Tentamen Econometrie 1, 28 januari 2004, 14.00-16.00 uur

Dit tentamen duurt 2 uur!

De uitslag komt uiterlijk vanaf 19 januari 2004 op Blackboard. Desgewenst kunt u daarna uw werk inzien bij de docent.

Vul bovenaan op ieder antwoordenblad de gevraagde gegevens in. Bij dit tentamen mag u gebruik maken van een zelfgemaakt formuleblad op A4-formaat. U mag ook de uitgereikte tabellen gebruiken (graag na afloop weer inleveren, samen met uw uitwerkingen).

In elk van de opgaven dient u ervan uit te gaan dat er voldaan is aan de basisveronderstellingen van het lineaire regressiemodel. Eventuele afwijkingen blijken uit de opgaven. Tussen haakjes worden onder de regressiecoëfficiënten hun standaardfouten vermeld. Voer eventuele toetsen uit met een significantieniveau van **5%**.

MOTIVEER UW ANTWOORDEN!

Opgave 1 (a=13, b=10, c=12, d=10, e=10, f=10, totaal 65)

Beschouw het model $y = Xb + e$ (in matrixnotatie), waarbij

$$y = \begin{bmatrix} 18 \\ 20 \\ 18 \\ 16 \\ 19 \end{bmatrix}, \quad X = \begin{bmatrix} 5 & 6 \\ 7 & 6 \\ 3 & 8 \\ 9 & 3 \\ 6 & 7 \end{bmatrix}.$$

a) Bereken de OLS schatter van b en de standaardfout van de regressie.

b) Laat zien dat onder de Gauss-Markov aannamen geldt $V[e \mid X] = s^2 I_n$, waarbij s^2 de variantie van de storingsterm is en n het aantal waarnemingen. Bewijs dat tevens $V[e] = s^2 I_n$ door gebruik te maken van $V[e] = E_X[V[e \mid X]] + V_X[E[e \mid X]]$.

c) Gegeven is $x_* = \begin{bmatrix} 4 & 5 \end{bmatrix}'$. Bepaal een puntvoorspelling van $y_* = (x_*)'b + e_*$. Bepaal tevens een 95% voorspellingsinterval voor y_*.

d) Bereken de OLS schatter van b onder de restrictie $b_2 = 2$.

e) Stel men doet OLS onder de (mogelijk foutieve) restrictie $b_2 = 0$. Leidt dan de vertekening af van de schatter van b_1 (conditioneel op X). Reken deze vertekening uit als in werkelijkheid geldt $b_2 = 2$.

f) Stel dat in werkelijkheid $b_2 = 0$. Bereken de procentuele toename van de standaardafwijking van de schatter van b_1, als men OLS zonder restricties zou toepassen in plaats van OLS onder de restrictie $b_2 = 0$.

Opgave 2 (a=7, b=10, c=8, d=10, totaal 35)

Bij hart katheterisatie wordt een katheter van 3 millimeter diameter in een slagader bij de lies ingebracht en verschoven tot in het hart. Vervolgens kan de katheter worden bewogen om informatie te verkrijgen over de fysiologie en het functioneren van het hart. Deze procedure wordt soms gevolgd bij kinderen met aangeboren hart afwijkingen. De arts moet daarbij de juiste lengte van de katheter van te voren inschatten.
Bij een kleine studie van 12 kinderen werd de juiste lengte van de katheter (*Y* in centimeters) gemeten en tevens werden de lengte (*L* in inches) en het gewicht (*G* in Engelse ponden) van het kind gemeten om te zien of deze gebruikt kunnen worden voor het voorspellen van de juiste katheter lengte. Hier onder worden enkele resultaten gegeven.

Correlaties

	Log *Y*	Log *L*	Log *G*
Log *Y*	1	0.90	0.91
Log *L*	0.90	1	0.97
Log *G*	0.91	0.97	1

Model 1: $$\widehat{\log Y_i} = \underset{(0.99)}{2.65} - \underset{(0.42)}{0.0458} \log L_i + \underset{(0.17)}{0.315} \log G_i \qquad R^2 = 0.816$$

Model 2: $$\widehat{\log Y_i} = \underset{(0.46)}{1.03} + \underset{(0.13)}{0.692} \log L_i \qquad R^2 = 0.749$$

Model 3: $$\widehat{\log Y_i} = \underset{(0.16)}{2.55} + \underset{(0.045)}{0.297} \log G_i \qquad R^2 = 0.816$$

a) Onderzoek multicollineariteit in model 1 en de gevolgen daarvan.
b) Ga uit van model 1. Stel dat het gewicht *G* in feite slechts fungeert als proxy voor de omvang (het volume) van een kind (*V* in kubieke centimeters) en neem aan dat er een exacte lineaire relatie $V_i = l\, G_i$ is. Welke effect heeft het gebruik van deze proxy op de kwaliteit van de OLS-schatters? Definieer de dikte van een kind als $D_i = V_i / L_i$. Bepaal een schatting van het effect van L_i op Y_i en van het effect van D_i op Y_i en interpreteer deze effecten.
c) Stel dat men zou willen toetsen of in model 1 de standaardafwijking van de storingen samenhangt met het gewicht G_i. Leg precies uit hoe de toetsingsgrootheid van Goldfeld-Quandt berekend zou moeten worden en bepaal het kritieke gebied bij een significantieniveau van 10%.
d) Welk model verdient de voorkeur? Motiveer uw antwoord.