Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Prof. Regina T. Riphahn, Ph.D.

Masterprüfung Wintersemester 2022/2023 - 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: **[20 Punkte]**

Sie interessieren sich für die Determinanten der monatlichen Sparquote von Haushalten. Ihnen steht ein Datensatz für 104 private Haushalte mit folgenden Informationen zur Verfügung:

$savingsrate_i$	Sparquote von Haushalt i, gemessen in % des monatlichen verfügbaren Einkommens
age_i	Alter des Haushaltsvorstandes von Haushalt i in Jahren
$kids_i$	Anzahl an Kindern unter 18 Jahren in Haushalt i
$female_i$	=1, wenn der Haushaltsvorstand von Haushalt i weiblich ist, =0, wenn der Haushaltsvorstand von Haushalt i männlich ist

Sie stellen folgendes Regressionsmodell auf und schätzen dieses anschließend mit Hilfe des Kleinst-Quadrate-Schätzers:

$$savingsrate_i = \beta_1 + \beta_2 age_i + \beta_3 kids_i + \beta_4 female_i + \epsilon_i$$

health	Coef.	Std. Err.	P>|t|
cons	4.219	1.352	0.003
age	.834	0.266	0.001
kids	-5.250	4.177	0.429
female	12.324	3.559	???

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

1.1 Interpretieren Sie b_3 inhaltlich und statistisch. (2 Punkte)

- Inhaltlich: Ein zusätzliches Kind im Haushalt führt c.p. im Mittel zu einer Verringerung der Sparquote von 5,25 Prozentpunkten. [1P]
- Statistisch: Der Koeffizient ist nicht statistisch signifikant von Null verschieden am 10%-Niveau [1P]

1.2 Berechnen und interpretieren Sie das 90%-Konfidenzintervall von $female$. Ist der Koeffizient statistisch signifikant am 10%-Niveau? Begründen Sie Ihre Antwort. (5 Punkte)

- $[b_k - t_{N-K;\alpha/2} se(b_k); b_k + t_{N-K;\alpha/2} se(b_k)]$
- $[b_4 - t_{100;0,05} se(b_4); b_2 + t_{100;0,05} se(b_4)]$ [1P]
- $[12,324 - 1,660 \cdot 3,559; 12,324 + 1,660 \cdot 3,559]$ [1P]
- $[6,416; 18,232]$ [1P]
- Interpretation: Bei wiederholten Stichproben enthalten 90% aller auf diese Weise berechneten Konfidenzintervalle den wahren, aber unbekannten Wert des Koeffizienten. [1P]
- Der Koeffizient ist statistisch signifikant am 10%-Niveau, da das Konfidenzintervall die Null nicht enthält. [1P]

1.3 Nennen Sie die Null- und Alternativhypothese eines F-Tests auf gemeinsame Signifikanz von b_2 und b_3. Die Teststatistik beträgt 20 und das R^2 des unrestringierten Modells beträgt 0,5. Berechnen Sie das R^2 des restringierten Modells. Wie lautet Ihre Testentscheidung bei einem 5%-Signifikanzniveau? Begründen Sie Ihre Antwort. (5,5 Punkte)

- Hypothesen: $H_0 : \beta_1 = \beta_2 = 0$, H_1 : mindestens ein $\beta_j \neq 0$ mit $j = 1, 2$ [1P]

$$F^{emp} = \frac{(R^2_{UR} - R^2_R)/J}{(1 - R^2_{UR})/(N - K)}$$

$$20 = \frac{(0,5 - R^2_R)/2}{(1 - 0,5)/(100)} \qquad [1P]$$

$$0,5 - R^2_R = 20 \cdot 0,005 \cdot 2 \qquad [1P]$$

$$R^2_R = 0,3 \qquad [1P]$$

- $F^{emp} = 20 > F^{krit}_{0,95,2,100} = 3,09$ [1P]
- Die Nullhypothese wird abgelehnt, die Koeffizienten sind gemeinsam statistisch signifikant. [0,5P]

1.4 Sie vermuten, dass ein umgekehrt U-förmiger Zusammenhang zwischen age_i und $savingsrate_i$ vorliegt. Schreiben Sie eine Schätzgleichung auf, mit der Sie diese Vermutung testen können. Für welche Werte der geschätzten Koeffizienten würde sich Ihre Vermutung bestätigen? Welche Werte der geschätzten Koeffizienten deuten auf einen exakt linearen und positiven Zusammenhang zwischen age_i und $savingsrate_i$ hin? Und welche Werte der geschätzten Koeffizienten deuten auf einen positiven Zusammenhang mit zunehmender Steigung zwischen age_i und $savingsrate_i$ hin? (4,5 Punkte)

- $savingsrate_i = \beta_1 + \beta_2 age_i + \beta_3 kids_i + \beta_4 female_i + \beta_5 age_i^2 + \epsilon_i$ [1,5P] (Hinweis: auch ohne $kids_i$ und $female_i$ korrekt)
- Umgekehrt U-förmig: $b_2 > 0$ und $b_5 < 0$ [1P]
- Exakt linear mit positiver Steigung: $b_2 > 0$ und $b_5 = 0$ [1P]
- Positive, zunehmende Steigung: $b_2 > 0$ und $b_5 > 0$ [1P]

1.5 Sie vermuten, dass sich der Zusammenhang zwischen $kids_i$ und $savingsrate_i$ unterscheidet, je nachdem ob der Haushaltsvorstand männlich oder weiblich ist. Schreiben Sie eine Schätzgleichung auf, mit der Sie diese Vermutung testen können. Bei welcher Parameterkonstellation ergibt sich ein negativer Zusammenhang zwischen Anzahl der Kinder und Ersparnis für weibliche Haushaltsvorstände und gleichzeitig ein positiver Zusammenhang für männliche Haushaltsvorstände? (3 Punkte)

- $savingsrate_i = \beta_1 + \beta_2 age_i + \beta_3 kids_i + \beta_4 female_i + \beta_5 female_i * kids_i + \epsilon_i$ [1,5P]
- $b_3 > 0$, $b_5 < 0$ und $|b_5| > b_3$ [1,5P]

Aufgabe 2: **[12 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen Gesundheit und der Anzahl der Arbeitsstunden. Ihren Analysen liegt ein Datensatz für 950 Arbeitnehmende mit folgenden Informationen zu Grunde:

$health_i$	Einschätzung des Gesundheitszustands von Person i auf einer Skala von 0 (sehr schlecht) bis 100 (sehr gut)
$hours_i$	Tatsächliche wöchentliche Arbeitszeit von Person i in Stunden
age_i	Alter von Person i in Jahren
$female_i$	=1, wenn Person i weiblich, =0, wenn Person i männlich ist

Sie stellen folgendes Regressionsmodell auf und schätzen dieses anschliessend mit Hilfe des Kleinst-Quadrate-Schätzers:

$$health_i = \beta_1 + \beta_2 hours_i + \beta_3 age_i + \beta_4 female_i + \epsilon_i$$

health	Coef.	Std. Err.	t
cons	35.313	11.804	2.992
hours	-0.536	0.089	-6.022
age	-1.255	0.259	3.861
female	11.645	3.112	3.742

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

2.1 Interpretieren Sie b_3 inhaltlich. Ist b_3 statistisch signifikant von Null verschieden auf dem 1%-Signifikanzniveau? (3 Punkte)

- Steigt das Alter um 1 Jahr an, so sinkt ceteris paribus im Mittel die Gesundheit um 1,255 Skalenpunkte. [1P]
- $t^{emp} = 3,861$
- $c = t_{(\alpha/2),n-k-1} = t_{0,005,950-3-1} = t_{0,005,946} = t_{0,005,\infty} = 2,576$ [1P]
- Da $t^{emp} > c$, kann die Nullhypothese, dass b_3 nicht signifkant von Null verschieden ist, auf dem 1%-Signifikanzniveau abgelehnt werden. [1P]

2.2 Nennen Sie zwei potentielle Ursachen für eine Korreliertheit der Arbeitszeit mit dem Fehlerterm und bewerten Sie, ob im vorliegenden Beispiel die Annahme $E(x_i\epsilon_i) = 0$ (A.7) verletzt ist. (3 Punkte)

- Zwei potentielle Ursachen: Messfehler in den erklärenden Variablen, Umgekehrte Kausalität [1P]
  Alternative Antworten, z.B. Omitted variable bias möglich.
- Im vorliegenden Beispiel könnte Endogenität durch Umgekehrte Kausalität vorliegen. Die wöchentliche Arbeitszeit kann nicht nur den Gesundheitszustand beeinflussen, es könnte auch sein, dass der Gesundheitszustand die wöchentliche Arbeitszeit beeinflusst. [2P]
  Alternative Antworten möglich, z.B. unberücksichtigte Variable *Anzahl der Kinder*. Je mehr Kinder man hat, desto wahrscheinlicher wird man krank und arbeitet weniger.

2.3 Sie vermuten, dass die Variable *hours* endogen ist und möchten die Variable *contract* als Instrumentvariable für *hours* nutzen. *contract* gibt die vertraglich vereinbarte wöchentliche Arbeitszeit in Stunden an. Nennen und erklären Sie kurz die beiden Bedingungen, die hier erfüllt sein müssen und diskutieren Sie, ob diese im vorliegenden Fall erfüllt sind. (4 Punkte)

- Relevanz: Die Instrumentvariable *contract* muss mit der endogenen Variable *hours* korrelieren: $cov(hours_i, contract_i) \neq 0$ [1P]
- Exogenität: Die Instrumentvariable *contract* darf nicht mit dem Fehlerterm der Schätzgleichung korrelieren: $cov(contract_i, \epsilon_i) = 0$ [1P]
- Im vorliegenden Fall ist die Relevanz-Bedingung vermutlich gegeben. Es ist wahrscheinlich, dass die tatsächlichen Arbeitsstunden *hours* positiv mit den vertraglichen Arbeitsstunden *contract* korrelieren. [1P]
- Im vorliegenden Fall ist die Exogenitäts-Bedingung vermutlich nicht gegeben. Es kann unbeobachtete Faktoren im Störterm geben, die mit der Instrumentvariable korrelieren, z.B. das Gesundheitsbewusstsein. [1P]
  Alternative Antworten möglich.

2.4 Erläutern Sie den Begriff schwacher Instrumente. Benennen Sie eine Folge, die die Verwendung schwacher Instrumente für die Schätzung hat. (2 Punkte)

- Man spricht von schwachen Instrumenten, wenn die Instrumente nur schwach mit der endogenen Variable korrelieren. [1P] Zusatz: Es liegen schwache Instrumente vor, wenn die F-Statistik des Tests auf gemeinsame Signifikanz der Instrumentvariablen als erkärende Variablen für die endogene Variable (Regression erster Stufe) kleiner als 10 ist (Daumenregel).
- Bei schwachen Instrumenten kann selbst eine geringe Verletzung der Exogenitätsbedingung der Instrumente dazu führen, dass die Parameterschätzer inkonsistent sind. [1P]
  Alternative Antworten möglich, z.B. die Standardfehler sind falsch berechnet.

Aufgabe 3: **[20 Punkte]**

Sie untersuchen anhand von Quartalsdaten eines Unternehmens die Determinanten des Absatzes der meistverkauften Teesorte. Das Unternehmen hat Ihnen einen Datensatz mit folgenden Informationen zu 24 Zeitpunkten zur Verfügung gestellt:

abs_tee_t	Absatz des Tees in t (gemessen in Tausend Kilogramm)
$temp_t$	Durchschnittliche Temperatur in t (in °C)
$prom_t$	Werbeausgaben in t (gemessen in Tausend Euro)
$q1_t$	=1, wenn 1.-Quartal, 0=sonst
$q2_t$	=1, wenn 2.-Quartal, 0=sonst
$q3_t$	=1, wenn 3.-Quartal, 0=sonst
$q4_t$	=1, wenn 4.-Quartal, 0=sonst

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

$$abs\_tee_t = \beta_1 + \beta_2\ temp_t + \beta_3\ prom_t + \beta_4\ q2_t + \beta_5\ q3_t + \beta_6\ q4_t + \epsilon_t$$

```
      Source |       SS           df       MS      Number of obs   =        24
-------------+----------------------------------   F(5, 18)        =     13.38
       Model |  8.21044506         5  1.64208901   Prob > F        =    0.0000
    Residual |  2.20955494        18  .122753052   R-squared       =    0.7880
-------------+----------------------------------   Adj R-squared   =    0.7290
       Total |       10.42        23  .453043478   Root MSE        =    .35036

------------------------------------------------------------------------------
     abs_tee | Coefficient  Std. err.      t    P>|t|     [95\% conf. interval]
-------------+----------------------------------------------------------------
        temp |    .000659   .0001243     5.30   0.000     .0003979    .0009201
        prom |    .134313   .0476885     2.82   0.011     .234504     .0341245
             |
          q2 |  -.9537988   .2239996    -4.26   0.000    -.483193    -1.424405
          q3 |  -.6069537   .2304583    -2.63   0.017    -.1227788   -1.091129
          q4 |   .7091535   .2568359     2.76   0.013     .1695613    1.248746
             |
       _cons |   6.778283   2.749391     2.47   0.024     1.002027    12.55454
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

3.1 Interpretieren Sie inhaltlich und statistisch den geschätzten Koeffizienten b_3. (2 Punkte)

- b_3: Erhöhen sich die Werbeausgaben um tausend Euro, so steigt der Absatz des Tees c.p. im Durchschnitt um 134,143 Kilo. [1P]
- Der Koeffizient ist am 5%-Signifikanzniveau signifikant. [1P]

3.2 Prognostizieren Sie den Absatz des Tees im dritten Quartal eines Jahres mit einer durchschnittlichen Temperatur von 19 Grad °C und Werbeausgaben von 4,230 Euro. (1,5 Punkte)

- $\widehat{Absatz\_tee} = 6,778 + 0,001 * 19 + 0,134 * 4,230 - 0,954 * 0 - 0,607 * 1 + 0,709 * 0 = 6,757$ [1,5P]

3.3 Erläutern Sie knapp verbal, was unter Autokorrelation zu verstehen ist und nennen Sie zwei Eigenschaften des KQ-Schätzers bei Autokorrelation. Nennen Sie zusätzlich einen Grund, warum im genannten Beispiel Autokorrelation vorliegen könnte. (3,5 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 konsistent. [0,5P] (Andere Antworten sind möglich)
- Positive Autokorrelation im Fehlerterm könnte vorliegen, wenn Werte der abhängigen Variable über die Zeit positiv korrelieren, d.h. einem Quartal mit hohem Absatz folgen weitere erfolgreiche Quartale, und diese Korrelation durch die Regressoren nicht aufgefangen wird.
- Ein Beispiel wäre ein Schock in Form eines besonders kalten Winters, der den Absatz des Tees über mehrere Quartale beeinflusst. (Andere Antworten möglich.) [1,5P]

3.4 Sie vermuten Autokorrelation und führen einen Breusch-Godfrey-Test auf Autokorrelation 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                   ?????
3            |          6.864               3                   ?????
4            |         13.267               4                   ?????
---------------------------------------------------------------------------
H0: no serial correlation
```

Beschreiben Sie knapp die Komponenten der Teststatistik $LM = (T-4) \cdot R^2$ für den Test 4. Ordnung. Geben Sie eine für den Test relevante Hilfsregression, die Null- und Alternativhypothese, kritischen Wert und Testergebnis für das vorliegende Beispiel an. (6 Punkte)

- Hilfsregression: $e_t = \beta_1 + \beta_2 temp_t + \beta_3 prom_t + \beta_4 q2_t + \beta_5 q3_t + \beta_6 q4_t + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \rho_4 e_{t-4} + \nu_t$; alternativ $e_t = \beta_1 + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \rho_4 e_{t-4} + \nu_t$. [2P]
- Komponenten der Teststatistik: $T-4$ entspricht der Anzahl an Messzeitpunkten - 4 (hier $24-4$) und R^2 ist das Bestimmtheitsmaß aus der Hilfsregression. [1P]
- Hypothesen: $H_0 : \rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$ (keine Autokorrelation 4. Ordnung); $H_1 : mindestens\ ein\ \rho \neq 0\ mit\ j = 1,2,3,4$ (Autokorrelation 4. Ordnung). [1P]
- Kritischer Wert: $\chi^2_{4;5\%} = 9,49$. [1P]
- Testergebnis: Da $\chi^2_{empirisch} = 13,267 > 9,49 = \chi^2_{kritisch}$ kann die Nullhypothese auf dem 5% Signifikanzniveau verworfen werden. Es gibt Evidenz für Autokorrelation 4. Ordnung. [1P]

3.5 Eine Alternative zum Breusch-Godfrey-Test ist der Durbin-Watson-Test. Erläutern Sie zwei Stärken oder Schwächen des Durbin-Watson-Tests, durch die Sie sich gegen bzw. für die Anwendung des Durbin-Watson-Tests 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 beim DW-Test sind manchmal keine Aussagen bezüglich der Autokorrelation möglich. [1P]
- Alternative Antworten möglich.

3.6 Nehmen Sie nun eine Autokorrelation 1. Ordnung im oben genannten Modell an. Zeigen Sie das Vorgehen für die Transformation zu einem Prais-Winsten-Schätzer am konkreten Modell. Wäre eine Transformation zum Prais-Winsten-Schätzer in diesem Fall geeignet? (5 Punkte)

- Für die Beobachtungen 2-24: $abs\_tee_t - \hat{\rho} * abs\_tee_{t-1} = \beta_1(1-\hat{\rho}) + \beta_2\ (temp_t - \hat{\rho} * temp_{t-1}) + \beta_3\ (prom_t - \hat{\rho} * prom_{t-1}) + \beta_4\ (q2_t - \hat{\rho} * q2_{t-1}) + \beta_5\ (q3_t - \hat{\rho} * q3_{t-1}) + \beta_6\ (q4_t - \hat{\rho} * q4_{t-1}) + \epsilon_t - \hat{\rho} * \epsilon_{t-1}$ [2,5P]
- Für die erste Beobachtung: $\sqrt{1-\hat{\rho}^2} * abs\_tee_1 = \beta_1\sqrt{1-\hat{\rho}^2} + \beta_2\ \sqrt{1-\hat{\rho}^2} * temp_1 + \beta_3\ \sqrt{1-\hat{\rho}^2} * prom_1 + \beta_4\ \sqrt{1-\hat{\rho}^2} * q2_1 + +\beta_5\ \sqrt{1-\hat{\rho}^2} * q3_1 + \beta_6\ \sqrt{1-\hat{\rho}^2} * q4_1 + \sqrt{1-\hat{\rho}^2} * \epsilon_1$. [1,5P]
- Ja, da der Prais-Winsten-Schätzer für Autokorrelationen bis zur 1. Ordnung korrigiert. [1P]

Aufgabe 4: **[8 Punkte]**

Sie interessieren sich für die Determinanten der Erwerbstätigkeit von verheirateten Frauen. Ihnen liegt ein Datensatz von 5 634 Frauen mit folgenden Informationen vor:

$employed_i$	=1, wenn Frau i erwerbstätig ist, =0, wenn Frau i nicht erwerbstätig ist
age_i	Alter in Jahren
age_i^2	quadriertes Alter in Jahren
$educ_i$	Anzahl der Bildungsjahre in Jahren
$kids_i$	=1, wenn Frau i ein Kind hat, =0, wenn Frau i kein Kind hat

Sie stellen folgendes Regressionsmodell auf und schätzen dieses anschliessend mit Hilfe des Kleinst-Quadrate-Schätzers:

$$employed_i = \beta_1 + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 educ_i + \beta_5 kids_i + \epsilon_i$$

employed	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.148802	.050409	2.95	0.003	.049982	.247623
age_sq	-.00215	.000465	-4.62	0.000	-.00406	-.00164
educ	.0340276	.0024484	13.90	0.000	.0292278	.0388274
kids	-.1824777	.0163445	-11.16	0.000	-.2145192	-.1504361
_cons	.0771875	.1016541	0.76	0.448	-.1220937	.2764688

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

4.1 Mit welchem Alter ist die Erwerbswahrscheinlichkeit c.p. maximal? (2 Punkte)

- $\frac{\delta \widehat{employed}}{\delta age} = 0$
- $\beta_2 + 2\beta_3 \cdot age = 0$
- $0,149 + 2 \cdot (-0,002) \cdot age = 0$
- $age = \frac{-0,149}{2 \cdot (-0,002)}$
- $age = 37,25$

4.2 Nennen Sie zwei Nachteile des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

- Die vorhergesagten Werte für die abhängige Variable können Werte außerhalb des Intervalls [0;1] liegen. [1P]
- Die Varianz der Störterme ist nicht konstant, es liegt Heteroskedastie vor. [1P]
- Alternative Antworten möglich.

4.3 Sie entscheiden sich nun für die Schätzung eines Probit-Modells anstatt des obigen linearen Wahrscheinlichkeitsmodells.

```
Iteration 0:   log likelihood =  -3826.743
Iteration 1:   log likelihood = -3627.6167
Iteration 2:   log likelihood = -3626.9417
Iteration 3:   log likelihood = -3626.9417

Probit regression                               Number of obs     =      5,634
                                                LR chi2(4)        =     399.60
                                                Prob > chi2       =     0.0000
Log likelihood = -3626.9417                     Pseudo R2         =     0.0522

------------------------------------------------------------------------------
    employed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0364971   .0136693     2.67   0.008     .0097057    .0632885
      age_sq |  -.0007247   .0001669    -4.34   0.000    -.0010519   -.0003975
        educ |   .0935253   .0068841    13.59   0.000     .0800327    .1070178
        kids |  -.4971213   .0446753   -11.13   0.000    -.5846833   -.4095594
       _cons |  -1.094283   .2773115    -3.95   0.000    -1.637803   -.5507622
------------------------------------------------------------------------------
```

Interpretieren Sie b_4 inhaltlich und statistisch. (2 Punkte)

- Nur Vorzeichen und Signifianz sind interpretierbar.
- Es besteht ein positiver Zusammenhang zwischen der Erwerbstätigkeit von Frauen und der Anzahl der Bildungsjahre. [1P]
- Dieser Zusammenhang ist statistisch signifkant auf dem 1%-Signifikanzniveau. [1P]

4.4 Sie überlegen die Variable *huswage* mit ins Modell aufzunehmen Diese gibt das monatliche Einkommen des Ehemanns von Frau i in Euro an. Erklären Sie die Intuition des Likelihood-Ratio Tests, um zu überprüfen, ob *huswage* einen signifikanten Einfluss auf die Erbwerbstätigkeit von verheirateten Frauen hat. (2 Punkte)

- Man schätzt das Modell restringiert, also ohne *huswage*, und unrestringiert, also mit *huswage*. [1P] Im Anschluss vergleicht man die Werte der Log-Likelihoodfunktion der beiden Modelle. Unterscheiden diese sich signifikant, so trägt *huswage* signifikant zum Erklärungsgehalt der Erwerbstätigkeit von verheirateten Frauen bei. [1P]

Aufgabe 5: [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.	Was wurde in diesem Stata-Code geschätzt?

```
reg educ fatheduc c.exper##c.exper ,r
predict educ_hat, xb
reg lwage educ_hat c.exper##c.exper, r
```

a.	Eine logistische Regression.
b	Eine Maximum-Likelihood-Schätzung.
c	Eine GLS-Schätzung.
d	Eine IV-Schätzung. **X**

2.	Was wird in diesem Stata-Code erreicht?

```
quietly tab year, gen(y)
```

a	Jahres-Dummies werden erstellt, Tabellierung von der Variable year wird ausgegeben.
b	Jahres-Dummies werden erstellt, Tabellierung von der Variable year wird nicht ausgegeben. **X**
c	Es werden Kreuztabellen der einzelnen Jahre ausgegeben.
d	Es wird ein Kernel-Density-Plot ausgegeben.

3.	Was wird in diesem Stata-Code berechnet?

```
reg hwage yedu workexp
predict e, residual
gen e2 = e^2
reg e2 yedu workexp
scalar test_stat = e(N)*e(r2)
```

a	Teststatistik von dem Breusch-Pagan-Test wird berechnet. **X**
b	Teststatistik von dem White-Test wird berechnet.
c	Teststatistik von dem Durbin-Watson-Test wird berechnet.
d	Teststatistik von dem Breusch-Godfrey-Test wird berechnet.

4.	Ein RESET-Test überprüft,
a	auf Basis eines LM-Tests, ob ein Problem ausgelassener Variablen vorliegt.
b	ob durch Änderung der Modellspezifikation das angepasste Bestimmtheitsmaß $\overline{R}^2$ signifikant steigt.
c	auf Basis eines Signifikanztests, ob ein Problem ausgelassener Variablen vorliegt. **X**
d	ob sich die marginalen Effekte im Modell für Teilgruppen unterscheiden.

5.	Ein Konfidenzintervall
a	wird mit c.p. steigender Präzision schmaler. **X**
b	wird mit c.p. steigender Stichprobengröße breiter.
c	wird mit c.p. sinkendem Signifikanzniveau α schmaler.
d	hat stets negative Intervallgrenzen, wenn $b_j < 0$.

6.	Der Breusch-Pagan Test
a	kann die falsche H_0 mit höherer Wahrscheinlichkeit verwerfen als der White-Test. **X**
b	ist allgemeiner als der White-Test.
c	prüft in der Hilfsregression, ob e^2 durch die ursprünglichen Regressoren und deren Quadrate erklärt werden kann.
d	hat eine höhere Anzahl an Freiheitsgraden J als der White-Test.

7.	Es sei $V(\epsilon_i) = \sigma^2 \cdot Bildung_i^4$. Wie muss der transformierte Störterm $\tilde{\epsilon}$ aussehen, um eine konstante Störtermvarianz aufzuweisen?
a	$\tilde{\epsilon} = \epsilon_i / Bildung_i^{16}$.
b	$\tilde{\epsilon} = \epsilon_i / Bildung_i^{8}$.
c	$\tilde{\epsilon} = \epsilon_i \cdot Bildung_i^{4}$.
d	$\tilde{\epsilon} = \epsilon_i / Bildung_i^{2}$. **X**

8.	Wenn für zwei Zufallsvariablen X und Y gilt, dass $E(Y\|X) = E(Y)$, dann
a	ist der Korrelationskoeffizent zwischen X und Y gleich 0. **X**
b	ist die Kovarianz zwischen X und Y gleich 1.
c	ist der auf X bedingte Erwartungswert von Y gleich 0.
d	sind X und Y statistisch unabhängig.

9.	Die Annahme $\epsilon_t \sim i.i.d.(0, \sigma^2)$ impliziert, dass die Störterme
a	nicht heteroskedastisch und nicht autokorreliert sein können. **X**
b	heteroskedastisch und autokorreliert sein können.
c	autokorreliert sein können.
d	heteroskedastisch sein können.

10.	Die Alternativhypothese im Durbin-Watson Test besagt, dass
a	Heteroskedastie vorliegt.
b	Autokorrelation vorliegt. **X**
c	sowohl Heteroskedastie als auch Autokorrelation vorliegen.
d	weder Heteroskedastie noch Autokorrelation vorliegen.

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

12.	Sie führen einen Chow-Test auf Geschlechterunterschiede für die KQ-Schätzung des Modells $Wage_i = \beta_1 + \beta_2 female_i + \beta_3 educ_i + \epsilon_i$ durch. Die Teststatistik des Chow-Tests hat den p-Wert 0,0134. Was können Sie daraus schließen?
a	Männer haben im Mittel einen höheren Stundenlohn als Frauen auf dem 1%-Signifikanzniveau.
b	Der mittlere Stundenlohn von Frauen und Männern unterscheidet sich nicht signifikant auf dem 10%-Signifikanzniveau.
c	Die Regressionsparameter unterscheiden sich signifikant zwischen Männern und Frauen auf dem 5%-Signifikanzniveau. **X**
d	Keine der genannten Antworten.

13.	Die irrelevante Variable z in der KQ-Schätzung des Modells $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \epsilon_i$
a	hat keinen Einfluss auf die Effizienz des Koeffizientenschätzers b_2.
b	senkt die Störtermvarianz.
c	führt zu einer Verzerrung des Koeffizientenschätzers b_2.
d	erhöht die Varianz des Koeffizientenschätzers b_2, wenn $cov(x, z) \neq 0$. **X**

14.	In der ersten Stufe einer 2SLS-Schätzung
a	wird das Instrument auf die exogene Variable regressiert.
b	wird die endogene Variable auf das Instrument und die exogenen Variablen regressiert. **X**
c	wird die exogene Variable auf die endogene Variable regressiert.
d	keine der genannten Antworten.

15.	Ein konsistenter Schätzer
a	konvergiert bei steigender Stichprobengröße gegen den zu schätzenden Populationsparameter. **X**
b	ist auch ein unverzerrter Schätzer.
c	benötigt die Annahme homoskedastischer Störterme.
d	ist der Schätzer mit der geringsten Varianz.

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

17.	Die Varianz-Kovarianzmatrix der Störterme eines Regressionsmodells sei $\sigma^2\Psi$, wobei $\Psi = \begin{pmatrix} 3 & 0 & 6 \\ 0 & 3 & 0 \\ 6 & 0 & 3 \end{pmatrix}$. Hier liegt vor
a	Homoskedastie und positive Autokorrelation. **X**
b	Heteroskedastie und positive Autokorrelation.
c	Homoskedastie und keine Autokorrelation.
d	Heteroskedastie und keine Autokorrelation.

18.	Wenn $\Psi = I$ gilt, dann impliziert $Var(\epsilon) = \sigma^2\Psi$,
a	dass Heteroskedastie vorliegt.
b	dass $cov(\epsilon_i, \epsilon_j) \neq 0$ für alle $i \neq j$.
c	Autokorrelation vorliegt.
d	dass $cov(\epsilon_i, \epsilon_j) = 0$ für alle $i \neq j$. **X**

19.	Der Wert der Likelihoodfunktion für einen Münzwurf mit einer fairen Münze und 3 Wiederholungen, bei denen dreimal Zahl vorkommt ist gleich
a	0,125. **X**
b	0,5.
c	1,5.
d	1.

20.	Für die Beobachtung $(y_i, x_i) = (1, -3)$ beträgt das Residuum für die Schätzgleichung $\hat{y_i} = 8 - 0,25x_i$
a	7.
b	-0,75.
c	-7,75. **X**
d	2,25.

21.	Man spricht von einem überidentifizierten Modell, wenn
a	mehr Instrumente als exogene Variablen vorliegen.
b	die Nullhypothese des Durbin-Wu-Hausman Tests abgelehnt wird.
c	mehr Instrumente als endogene Variablen vorliegen. **X**
d	die F-Statistik in der ersten Stufe einen Wert größer als 10 annimmt.

22.	Wenn die Matrix $X'X$ für ein lineares Regressionsmodel nicht den vollen Rang hat, ist
a	die Matrix invertierbar.
b	die Matrix regulär.
c	die Korrelation unter den erklärenden Variablen gleich Null.
d	die Matrix singulär. **X**

23.	Wenn A eine Matrix mit der Dimension $[12 \times 5]$ ist und B eine Matrix mit der Dimension $[12 \times 8]$, dann
a	ist AB nicht definiert. **X**
b	hat AB die Dimension $[8 \times 5]$.
c	hat AB die Dimension $[5 \times 8]$.
d	hat AB die Dimension $[12 \times 12]$.

24.	Ein Typ II Fehler liegt vor, wenn man
a	eine zutreffende Nullhypothese ablehnt.
b	eine falsche Nullhypothese nicht ablehnt. **X**
c	eine falsche Nullhypothese ablehnt.
d	eine zutreffende Nullhypothese nicht ablehnt.

25.	Für die Hilfsregression des RESET Tests bildet man das Quadrat und höhere Polynome der
a	potenziell endogenen Variablen.
b	vorhergesagten Werte der abhängigen Variable. **X**
c	abhängigen Variable.
d	Residuen.

26.	Sie schätzen folgendes Modell mit dem KQ-Verfahren: $Lebenszufriedenheit_i = \beta_1 + \beta_2 Monatseinkommen(inEuro)_i + \epsilon_i$. b_2 nimmt den Wert 0,024 an. Welchen Wert würde b_2 annehmen, wenn das Monatseinkommen in 100 Euro anstatt in Euro gemessen wäre?
a	2,4. **X**
b	0,00024.
c	0,24.
d	100,024.

27.	Sie schätzen das Modell $ln\_Monatseinkommen_i = \beta_1 + \beta_2 male_i + \beta_3 age_i + \epsilon_i$ mittels KQ-Schätzer. Der geschätzte Koeffizient b_2 ist $0,286$. Wie hoch ist der erwartete Unterschied im monatlichen Einkommen zwischen Männern und Frauen mit gleichem Alter genau?
a	$28,60\%$
b	$33,11\%$ **X**
c	$38,97\%$
d	$52,20\%$

28.	Im linearen Modell $wage_i = \beta_1 + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 educ_i + \epsilon_i$ seien age_i und age_i^2 exogen und $educ_i$ endogen, wobei $educ_i$ durch eine geeignete Instrumentvariable z_i instrumentiert wird. Welche der folgenden Bedingungen benötigt man zur Herleitung des IV-Schätzers <u>nicht</u>:
a	$\frac{1}{N}\sum_{i=1}^{N}(wage_i - b_1 - b_2 age_i - b_3 age_i^2 - b_4 educ_i) age_i^2 = 0.$
b	$\frac{1}{N}\sum_{i=1}^{N}(wage_i - b_1 - b_2 age_i - b_3 age_i^2 - b_4 educ_i) = 0.$
c	$\frac{1}{N}\sum_{i=1}^{N}(wage_i - b_1 - b_2 age_i - b_3 age_i^2 - b_4 educ_i) educ_i = 0.$ **X**
d	$\frac{1}{N}\sum_{i=1}^{N}(wage_i - b_1 - b_2 age_i - b_3 age_i^2 - b_4 educ_i) z_i = 0.$

29.	Wenn der Störterm einer Regression nicht normalverteilt ist, dann
a	ist die Annahme $E[\mathbf{x}\epsilon] = 0$ verletzt.
b	sinkt die Effizienz der Schätzung.
c	minimiert das KQ-Verfahren nicht die Summe der quadrierten Residuen.
d	sind bei KQ-Schätzern die t- und F-Tests asymptotisch zutreffend. **X**

30.	Das Informationskriterium AIC
a	wird verwendet, um zu überprüfen, ob die Log-Likelihood Funktion korrekt spezifiert ist.
b	hat weniger Informationsgehalt als das R^2.
c	bewertet die Effizienz einer Schätzung.
d	kann zum Vergleich genesteter Modelle verwendet werden. **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 $\left[\chi_n^2 \le c\right] = P$, mit n Freiheitsgraden

P / n	**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 4a: 95% Perzentile der *F*-Verteilung

Zelleneintrag: f, sodass Prob $\left[F_{n1,n2} \leq f\right] = 0.95$

	n1 = Freiheitsgrade des Zählers								
n2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
∞	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.88

	n1 = Freiheitsgrade des Zählers								
n2	**10**	**12**	**15**	**20**	**30**	**40**	**50**	**60**	**∞**
1	241.88	243.91	245.95	248.01	250.10	251.14	251.77	252.20	254.19
2	19.40	19.41	19.43	19.45	19.46	19.47	19.48	19.48	19.49
3	8.79	8.74	8.70	8.66	8.62	8.59	8.58	8.57	8.53
4	5.96	5.91	5.86	5.80	5.75	5.72	5.70	5.69	5.63
5	4.74	4.68	4.62	4.56	4.50	4.46	4.44	4.43	4.37
6	4.06	4.00	3.94	3.87	3.81	3.77	3.75	3.74	3.67
7	3.64	3.57	3.51	3.44	3.38	3.34	3.32	3.30	3.23
8	3.35	3.28	3.22	3.15	3.08	3.04	3.02	3.01	2.93
9	3.14	3.07	3.01	2.94	2.86	2.83	2.80	2.79	2.71
10	2.98	2.91	2.85	2.77	2.70	2.66	2.64	2.62	2.54
15	2.54	2.48	2.40	2.33	2.25	2.20	2.18	2.16	2.07
20	2.35	2.28	2.20	2.12	2.04	1.99	1.97	1.95	1.85
25	2.24	2.16	2.09	2.01	1.92	1.87	1.84	1.82	1.72
30	2.16	2.09	2.01	1.93	1.84	1.79	1.76	1.74	1.63
40	2.08	2.00	1.92	1.84	1.74	1.69	1.66	1.64	1.52
50	2.03	1.95	1.87	1.78	1.69	1.63	1.60	1.58	1.45
70	1.97	1.89	1.81	1.72	1.62	1.57	1.53	1.50	1.36
100	1.93	1.85	1.77	1.68	1.57	1.52	1.48	1.45	1.30
∞	1.83	1.75	1.67	1.57	1.46	1.39	1.34	1.31	1.30

Quelle: In R generiert

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 symmetrisch 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:

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

Korrelationskoeffizient:

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

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

***Rechenregeln*:**

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

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

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

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 $\text{cov}\{\varepsilon_i, \varepsilon_j\} = 0$ 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 $V\{\varepsilon \mid X\} = \sigma^2\, \text{Diag}\{h_i^2\} = \sigma^2 \Psi$

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:

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

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

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

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

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