Standaarduitwerkingen tentamen V&O: Econometrie, 30 oktober 2008.

Vraag 1.
a). Cross-sectie gegevens hebben betrekking op waarnemingen (economische agenten volgens de syllabus) die worden gedaan op één moment. Dezelfde informatie over verschillende waarnemingen worden op een bepaald moment waargenomen. Tijdsreeksgegevens betreft informatie van gegevens (een economische agent) over een bepaalde periode.
b). De volgorde van cross-sectie gegevens doet er niet toe. 1000 waarnemingen uit de Nederlandse beroepsbevolking mogen best door elkaar gegooid worden. Als er een relatie bestaat tussen de waarnemingen (bijvoorbeeld een tijdsdimensie) dan wordt door het herschikken van de gegevens deze relatie verstoord, denk aan de waarneming van de inflatie over een bepaalde tijdsperiode. Bij tijdreeksgegevens kun je dus niet zomaar herschikken, er gaat dan mogelijk cruciale informatie verloren.

Vraag 2.
a). De variantie is een maat voor de spreiding van een variabele.
b). De correlatie is een tussen -1 en 1 geschaalde maat voor de mate van lineaire afhankelijkheid van twee variabelen.
c). $\operatorname{var}(y) = \operatorname{var}(2-3x) = \operatorname{var}(3x) = 3^2 \operatorname{var}(x) = 9\operatorname{var}(x)$ of

$$\operatorname{var}(y) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\sum(2-3x_i-(2-3\bar{x}))^2 = \frac{1}{n-1}\sum(3(x_i-\bar{x}))^2 = 9\operatorname{var}(x)$$

d). x en y hebben een volmaakt lineair verband. De correlatie is dus 1 of -1. Daar de richtingscoëfficiënt negatief is moet dat -1 zijn. Alternatief:

$$cor(x,y) = cor(x,2-3x) = \frac{\operatorname{cov}(x,2-3x)}{\sqrt{\operatorname{var}(x)}\sqrt{\operatorname{var}(2-3x)}} = \frac{\operatorname{cov}(x,2)-\operatorname{cov}(x,3x)}{\sqrt{\operatorname{var}(x)}\sqrt{\operatorname{var}(2-3x)}} =$$

$$\frac{-3\operatorname{cov}(x,x)}{\sqrt{\operatorname{var}(x)}\sqrt{9\operatorname{var}(x)}} = \frac{-3\operatorname{var}(x)}{3\operatorname{var}(x)} = -1$$

Vraag 3.
Je dient hier twee modellen met elkaar te vergelijken:
$y_i = \alpha + \beta x_i + \varepsilon_i$ en $y_i = \tilde{\alpha} + \tilde{\beta}\tilde{x}_i + \tilde{\varepsilon}_i$
a). De OLS-schatter van de constante is (p. 43 syllabus):
Voor model(1): $a = \bar{y} - b\bar{x}$. Voor model (2): $\tilde{a} = \bar{y} - \tilde{b}\bar{\tilde{x}} = \bar{y} - \tilde{b}\,\theta\bar{x}$.
b). De OLS-schatter voor de helling is (p. 43 syllabus):
Voor model (1):

$$b = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$$

.Voor model (2):

$$.\ \tilde{b} = \frac{\sum(\tilde{x}_i-\bar{\tilde{x}})(y_i-\bar{y})}{\sum(\tilde{x}_i-\bar{\tilde{x}})^2} = \frac{\sum(\theta x_i-\theta\bar{x})(y_i-\bar{y})}{\sum(\theta x_i-\theta\bar{x})^2} = \frac{\theta\sum(x_i-\bar{x})(y_i-\bar{y})}{\theta^2\sum(x_i-\bar{x})^2} = \frac{b}{\theta}$$

[Uit a). volgt nu dat: $\tilde{a} = a$]

c).R^2 meet de proportie verklaarde variantie. Die verandert natuurlijk niet bij schaling. Dit staat direct aangegeven op pagina 54 van de syllabus.

Vraag 4.

a). Het gaat hier om een log-lin verband en dus geldt voor het marginale effect:

$\frac{\partial w}{\partial x} = \frac{\partial}{\partial x}\exp(b_1 + b_2 LEDUC + b_3 IQ + b_4 FEMALE)$ waarbij x = LEDUC, IQ of FEMALE.

$$\frac{\partial w}{\partial x} = \exp(b_1 + b_2 LEDUC + b_3 IQ + b_4 FEMALE) b_x$$

of $\frac{\Delta w}{w} \approx b_x \Delta x$.

Een reële verandering in x met +1 levert een procentuele verandering van w op van b_x*100%.

Dus:

- Een niveau hoger opleidingsniveau (gegeven dat de andere verklarende variabelen gelijk blijven) geeft een hoger uurloon van 11.7%.
- Eén punt hoger IQ geeft een hoger uurloon van 0.6% (gelijkblijvende andere variabelen).
- Vrouwen verdienen 88.5% minder dan mannen (bij zelfde opleiding en IQ).

b). [0.623 +/- 2*0.130] = [0.363,0.883]

c). Gegeven een correcte specificatie van het model, geeft het betrouwbaarheidsinterval het interval weer waarin de werkelijke waarde van de (populatie)parameter met 95% kans ligt.

d). Een R^2 is een statistische maat en geeft geen causaliteit aan. Een R^2 van een geeft een 100% ‘statistische’ verklaring aan. Een R^2 van 0 wijst op volledige afwezigheid van ‘statistische’ verklaring van de gebruikte verklarende variabelen. De beoordeling van een model moet vooral geschieden vanuit de economische theorie. Statistiek is slechts een hulpmiddel hierbij.