PROEFEXAMEN SOCIALE STATISTIEK

November 2011 – REEKS 1

Score **…/20**

Naam: ....

Voornaam: ....

Studierichting: ………………………………………………….

Studentennummer: ....

Studierichting (laatste) middelbaar:…………………………………….

Uren wiskunde per week (laatste middelbaar): …….

Enkele belangrijke opmerkingen:

• Kijk vooraleer je begint kort de hele bundel door. Je zou een vragenbundel met 9 bladzijden moeten hebben. Ga ook na of je het formuleblad en kladpapier ontvangen hebt.

• Na iedere vraag staat tussen haakjes aangegeven voor hoeveel punten deze meetelt in het eindtotaal (**op 20 punten**). Indien je een bepaalde vraag niet kan oplossen, blijf dan geen half uur tobben over dat probleem, maar los **eerst** de andere vragen op. Vaak kan je dan achteraf de opengelaten vraag verder oplossen. Indien je dan toch een bepaalde vraag niet kan oplossen betekent dit niet noodzakelijk ook een slecht eindresultaat!

• Maak indien nodig je bewerkingen eerst op kladpapier en schrijf de definitieve bewerkingen, tussentijdse resultaten en antwoorden **netjes** over op de nette bladen. Zorg dat alle stappen in de juiste volgorde en ordelijk worden weergegeven. Let wel: enkel de antwoorden op de nette bladen tellen, er wordt niet gekeken naar antwoorden op het kladpapier.

• Bij **plaatsgebrek**, maak je gebruik van de achterzijde van de laatste pagina. Geef dan wel duidelijk aan over welke vraag het gaat.

• Rond de resultaten af tot op **drie cijfers na de komma**. Tussenbewerkingen worden best niet afgerond.

• Gelieve niet op de formulebladen te schrijven.

VEEL SUCCES!

Vraag 1 (2 punten)

a) Bereken:

$$\sum_{i=1}^{2}\sum_{j=2}^{4}\frac{j}{2i}+5=$$

$$\boxed{\sum_{i=1}^{2}\sum_{j=2}^{4}\frac{j}{2i}+5=\frac{2}{2*1}+\frac{3}{2*1}+\frac{4}{2*1}+\frac{2}{2*2}+\frac{3}{2*2}+\frac{4}{2*2}+5=1+1,5+2+0,5+0,75+1+5=11,75}$$

b) Gegeven onderstaande 4x4 kruistabel van de kenmerken x en y. Voor de variabele x loopt i van 1 tot 4. Voor de variabele y loopt j van 1 tot 5.

		y					
		1	2	3	4	5	
x	1	15	26	3	12	19	75
	2	5	31	19	8	14	77
	3	17	11	20	16	3	67
	4	24	18	12	12	19	85
		61	86	54	48	55	

Bereken:

$$\sum_{i=2}^{3}\sum_{j=3}^{4}5*f_{ij}^{2}+3*f_{\bullet\bullet}=$$

$$\boxed{\begin{aligned}&\sum_{i=2}^{3}\sum_{j=3}^{4}5*f_{ij}^{2}+3*f_{\bullet\bullet}=5*(f_{23}^{2}+f_{24}^{2}+f_{33}^{2}+f_{34}^{2})+3*f_{\bullet\bullet}\\&=5*(19^{2}+8^{2}+20^{2}+16^{2})+3*304=6317\end{aligned}}$$

Vraag 2 (3 punten)

Een belangrijke voedselproducent maakt gebruik van een smaakpanel om de smaak van zijn nieuwe producten te evalueren alvorens ze op de markt te brengen. De producent ontwikkelde nieuwe bereide maaltijden die op een aantal smaakaspecten (zout, zoet en zuur) geëvalueerd worden op een schaal van 0 tot 10. Een hoge score voor een smaakaspect wijst erop dat die smaak erg aanwezig is, een lage score wijst erop dat die smaak (bijna) niet aanwezig is. De leden van het smaakpanel geven ook een algemene smaakbeoordeling van de bereide maaltijd op de volgende schaal: 'zeer lekker, lekker, neutraal, niet lekker, helemaal niet lekker'. Daarnaast werd ook genoteerd of de bereide maaltijden al dan niet vegetarisch waren.

a) Wat zijn de onderzoekseenheden?

Bereide maaltijden

b) Maak een opsomming van de variabelen die gemeten zijn bij deze onderzoekseenheden en vermeld telkens hun meetniveau.

Variabelen	Meetniveau
Zout	Metrisch
Zoet	Metrisch
Zuur	Metrisch
Algemene smaakbeoordeling	Ordinaal
Vegetarisch	Nominaal

Vraag 3 (3 punten)

Gegeven volgend stamgram:

Stam	Blad
5	1 4 8 9
6	1 1 2 4 5 5
7	0 1
8	5 9

a) Bereken het gemiddelde van het kenmerk dat in dit stamgram gepresenteerd wordt.

Berekening + antwoord:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\overline{x} = \frac{1}{14}(51+54+58+59+61+61+62+64+65+65+70+71+85+89)$$

$$\overline{x} = 65,357$$

b) Bereken de variantie van het kenmerk dat in dit stamgram gepresenteerd wordt.

Berekening + antwoord:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$s^2 = \frac{1}{13}((51-65,357)^2+(54-65,357)^2+(58-65,357)^2+(59-65,357)^2+(61-65,357)^2$$

$$+(61-65,357)^2+(62-65,357)^2+(64-65,357)^2+(65-65,357)^2+(65-65,357)^2+(70-65,357)^2$$

$$+(71-65,357)^2+(85-65,357)^2+(89-65,357)^2)$$

$$s^2 = 113,786$$

c) Wat is de modus van het kenmerk dat in dit stamgram gepresenteerd wordt?

Antwoord: Er zijn 2 modi: **61** en **65**

<u>Vraag 4</u> (3 punten)

Bij een steekproef van 800 personen werd gevraagd of men al dan niet rookte. Verder worden er 2 leeftijdscategorieën onderscheiden: jongere (= jonger dan 30 jaar) en oudere rokers (= 30 jaar en ouder). Uit de resultaten blijkt dat 26% van de bevraagde personen rookt. Het verschil in celfrequentie tussen jongere rokers en oudere rokers is 2% (uitgedrukt in een percentage op het totaal aantal elementen van de steekproef). In absolute frequenties zijn er meer oudere rokers dan jongere rokers. Bepaal het aantal personen dat rookt in beide leeftijdscategorieën.

<u>Berekening</u>

		Roker		
		ja	nee	
Leeftijd	<30	x1		
	≥30	x2		
		26%		800

$$\begin{cases} x_1 + x_2 = 0,26 * 800 = 208 \\ x_2 - x_1 = 0,02 * 800 = 16 \end{cases}$$

➔ In totaal zijn er 208 rokers. Er zijn 16 meer oudere rokers dan jongere rokers.

$$\begin{cases} x_1 + x_2 = 208 \\ x_2 - x_1 = 16 \end{cases}$$

$$x_1 + x_2 = 208$$
$$x_2 = 16 + x_1$$
$$x_1 + (16 + x_1) = 208$$
$$2x_1 = 208 - 16$$

$$\begin{cases} x_1 = 96 \\ x_2 = 16 + 96 = 112 \end{cases}$$

➔ Er zijn 96 jongere rokers en 112 oudere rokers.

<u>Antwoord</u>:

Aantal jongere rokers: 96

Aantal oudere rokers: 112

Vraag 5 (3 punten)

We beschouwen de puntenverdeling voor een bepaald examen dat gescoord werd op 20. De examenresultaten hebben een klokvormige verdeling. Het gemiddelde bedraagt 10/20 en 95% van de observaties valt in het interval [7;13]. Welke transformatie moeten we op deze punten uitvoeren om ervoor te zorgen dat 95% van de observaties in het interval [2;18] valt terwijl het gemiddelde niet verandert?

Berekening

$$y_i = ax_i + b$$
$$\overline{x} = 10$$
$$\overline{y} = 10$$
$$2s_x = 3 \Rightarrow s_x = 1,5$$
$$2s_y = 8 \Rightarrow s_y = 4$$

$$s_y = a \cdot s_x$$
$$4 = a \cdot 1,5$$
$$a = 2,6667$$

$$y_i = 2,6667 \cdot x_i + b$$
$$10 = 2,6667 \cdot 10 + b$$
$$b = 10 - 26,6667 = -16,6667$$

$$y_i = 2,6667 \cdot x_i - 16,6667$$

Antwoord:

Uit te voeren transformatie op het kenmerk punten:

$$y_i = 2,6667 \cdot x_i - 16,6667$$

Vraag 6 (1 punt)

Één seizoen van het programma “de slimste mens” bestaat uit 10 afleveringen. In elke aflevering dingen drie kandidaten mee naar de titel van slimste mens. Van deze drie kandidaten valt degene met de laagste score echter af. In de volgende aflevering wordt deze dan vervangen door een nieuwe kandidaat. Hoeveel keer spelen de kandidaten gemiddeld mee?

Berekening

$$\overline{x} = \frac{30}{12} = 2,5$$

Onafhankelijk van de prestaties van de specifieke kandidaten, blijft de totale som van beurten gelijk. In totaal komen er 12 kandidaten aan de beurt. Per aflevering zijn er drie ‘beurten’ (3*10=30 beurten).

Antwoord:

2,5 keer

<u>Vraag 7</u> (2 punten)

Voor een bepaald examen krijgen de studenten lettercodes in de plaats van punten toegekend. A is hierbij de beste score, F is de slechtste score. De oorspronkelijke puntenverdeling ziet er als volgt uit:

Punten	f_i
A	35
B	55
C	25
D	45
E	15
F	25

Men besluit echter om de puntenverdeling verder samen te vatten door een aantal categorieën samen te voegen.

Punten	f_i
A-B	90
C-D	70
E-F	40

Wat gebeurt er met het oorspronkelijke gemiddelde door het samenvoegen van deze categorieën? Omcirkel het juiste antwoord en leg kort uit waarom!

a. Het oorspronkelijke gemiddelde *daalt* wanneer men deze categorieën samenvoegt

b. Het oorspronkelijke gemiddelde *stijgt* wanneer men deze categorieën samenvoegt

c. Het oorspronkelijke gemiddelde *verandert niet* wanneer men deze categorieën samenvoegt

<u>Motivering keuze</u>:

Antwoord: B: het gemiddelde wordt hier eigenlijk overschat. Het klassenmidden wordt gebruikt als benadering voor het gemiddelde van elke klasse. Wanneer in elke klasse systematisch meer observaties onder het klassenmidden liggen is deze benadering van het gemiddelde een overschatting en zal het bekomen gemiddelde over de klassen heen dus ook een overschatting zijn.

Vraag 8 (3 punten)

(fictieve gegevens) Uit onderzoek bij 445 bachelor studenten blijkt dat slechts 20% van hen ooit al bloed heeft gedoneerd.

a) Vul de gegevens verder aan in onderstaande kruistabellen.

Frequenties:

	Bloeddonatie		
Studiejaar	Wel	Niet	**Totaal**
1e bachelor			
2e bachelor	15		**107**
3e bachelor			
Totaal			

Conditionele proporties (afgerond op 4 decimalen):

	Bloeddonatie			
Studiejaar	Wel	Niet	**Totaal**	n
1e bachelor		0,8455		123
2e bachelor			**1**	
3e bachelor	0,2558			

	Bloeddonatie		
Studiejaar	Wel	Niet	**Totaal**
1e bachelor	19	104	**123**
2e bachelor	15	92	**107**
3e bachelor	55	160	**215**
Totaal	**89**	**356**	**445**

	Bloeddonatie			
Studiejaar	Wel	Niet	**Totaal**	n
1e bachelor	0,1545	0,8455	**1**	123
2e bachelor	0,1402	0,8598	**1**	107
3e bachelor	0,2558	0,7442	**1**	215

b) Wat is de afhankelijke variabele en wat is de onafhankelijke variabele?

Antwoord:

Afhankelijke variabele = response variabele = bloeddonatie

Onafhankelijke variabele = verklarende variabele = studiejaar