

## Uitwerkingen oefentoets V&O Econometrie

### Vraag 1

- Seizoen is een nominale variabele aangezien er geen hoeveelheidsinterpretatie is. Met andere woorden, de variabele klassificeert alleen.
- Kredietwaardigheidsscore is een ordinale variabele aangezien er nu sprake is van een rangorde. Een hogere score impliceert een minder risicovolle lening.

### Vraag 2

- Temperatuur is een interval variabele aangezien het nulpunt arbitrair gekozen is. Er is wel een verschilinterpretatie, maar geen hoeveelheidsinterpretatie. Bij 30 graden is het 15 graden warmer dan bij 15 graden, maar niet twee keer zo warm.
- Aantal jaren scholing is een ratio variabele aangezien het nulpunt niet arbitrair gekozen is. Twaalf jaar scholing is twee keer zes jaar scholing.

### Vraag 3

- Correct, de variabele klassificeert alleen maar, je hoeft er dus alleen voor te zorgen dat voor elke categorie een andere waarde is gekozen.
- Nee, er is geen verschilinterpretatie in de kwalitatieve variabele geslacht.

### Vraag 4

- We hebben  $y_i - \bar{y} = a + bx_i - (a + b\bar{x}) = b(x_i - \bar{x})$  dus

$$\text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})b(x_i - \bar{x}) = bs_x^2$$

- $s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (b(x_i - \bar{x}))^2 = b^2 s_x^2$  dus  $s_y = \sqrt{b^2 s_x^2} = bs_x$

- $\text{cor}(x, y) = \frac{bs_x^2}{s_x bs_x} = 1$

### Vraag 5

De variantie gebruikt alle waarnemingen en is daarom informatiever over de spreiding dan de variatiebreedte, die slechts twee waarnemingen gebruikt. Voorts is de variatiebreedte nog meer dan de variantie gevoelig voor afwijkende waarnemingen.

## Vraag 6

Met een lineair regressiemodel modelleren de voorwaardelijke verwachting van een afhankelijke variabele als lineaire functie van een aantal verklarende factoren. Wanneer de afhankelijke variabele binair (0-1) is is deze voorwaardelijke verwachting gelijk aan  $\Pr(Y=1)$  d.w.z. we modelleren een kans. Het probleem met kleinste kwadraten is dat voorspelde waarden kleiner dan 0 of groter dan 1 kunnen zijn.

## Vraag 7

- a. Wanneer initieel inkomen in 1960 1 procent lager is dan is verwachte jaarlijkse groei over 1960-1985 0.01% hoger.
- b. Verschil is dat in tweede model het een partieel effect betreft d.w.z. controlerend voor het effect van investeringen en populatie. Met andere woorden, wanneer initieel inkomen in 1960 1 procent lager is dan is verwachte jaarlijkse groei over 1960-1985 0.18% hoger gegeven dat investeringen en populatie niet veranderen.
- c. Betrouwbaarheidsintervallen zijn  $[-0.01-2*0.05, -0.01+2*0.05]=[-0.11, 0.09]$  en  $[-0.18-2*0.04, -0.18+2*0.04]=[-0.26, -0.10]$  respectievelijk. In het eerste model is de coefficient niet significant verschillend van nul, want deze waarde ligt in het 95% betrouwbaarheidsinterval. Daarentegen ligt nul niet in het interval van het tweede model, die coefficient is dus wel significant verschillend van nul.
- d. Het tweede model is meer geschikt want er is zeer waarschijnlijk in het eerste model sprake van weggelaten relevante factoren. Zoals te zien valt in het tweede model is de coefficient van in ieder geval de investeringsquote significant, dus weglating van deze relevante regressor zal leiden tot een onzuivere schatter. Voorts zijn betrouwbaarheidsintervallen gebaseerd op onzuivere schatter niet langer betrouwbaar.