

## Oefentoets V&O Econometrie

*Bij het maken van deze toets mag geen boek, syllabus of formulebriefje gebruikt worden. In totaal zijn er 100 punten te verdienen. Bij elke vraag staat tussen accolades het maximaal te behalen punten. Zorg voor een goede motivatie van je antwoorden.*

### Vraag 1

Geef voor de volgende kwalitatieve variabelen aan of het een nominale of ordinale variabele betreft:

- {5} seizoen
- {5} kredietwaardigheidsscore

### Vraag 2

Geef voor de volgende kwantitatieve variabelen aan of het een interval of ratio variabele betreft:

- {5} temperatuur
- {5} aantal jaren scholing

### Vraag 3

Stel we kwantificeren de variabele geslacht met 1 voor man en 2 voor vrouw.

Becommentarieer de volgende stellingen:

- {5} We hadden net zo goed 1000 voor man en 0 voor vrouw kunnen kiezen.
- {5} De codering impliceert dat twee mannen gelijk zijn aan één vrouw.

### Vraag 4

Stel we hebben de beschikking over  $n$  waarnemingen  $x_1, x_2, \dots, x_n$  en  $y_1, y_2, \dots, y_n$ . Er geldt dat bovendien dat  $y_i = a + bx_i$  voor alle  $i$  met  $a$  en  $b$  willekeurige positieve getallen. Voorts gelden de volgende definities:

Variantie van  $x$ : 
$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

Covariantie van  $x$  met  $y$ : 
$$\text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

Correlatiecoëfficiënt: 
$$\text{cor}(x, y) = \frac{\text{cov}(x, y)}{s_x s_y}.$$

Bewijs dat geldt:

- {10}  $\text{cov}(x, y) = b s_x^2$
- {10}  $s_y = b s_x$
- {5}  $\text{cor}(x, y) = 1$

### Vraag 5

{10} De variatiebreedte is gedefinieerd als het verschil tussen maximale en minimale waarde. Waarom is de variantie een betere spreidingsmaat dan de variatiebreedte?

### Vraag 6

{10} Het linear probability model is een lineair regressiemodel met een kwalitatieve (0-1) variabele als afhankelijke variabele. Leg uit wat precies gemodelleerd wordt. En wat is het probleem bij het schatten met de kleinste kwadratenmethode?

### Vraag 7

We willen de determinanten van economische groei analyseren. Cross-sectie gegevens zijn beschikbaar voor 104 landen over:

- gemiddelde jaarlijkse inkomensgroei 1960-1985 (G)
- logaritme van nationaal inkomen in 1960 (Y60)
- gemiddelde jaarlijkse investeringen als percentage van inkomen 1960-1985 (I)
- groei populatie 1960-1985 (P)

We willen onderzoeken of er convergentie in inkomen optreedt tussen landen. Met andere woorden, we willen de samenhang analyseren tussen initieel inkomen (Y60) en toekomstige groei (G). De volgende regressieresultaten zijn beschikbaar:

$$\hat{G}_i = 0.44 - 0.01 Y_i,$$

(0.37)      (0.05)

$$\hat{G}_i = -1.19 - 0.18 Y_i + 0.69 I_i - 0.44 P_i,$$

(0.69)      (0.04)      (0.08)      (0.26)

waarbij de getallen tussen haakjes geschatte standaardfouten zijn.

- {5} Geef in woorden een interpretatie aan de coefficient van Y in het eerste model.
- {5} Beschrijf het verschil in interpretatie van de coefficient van Y tussen beide modellen.
- {10} Een negatieve coefficient voor de regressor Y impliceert convergentie in inkomen. Construeer voor beide modellen een 95% betrouwbaarheidsinterval voor deze coefficient en bepaal voor elk model of de geschatte coefficient significant afwijkt van nul.
- {5} Welk model is het meest geschikt om de convergentie hypothese uit de vorige vraag mee te toetsen?