χ^2–test i matematikundervisningen – en replik

Bjørn Felsager, Midtsjællands Gymnasieskoler

Ole Witt–Hansen har i sidste nummer af LMFK–bladet endnu et indlæg, hvor han denne gang harcelerer over at *'regression og især χ^2–fordeling havde fundet indpas i matematikundervisningen samtidigt med, at sandsynlighedsregningen helt forsvandt fra matematikundervisningen efter reformen i 2005'*. Lad gå med, at han ikke har forstået reformen: Sandsynlighedsregning forsvandt ikke med reformen, men blev gjort til et obligatorisk supplerende emne, dvs. et emne, der ikke længere indgår i den skriftlige eksamen; men det indgår stadigvæk i den mundtlige eksamen. Det gav lærerne en betydelig frihed, fordi de nu selv kunne vælge to emner indenfor sandsynligheds-regning og statistik, som de fandt særligt relevante for deres undervisning. Denne frihed er ved justeringen af reformen i 2010 indskrænket til et enkelt obligatorisk supplerende emne, idet χ^2–test nu er gjort til kernepensum og dermed også til et emne for den skriftlige eksamen. Men desværre har Ole Witt–Hansen heller ikke forstået χ^2–testen. Det er længe siden jeg har læst en artikel i LMFK–bladet, der er fyldt med så mange elementære faglige fejl, mangler og misforståelser. Det er også bemærkelsesværdigt, at artiklen i vidt omfang ikke omtaler nogle af de centrale begreber, der indgår i pensum. Fx undgås fuldstændigt begreberne nulhypotese, frihedsgrader og p–værdi.

Af artiklens start fremgår det, at Ole Witt–Hansen er af den opfattelse, at *'man i matematik skal kunne forklare det, der står i bøgerne for eleverne, hvilket tidligere jo også havde den forudsætning, at læreren selv forstod det'*. Jeg har ingen ide om, hvem han tror, der kan være uenig med ham i disse to trivielle forudsætninger for undervisning, så uenigheden må bunde et andet sted: To ting er væsentlige for at forstå muligheden for en uenighed: Vendingen *'det, der står i bøgerne'* antyder, at pensum i matematik er lig med det, der står i bøgerne og ikke det, der står i læreplan og undervisningsvejledning. Men det må vel være lærerens ansvar, at der vælges bøger, hvor læreren kan forklare, det der står i bøgerne. Lige netop det kan vel ikke være et problem ved reformen som sådan? Det står jo lærerne frit for at vælge undervisningsmaterialer herunder selv at skrive noter, hvis de ikke er tilfredse med bogen. Dernæst er der lærerens *forståelse* af emnet. Selvfølgelig skal læreren kunne forstå det, der undervises i. Det er der ikke på nogen måde ændret ved som følge af reformen.

Derudover er det såvel muligt som realistisk, at vi er uenige om, hvad det vil sige at forstå et matematisk emne. Jeg er fx ikke så fikseret på formler som Ole Witt–Hansen, men lægger mere fokus på begrebsforståelsen.

Ole Witt–Hansen tager udgangspunkt i normalfordelingen. Det er fint nok, for normalfordelingen er et klassisk emne, vi har undervist i på mange niveauer – også uden, at det har været nødvendigt for eleverne at kende formlen for tæthedsfordelingen og også uden, at mange lærere har været præsenteret for udledningen af tæthedsfordelingen i deres universitetsstudier. Hvis man skal følge Ole Witt–Hansens logik burde man derfor vel heller ikke undervise i normalfordelingen? Derefter følger et langt universitetsbevis for tæthedsfordelingen for χ^2–fordelingen. Bevisets udgangspunkt er desværre helt forkert, men selv om det rettes op er det tvivlsomt, om det egner sig til undervisning på A–niveau. Men for lærere, der kan huske teorien for multidimensionale integraler og de tilhørende Jacobideterminanter, giver det altså mulighed for at udlede formlen. Giver det dem så en bedre forståelse for χ^2–fordelingen? Jeg tvivler.

Men lad os lige se på, hvorfor beviset kikser: Det grundlæggende argument går på, at den samlede fordeling er et produkt at de enkelte tæthedsfordelinger:

$$P(x_1, x_2, \ldots, x_n)\cdot dx_1 \cdot dx_2 \cdot \ldots \cdot dx_n = \prod \frac{1}{\sqrt{2\pi}\sigma_i} \cdot e^{-\frac{\left(x_i-\bar{x}\right)^2}{2\sigma_i}} dx_i$$

Men det må man selvfølgelig kun, hvis de involverede stokastiske variable er uafhængige. Og lige præcis her går det galt: Ole Witt–Hansen betragter n normalfordelte målinger $x_1, x_2, \ldots, x_n$, hvor spredningerne $\sigma_1, \sigma_2, \ldots, \sigma_n$ er kendte. Det er forvirrende, at vi intet får oplyst om middelværdierne, men man må vel gå ud fra, at der er tale om en fælles middelværdi μ, da vi jo vel forsøger at måler den samme størrelse? Disse n målinger $x_1, x_2, \ldots, x_n$ repræsenterer derfor værdierne for n uafhængige stokastiske variable. Så langt er alt godt.

Men derefter betragter Ole–Witt Hansen de stokastiske variable

$$q_i = \frac{x_i - \bar{x}}{\sigma_i}$$

og de er *ikke* uafhængige, eftersom de jo tydeligvis opfylder den lineære sammenhæng (summen af residualerne er nul)

$$\begin{aligned}\sigma_1 \cdot q_1 + \sigma_2 \cdot q_2 + \ldots + \sigma_n \cdot q_n \\ = \quad (x_1 - \bar{x}) + \ldots + (x_n - \bar{x}) = 0\end{aligned}$$

Dermed er det på forhånd klart, at man ikke kan tillade sig at gange de enkelte tæthedsfordelinger sammen og resten af udledningen frem til tætheds-funktionen for χ^2 er derfor noget vrøvl. Tæthedsfordelingen for $(q_1, q_2, \ldots, q_n)$ er da heller ikke som påstået en χ^2–fordeling med n frihedsgrader. Man kan dog vise, omend det er langt fra trivielt, at der i stedet er tale om en χ^2–fordeling med $n-1$frihedsgrader, idet vi mister én frihedsgrad på grund af den lineære sammenhæng. Kun hvis Ole Witt–Hansen i stedet tager udgangspunkt i de stokastiske variable

$$q_i = \frac{x_i - \mu}{\sigma_i}$$

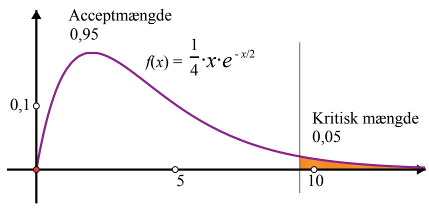
kan det skitserede bevis gennemføres, som det står.

Derefter kommer så følgende svada: *'χ^2 anvendes som bekendt til at vurdere, hvorvidt resultatet af N målinger ligger indenfor den statistiske usikkerhed'*. Nu ved jeg ikke, hvor bekendt dette resultat er, men det forbigår fuldstændig at χ^2–fordelingen blev indført med et helt andet sigte omkring 1900 af statistikeren Karl Pearson, nemlig for at kunne foretage et Goodness–of–fit test af *kategoriske* variable, hvilket er noget ganske andet end vurderingen af middelværdien for en *numerisk* variabel. Senere blev det inddraget i Fishers teori for statistisk signifikans, som et af flere klassiske test, der nu kunne udføres som hypotesetest. Ole Witt–Hansen kan selvfølgelig gennemgå χ^2–testen som det passer ham, men det forplumrer unægtelig diskussionen, at han vælger at fokusere på et perifert χ^2–test, som oven i købet intet har at gøre med de χ^2–test, vi indfører i gymnasiet.

Ole Witt–Hansen afslutter sin diskussion af χ^2–testen med følgende bemærkninger:

> Fordelingsfunktionen for χ^2 er kendt som chi–square fordelingsfunktionen. Sandsynligheden for at få en værdi af χ^2, som ikke overstiger χ_0^2 er givet ved:
>
> $$P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} F(\chi^2)d\chi^2$$
>
> Denne funktion er tabuleret og kan i øvrigt findes på en CAS. Hvis χ^2 er lig med 0, fordi alle observationer er lig med middelværdien, så er sandsynligheden $P = 1$. Jo større $P(\chi^2 > \chi_0^2)$ er, jo bedre er observationerne statistisk set. Skal man foretage en test med et signifikansniveau på 5%, er acceptbetingelsen altså at
>
> $$P(\chi^2 > \chi_0^2) > 0{,}95$$
> $$\Leftrightarrow \; P(\chi^2 < \chi_0^2) < 0{,}05$$
>
> Især det sidste har jeg erfaret giver vanskeligheder, når det skal forklares til eleverne – afsluttende med *"Sådan er det bare"*, men sådanne "forklaringer" synes jeg, man bør overlade til andre fag end matematik.

Jeg har prøvet på at læse det ovenstående adskillige gange, men er bange for, at jeg stadigvæk ikke har fattet pointen, så jeg kan sagtens forstå eleverne, der har haft vanskeligheder med at forstå den ovenstående sorte tale. I den første sætning er der vel et 'ikke' for meget? En værdi af χ^2, som ikke overstiger χ_0^2 svarer vel til uligheden $\chi^2 < \chi_0^2$? I den næste sætning 'Hvis χ^2 er lig med 0, skal der vel stå χ_0^2? Jeg ved ikke, hvad der menes med 'jo bedre er observationerne statistisk set', men det er vel noget i retning af, at jo mindre måleusikkerheden er, jo bedre er forsøget? Nu handlede forsøget jo om gentagne målinger af en størrelse, hvor spredningen σ er kendt (men i øvrigt kan variere fra måling til måling – øh?). Hvis det er tilfældet er vi da lige så bekymrede, hvis vores målinger ligger alt for tæt på middelværdien, som hvis de ligger for langt fra middelværdien! Der mangler også en eksplicit nulhypotese (når man begynder at diskutere en test uden at forklare, hvad det er for en hypotese man tester, er det ganske enkelt dårlig skik), men det må vel være den oprindelige antagelse om, at $x_1, x_2, \ldots, x_n$ er normalfordelte med kendte spredninger og – formoder jeg – en fælles middelværdi μ, som vi ønsker at estimere? Under alle omstændigheder er det eneste, vi kan opnå med *en statistisk test*, at forkaste nulhypotesen, hvilket er noget ganske andet end en vurdering af 'hvorvidt resultatet af N målinger ligger indenfor den statistiske usikkerhed'. Endelig er der acceptbetingelsen. Det ville have hjulpet meget, hvis Ole Witt–Hansen havde tilføjet en illustration. Hvis vi ser på grafen for en typisk χ^2–tæthedsfordeling (her med fire frihedsgrader), så svarer acceptbetingelsen til, at teststørrelsen er lille og ligger indenfor de første 95%:



Acceptbetingelsen må vel derfor svare til uligheden: $P(\chi^2 < \chi_0^2) < 0{,}95$, hvilket er det modsatte af Ole Witt–Hansens konklusion?

Herefter vender Ole Witt–Hansen sig mod den χ^2–test, man anvender i samfundsfag og biologi. Men da han sammenblander det med det χ^2–test, der anvendes i fysik, er det klart det går helt galt. Han nævner fx formlen for de enkelte led i χ^2, der angives på formen

$$\frac{(y_i - x_i)^2}{x_i}$$

hvor x_i er det observerede antal. Det er helt elementært forkert: I Pearsons formel for χ^2–teststørrelsen er de enkelte led givet ved formlen

$$\frac{\left(observeret_i - forventet_i\right)^2}{forventet_i}$$

dvs. man dividerer med den forventede værdi og netop ikke med den observerede værdi. Og når Pearson netop *ikke* dividerer med kvadratet på spredningen er det med velberåd omhu! Det er nemmest at forklare i et simpelt tilfælde. Hvis en kategorisk stokastisk variabel kun antager to værdier er der tradition for at kalde disse to værdier succes og fiasko. Sandsynligheden for succes betegnes traditionelt med p. Sandsynligheden for fiasko er da tilsvarende givet ved $1 - p$. Den stokastiske variabel X, der angiver antallet af succeser ved n gentagne målinger af værdien for den kategoriske variabel, er derfor binomialfordelt med antalsparameter n og sandsynlighedsparameter p. Vi kan derfor opstille den følgende antalstabel for de n gentagne målinger:

Kategori	Observeret værdi	Forventet værdi
Succes	X	$n \cdot p$
Fiasko	$n - X$	$n \cdot (1 - p)$

Pearsons teststørrelse er derfor givet ved

$$\chi^2 = \sum_{i=1}^{2} \frac{\left(observeret_i - forventet_i\right)^2}{forventet_i}$$
$$= \frac{(X - n \cdot p)^2}{n \cdot p} + \frac{\left((n - X) - n \cdot (1 - p)\right)^2}{n \cdot (1 - p)}$$

Ved hjælp af elementær algebra omskrives det på formen

$$\chi^2 = \frac{(X - n \cdot p)^2}{n \cdot p \cdot (1 - p)}$$

Men ligesom middelværdien for en binomialfordeling er givet ved $n \cdot p$ er kvadratet på spredningen givet ved $n \cdot p \cdot (1 - p)$. Vi har derfor fundet formlen

$$\chi^2 = \frac{(X - \mu)^2}{\sigma^2} = \left(\frac{X - \mu}{\sigma}\right)^2$$

I præcis det samme omfang, som man ifølge den centrale grænseværdisætning kan approksimere tæthedsfordelingen for binomialfordelingen med tæthedsfordelingen for normalfordelingen (med den samme middelværdi og spredning), kan man altså approksimere tæthedsfordelingen for Pearsons teststørrelse med kvadratet på standardnormalfordelingen med middelværdi 0 og spredning 1. Men det er præcis definitionen på χ^2–fordelingen med 1 frihedsgrad. Argumentet viser derfor, hvordan den eksakte binomialtest kan erstattes af en *approksimativ* χ^2–test.

Hvis det var hele historien, var der selvfølgelig slet ingen grund til at indføre χ^2–testen! Men pointen er nu, at for kategoriske variable med k kategorier kan den approksimative χ^2–test umiddelbart generaliseres (idet antallet af frihedsgrader denne gang bliver $k - 1$) i modsætning til multinomialtesten, der stadigvæk selv med brug af slagkraftige computere ikke kan udregnes eksakt. Ydermere kan den approksimative χ^2–test udvides til også at kunne teste uafhængighed af to kategoriske variable, hvilket Pearson selv opdagede umiddelbart efter han havde fundet Goodness–of–fit testen.

I artiklen hiver Ole Witt–Hansen i stedet Poissonfordelingen ind på scenen, hvor den så absolut ikke hører hjemme, idet han skriver: *'Som beskrevet men ikke redegjort for i nogle matematikbøger anvender man imidlertid spredningen for Poissonfordelingen … Dette er den eneste forklaring jeg kan give på de χ^2–test, man anvender i samfundsfag og biologi, og som er refereret i matematikbøger (uden forklaring – heller ikke for læreren), …'*

Nu er det jo svært at vide præcis hvilke matematikbøger Ole Witt–Hansen har i tankerne, men da χ^2–testen er nyt pensum, kan man selvfølgelig ikke forvente at finde en detaljeret gennemgang i standardlærebøgerne til brug for gymnasieundervisningen. I forbindelse med justeringen af reformen har undervisningsministeriet derfor foranlediget, at der blev udarbejdet eksemplarisk undervisningsmateriale til brug for χ^2–testen. Susanne Christensen fra Aalborg Universitet har på opfordring af fagkonsulenten Bjørn Grøn skrevet noten *At træffe sine valg i en usikker verden*, der gennemgår de centrale begreber der ligger til grund for χ^2–testen, og en gruppe gymnasielærere har på baggrund af denne note udarbejdet en supplerende undervisningsnote, det såkaldte kursusmateriale. Disse materialer er frit stillet til rådighed på EMU'en, så enhver gymnasielærer kan hente materialet og læse det i fred og ro, se på:

emu.dk/gym/fag/ma/undervisnings-forloeb/statistik/index.html

Jeg kan ikke umiddelbart genkende noget af materialet i Ole Witt–Hansens gennemgang, så han har åbenlyst valgt at ignorere tilbuddet. Det er da heller ikke sikkert, at han vil sympatisere med den faglige tilgang i disse materialer. Der optræder ingen integraler og der trækkes ikke på avanceret sandsynlighedsregning. I stedet lægges hovedvægten på at forstå de centrale begreber bag statistiske test. Her følger Susanne Christensen en hæderkronet tradition: Den moderne testteori er udviklet i 1920'erne af Ronald Fisher og fremstillet i den epokegørende lærebog *Statistical Methods for Research Workers* fra 1925. I denne fremstilling har Fisher netop fravalgt den avancerede matematik, selv om han faktisk var en af de førende eksperter indenfor multidimensionale integraler og fx står bag udledningen af en række centrale tæthedsfordelinger i avanceret statistik. I stedet har han lagt hovedvægten på begrebsforståelsen og illustreret diskussionerne med instruktive eksempler. Det er et bevidst (og meget heldigt) pædagogisk valg, for det har betydet en enorm udbredelse af hans begreber og metoder indenfor andre fag som biologi og samfundsfag.

Man kan selvfølgelig stille spørgsmål ved, om der så stadigvæk er tale om matematik, når man nedtoner sætninger og beviser så voldsomt indenfor et givet emne. Det kan man i sagens natur ikke svare entydigt på, eftersom der ikke findes et entydigt svar på, hvad matematik er for noget, og afgrænsningen varierer meget alt efter hvilket land og hvilken undervisningstradition, man kigger på. Men hvis vi holder os til den angelsaksiske verden, kan man fx tage udgangspunkt i den nyligt udkomne bog *'Landmarks Writing in Western Mathematics 1640–1940'* (Elsevier 2005), der rummer en gennemgang af 77 epokegørende artikler og bøger fra den nævnte periode fra Descartes *Geometri* (1649) til Hilbert og Bernays *Grundlagen der Mathematik* (1934–39). I dette oversigtsværk har de også valgt at inddrage tre gennembrudsværker om statistik: Gauss artikel fra 1807 om himmelmekanikken, hvor han introducerer mindste kvadraters metode og normalfordelingen, Pearsons artikel fra 1900, hvor han introducerer χ^2–testen som et Goodness–of–fit test og Fishers bog *Statistical Methods for Research Workers*, hvor han introducerer den moderne hypotesetest med nulhypotese, signifikansniveau osv. Blandt de 77 værker, der blev fundet afgørende for matematikkens udvikling i den moderne vestlige matematik, finder vi også D'Arcy Thompsons bog *On Growth and form*, Volterras bog om matematiske metoder i biologi samt Shewharts artikel om *Economic Control of Quality of manufactured products*. Det er altså ikke alle matematikere, der afgrænser matematik så snævert som Ole Witt–Hansen.

Ole Witt–Hansen slutter sin artikel med følgende svada: *'Efter reformen står der en del i lærebøgerne i fysik og matematik, som ikke længere bliver forklaret for eleverne, blandt andet på grund af manglende forudsætning. Og det kan man jo have sin mening om. At de bliver undervist i formler, som læreren heller ikke forstår, kan man vist kun have én mening om.'*

Jeg ved ikke hvem Ole Witt–Hansen forestiller sig kan være uenige i synspunktet, at der skal undervises i formler, som læreren selv forstår. I tilfældet med χ^2–testen, drejer det sig om et område indenfor sandsynlighedsregning og statistik, der som vi har set er snævert forbundet med binomialfordelinger og normalfordelinger, som vi har mangeårige og gode erfaringer med at undervise i. Jeg har svært ved at forstå, hvorfor det så lige netop skulle være så meget sværere at undervise i χ^2–testen på en forsvarlig måde? Det kan gøres på mange forskellige måder: Susanne Christensen har anvist en vej ind i emnet, der bygger på grundlæggende statistik. Andre vil foretrække en vej ind i emnet, der bygger på grundlæggende sandsynlighedsregning og fx tager udgangspunkt i binomialfordelingstesten og normalfordelingsapproksimationen. Det står dem helt frit – i det første tilfælde vil de så stadigvæk mangle et valgfrit emne indenfor sandsynlighedsregning og statistik, i det andet tilfælde vil man kunne 'slå to fluer med et smæk' idet man som det valgfri emne netop kan benytte binomialfordelingen og normalfordelingen. Ole Witt–Hansen har peget på en tredje vej ind i emnet via usikkerhedsteori i fysiske målinger. Det står ham frit for at følge denne vej, og der er utvivlsom mange andre måder at gøre det på. Det er helt op til den enkelte lærer. Og hvis læreren insisterer, kan han også som Ole Witt–Hansen introducere formlen for tæthedsfordelingen for χ^2–fordelingen som universitetsformel

$$F(\chi^2)d\chi^2 = \frac{1}{2^{\frac{N}{2}}\Gamma\left(\frac{N}{s}\right)} e^{-\frac{1}{2}\cdot\chi^2} \left(\chi^2\right)^{\frac{N}{2}-1} d\chi^2$$

Men for to frihedsgrader er det altså bare eksponentialfunktionen

$$\frac{1}{2}\cdot e^{-x/2}$$

og for fire frihedsgrader er der tale om funktionen med forskriften

$$\frac{1}{4}\cdot x\cdot e^{-x/2}$$

så mere indviklet er det altså ikke. Og på A–niveau findes der mange andre måder at udlede den generelle formel på, end den af Ole Witt–Hansen foreslåede. Men som sagt er det altså ikke noget krav, at man inddrager forskriften for tæthedsfordelingen.

Hvis man ikke på forhånd kender til χ^2–testen, kan man selvfølgelig godt føle sig usikker på, hvordan man skal håndtere emnet. Udover at få fat i nogle af de undervisningsmaterialer, der er lagt ud til fri afbenyttelse, kan det da også være en god ide at arrangere et inspirationskursus på skolen om det nye statistikpensum. Vil man gerne have hjælp udefra til afholdelsen af et sådant kursus har matematiklærerforeningen nu et helt år tilbudt et sådant kursus, et tilbud som tyve gymnasier allerede har taget imod. Til næste skoleår er der udbudt en ny runde skolebaserede statistikkurser, så man vil kunne få inspiration til at komme i gang med støtte fra en kollega fra et nabogymnasium, se fx kursusbeskrivelsen på:

lmfk.dk/tilmelding/rekkurser/1/index.php