

Opdracht 11a

chi-kwadraattoets

In Januari 1975 sprak de geneesmiddelencommissie van de Amerikaanse Academie voor Kindergeneeskunde de aanbeveling uit, geneesmiddelen met tetracycline niet te verstrekken aan kinderen onder de 8 jaar. Een twee jaar durende studie onderzocht in welke mate de artsen tussen 1973 en 1975 dergelijke geneesmiddelen hadden voorgeschreven. Het onderzoek classificeerde artsen met een huisartsenpraktijk volgens de aard van het gewest waar hun praktijk was gevestigd: stedelijk, voorstedelijk of plattelands. De onderzoekers gingen voor elk van deze categorieën na hoeveel artsen aan tenminste een patient onder de 8 jaar tetracycline hadden voorgeschreven. Hieronder volgt een tabel met de waargenomen aantallen. (Gegevens uit Wayne A. Ray et al., 'Prescribing of tetracycline to children of less than 8 years old', Journal of the American Medical Association, 237 (1977), blz. 2069-2074.)

Tetra	Type gewest		
	Stedelijk	Voorstedelijk	Plattelands
Tetracycline	65	90	172
Geen tetracycline	149	136	158

Voer de gegevens in in een tabel. Definieer de drie kolommen van de tabel en kies als kolomnamen 'gewest', 'tetra' en 'freq'.

- Maak een staafdiagram met voor ieder gewest twee staven, namelijk een voor artsen die wel, en een voor artsen die geen tetracycline voorgeschreven hadden.
- Bepaal per cel, rij en kolom de geobserveerde aantallen.
- Bepaal per cel, rij en kolom de verwachte aantallen.
- Bepaal per cel en kolom de rijpercentages. Wat laten ze zien?
- Bepaal per cel en rij de kolompercentages. Wat laten ze zien?
- Bepaal per cel, rij en kolom de totaalpercentages.
- Bepaal chi-kwadraat.
- Is er een significant verband tussen het al of niet voorschrijven van tetracycline door een arts en het gewest waar de arts gevestigd was? Formuleer H_0 en H_a . Wat is je conclusie?

Opdracht 11a - S-PLUS

Voer eerst de gegevens in in SPSS. Maak een tabel met drie kolommen: 'tetra', 'gewest' en 'freq'. Gebruik de volgende coderingen:

'tetra' : 1 = wel tetracycline voorgeschreven
 2 = geen tetracycline voorgeschreven

'gewest' : 1 = Stedelijk

2 = Voorstedelijk
3 = Plattelands

- a. Kies >Statistics >Data Summaries >Crosstabulations. Selecteer onder Data en achter Variables eerst de variabele 'gewest' door er een keer met de linker muisknop op te klikken. Selecteer vervolgens de variabele 'tetra' door de Ctrl-toets ingedrukt te houden en met de linker muisknop een keer op de naam van die variabele te klikken. Selecteer achter Counts Variable de variabele 'freq'. Klik nu bovenaan het venster op >Options. Zet de opties >Drop Unused Levels, Show Cell Proportions, Show Marginal Totals en >Run Chi-Square Test uit. Vul achter Decimal Places een 3 in. Klik op >OK. In het Report-venster wordt de kruistabel nu gegeven.

We moeten nu een echte gegevenstabel maken waarin we de gegevens van de kruistabel invoeren. Kies >File >New en selecteer >Data Set. We kunnen nu de gegevens uit de kruistabel kolom voor kolom invoeren. Het is van belang om per kolom steeds eerst de gegevens in te voeren, en pas daarna via >Format >Selected Object(s) de kolomnaam te geven door achter Name de door S-PLUS voorgestelde naam te vervangen door de gewenste naam.

De eerste kolom van de kruistabel bevat de gewesten. In onze nieuwe gegevenstabel vullen we als gegevens van boven naar onder achtereenvolgens in: stedelijk, voorstedelijk, plattelands. Daarna geven we als kolomnaam 'gewest'. Voor elk van de tweede en derde kolom we steeds eerst de frequenties in zoals gegeven in de kruistabel, en daarna geven we als kolomnaam respectievelijk 'wel tetra' en 'geen tetra'.

De tabel moet er als volgt uit komen te zien:

gewest	wel.tetra	geen.tetra
stedelijk	65	149
voorstedelijk	90	136
plattelands	172	158

Controleer na het intypen van de tabel de kolom 'gewest'. Selecteer de kolom 'gewest' door op de kolomnaam te klikken. Kies >Format >Selected Objects. Verwijder achter Factor Levels de gewesten die je niet wilt hebben. Klik op >OK.

We maken nu het geclusterde staafdiagram. Kies >Graph >2D Plot. Kies onder Axes Type voor Linear, en onder Plot Type voor Bar - Grouped (x, y1..yn). Klik op >OK. Selecteer onder Data Columns en achter x Columns de variabele 'gewest', en selecteer achter y Columns de variabelen 'wel.tetra' en 'geen.tetra' als volgt: klik met de linker muisknop op de variabele 'wel.tetra'. Houd vervolgens de Ctrl-toets ingedrukt en klik met de linker muisknop vervolgens op de variabele 'geen.tetra'. Klik op >OK.

De onderste toolbar is de Graph Toolbar. Klik in deze toolbar op het icoontje dat precies onder Linear in de Standard Toolbar zit. Als je de cursor op dit icoontje plaatst, moet als bijschrift verschijnen: Auto Legend. Klik op dit icoontje. De legenda wordt nu in de grafiek geplaatst.

Kies >Statistics >Data Summaries >Crosstabulations. Selecteer onder Data en achter Data Set de oorspronkelijke tabel. Selecteer achter Variables eerst de variabele 'gewest' door er een keer met de linker muisknop op te klikken. Selecteer vervolgens de variabele 'tetra' door de Ctrl-toets ingedrukt te houden en met de linker muisknop een keer op de naam van die variabele te klikken. Selecteer achter Counts Variable de variabele 'freq'. Klik nu bovenaan het venster op >Options. De opties >Drop Unused Levels, Show Cell

Proportions, Show Marginal Totals en >Run Chi-Square Test moeten aan staan. Vul achter Decimal Places een 3 in. Klik op >OK. In het Report-venster wordt de kruistabel nu gegeven.

In het Report-venster wordt de kruistabel nu gegeven. Deze kruistabel bevat de gegevens voor b., d., e. en f.. Onder de kruistabel vinden we de gegevens voor g. en h..

- b. De eerste waarde in een cel of in de totaalkolom of in de totaalrij bevat de geobserveerde aantallen voor die cel, rij of kolom.
- c. De verwachte aantallen moet je zelf handmatig voor iedere cel berekenen. Het verwachte aantal voor een cel is gelijk aan:

$$\frac{\text{rijtotaal} * \text{kolomtotaal}}{\text{tabeltotaal}}$$

Rijtotalen, kolomtotalen, en tabeltotaal kun je vinden in de kruistabel in het Report-venster.

- d. De tweede waarde in een cel bevat de rijproportie voor die cel, en de tweede waarde in de totaalkolom bevat de rijproportie voor die rij. Proporties worden omgerekend naar percentages door ze te vermenigvuldigen met 100.
- e. De derde waarde in een cel bevat het kolomproportie voor die cel, en de tweede waarde in de totaalrij bevat het kolomproportie voor die kolom. Proporties worden omgerekend naar percentages door ze te vermenigvuldigen met 100.
- f. De vierde waarde in een cel bevat de totaalproportie voor die cel, de tweede waarde in de totaalrij bevat de totaalproportie voor die kolom, en de tweede waarde in de totaalkolom bevat de totaalproportie voor die rij. Proporties worden omgerekend naar percentages door ze te vermenigvuldigen met 100.
- g. Onder 'Test for independence of all factors' is gegeven: $\text{Chi}^2=26.04449$.
- h. Onder 'Test for independence of all factors' is gegeven: $p=2.210598e-006$ wat hetzelfde is als $p=0.000002210598$.

Opdracht 11a - SPSS

Voer eerst de gegevens in in SPSS. Maak een tabel met drie kolommen: 'tetra', 'gewest' en 'freq'. Gebruik de volgende coderingen:

'tetra' : 1 = wel tetracycline voorgeschreven
2 = geen tetracycline voorgeschreven

'gewest' : 1 = Stedelijk
2 = Voorstedelijk
3 = Plattelands

- a. Kies >Graphs >Bar. Kies nu >Clustered en zet >Summaries for groups of cases aan. Klik nu op >Define. Plaats 'gewest' in het kader bij Category Axis en plaats 'tetra' in het kader bij Define Clusters. Zet onder Bars Represent >Other summary function aan, en plaats 'freq' in het kader bij Variable. Klik op >OK.

De kolom 'freq' bevat de frequenties van de combinaties. Om deze als gewichten voor de records te laten fungeren moeten we nu het volgende doen. Kies >Data >Weight Cases. Zet 'Weight cases by' aan en plaats 'freq' in het kader. Klik vervolgens op >OK.

Om SPSS chi-kwadraat te laten berekenen moeten we het volgende doen. Klik op >Statistics >Summarize >Crosstabs. Plaats 'tetra' in het kader 'Row(s)' en 'gewest' in het kader 'Column(s)'.

Klik nu op >Cells. Zet achtereenvolgens 'Observed', 'Expected', 'Row', 'Column' en 'Total' aan. Klik op >Continue. Klik daarna op >Statistics. Zet 'Chi-square' aan. Klik op >Continue. Klik tenslotte op >OK.

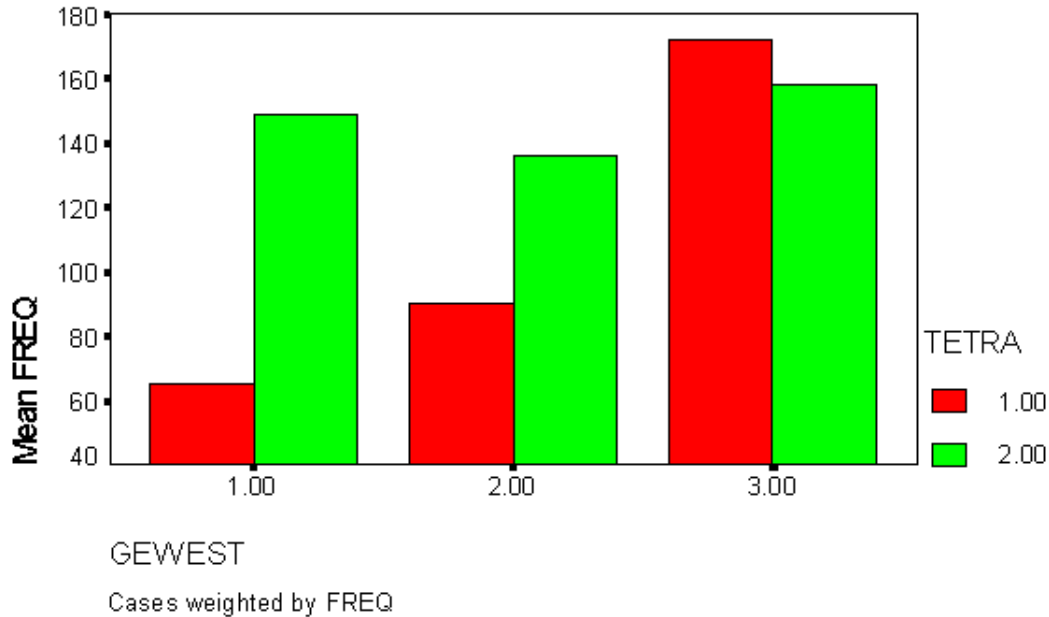
In het Output-window vinden we de resultaten. De tweede tabel heet 'Tetra* Gewest Crosstabulation' en bevat de gegevens voor b., c., d., e. en f.. De derde tabel heet 'Chi-Square Tests' en bevat de gegevens voor g. en h..

- b. De eerste waarde in een cel of in de totaalkolom of in de totaalrij bevat de geobserveerde aantallen voor die cel, rij of kolom.
- c. De tweede waarde in een cel of in de totaalkolom of in de totaalrij bevat de verwachte aantallen voor die cel, rij of kolom.
- d. De derde waarde in een cel bevat het rijpercentage voor die cel, en de derde waarde in de totaalkolom bevat het rijpercentage voor die rij.
- e. De vierde waarde in een cel bevat het kolompercentage voor die cel, en de vierde waarde in de totaalrij bevat het kolompercentage voor die kolom.
- f. De vijfde waarde in een cel of in de totaalkolom of in de totaalrij bevat het totaalpercentage voor die cel, rij of kolom.
- g. In de rij 'Pearson Chi-Square' vinden we in de kolom 'Value' de chi-kwadraat-waarde: 26.044.
- h. In de rij 'Pearson Chi-Square' vinden we in de kolom 'Asymp. Sig. (2-sided)' de P-waarde: 0.000.

Opdracht 11a - verslag

In Januari 1975 sprak de geneesmiddelencommissie van de Amerikaanse Academie voor Kindergeneeskunde de aanbeveling uit, geneesmiddelen met tetracycline niet te verstrekken aan kinderen onder de 8 jaar. Een twee jaar durende studie onderzocht in welke mate de artsen tussen 1973 en 1975 dergelijke geneesmiddelen hadden voorgeschreven. Het onderzoek classificeerde artsen met een huisartsenpraktijk volgens de aard van het gewest waar hun praktijk was gevestigd: stedelijk, voorstedelijk of plattelands. De onderzoekers gingen voor elk van deze categorieën na hoeveel artsen aan tenminste een patient onder de 8 jaar tetracycline hadden voorgeschreven.

- a. Maak een staafdiagram met voor ieder gewest twee staven, namelijk een voor artsen die wel, en een voor artsen die geen tetracycline voorgeschreven hadden.



b. Bepaal per cel, rij en kolom de geobserveerde aantallen.

Tetra	Type gewest			Totaal
	Stedelijk	Voorstedelijk	Plattelands	
Tetracycline	65	90	172	327
Geen tetracycline	149	136	158	443
Totaal	214	226	330	770

c. Bepaal per cel, rij en kolom de verwachte aantallen.

Tetra	Type gewest			Totaal
	Stedelijk	Voorstedelijk	Plattelands	
Tetracycline	90.9	96.0	140.1	327.0
Geen tetracycline	123.1	130.0	189.9	443.0
Totaal	214.0	226.0	330.0	770.0

Voor tabellen groter dan 2×2 mag de chi-kwadraattoets gebruikt worden wanneer het gemiddelde van de verwachte aantallen groter dan of gelijk is aan 5, en het kleinste verwachte aantal 1 of meer bedraagt. Het kleinste verwachte aantal is 90.9, zodat aan beide eisen wordt voldaan en de chi-kwadraattoets toegepast mag worden.

d. Bepaal per cel en kolom de rijpercentages. Wat laten ze zien?

Type gewest

Tetra	Stedelijk	Voorstedelijk	Plattelands	Totaal
Tetracycline	19.9%	27.5%	52.6%	100.0%
Geen tetracycline	33.6%	30.7%	35.7%	100.0%
Totaal	27.8%	29.4%	42.9%	100.0%

De tabel geeft voor elk type arts het percentage stedelijke artsen, het percentage voorstedelijke artsen en het percentage plattelands artsen. We zien dat zowel de artsen die wel als geen tetracycline verstrekten het meest op het platteland te vinden zijn. De rijpercentages per kolom laten zien dat de meeste artsen op het platteland te vinden zijn.

- e. Bepaal per cel en rij de kolompercentages. Wat laten ze zien?

Tetra	Type gewest			
	Stedelijk	Voorstedelijk	Plattelands	Totaal
Tetracycline	30.4%	39.8%	51.1%	42.5%
Geen tetracycline	69.6%	60.2%	47.9%	57.5%
Totaal	100.0%	100.0%	100.0%	100.0%

De tabel geeft voor elk type gewest het percentage artsen dat tetracycline voorschreef en het percentage dat dat niet deed. We zien dat in stedelijke gebieden tetracycline het minst werd voorgeschreven en in plattelandsgebieden het vaakst. De kolompercentages per rij laten zien dat de meeste artsen geen tetracycline voorschreven.

- f. Bepaal per cel, rij en kolom de totaalpercentages.

Tetra	Type gewest			
	Stedelijk	Voorstedelijk	Plattelands	Totaal
Tetracycline	8.4%	11.7%	22.3%	42.5%
Geen tetracycline	19.4%	17.7%	20.5%	57.5%
Totaal	27.8%	29.4%	42.9%	100.0%

- g. Bepaal chi-kwadraat.

Chi-kwadraat = 26.044. Hoe hoger chi-kwadraat, hoe sterker de verwachte aantallen en de waargenomen aantallen verschillen, hoe sterker het verband tussen het al of niet voorschrijven van tetracycline door een arts en het gewest waar de arts gevestigd was, en hoe sterker het bewijs tegen H_0 .

- h. Is er een significant verband tussen het al of niet voorschrijven van tetracycline door een arts en het gewest waar de arts gevestigd was? Formuleer H_0 en H_a . Wat is je conclusie?

H_0 : Er is geen verband tussen het al of niet voorschrijven van tetracycline door een arts en het gewest waar de arts was gevestigd.

H_a: Er is wel verband tussen het al of niet voorschrijven van tetracycline door een arts en het gewest waar de arts was gevestigd.

De P-waarde is gelijk aan 0.000. De kans, berekend onder de aanname dat H₀ waar is, dat de toetsingsgrootte χ^2 een waarde zou aannemen die even extreem is als of nog extremer is dan 26.044, is gelijk aan 0.000. Hoe kleiner de P-waarde, hoe sterker het door de data tegen H₀ geleverde bewijs. De P-waarde is kleiner dan α , want $0.000 < 0.05$, dus wordt H₀ verworpen. Er is een significant verband tussen het al of niet voorschrijven van tetracycline en het gewest.