Pearsons -test

# Goodness of fit

## To mulige udfald

Betragt ens og uafhængige Bernouilli-eksperimenter med sandsynlighed . Altså et binomial-eksperiment. Det forventede antal succeser (mulighed 1) er da , mens det tilsvarende forventede antal fiaskoer (mulighed 2) er . Variansen af begge er .

Hvis og er de observerede antal hhv. succeser og fiaskoer defineres teststørrelsen således:

Ved at indsætte fås:

Her er brugt, at . Reducer indholdet af anden tæller:

Altså i alt:

Når er stor vil indholdet af parentesen iflg. den centrale grænseværdisætning (under de sædvanlige antagelser) tilnærmelsesvist være standardnormalfordelt. Derfor er tilnærmelsesvis -fordelt med 1 frihedsgrad.

## mulige udfald

Betragt ens og uafhængige eksperimenter med udfald (1, 2 og 3) med tilhørende sandsynligheder . Da den samlede sandsynlighed er må der gælde . Den samlede fordeling af udfaldene udgør en multinomialfordeling. De forventede antal er . Varianserne for de tilhørende stokastiske variable er .

For hvert udfald vil vi desuden definere en stokastisk variabel svarende til det ’nde eksperiment. sættes til 1 hvis udfaldet af eksperiment faktisk var og 0 ellers. er dermed Bernoulli-fordelt med . og er uafhængige når og er forskellige, men ellers ikke.

Lad nu og være de observerede hyppigheder af de tre udfald. Definer nu teststørrelsen:

Lad være den stokastiske variabel der svarer til . Ifølge den centrale grænseværdisætning er disse stokastiske variable tilnærmelsesvist standardnormalfordelte:

Alternativt kan man sige, at følgende stokastiske variable skal være normalfordelte med middelværdi og varians :

Teststørrelsen (som stokastisk variabel) kan udtrykket vha. disse:

Definer nu en stokastiske vektor . Lad desuden være en standardnormaltfordelt stokastisk vektor i dimensioner. Sæt . Definer nu flg. stokastiske vektor:

Her er . Da kunne man også skrive . I denne formulering er det tydeligt, at er projektionen af på . ligger dermed i det ortogonale komplement til underrummet udspændt af . har altså koordinater på formen ifht. denne basis.

Suppler nu op til en ortonormalbasis for : og konstruer en matrix bestående af disse basisvektorer som søjler: . Sæt nu . er dermed standardnormaltfordelt i dimensioner. er samme vektor, bortset fra at første koordinat er projiceret væk:

Da er ortogonal er , så:

Dermed er . Hvis vi kan vise er vi færdige.

**Lemma**: og som defineret ovenfor har kovariansmatrix, givet ved og .

Bevis: Til at starte med bemærkes, at begge matricer har identisk forventningsværdi, nemlig nulvektoren. Resultatet skal vises for begge matricer. Først . Her får vi brug for flg. forventningsværdi når :

Første led er altid nul, da udfaldet af hvert enkelt Bernoulli-eksperiment kun kan lande i én kategori. Tilbage er:

Her er uafhængigheden mellem de enkelte eksperimenter benyttet. De er i alt led i summen:

Nu kan kovariansen beregnes (husk at alle forventningsværdier er nul):

Varianserne er blot forventningsværdierne af kvadraterne:

Her skal vi igen bruge at , samt:

Så:

Det samme skal vises for . Den ’te koordinat af denne vektor er altså:

Kovarianserne bliver:

Da ’erne er uafhængige og standardfordelte er . Derfor reducerer ovenstående til:

For giver dette netop og for giver dette . Bevis slut.

og har altså samme forventningsværdi og kovariansmatrix. I grænsen hvor er stor er tilnærmelsesvis normalfordelt, ligesom . Altså må begge vektorer have samme længde i denne grænse.