

Test for Goodness of Fit og test for uafhængighed er basalt set absurde

JAN B. SØRENSEN, Aalborg Studenterkursus

Lad os sige, at vi ønsker at teste, om fordelingen i befolkningen i de tre nordiske lande er 10 % for, 70 % imod og 20 % ligeglade mht. et eller andet.

Vi udtager så en stikprøve, og hvis fordelingen i stikprøven ikke er for langt fra den forventede fordeling, så accepterer vi, at fordelingen i hele populationen godt kan være *perfekt*, den forventede fordeling.

Lad os nu sige, at der i populationen er 20 mio. +1 individer. Så skal der være 2 mio. + 0,1 individer for, men eftersom det ikke er muligt med 0,1 individ, kan det ikke give mening at teste, om fordelingen passer *matematisk perfekt*.

Vi må altså i stedet reelt mene, at vi tester, om fordelingen er *passende tæt på den forventede fordeling*, men uden at angive, hvad vi mener med *passende tæt på*. Men det giver et andet problem. Lad os sige, at vi underforstået mener 9,9 % – 10,1 % for.

Hvor stor en stikprøve er anvendelig? Under 50 giver et problem med tommelfingerreglen om, at alle de forventede værdier skal være mindst 5, og over 2 mio. giver et problem med

en tommelfingerregel om, at mere end 10 % af populationen begrænser variationsmulighederne, så man mere bevæger sig over i deskriptiv statistik. Inden for dette råderum er den gængse tanke, at en større stikprøve er bedre, så lad os tage en stikprøve på 1 mio. individer (i nogle sammenhænge ikke urealistisk i takt med automatisk dataopsamling).

For at udfordre GoF-testet tager vi nu en stikprøve på 1 mio. individer med 9,91 % for, 70,09 % imod og 20,00 % ligeglade og gennemfører χ^2 GoF-testet, her med Maple. Vi vil gerne have, at hypotesen accepteres, da vi jo er *passende tæt på fordelingen*, men testet giver en *p-værdi* på lige under 1 %, så på 1 % signifikansniveau forkastes hypotesen!

Helt tilsvarende argumenter kan gennemføres for χ^2 -test for uafhængighed.

Om man så mener, at dette gør testene decideret absurde, eller blot at man skal være meget opmærksom på, at et forkastet resultat ikke siger noget om, hvor stor afvigelsen er, og dermed heller ikke, om den har nogen som helst praktisk betydning, overlades til læseren. Blot opfordres hermed til, at man tænker sig godt om, hvis man har en stor stikprøve.

```
restart : with(Gym) :
n := 106 :
forv := [0.10, 0.70, 0.20] · n = [1.0000000 105, 7.0000000 105, 2.0000000 105]
obs := [0.0991, 0.7009, 0.20] · n = [99100.0000, 7.009000000 105, 2.0000000 105]
```

χ^2 -teststørrelse = 9.2571
Frihedsgrader = 2
Kritisk værdi = 5.9915
p-værdi = 0.0097687

ChiKvadratGOFtest(obs, forv) =

