

Hvordan undgås det,

at undervisningen i χ^2 -test bliver black-box?

JAN B. SØRENSEN, Aalborg Studenterkursus

Som censor til mundtlig eksamen skal man lægge øre til meget sludder – ikke mindst, når det kommer til forståelse af begreber som frihedsgrader, kritisk værdi, p -værdi osv. i forbindelse med χ^2 -test. Det er sådan set ikke overraskende, for det er svære begreber, og χ^2 -fordelingen er jo praktisk umulig at forklare teoretisk til bunds på gymnasieniveau.

Men – hvad kan man så gøre i stedet, så eleverne alligevel får en forståelse af begreberne? Det vil jeg give mit bud på i denne artikel.

Først og fremmest – glem alt om CAS og om χ^2 -fordelingen til at starte med. I stedet mener jeg, at det rigtige sted at starte er simulering ved håndkraft. Det følgende er således tænkt som de første lektioner i emnet.

1. Opstil et scenarie, fx: Et blomsterfirma sælger store sække med 3 slags blomsterfrø. De påstår, at 1/2 af blomsterne bliver hvide, 1/3 bliver blå og 1/6 røde. Et biologihold ønsker at undersøge dette, så de planter 36 frø (som alle spirer). Det giver konkret 12 hvide, 15 blå og 9 røde blomster. Det stemmer jo ikke med 1/2, 1/3 og 1/6, men ...
2. Debatter med eleverne, hvor følgende begreber kan og bør komme på banen: Population, stikprøve, nulhypotese, forventet fordeling, observerede værdier og forventede værdier.
3. Helt naturligt kommer spørgsmålet undervejs, eller måske nærmere helt fra starten, om 12, 15, 9 er langt fra den forventede fordeling eller ej, og dermed kommer snakken om teststørrelsen. Lad endeligt være med straks at give eleverne formelen for, hvordan vi har tænkt os at regne denne, men læg i stedet op til en åben debat om, hvordan man kan måle "afstanden" mellem forventet og observeret. Det kunne jo være gjort på mange måder.
4. Når I så er blevet enige om formelen til at beregne teststørrelsen, så melder spørgsmålet sig – er den størrelse stor og dermed langt fra eller lille og dermed tæt på? Det giver anledningen til at finde en kasse med terninger frem (mange terninger) og tale om simulering.
5. Giv grupper af 2 – 3 elever 36 terningerne hver, og tal grundigt om, at de nu skal simulere antal hvide, blå og røde blomster under antagelse af, at nulhypotesen er sand. Understreg det sidste meget nøje, da det jo er hele ideen, så brug tid på at tale om, hvad det betyder.
6. Lad så eleverne kaste de 36 terninger, dele dem ind i 1 – 3 (hvide), 4 – 5 (blå) og 6 (røde), optælle de derved simulerede observerede værdier og udregne den simulerede teststørrelse. Anvend begreberne meget rundhåndet i hele forløbet – formålet er netop, at eleverne kommer til at kende og forstå disse.
7. Lav et fælles dokument, som alle kan se via projektor, og få eleverne til at indskrive de udregnede simulerede teststørrelser. Tal sammen med klassen om det første sæt af værdier, når alle grupper har lavet én simulering. Tal om den naturlige spredning, der vil være i disse tal, og sammenlign med teststørrelsen for observationen 12, 15, 9.
8. Fortsæt simuleringerne indtil der samlet er 100 simulerede teststørrelser. Det tager 15 – 20 min, og eleverne synes ofte, at det er lidt sjovt at se, om de kan få meget store værdier.
9. Bed så eleverne om at sige, hvad de nu synes om teststørrelsen for de observerede værdier kontra teststørrelserne for de simulerede observerede værdier under antagelse af nulhypotesen. Der vil nok være elever, som foreslår at sortere de 100 simulerede teststørrelser, ellers må man som lærer selv foreslå dette. De sorterede værdier giver så anledningen til at tale om, hvilke værdier er kritiske for hypotesen (store værdier) og især om p -værdien. Tæl ganske enkelt antal simulerede teststørrelser, der er lige så store eller større end den observerede teststørrelse.
10. Så er det tid til at tale om signifikansniveau, at forkaste eller ikke forkaste hypotesen og dermed fejl af type 1 (og måske lidt om type 2), hvilket fører til kritisk område og kritisk værdi. Her ser eleverne naturligvis blot på den 5. største simulerede teststørrelse, hvis signifikansniveauet er sat til 5 %.
11. Foreslå så eleverne, at de tegner et histogram/pindediaagram over de 100 simulerede observerede værdier, og tal så om, hvad teststørrelse, p -værdi, signifikansniveau, kritisk værdi og kritisk område er rent grafisk.
12. Giv eleverne noget kode til at simulere på computer, og lav samme simulering bare med 10.000 gentagelser. Find igen de simulerede værdier for p -værdi og kritisk værdi, grupper data og tegn et histogram. Maple kode hertil kan hentes her: tinyurl.com/hfw5bbo
13. Forklar om χ^2 -fordelingen og tegn grafen for denne med 2 frihedsgrader sammen med histogrammet. Det skulle gerne give en ide om, at de to ting nogenlunde passer sammen, så vi måske kan bruge den fordeling i stedet for at simulere. Tal om, at hverken den simulerede fordeling eller χ^2 -fordelingen er eksakt korrekt, begge er approksimationer.
14. Det giver anledningen til at tale om antal frihedsgrader. Ændr i koden til simulering, så der kun er 2 slags blomster (1 frihedsgrad) og se, hvad der sker med histogrammet og grafen for χ^2 -fordelingen. Ændr ligeledes til 4 slags blomster (3 frihedsgrader). Tal om, hvorfor et større antal frihedsgrader intuitivt flytter fordelingen mod højre.
15. Hvis eleverne har haft om integralregning – eller måske endda alligevel – kan man så tale om, hvordan den kritiske værdi og p -værdien kan udregnes vha. integralregning.

I Maple kunne det se således ud:

```

forvP := [1/2, 1/3, 1/6]:
obs := [12, 15, 9]:
n := 12 + 15 + 9 = 36
forv := forvP · n = [18, 12, 6]
teststr := sum_{i=1}^3 (obs[i] - forv[i])^2 / forv[i] = 17/4
kritisk := fsolve(int_k^inf chipdf(2, x) dx = 0.05, k) = 5.991464547
pværdi := int_teststr^inf chipdf(2, x) dx = 0.1194329683

```

Illustrer dette grafisk i hånden og evt. på computer.

Hermed er hypotesen om fordelingen af blomsterfrøene testet – først vha. simulering i hånden, så vha. simulering på computer, og endelig vha. en praktisk begrundet anvendelse af χ^2 -fordelingen. Undervejs er de centrale begreber blevet introduceret og begrundet uden det er blevet (ret meget) black-box.

Find så 1 (eller flere) tidligere eksamensopgaver med Goodness-of-fit – og bed eleverne om i små grupper at anvende både begreber og metoder på disse. Det vil sige, at de skal

- Forklare, hvad der er population, stikprøve, nulhypotese, observerede værdier, forventet fordeling og forventede værdier.
- Udregne teststørrelsen.
- Simulere (på computer) under antagelse af, at nulhypotesen er sand, og derved finde den simulerede kritiske værdi og den simulerede p -værdi. Ud fra disse forkaste eller ikke forkaste hypotesen, og skrive hvad det betyder i opgavens kontekst.
- Tegne et histogram over en gruppering af de simulerede teststørrelser og ud fra dette forklare, hvad signifikansniveau, kritisk værdi, kritisk område, teststørrelse og p -værdi er grafisk.
- Udregne antal frihedsgrader og derefter kritisk værdi og p -værdi vha. χ^2 -fordelingen, og tegne grafen for denne fordeling. Ud fra denne graf igen forklare, hvad signifikansniveau, kritisk værdi, kritisk område, teststørrelse og p -værdi er. Desuden forkaste eller ikke forkaste hypotesen, og skrive, hvad det betyder i opgavens kontekst.

Senere, evt. meget senere, kan man så vise eleverne, at deres CAS værktøj har en black-box kommando, der laver meget af arbejdet for dem.

