Aufgabe 1: **[20 Punkte]**

Sie führen eine Regression der Arbeitsnachfrage (*labour*, Anzahl der Beschäftigten) auf die erklärenden Variablen Einkommen (*wage*, pro-Kopf-Löhne in 1000 €), Produktionsleistung (*output*, in Mio. €) und Anlagekapital (*capital*, in Mio. €) mit Daten für 569 belgische Unternehmen durch und erhalten folgenden Regressionsoutput.

```
Call:
lm(formula = labour ~ wage + output + capital)

Coefficients:
              Estimate   Std. Error
(Intercept)   287.7186      19.6418
wage           -6.7419       0.5014
output         15.4005       0.3556
capital        -4.5905       2.7000
---

Residual standard error: ? on 565 degrees of freedom
Multiple R-Squared: ?
F-statistic:  2716 on 3 and 565 DF,  p-value: < 2.2e-16
```

a) Überprüfen Sie, ob die geschätzten Koeffizienten für *wage*, *output* und *capital* statistisch signifikant sind und erläutern Sie am Beispiel des Koeffizienten für *wage* ausführlich Ihre Vorgehensweise. Geben Sie hierzu H_0, H_1, die genaue Teststatistik, die Freiheitsgrade und Ihre Schlusslogik an. (4 Punkte)

- wage signifikant am 5% Niveau
- output signifikant am 5% Niveau
- capital signifikant am 10% Niveau *(Teilpunkte für: capital nicht signifikant am 5% Niveau)*

- $H_0 : \beta_{wage} = 0$; $H_1 : \beta_{wage} \neq 0$
- $t = \dfrac{b_{wage} - 0}{se\left(b_{wage}\right)} = \dfrac{-6.7419}{0.5014} = -13.44615$
- $t_{krit,df=565,5\%} = 1.96$ (Freiheitsgrade = T - Anzahl der geschätzten Parameter = 596 - 4 = 565)
- Schlusslogik: Wenn | t | > t_{krit}, dann verwerfe H_0.

  Hier: 13.45 > 1.96 → H_0 verwerfen.

b) Interpretieren Sie die Steigungsparameter inhaltlich. (3 Punkte)

- b_{wage}: Einkommen/Kopf + 1000 € reduziert A^N um 6.7 Personen.
- b_{output}: Wertschöpfung + 1 Mio. € erhöht A^N um 15.4 Personen.
- $b_{capital}$: Anlagekapital + 1 Mio. € reduziert A^N um 4.6 Personen.

c) In R wird für das Modell der folgende ANOVA-Output ausgegeben. Berechnen Sie aufgrund dieser Informationen das R^2 für obige Regression. Wie hoch ist der Wert der Standardfehler der Residuen? (4 P)

```
Analysis of Variance Table

Response: labour
              Df      Sum Sq     Mean Sq    F value      Pr(>F)
wage           1    14330208    14330208     587.16   < 2.2e-16 ***
output         1   177494090   177494090    7269.63   < 2.2e-16 ***
capital        1     7110022     7110022     291.28   < 2.2e-16 ***
Residuals    565    13795027       24416
```

- $R^2 = \dfrac{14330208 + 177494090 + 7110022}{14330208 + 177494090 + 7110022 + 13795027} = 0.9352$
- *oder:* $R^2 = 1 - \dfrac{13795027}{14330208 + 177494090 + 7110022 + 13795027} = 0.9352$
- $\sigma = \sqrt{24416} = 156.256$
- *oder:* $\sigma = \sqrt{13795027 / 565} = 156.256$

d) Sie wollen testen, ob es ausreicht, nur *output* als erklärende Variable zu berücksichtigen. Führen Sie den Test mit den Informationen aus folgender ANOVA-Tabelle durch. Geben Sie hierzu H_0 und H_1 an und erläutern Sie kurz Ihre Vorgehensweise. (2 Punkte)

```
Analysis of Variance Table

Model 1: labour ~ output
Model 2: labour ~ wage + output + capital

   Res.Df         RSS    Df    Sum of Sq        F       Pr(>F)
1     567    23473922
2     565    13795027     2      9678895   198.21   < 2.2e-16
```

- $H_0 : \beta_{wage} = \beta_{capital} = 0$
  $H_1 : \beta_{wage} \neq 0 \;\; od. \;\; \beta_{capital} \neq 0$
- Vorgehensweise: F-Test auf gemeinsame Signifikanz

  Vergleich F_{emp} mit F_{krit}: $F_{emp} = 198.21$; $F_{krit, 5\%, 2, 565} = 3.00$

  Da $| F_{emp} | > | F_{krit} |$ verwerfe H_0.

e) Ein einseitiger Goldfeld-Quandt Test wird in R durchgeführt und liefert das folgende Ergebnis:

```
Goldfeld-Quandt test

data:  mod.kq
GQ = 15.0927, df1 = 281, df2 = 280, p-value < 2.2e-16
```

Wie ist die Teststatistik unter der Nullhypothese verteilt? Geben Sie ein Signifikanzniveau, den kritischen Wert und die Testentscheidung an. Interpretieren Sie auch den p-Wert. (4 Punkte)

- Teststatistik ist F-verteilt.
- $F_{krit, 5\%, 281, 280} = 1.00$ ($= F_{krit, 5\%, \infty, \infty}$)
- Da 15.09 > 1, verwerfe H_0, dass $\sigma_1^2 = \sigma_2^2$.

- p-Wert: Nur 2.2*e^{-16} % der Werte einer F-verteilten Zufallsvariable liegen rechts von F_{emp}. Daher sehr ungewöhnlicher Wert unter H_0. → H_0 verwerfen.

f) Was besagen die Ergebnisse aus Teilaufgabe e) für die Eigenschaften des vorliegenden KQ-Schätzers? (3 Punkte)

- falsche Standardfehler
- ineffizient
- nicht BLUE

Aufgabe 2: **[19 Punkte]**

a) Wie lässt sich Autokorrelation grafisch erkennen? **(2 Punkte)**

- Plot der KQ - Residuen gegen die Zeit (auch Zeichnung)
- bei positiver AK folgen überdurchschnittlich viele Residuen mit gleichem Vorzeichen direkt aufeinander

b) Ihnen liegt mit $e_t = \rho e_{t-1} + \upsilon_t$ und $\upsilon_t = \theta\upsilon_{t-1} + u_t$ ein so genannter AR(2) Prozess vor, wobei $u_t \sim \left(0, \sigma_u^2\right)$. Zeigen Sie, dass gilt $e_t = \left(\rho + \theta\right) e_{t-1} - \theta\rho e_{t-2} + u_t$. **(3 Punkte)**

- Einsetzen von $\upsilon_t = \theta\upsilon_{t-1} + u_t$ in $e_t = \rho e_{t-1} + \upsilon_t$:

$$e_t = \rho e_{t-1} + \theta\upsilon_{t-1} + u_t$$

- $e_t = \rho e_{t-1} + \upsilon_t$ ‚verzögern', mit θ multiplizieren und nach $\theta\upsilon_{t-1}$ auflösen:

$$\theta\upsilon_{t-1} = \theta e_{t-1} - \theta\rho e_{t-2}$$

- Einsetzen ergibt $e_t = \rho e_{t-1} + \theta e_{t-1} - \theta\rho e_{t-2} + u_t = \left(\rho + \theta\right) e_{t-1} - \theta\rho e_{t-2} + u_t$

c) Mit Quartalsdaten zum Bergbau in den USA von 1972, Quartal 1 bis 1996, Quartal 4 wird eine Funktion der Stromnachfrage (*POW*, in Mio. US-$) geschätzt: $ln\left(POW_t\right) = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 ln\left(PRO_t\right) + e_t$. Hierbei ist *PRO* die Produktivität (in Mio. US-$) und *t* ein Zeittrend. Eine KQ-Schätzung mit R liefert folgende Ergebnisse: **(11 Punkte)**

```
Call:
lm(formula = log(POW) ~ t + t2 + log(PRO))

Coefficients:
             Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  -16.1200      7.7670  -2.075    0.0407 *
t              0.6137      0.0514  11.939   < 2e-16 ***
t2            -0.0239      0.0051  -4.686  7.34e-06 ***
log(PRO)       0.9588      0.1012   9.474  1.98e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0378 on 96 degrees of freedom
Multiple R-Squared: 0.935,      Adjusted R-squared: 0.933
F-statistic: 460.3 on 3 and 96 DF,  p-value: < 2.2e-16
```

c1) Interpretieren Sie die Steigungsparameter inhaltlich und statistisch und berechnen Sie den Effekt von *t* auf die abhängige Variable im 10. Quartal.

- Alle Steigungsparameter (t, t^2, log(PRO)) sind statistisch signifikant.
- Die Stromnachfrage hat einen konkaven Verlauf über die Zeit.
- Steigt die Produktivität um ein Prozent, so steigt die Stromnachfrage um 0.96%.
- Marginaler Effekt: $\frac{\partial \log(POW)}{\partial t} = 0.6137 - 2 \cdot 0.0239 \cdot t$
- Der Effekt des Zeittrends auf log(POW) beträgt im 10. Quartal (also $t = 10$) 0,1357.

c2) Sie vermuten, dass ein AR(1) Prozess vorliegt. Aus der Schätzung ergibt sich, dass $\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1} = 0.10905$. Berechnen Sie einen approximativen Schätzwert für $\hat{\rho}$.

- $\hat{\rho}_{approx.} = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2}$; zunächst ist also $\sum_{t=1}^{T} \hat{e}_t^2$ zu berechnen:
- $\sum_{t=1}^{T} \hat{e}_t^2 = 0.0378^2 \cdot 96 = 0.1372$
- $\hat{\rho}_{approx.} = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} = \frac{0.10905}{0.1372} = 0.795$

c3) Verwenden Sie das Ergebnis aus Teilaufgabe c2 und berechnen Sie die approximative Durbin-Watson Teststatistik *d*. Testen Sie am 5%-Signifikanzniveau, ob Autokorrelation vorliegt (Falls Sie in c2 kein Ergebnis ermittelt haben, zeigen Sie den Rechenweg und verwenden Sie *d* = 0.4). Geben Sie dazu auch die kritischen Werte der Teststatistik an.

- $d = 2 \cdot (1 - \hat{\rho}) = 2 \cdot (1 - 0.795) = 0.41$.
- K = 4; N = 100; $d_L^* = 1.61$; $d_U^* = 1.74$
- Wenn $d < 1.61$, dann positiver AR(1) Prozess; Hier: H_0: $\rho = 0$ verwerfen.

d) Betrachten Sie ein einfaches lineares Modell $y_t = \beta_0 + \beta_1 x_t + e_t$. Sie vermuten, dass ein AR(2) Prozess vorliegt: $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \upsilon_t$. Welches ist ein geeigneter Test auf Autokorrelation? Beschreiben Sie ausführlich seine Vorgehensweise. (3 Punkte)

- LM-Test
- KQ schätzen, verzögerte Residuen $\hat{e}_{t-1}$, $\hat{e}_{t-2}$ ermitteln
- verzögerte Residuen als erklärende Größen in Modell aufnehmen:

$y_t = \beta_0 + \beta_1 x_t + \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \upsilon_t$; F-Test auf gemeinsame Signifikanz von $\hat{\rho}_1$ und $\hat{\rho}_2$; wenn signifikant, kann AR(2) nicht verworfen werden

Aufgabe 3: **[19 Punkte]**

Sie möchten herausfinden, wie sich das Einschulungsalter auf den Lernerfolg von Schülern auswirkt. Anhand einer Untersuchung der Leistungen im Rechnen von 50 Grundschülern betrachten Sie den Einfluss folgender Faktoren auf die im Test erzielten Punkte:

EA:	Alter zum Zeitpunkt der Einschulung in Monaten
Ek:	Einkommen der Eltern
M:	Geschlecht (Mädchen=1, Junge=0)
Mi:	Schüler kommt aus einer Migrantenfamilie (ja = 1, nein = 0)
TV:	durchschnittlicher täglicher TV-Konsum in Stunden
B:	Zahl der Bücher im Haushalt der Eltern

Sie formulieren folgendes Modell:

$$Punkte_t = \beta_1 + \beta_2 * EA_t + \beta_3 * \ln(Ek_t) + \beta_4 * B_t + \beta_5 * TV_t^2 + \beta_6 * M_t + \beta_7 * Mi_t + \beta_8 * (M_t * Mi_t) + e_t$$

Die Auswertung der Daten mit R ergibt folgenden Output:

```
Call:
lm(formula = Punkte ~ EA + log(Ek) + B + TV^2 + M + Mi + I(M * Mi))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  55.0709     5.1487  10.696 1.45e-13 ?
EA            0.1461     0.1092   1.339 0.187885 ?
log(Ek)       6.5471     1.7624   3.715 0.000594 ?
B             0.6120     0.4223   1.449 0.154713 ?
TV^2         -2.5340     1.5374  -1.648 0.106763 ?
M             1.7346     0.2723   6.370 1.16e-07 ?
Mi           -4.5591     0.5176  -8.808 4.30e-11 ?
I(M * Mi)     0.3415     0.6892   0.496 0.622819 ?
---
Signif. codes:  ?

Residual standard error: 0.7773 on 42 degrees of freedom
Multiple R-Squared: 0.8553,     Adjusted R-squared: 0.8312
F-statistic: 35.46 on 7 and 42 DF,  p-value: 1.201e-15
```

a) Betrachten sie die p-Werte: Welche Koeffizienten sind am 5%-Niveau signifikant? (2 Punkte)

- Konstante; log(Einkommen); M und Mi
- *oder:* b_1 , b_3 , b_6 *und* b_7

b) Wie groß ist unter den Kindern aus Migrantenfamilien der zu erwartende Punktunterschied zwischen Jungen und Mädchen, die sonst in allen Merkmalen übereinstimmen? (2 Punkte)

- $b_6 \cdot 1 + b_8 \cdot 1 = 1.7346 + 0.3415 = 2.0761$

c) Ist der Effekt aus b) bei einem Signifikanzniveau von α=0.05 signifikant verschieden von Null (Hinweis: cov(b_6, b_7)=0.054; cov(b_6, b_8)= –0.070; cov(b_7, b_8)= –0.256)? Erläutern Sie Ihre Vorgehensweise. (5 Punkte)

- $H_0: \beta_6 + \beta_8 = 0$
  $H_1: \beta_6 + \beta_8 \neq 0$
- $t = \dfrac{b_6 + b_8}{se(b_6 + b_8)} = \dfrac{2.0761}{\left[V(b_6) + V(b_8) + 2 \cdot Cov(b_6, b_8)\right]^{1/2}}$
- $= \dfrac{2.0761}{\left(0.2723^2 + 0.6892^2 + 2 \cdot -0.070\right)^{1/2}} = ... = \dfrac{2.0761}{0.6396} = 3.2$
- t_{krit} bei 42 df nicht gegeben; $t_{krit,\,5\%,\,45} = 2.014 < 3.2$ → H_0 verwerfen

d) Wie groß ist der marginale Effekt einer Stunde TV-Konsums für Kinder, die pro Tag 1 Stunde bzw. 4 Stunden TV schauen? (2 Punkte)

- $\dfrac{\partial\, Punkte}{\partial\, TV} = \dfrac{\partial\, Punkte}{\partial\, TV^2} \cdot \dfrac{\partial\, TV^2}{\partial\, TV} = \beta_5 \cdot 2 \cdot TV$
- $-2.534 \cdot 2 \cdot 1 = -5.068$
- $-2.534 \cdot 2 \cdot 4 = -20.272$

e) Sie vermuten, das Einschulungsalter könnte endogen sein und möchten daher den Geburtsmonat (*GMon*) als Instrument nutzen. Unter welchen Bedingungen wäre der Geburtsmonat ein valides Instrument? Halten Sie diese Bedingungen für erfüllt? Begründen Sie Ihre Einschätzung.(4 Punkte)

- i) korreliert mit EA

  ii) ist unkorreliert mit Störterm

  ad i) wahrscheinlich erfüllt, da GMon EA determiniert

  ad ii) nicht sicher erfüllt: Bei gegebenem EA könnte GMon wg. körperlicher Entwicklung noch eine Rolle spielen

f) Nehmen Sie an, die notwendigen Bedingungen seien erfüllt. Sie erhalten folgenden R-Output:

```
2SLS Estimates

Model Formula: Punkte ~ EA + log(Ek) + B + TV^2 + M + Mi + I(M * Mi)

Instruments: ~GMon + log(Ek) + B + TV^2 + M + Mi + I(M * Mi)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  57.2009     6.7217   8.510 1.100e-10
EA            0.6967     0.4399   1.584 1.207e-01
log(Ek)       6.4955     2.2337   2.908 5.790e-03
B             0.6466     0.5358   1.207 2.343e-01
TV^2         -5.1716     2.7923  -1.852 7.105e-02
M             1.9666     0.3873   5.077 8.277e-06
Mi           -4.6372     0.6586  -7.041 1.270e-08
I(M * Mi)     0.4068     0.8748   0.465 6.444e-01

Residual standard error: 0.985 on 42 degrees of freedom
```

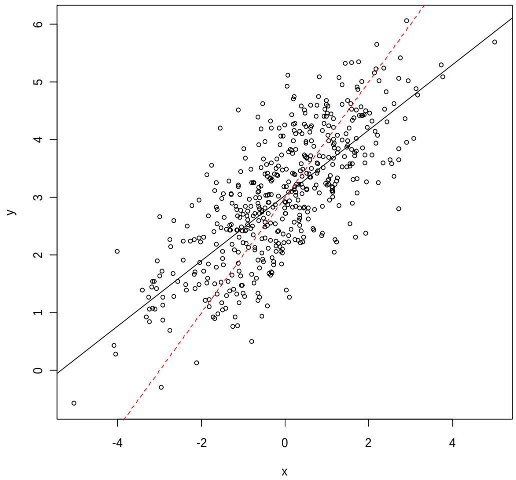
Was können Sie aus dem Vergleich der Koeffizienten über die Korrelationsrichtung zwischen Einschulungsalter und Störterm schließen? Erläutern Sie Ihr Ergebnis. (4 Punkte)

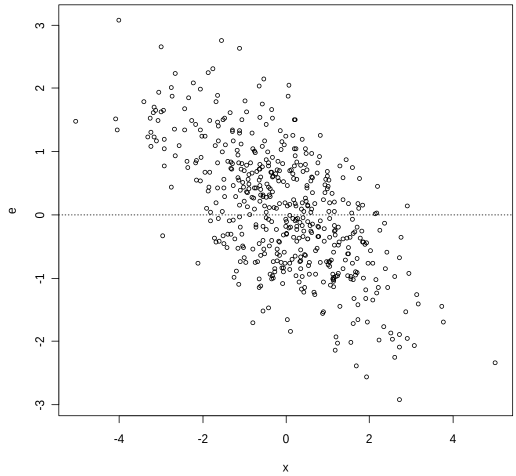
- KQ: $b_2 = 0.1461$

2SLS: $b_2 = 0.6967$

→ Wahrer Zusammenhang steiler als KQ (& Punktwolke)

- grafische Veranschaulichung (nicht erwartet):





- Grafik links (KQ – schwarze Gerade; rot-gestrichelt: ‚wahrer' Zusammenhang)
- Grafik rechts (Plot von x (also EA) und e)

- Steilere Steigung bei korrigierter Schätzung legt nahe, dass bei hohen EA-Werten negative Störterme auftreten und umgekehrt → corr(e, EA) < 0

- oder:

$$\hat{\beta} = \frac{cov(X,Y)}{Var(X)} - \frac{cov(X,e)}{Var(X)}$$

Wenn $\hat{\beta}^{KQ} = \frac{cov(X,Y)}{Var(X)} < \hat{\beta}^{2SLS}$, dann muss $cov(X,e) < 0$.

Aufgabe 4: **[12 Punkte]**

In R wurde folgende Funktion programmiert:

```
my.plot <- function(x)
 {
  a <- seq(x:5*x)
  b <- seq(x,10*x,by=2)
  plot(a,b)
  abline(v=mean(a),h=mean(b))
  return(sum(a),sum(b))
}
```

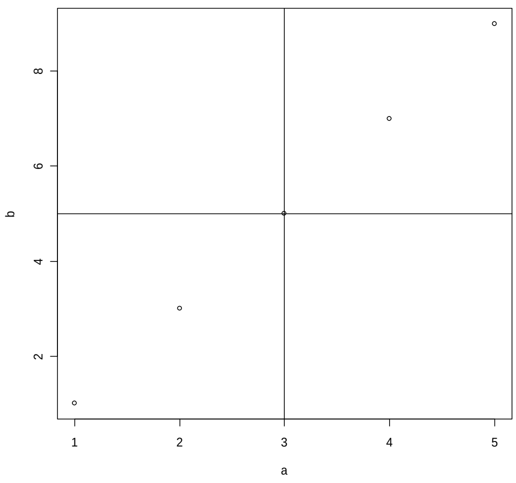
a) Welchen R-Befehl müssen Sie eingeben, um die Funktion auszuführen und welchen, um Änderungen an der Funktion vorzunehmen? **(2 Punkte)**

- `my.plot(x)`

```
- fix(my.plot)
```

b) Stellen Sie alle Ausgaben so dar, wie sie mit dieser Funktion für x = 1 erzeugt werden. (8 Punkte)

Ausgaben der Funktion:




```
[1] 15
[1] 25
```

c) Geben Sie eine zusätzliche Befehlszeile an, mit der die Mittelwerte der Vektoren ausgegeben werden.

```
- return(mean(a),mean(b))
```

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.	Mit welchem R-Befehl berechnet man den Korrelationskoeffizienten zwischen zwei Vektoren X und Y?	
	☐	`> kor(X,Y)`
	☒	`> cor(X,Y)`
	☐	`> corr(X,Y)`

2.	Mit welcher Option kontrolliert man in R im RESET Test für quadrierte und kubische Werte der vorhergesagten Werte für Y?	
	☐	`> , power(2-3)`
	☐	`> , Power((2&3)=T)`
	☒	`> , power(2:3)`
3.	Welcher der folgenden R-Befehle ist nicht zum Aktivieren des Pakets `tseries` geeignet?	
	☐	`> library(tseries)`

	☒	`> package(tseries)`
	❑	`> library(package=tseries)`

4.	Mit welchem R-Befehl kontrolliert man für die Interaktion zwischen der Dummyvariable d und der kontinuierlichen Variable x ?	
	☒	`> lm(y ~ x + d + I(x*d))`
	❑	`> lm(y ~ x + d + I(xd))`
	❑	`> lm(y ~ x + d + x%*%d)`

5.	Welche Kennzahl berechnet man mit folgender Formel: `sum((X-mean(X))*(Y-mean(Y)))/(length(X)-1)`?	
	❑	Konfidenzintervall zwischen X und Y
	☒	Kovarianz zwischen X und Y
	❑	Korrelationskoeffizient zwischen X und Y

6.	Mit welchem R-Befehl bestimmt man den kritischen Wert einer t-Verteilung mit 38 Freiheitsgraden bei einem Signifikanzniveau von 1%?	
	❑	`> pt(0.01,38,lower.tail=F)`
	☒	`> qt(0.01,38,lower.tail=F)`
	❑	`> dt(0.99,38,lower.tail=F)`

7.	Mit welchem der folgenden R-Befehle kann man nicht die Residuen eines vorher geschätzten Modells `mod.kq` auslesen?	
	❑	`> resid(mod.kq)`
	❑	`> mod.kq$residuals`
	☒	`> mod.kq[resid]`

8.	Welchen R-Befehl müssen Sie anwenden, um die Funktion `my_function.R` einzulesen?	
	❑	`> get.source(“path/my_function.R“)`
	☒	`> source(“path/my_function.R“)`
	❑	`> get.code(“path/my_function.R“)`

9.	Wie lautet der R-Befehl, um eine KQ-Schätzung für Beobachtungen durchzuführen, für die eine Dummyvariable d den Wert 1 annimmt?	
	❑	`> lm(y(d==1) ~ x(d==1))`
	❑	`> lm(y[d=1] ~ x[d=1])`
	☒	`> lm(y[d==1] ~ x[d==1])`

10.	Welche Größe wird ausgegeben, wenn der Befehl `s$coef[1]/s$coef[1,2]` aufgerufen wird (s ist der Modelloutput `summary(...)`).	
	☒	t-Wert der Konstante
	❑	p-Wert der Konstante
	❑	Standardfehler der Konstante

Aufgabe 6 **[25 Punkte]**

Wahr oder falsch? …

W	Je stärker die Konzentration eines Merkmals, umso größer der Abstand zwischen Lorenzkurve und Diagonale.
W	Die H_0: cov(x,e) = 0 kann mittels eines t-Tests getestet werden.
F	Heteroskedastie führt zu verzerrten Schätzern für den Achsenabschnittsparameter.
F	Um eine saisonbereinigte Zeitreihe zu erstellen, können lineare, exponentielle oder logistische Saisonmodelle genutzt werden.
F	Eine Variable *z* kann als Instrument für eine stochastische erklärende Variable genutzt werden, wenn ihre Standardabweichung nicht mit *x* korreliert.
W	Konsistente Schätzer können verzerrt sein.
F	Bei einem Gini-Koeffizienten von 0.23 ist die Konzentration der Verteilung höher als bei einem Gini-Koeffizienten von 0.67.
F	Bei einem Wert der Jarque Bera Teststatistik von 0 wird die Nullhypothese, dass die Störterme normalverteilt sind, verworfen.
F	Der White Schätzer stellt eine approximative Korrektur für Situationen mit stochastischer Fehlertermvarianz dar.
F	Bei linear homogenen Indizes ändern sich die Werte nicht, wenn Güter in einer anderen Maßeinheit gemessen werden.
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 positiv sein.
F	Bei einem Wert der Durbin-Watson Statistik von 2 würde man vermuten, dass negative Autokorrelation erster Ordnung vorliegt.
W	Kollinearität unter erklärenden Variablen führt zu reduzierter Präzision von Schätzungen.
W	Der Herfindahl Index ist ein Maß absoluter Konzentration, das als gewichtete mittlere Steigung der Konzentrationskurve interpretiert werden kann.
F	Beim einseitigen t-Test liegt die Ablehnungsregion im Bereich positiver t-Werte.
W	Der p-Wert beschreibt, wie wahrscheinlich eine Ausprägung der Teststatistik jenseits des empirisch beobachteten Wertes ist.
F	Die F-Verteilung beschreibt quadrierte, χ^2-verteilte Zufallsvariablen.
F	Das Produkt von Mengen- und Preisindex entspricht bei Laspeyres- und Paasche-Index der Umsatzmesszahl.
W	Die Varianz von Parameterschätzern im einfachen Regressionsmodell ist umso kleiner, je breiter die erklärende Variable gestreut ist.
F	Um einen Chow Test durchzuführen, sind zwei Schätzungen erforderlich.
F	Wenn es 3 Merkmalsträger gibt, ist die Konzentrationsquote K_4 informativ.

F	Unter Heteroskedastie können bessere Vorhersagen gemacht werden, als ohne Heteroskedastie.
F	Wenn Autokorrelation vorliegt, sind die KQ-Schätzer nicht mehr BLUE, aber die Intervallschätzer können noch verwendet werden.
W	Das Verfahren gleitender Durchschnitte kann zur Trendbereinigung verwendet werden.

Aufgabe 7 **[15 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.	Der Steigungsparameter einer einfachen Regressionsschätzung, in der die abhängige Variable linear und die erklärende Variable logarithmisch kodiert ist, beschreibt	
	☐	um wie viel Prozent y steigt, wenn sich x um eine Einheit ändert;
	☒	um wie viele Einheiten y steigt, wenn sich x um ein Prozent ändert;
	☐	wie hoch die Elastizität von y hinsichtlich x ist.

2.	Das Auslassen relevanter erklärender Variablen führt zu	
	☐	stets verzerrten KQ-Schätzern für die Steigungsparameter des Modells;
	☐	falsch ausgewiesenen Standardfehlern im KQ-Schätzer;
	☒	keinen Konsequenzen, wenn die ausgelassene Variable nicht mit den berücksichtigten Variablen des Modells korreliert ist.

3.	Der Typ II Fehler	
	☐	tritt auf, wenn die Nullhypothese verworfen wird, obwohl sie zutrifft;
	☐	ist umso wahrscheinlicher, je größer die Stichprobe ist;
	☒	wird unwahrscheinlicher, wenn der Typ I Fehler wahrscheinlicher wird.

4.	Intervallschätzer sind	
	☒	informativer als Punktschätzer;
	☐	nicht auf Basis von Stichproben interpretierbar;
	☐	umso verlässlicher, je kleiner der geschätzte Parameterwert ist.

5.	Unter den Standardannahmen gilt für den Vorhersagefehler im einfachen linearen Regressionsmodell:	
	☒	Er hat einen Erwartungswert von 0;
	☐	Er hat eine von T unabhängige Varianz;
	☐	Er ist umso kleiner, je näher sich die Beobachtung am Mittelwert der erklärenden Variablen befindet.

6.	Eine Division der abhängigen Variablen durch 1000 führt zu	
	☐	einem um den Faktor 1000 erhöhten Achsenabschnittsparameter;
	☐	einem nicht veränderten Steigungsparameter;

	☒	um den Faktor 1000 reduzierten Werten für alle Achsen- und Steigungsparameter.

7.	Der R^2 Wert einer Schätzung ist umso höher	
	☒	je geringer der Anteil der unerklärten Variation der abhängigen Variable;
	❑	je höher der Anteil der erklärten Variation der erklärenden Variablen;
	❑	je höher die Kovarianz des Störterms mit der erklärenden Variable.

8.	Der Test auf Gesamtsignifikanz eines Modells	
	☒	nutzt die F-Statistik;
	❑	kann ohne Achsenabschnittsparameter nicht durchgeführt werden;
	❑	kann als RESET Test durchgeführt werden.

9.	Eine hohe Varianz geschätzter Parameter	
	❑	kann durch Auslassen von Beobachtungen gesenkt werden;
	☒	kann durch Berücksichtigung externer Information im Regressionsmodells reduziert werden;
	❑	führt zu hohen t-Werten.

10.	Bei einem Test auf positive Autokorrelation am 5 Prozent Niveau, mit 20 Beobachtungen und 3 geschätzten Parametern bedeutet eine *d*-Statistik in Höhe von 1.4,	
	❑	dass die Nullhypothese verworfen wird;
	❑	dass die Nullhypothese nicht verworfen werden kann;
	☒	dass keine Aussage möglich ist.

11.	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.

12.	Bei Messfehlern in den erklärenden Variablen	
	❑	ist die Präzision der Schätzung reduziert;
	☒	sind die Parameterschätzer inkonsistent;
	❑	sollten generalisierte Kleinstquadrateschätzer verwendet werden.

13.	Der Langrange 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.

14.	Bei einem t-Test der Nullhypothese $\beta \geq k$ am 5 Prozent Signifikanzniveau	
	☒	muss die Nullhypothese verworfen werden, wenn $p < 0.05$;
	❑	ist die t-Verteilung von der Stichprobengröße unabhängig;
	❑	muss die Nullhypothese verworfen werden, wenn $p < 0.10$.

15.	Bei AR(1) Störtermen mit $\rho = 0.30$	
	❑	liegt negative Autokorrelation vor;
	❑	ist die Kovarianz zwischen zeitlich benachbarten Störtermen 0.30;
	☒	beträgt der Korrelationskoeffizient corr (e_t, e_{t-2}) = 0.09.