

χ^2 for gymnasielærere

FRANS MORVILLE, VoksenUddannelsescenter Frederiksberg

Nedenfor først en teoretisk begrundelse for χ^2 -teststørrelsens fordeling i en bestemt "2x2"-situation og bagefter (uafhængigt heraf) PC-simulationer af forskellige scenarier.

Ole Witt-Hansen udledte i LMFK nr 3 (maj 2011) en formel for χ^2 -fordelingernes sandsynlighedsfunktion (tæthedsfunktion). Man kan sige, at det er på grundlag af den definition, som Wikipedia giver:

If Z_1, \dots, Z_k are independent, standard normal random variables, then the sum of their squares,

$$Q = \sum_{i=1}^k Z_i^2$$

is distributed according to the chi-square distribution with k degrees of freedom.

en.wikipedia.org/wiki/Chi-square_distribution#Definition

Jeg vil nedenfor bruge dette ved 1 frihedsgrad:

Hvis Z er standardnormalfordelt, så gælder om den stokastiske variabel $Q = Z^2$, at den er χ^2 -fordelt med 1 frihedsgrad, og derfor har den følgende fordelingsfunktion:

$$F(t) = P(Q \leq t) = P(Z^2 \leq t) = P(-\sqrt{t} \leq Z \leq \sqrt{t}) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t})$$

χ^2 -tests bruges i flere forskellige forbindelser, og de kan ikke behandles ens matematisk. Jeg vil nedenfor begrunde rigtigheden af χ^2 -teststørrelsens fordeling i et enkelt – ret hyppigt forekommende – tilfælde på en måde, der er helt inden for matematiklæreres rækkevidde. Behandlingen knyttes til et opdigtet taleksempel, men generelle formler opstilles:

En ny medicin afprøves på $n_1 = 40$ patienter. $x_1 = 30$ patienter bliver raske.

Som sammenligningsgrundlag bruges en gammel medicin på $n_2 = 60$ patienter. $x_2 = 40$ af dem bliver raske.

Er den nye medicin bedre end den gamle?

Nulhypotese: Ny og gammel medicin har samme helbreds-sandsynlighed, p .

Under denne nulhypotese er nedenstående stokastiske variable X_1 og X_2 binomialfordelte med sandsynlighedsparameter p og antalsparametre hhv. n_1 og n_2 .

X_1 = antal raske ud af de n_1 , der fik ny medicin.

Middelværdi $n_1 \cdot p$ og varians $n_1 \cdot p \cdot (1 - p)$

X_2 = antal raske ud af de n_2 , der fik gammel medicin.

Middelværdi $n_2 \cdot p$ og varians $n_2 \cdot p \cdot (1 - p)$

Begge er tilnærmelsesvis normalfordelte. Det er også helbredsprocenterne blandt de undersøgte (i taleksemplet $30/40 = 0,75$ hhv. $40/60 = 0,667$).

Efter NY medicin:

$$\frac{X_1}{n_1} \text{ har middelværdi } p \text{ og varians } \frac{p \cdot (1 - p)}{n_1}$$

Efter GAMMEL medicin:

$$\frac{X_2}{n_2} \text{ har middelværdi } p \text{ og varians } \frac{p \cdot (1 - p)}{n_2}$$

Disse to stokastiske variable er uafhængige. Derfor er også differensen tilnærmelsesvis normalfordelt, og varianserne adderes: Forskellen i helbredsprocenterne er med ovennævnte tal $66,7\% - 75\% = -8,3\%$.

Betragtet som stokastisk variabel gælder om denne forskel:

$$\frac{X_2}{n_2} - \frac{X_1}{n_1} \text{ har middelværdi } p - p = 0$$

$$\text{og varians } \frac{p \cdot (1 - p)}{n_2} + \frac{p \cdot (1 - p)}{n_1} = \left(\frac{1}{n_2} + \frac{1}{n_1} \right) \cdot p \cdot (1 - p)$$

Vi dividerer med spredningen for at få en tilnærmelsesvis standard-normalfordelt stokastisk variabel:

$$Z = \frac{\frac{X_2}{n_2} - \frac{X_1}{n_1}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \cdot p \cdot (1 - p)}}$$

Som anført allerførst: Hvis man kvadrerer en standardnormalfordelt stokastisk variabel, får man en stokastisk variabel, der er χ^2 -fordelt med 1 frihedsgrad. Den kvadrerede størrelse, $Z^2 = z_{\text{teo}}^2$, er opskrevet og omskrevet nedenfor. Den må være tilnærmelsesvis χ^2 -fordelt med 1 frihedsgrad.

Når man blot har set forsøgsresultaterne (NY: 30 raske ud af 40; GAMMEL: 40 raske ud af 60), kender man ikke den sande værdi af basissandsynligheden p – hvis den overhovedet findes. Og det gør den kun, hvis nulhypotesen holder, dvs. hvis ny og gammel medicin er lige gode anvendt på en meget stor population af patienter.

Under nulhypotesen vil det bedste bud på en fælles helbreds-

$$\text{sandsynlighed være } p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{30 + 40}{40 + 60} = \frac{70}{100} = 0,70.$$

Med dette estimat indsat beregnes nedenstående udtryk for z_{estim}^2 :

$$z^2_{\text{teo}} = \frac{\left(\frac{x_2}{n_2} - \frac{x_1}{n_1}\right)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot p \cdot (1-p)} = \frac{-((n_1 \cdot x_2 - n_2 \cdot x_1)^2)}{p \cdot (p-1) \cdot n_1 \cdot (n_1 + n_2) \cdot n_2}$$

og med $p = \frac{x_1 + x_2}{n_1 + n_2}$ fås estimatet

$$z^2_{\text{estim}} = \frac{(n_1 + n_2) \cdot (n_1 \cdot x_2 - n_2 \cdot x_1)^2}{n_1 \cdot (n_1 + n_2 - x_1 - x_2) \cdot n_2 \cdot (x_1 + x_2)}$$

Med moderate stikprøvestørrelser som i eksemplerne ovenfor kan der sagtens optræde forskelle fra -20 % til +20 % mellem enkelte værdier af z^2_{teo} og z^2_{estim} . Men genererer man ved PC-simulation et par hundrede værdier af både z^2_{teo} og z^2_{estim} , er der ingen synlig forskel mellem de to seriers statistiske fordeling. Den χ^2 -størrelse, som vi plejer at beregne ved at summere 4 led af typen

$$\frac{(\text{observeret} - \text{forventet})^2}{\text{forventet}}$$

reduceres til nøjagtig den samme beregningsformel som z^2_{estim} ovenfor:

$$\begin{aligned} \text{chi}^2 &:= \frac{(x_1 - p \cdot n_1)^2}{p \cdot n_1} + \frac{(x_2 - p \cdot n_2)^2}{p \cdot n_2} + \frac{((n_1 - x_1) - (n_1 - p \cdot n_1))^2}{n_1 - p \cdot n_1} \\ &+ \frac{((n_2 - x_2) - (n_2 - p \cdot n_2))^2}{n_2 - p \cdot n_2} \quad \wedge \quad p = \frac{x_1 + x_2}{n_1 + n_2} \\ &= \frac{(n_1 + n_2) \cdot (n_1 \cdot x_2 - n_2 \cdot x_1)^2}{n_1 \cdot (n_1 + n_2 - x_1 - x_2) \cdot n_2 \cdot (x_1 + x_2)} \end{aligned}$$

Her har vi altså en forklaring på at χ^2 -teststørrelsen i dette simple 2 x 2-tilfælde er tilnærmelsesvis χ^2 -fordelt med 1 frihedsgrad. Forståelsen? Her er vi nok i samme båd som vores elever. En lang bogstavregning er måske nok et slags bevis, men kan den overbevise?

PC-simulationer

Ovenfor så vi på følgende konkrete eksempel: En ny medicin afprøves på $n_1 = 40$ patienter. $x_1 = 30$ patienter bliver raske. Som sammenligningsgrundlag bruges en gammel medicin på $n_2 = 60$ patienter. $x_2 = 40$ af dem bliver raske. Er den nye medicin bedre end den gamle?

Ud fra en antagelse om at x_1 og x_2 er uafhængige binomialfordelte stokastiske variable, begrundede jeg ad algebraisk vej, at den velkendte χ^2 -teststørrelse (med 4 led) er tilnærmelsesvis χ^2 -fordelt med 1 frihedsgrad, dvs. fordelt ligesom kvadratet på en standardnormaldelt stokastisk variabel.

Her vil jeg belyse dette og beslægtede ting ved PC-simulation, og her vil også vores elever kunne følge med. Jeg bryder mig ikke om ”omrøring” i dette tilfælde. Omrøring er ikke helt intuitiv, og den giver talresultater der adskiller sig en del fra χ^2 -fordelingen.

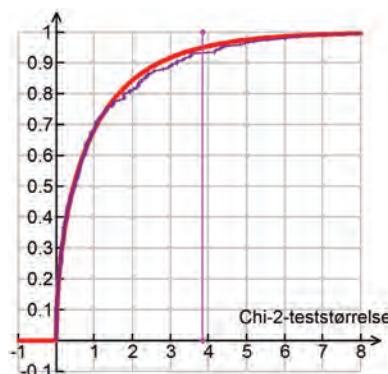
Jeg vil hellere tage udgangspunkt i, at det, der giver den tilfældige variation, er den stikprøveudtagning, der foregår ved at man udvælger 40 hhv. 60 personer af en stor population af patienter.

Hvis en anden reservelæge på tilfældig måde udvalgte og behandlede 40 og 60 andre patienter, ville hans antal raske sandsynligvis komme til at variere noget fra den første læges resultat. Måske kommer det til at se sådan ud:

Vi simulerer nu, at 200 læger foretager et tilsvarende udvalg af 40 + 60 patienter og undersøger dem efter behandling med hhv. ny og gammel medicin. I første omgang sættes sandsynligheden for helbredelse til 0,70 både ved NY og ved GAMMEL medicin, senere ændres dette.

”Antal raske” er simuleret i TI-Interactive med ordrene RandBin(60, 0.70) hhv. RandBin(40, 0.70). Hver af de 200 læger beregner en χ^2 -værdi ud fra sine observationer. Vi sorterer de 200 χ^2 -værdier efter størrelse, og afbilder dem i et diagram – en sumkurve, hvor χ^2 -værdien er afsat ud ad den vandrette akse, og man kan sige at man har de forskellige læger op ad den lodrette akse.

I diagrammet nedenfor er desuden indtegnet den teoretiske fordelingsfunktion for χ^2 -fordelingen med 1 frihedsgrad.



Læge 1	Raske	Syge
NY	30 raske	10 syge
GAMMEL	40 raske	20 syge

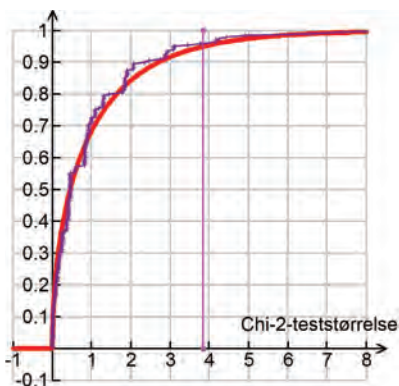
χ^2 -teststørrelse: 0,794

Læge 2	Raske	Syge
NY	33 raske	7 syge
GAMMEL	48 raske	12 syge

χ^2 -teststørrelse: 0,097

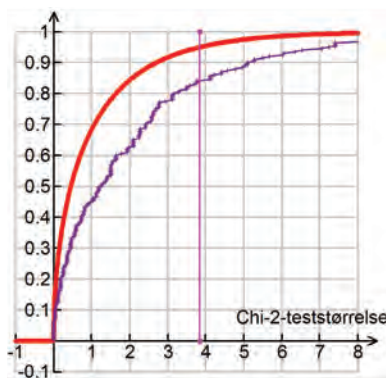
Ved signifikansniveau 5 % og 1 frihedsgrad er den kritiske værdi i χ^2 -fordelingen 3,84. I sumkurven ser vi, at 6 – 7 % af de 200 læger fik en χ^2 -værdi på mere end 3,84 og derfor forkastede nulhypotesen om, at ny og gammel medicin var lige gode. Det passer pænt med de 5%.

Man kan nu ændre forudsætningerne for simulationen. I stedet for 70% sættes helbredelsessandsynligheden fx til 10 % for både ny og gammel medicin. Det ændrer ikke billedet afgørende, dog ser man effekter af ”små” forventede værdier (under 5).



Anderledes bliver det, hvis vi gør nulhypotesen falsk i populationen. Fx sådan,

at 70 % bliver raske med den ny medicin, men kun 60 % med den gamle. De 200 læger får nu følgende resultater:

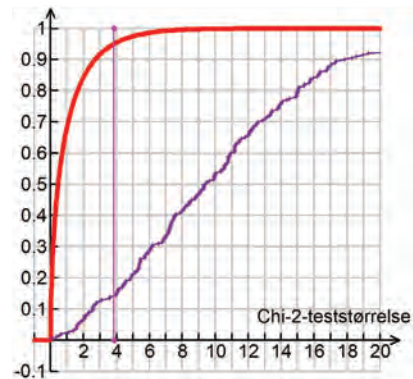


Ca. 17 % af de 200 læger vil skrive: *”På niveau 5 % har min undersøgelse vist, at den nye medicin giver større chance for helbredelse end den gamle medicin”*

De resterende 83 % af lægerne skriver stadig: *”Min undersøgelse giver ikke nogen anledning til at antage, at der er forskel i virkning mellem de to typer medicin”*.

Til sidst lader vi de 200 læger undersøge patientgrupper, der er 10 gange så store, stadig med 70 % helbredelsessandsynlighed for den ny medicin og kun 60 % for den gamle.

Hver af de 200 læger udvælger altså 400 patienter, som behandles med ny medicin, og 600 som behandles med den gamle:



Nu er det kun ca. 13 % , altså cirka 26 af lægerne, der svarer *”Ingen forskel påvist”*, mens de 87 % af lægerne skriver, at tilfældigheder alene har mindre end 5 % chance for at give så stor forskel i helbredelsesprocent mellem den nye og den gamle medicin: **Køb den nye!**

Læsere, der er interesserede i at simulere videre eller læse lidt uddybning, kan downloade en fil (TI-interactive eller Word) på fmorville.dk/LMFK-chi2/.