

Les 5 Vergelijken van verdelingen

In de vorige les hebben we naar toetsen voor hypothesen gekeken, waarbij de hypothese een uitspraak over een parameter van een kansverdeling was, bijvoorbeeld over het gemiddelde of een relatieve frequentie. Maar als we bijvoorbeeld willen toetsen, of een dobbelsteen eerlijk is, zullen we na 120 worpen niet alleen maar het gemiddelde en de variantie bepalen, maar kijken of de getallen 1 t/m 6 alle ongeveer 20 keer gevallen zijn. Op deze manier zouden we natuurlijk onmiddellijk zien, dat de stochast X met

$$P(X = 1) = \frac{5}{24}, \quad P(X = 2) = \frac{1}{6}, \quad P(X = 3) = \frac{1}{12}, \\ P(X = 4) = 0, \quad P(X = 5) = \frac{13}{24}, \quad P(X = 6) = 0$$

geen eerlijke dobbelsteen beschrijft, terwijl $E[X] = 3\frac{1}{2}$ en $Var(X) = \frac{35}{12}$, net zo als bij een eerlijke dobbelsteen (ga dit na).

We zouden dus met toetsen op het gemiddelde en de variantie niet aan het licht kunnen brengen dat de dobbelsteen oneerlijk is, maar natuurlijk zouden we dit ook niet op zo'n stomme manier proberen te toetsen.

De vraag of een dobbelsteen eerlijk of oneerlijk is, is een voorbeeld van een vraagstuk, waar we niet alleen maar een parameter van een kansverdeling willen toetsen, maar waar we de volledige verdeling willen bekijken. De nulhypothese, die we in dit geval zouden toetsen is

$$H_0 : P(X = 1) = \frac{1}{6}, \quad P(X = 2) = \frac{1}{6}, \quad \dots, \quad P(X = 6) = \frac{1}{6}$$

en de alternatieve hypothese luidt, dat niet alle van deze kansen gelijk aan $\frac{1}{6}$ zijn.

Natuurlijk kunnen we niet verwachten, dat we bij een steekproef precies de kansen van de nulhypothese vinden, maar naarmate de steekproef groter wordt, zouden we steeds kleinere afwijkingen verwachten.

Het vergelijken van de onder de nulhypothese verwachte aantallen en de daadwerkelijk waargenomen aantallen geeft aanleiding tot een belangrijke klasse van toetsen voor hypothesen over kansverdelingen, namelijk de χ^2 -toetsen, die we in deze les gaan bekijken.

5.1 De χ^2 -aanpassingstoets

De situatie die we nu gaan bekijken is als volgt: Gegeven is een stochast X met een zekere kansverdeling, bijvoorbeeld de uniforme verdeling voor een eerlijke dobbelsteen. De nulhypothese luidt, dat een steekproef door de stochast X is voortgebracht en we willen toetsen of deze hypothese plausibel is.

De algemene aanpak is, de mogelijke uitkomsten van de stochast X in een aantal klassen in te delen. Voor een stochast met een discrete kansverdeling zijn de klassen vaak de verschillende mogelijke uitkomsten, maar soms is het handig verschillende uitkomsten in één klasse samen te vatten.

Voor continue kansverdelingen kiest men als klassen meestal intervallen, deze zijn vaak van dezelfde breedte, maar dit is niet noodzakelijk zo.

Voorbeeld: Voor een stochast $X \in \mathcal{N}(\mu, \sigma^2)$ waarvoor men een normale verdeling met verwachtingswaarde μ en variantie σ^2 veronderstelt, worden de intervalgrenzen vaak op veelvoudigen van de standaardafwijking σ vastgelegd. Men krijgt zo bijvoorbeeld de klassen

$$\begin{aligned} K_1 : -\infty < X < \mu - 3\sigma, & \quad K_2 : \mu - 3\sigma \leq X < \mu - 2\sigma, \\ K_3 : \mu - 2\sigma \leq X < \mu - \sigma, & \quad K_4 : \mu - \sigma \leq X < \mu, \\ K_5 : \mu < X \leq \mu + \sigma, & \quad K_6 : \mu + \sigma \leq X < \mu + 2\sigma, \\ K_7 : \mu + 2\sigma \leq X < \mu + 3\sigma, & \quad K_8 : \mu + 3\sigma \leq X < \infty. \end{aligned}$$

Als de mogelijke uitkomsten van X in k klassen ingedeeld zijn, wordt voor iedere klasse de kans p_i bepaald, dat X een uitkomst in de i -de klasse produceert. Bij een steekproef van n stuks zullen we dan (onder de aanname van de nulhypothese) np_i waarden in de i -de klasse verwachten.

In het voorbeeld van de normale verdeling met 8 klassen kunnen we uit de standaard-normale verdeling de volgende kansen afleiden:

i	1	2	3	4	5	6	7	8
p_i	0.0013	0.0214	0.1359	0.3413	0.3413	0.1359	0.0214	0.0013

We beschrijven nu met een stochast X_i het aantal uitkomsten in een steekproef van n stuks, die in de i -de klasse vallen. Uit de verschillen van X_i en np_i moeten we nu een toets afleiden, die aangeeft of het plausibel is dat de steekproef volgens de veronderstelde kansverdeling is voortgebracht.

Afwijkingen van de verwachte aantallen

Voor het speciale geval van slechts 2 klassen hebben we dit probleem al eerder bekeken, in dit geval vallen de uitkomsten met kans p in de eerste klasse en met kans $q = 1 - p$ in de tweede klasse. Maar dit betekent, dat X de stochast van een Bernoulli-experiment met kans p is en de stochast X_1 die het aantal uitkomsten in de eerste klasse (het aantal successen bij n pogingen) beschrijft, is binomiaal verdeeld met parameters n en p . Evenzo is de stochast X_2 die het aantal uitkomsten in de tweede klasse (het aantal mislukkingen bij n pogingen) beschrijft, binomiaal verdeeld met parameters n en $q = 1 - p$.

De relatieve frequentie p van een binomiale verdeling hadden we in de vorige les getoetst, door X_1 op een (bij benadering) standaard-normale verdeling te transformeren, namelijk door

$$Z := \frac{X_1 - np}{\sqrt{np(1-p)}}.$$

Als Z standaard-normaal verdeeld is, heeft Z^2 een χ^2 -verdeling met 1 vrijheids-

graad en we kunnen Z^2 als volgt herschrijven:

$$\begin{aligned} Z^2 &= \frac{(X_1 - np)^2}{np(1-p)} = (1-p) \frac{(X_1 - np)^2}{np(1-p)} + p \frac{(X_1 - np)^2}{np(1-p)} \\ &= \frac{(X_1 - np)^2}{np} + \frac{((n - X_1) - n(1-p))^2}{n(1-p)} \\ &= \frac{(X_1 - np)^2}{np} + \frac{(X_2 - nq)^2}{nq}. \end{aligned}$$

We zien dus dat we Z^2 kunnen beschrijven als som van de kwadratische afwijkingen tussen waargenomen aantallen in de twee klassen en verwachte aantallen voor deze klassen, waarbij de kwadratische afwijkingen op de verwachte aantallen genormeerd worden.

In plaats van de waarde van Z met de z -waarden van de standaard-normale verdeling te vergelijken, kunnen we de waarde van Z^2 tegen de waarden χ_α^2 van een χ^2 -verdeling met 1 vrijheidsgraad toetsen die gedefinieerd zijn door

$$P(Z^2 > \chi_\alpha^2) = \alpha$$

want er geldt $P(Z^2 > \chi_\alpha^2) = P(Z > z_\alpha) = \alpha$.

De veralgemening van 2 tot k klassen is nu enigszins voor de hand liggend: De gekwadrateerde afwijkingen van de waargenomen aantallen van de verwachte aantallen worden door de verwachte aantallen gedeeld en deze hoeveelheden worden voor de verschillende klassen bij elkaar opgeteld. Het idee achter de normering op het aantal verwachte uitkomsten in een klasse is dat bij een verwacht aantal van 100 uitkomsten een afwijking van 3 minder sterk weegt dan bij een verwacht aantal van 10 uitkomsten.

Definitie: Bij een kansexperiment met k mogelijke (klassen van) uitkomsten zij p_i de kans op een uitkomst in de i -de klasse. Het aantal van uitkomsten in de i -de klasse bij n pogingen wordt door de stochast X_i beschreven. De afwijking tussen de waargenomen verdeling en de verwachte verdeling wordt beschreven door de stochast χ^2 die gedefinieerd is door

$$\chi^2 := \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} = \frac{(X_1 - np_1)^2}{np_1} + \dots + \frac{(X_k - np_k)^2}{np_k}.$$

De naam χ^2 voor deze stochast is natuurlijk met opzet gekozen, er laat zich aantonen dat χ^2 voor $n \rightarrow \infty$ inderdaad een χ^2 -verdeling met $k - 1$ vrijheidsgraden heeft.

Voor het geval $k = 2$ hebben we dit boven ingezien, want we hebben aangetoond dat

$$\frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = \left(\frac{X_1 - np_1}{\sqrt{np_1(1-p_1)}} \right)^2$$

en het laatste heeft voor $n \rightarrow \infty$ inderdaad een χ^2 -verdeling met 1 vrijheidsgraad. Het bewijs voor algemene k vergt behoorlijk meer moeite en wordt hier onderdrukt.

We geven wel een iets handigere manier aan om χ^2 uit te rekenen:

Uit $(X_i - np_i)^2 = X_i^2 - 2X_i np_i + n^2 p_i^2$ volgt dat $\frac{(X_i - np_i)^2}{np_i} = \frac{X_i^2}{np_i} - 2X_i + np_i$.

We hebben $\sum_{i=1}^k p_i = 1$ en omdat de som van de X_i het totaal aantal n van waargenomen uitkomsten aangeeft, geldt $\sum_{i=1}^k X_i = n$. Hiermee krijgen we

$$\begin{aligned}\chi^2 &:= \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} = \sum_{i=1}^k \frac{X_i^2}{np_i} - \sum_{i=1}^k 2X_i + \sum_{i=1}^k np_i \\ &= \left(\sum_{i=1}^k \frac{X_i^2}{np_i} \right) - 2n + n = \left(\sum_{i=1}^k \frac{X_i^2}{np_i} \right) - n.\end{aligned}$$

De kansverdeling die de verdeling van n uitkomsten over k klassen beschrijft, waarbij een uitkomst met kans p_i in de i -de klasse valt, heet de *multinomiale verdeling* met parameters p_1, \dots, p_k (die aan $p_1 + \dots + p_k = 1$ moeten voldoen). Er geldt

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

waarbij $n_1 + \dots + n_k = n$ is. De multinomiale verdeling voor het speciale geval $k = 2$ is natuurlijk juist de binomiale verdeling.

Toets op de afwijkingen

Het idee van een toets op de afwijkingen tussen waargenomen en verwachte aantallen, de zogeheten χ^2 -aanpassingstoets of kort χ^2 -toets, is in principe hetzelfde als bij de toetsen die we in de vorige les hebben gezien:

Voor de verschillende aantallen ν van vrijheidsgraden en de verschillende levels α van onbetrouwbaarheid worden waarden $\chi_{\nu, \alpha}^2$ bepaald zo dat

$$P(\chi^2 > \chi_{\nu, \alpha}^2) = \alpha.$$

Onder de aanname van de nulhypothese geeft een steekproef dus (slechts) met kans α een χ^2 -waarde die zo groot of groter is dan $\chi_{\nu, \alpha}^2$ en de nulhypothese wordt verworpen als een waarde χ^2 wordt gevonden die groter is dan $\chi_{\nu, \alpha}^2$ voor de gekozen level α .

Vaak wordt ook in het kader van χ^2 -toetsen de P -waarde van χ^2 bepaald, dus de kans waarmee de stochast X van de nulhypothese een steekproef produceert die een χ^2 -waarde heeft die groter is dan de gevonden waarde χ^2 .

Merk op: Een belangrijke voorwaarde voor de toepasbaarheid van de χ^2 -toets is, dat voor iedere klasse de verwachte aantallen $np_i \geq 5$ zijn, want anders wordt de verdeling van de χ^2 -waarden niet nauwkeurig genoeg door een χ^2 -verdeling benaderd. Dit eist soms dat men klassen samenvoegt die anders te weinig waarnemingen laten verwachten.

In het voorbeeld van de normale verdeling heeft de klasse K_1 de verwachte relatieve frequentie $p_1 = 0.0013$: Om hier op $np_1 \geq 5$ te komen, moeten we een steekproef van grootte $n \geq 3847$ hebben. Als dit niet haalbaar is, kunnen we bijvoorbeeld de klassen K_1 en K_2 samenvoegen, de gecombineerde kans voor deze twee klassen is $p'_1 = 0.02275$ en om nu aan de voorwaarde $np'_1 \geq 5$ te voldoen is al een steekproef van grootte $n \geq 220$ voldoende.

Voorbeeld: We nemen aan dat we voor onze oneerlijke dobbelsteen met kansen $(\frac{5}{24}, \frac{1}{6}, \frac{1}{12}, 0, \frac{13}{24}, 0)$ bij een steekproef met $n = 120$ worpen precies de juiste aantallen vinden, dus $(25, 20, 10, 0, 65, 0)$. Bij een eerlijke dobbelsteen is $p_1 = \dots = p_6 = \frac{1}{6}$ en we zouden dus voor elke klasse 20 uitkomsten verwachten. De waarde voor χ^2 is in dit geval

$$\begin{aligned}\chi^2 &= \frac{(25 - 20)^2}{20} + \frac{(20 - 20)^2}{20} + \frac{(10 - 20)^2}{20} + \frac{(0 - 20)^2}{20} + \frac{(65 - 20)^2}{20} + \frac{(0 - 20)^2}{20} \\ &= \frac{1}{20}(25 + 0 + 100 + 400 + 2025 + 400) = 147.5.\end{aligned}$$

Voor $\alpha = 0.01$ vindt men in de tabellen voor een χ^2 -verdeling met 5 vrijheidsgraden de waarde $\chi^2_{5,0.01} = 15.1$ en zelfs voor $\alpha = 0.001$ is $\chi^2_{5,0.001} = 20.5$ veel kleiner dan de gevonden waarde voor χ^2 . De P -waarde voor $\chi^2 = 147.5$ is in feite $4.5 \cdot 10^{-30}$ dus is het nagenoeg uitgesloten dat een resultaat met zo'n grote waarde voor χ^2 toevallig door een eerlijke dobbelsteen opgeleverd zou worden.

Voorbeeld: Van een bepaalde plantensoort komen volgens de wetten van Mendel vier variaties voor in de verhouding $9 : 3 : 3 : 1$. De verwachte relatieve frequenties zijn dus $p_1 = \frac{9}{16}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{3}{16}$ en $p_4 = \frac{1}{16}$. In een steekproef van 160 exemplaren vindt men de volgende aantallen n_i , de met de verwachte aantallen np_i vergeleken worden:

	variatie				totaal
	1	2	3	4	
n_i	88	35	24	13	160
np_i	90	30	30	10	160

Omdat de verdeling 4 klassen bevat, hebben we de kritieke waarden van de χ^2 -verdeling met 3 vrijheidsgraden nodig. Voor $\alpha = 0.1$ is $\chi^2_{3,0.1} = 6.25$ en voor $\alpha = 0.05$ is $\chi^2_{3,0.05} = 7.81$. Als waarde voor χ^2 krijgen we

$$\chi^2 = \frac{(88 - 90)^2}{90} + \frac{(35 - 30)^2}{30} + \frac{(24 - 30)^2}{30} + \frac{(13 - 10)^2}{10} \approx 2.98$$

dus geeft dit experiment niet eens op een onbetrouwbaarheidslevel van 10% evidentie tegen de wetten van Mendel. De P -waarde van $\chi^2 = 2.98$ is 0.395, dit betekent dat 39.5% van de steekproeven minstens een χ^2 -waarde van 2.98 zou opleveren, dus is onze steekproef zeker geen atypisch resultaat.

Tweezijdige χ^2 -toetsen

Meestal wordt de χ^2 -aanpassingstoets als rechtséénzijdige toets toegepast, die aangeeft wat de kans is dat een steekproef in het geval van de nulhypothese een

zo grote χ^2 -waarde geeft. Er zijn echter ook gevallen waarbij een tweezijdige χ^2 -toets toegepast wordt, omdat men steekproeven ook verdacht vindt, als ze *te goed* bij de nulhypothese passen.

Een voorbeeld hiervoor is het toetsen van een implementatie van een *pseudo-randomgenerator* die toevalsgetallen moet voortbrengen. Voor toevalsgetallen tussen 0 en 1 kan men als klassen bijvoorbeeld de deelintervallen van lengte 0.1 kiezen. Als een toevalsgenerator nu 10000 toevalsgetallen produceert, zou men ongeveer 1000 getallen in ieder deelinterval verwachten en men berekent hiervoor de waarde van χ^2 .

Natuurlijk mag χ^2 in dit geval niet te groot zijn, omdat dit evidentie tegen de nulhypothese geeft dat de toevalsgenerator onbevooroordeeld (uniform verdeeld) is.

Maar omgekeerd geeft een te kleine χ^2 -waarde aanleiding tot de aanname dat er te veel regelmaat in de toevalsgetallen zit en de rij toevalsgetallen voorspelbaar is. Dit is evidentie tegen de nulhypothese dat de toevalsgenerator de getallen onafhankelijk van elkaar produceert. Men zou in dit geval de toevalsgenerator als ongeschikt verwerpen als de χ^2 -waarde niet tussen $\chi_{0.05}^2$ en $\chi_{0.95}^2$ ligt.

Een van de grondleggers van de statistiek, R.A. Fisher, heeft de χ^2 -toets op de experimenten van Gregor Mendel met erwten toegepast, waardoor deze tot de ontdekking van de genen werd geleid (zonder ze zo te noemen). Fisher kwam tot het resultaat dat χ^2 een P -waarde van 0.99996 had, dus slechts 4 in 100000 steekproeven zouden een zo kleine χ^2 -waarde opleveren. Het lijkt erop dat Mendel's tuin assistent precies wist, welke uitslag Mendel bij zijn experimenten verwachtte en hier een handje bij heeft geholpen.

De waarden $\chi_{\nu,\alpha}^2$

De $\chi_{\nu,\alpha}^2$ -waarden zijn net zo als de z -waarden en t -waarden voor verschillende parameters ν en α in tabellen opgeslagen of worden door software pakketten berekend. Voor grotere aantallen van vrijheidsgraden zijn er zekere benaderingen die op het verband van de χ^2 -verdeling met de normale verdeling berusten.

- (1) Voor een stochast χ^2 met een χ^2 -verdeling met ν vrijheidsgraden is

$$Z := \sqrt{2\chi^2} - \sqrt{2\nu - 1}$$

bij benadering standaard-normaal verdeeld, waarbij deze benadering zeker voor $\nu > 100$ toegepast mag worden. Door dit naar χ^2 op te lossen, volgt dat men $\chi_{\nu,\alpha}^2$ met behulp van de z_α -waarden kan benaderen door

$$\chi_{\nu,\alpha}^2 \approx \frac{1}{2} (z_\alpha + \sqrt{2\nu - 1})^2.$$

- (2) Een betere benadering krijgt men uit het feit dat ook

$$Z := \frac{\sqrt[3]{\frac{\chi^2}{\nu}} - (1 - \frac{2}{9\nu})}{\frac{2}{9\nu}}$$

bij benadering standaard-normaal verdeeld is. Oplossen hiervan naar χ^2 geeft de benadering

$$\chi_{\nu,\alpha}^2 \approx \nu \cdot \left(1 - \frac{2}{9\nu} + z_\alpha \sqrt{\frac{2}{9\nu}} \right)^3.$$

Er wordt soms aangegeven dat de benadering (1) voor $\nu > 100$ toegepast mag worden en de betere benadering (2) voor $\nu > 30$, maar met deze grenzen speelt men zeker aan de veilige kant.

Voor $\nu = 50$ en $\alpha = 0.05$ is bijvoorbeeld de juiste waarde $\chi_{50,0.05} = 67.5048$, benadering (1) geeft $\chi_{50,0.05} \approx 67.2189$ en benadering (2) $\chi_{50,0.05} \approx 67.5006$. Zelfs voor $\nu = 10$ en $\alpha = 0.05$ is de fout van de twee benaderingen nog klein, de juiste waarde is hier $\chi_{10,0.05} = 18.3070$, benadering (1) geeft $\chi_{10,0.05} \approx 18.0225$ en benadering (2) $\chi_{10,0.05} \approx 18.2918$.

Verschillende kritieke waarden $\chi_{\nu,\alpha}^2$ zijn in Tabel 3 te vinden. Merk op dat in deze tabel het aantal vrijheidsgraden met n (in plaats van ν) aangegeven is en dat de kritieke waarden in de vorm $\chi_{n,1-\alpha}^2$ aangegeven zijn, d.w.z. de waarde $\chi_{2,0.05}^2$ vindt men bijvoorbeeld in de kolom onder $1 - \alpha = 0.95$.

Voor aantallen van vrijheidsgraden die niet in de tabel genoteerd zijn, kan men (voor voldoende grote ν) de boven aangegeven benaderingen toepassen, of een waarde voor een hoger aantal vrijheidsgraden kiezen, die wel genoteerd is. Op deze manier wordt in ieder geval de kans op een type I fout niet vergroot.

Onbekende parameters

In veel gevallen wil men toetsen of een steekproef door een stochast met een zeker *type* van kansverdeling geproduceerd is, bijvoorbeeld met een binomiale verdeling of een normale verdeling. In dit geval hangt de verdeling voor de nulhypothese van onbekende parameters af die uit de steekproef geschat moeten worden. Bij een schatter voor het gemiddelde van een kansverdeling hebben we gezien dat door het vervangen van de variantie door een schatting de verdeling breder wordt, omdat er meer onzekerheid in de schatting zit. We moesten daarom de normale verdeling door de Student-*t* verdeling vervangen.

Iets soortgelijks gebeurt ook bij de χ^2 -toetsen. Als we de parameters van de verdeling waarmee we de verwachte kansen p_i voor de klassen berekenen door schattingen vervangen, passen we de kansen p_i in feite al aan de steekproef aan. Hierdoor wordt de afwijking tussen waargenomen en verwachte aantallen kleiner tegenover het geval van bekende parameters. Op een gegeven onbetrouwbaarheidsniveau α moeten de kritieke waarden vanaf waar we de nulhypothese verwerpen dus scherper gekozen worden.

Chi-kwadraatverdeling. Waarden van $\chi^2_{1-\alpha}[n]$

n	1 - α =														n
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995	.999	
1	-	-	.001	.004	.016	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88	10.8	1
2	.010	.020	.051	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	13.8	2
3	.072	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	16.3	3
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	18.5	4
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	20.5	5
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5	22.5	6
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3	24.3	7
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0	26.1	8
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6	27.9	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2	29.6	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8	31.3	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3	32.9	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8	34.5	13
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3	36.1	14
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8	37.7	15
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3	39.3	16
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7	40.8	17
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2	42.3	18
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6	43.8	19
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0	45.3	20
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4	46.8	21
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8	48.3	22
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2	49.7	23
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6	51.2	24
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9	52.6	25
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	54.1	26
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	55.5	27
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	56.9	28
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	58.3	29
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	59.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	73.4	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	86.7	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	91.6	99.6	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	112.3	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	124.8	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	137.2	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	149.4	100

Tabel 3: Kritieke waarden $\chi_{n,1-\alpha}$ voor de χ^2 -verdelingen met n vrijheidsgraden.

Gelukkig laat zich bewijzen dat de aanpassing van de kritieke waarden op een overzichtelijke manier gebeurt, er moet namelijk voor elke parameter die we uit de steekproef schatten één vrijheidsgraad afgetrokken worden. Er geldt:

Stelling: Als voor het berekenen van de verwachte kansen p_i voor een uitkomst in de i -de klasse r parameters voor de kansverdeling van X met een maximum likelihood schatting worden bepaald, dan heeft $\chi^2 := \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i}$ voor $n \rightarrow \infty$ een χ^2 -verdeling met $k - 1 - r$ vrijheidsgraden.

Merk op: Voor het gemiddelde μ van een verdeling is de maximum likelihood schatting gewoon het steekproefgemiddelde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ en voor de parameter p van een binomiale verdeling is $\bar{p} = \frac{k}{n}$ de maximum likelihood schatting, waarbij k het aantal successen bij n pogingen is.

Aan de andere kant geldt dat de maximum likelihood schatting voor de variantie niet de steekproefvariantie $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ is, maar $\frac{1}{n} \sum_{i=1}^n (x_i -$

$$\bar{x})^2 = \frac{n-1}{n} s^2.$$

Maar omdat de verdeling van χ^2 toch alleen maar voor $n \rightarrow \infty$ een χ^2 -verdeling geeft, maakt het niet zo veel uit of we de variantie σ^2 door de (zuivere) schatting s^2 of door de asymptotisch zuivere maximum likelihood schatting $\frac{n-1}{n} s^2$ vervangen. Vaak wordt daarom in de literatuur ook alleen maar aangegeven, dat een parameter door *een* schatting wordt vervangen, maar niet of door de maximum likelihood schatting of door een andere schatting.

Voorbeeld: Om het uur worden uit een productieproces steekproeven genomen van 5 stuks en het aantal defecte stukken wordt genoteerd. In 200 zulke steekproeven zijn de volgende resultaten gevonden:

aantal defecte stukken	0	1	2	3	4	5
aantal steekproeven	104	58	26	8	4	0

We willen toetsen of het aantal defecte stukken een binomiale verdeling heeft omdat dit het geval zou zijn als de kans op defecte stukken over de tijd constant gebleven is. Omdat de parameter p van de binomiale verdeling niet bekend is, moeten we deze uit de steekproeven schatten. We krijgen hiervoor

$$\bar{p} = \frac{1}{1000}(104 \cdot 0 + 58 \cdot 1 + 26 \cdot 2 + 8 \cdot 3 + 4 \cdot 4 + 0 \cdot 5) = \frac{150}{1000} = 0.15.$$

Als indeling van de steekproeven in klassen kiezen we de aantallen defecte stukken in een steekproef (van 5 stuks). De verwachte relatieve frequentie p_i voor de i -de klasse (met i defecte stukken) is dan volgens de binomiale verdeling met parameters $m = 5$ en $p = \bar{p} = 0.15$ gegeven door

$$p_i = \binom{m}{i} \bar{p}^i \cdot (1 - \bar{p})^{m-i} = \binom{5}{i} 0.15^i \cdot 0.85^{5-i}$$

en voor de $n = 200$ steekproeven krijgen we als verwachte aantallen voor de klassen

defect	0	1	2	3	4	5
p_i	0.444	0.392	0.138	0.024	0.002	0.0001
np_i	88.74	78.30	27.64	4.88	0.43	0.02

Omdat de verwachte aantallen voor de klassen met 3, 4 en 5 defecte stukken te klein zijn, voegen we deze samen tot één klasse met ≥ 3 defecte stukken. We krijgen zo de volgende statistiek waarvoor we de χ^2 -waarde moeten bepalen:

defect	0	1	2	≥ 3
n_i	104	58	26	12
np_i	88.74	78.30	27.64	5.32

Omdat we de parameter \bar{p} van de binomiale verdeling uit de steekproeven hebben geschat, heeft de χ^2 -verdeling $4 - 1 - 1 = 2$ vrijheidsgraden. Op de levels $\alpha = 0.05$ en $\alpha = 0.01$ hebben we de kritieke waarden $\chi_{2,0.05}^2 = 5.99$ en $\chi_{2,0.01}^2 = 9.21$. Er geldt nu

$$\chi^2 = \frac{(104 - 88.74)^2}{88.74} + \frac{(58 - 78.30)^2}{78.30} + \frac{(26 - 27.64)^2}{27.64} + \frac{(12 - 5.32)^2}{5.32} \approx 16.37$$

dus kunnen we de nulhypothese van een binomiale verdeling zelfs op de onbetrouwbaarheidslevel $\alpha = 0.01$ veilig verwerpen. De P -waarde van $\chi^2 = 16.37$ is in feite 0.0003, een veel te lage waarde voor de aanname dat de afwijking van de binomiale verdeling toevallig is. We zouden dus concluderen, dat de kans p op defecte stukken in het productieproces over de tijd niet constant was.

5.2 χ^2 -toets voor contingentietabellen

We hebben met de χ^2 -aanpassingstoets getoetst of een steekproef bij een zekere kansverdeling past. Vaak komt men echter een iets andere vraag tegen, namelijk of twee of meer steekproeven bij een gemeenschappelijke kansverdeling horen, waarbij het niet nodig is deze gemeenschappelijke verdeling nader te bepalen. Dit probleem wordt meestal met een variatie van de χ^2 -toets uit de vorige sectie aangepakt, waarbij men de verwachte aantallen uit de steekproeven bepaald. Hierbij gebruikt men een *contingentietabel*.

Stel we hebben r steekproeven met omvangen n_1, \dots, n_r . Ieder van de steekproeven wordt op k klassen verdeeld, dit geeft de aantallen n_{ij} van elementen in de i -de steekproef, die in de j -de klasse vallen. We krijgen zo een $r \times k$ -matrix met als elementen de hoeveelheden van elementen in de doorsnede van een steekproef en een klasse en deze matrix noemen we een *contingentietabel*.

Met $n := \sum_{i=1}^r n_i = n_1 + \dots + n_r$ noteren we de gemeenschappelijke omvang van alle steekproeven. We definiëren nu

$$p_j := \frac{n_{1j} + \dots + n_{rj}}{n}$$

als kans voor een uitkomst in de j -de klasse, dit is juist de relatieve frequentie van uitkomsten die in de j -de klasse vallen, bepaald over alle steekproeven.

Met de kansen p_j krijgen we als verwachte waarde op positie (i, j) in de contingentietabel de waarde $n_i \cdot p_j$, want dit is het aantal uitkomsten in de j -de klasse die we bij een steekproef van omvang n_i zouden verwachten. We vatten nu de cellen van de contingentietabel als nieuwe klassen op en berekenen voor deze klassen de χ^2 -waarde, dus

$$\chi^2 := \sum_{i=1}^r \sum_{j=1}^k \frac{(n_{ij} - n_i \cdot p_j)^2}{n_i \cdot p_j}.$$

Er laat zich ook in dit geval aantonen, dat χ^2 voor $n \rightarrow \infty$ een χ^2 -verdeling heeft, en het aantal vrijheidsgraden van deze χ^2 -verdeling is

$$\nu = (r - 1)(k - 1).$$

Dit kunnen we als volgt inzien: Als de p_j bekend waren, hadden we voor iedere steekproef $k - 1$ vrijheidsgraden, dus in het geheel $r(k - 1)$ vrijheidsgraden. Maar omdat we de p_j uit de steekproeven schatten, moeten we hiervan $k - 1$ aftrekken (niet k , want p_k laat zich door $p_k = 1 - p_1 - \dots - p_{k-1}$ uit de andere schattingen berekenen). Dit geeft dus $\nu = r(k - 1) - (k - 1) = (r - 1)(k - 1)$ vrijheidsgraden.

Voorbeeld: Bij een enquête in drie steden A , B en C werd een contigentietabel met de volgende resultaten gevonden:

stad	voor	tegen	neutraal	geen antwoord	totaal
A	105	61	87	167	420
B	118	60	130	145	453
C	88	58	62	101	309
totaal	311	179	279	413	1182

We hebben dus

$$n_1 = 420, \quad n_2 = 453, \quad n_3 = 309, \quad n = 1182,$$

$$p_1 = \frac{311}{1182} \approx 0.263, \quad p_2 = \frac{179}{1182} \approx 0.151, \quad p_3 = \frac{279}{1182} \approx 0.236, \quad p_4 = \frac{413}{1182} \approx 0.349$$

en dit geeft als tabel met de verwachte aantallen $n_i \cdot p_j$:

stad	voor	tegen	neutraal	geen antwoord
A	110.5	63.6	99.1	146.8
B	119.2	68.6	106.9	158.3
C	81.3	46.8	72.9	108.0

Als we nu de waarde van χ^2 berekenen, zijn de cellen van de tabellen de nieuwe klassen en we krijgen

$$\chi^2 = \frac{(105 - 110.5)^2}{110.5} + \frac{(61 - 63.6)^2}{63.6} + \dots + \frac{(101 - 108.0)^2}{108.0} \approx 17.2.$$

Dit moeten we vergelijken met de kritieke waarden van de χ^2 -verdeling met $(3 - 1) \cdot (4 - 1) = 6$ vrijheidsgraden. We hebben $\chi_{6,0.05}^2 = 12.6$ en $\chi_{6,0.01}^2 = 16.8$, dus zijn de resultaten van de drie steden op de level $\alpha = 0.01$ significant verschillend.

In het geval van $r = 2$ steekproeven hebben we natuurlijk al eerder toetsen op verschillen van de verdelingen gezien, bijvoorbeeld toetsen op hetzelfde gemiddelde. Het hangt vaak van de vraagstukken af, of een χ^2 -toets hier beter geschikt zou zijn. In het algemeen is de χ^2 -toets minder scherp dan een toets op verschillen van de gemiddelden, aan de andere kant kan deze ook nog verschillen detecteren als de gemiddelden wel overeenkomen. In het bijzonder is de χ^2 -toets ook toepasbaar, als de veronderstelling van een normaal verdeelde schatter niet meer houdbaar is.

Voorbeeld: Bij een niet nader toegelicht experiment met mogelijke uitslagen $1, \dots, 10$ worden met twee verschillende methoden I en II de volgende aantallen uitslagen bereikt:

methode	1	2	3	4	5	6	7	8	9	10	totaal
I	6	16	22	38	44	30	18	12	8	6	200
II	2	6	12	22	29	30	21	16	8	4	150
totaal	8	22	34	60	73	60	39	28	16	10	350

Als geschatte kansen p_j voor de uitkomsten krijgen we

j	1	2	3	4	5	6	7	8	9	10
p_j	0.023	0.063	0.097	0.171	0.209	0.171	0.111	0.080	0.046	0.029

en als we hiermee de χ^2 -waarde berekenen, krijgen we $\chi^2 \approx 11.12$. Voor een χ^2 -verdeling met $(2-1) \cdot (10-1) = 9$ vrijheidsgraden hebben we $\chi_{9,0.1} = 14.7$, dus geeft de χ^2 -toets met onbetrouwbaarheid $\alpha = 0.1$ geen evidentie voor een verschil van de twee methoden. De P -waarde van $\chi^2 = 11.12$ is 0.268.

Maar we kunnen met onze kennis uit de vorige les natuurlijk ook toetsen, of de twee methoden hetzelfde gemiddelde hebben. Hiervoor kijken we naar de steekproefgemiddelden \bar{x}_I en \bar{x}_{II} en de steekproefvarianties s_I^2 en s_{II}^2 voor de twee steekproeven met omvang $n_I = 200$ en $n_{II} = 150$. We hebben

$$\bar{x}_I = \frac{1}{200}(6 \cdot 1 + \dots + 6 \cdot 10) = 5.05, \quad \bar{x}_{II} = \frac{1}{150}(2 \cdot 1 + \dots + 4 \cdot 10) = 5.67$$

$$s_I^2 = 4.29, \quad s_{II}^2 = 3.86$$

en hieruit bepalen we als waarden voor de gepoolde variantie s^2 en standaardafwijking s :

$$s^2 = \frac{(n_I - 1)s_I^2 + (n_{II} - 1)s_{II}^2}{n_I + n_{II} - 2} = \frac{199 \cdot s_I^2 + 149 \cdot s_{II}^2}{348} = 4.11, \quad s = 2.03.$$

Als t -waarde die we met de kritieke waarden van de Student- t verdeling met 348 vrijheidsgraden moeten toetsen, hebben we

$$t = \frac{\bar{x}_{II} - \bar{x}_I}{s \sqrt{\frac{1}{n_I} + \frac{1}{n_{II}}}} \approx 2.82.$$

De verdeling van t is nagenoeg een standaard-normale verdeling en als P -waarde voor $t = 2.82$ vinden we 0.0024, dus vinden we met deze toets een significant verschil voor de gemiddelden van de twee methoden.

Toets op onafhankelijkheid van kenmerken

Een variatie op het vergelijken van r steekproeven geeft een toets op onafhankelijkheid van twee kenmerken in een steekproef. Bijvoorbeeld wil men weten, of het interesse in verschillende studievakken onafhankelijk is van het geslacht van de student. Men interpreteert nu de studenten van de verschillende studievakken als verschillende steekproeven en de indeling vrouw/man als indeling in klassen. De nulhypothese is, dat de kenmerken studievak en geslacht onafhankelijk zijn, in dit geval zou de kansverdeling voor iedere steekproef hetzelfde zijn en we zijn terug bij de situatie van de vorige sectie.

Voor het gemak nemen we aan dat het eerste kenmerk de waarden $\{1, \dots, r\}$ kan hebben en het tweede kenmerk de waarden $\{1, \dots, k\}$. Als n elementen in de steekproef zitten, noteren we met n_{ij} het aantal elementen met waarde i voor het eerste kenmerk en waarde j voor het tweede kenmerk. Als schatting

p_{i*} voor de relatieve frequentie van elementen met waarde i voor het eerste kenmerk krijgen we

$$p_{i*} := \frac{n_{i1} + \dots + n_{ik}}{n}$$

en als schatting p_{*j} voor de relatieve frequentie van elementen met waarde j voor het tweede kenmerk krijgen we

$$p_{*j} := \frac{n_{1j} + \dots + n_{rj}}{n}.$$

De kansen p_{i*} en p_{*j} heten ook *marginale kansen*, omdat ze met de totale aantallen corresponderen die we aan de rand van de contingentietabel schrijven.

Onder de aanname van de nulhypothese zijn de twee kenmerken onafhankelijk, dus is de kans op een uitkomst in de cel (i, j) van de contingentietabel gelijk aan $p_{i*} \cdot p_{*j}$ en het verwachte aantal uitkomsten voor deze cel is dus $n \cdot p_{i*} \cdot p_{*j}$.

Om de kenmerken op onafhankelijkheid te toetsen, kijken we dus in dit geval naar de χ^2 -waarde

$$\chi^2 := \sum_{i=1}^r \sum_{j=1}^k \frac{(n_{ij} - np_{i*}p_{*j})^2}{np_{i*}p_{*j}}.$$

Net als boven laat zich aantonen dat de verdeling van deze waarden χ^2 over alle mogelijke steekproeven voor $n \rightarrow \infty$ een χ^2 -verdeling heeft. Omdat het schatten van de p_{i*} uit de steekproef $r - 1$ vrijheidsgraden wegneemt en het schatten van de p_{*j} het aantal vrijheidsgraden om $k - 1$ reduceert, hebben we het met een χ^2 -verdeling met

$$\nu = rk - 1 - (r - 1) - (k - 1) = (r - 1)(k - 1)$$

vrijheidsgraden te maken.

Voorbeeld: In een onderzoek werd getoetst of de prestaties van leerlingen in de vakken Engels en Wiskunde onafhankelijk van elkaar zijn. Men deelt de resultaten in 3 klassen, cijfers 6 en lager, cijfers 7 en 8 en cijfer 9 en 10.

Engels	Wiskunde			totaal
	≤ 6	7, 8	9, 10	
≤ 6	85	42	14	141
7, 8	38	163	47	248
9, 10	12	71	56	139
totaal	135	276	117	528

Hieruit krijgen we voor de marginale kansen:

$$p_{1*} = \frac{141}{528} = 0.267, \quad p_{2*} = \frac{248}{528} = 0.470, \quad p_{3*} = \frac{139}{528} = 0.263$$

$$p_{*1} = \frac{135}{528} = 0.256, \quad p_{*2} = \frac{276}{528} = 0.523, \quad p_{*3} = \frac{117}{528} = 0.222$$

Onder de aanname van de nulhypothese dat de twee kenmerken onafhankelijk zijn, zouden we voor de combinatie (i, j) van de kenmerken $n \cdot p_{i*} \cdot p_{*j}$ leerlingen in de steekproef verwachten. Dit geeft de verwachte waarden in de volgende tabel:

Engels	Wiskunde		
	≤ 6	7, 8	9, 10
≤ 6	36.1	73.7	31.2
7, 8	63.4	129.6	55.0
9, 10	35.5	72.7	30.8

We zien al dat dit behoorlijk afwijkt van de gevonden waarden. Als we hiervoor de χ^2 -waarde berekenen, krijgen we

$$\chi^2 = \frac{(85 - 36.1)^2}{36.1} + \frac{(42 - 73.7)^2}{73.7} + \dots + \frac{(56 - 30.8)^2}{30.8} \approx 145.8$$

terwijl we voor een χ^2 -verdeling met $(3 - 1) \cdot (3 - 1) = 4$ vrijheidsgraden op significantie level $\alpha = 0.001$ de waarde $\chi_{4,0.001}^2 = 18.5$ vinden. Het is dus duidelijk dat de resultaten in de twee vakken niet onafhankelijk van elkaar zijn.

Yates-correctie

In het speciale geval van een 2×2 contingentietabel wordt vaak de *Yates-correctie* toegepast, die rekening ermee houdt, dat in klassen ingedeelde gegevens discreet zijn, terwijl de χ^2 -verdeling een continue kansverdeling is. In het algemeen wordt de χ^2 -waarde met Yates-correctie bij l klassen met kansen p_1, \dots, p_l berekend door

$$\chi^2 := \sum_{i=1}^l \frac{(|X_i - np_i| - \frac{1}{2})^2}{np_i}$$

maar dit wordt eigenlijk alleen maar in het geval van 1 vrijheidsgraad toegepast, en dit is juist het geval voor $r = 2$ en $k = 2$.

De Yates-correctie heeft het effect dat de χ^2 -waarde die berekend wordt iets lager is dan zonder de correctie. Dit leidt ertoe dat de nulhypothese met Yates-correctie minder snel verworpen wordt dan zonder Yates-correctie.

Voor grote steekproeven maakt de Yates-correctie bijna geen verschil en inmiddels wordt soms aanbevolen, de Yates-correctie *niet* toe te passen. Als alle gevonden aantallen van de cellen klein zijn (bijvoorbeeld tussen 5 en 10 liggen) is het verstandig om de χ^2 -waarde met en zonder Yates-correctie te bepalen. Als de twee versies tot verschillende conclusies leiden (verwerpen van de nulhypothese bij de ene, niet verwerpen bij de andere), zou men de steekproef moeten vergroten om tot een duidelijke beslissing te kunnen komen.

Voorbeeld: In een proef wordt aan een groep van mensen met een bepaalde ziekte een nieuwe medicijn gegeven, terwijl een tweede groep met dezelfde ziekte een placebo krijgt. Er wordt nu gekeken hoe veel van de mensen binnen een bepaalde periode gezond zijn geworden.

	gezond	ziek	totaal
medicijn	75	25	100
placebo	65	35	100
totaal	140	60	200

Als marginale kansen krijgen we hieruit

$$p_{1*} = p_{2*} = \frac{100}{200} = 0.5 \quad \text{en} \quad p_{*1} = \frac{140}{200} = 0.7, \quad p_{*2} = \frac{60}{200} = 0.3.$$

De aanname van onafhankelijkheid betekent in dit geval dat de nieuwe medicijn hetzelfde effect heeft als het placebo. Omdat de groepen even groot zijn, zouden we onder de aanname van onafhankelijkheid verwachten dat in beide groepen $200 \cdot 0.5 \cdot 0.7 = 70$ mensen gezond worden en dat $200 \cdot 0.5 \cdot 0.3 = 30$ ziek blijven.

Zonder Yates-correctie krijgen we hieruit de χ^2 -waarde

$$\chi^2 = \frac{(75 - 70)^2}{70} + \frac{(25 - 30)^2}{30} + \frac{(65 - 70)^2}{70} + \frac{(35 - 30)^2}{30} \approx 2.38$$

en met Yates-correctie

$$\begin{aligned} \chi^2 &= \frac{(|75 - 70| - 0.5)^2}{70} + \frac{(|25 - 30| - 0.5)^2}{30} \\ &+ \frac{(|65 - 70| - 0.5)^2}{70} + \frac{(|35 - 30| - 0.5)^2}{30} \approx 1.93. \end{aligned}$$

In beide gevallen kunnen we de nulhypothese op onafhankelijkheid op een level van $\alpha = 0.1$ niet verwerpen, want voor een χ^2 -verdeling met 1 vrijheidsgraad vinden we $\chi_{1,0.1}^2 = 2.71$. De P -waarde zonder Yates-correctie is 0.123 en de P -waarde met Yates-correctie is 0.165 en dit zijn allebij geen afzonderlijk kleine waarden. Om aan te tonen dat de nieuwe medicijn wel een effect heeft, zijn dus verdere experimenten nodig.

2 × 2-tabellen

In het voorbeeld hierboven hebben we kunnen zien, dat bij een 2×2 -contingentietabel de tellers in de som voor χ^2 alle hetzelfde zijn (in het voorbeeld 5^2). Dit is geen toeval, maar in feite altijd het geval voor 2×2 -tabellen en heeft tot gevolg dat we voor dit belangrijke speciaal geval de χ^2 -waarde op een veel makkelijkere manier kunnen uitrekenen.

Het zal geen verrassing zijn, dat een 2×2 -tabel een speciaal geval is, want hier gaan we toetsen of twee relatieve frequenties hetzelfde zijn. In de vorige les hebben we gezien, dat we dit voor twee relatieve frequenties p_1 en p_2 kunnen doen, door de z -waarde

$$z := \frac{p_1 - p_2}{\sqrt{p_0(1 - p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

met $p_0 := \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ te berekenen, die onder de aanname van de nulhypothese $p_1 = p_2$ standaard-normaal verdeeld is. De waarde χ^2 voor de χ^2 -toets die we nu gaan berekenen is in dit speciaal geval juist het kwadraat van z .

We noteren de 2-contingentietabel als volgt:

	A	B	totaal
1	a	b	n_1
2	c	d	n_2
totaal	n_A	n_B	n

Hiervoor berekenen we de χ^2 -waarde door

$$\begin{aligned}
 \chi^2 &= \frac{(a - \frac{n_1 n_A}{n})^2}{\frac{n_1 n_A}{n}} + \frac{(b - \frac{n_1 n_B}{n})^2}{\frac{n_1 n_B}{n}} + \frac{(c - \frac{n_2 n_A}{n})^2}{\frac{n_2 n_A}{n}} + \frac{(d - \frac{n_2 n_B}{n})^2}{\frac{n_2 n_B}{n}} \\
 &= \frac{n}{n_1 n_2 n_A n_B} \left(n_2 n_B (a - \frac{n_1 n_A}{n})^2 + n_2 n_A (b - \frac{n_1 n_B}{n})^2 \right. \\
 &\quad \left. + n_1 n_B (c - \frac{n_2 n_A}{n})^2 + n_1 n_A (d - \frac{n_2 n_B}{n})^2 \right) \\
 &= \frac{n}{n_1 n_2 n_A n_B} \left(\frac{n_2 n_B}{n^2} (n a - n_1 n_A)^2 + \frac{n_2 n_A}{n^2} (n b - n_1 n_B)^2 \right. \\
 &\quad \left. + \frac{n_1 n_B}{n^2} (n c - n_2 n_A)^2 + \frac{n_1 n_A}{n^2} (n d - n_2 n_B)^2 \right).
 \end{aligned}$$

Dit ziet nog niet naar een verbetering uit, maar nu vullen we in dat $n = a + b + c + d$, $n_1 = a + b$, $n_2 = c + d$, $n_A = a + c$ en $n_B = b + d$. Dit geeft

$$\begin{aligned}
 n a - n_1 n_A &= (a + b + c + d)a - (a + b)(a + c) \\
 &= a^2 + ab + ac + ad - a^2 - ab - ac - bc = ad - bc =: \Delta.
 \end{aligned}$$

Op een soortgelijke manier zien we in, dat ook

$$n b - n_1 n_B = \Delta, \quad n c - n_2 n_A = \Delta, \quad n d - n_2 n_B = \Delta.$$

Dit is in feite het bewijs, dat we in de tellers van de termen voor χ^2 altijd hetzelfde getal vinden, namelijk $(\frac{\Delta}{n})^2$.

Als we nu nog invullen dat $n_1 + n_2 = n$ en $n_A + n_B = n$, zien we dat $n_2 n_B + n_2 n_A + n_1 n_B + n_1 n_A = n_2(n_B + n_A) + n_1(n_B + n_A) = (n_2 + n_1)n = n^2$ en daarom geldt

$$\frac{n_2 n_B}{n^2} \Delta^2 + \frac{n_2 n_A}{n^2} \Delta^2 + \frac{n_1 n_B}{n^2} \Delta^2 + \frac{n_1 n_A}{n^2} \Delta^2 = \Delta^2 = (ad - bc)^2.$$

Alles bij elkaar genomen, hebben we dus aangetoond dat

$$\chi^2 = \frac{n}{n_1 n_2 n_A n_B} (ad - bc)^2$$

en dit is voor 2×2 -contingencietabellen inderdaad veel handiger dan de algemene formule van boven.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- χ^2 -aanpassingstoets
- kritieke waarden $\chi^2_{\nu, \alpha}$

- χ^2 -toets bij onbekende parameters
- contingentietabel
- χ^2 -toets op onafhankelijkheid
- Yates-correctie

OPGAVEN

29. Er wordt 120 keer met een dobbelsteen geworpen. De aantallen voor de verschillende uitkomsten zijn:

$$1 : 12, \quad 2 : 21, \quad 3 : 27, \quad 4 : 22, \quad 5 : 20, \quad 6 : 18.$$

Is dit een zuivere dobbelsteen?

30. Bij een reukproef werd aan 50 willekeurig gekozen vrouwen gevraagd of zij parfum *A* lekkerder vonden dan *B* of omgekeerd. Aan *A* gaven 37 vrouwen de voorkeur, de overige vonden *B* lekkerder. Toets op de significantie level $\alpha = 0.1$ de nulhypothese dat er geen voorkeur voor één van de twee merken bestaat. Voer de toets zonder en met Yates-correctie uit.
31. In een weverij zijn in het verleden gemiddeld 2 weeffouten per $100m^2$ geweven doek opgetreden. Een recente steekproef op 100 stukken doek van $100m^2$ heeft het volgende resultaat opgeleverd:

fouten	0	1	2	3	4	5	6	7	8	9	10
aantal doeken	16	22	28	15	8	3	3	1	2	1	1

- (i) Toets op een significantie level van $\alpha = 0.05$ de nulhypothese dat het aantal fouten Poisson-verdeeld met parameter $\lambda = 2$ is.
- (ii) Toets op een significantie level van $\alpha = 0.05$ de nulhypothese dat het aantal fouten überhaupt Poisson-verdeeld is.
32. Van 1000 aselect gekozen personen is nagegaan of ze kleurenblind zijn. Van de 480 mannen bleken dit er 38 te zijn, bij de vrouwen was het aantal 6.
- (i) Toets op de level $\alpha = 0.1$ of kleurenblindheid onafhankelijk is van het geslacht.
- (ii) Wat is het minimale aantal vrouwen dat kleurenblind mag zijn, waarvoor de nulhypothese op level $\alpha = 0.1$ niet verworpen wordt (waarbij we nog steeds van 38 kleurenblinde mannen uit gaan)?
33. Twee groepen *A* en *B* van elk 100 patiënten hebben een bepaalde ziekte. Groep *A* wordt behandeld met een zeker serum, groep *B* met een ander serum. Na een bepaalde tijd zijn 75 patiënten van groep *A* en 65 patiënten van groep *B* genezen. Toets met onbetrouwbaarheid $\alpha = 0.05$ of beide sera evenveel effect hebben.
34. Bij een computerbedrijf wordt in 3 ploegen (ochtend, middag, nacht) op vier verschillende types van computers (*A*, *B*, *C*, *D*) gewerkt. De manager vraagt zich af of er bij het aantal reboots van computers een samenhang tussen de ploeg en de type computer bestaat. Hij heeft de volgende contingentietabel voor reboots gemaakt:

	type computer			
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
ochtend	5	3	2	7
middag	7	12	9	16
nacht	1	2	4	2

Wat kan hij op een onbetrouwbaarheidslevel van $\alpha = 0.05$ zeggen?