Aufgabe 1: ***(20 Punkte)***

Sie sind am Konsumverhalten in Ihrem persönlichen Umfeld interessiert und führen hierzu eine eigene kleine Erhebung durch. Sie fragen 40 Bekannte nach der Höhe Ihrer Ausgaben für Luxusartikel (L) im Jahr 2007. Zusätzlich erfragen Sie die Einflussfaktoren:
E_t = Jahreseinkommen (in Euro), A_t = Alter (in Jahren), G_t = Geschlecht (Mann = 1, Frau=0).
Sie schätzen folgendes Modell:

$$L_t = \beta_0 + \beta_1 \cdot E_t + \beta_2 \cdot A_t + \beta_3 \cdot A_t^2 + \beta_4 \cdot G_t + \beta_5 \cdot (A_t \cdot E_t) + e_t$$

```
Call:
lm(formula = L ~ E + A + A^2 + G + A*E)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     628.0        ?       2.626  0.01286 *
E               0.013     0.0043   2.997  0.00506 **
A             -33.338      15.68      ?     0.04060 *
A^2             0.480      0.240   1.997  0.05388 .
G                 ?        41.04  -1.965  0.05764 .
A*E           -0.0003     0.0001  -2.551  0.01542 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 119 on 34 degrees of freedom
Multiple R-squared: ?,     Adjusted R-squared:  0.4176
F-statistic: 6.595 on 5 and 34 DF,  p-value: 0.0002177
```

a) Berechnen Sie die Werte der im Regressionsoutput fehlenden Angaben für *(1,5 Punkte)*

a1) den t-Wert zu b_2,

$$t_2 = \frac{b_2}{se(b_2)} = \frac{-33{,}338}{15{,}68} = -2{,}126$$

a2) den Standardfehler von b_0 sowie

$$se(b_0) = \frac{b_0}{t_0} = \frac{628{,}0}{2{,}626} = 239{,}147$$

a3) b_4.

$$b_4 = t_4 \cdot se(b_4) = 41{,}04 \cdot (-1{,}965) = -80{,}64$$

b) Berechnen Sie das multiple R^2 mit Hilfe der Informationen aus dem Regressionsoutput. *(3 Punkte)*

$$R^2 = 1 - \frac{SSE}{SST}$$

$$\overline{R}^2 = 1 - \frac{SSE/(T-K)}{SST/(T-1)}$$

$$\Rightarrow R^2 = 1 - \left[\left(1 - \overline{R}^2\right) \cdot \frac{T-K}{T-1}\right]$$

$$R^2 = 1 - \left[\left(1 - 0{,}4176\right) \cdot \frac{34}{39}\right] = 0{,}492$$

c) Berechnen Sie den marginalen Effekt

c1) des Jahreseinkommens für eine Person im Alter von 36 Jahren, *(2 Punkte)*

$$\frac{\partial L}{\partial E} = \beta_1 + \beta_5 \cdot A$$
$$0{,}013 + (-0{,}0003) \cdot 36 = 0{,}0022$$

c2) des Alters, wenn eine Person 36 Jahre alt ist und das Jahreseinkommen 90000 € beträgt. *(2,5 Punkte)*

$$\frac{\partial L}{\partial A} = \beta_2 + 2\beta_3 \cdot A + \beta_5 \cdot E$$
$$-33{,}338 + 2 \cdot 0{,}48 \cdot 36 + (-0{,}0003) \cdot 90000 = -25{,}778$$

d) Hat das Alter einen statistisch signifikanten Einfluss? Testen Sie auf dem 5%-Signifikanzniveau. Benutzen Sie folgende Werte: SSE_U = 480000 und SSE_R = 710000. Geben Sie an, welchen Test Sie benutzen. Nennen Sie die Null- und Alternativhypothese, den Wert der Teststatistik, den kritischen Wert sowie die Testentscheidung. *(4 Punkte)*

- *F-Test 0,5P*

$H_0 : \beta_2 = \beta_3 = \beta_5 = 0 \qquad H_1 : \textit{mind. ein Koeffizient} \neq 0$

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)} = 5{,}43$$

$F_{c(3;34;0,05)} \approx F_{c(3;30;0.05)} = 2{,}92$

$\rightarrow F > F_c \rightarrow H_0$ *verwerfen* $\rightarrow$ *Alter hat einen signifikanten Einfluss*

e) Sie sind nicht sicher, ob Sie vergessen haben, aussagekräftige Variablen in Ihr Modell aufzunehmen. Sie machen einen Test und erhalten den folgenden Regressionsoutput.

```
Coefficients:
             Estimate Std. Error
(Intercept)     627.4      239.8
E               0.012     0.0047
A             -33.418      16.29
A^2             0.482      0.239
G             -81.729      40.39
A*E           -0.0003     0.0001
L̂^2           0.00214      0.002
---

Residual standard error: 118.2 on 33 degrees of freedom
Multiple R-squared: 0.523,     Adjusted R-squared:   0.436
F-statistic: 5.923 on 6 and 33 DF,  p-value: 0.0003189
```

e1) Welchen Test haben Sie durchgeführt? Erläutern Sie knapp aber präzise die Vorgehensweise und die Testidee. *(4 Punkte)*

- *RESET-Test*
- *vorhergesagte Werte erzeugen*
- *Die für die abhängige Variable vorhergesagten Werte werden in der zweiten, dritten… Potenz als erklärende Variable in die Regressionsgleichung aufgenommen.*
- *Die Koeffizienten der hinzugefügten Variablen werden auf gemeinsame Signifikanz getestet.*
- *Wenn die Koeffizienten gemeinsam signifikant sind, war das ursprüngliche Modell falsch spezifiziert. Denn gibt es nicht berücksichtigte Variablen, welche mit den Variablen im Modell korreliert sind, dann wird dies durch das Polynom der vorhergesagten Werte dargestellt.*

e2) Geben Sie die Null- und Alternativhypothese an, die Teststatistik sowie den kritischen Wert am 5%-Signifikanzniveau. Was ist Ihr Testergebnis und was bedeutet es?

(3 Punkte)

$H_0 : \gamma = 0 \quad H_1 : \gamma \neq 0$

$t = \dfrac{\gamma}{se(\gamma)} = 1{,}07$

$t_{c(33;0,05)} = 2{,}03$

$t < t_c \rightarrow H_0$ *kann nicht verworfen werden* $\rightarrow$ *Kein Hinweis auf Fehlspezifikation.*

Aufgabe 2: ***(23 Punkte)***

Ihnen liegen Daten über die Erwerbsbeteiligung von Frauen in 180 Landkreisen Deutschlands vor. Weiterhin können Sie Informationen zu einigen soziodemografischen Charakteristika der Kreise verwenden. Die in Ihrem Datensatz enthaltenen Variablen lauten im Einzelnen:

E_t: Erwerbsbeteiligung der Frauen in Kreis t (Anteil in Prozent)
K_t: durchschnittliche Kinderzahl pro Frau in Kreis t
L_t: Dummy-Variable, die anzeigt, ob der Kreis t überwiegend ländlich oder überwiegend städtisch geprägt ist (1=ländlich, 0=städtisch)
$RELIGION_t$: Anteil der Personen an der Gesamtbevölkerung in Kreis t, die sich selbst als religiös einstufen (in Prozent)
$ALTER_t$: Anteil der Frauen im gebärfähigen Alter (15 bis 45 Jahre) an allen Frauen in Kreis t (in Prozent)

Sie möchten die Determinanten der Erwerbsbeteiligung von Frauen untersuchen. Sie schätzen zunächst das Modell: $E_t = \beta_0 + \beta_1 \cdot K_t + \beta_2 \cdot L_t + e_t$

a) Erläutern Sie kurz verbal am Beispiel, unter welchen Bedingungen ein Multikollinearitätsproblem im Modell vorliegen könnte. Welche Konsequenzen würden sich für die Schätzung ergeben? *(2 Punkte)*
   - *könnte auftreten, wenn die Variablen „Kinder“ und „Land“ mind. approximative Lineartransformationen voneinander sind (linear abhängig sind, hoch korreliert sind)*
   - *die Schätzer hätten hohe Standardfehler und niedrige t-Werte (werden unpräzise)*

b) Sie interessieren sich für Unterschiede zwischen städtischen und ländlichen Kreisen. *(4 Punkte)*
   b1) Wie können Sie testen, ob sich die erwartete Erwerbsbeteiligung in einem ländlichen und einem städtischen Kreis mit jeweils identischer Kinderzahl unterscheidet?
   - *t-Test auf* $\beta_2=0$

   b2) Der marginale Effekt der Kinderzahl auf die Erwerbsbeteiligung in ländlichen und städtischen Kreisen kann sich unterscheiden. Wie würden Sie vorgehen, um dies zu prüfen? Stellen Sie hierzu eine geeignete Modellgleichung auf und geben Sie Ihre Nullhypothese an.
   - *Einfügen eines Interaktionsterms K*L und Signifikanztest*
     $E_t = \beta_0 + \beta_1 \cdot K_t + \beta_2 \cdot L_t + \beta_3 \cdot (K \cdot L)_T + e_t$
     t-Test auf H_o: $\beta_3=0$

c) Sie vermuten ein Endogenitätsproblem bei der Verwendung der Variable *K.* *(5 Punkte)*
   c1) Erläutern Sie kurz am Beispiel, unter welchen Umständen ein solches Problem auftreten könnte. Geben Sie eine mögliche Erklärung, wieso dies hier der Fall sein könnte.
   - *Wenn die Variable K und die Variable E gemeinsam durch etwas unbeobachtetes Drittes determiniert werden*
     (alternativ: weil sowohl die Zahl der Kinder die Erwerbsbeteiligung bestimmen könnte als auch umgekehrt; weil die Variable Kinder mit einer ausgelassenen Variable korreliert ist)

· *eine solche unbeobachtete Determinante könnte z.B. die Einstellung zu Beruf und Familie sein*

c2) Welche formale Annahme wäre verletzt?

· $E(K \cdot e) = 0$ *(oder: $Cov(K,e) = 0$) ist nicht erfüllt*

c3) Vermuten Sie im vorliegenden Beispiel bei der Verwendung eines KQ-Schätzers einen korrekten, einen *betraglich* zu großen oder zu kleinen Schätzwert für den Steigungsparameter β_1? Warum? Erläutern Sie Ihre Vermutung.

· *zu klein (bzw. betragsmäßig zu groß, zu stark negativ)*

· *Dies ist vermutlich auf negative Korrelation zwischen Kinderzahl und Störterm zurück zuführen (kleine/negative Störterme bei großer Kinderzahl und umgekehrt)*

d) Jemand schlägt vor, die Variable RELIGION als Instrument für K zu nutzen. *(6 Punkte)*

d1) Welche formalen Bedingungen an ein Instrument sollte RELIGION hierzu erfüllen?

· $E(RELIGION \cdot e) = 0$ oder $Cov(RELIGION, e) = 0$

· $COV(RELIGION \cdot K) \neq 0$

d2) Nennen Sie eine empirische Maßzahl, die Sie heranziehen können, um die Qualität dieses Instruments zu prüfen. Gibt es Anforderungen an ein Instrument, deren Erfüllung Sie nicht prüfen können?

· *Korrelationskoeffizient zwischen erklärender Variable und Instrument prüft zweite Bedingung*

· *erste Bedingung nicht prüfbar*

d3) Argumentieren Sie *kurz* verbal, ob eine der in d1) genannten Bedingungen in diesem Beispiel verletzt sein könnte.

· *möglich, dass RELIGION gerade die bislang unbeobachtete Einstellung zu Beruf und Familie misst – weitere Korrelation zwischen Religion und Störterm nicht zu erwarten; oder: Genau wie K und E gemeinsam z.B. durch unbeobachtete Werte determiniert werden könnten, könnte das auch für RELIGION und E gelten – weitere Korrelation zwischen Religion und Störterm zu erwarten; oder irgendetwas anderes Plausibles*

e) Sie entscheiden sich, die Variable ALTER als Instrument zu nutzen. Erklären Sie, wie Sie anhand eines Hausman-Tests prüfen können, ob die Instrumentierung notwendig ist. Beschreiben Sie die dazu notwendigen Schritte zur Durchführung, nennen Sie die konkret verwendete Null- und Alternativhypothese und treffen Sie eine Entscheidung anhand des folgenden Outputs: *(6 Punkte)*

```
Hausman Test

Model Formula: E ~ K + L + vhat

            Estimate   Std. Error t value Pr(>|t|)
(Intercept) 73.633241   4.879251   15.091    0.412
K           -0.128279   0.047284   -2.713 5.71e-12
L           -0.132536   0.030376   -4.309 3.62e-05
vhat        -0.036455   0.011524   -3.163 4.14e-03
```

· *Schätzung der Hilfsregression:* $KINDER_t = a_0 + a_1 \cdot ALTER_t + v_t$

· *Vorhersage für die Variable K und Ermittlung des Vektors* $\hat{v}_t$

· *Schätzung der Regression:*

$$ERWERB_t = \beta_0 + \beta_1 \cdot KINDER + \beta_2 \cdot LAND_t + \delta \cdot \hat{v}_t + \mu_t$$

· *t-Test der* $H_0: \delta = 0$ *gegen* $H_1: \delta \neq 0$

· *führt hier zum Verwerfen der H_0, da* `vhat` *statistisch signifikant ist.*

- *Das Ergebnis des Tests ist, dass der IV-Schätzer dem KQ-Schätzer vorzuziehen ist.*

Aufgabe 3: ***(22 Punkte)***

Sie beschäftigen sich mit der Frage, welche Determinanten für die Entscheidung verantwortlich sind, häufig den Arbeitgeber zu wechseln. In Ihrem Datensatz haben Sie folgende Variablen:

JOB_t: die Anzahl der Arbeitgeberwechsel pro Person innerhalb der letzten 5 Jahre
ABI_t: Abitur als höchster Bildungsabschluss = 1, sonst = 0
UNI_t: Hochschulabschluss = 1, sonst = 0
AGE_t: Alter der Person in Jahren

Sie führen folgende Regression durch: $JOB_t = \beta_0 + \beta_1 \cdot ABI_t + \beta_2 \cdot UNI_t + \beta_3 \cdot AGE_t + e_t$

Als Schätzoutput erhalten Sie:

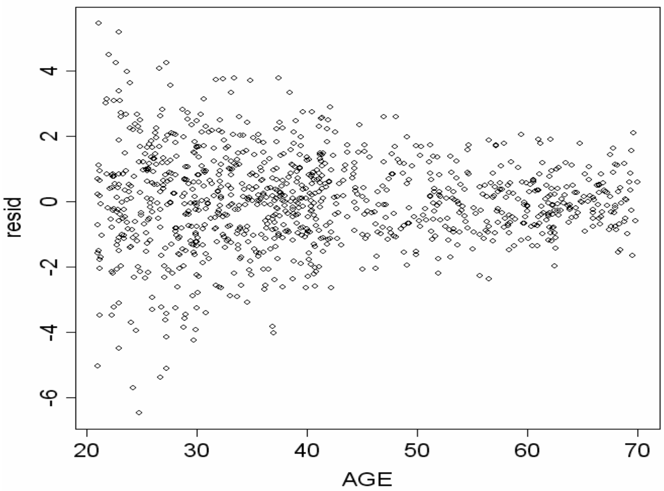
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.979454   0.149469  13.243  < 2e-16
ABI          0.226281   0.125171   1.808    0.071
UNI          0.741697   0.139332   5.323 1.26e-07
AGE         -0.030382   0.003351  -9.066  < 2e-16
---

Residual standard error: 1.449 on 996 degrees of freedom
Multiple R-Squared: 0.1402,     Adjusted R-squared: 0.1376
F-statistic: 54.14 on 3 and 996 DF,  p-value: < 2.2e-16
```

a) Interpretieren Sie die geschätzten Steigungsparameter statistisch und inhaltlich. Achten Sie auf die Exaktheit Ihrer Aussagen. *(4,5 Punkte)*

- *b1 auf 10% Signifikanzniveau von Null verschieden, im Vergleich zu Personen deren höchster Bildungsabschluss weniger als Abitur ist, haben Personen mit Abitur als höchstem Abschluss innerhalb von 5 Jahren im Erwartungswert 0,23 mehr Jobwechsel*
- *b2 auf 1% Signifikanzniveau von Null verschieden, im Vergleich zu Personen deren höchster Bildungsabschluss weniger als ein Hochschulabschluss ist, haben Personen mit Hochschulabschluss innerhalb von 5 Jahren im Erwartungswert 0,74 mehr Jobwechsel*
- *b3 auf 1% Signifikanzniveau von Null verschieden, Personen die ceteris paribus ein Jahr älter sind, haben innerhalb von 5 Jahren im Erwartungswert um 0,03 weniger Jobwechsel*

b) Sie bilden die geschätzten Residuen in Abhängigkeit von den exogenen Variablen ab. Für AGE erhalten Sie folgende Grafik:



Sie folgern, dass Sie möglicherweise mit Heteroskedastie konfrontiert sind. Erklären Sie, was Heteroskedastie ist, welche Annahme des klassischen Modells verletzt wird und nennen Sie die Konsequenz für den KQ-Schätzer. *(3 Punkte)*

- *Heteroskedastie liegt vor, wenn die Störterme keine einheitliche Varianz aufweisen:*
- *$var(e_t) = \sigma_t^2$*
- *dadurch ist die Annahme $var(e_t) = \sigma^2$ verletzt*
- *eine Kleinstquadrateschätzung bei Vorliegen heteroskedastischer Störterme ist nicht effizient; oder: eine Kleinstquadrateschätzung liefert falsche Standardfehler*

c) Um sich Gewissheit zu verschaffen, führen Sie einen Goldfeld-Quandt-Test durch. Erläutern Sie die Testidee und die einzelnen Schritte. Geben Sie Nullhypothese, Teststatistik, Freiheitsgrade und Schlusslogik an. Erläutern Sie Ihr Vorgehen detailliert am Beispiel der Aufgabenstellung. *(7,5 Punkte)*

- *für den GQ Test wird die Stichprobe in zwei annähernd gleich große Teilstichproben aufgeteilt*
- *dabei wird das Ziel verfolgt, in einer Teilstichprobe möglichst die Beobachtungen mit der kleinsten Störtermvarianz und in der anderen Teilstichprobe die mit der größten Störtermvarianz zu haben*
- *im Beispiel nimmt die Varianz der Residuen mit zunehmendem Alter ab - daher sollten die Beobachtungen nach dem Alter geordnet werden und in zwei Teilstichproben mit jeweils 500 Beobachtungen aufgeteilt werden*
- *für beide Teilstichproben werden die Fehlertermvarianzen σ_t^2 in zwei getrennten Regressionen geschätzt*
- *unter der Nullhypothese H_0: $\sigma_t^2 = \sigma^2$ liegt Homoskedastie vor*
- *die Teststatistik $GQ = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ folgt einer F-Verteilung mit jeweils $(T_i - K) = 496$ Freiheitsgraden*
- *in der Teststatistik steht die größere Fehlertermvarianz im Zähler*
- *ist GQ größer als der kritische Wert der F-Verteilung wird H_0 abgelehnt und wir gehen vom Vorliegen von Heteroskedastie aus*

d) Unterstellen Sie, dass der Goldfeld-Quandt-Test auf Heteroskedastie hinweist. Sie beschließen, einen generalisierten KQ-Schätzer (GLS-Schätzer) anzuwenden. *(5 Punkte)*

d1) Treffen Sie eine für die GLS Transformation erforderliche sinnvolle Annahme und geben Sie die transformierte Schätzgleichung vollständig an. Orientieren Sie sich dabei wieder am Beispiel der Aufgabenstellung.

- *eine sinnvolle Annahme über die Art der Heteroskedastie (vgl. Residuenplot) wäre z.B.* $\operatorname{var}(e_t) = \sigma_t^2 = \sigma^2 / AGE_t$ *oder* σ^2 / AGE_t^2
- *Transformation:*

$$JOB_t \cdot \sqrt{AGE_t} = \beta_0 \cdot \sqrt{AGE_t} + \beta_1\left(ABI_t \cdot \sqrt{AGE_t}\right) + \beta_2\left(UNI_t \cdot \sqrt{AGE_t}\right) + \beta_3 AGE_t^{\frac{3}{2}} + e_t \cdot \sqrt{AGE_t}$$

d2) Zeigen Sie algebraisch, dass die Transformation ihr Ziel erreicht hat.

- *der neue Störterm e_t^* ist homoskedastisch, da:*

$$\operatorname{var}\left(e_t^*\right) = \operatorname{var}\left(e_t \cdot \sqrt{AGE_t}\right) = AGE_t \cdot \operatorname{var}(e_t) = AGE_t \cdot \frac{\sigma^2}{AGE_t} = \sigma^2$$

e) Welche Alternative zum GLS Verfahren kennen Sie, um mit Heteroskedastie umzugehen? Erläutern Sie kurz das Vorgehen. *(2 Punkte)*

- *die Alternative ist der White – Schätzer für die Varianz des Kleinstquadrate Schätzers*

- *hierbei werden die unbekannten σ_t^2 durch die Quadrate der Kleinstquadrate Residuen $\hat{e}_t^2$ ersetzt und letztere dann für die Varianzberechnung verwendet*

Aufgabe 4: ***(10 Punkte)***

In R wurde folgende Funktion programmiert:

```
my.function <- function(x,y)
 {
  kq <- lm(y~x)
  plot(x,y)
  lines(x,y)
  abline(h=0)
  r <- cor(x,y)
  return(r)
}
```

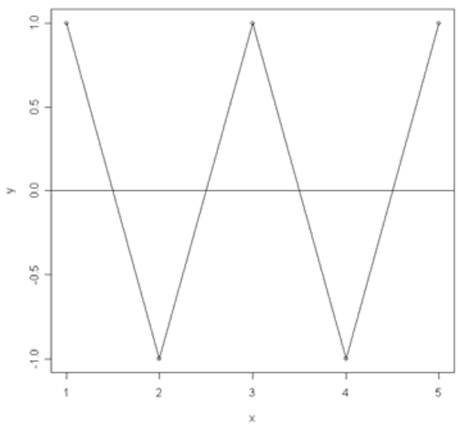
Der Datensatz, auf den diese Funktion angewendet werden soll, enthält die beiden Variablen *a* und *b*, die die folgenden Ausprägungen haben:

i	a_i	b_i
1	1	1
2	2	-1
3	3	1
4	4	-1
5	5	1

a) Mit welchen Befehlen können Sie *a* und *b* in R eingeben? *(2 Punkte)*
- `> a <- c(1,2,3,4,5)` oder `a <- seq(1:5)`
  (statt <− kann auch = verwendet werden)
- `> b <- c(1,-1,1,-1,1)`

b) Welchen R-Befehl müssen Sie eingeben, um die Funktion auszuführen und welchen, um Änderungen an der Funktion vorzunehmen? *(2 Punkte)*
- `> my.function(a,b)`
- `> fix(my. function)`

c) Geben Sie **alle** Ausgaben an, die mit dieser Funktion für den Datensatz erzeugt werden. *(6 Punkte)*




```
[1] 0
```

Bepunktung:
- *- Punktepaare zeichnen*
- *- Punktepaare mit Linie verbinden*

- *- Null-Linie einzeichnen*
- *- Korrelationskoeffizient (=0) ausgeben*

Aufgabe 5: ***(10 Punkte)***

Welche Antwort ist richtig? Bitte kreuzen Sie die zutreffende Antwort an. Zu jeder Frage gibt es nur eine richtige Antwort. Für jede korrekt angekreuzte Antwort gibt es 1 Punkt, für jede falsch angekreuzte Antwort wird 1 Punkt abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

1.	Welchen R-Befehl können Sie **nicht** verwenden, um die Summe der quadrierten Werte einer Variable *x* zu generieren?	
	☐	`> sum(x^2)`
	☒	`> I(sum(x)^2)`
	☐	`> I(sum(I(x^2)))`
2.	Mit welchem R-Befehl berechnen Sie die auf den Wert 1 einer Dummy-Variablen *dumvar* bedingten Mittelwerte einer Variable *x*?	
	☐	`> mean(x(dumvar==1))`
	☐	`> mean(x{dumvar==1})`
	☒	`> mean(x[dumvar==1])`
3.	Welchen R-Befehl kann man zur Berechnung des Stichproben-Korrelationskoeffizienten heranziehen?	
	☐	`> cor(x,y)/sqrt((var(x)*var(y)))`
	☒	`> cov(x,y)/sqrt((var(x)*var(y)))`
	☐	`> var(x,y)/sqrt((cov(x)*cov(y)))`
4.	Mit welchem R-Befehl kann man die Fehlertermvarianz eines vorab geschätzten und als Objekt `kq.mod` vorliegenden Regressionsmodells generieren?	
	☐	`> summary(kq.mod)$sigma2`
	☐	`> summary(kq.mod$sigma^2)`
	☒	`> summary(kq.mod)$sigma^2`
5.	Welches Objekt generiert man mit dem R-Befehl `seq(1:100)`?	
	☒	Vektor, der die Zahlen 1 bis 100 enthält
	☐	Vektor, der eine Sequenz mit dem Bruch 1/100 enthält
	☐	Vektor mit 100 Einträgen der Ziffer 1
6.	Mit welchem R-Befehl kann man eine Matrixmultiplikation der beiden Matrizen A und B durchführen?	
	☐	`> A $%$ B`
	☒	`> A %*% B`
	☐	`> A &%& B`
7.	Welche Größe wird mit folgendem Ausdruck berechnet (`n`: Anzahl der Beobachtungen, `res`: Objekt mit vorhergesagten Residuen): `sum((res[2:n] - res[1:n-1])^2)/sum(res^2)`	
	☐	Autokorrelationskoeffizient
	☒	Durbin-Watson Teststatistik
	☐	Lagrange-Multiplier Teststatistik
8.	Welchen R-Befehl kann man verwenden, um die Anzahl der Beobachtungen in einem Vektor *x* zu ermitteln?	

	☐	> count(x)
	☐	> number(x)
	☒	> length(x)
9.	Welchen Größe wird mit einem R-Befehl `as.numeric(var==2)` erzeugt?	
	☒	Dummyvariable, die 1 ist, wenn die Variable `var` den Wert 2 annimmt, 0 sonst
	☐	Kategoriale Variable `var`, die den Wert 2 für alle numerischen Größen annimmt
	☐	Numerische Größe, die die Variable `var` beim Wert 2 auf TRUE setzt
10.	Mit welchem R-Befehl bestimmt man **nicht** den kritischen Wert einer F-Verteilung mit 3 und 40 Freiheitsgraden bei einem Signifikanzniveau von 5%?	
	☐	`> qf(.05,3,40,lower.tail=F)`
	☐	`> qf(.95,3,40,lower.tail=T)`
	☒	`> qf(.95,3,40,lower.tail=F)`

Aufgabe 6: ***(24 Punkte)***

Wahr oder falsch? Tragen Sie für jede der folgenden Aussagen ein „w“ für „wahr“ oder ein „f“ für „falsch“ ein. Für jede richtige Antwort gibt es 1 Punkt, für jede falsche Antwort wird 1 Punkt abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

F	Der Ausdruck $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ kann nur für stetige Zufallsvariablen berechnet werden.
W	Eine Vorhersage auf Basis eines linearen Modells ist genau dann unverzerrt, wenn der Erwartungswert des Vorhersagefehlers 0 beträgt.
F	Grundidee des KQ Schätzers ist es, eine Linie so durch eine Punktwolke zu legen, dass die Summe der vertikalen Abweichungen der beobachteten Werte von der Linie minimiert wird.
F	Je schwächer die Konzentration eines Merkmals, umso größer der Abstand zwischen Lorenzkurve und Diagonale.
W	Auf der Hauptdiagonale der Varianz-Kovarianz Matrix der geschätzten Parameter befinden sich ausschließlich die Varianzen der einzelnen Parameter.
W	Konsistente Schätzer können verzerrt sein.
F	Bei einem Wert der Jarque-Bera-Teststatistik von 0 wird die Nullhypothese, dass die Störterme normalverteilt sind, verworfen.
W	Die Hypothese positiver Autokorrelation erster Ordnung kann mittels eines t-Tests getestet werden.
F	Werden logarithmierte erklärende Variablen genutzt, so muss der geschätzte Steigungsparameter negativ sein.
F	Bei einem Wert der Durbin-Watson Teststatistik von kleiner als 0,5 würde man vermuten, dass negative Autokorrelation erster Ordnung vorliegt.
W	Der Herfindahl-Index ist ein Maß absoluter Konzentration, das als gewichtete mittlere Steigung der Konzentrationskurve interpretiert werden kann.
F	Der p-Wert beschreibt, wie wahrscheinlich eine Ausprägung der Teststatistik ist, die kleiner ist als der empirisch beobachtete Wert.
F	Bei Autokorrelation sind die KQ-Intervallschätzer im Gegensatz zu den Punktschätzern BLUE.
W	Das Verfahren gleitender Durchschnitte kann zur Trendbereinigung verwendet werden.

F	Der F-Test kann nicht verwendet werden, um nur eine einzige Restriktion zu testen.
W	Das angepasste R^2 berücksichtigt die Zahl der geschätzten Steigungsparameter.
F	Eine Variable *z* kann als Instrument für eine stochastische erklärende Variable *x* genutzt werden, wenn ihre Standardabweichung nicht mit *x* korreliert.
F	Nimmt eine Zufallsvariable den Wert x = 2 an, so ist der an dieser Stelle berechnete Wert der Wahrscheinlichkeitsfunktion größer als 2.
W	Die Normalverteilung ist symmetrisch.
F	Im linearen Modell gibt die Regressionskonstante den Mittelwert der abhängigen Variablen an.
F	Wenn Dummyvariablen als erklärende Variablen im linearen Modell betrachtet werden, muss die Konstante entfallen.
W	Mit dem Lagrange-Multiplier Test kann getestet werden, ob der Störterm eines Modells durch einen autoregressiven Prozess zweiter Ordnung determiniert wird.
F	Zur Durchführung eines Instrumentvariablenschätzverfahrens benötigt man eine erklärende Variable, die mit keiner anderen erklärenden Variablen korreliert ist.
W	Multikollinearitätsprobleme lassen sich über eine Erhöhung der Beobachtungszahl reduzieren.

Aufgabe 7: ***(11 Punkte)***

Welche Antwort ist richtig? Bitte kreuzen Sie die zutreffende Antwort an. Zu jeder Frage gibt es nur eine richtige Antwort. Für jede korrekt angekreuzte Antwort gibt es 1 Punkt, für jede falsch angekreuzte Antwort wird 1 Punkt abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

1.	Ein Parameter β_1 im einfachen linearen Modell wird umso präziser geschätzt,	
	☒	je größer die Streuung in x ist.
	❑	je kleiner der t-Wert von β_1 ist.
	❑	je größer die Streuung von y ist.
2.	Der Kleinstquadratschätzer	
	❑	sollte nur bei normalverteilten abhängigen Variablen verwendet werden.
	❑	sollte nur bei normalverteilten Koeffizienten verwendet werden.
	☒	kann auch bei nicht normalverteilten Residuen das Gauss-Markov-Theorem erfüllen.
3.	Welche Interpretation gilt für β_1 im Modell für die Einfachregression $\ln y = \beta_0 + \beta_1 \cdot \ln x + e$?	
	❑	bei einer Änderung von x um eine Einheit ändert sich y um β_1 Einheiten.
	❑	bei einer Änderung von x um eine Einheit ändert sich y um β_1 Prozent.
	☒	bei einer Änderung von x um ein Prozent ändert sich y um β_1 Prozent.
4.	Ein RESET Test ergibt eine Teststatistik von 4,8 mit einem p-Wert von 0,067. Dies bedeutet:	
	❑	Das Modell sollte in logarithmierter Form geschätzt werden.
	❑	Am Signifikanzniveau von 10% wird H_0 nicht verworfen.
	☒	Am Signifikanzniveau von 5% ist das Modell nicht fehlspezifiziert.
5.	Bei einem Hypothesentest ist die Ablehnungsregion	
	❑	umso größer, je kleiner das Signifikanzniveau α.

	❑	unabhängig von der Typ-I Fehlerwahrscheinlichkeit.
	☒	abhängig von der Anzahl der Beobachtungen.
6.	Bei AR(1) Störtermen mit $\rho = 0{,}30$	
	❑	liegt negative Autokorrelation vor.
	❑	ist die Kovarianz zwischen zeitlich benachbarten Störtermen 0,30.
	☒	beträgt die Korrelation corr $(e_t, e_{t-2}) = 0{,}09$.
7.	Der Lagrange-Multiplier-Test auf Autokorrelation	
	❑	ist nur bei Autokorrelation erster Ordnung verwendbar.
	❑	führt bei verzögerten abhängigen Variablen unter den Regressoren zu verzerrten Ergebnissen.
	☒	kann als F-Test durchgeführt werden.
8.	Der Two-Stage-Least-Squares-Schätzer	
	❑	schätzt das gleiche lineare Regressionsmodell zweimal.
	☒	nutzt vorhergesagte Werte auf der zweiten Stufe.
	❑	berücksichtigt ein Polynom zweiter Ordnung der erklärenden Variable.
9.	Für die Durchführung eines Chow Tests	
	❑	werden Interaktionsterme und ihr quadrierter Wert benötigt.
	❑	müssen die Störterme standardnormalverteilt sein.
	☒	kann man die Stichprobe in zwei Teile teilen.
10.	Ausgelassene Variablen	
	❑	sollte man vermeiden, indem man möglichst viele Variablen in das Modell aufnimmt.
	☒	können zu verzerrten Parameterschätzern führen.
	❑	führen zu einer höheren Varianz der geschätzten Parameter.
11.	Der R^2-Wert einer Schätzung ist umso geringer,	
	❑	je größer der Achsenabschnitt geschätzt wird.
	❑	je geringer die erklärte Variation der erklärenden Variablen.
	☒	je höher die unerklärte Variation der abhängigen Variable.