

Een samenvatting over de vijf tools om uit te maken of een bepaald verschijnsel een grote, gemiddelde of kleine verandering is.

Deze vijf zijn:

Phi-coefficient	Verschillen bij nominale waarden.
MaxVCP	Verschillen bij Ordinale waarden
Effectgrootte	Verschillen bij kwantitatieve waarden.

Box-plots

Phi-coefficient.

Deze gebruiken we in situaties waarbij twee zaken tegen elkaar in een tabel worden gezet.

We hebben deze bekeken:

	tevredenheid		
	ontevreden	tevreden	
huurwoning	485	3172	3657
koopwoning	180	6163	6343
	665	9335	10000

De vraag die hier gesteld kan worden is:

“is er een verschil in tevredenheid van woning of het om een huurhuis of een koophuis gaat”.

Door deze 8 getallen (10000 doet niet mee) in de Phi-formule te zetten ontstaat een phi-waarde.

$$\text{phi} = \frac{485 \cdot 6163 - 180 \cdot 3172}{\sqrt{3657 \cdot 6343 \cdot 665 \cdot 9335}} \approx 0,202$$

De waarde van phi kent drie gebieden:

1. Als phi < -0,4 of groter is dan 0,4 dan wordt het verschil als “groot” aangemerkt.
2. Als Phi zit tussen 0,2 en 0,4 of -0,2 en -0,4 dan is het verschil “middelmatic”.
3. Als phi rond de nul zit (tussen -0,2 en 0,2) dan is het verschil gering.

Deze manier van ‘kwalificeren’ is om een idee te hebben hoe ‘groot’ een verschil is.

Waar herken ik aan dat de phi-formule moet worden ingezet?

1. Er is vaak sprake van een 2x2 tabel met sprekende verschillen: “man-vrouw” “aantal voorstanders – aantal tegenstanders” of dergelijke aantallen.
2. Het gaat om nominale waarden.

Pas op : niet elke 2x2 tabel moet zo worden behandeld:

INHOUD 1 LITERPAK APPELSAP

	gem.	stand. afw.
merk A	1007 mL	3 mL
merk B	1012 mL	10 mL

Hier is sprake van kenmerken die bij elkaar horen.

MaxVcP is de tweede.

We willen weten of twee trends ergens een grootste verschil hebben.

Daarvoor moeten we de (meestal langere) tabel eerst uitbreiden.

Voorbeeld (hebben we ook bekeken):

inkomens- categorie	werksector	
	overheid	bedrijf
1	25	11
2	58	34
3	24	32
4	7	26
5	6	12

De vraag hier is, bij welke categorie inkomen is het verschil tussen overheid en bedrijfsleven het grootst?

	werksector				
	overheid		bedrijf		
inkomens- categorie	cum. freq.	cum. perc.	cum. freq.	cum. perc.	Vcp
1	25	20,8%	11	9,6%	11,2%
2	83	69,2%	45	39,1%	30,1% ←..... max.Vcp
3	107	89,2%	77	67,0%	22,2%
4	114	95,0%	103	89,6%	5,4%
5	120	100%	115	100%	0%

Bij de max.Vcp hanteren we de volgende vuistregels om een conclusie te trekken over de grootte van het verschil tussen twee groepen.

- Bij $\text{max.Vcp} > 40$ zeggen we *het verschil is groot*.
- Bij $20 < \text{max.Vcp} \leq 40$ zeggen we *het verschil is middelmatig*.
- Bij $\text{max.Vcp} \leq 20$ zeggen we *het verschil is gering*.

Wat zien we?

1. We hebben twee extra kolommen nodig: percentage en cumulatief (doortellend) percentage.
2. .. en dat voor de beide kolommen: overheid en bedrijfsleven. In het voorbeeld hierboven worden die in één keer uitgerekend: $25/120 = 20,8\%$ deze wordt opgeteld bij $83/120 \cdot 100\%$ en levert zo: $69,2\%$ op. Etc.
3. Dan ontstaat de vijfde kolom: het verschil: Bij categorie 2 is het verschil tussen overheid en bedrijfsleven dus blijkbaar het grootst.
4. De drie criteria eronder $\text{MaxVcp} > 40$ noemen we "groot". Als Maxvcp tussen 20 en 40 ligt heet het middelmatig en zoals gezegd: onder de 20 is het verschil gering.

De effectgrootte.

De derde in de rij is de effectgrootte.

Hierbij worden altijd twee gemiddelden met hun standaardafwijkingen vergeleken.

INHOUD 1 LITERPAK APPELSAP

	gem.	stand. afw.
merk A	1007 mL	3 mL
merk B	1012 mL	10 mL

De gehanteerde formule om te onderzoeken hoe groot het verschil tussen merk A en B is, is deze:

$$E = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{1}{2}(S_1^2 + S_2^2)}}$$

Je moet dus (uit de tekst) vaststellen wat de gemiddelden zijn (hier 1007 en 1012) en hun standaard afwijkingen (hier 3 resp 10 ml).

$$E = (1012 - 1007) / 0,5(3 + 10) = 0,76.$$

Nu geldt:

- Bij $E > 0,8$ zeggen we *het verschil is groot*.
- Bij $0,4 < E \leq 0,8$ zeggen we *het verschil is middelmatig*.
- Bij $E \leq 0,4$ zeggen we *het verschil is gering*.

In dit geval is het verschil dus middelmatig.

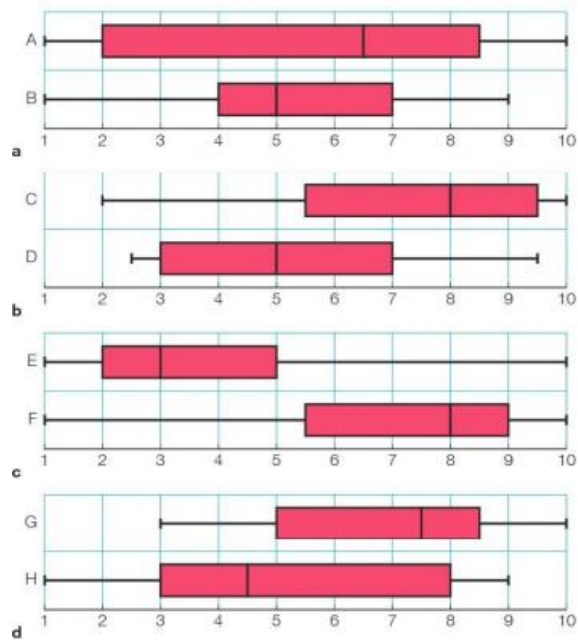
Deze theorie en vragen vinden we in par. 2.5

De vierde is het vergelijken van twee boxplots:

Een enkele boxplot zegt eigenlijk niks. Je hebt er altijd 2 (of meer) nodig om uitspraken te doen.

Hierbij let men alleen op het 'doosje' van de boxplot. Dus lees voor boxen, steeds die (rode) doosjes.

- Als de boxen elkaar niet overlappen, dan zeggen we *het verschil is groot*.
- Als de boxen elkaar wel overlappen en een mediaan van een boxplot ligt buiten de box van de andere boxplot, dan zeggen we *het verschil is middelmatig*.
- In alle andere gevallen zeggen we *het verschil is gering*.



Bij a zie je dat box 2 geheel binnen box 1 valt: het verschil is gering.

Bij b zie je dat de boxen wel overlappen, doch hun medianen steken niet in elkaars box. Dus is het verschil middelmatig.

Bij c overlappen de boxen totaal niet: verschil is groot.

Bij de set bij d zie we dat de mediaan van de bovenste steekt in de box van een ander, maar niet andersom: Dan wordt het verschil alsnog middelmatig.

De Odd-ratio.

Ook een vrij simpel criterium dat met een 2x2 tabel werkt.

	tevredenheid		
	ontevreden	tevreden	
huurwoning	485	3172	3657
koopwoning	180	6163	6343
	665	9335	10 000

We nemen expres deze die we al eerder zagen bij de phi-coefficient.
De odd-ratio deelt van één rij de ene kolom door de andere.

De odds-ratio voor deze tabel:

$$485/3657 = 0,1326$$

$$180/6343 = 0,028$$

$$\text{Oddsratio} = \text{grootste} / \text{kleinste} = 4,67$$

<i>PV</i>	<i>OR</i>	<i>phi</i>	het verschil is
$PV > 30\%$	$OR > 3$	$phi < -0,4$ of $phi > 0,4$	groot
$15\% < PV \leq 30\%$	$2 \leq OR \leq 3$	$-0,4 \leq phi < -0,2$ of $0,2 < phi \leq 0,4$	middelmatig
$PV \leq 15\%$	$OR < 2$	$-0,2 \leq phi \leq 0,2$	gering

OR is in dit geval > 3 , dus het verschil is groot.

Grappig is dat als we de phi-coeff gebruiken we de conclusie trekken dat het verschil middelmatig is.
Let op!! DAT KAN DUS. Het kan dus zijn dat de ene methode net iets anders concludeert dan de andere. Typische examenvraag!

SAMENGEVAT:

1. ODDs ratio. Vergelijkt onderlinge verhoudingen in een 2x2 tabel
2. Phi-coeff. Vergelijkt de samenhang van tabelelementen met hun totalen in een 2x2
3. Effectgrootte Werkt altijd met twee gemiddelden en hun standaardafwijkingen.
Pas op: deze gegevens staan vaak ook in een soort 2x2 tabel. Niet hetzelfde!!!
4. Boxplots Er zijn criteria voor het vergelijken van twee boxplots
5. MaxVcP Een manier om een langere tabel % uit te rekenen, cumulatief en het grootste verschil te duiden.

- Bij $\text{max.Vcp} > 40$ zeggen we *het verschil is groot*.
- Bij $20 < \text{max.Vcp} \leq 40$ zeggen we *het verschil is middelmatig*.
- Bij $\text{max.Vcp} \leq 20$ zeggen we *het verschil is gering*.

Dit was m.