

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 ,
 - a2) den Standardfehler von b_0 sowie
 - a3) b_4 .
- b) Berechnen Sie das multiple R^2 mit Hilfe der Informationen aus dem Regressionsoutput. (3 Punkte)
- c) Berechnen Sie den marginalen Effekt (4,5 Punkte)
 - c1) des Jahreseinkommens für eine Person im Alter von 36 Jahren,
 - c2) des Alters, wenn eine Person 36 Jahre alt ist und das Jahreseinkommen 90000 € beträgt.
- 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)
- e) Sie sind nicht sicher, ob Sie vergessen haben, aussagekräftige Variablen in Ihr Modell aufzunehmen. Sie machen einen Test und erhalten folgenden Regressionsoutput: (7 Punkte)

```
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
L^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.
- 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?

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)
- b) Sie interessieren sich für Unterschiede zwischen städtischen und ländlichen Kreisen. (4 Punkte)
- b1) Wie könnten Sie testen, ob sich die erwartete Erwerbsbeteiligung in einem ländlichen und einem städtischen Kreis mit jeweils identischer Kinderzahl unterscheidet?
- 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.
- 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.
- c2) Welche formale Annahme wäre verletzt?
- 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 ? Erläutern Sie Ihre Vermutung.
- 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?
- d2) Nennen Sie eine empirische Maßzahl, die Sie heranziehen können, um die Qualität dieses Instruments zu prüfen. Gibt es eine Anforderung an ein Instrument, deren Erfüllung Sie nicht prüfen können?
- d3) Argumentieren Sie *kurz* verbal, ob eine der in d1) genannten Bedingungen in diesem Beispiel verletzt sein könnte.
- 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, nennen Sie die konkret verwendete Null- und Alternativhypothese und treffen Sie eine Entscheidung anhand des folgenden Outputs:

(6 Punkte)

Hausman Test				
Model Formula: Erwerb ~ Kinder + Land + vhat				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	73.633241	4.879251	15.091	0.412
Kinder	-0.128279	0.047284	-2.713	5.71e-12
Land	-0.132536	0.030376	-4.309	3.62e-05
vhat	-0.036455	0.011524	-3.163	4.14e-03

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

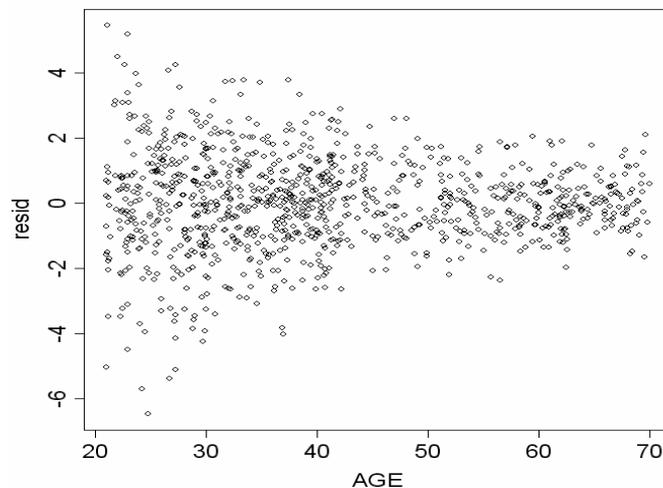
Sie führen folgende Regression durch: $JOB_t = \beta_0 + \beta_1 \cdot ABI_t + \beta_2 \cdot UNI_t + \beta_3 \cdot AGE_t + e_t$

Als Schätzwert 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)
- 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)

- 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)*
- 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.
- d2) Zeigen Sie algebraisch, dass die Transformation ihr Ziel erreicht hat.
- e) Welche Alternative zum GLS Verfahren kennen Sie, um mit Heteroskedastie umzugehen? Erläutern Sie kurz das Vorgehen. *(2 Punkte)*

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)*
- b) Welchen R-Befehl müssen Sie eingeben, um die Funktion auszuführen und welchen, um Änderungen an der Funktion vorzunehmen? *(2 Punkte)*
- c) Geben Sie **alle** Ausgaben an, die mit dieser Funktion für den Datensatz erzeugt werden. *(6 Punkte)*

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

9.	Welchen Größe wird mit einem R-Befehl <code>as.numeric(var==2)</code> erzeugt?	
	<input type="checkbox"/>	Dummyvariable, die 1 ist, wenn die Variable <code>var</code> den Wert 2 annimmt, 0 sonst
	<input type="checkbox"/>	Kategoriale Variable <code>var</code> , die den Wert 2 für alle numerischen Größen annimmt
	<input type="checkbox"/>	Numerische Größe, die die Variable <code>var</code> 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%?	
	<input type="checkbox"/>	<code>> qf(.05, 3, 40, lower.tail=F)</code>
	<input type="checkbox"/>	<code>> qf(.95, 3, 40, lower.tail=T)</code>
	<input type="checkbox"/>	<code>> qf(.95, 3, 40, lower.tail=F)</code>

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.

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

	Eine Variable z kann als Instrument für eine stochastische erklärende Variable x genutzt werden, wenn ihre Standardabweichung nicht mit x korreliert.
	Nimmt eine Zufallsvariable den Wert $x = 2$ an, so ist der an dieser Stelle berechnete Wert der Wahrscheinlichkeitsfunktion größer als 2.
	Die Normalverteilung ist symmetrisch.
	Im linearen Modell gibt die Regressionskonstante den Mittelwert der abhängigen Variablen an.
	Wenn Dummyvariablen im linearen Modell betrachtet werden, muss die Konstante entfallen.
	Mit dem Lagrange-Multiplier Test kann getestet werden, ob der Störterm eines Modells durch einen autoregressiven Prozess zweiter Ordnung determiniert wird.
	Zur Durchführung eines Instrumentvariablenschätzverfahrens benötigt man eine erklärende Variable, die mit keiner anderen erklärenden Variablen korreliert ist.
	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,
	<input type="checkbox"/> je größer die Streuung in x ist.
	<input type="checkbox"/> je kleiner der t-Wert von β_1 ist.
	<input type="checkbox"/> je größer die Streuung von y ist.
2.	Der Kleinstquadratschätzer
	<input type="checkbox"/> sollte nur bei normalverteilten abhängigen Variablen verwendet werden.
	<input type="checkbox"/> sollte nur bei normalverteilten Koeffizienten verwendet werden.
	<input type="checkbox"/> kann auch bei nicht normalverteilten Residuen das Gauss-Markov-Theorem erfüllen.
3.	Welche Interpretation gilt für β_1 im Modell für die Einfachregression $\ln y = \beta_0 + \beta_1 \cdot \ln x + e$?
	<input type="checkbox"/> bei einer Änderung von x um eine Einheit ändert sich y um β_1 Einheiten.
	<input type="checkbox"/> bei einer Änderung von x um eine Einheit ändert sich y um β_1 Prozent.
	<input type="checkbox"/> 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:
	<input type="checkbox"/> Das Modell sollte in logarithmierter Form geschätzt werden.
	<input type="checkbox"/> Am Signifikanzniveau von 10% wird H_0 nicht verworfen.
	<input type="checkbox"/> Am Signifikanzniveau von 5% ist das Modell nicht fehlspezifiziert.
5.	Bei einem Hypothesentest ist die Ablehnungsregion
	<input type="checkbox"/> umso größer, je kleiner das Signifikanzniveau α .
	<input type="checkbox"/> unabhängig von der Typ-I Fehlerwahrscheinlichkeit.
	<input type="checkbox"/> abhängig von der Anzahl der Beobachtungen.

6.	Bei AR(1) Störtermen mit $\rho = 0,30$
	<input type="checkbox"/> liegt negative Autokorrelation vor.
	<input type="checkbox"/> ist die Kovarianz zwischen zeitlich benachbarten Störtermen 0,30.
7.	Der Lagrange-Multiplier-Test auf Autokorrelation
	<input type="checkbox"/> ist nur bei Autokorrelation erster Ordnung verwendbar.
	<input type="checkbox"/> führt bei verzögerten abhängigen Variablen unter den Regressoren zu verzerrten Ergebnissen.
8.	Der Two-Stage-Least-Squares-Schätzer
	<input type="checkbox"/> schätzt das gleiche lineare Regressionsmodell zweimal.
	<input type="checkbox"/> nutzt vorhergesagte Werte auf der zweiten Stufe.
9.	Für die Durchführung eines Chow Tests
	<input type="checkbox"/> werden Interaktionsterme und ihr quadrierter Wert benötigt.
	<input type="checkbox"/> müssen die Störterme standardnormalverteilt sein.
10.	Ausgelassene Variablen
	<input type="checkbox"/> sollte man vermeiden, indem man möglichst viele Variablen in das Modell aufnimmt.
	<input type="checkbox"/> können zu verzerrten Parameterschätzern führen.
11.	Der R^2 -Wert einer Schätzung ist umso geringer,
	<input type="checkbox"/> je größer der Achsenabschnitt geschätzt wird.
	<input type="checkbox"/> je geringer die erklärte Variation der erklärenden Variablen.
	<input type="checkbox"/> je höher die unerklärte Variation der abhängigen Variable.