

De chi-kwadraattoets

In deze aflevering bespreek ik de χ^2 -toets (uitgesproken en vaak ook geschreven als chi-kwadraattoets).

Deze toets geeft een goed beeld van de algemene principes van statistisch toetsen, en hij is met behulp van een eenvoudige rekenmachine toe te passen.

Simone Buitendijk

Het principe achter de chi-kwadraattoets is dat er twee situaties met elkaar worden vergeleken:

- de *verwachte* situatie;
- de *geobserveerde* situatie.

Verwacht betekent in dit geval: verwacht onder de nulhypothese. Deze wordt voor de start van een onderzoek geformuleerd en luidt dat twee te onderzoeken interventies of twee vergeleken onderzoeksgroepen geen verschil in uitkomst zullen laten zien.* Meestal zullen onderzoekers voor het begin van hun onderzoek natuurlijk hopen dat een nieuwe behandeling wel degelijk beter is dan de oude, maar die hoop heeft geen invloed op het formuleren van de nulhypothese.

Wanneer de nulhypothese wordt verworpen op basis van de onderzoeksresultaten, besluit men dus dat er wél een verschil is tussen de twee groepen. Naarmate de verschillen tussen de uitkomsten van beide groepen groter zijn, zal het meer voor de hand liggen te concluderen dat de nulhypothese ('geen verschil') niet kan worden gehandhaafd.

Met behulp van een statistische toets – in geval de chi-kwadraattoets – kan worden vastgesteld hoe groot de kans is dat de bevindingen nog op toeval zouden kunnen

berusten. Die kans wordt uitgedrukt in de p-waarde: hoe kleiner de p-waarde, des te kleiner de kans dat een gevonden verschil een toevalsbevinding kan zijn, en des te groter dus de neiging om te concluderen dat er echt een verschil bestaat tussen de twee onderzochte groepen.

Een voorbeeld

Als voorbeeld nemen we een hypothetisch onderzoek naar het effect van externe versie op de kans dat een kind dat à terme in stuit lag, bij de baring in hoofdligging ligt. We kunnen voor dit onderzoek beschikken over 200 zwangeren met een kind in stuit bij ongeveer 37 weken zwangerschapsduur. Deze vrouwen worden verdeeld over twee groepen: bij 100 vrouwen wordt de externe versie uitgevoerd (interventiegroep), bij de andere 100 niet (controlegroep). Uiteindelijk blijken na het onderzoek 80 vrouwen een kind in hoofdligging te hebben bij de aanvang van de baring.

De nulhypothese is dat er *geen* verschil is in kans op hoofdligging bij de baring tussen beide groepen. 'Onder de nulhypothese' verwacht je dus 40 vrouwen met een kind in hoofdligging in de interventiegroep en 40 in de controlegroep. Dit is de verwachte (*expected*) situatie. Deze cijfers brengen we onder in een zogenaamde 2x2-tabel (*tabel 1*).

In *tabel 2* is de situatie weergegeven zoals die in werkelijkheid wordt gevonden: de *observed* situatie. Van de 80 vrouwen met een kind in hoofdligging bleken er 30 in de controlegroep en 50 in de interventiegroep te zitten. De kans op een kind in stuitligging is dus 70% in de controlegroep en 50% in de interventiegroep. Het relatieve risico op een kind in stuitligging in de controlegroep ten opzichte van de interventiegroep is dus $50/30 = 1,4$.

De vraag is nu: kan dit verschil nog toeval zijn, of is er echt een gunstig effect van de externe versie? Als het verschil eigenlijk op toeval berust, betekent dit dat er in werkelijkheid geen effect is van de externe versie.

* In deze aflevering wordt voortgebouwd op eerdere artikelen in deze serie.^{1,2} Zo worden de begrippen nulhypothese, p-waarde en relatief risico bekend verondersteld.

Mw.dr. S.E. Buitendijk,
TNO Preventie en
Gezondheid,
Postbus 2215,
2301 CE Leiden;
e-mail
se.buitendijk@pg.tno.nl

Tabel 1

Aantallen in de cellen verwacht onder de nulhypothese ('expected')			
Uitkomst	Versie		Totaal
	Ja	Nee	
Hoofd	40	40	80
Stuit	60	60	120
Totaal	100	100	200

Tabel 2

Aantallen in de cellen geobserveerd in (hypothetisch) onderzoek ('observed')			
Uitkomst	Versie		Totaal
	Ja	Nee	
Hoofd	50	30	80
Stuit	50	70	120
Totaal	100	100	200

De formule

Met behulp van de chikwadraat-toets kan worden uitgerekend hoe groot de kans is dat het gevonden verschil geen toeval is. Daartoe gebruiken we de volgende formule:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Voor elk van de vier cellen in de 2x2-tabelen trekken we de *expected* waarde (in tabel 1) af van de *observed* waarde (in tabel 2). Deze getallen worden vervolgens gekwadraterd. Het kwadrateren is nodig om uitsluitend positieve getallen te krijgen; anders zouden de positieve en de negatieve verschillen binnen de cellen elkaar neutraliseren, en zou de waarde van de chi-kwadraat nul zijn.

Vervolgens delen we de vier uitkomsten door de bijbehorende *expected* waarde. Voor de cel linksboven zou dit dus worden: $(50 - 40)^2 / 40$. En voor de cel rechtsboven: $(30 - 40)^2 / 40$.

Deze vier getallen worden vervolgens bij elkaar opgeteld. De uitkomst van deze berekening is dan de waarde van de chi-kwadraat. In dit geval is die waarde 8,34. De waarde van de chi-kwadraat wordt groter naarmate de verschillen tussen de observed en de expected waarden groter worden. Dat is logisch: hoe groter die verschillen, des te groter de kans dat ze niet op toeval berusten.

Van chi-kwadraat naar p-waarde

De waarde van de chi-kwadraat is vervolgens met behulp van een tabel (tabel 3) te 'vertalen' naar een p-waarde.

Om te weten op welke regel in de tabel we moeten kijken, moeten we nog één ding vaststellen, namelijk het aantal 'vrijheidsgraden' (*degrees of freedom*, DF). Dat aantal vrijheidsgraden is afhankelijk van het aantal rijen en kolommen in de tabel met *observed* en *expected* waarden. We gebruiken hiervoor de volgende berekeningswijze:

- Het aantal rijen en het aantal kolommen wordt verminderd met 1.
- Beide uitkomsten worden met elkaar vermenigvuldigd.

Bij een 3x3-tabel is het aantal vrijheidsgraden dus $(3-1) \times (3-1) = 2 \times 2 = 4$.

Bij een 2x2-tabel (zoals in dit geval) is er één vrijheidsgraad: $(2-1) \times (2-1) = 1 \times 1 = 1$.

Tabel 3

Van chi-kwadraat naar p-waarde							
DF	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	2,706	3,841	5,024	6,635	7,879	10,828	12,116
2	4,605	5,991	7,378	9,210	10,597	13,816	15,202
3	6,251	7,815	9,348	11,345	12,838	16,266	17,730
4	7,779	8,488	11,143	13,277	14,860	18,467	19,998
5	9,236	11,070	12,832	15,086	16,750	20,515	22,108

DF = degrees of freedom (vrijheid)

In dit geval is de p-waarde die hoort bij de chi-kwadraat dus net iets groter dan 0,005. Dit betekent dat het verschil tussen de interventie- en de controlegroep in ieder geval statistisch significant is. Het betekent ook dat de kans dat het gevonden relatieve risico van 1,4 op toeval berust (en dus eigenlijk 1,0 is) ongeveer 0,005 (0,5%) is. Waarschijnlijk is er dus wel degelijk een effect van externe versie op kans op een hoofdligging atermo. De nulhypothese kan dus worden verworpen.

Toepassingsmogelijkheden

De chi-kwadraattoets is te gebruiken wanneer zowel de expositiefactor als de onderzochte uitkomsten nominale variabelen zijn. Dat betekent dat ze in te delen zijn in categorieën die een naam kunnen hebben, zonder dat daar een bepaalde volgorde bij hoort. Een nominale variabele is bijvoorbeeld 'geslacht' (manlijk of vrouwelijk) of 'ligging van de baby bij de geboorte'

(hoofdligging, stuitligging, dwarsligging). In principe kunnen bij nominale variabelen veel verschillende categorieën worden onderscheiden. Een chi-kwadraattoets kan dan ook voor grotere tabellen dan een 2x2-tabel worden uitgerekend. Wel geldt daarbij als regel dat er in de *expected* 2x2-tabel geen cellen mogen zijn met minder dan vijf observaties. Let er ook op dat altijd absolute getallen worden ingevuld, geen percentages.

Met een simpele rekenmachine kunnen zo bij bepaalde onderzoeksresultaten de bijbehorende p-waarden uitgerekend worden. Vergeet echter nooit dat ook bij onzinnige onderzoeksvragen en -uitkomsten heel nette p-waarden kunnen worden uitgerekend. ¹

Literatuur

- 1 Buitendijk SE. Risicomaten. Tijdschr Verlosk 2001;26(11):87-99.
- 2 Buitendijk SE. Statistisch toetsen. Tijdschr Verlosk 2002;27(2):97-101.