

Erlaubte Hilfsmittel: Tabelle der statistischen Verteilungen, 4 DIN A4-Seiten eigene Notizen, Taschenrechner, Fremdwörterbuch.

Bearbeitungszeit: 90 Minuten

Die Klausur besteht aus 5 Aufgaben, in denen insgesamt 90 Punkte erworben werden können. Die Punktzahl je Aufgabe, bzw. je Teilaufgabe, ist in eckigen Klammern angegeben und entspricht der für die Aufgabe aufzuwendenden Zeit in Minuten.

Das Team der Abteilung Statistik und Ökonometrie wünscht Ihnen viel Erfolg!

Klausurergebnis:

Aufgabe	1	2	3	4	5
Punkte:					

Aufgabe 1:

[34]

Als wissenschaftlicher Mitarbeiter eines Schlafinstituts versuchen Sie schon seit längerem, geeignete Determinanten zu finden, die menschliches Schlafverhalten erklären können. Nach jahrelanger Forschung und vielen schlaflosen Nächten sind Sie auf folgende Gleichung gekommen:

$$SLEEP = \beta_1 + \beta_2 * TOTWRK + \beta_3 * EDUC + \beta_4 * AGE + \beta_5 * MARR$$

Um Ihre Gleichung schätzen zu können, nutzen Sie amerikanische Querschnittsdaten aus dem Jahre 1990 mit 532 Beobachtungen. Die Variablen sind wie folgt definiert:

$SLEEP_i$	Schlaf pro Woche einer Person i , gemessen in Minuten
$TOTWRK_i$	Totale Arbeitszeit pro Woche der Person i , gemessen in Minuten
$EDUC_i$	Ausbildung der Person i , gemessen in Anzahl der Schuljahre
AGE_i	Alter der Person i , gemessen in Jahren
$MARR_i$	Dummy-Variable mit Ausprägung 1, wenn Person i verheiratet, sonst 0
$MALE_i$	Dummy-Variable mit Ausprägung 1, wenn Person i männlich, sonst 0

Gesamthalt führen Sie drei Schätzungen durch. Die erste (Tabelle 1) beinhaltet sämtliche 532 Beobachtungen. Da Sie der Überzeugung sind, dass das Schlafverhalten zwischen Männern und Frauen jedoch grundsätzlich anders ist, schätzen Sie erneut nur für Männer (Tabelle 2) und danach nur für Frauen (Tabelle 3).

- Interpretieren Sie den Regressionsparameter b_5 der Tabelle 1 statistisch und inhaltlich. [2]
- Testen Sie $H_0 : \beta_2 - \beta_3 + \beta_4 = 10$ für Tabelle 1 auf dem 5%-Signifikanzniveau zweiseitig. Stellen Sie die Alternativhypothese auf. Sie wissen, dass $Cov(b_2, b_3) = 0$, $Cov(b_2, b_4) = 0$ und $Cov(b_3, b_4) = 0$. Geben Sie die korrekte Anzahl der Freiheitsgrade an, verwenden Sie jedoch $DF = \infty$. [5]
- Die berechnete Durbin-Watson Statistik hat den Wert 1.917767 (Tabelle 1). Testen Sie $H_0 : \rho \geq 0$ auf dem 5% Signifikanzniveau und erläutern Sie Ihre Vorgehensweise detailliert. Erläutern Sie die Alternativhypothese in Worten. [4]
- Sie vermuten, dass Sie entscheidende Variablen ausgelassen haben. Deshalb schätzen Sie

Ihr Modell erneut, diesmal zusätzlich mit den prognostizierten Werten der abhängigen Variablen in zweiter und dritter Potenz. Warum kann das nützlich sein? Die nun berechnete Fehlerquadratsumme ist 85774437. Welchen Schluss können Sie daraus ziehen? Erläutern Sie Ihr Vorgehen (verwenden Sie Tabelle 1). [5]

- e) Sie glauben weiter, dass der Einfluss der Arbeitsstunden auf das Schlafbedürfnis mit der Ausbildung des Arbeitnehmers variiert. Wie können Sie diese Hypothese testen? [3]
- f) Führen Sie einen Chow-Test bei einem Signifikanzniveau von 5% durch, um zu prüfen, ob es signifikante Unterschiede in den Regressionsergebnissen für Männer und Frauen gibt. Zu welchem Ergebnis kommen Sie? Wie sieht das unrestringierte Modell formal aus? [8]
- g) Sie befürchten, dass die Variable *TOTWRK* mit dem Störterm korreliert. Was hätte das für Auswirkungen? Sie führen einen Hausman Test mit Hilfe von Instrument-Variablen durch. Erklären Sie genau Ihre Vorgehensweise. Welche Bedingungen müssen die Instrumente erfüllen? [7]

Tabelle 1:

Dependent Variable: SLEEP Method: Least Squares				
Sample (adjusted): 1 532: Included observations: 532 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3609.8	138.3	26.1012292	0
TOTWRK	-0.15	0.2	-0.75	0
EDUC	-10	7.2	-1.3888889	0.1651
AGE	1.5	1.6	0.9375	0.539
MARR	64.96	41.9	1.550358	0.0583
R-squared	0.113715	Mean dependent var		3259.466
Adjusted R-squared	0.10529	S.D. dependent var		430.9729
S.E. of regression	407.6534	Akaike info criterion		14.86993
Sum squared resid	87411354	Schwarz criterion		14.91816
Log likelihood	-3949.4	F-statistic		13.4977
Durbin-Watson stat	1.917767	Prob(F-statistic)		0

Tabelle 2:

Dependent Variable: SLEEP Method: Least Squares				
Sample: 1 532 IF MALE=1: Included observations: 273				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3740.654	199.2318	18.77538	0
TOTWRK	-0.193726	0.029113	-6.654264	0
EDUC	-9.834978	9.557494	-1.029033	0.3043
AGE	1.396035	2.299914	0.606994	0.5443
MARR	56.52929	71.36427	0.792123	0.4289
R-squared	0.144577	Mean dependent var		3225.939
Adjusted R-squared	0.129675	S.D. dependent var		429.4095
S.E. of regression	400.6014	Akaike info criterion		14.84408
Sum squared resid	46058189	Schwarz criterion		14.91944
Log likelihood	-2168.657	F-statistic		9.70134
Durbin-Watson stat	1.800945	Prob(F-statistic)		0

Tabelle 3:

Dependent Variable: SLEEP Method: Least Squares				
Sample: 1 532 IF MALE=0: Included observations: 259				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3579.708	203.2932	17.6086	0
TOTWRK	-0.119563	0.030968	-3.860808	0.0001
EDUC	-10.32447	10.98434	-0.939926	0.3482
AGE	0.200901	2.485346	0.080834	0.9356
MARR	70.98971	63.54732	1.117116	0.2651
R-squared	0.077318	Mean dependent var		3300.569
Adjusted R-squared	0.057518	S.D. dependent var		430.2185
S.E. of regression	417.6626	Akaike info criterion		14.93201
Sum squared resid	40645002	Schwarz criterion		15.01928
Log likelihood	-1778.375	F-statistic		3.904933
Durbin-Watson stat	1.950652	Prob(F-statistic)		0.002033

Aufgabe 2:**[21]**

In einem einfachen Regressionsmodell $y_t = \beta_1 + \beta_2 x_t + e_t$ hat die Varianz des Störterms folgende Struktur:

$$\sigma_{e_t}^2 = \sigma^2 x_t^3$$

- Welche Auswirkungen hat dies auf die Eigenschaften des Kleinstquadrateschätzers? [2]
- Wie würden Sie vorgehen, um eine Schätzung mit homoskedastischen Fehlern zu erhalten? Erläutern Sie Ihre Vorgehensweise und zeigen Sie, dass der in Ihrem Verfahren resultierende Störterm homoskedastisch ist. [4]
- Gibt es ein alternatives Verfahren, das Problem anzugehen? Erläutern Sie dieses verbal und formal. [4]
- Erläutern Sie detailliert, wie man am Beispiel der oben genannten Störtermvarianz formal auf Heteroskedastie testen kann. [6]
- Wie lautet Ihr Schluss am 5 Prozent Signifikanzniveau für einen einseitigen Test der Homoskedastie als Nullhypothese, wenn die Teststatistik für den Test aus Aufgabe (d) den Wert 2,08 annimmt und Sie mit insgesamt 26 Beobachtungen gearbeitet haben? [2]
- Sie erfahren nachträglich, dass die Varianz des Störterms die Form $\sigma_{e_t}^2 = \sigma^2 x_t^2$ annimmt. Was bedeutet dies für Ihre Schätzergebnisse aus Teilaufgabe (b)? [3]

Aufgabe 3:**[5]**

Zeigen Sie, dass folgender Zusammenhang zwischen der Teststatistik des F-Tests (F) und dem Gütemass R^2 gilt:

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (T - k)}$$

wobei k die Anzahl der geschätzten Steigungsparameter und T die Anzahl der Beobachtungen angibt.

Aufgabe 4:**[15]**

Wahr oder Falsch? Tragen Sie für zutreffende Aussagen den Buchstaben w (für wahr), für nicht zutreffende f (für falsch) ein.

(Für jede richtige Antwort gibt es 0,75 Punkte, für jede falsche Antwort werden 0,75 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.)

	Die Chi-Quadrat Verteilung ist eine einparametrische Verteilung.
	Bei immer grösser werdenden Stichproben tendiert die Varianz der mit dem Kleinstquadratverfahren geschätzten Parameter gegen Null.
	Das Auslassen relevanter erklärender Variablen führt in manchen, aber nicht in allen Fällen zu verzerrten Koeffizientenschätzungen für die berücksichtigten erklärenden Variablen.
	Wenn die Standardfehler von Koeffizienten unterschätzt werden, sind die geschätzten Konfidenzintervalle der Koeffizienten zu breit.
	Bei gegebener Typ I Fehlerwahrscheinlichkeit (α) ist die Wahrscheinlichkeit eines Typ II Fehlers positiv mit der Stichprobengrösse korreliert.
	Regressiert man eine logarithmierte abhängige Variable ($\ln Y$) auf eine logarithmierte erklärende Variable ($\ln X$), so gibt der Koeffizient an, um wieviel Prozent sich Y ändert, wenn X um eine Einheit steigt.
	Das Gauss-Markov Theorem gilt nur für kleine Stichproben.
	Der Kleinstquadrateschätzer ist umso unverzerrter, je grösser die Stichprobe.

	Heteroskedastie führt zu ineffizienten aber unverzerrten Schätzergebnissen.
	Im Rahmen des "Two stage least squares" Verfahren wird im zweiten Schritt die abhängige Variable auf den vorhergesagten Wert einer oder mehrerer erklärender Variablen regressiert.
	Der Hausman Test vergleicht die Ergebnisse verschiedener Schätzverfahren.
	Um nichtlineare Zusammenhänge zwischen erklärenden und abhängigen Variablen abzubilden, muss mehr als eine erklärende Variable im Modell sein.
	Der Jarque-Bera Test prüft, ob eine Zufallsvariable standardnormalverteilt ist.
	Der k-te Moment einer Zufallsvariable (Y) entspricht dem Erwartungswert von Y^k .
	Der Durbin-Watson Test ist nicht geeignet, um auf Autokorrelation zu testen, wenn der Störterm von mehr als einem seiner verzögerten Werte bestimmt wird.
	Es gibt Situationen in denen der Durbin-Watson Test nicht zu einem klaren Testergebnis führt.
	Der Goldfeld-Quandt Test prüft genau wie der Chow Test, ob zwei Teilstichproben mit dem gleichen Modell geschätzt werden sollten.
	Solange im Modell eine Konstante berücksichtigt ist, werden die Koeffizienten kategorischer erklärender Variablen hinsichtlich einer Referenzkategorie interpretiert.
	Das angepasste R^2 sinkt, wenn die gleiche Erklärungsgüte mit weniger geschätzten Parametern erzielt werden kann.
	Bei Messfehlern in den erklärenden Variablen ist der Kleinstquadrateschätzer inkonsistent.

Aufgabe 5:

[15]

Sind folgende Aussagen richtig? Erläutern Sie **stichwortartig** Ihre Auffassung (Bsp.: "Stimmt, weil..." bzw. "Stimmt nicht, weil..."). Nur bei korrekter Begründung wird die Antwort mit 1,5 Punkten pro Frage honoriert.

	Durch die Berücksichtigung externer Informationen im Rahmen eines restricted least squares Schätzers kann sich die Qualität der Schätzung verschlechtern.
	Der Lagrange Multiplier Test auf Autokorrelation ist nur bei Autokorrelation erster Ordnung anwendbar.
	Monte Carlo Simulationen nutzen vorliegende Datensätze, um zu überprüfen ob vorgegebene Beziehungen in den Daten gelten.

	Die Varianz des Vorhersagefehlers bei Vorhersagen auf Basis eines linearen Modells hängt nicht von der Kovarianz zwischen geschätzter Regressionskonstante und Steigungsparameter ab.
	Der Hausman Test kann in der Form eines F-Tests durchgeführt werden.
	Darstellungen der Residuen erlauben Rückschlüsse auf mögliche Autokorrelation, aber nicht auf mögliche Heteroskedastie.
	Beim RESET Test wird der vorhergesagte Wert des Störterms als zusätzlicher Regressor im Modell verwendet.
	Die Auswirkungen von Autokorrelation und Heteroskedastie auf den Kleinstquadrateschätzer sind gleich.
	Die Nullhypothese $\rho \leq 0$ wird im Durbin-Watson Test verworfen, wenn die Durbin-Watson Teststatistik grösser ist als $4 - d_L$, wobei d_L den unteren kritischen Wert des Durbin-Watson Tests angibt.
	Der Wert der geschätzten Koeffizienten ändert sich nicht, wenn alle Variablen des Modells (abhängige und unabhängige) durch einen Faktor 1'000 dividiert werden.