residual |  992.318274   361  2.74880408           R-squared     =  0.1150
-------------+------------------------------           Adj R-squared =  0.1076
       Total |  1121.24021   364  3.08033025           Root MSE      =   1.658

------------------------------------------------------------------------------
  log_unfall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        temp |  -.1006878   .0166068    -6.06   0.000     -.133346   -.0680295
      tempsq |   .0044969   .0007287     6.17   0.000     .0030639      .00593
  wochenende |  -.5620591   .1924058    -2.92   0.004    -.9404361   -.1836821
       _cons |   7.190896   .1288277    55.82   0.000     6.937549    7.444243
------------------------------------------------------------------------------
```

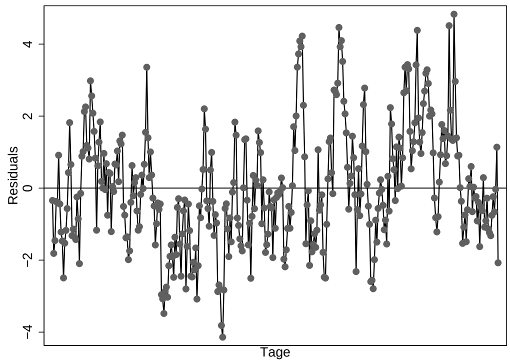
Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- Ausgehend von einer Temperatur von zehn Grad ist ein Anstieg der Temperatur um 1 Grad c. p. im Mittel mit einem Rückgang der Unfälle um 2% verbunden. [2P]

$$\frac{\partial \log(unfall_t)}{\partial temp_t} = \beta_2 + 2\beta_3 \cdot 10 = -0.101 + 2 \cdot 0.004 \cdot 10 = -0.021$$

- Am Wochenende gibt es c. p. im Mittel $(\exp(-0.562) - 1)100\% = 43\%$ weniger Unfälle. [2P]

3.3 Die Abbildung zeigt die Residuen der Regression aus Teilaufgabe 3.2. Interpretieren Sie das Muster der Residuen. [2 Punkte]



- Zeitlich aufeinanderfolgende Residuen e_t und e_{t-1} nehmen häufig gemeinsam negative oder positive Werte an(, sodass $Cov(\varepsilon_t, \varepsilon_{t-1}) = 0$ verletzt sein könnte). [1P]
- Der Plot deutet auf positive Autokorrelation hin. [1P]

3.4 Führen Sie unter Verwendung des folgenden Stata-Outputs einen Breusch-Godfrey-Test auf Autokorrelation 1. Ordnung am 5% Signifikanzniveau durch. Geben Sie Null- und Alternativhypothese, Teststatistik, Freiheitsgrade, Schlusslogik und Testergebnis an. *Hinweis*: Die Variable `L1.e` im Stata-Output bezeichnet das verzögerte Residuum e_{t-1}. [6 Punkte]

```
      Source |       SS       df       MS              Number of obs =     364
-------------+------------------------------           F(  4,   359) =  160.36
       Model |  636.157566     4  159.039391           Prob > F      =  0.0000
    Residual |  356.042705   359  .991762409           R-squared     =  0.6412
-------------+------------------------------           Adj R-squared =  0.6372
       Total |  992.200271   363  2.73333408           Root MSE      =  .99587

------------------------------------------------------------------------------
           e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           e |
         L1. |   .8033894   .0317211    25.33   0.000      .741007    .8657719
             |
        temp |   .0034219   .0099779     0.34   0.732    -.0162006    .0230443
      tempsq |  -.0000395   .0004385    -0.09   0.928    -.0009019    .0008228
  wochenende |   .1284414   .1157486     1.11   0.268    -.0991891    .3560719
       _cons |  -.0647561   .0774248    -0.84   0.404    -.2170193    .0875071
------------------------------------------------------------------------------
```

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

- Hypothesen: H0: $\rho = 0$ gegen H1: $\rho \neq 0$ [1P]
- Teststatistik: Mit R^2 aus der Hilfsregression für $t = 2, ..., T$

  $$e_t = \alpha + \rho e_{t-1} + \mathbf{x}_t' \beta + \nu_t$$

  kann die Teststatistik wie folgt berechnet werden:

  $$\chi^2 = (T-1)R^2 \sim \chi^2_1,$$

  wobei 1 Freiheitsgrad berücksichtigt wird. [2P]
- Entscheidungsregel: H0 verwerfen, falls $\chi^2 > \chi^2_{1,0.05} = 3.84$ [1P]
- Berechnung: $364 \cdot 0.64 = 232.96$ [1P]
- Entscheidung: H0 verwerfen, da $232.96 > 3.84$. [1P]

Aufgabe 4 **[15 Punkte]**

Die Wahrscheinlichkeit für den Bezug von Arbeitslosengeld II hängt von den sozioökonomischen Charakteristika einer Person. Dieser Zusammenhang wird mit einem binären Modell analysiert:

$$P(alg_i = 1|\mathbf{x}_i) = F(\beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 immi_i)$$

Dabei wird für F die Verteilungsfunktion der logistischen Verteilung angenommen. Die Variablen sind wie folgt definiert:

alg	Arbeitslosengeld-II-Bezieher (= 1, sonst= 0)
educ	Schulausbildung (in Jahren)
exper	Berufserfahrung (in Jahren)
immi	Migrationshintergrund (= 1, sonst= 0)

4.1 Nennen Sie zwei Gründe, die gegen eine Schätzung des Modells mit dem KQ-Verfahren sprechen. [2 Punkte]

- Die Varianz des KQ-Schätzers ist im Fall einer binären abhängigen Variable nicht konstant. Es liegen also heteroskedastische Störterme vor. [1P]
- Beim linearen Modell können die prognostizierten Wahrscheinlichkeiten außerhalb des Intervalls $[0,1]$ liegen. [1P]

4.2 Im Folgenden finden Sie die Ergebnisse einer Logit-Schätzung des Modells. Bestimmen und interpretieren Sie den marginalen Effekt eines weiteren Jahres Schulausbildung für einen Einheimischen mit 10 Jahren Schulausbildung und 10 Jahren Berufserfahrung. [5 Punkte]

```
Logistic regression                               Number of obs   =       4475
                                                  LR chi2(3)      =     281.33
                                                  Prob > chi2     =     0.0000
Log likelihood =  -1245.887                       Pseudo R2       =     0.1014

------------------------------------------------------------------------------
         alg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.2877139   .0260273   -11.05   0.000    -.3387264   -.2367014
       exper |  -.0505059   .0049531   -10.20   0.000    -.0602138    -.040798
        immi |   .5264153   .1324786     3.97   0.000      .266762    .7860686
       _cons |   1.798023   .3160041     5.69   0.000     1.178666     2.41738
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- Bestimmung des marginalen Effekts im Logit-Modell für einen Einheimischen mit 10 Jahren Schulausbildung und 10 Jahren Berufserfahrung:

$$\bar{\mathbf{x}}'\boldsymbol{\beta} = -0.288 \cdot 10 - 0.051 \cdot 10 + 0.526 \cdot 0 + 1.798 = -1.592 \text{ [2P]}$$

$$\frac{\partial F(\mathbf{x}_i'\boldsymbol{\beta})}{\partial educ} = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i'\boldsymbol{\beta}))^2}\beta_{educ} = \frac{\exp(-1.592)}{(1+\exp(-1.592))^2}(-0.288) = -0.040 \text{ [2P]}$$

- Ein weiteres Jahr Schulausbildung (des Haushaltsvorstandes) reduziert für einen Einheimischen mit 10 Jahren Schulausbildung und 10 Jahren Berufserfahrung c. p. im Mittel die Wahrscheinlichkeit, Alg II zu beziehen, um 4.0 Prozentpunkte. [1P]

4.3 Testen Sie, ob Migranten ihr höheres Risiko für Transferbezug durch ein zusätzliches Bildungsjahr kompensieren können. Führen Sie einen Likelihood-Ratio-Test am 5% Signifikanzniveau durch. Geben Sie die Nullhypothese, Spezifikation des restringierten Modells, Freiheitsgrade und kritischen Wert an. Berechnen Sie den empirischen Wert der Teststatistik und treffen Sie eine Testentscheidung. *Hinweis*: Der Log-Likelihood-Wert des restringierten Modells beträgt -1247,3104. [8 Punkte]

- Nullhypothese: $H_0 : \beta_2 = -\beta_4$ oder $H_0 : \beta_2 + \beta_4 = 0$ [1P]
- Spezifikation des restringierten Modells: [2P]

$$P(alg_i = 1|\mathbf{x}_i) = F(\beta_1 + \beta_3 exper_i + \beta_4(immi_i - educ_i))$$

- Freiheitsgrade: 1 [1P]
- kritischer Wert: $\chi^2_{1,0.05} = 3.84$ [1P]
- Berechnung: $LR = -2(\ln L_R - \ln L_U) = -2(-1247.3104 + 1245.887) = 2(-1.4234) = 2.847$ [2P]
- Entscheidung: Die H0 kann auf dem 5% Signifikanzniveau nicht verworfen werden. Migranten können ihr höheres Transferrisiko wohl durch ein zusätzliches Bildungsjahr kompensieren. [1P]

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.	Sie schätzen folgende Konsumfunktion mit KQ: $konsum_i = \beta_1 + \beta_2 log(einkommen_i) + u_i$. (*Hinweis*: *konsum* und *einkommen* sind in Euro gemessen.) b_2 beträgt 400. Welche Aussage ist richtig?
a	Wenn das Einkommen um 1000 Euro ansteigt, steigt der Konsum um 400 Euro.
b	Wenn das Einkommen um 1 Euro ansteigt, steigt der Konsum um 0,4%.
c	**X** Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 4 Euro.
d	Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 0,4%.

2.	Im Modell $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ kommt es bei Weglassen von x_{3i} zu einer negativen Verzerrung von β_2, wenn
a	**X** $\beta_3 > 0$ und $Cov(x_{2i}, x_{3i}) < 0$.
b	$\beta_3 > 0$ und $Cov(x_{2i}, x_{3i}) > 0$.
c	$\beta_3 < 0$ und $Cov(x_{2i}, x_{3i}) < 0$.
d	$\beta_3 > 0$ und $Cov(x_{2i}, x_{3i}) = 0$.

3.	Für das Modell $lohn_i = \beta_1 + \beta_2 alter_i + \beta_3 alter_i^2 + u_i$ ergibt eine KQ-Schätzung $b_2 = 0,28$ und $b_3 = -0,004$. In welchem Alter wird der durchschnittliche Lohn maximiert?
a	32.
b	**X** 35.
c	38.
d	40.

4.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i$. Man testet $H_0 : \beta_2 \leq 2$ vs. $H_1 : \beta_2 > 2$. Bei einem Signifikanzniveau von 5% führt eine Teststatistik von $1,71$
a	**X** bei $n = 30$ zur Ablehung von H_0.
b	bei $n = 20$ zur Ablehung von H_0.
c	bei $n = 35$ nicht zur Ablehung von H_0.
d	bei $n = 40$ nicht zur Ablehung von H_0.

5.	Welcher Ausdruck beschreibt den marginalen Effekt von exp für Männer gegeben das Modell $lohn_i = \beta_1 + \beta_2 frau_i + \beta_3 exp_i + \beta_4 age_i + \beta_5 exp_i^2 + \beta_6 exp_i \cdot frau_i + u_i$?
a	β_3.
b	**X** $\beta_3 + 2\beta_5 exp_i$.
c	$\beta_3 + 2\beta_5 exp_i + \beta_6 frau_i$.
d	$\beta_1 + \beta_3 + 2\beta_5 exp_i + \beta_6 frau_i$.

6.	Bei Vorliegen von perfekter Multikollinearität ist die Kreuzproduktmatrix $\mathsf{X}'\mathsf{X}$
a	invertierbar.
b	**X** singulär.
c	ein Skalar.
d	idempotent.

7.	Bei Vorliegen von Heteroskedastie
a	**X** können die Koeffizienten konsistent geschätzt sein.
b	müssen die Parameter kausal interpretiert werden.
c	sind die Konfidenzintervalle korrekt berechnet.
d	gilt das Gauss-Markov-Theorem.

8.	Wird die Nullhypothese eines Strukturbruchtests nicht abgelehnt, so
a	liegt ein Problem ausgelassener Variablen vor.
b	liegen signifikante Unterschiede in den Steigungsparametern der verschiedenen Gruppen vor.
c	liegt Heteroskedastie vor.
d	**X** ist das restringierte Modell nicht zu verwerfen.

9.	Ein Typ 2-Fehler tritt auf, wenn man die Nullhypothese
a	**X** nicht verwirft, obwohl diese falsch ist.
b	verwirft, obwohl diese falsch ist.
c	nicht verwirft, obwohl diese zutrifft.
d	verwirft, obwohl diese zutrifft

10.	Um einen RESET-Test durchzuführen, regressiert man in einer Hilfsregression
a	Residuen auf alle unabhängigen Variablen und zeitlich verzögerte Residuen.
b	**X** die abhängige Variable auf Polynome der vorhergesagten abhängigen Variable und alle unabhängigen Variablen.
c	quadrierte Residuen auf alle unabhängigen Variablen, deren Quadrate und deren Interaktionen.
d	die abhängige Variable auf die quadrierten Residuen.

11	Die Annahme $\varepsilon \sim i.i.d.(0, \sigma^2)$ impliziert, dass
a	**X** Heteroskedastie und Autokorrelation ausgeschlossen werden können.
b	die Störterme autokorreliert sein können.
c	die Störterme heteroskedastisch sein können.
d	die Störterme heteroskedastisch und autokorreliert sein können.

12.	Die FGLS-Schätzung bei heteroskedastischen Störtermen beruht darauf, dass
a	nur die abhängige Variable so transformiert wird, dass Homoskedastie vorliegt.
b	die Standardfehler des OLS Schätzers neu berechnet werden.
c	Beobachtungen mit großer Störtermvarianz ein größerers Gewicht erhalten als Beobachtungen mit kleiner Störtermvarianz.
d	**X** Beobachtungen mit großer Störtermvarianz ein kleineres Gewicht erhalten als Beobachtungen mit kleiner Störtermvarianz.

13.	Im einfachen linearen Regressionsmodell $y_i = \beta_1 + \beta_2 x_{2i} + u_i$ ist β_2
a	ein Schätzer.
b	ein mehrdimensionaler Vektor.
c	eine Zufallsvariable.
d	**X** eine Konstante.

14.	Die Summe quadrierter, standardnormalverteilter, unabhängiger Zufallsvariablen ist
a	normalverteilt.
b	**X** χ^2-verteilt.
c	F-verteilt.
d	t-verteilt.

15.	Wird ein irrelevanter Regressor (d.h. ein Regressor, dessen wahrer Koeffizient exakt gleich 0 ist) einem multiplen Regressionsmodell hinzugefügt, so
a	verändern sich die Regressionsergebnisse nicht.
b	**X** kann sich die Präzision der Schätzung verringern.
c	erhöht sich automatisch das R^2.
d	kann das Problem einer Verzerrung durch ausgelassene Variablen vermieden werden.

16.	Bei Querschnittsdaten kann
a	**X** es nicht zu zeitlich bedingter Autokorrelation kommen.
b	es nicht zu Heteroskedastie kommen.
c	der FGLS-Schätzer nicht angewendet werden.
d	das Maximum-Likelihood-Verfahren nicht angewendet werden.

17.	Bei einer starken Krümmung der Log-Likelihood-Funktion an der Stelle $\hat{\theta}$ ist die Schätzung des Parameters θ
a	unpräzise.
b	**X** präzise.
c	inkonsistent.
d	ineffizient.

18.	Ein niedriger AIC-Wert weist auf
a	eine Fehlspezifikation des Modells hin.
b	einen niedrigen Erklärungsgehalt des Modells hin.
c	**X** einen hohen Erklärungsgehalt des Modells hin.
d	ausgelassene Variablen hin.

19.	Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Tests
a	erlauben keinen Test von Linearkombinationen der Parameter.
b	haben eine approximativ normalverteilte Teststatistik.
c	haben eine approximativ standardnormalverteilte Teststatistik.
d	**X** sind asymptotisch äquivalent.

20.	Das Maximum-Likelihood-Verfahren
a	**X** benötigt Annahmen über die Verteilung der Störterme.
b	liefert für ein lineares Modell dieselben Ergebnisse wie das KQ-Verfahren.
c	hat bei Gültigkeit der Gauß-Markov-Annahmen die BLUE-Eigenschaft.
d	benötigt Annahmen über das präzise Konfidenzniveau der Parameter.

21.	Das Informationskriterium BIC
a	kann nur zum Vergleich genesteter Modelle verwendet werden.
b	**X** bestraft die Hinzunahme zusätzlicher Regressoren.
c	nimmt ausschließlich negative Werte an.
d	berechnet sich aus einer Funktion der Fehlerquadratsumme und Steigungsparameter.

22.	Logit- und Probit-Schätzer
a	ergeben identische Koeffizienten.
b	ergeben identische marginale Effekte.
c	**X** werden nicht bei stetigen abhängigen Variablen angewendet.
d	unterliegen identischen Annahmen bezüglich der Fehlertermverteilung.

23.	Das McFadden R^2
a	**X** liegt immer im Intervall [0;1].
b	nimmt im schlechtesten Fall den Wert 1 an.
c	nimmt im besten Fall den Wert 0 an.
d	Alle Aussagen sind richtig.

24.	Welche Aussage ist richtig?
a	Der Likelihood-Ratio-Test hat N Freiheitsgrade.
b	Zur Durchführung eines Wald-Tests ist die zweimalige Schätzung des Modells (d.h. mit bzw. ohne Restriktion) notwendig.
c	In kleinen Stichproben sind Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Tests äquivalent.
d	**X** Die Teststatistik des Lagrange-Multiplier-Tests folgt asymptotisch der χ^2-Verteilung.

25.	Autokorrelation im Störterm kann behoben werden durch
a	die Quadrierung der abhängigen Variablen.
b	die Aufnahme zusätzlicher Beobachtungen.
c	**X** eine GLS Transformation.
d	a und b.

26.	Bei Autokorrelation mit moving average Störprozessen der Form $\varepsilon_t = \nu_t + \nu_{t-1}$ (wobei ν_t den Gauß-Markov-Bedingungen genügt)
a	gibt es Fehlerterme ε_t und ε_s ($t \neq s$), die miteinander korrelieren.
b	gibt es Fehlerterme ε_t und ε_s ($t \neq s$), die nicht miteinander korrelieren.
c	ist die Varianz $Var(\varepsilon_t)$ konstant.
d	**X** Alle Aussagen sind richtig.

27.	Der Prais-Winsten-Schätzer
a	muss vor der Interpretation transformiert werden.
b	**X** nutzt mehr Beobachtungen als der Cochrane-Orcutt-Schätzer.
c	ist bei Zeitreihendaten nicht anwendbar.
d	b und c.

28.	Der Durbin-Watson-Test
a	verallgemeinert den Breusch-Godfrey-Test.
b	hat eine t-verteilte Teststatistik.
c	ist bei Zeitreihendaten nicht anwendbar.
d	**X** ist bei kleinen Stichproben gültig.

29.	Das Problem perfekter Multikollinearität
a	kann durch Vergrößerung der Stichprobe reduziert werden.
b	führt zu inkonsistenten Schätzergebnissen.
c	**X** kann durch Auslassen der linear abhängigen Variablen gelöst werden.
d	führt zu ineffizienten Schätzergebnissen.

30.	Mithilfe eines linearen Regressionsmodells lassen sich
a	keine Elastizitäten schätzen.
b	keine Semielastizitäten schätzen.
c	keine Koeffizienten schätzen.
d	**X** Keine der Aussagen ist richtig.