

PROEFEXAMEN SOCIALE STATISTIEK

November 2010 – REEKS 1

Score .../15

Vraag 1 (3 punten)

Gegeven: een bivariate frequentietabel voor de variabelen x en y.
x en y worden als metrische variabelen beschouwd.

x \ y	1	2	3	4
1	0	3	0	0
2	3	4	1	0
3	0	2	5	4
4	1	0	2	6

Gevraagd: Bereken de covariantie tussen x en y.

Berekening

x	y	freq	x * freq	y * freq	x gem	y gem	x-xgem	y-ygem	(x-xgem)*(y-ygem)*freq
1	1	0	0	0	2,774194	2,83871	-1,77419	-1,83871	0
1	2	3	3	6	2,774194	2,83871	-1,77419	-0,83871	4,464099896
1	3	0	0	0	2,774194	2,83871	-1,77419	0,16129	0
1	4	1	1	4	2,774194	2,83871	-1,77419	1,16129	-2,060353798
2	1	3	6	3	2,774194	2,83871	-0,77419	-1,83871	4,270551509
2	2	4	8	8	2,774194	2,83871	-0,77419	-0,83871	2,597294485
2	3	2	4	6	2,774194	2,83871	-0,77419	0,16129	-0,249739854
2	4	0	0	0	2,774194	2,83871	-0,77419	1,16129	0
3	1	0	0	0	2,774194	2,83871	0,225806	-1,83871	0
3	2	1	3	2	2,774194	2,83871	0,225806	-0,83871	-0,189386056
3	3	5	15	15	2,774194	2,83871	0,225806	0,16129	0,182101977
3	4	2	6	8	2,774194	2,83871	0,225806	1,16129	0,524453694
4	1	0	0	0	2,774194	2,83871	1,225806	-1,83871	0
4	2	0	0	0	2,774194	2,83871	1,225806	-0,83871	0
4	3	4	16	12	2,774194	2,83871	1,225806	0,16129	0,790842872
4	4	6	24	24	2,774194	2,83871	1,225806	1,16129	8,541103018
		31	86	88					18,87096774
			2,774194	2,83871					0,629032258

Antwoord:

0,629

Vraag 2 (1,5 punten)

Het aantal gerealiseerde verschillen tussen observaties bij een categorisch kenmerk met 4 categorieën bedraagt 3500. De waarde van de index van kwalitatieve variatie bedraagt 0.933. Bepaal het aantal observaties waarbij het categorisch kenmerk werd gemeten.

Berekening

$$IKV = \frac{\text{aantal gerealiseerde verschillen}}{\frac{m(m-1)}{2} * \left(\frac{n}{m}\right)^2} = \frac{3500}{\frac{4(4-1)}{2} * \left(\frac{n}{4}\right)^2} = 0,933$$

$n?$

$$n = 4 * \sqrt{\frac{3500}{\frac{4(4-1)}{2} * 0,933}} = 100,0179 \approx 100$$

Antwoord:

100

Vraag 3 (1 punt)

Bereken:

$$\sum_{i=1}^4 \sum_{j=1}^3 i^j = 1^1 + 1^2 + 1^3 + 2^1 + 2^2 + 2^3 + 3^1 + 3^2 + 3^3 + 4^1 + 4^2 + 4^3 = 140$$

$$\begin{aligned} \sum_{i=1}^3 \sum_{j=2}^3 (5j+i) &= \sum_{i=1}^3 \sum_{j=2}^3 5j + \sum_{i=1}^3 \sum_{j=2}^3 i = 3 * \sum_{j=2}^3 5j + 2 * \sum_{i=1}^3 i = 3 * 5 * \sum_{j=2}^3 j + 2 * \sum_{i=1}^3 i \\ &= 15 * (2+3) + 2 * (1+2+3) = 87 \end{aligned}$$

Vraag 4 (2 punten)

In een frequentieverdeling is de oorspronkelijke waarde 36 na een z-transformatie gelijk aan 1 en de z-transformatie van de oorspronkelijke waarde 15 is gelijk aan -2. Bepaal het gemiddelde en de variantie van de verdeling.

Berekening

$$z_x = \frac{x_i - \bar{x}}{s_x} \quad 1 = \frac{36 - \bar{x}}{s_x} \quad -2 = \frac{15 - \bar{x}}{s_x}$$

$$s_x = 36 - \bar{x} \quad \text{en} \quad s_x = \frac{15 - \bar{x}}{-2} \quad \text{dus} \quad 36 - \bar{x} = \frac{15 - \bar{x}}{-2}$$

$$36 - \bar{x} = \frac{15 - \bar{x}}{-2} \quad \Leftrightarrow \quad -2 * (36 - \bar{x}) = 15 - \bar{x} \quad \Leftrightarrow \quad 2\bar{x} + \bar{x} = 15 + 72$$
$$\Leftrightarrow \quad \bar{x} = 29$$

$$36 - \bar{x} = s_x \quad \text{en} \quad \bar{x} = 29 \quad \text{dus} \quad 36 - 29 = s_x \quad \Leftrightarrow \quad 7 = s_x$$

$$\bar{x} = 29 \quad s_x = 7 \quad s_x^2 = 49$$

Antwoord:

Gemiddelde: 29

Variantie: 49

Vraag 5 (2,5 punten)

De examenresultaten van een vak (percentages) hebben een klokvormige verdeling met een gemiddeld percentage van 53. 95% van de studenten behaalde een resultaat in het interval [31;75]. Welke lineaire transformatie moet men uitvoeren op de oorspronkelijke resultaten zodat 95% van de studenten een resultaat behaalt in het interval [20;86] en het gemiddelde niet verandert.

Berekening

$$\bar{x} = 53$$

$$\bar{x} + 2 * s_x = 53 + 2 * s_x = 75$$

$$\bar{x} - 2 * s_x = 53 - 2 * s_x = 31$$

$$s_x = 11$$

$$\bar{y} = 53$$

$$\bar{y} + 2 * s_y = 53 + 2 * s_y = 86$$

$$\bar{y} - 2 * s_y = 53 - 2 * s_y = 20$$

$$s_y = 16.5$$

$$y_i = a + bx_i$$

$$s_y = b * s_x$$

$$16.5 = b * 11$$

$$b = 1.5$$

$$\bar{y} = a + b\bar{x}$$

$$53 = a + 1.5 * 53$$

$$a = -26.5$$

Antwoord:

$$y_i = -26.5 + 1.5x_i$$

Vraag 6 (1,5 punten)

In een bedrijf met 1000 werknemers krijgen we onderstaande tabel met de verdeling naar geslacht en statuut en telkens de vermelding van het gemiddeld netto maandloon. (bovenste cijfer in een cel is de frequentie, onderste cijfer is het gemiddeld netto maandloon)

	Arbeiders	Bedienden	Kader	
Man	530 2100	150 2500	70 2800	750
Vrouw	120 2000	100 2200	30 2900	250
	650	250	100	

Bereken het verschil tussen mannen en vrouwen wat betreft het gemiddeld netto maandloon.

Berekening

$$\bar{x}_{\text{man}} = \frac{530 * 2100 + 150 * 2500 + 70 * 2800}{530 + 150 + 70} = \frac{1684000}{750} = 2245,333$$

$$\bar{x}_{\text{vrouw}} = \frac{120 * 2000 + 100 * 2200 + 30 * 2900}{120 + 100 + 30} = \frac{547000}{250} = 2188$$

$$\text{verschil} = \bar{x}_{\text{vrouw}} - \bar{x}_{\text{man}} = 2245,333 - 2188 = 57,333$$

Antwoord:

57,333

Vraag 7 (2,5 punten) (-0,5 per fout/blanco)

Kruis de maten aan die gevoelig zijn voor outliers en geef voor elke maat het laagste meetniveau waarvoor de maat berekend kan worden.

maat	Deze maat is gevoelig voor outliers	Dit is het laagste meetniveau waarvoor deze maat berekend kan worden.
Modus		Nominaal
Interkwartielafstand		Metrisch/kwantitatief
Gemiddelde	x	Metrisch/kwantitatief
Range	x	Metrisch/kwantitatief
Standaarddeviatie	x	Metrisch/kwantitatief
Mediaan		Ordinaal

Vraag 8 (1 punt) (-0,5 per fout)

Kruis alle juiste uitspraken aan:

- Als de correlatie tussen twee variabelen gelijk is aan nul, is er geen samenhang tussen deze variabelen.
- Conditionele proporties zijn een goede methode om de samenhang tussen twee categorische variabelen te bepalen.
- De covariantie is de gestandaardiseerde correlatie.
- Op basis van een histogram kan je altijd de oorspronkelijke datamatrix van de variabele volledig reconstrueren.