

## PARAGRAAF 10.1 : POPULATIE EN STEEKPROEF

## LES 1 : HERHALING

## DEFINITIE BETROUWBAARHEIDSINTERVAL (BI)

- Betrouwbaarheidsinterval (BI) = { de waarden waarvoor de uitkomst van de steekproef betrouwbaar is }
- We gaan betrouwbaarheidsintervallen berekenen. Daarvoor hebben we de volgende gegevens nodig :

(1)  $p = \{ \text{populatieproportie} \} = \{ \text{Kans dat iemand dit kiest} \}$

(2)  $n = \{ \text{Steekproeflengte} \} = \{ \text{Hoeveel producten je gekozen hebt in een steekproef} \}$

$$(3) \sigma = \sqrt{\frac{p(1-p)}{n}}$$

(4) 68% - BI =  $[p - \sigma ; p + \sigma]$

(5) 95% - BI =  $[p - 2\sigma ; p + 2\sigma]$

## VOORBEELD 1.

Bij de Top 2000 stemt 30% op de Stones. Jan voert een steekproef uit. Hij vraagt aan 40 mensen op wie ze gaan stemmen.

- Bereken  $\sigma$ .
- Bereken het 95%-BI.

## OPLOSSING 1

a.  $\sigma = \sqrt{\frac{0,3(1-0,3)}{40}} = 0,07$

b. 95% - BI =  $[0,3 - 2 \cdot 0,07 ; 0,3 + 2 \cdot 0,07] = [0,16 ; 0,44]$

**LES 2 : POPULATIEGEMIDDELDE OF STEEKPROEFGEMIDDELDE****DEFINITIES**

- $p$  = Populatiegemiddelde = { Het gemiddelde is uit de hele populatie genomen }
- $\sigma$  = populatiestandaardafwijking
- $\bar{X}$  = Steekproefgemiddelde = { Het gemiddelde is uit de steekproef genomen }
- $S$  = Steekproefstandaardafwijking

**DEFINITIES BI**

- Bij populatie : 95% - BI =  $[p - 2\sigma ; p + 2\sigma]$
- Bij steekproef : 95% - BI =  $[\bar{X} - 2\frac{S}{\sqrt{n}} ; \bar{X} + 2\frac{S}{\sqrt{n}}]$

**OPMERKING BI**

- Bij een steekproef-BI wordt de  $\bar{X}$  en  $S$  altijd gegeven !!!!!

**VOORBEELD 1**

In Nederland zijn 100000 Jan Smit fans. In een steekproef van 500 personen zijn 20 Jan Smit fans.

- Bereken het 95%-BI.
- Bereken hoeveel personen uit deze steekproef met 95% betrouwbaarheid fan zijn van Jan Smit.

Sam doet een onderzoek naar overgewicht. Hij voert een steekproef uit onder 270 mensen. Het steekproefgemiddelde is 16 kilo overgewicht en de steekproefstandaardafwijking is 3 kilo.

- Bereken het 95%-BI.

Bij een steekproef onder 80 mensen is het 95%-BI gelijk aan  $[0,20 ; 0,28]$ .

- Bereken het steekproefgemiddelde en de steekproefstandaardafwijking.

**OPLOSSING 1**

- a. Het gaat om een deel van de hele groep. Dus populatie !!! (Je kunt ook zien dat  $\bar{X}$  en S niet gegeven zijn, dus moet het wel populatie zijn)

$$(1) p = \frac{20}{500} = 0,04 \rightarrow \sigma = \sqrt{\frac{0,04(1-0,04)}{500}} = 0,0088$$

$$(2) 95\% - \text{BI} = [p - 2\sigma ; p + 2\sigma] = [0,04 - 2 \cdot 0,0088; 0,04 + 2 \cdot 0,0088]$$

$$95\% - \text{BI} = [0,0224 ; 0,0576]$$

- b.  $0,0224 \times 500 = 11,2 = 11$  en  $0,0576 \times 500 = 28,8 = 29 \rightarrow$  Dus tussen de 11 en de 29 personen

- c. Je weet dat  $\bar{X}$  en S wel gegeven zijn, dus moet het wel steekproef zijn !!

$$(1) \bar{X} = 16 \quad S = 3 \quad n = 270$$

$$(2) 95\% - \text{BI} = [\bar{X} - 2 \frac{S}{\sqrt{n}} ; \bar{X} + 2 \frac{S}{\sqrt{n}}] = [16 - 2 \cdot \frac{3}{\sqrt{270}} ; 16 + 2 \cdot \frac{3}{\sqrt{270}}]$$

$$95\% - \text{BI} = [15,63 ; 16,37]$$

**Opmerking :**

Dit betekent dat 95% van de mensen uit de steekproef een overgewicht tussen de 15,63 en 16,37 kg hebben.

- d.  $\bar{X} - 2 \frac{S}{\sqrt{n}} = 0,20$  en  $\bar{X} + 2 \frac{S}{\sqrt{n}} = 0,28$  en  $n = 80$

$$(1) 4 \frac{S}{\sqrt{n}} = 0,28 - 0,20 = 0,08$$

$$\frac{S}{\sqrt{n}} = 0,02$$

$$(2) n = 80 \rightarrow \frac{S}{\sqrt{80}} = 0,02$$

$$S = 0,02 \cdot \sqrt{80} = 0,179$$

$$(3) \bar{X} = 0,20 + 2 \frac{S}{\sqrt{n}} = 0,20 + 2 \cdot 0,02 = 0,24$$

## PARAGRAAF 10.2 : GROEPEN VERGELIJKEN

## LES 1 : GROEPEN VERGELIJKEN BIJ EEN 2X2 KRUISTABEL

## DEFINITIES

Er zijn een aantal manieren om groepen gegevens met elkaar te vergelijken. De onderstaande gebruik je als je een 2x2 kruistabel hebt / kunt maken :

**(1) Percentageverschil (PV)**

$$PV = \text{Percentage A} - \text{Percentage B}$$

**(2) Odds-ratio (OR)**

$$OR = \frac{\text{Grootste kruisproduct}}{\text{Kleinste kruisproduct}}$$

**(3) Phi-coëfficiënt (phi)**

$$\text{phi} = \frac{\text{Grootste kruisproduct} - \text{Kleinste kruisproduct}}{\sqrt{T1 \times T2 \times T3 \times T4}}$$

met T1 is totaal van gegeven 1 etc.

Je kunt op basis van onderstaande tabel dan de volgende conclusies trekken :

PV	OR	phi	Verskil is
PV > 30%	OR > 3	phi < -0,4 of phi > 0,4	Groot
15% < PV ≤ 30%	2 ≤ OR ≤ 3	-0,4 < phi < -0,2 of 0,2 < phi < 0,4	Middelmatig
PV ≤ 15%	OR < 2	-0,2 < phi < 0,2	Gering

**VOORBEELD 1**

Gegeven is het aantal Havo 4 leerling dat een smartphone heeft :

Smartphone	Wel	Niet	Totaal
Jongens	38	5	43
Meisjes	43	14	57
<b>Totaal</b>	<b>81</b>	<b>19</b>	<b>100</b>

Bereken hoe groot het verschil is als je gebruik maakt van :

- Het Percentageverschil (PV)
- De Odds-ratio (OR)
- Phi-coëfficiënt (phi)

**OPLOSSING 1**

- Percentage Jongens =  $38 / 43 \times 100\% = 88,4\%$   
 Percentage Meisjes =  $43 / 57 \times 100\% = 75,4\%$   
 PV = Percentage Jongens – Percentage Meisjes =  $88,4 - 75,4 = 13\%$   
 Op basis van tabel is  $PV < 15\%$  dus gering verschil
- Kruisproduct A =  $38 \times 14 = 532$   
 Kruisproduct B =  $43 \times 5 = 215$   
 PV = Grootste Kruisproduct / Kleinste Kruisproduct =  $532 / 215 = 2,47$   
 Op basis van tabel is  $2 \leq OR \leq 3$  dus middelmatig verschil
- $$phi = \frac{Grootste\ kruisproduct - Kleinste\ kruisproduct}{\sqrt{T_1 \times T_2 \times T_3 \times T_4}} = \frac{532 - 214}{\sqrt{43 \times 57 \times 81 \times 19}} = \frac{217}{\sqrt{3772089}} = 0,11$$
 Op basis van tabel is  $-0,2 \leq phi \leq 0,2$  dus gering verschil

---

**OPMERKINGEN**

- Je ziet dat het afhangt van de ratio, wat de statistische conclusie. (Een vaak gehoord probleem bij statistisch onderzoek)
- In CE kom je alleen de phi tegen (belangrijkste !!!)
- Als je voor phi de formule  $phi = \frac{\text{Grootste kruisproduct} - \text{Kleinste kruisproduct}}{\sqrt{T_1 \times T_2 \times T_3 \times T_4}}$  gebruikt, kun je de tabel met conclusies wat vereenvoudigen :

phi	Vershil is
phi > 0,4	Groot
0,2 < phi < 0,4	Middelmatig
phi < 0,2	Gering

**LES 2 : GROEPEN VERGELIJKEN BIJ EEN GROTERE TABELLEN****DEFINITIE GROTE TABELLEN VERGELIJKEN**

Er zijn ook een aantal manieren om groepen gegevens met grotere tabellen met elkaar te vergelijken :

**(1) Maximale Verschil cumulatief percentage (Max Vcp)**

Max Vcp = Maximale [ Cumulatief Percentage A – Cumulatief Percentage B ]

**(2) Effectgrootte (E)**

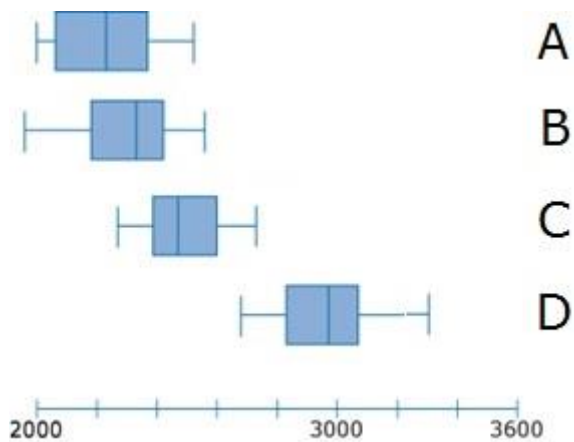
$$E = \frac{\text{Grootste } \bar{X}_1 - \text{Kleinste } \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$$

**(3) Boxplot**

Soms krijg je de boxplot van meerdere gegevens

Je kunt op basis van onderstaande tabel dan de volgende conclusies trekken :

Max. Vcp	E	Boxplot	Verskil is
Max. Vcp > 40%	$E > 0,8$	Geen overlapping	Groot
$20\% < \text{Max. Vcp} \leq 40\%$	$0,4 < E \leq 0,8$	Wel overlapping en minstens één mediaan ligt buiten de andere box	Middelmatig
Max. Vcp $\leq 20\%$	$E \leq 0,4$	Alle andere gevallen	Gering

**VOORBEELD 1**

Bij lampen wordt onderzoek gedaan naar de branduren. Bekijk de boxplots van branduren van lampen. Die laten het resultaat van steekproeven van vier typen lampen zien. Vergelijk de branduren van type A met de andere types.

**OPLOSSING 1**

Het verschil tussen

- A en B is gering, want de boxen overlappen en er ligt geen mediaan buiten de andere box;
- A en C is middelmatig, want de boxen overlappen en een mediaan (zelfs beide) ligt buiten de andere boxen;
- A en D is groot, want de boxen overlappen niet.



**VOORBEELD 2**

Om het effect van het taalonderwijs te onderzoeken is van twee even grote groepen Nederlanders en Belgen gekeken naar het aantal vreemde talen dat ze spreken. Bekijk de tabel met resultaten.

Aantal talen	Belgen	NL-ers	Totaal
0	122	19	141
1	168	156	324
2	184	272	456
3	103	146	249
4	34	18	52
Totaal	611	611	1222

Welk verschil is er, statistisch gezien, tussen het aantal gesproken talen van Belgen en Nederlanders als je kijkt naar :

- Verskil cumulatief percentage
- Effectgrootte als je met de GR berekend hebt dat :

$$\bar{X}_B = 1,606 \text{ en } \bar{X}_N = 1,980 \text{ en } S_B = 1,144 \text{ en } S_N = 0,868$$

## OPLOSSING 2

a. (1) Je hebt eerst van beide landen de cumulatieve percentages nodig :

Aantal talen	Belgen	Perc	Cum Perc
0	122	20,0	20
1	168	27,5	47,5
2	184	30,1	77,6
3	103	16,9	94,5
4	34	5,6	100,0
Totaal	611		

Aantal talen	NL-ers	Perc	Cum Perc
0	19	3,1	3,1
1	156	25,5	28,6
2	272	44,5	73,2
3	146	23,9	97,1
4	18	2,9	100,0
Totaal	611		

(2) Vervolgens bereken je de verschillen van de cumulatieve percentages

Cum Perc	Cum Perc	Vershil
20	3,1	16,9
47,5	28,6	<b>18,9</b>
77,6	73,2	4,5
94,5	97,1	-2,6
100,0	100,0	0,0

Je ziet dat max  $V_{cp} = 18,9$ . Omdat max.  $V_{cp} < 20\%$  is het verschil gering.

$$b. E = \frac{\text{Grootste } \bar{X}_1 - \text{Kleinste } \bar{X}_2}{\frac{1}{2}(S_1 + S_2)} = \frac{1,980 - 1,606}{\frac{1}{2}(1,144 + 0,868)} = 0,37$$

$E \leq 0,40$  dus het effect is gering.