

PROEFEXAMEN SOCIALE STATISTIEK

November 2009 – REEKS 1

Score .../15

1. (2 punten) Beantwoord volgende vraag aan de hand van onderstaande SPSS-output:
Omcirkel de juiste waarde voor A en voor B als je weet dat deze verdeling berekend is op basis van 250 observaties. Leg uit hoe je dit berekend hebt.

VAR00001

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 5,00				22,0
10,00	A			34,0
15,00	B			42,0
20,00				84,0
25,00				100,0
Total				

Antwoord:	A:	12	22	<input type="text" value="30"/>	34
	B:	8	<input type="text" value="20"/>	34	42

Toelichting

Berekening van het gewone percentage voor A:

$$34 \% - 22 \% = 12 \%$$

Berekening van de frequentie voor A:

$$250 / 100 * 12 = 30$$

Berekening van het gewone percentage voor B:

$$42 \% - 34 \% = 8 \%$$

Berekening van de frequentie voor B:

$$250 / 100 * 8 = 20$$

Puntenverdeling

1 punt voor A bij juiste antwoord en toelichting

1 punt voor B bij juiste antwoord en toelichting

2. (2 punten) Starttermen Wiskunde: Gegeven onderstaande 4x4 kruistabel van de kenmerken x en y

		Y				
		1	2	3	4	
X	1	<u>19</u>	<u>5</u>	7	21	52
	2	<u>23</u>	<u>25</u>	11	17	76
	3	<u>31</u>	<u>34</u>	29	27	121
	4	11	13	17	19	60
		84	77	64	84	

Bereken $\sum_{i=1}^3 \sum_{j=1}^2 \left(\frac{f_{ij}}{f_{i\cdot}} \right)^2 =$

Oplossing

$$\begin{aligned} \sum_{i=1}^3 \sum_{j=1}^2 \left(\frac{f_{ij}}{f_{i\cdot}} \right)^2 &= \left(\frac{f_{11}}{f_{1\cdot}} \right)^2 + \left(\frac{f_{12}}{f_{1\cdot}} \right)^2 + \left(\frac{f_{21}}{f_{2\cdot}} \right)^2 + \left(\frac{f_{22}}{f_{2\cdot}} \right)^2 + \left(\frac{f_{31}}{f_{3\cdot}} \right)^2 + \left(\frac{f_{32}}{f_{3\cdot}} \right)^2 \\ &= \left(\frac{19}{52} \right)^2 + \left(\frac{5}{52} \right)^2 + \left(\frac{23}{76} \right)^2 + \left(\frac{25}{76} \right)^2 + \left(\frac{31}{121} \right)^2 + \left(\frac{34}{121} \right)^2 = 0,487 \end{aligned}$$

Puntenverdeling

- 1 punt indien alle juiste tellers geselecteerd werden (19, 5, 23, 25, 31, 34)
2 punten bij juiste einduitkomst en toelichting

Antwoord:

3. (1 punt) Vijf KMO's (A, B, C, D, E) hebben arbeiders in dienst. De informatie hierover is weergegeven in onderstaande tabel.

KMO	frequentie
A	2
B	5
C	8
D	9
E	11
	35

Wat zijn hier de onderzoekseenheden?

Arbeiders

Wat is het kenmerk en het meetniveau van dat kenmerk?

Kenmerk: KMO

Meetniveau: nominaal

Puntenverdeling

0.25 voor arbeiders

0.25 voor KMO

0.5 voor nominaal

4. (2 punten) Geef de vier eigenschappen van een correlatie.

Zie handboek p 108 – 109: Properties of the correlation

Mogelijke eigenschappen telkens goed voor 0.5 punten met maximum van 2 punten:

- De correlatie ligt altijd tussen -1 en +1.
- Hoe dichter de absolute waarde van de correlatie bij 1 ligt, hoe sterker het LINEAIRE verband. Hoe dichter de absolute waarde van de correlatie bij 0 ligt, hoe zwakker het LINEAIRE verband.
- Een positieve correlatie wijst op een positieve associatie ; een negatieve correlatie op een negatieve associatie.
- De waarde van de correlatie is niet meeteenheidafhankelijk
- De waarde van de correlatie is gevoelig voor outliers.
- Twee variabelen hebben dezelfde correlatie ongeacht welke variabele als outcome/verklarende variabele gekozen wordt : $r_{xy} = r_{yx}$. De correlatie is symmetrisch.

5. (2 punten) Een onderzoeker onderscheidt bij het opstellen van een frequentieverdeling in verband met de houding tegenover migranten, de volgende categorieën: zeer negatief, negatief, noch negatief – noch positief, positief en zeer positief. Deze vijf categorieën samen bevatten 2534 steekprofeenheden.

Tabel: relatieve frequentieverdeling

zeer negatief	0,135359
negatief	0,146014
noch negatief - noch positief	0,222573
positief	0,310576
zeer positief	0,185478

De categorie ‘geen mening’ is niet in de bovenstaande tabel opgenomen, maar omvat 12,5% van het totaal aantal steekprofeenheden (= de vijf categorieën uit de tabel + de categorie ‘geen mening’). Bepaal het percentage dat overeenstemt met de categorie ‘zeer negatief’ wanneer de categorie ‘geen mening’ wel in de tabel zou worden opgenomen.

Oplissing

Absoluut aantal ‘zeer negatief’ is $0,135359 \times 2534 = 343$

87,5% (= $100 - 12,5$) van het aantal steekprofeenheden is 2534.

100% van het totaal aantal steekprofeenheden is $2534 / 87,5 * 100 = 2896$

Relatief aantal ‘zeer negatief’ is dan $\frac{343}{2896} = 0,118$ of 11,844%

Puntenverdeling

1 punt indien waarde voor de teller of voor de noemer gevonden werd.

2 punten bij juiste einduitkomst en toelichting

Antwoord:

6. (2 punten) Geef een waarde voor x_5 en y_5 zodanig dat de waarde van de correlatie berekend op basis van de eerste 4 observaties, niet verandert.

	x	y
1	3	8
2	5	12
3	6	14
4	7	16
5	x_5	y_5

Oplossing

Methode 1

Het punt (\bar{x}, \bar{y}) ligt steeds op de regressierechte. Wanneer we dit punt toevoegen aan de gegevens, zal het de regressielijn, en dus ook de correlatie, niet beïnvloeden.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{3+5+6+7}{4} = 5,25$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{8+12+14+16}{4} = 12,5$$

Het punt (5.25 ; 12.5) zal de correlatie onveranderd laten.

Methode 2

Via grafisch rekenmachine:

Kolom x ingeven bij L1

Kolom y ingeven bij L2

LinReg (a+bx) L1, L2 geeft je de waarde van de correlatie, het intercept en de slope:

$r = 1$ dwz dat de punten op 1 rechte liggen

$$a = 2$$

$$b = 2$$

dus regressievergelijking: $\hat{y} = 2 + 2x_i$

Alle punten die aan deze vergelijking voldoen, liggen ook op de rechte en gaan de correlatie onveranderd laten. Bvb (0,2); (1,4); (2,6); (4,10); (8,18); (9,20); ...

Handmatig:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{3+5+6+7}{4} = 5,25$$

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{(3-5,25)^2 + (5-5,25)^2 + (6-5,25)^2 + (7-5,25)^2}{3}} = 1,707825128$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{8+12+14+16}{4} = 12,5$$

$$s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{\frac{(8-12,5)^2 + (12-12,5)^2 + (14-12,5)^2 + (16-12,5)^2}{3}} = 3,415650255$$

	x	y	z _x	z _y	z _x * z _y
1	3	8	-1,3174651	-1,3174651	1,73571429
2	5	12	-0,14638501	-0,14638501	0,02142857
3	6	14	0,439155033	0,439155033	0,19285714
4	7	16	1,024695076	1,024695077	1,05
5	x ⁵	y ⁵			

$$r_{xy} = \frac{\sum_{i=1}^n z_x * z_y}{n-1} = \frac{1,73571429 + 0,02142857 + 0,19285714 + 1,05}{3} = \frac{3}{3} = 1$$

De correlatie is 1: dwz dat de punten op een rechte liggen. Aan de hand van de vergelijking van de rechte kunnen we nog een punt van de rechte bepalen.

$$\hat{y} = a + bx_i \text{ met } b = r_{xy} * \left(\frac{s_y}{s_x} \right) \text{ en } a = \bar{y} - b\bar{x}$$

$$b = r_{xy} * \left(\frac{s_y}{s_x} \right) = 1 * \left(\frac{3.415650255}{1.707825128} \right) = 2$$

$$a = \bar{y} - b\bar{x} = 12,5 - 2 * 5,25 = 2$$

Dus: $\hat{y} = 2 + 2x_i$

Alle punten die aan deze vergelijking voldoen, liggen ook op de rechte en gaan de correlatie onveranderd laten. Bvb (0,2); (1,4); (2,6); (4,10); (8,18); (9,20); ...

Puntenverdeling

1 punt voor de redenering

1 punt voor een juiste waarden voor x₅ én y₅

7. (2 punten) Een onderzoeker stelt vast dat de afhankelijke variabele Y met 4 eenheden daalt als de onafhankelijk variabele met één eenheid toeneemt. De onderzoeker beslist vervolgens om de waarden van X te delen door vier. Omcirkel de wijziging in Y per eenheidstoename in X na de transformatie en leg uit hoe je dit berekend hebt.

Antwoord:	-4	<u>-16</u>	1	4
-----------	----	------------	---	---

Toelichting

$$y = -4x$$

$$x' = \frac{x}{4} \Rightarrow x = 4x'$$

$$y = -4 * 4x' = -16x'$$

Puntenverdeling

1 punt voor een juist antwoord zonder redenering of foute redenering

2 punten voor een juist antwoord met juiste redenering

8. (1 punt) Omcirkel het juiste antwoord: De verwachte vorm voor de verdeling van de tijd nodig om een moeilijk examen af te werken wanneer de maximumtijd 1 uur is, is

Antwoord:	<u>links scheef</u>	symmetrisch	rechts scheef
-----------	---------------------	-------------	---------------

Puntenverdeling

1 punt voor een juist antwoord

9. (1 punt) De resultaten van een examen lagen tussen 30 en 93 met een gemiddelde van 69. Omcirkel de meest realistische waarde voor de standaarddeviatie en leg uit waarom je deze waarde gekozen hebt.

Antwoord:	-11	2	<u>13</u>	63
-----------	-----	---	-----------	----

Toelichting

-11: een standaarddeviatie kan niet negatief zijn aangezien het de vierkantswortel van de variantie is.

2: te klein want dan zouden alle scores zeer dicht bij 69 moeten liggen

63: te groot want dit is bijna gelijk aan de range.

Puntenverdeling

0.25 punten voor een juist antwoord

0.25 punten per juiste toelichting voor de uitsluiting van de andere mogelijkheden.