

Ofte udføres eksperimenter, hvor man forventer et bestemt resultat. Byg, der har været udsat for radioaktiv bestråling, vil efter spiring fremkomme i to forskellige farver. Man kan undersøge hyppigheden mellem de to farver og sammenligne med det forventede resultat ifølge en biologisk model. Man kan også sammenligne blodtypers fordeling i en gymnasieklasse, med den kendte blodtypefordeling i Danmark. I begge tilfælde skal man vurdere, om afvigelsen mellem det observerede og det forventede resultat, er acceptabel eller ej, dvs. om afvigelsen skyldes en tilfældighed, eller om afvigelsen skyldes en virkelig forskel. For at forstå hvordan afgørelsen træffes, er det vigtig først at udføre metoden "håndholdt" for til sidst at få Excel til at udføre hele arbejdet. Metoden er enkel at udføre, men svær at fortolke.

Det græske bogstav χ udtales ki, og χ^2 udtales "ki i anden".

Eksempel 1. Test om en mønt er fair.

Hypotese. Mønten er fair, dvs. sandsynlighed for plat er $\frac{1}{2}$ og sandsynligheden for krone er $\frac{1}{2}$.

Observation. Ved 20 kast med en mønt fås 6 plat og 14 kroner.

Udfald	plat	krone
Observeret	6	14
Forventet	10	10

Hypotesen er vores model af virkeligheden og observationen er virkeligheden. For at afgøre om model og virkelighed afviger fra hinanden, skal vi opstille og vurdere et mål for denne afvigelse.

Teststørrelse. Teststørrelsen er et tal, der angiver afvigelsen mellem det forventede (Expected) og det observerede (Observed).

$$\text{Teststørrelse} = \frac{(\text{observeret antal plat} - \text{forventet antal plat})^2}{\text{forventet antal plat}} + \frac{(\text{observeret antal krone} - \text{forventet antal krone})^2}{\text{forventet antal krone}}$$

$$\text{Teststørrelse} = \frac{(O_{\text{plat}} - E_{\text{plat}})^2}{E_{\text{plat}}} + \frac{(O_{\text{krone}} - E_{\text{krone}})^2}{E_{\text{krone}}}$$

$$\text{Teststørrelse} = \frac{(6 - 10)^2}{10} + \frac{(14 - 10)^2}{10} = \frac{(-4)^2}{10} + \frac{4^2}{10} = 1,6 + 1,6 = 3,2$$

Hvis tælleren ikke sættes i anden, ville de to led ophæve hinanden og teststørrelsen ville give 0.

Jo større teststørrelsen er, jo mere afviger det observerede antal fra det forventede. Den svære afgørelse er, hvor stor teststørrelsen skal være (kritisk værdi), før afvigelsen er så betydningsfuld (signifikant), så vi vælger at forkaste (afvise) hypotesen om en fair mønt.

Bestemmelse af antallet af frihedsgrader. Hvis vi kaster 20 gange med en mønt, kan vi nøjes med at opskrive antallet af kroner. Hvis der er 12 kroner, ved vi der må være 8 plat. Når man tester en mønt, er antallet af frihedsgrader 1. Når man tester en terning, er antallet af frihedsgrader 5. Der gælder generelt, at antallet af frihedsgrader er 1 mindre end antallet af udfald.

Signifikansniveau. Er det sandsynligt, at en teststørrelse på 3,2 kan være fremkommet ved kast med en fair mønt? Hvis denne sandsynlighed er over 5 % accepteres hypotesen om en fair mønt og hvis sandsynligheden er under 5 % forkastes hypotesen. De 5 % kaldes signifikansniveauet.



Den kritiske værdi. Hvis en mønt er fair, er der en sandsynlighed på 5 % for at få en teststørrelse på 3,84 eller derover. Det er så usandsynligt, at vi vælger at forkaste hypotesen om, at mønten er fair, hvis teststørrelsen ligger i dette interval. Tallet 3,84 kaldes for den kritiske værdi.

Konklusion. Vi accepterer hypotesen om at mønten er fair, fordi teststørrelse på 3,2 ligger under den kritiske værdi på 3,84.

Opgave 1. Vil du forkaste eller acceptere, hvis stikprøveresultatet bliver 5 plat og 15 kroner?

Opgave 2. Udregn teststørrelsen, når man i 40 kast med en mønt får 16 plat og 24 kroner. Hvor mange gange mindre bliver teststørrelsen? Kommentar til opgaven: En afvigelse på 4 i forhold til en forventet værdi på 10 vægter mere end en afvigelse på 4 i forhold til en forventet værdi på 20. Det er ikke forskellen i sig selv, men forskellen i forhold til det forventede, som har betydning for teststørrelsens værdi.

Opgave 3. Udregn teststørrelsen, når man i 200 kast med en mønt får 60 plat og 140 kroner. Hvor mange gange større bliver teststørrelsen? Kommentar til opgaven: Hvis alle de observerede tal ganges med 10 bliver teststørrelsen også ganget med 10. Jo flere observationer jo nemmere får vi øje på en afvigelse.

Eksempel 2. Test om bygplanter følger en forventet fordeling.

Bygplanter fra et kryds mellem en grøn og en hvid mutant er udsået. Vi vil teste om antallet af grønne og hvide spirer vil fordele sig som 3 til 1. Der kommer i alt 96 bygplanter frem, hvor 81 er grønne og 15 er hvide. Vi vil forvente $3/4 \cdot 96 = 72$ grønne og $1/4 \cdot 96 = 24$ hvide.

Udfald	grønne	hvide
Observeret = O	81	15
Forventet = E	72	24

$$\text{Teststørrelse} = \frac{(O_{\text{grønne}} - E_{\text{grønne}})^2}{E_{\text{grønne}}} + \frac{(O_{\text{hvide}} - E_{\text{hvide}})^2}{E_{\text{hvide}}}$$

$$\text{Teststørrelse} = \frac{(81 - 72)^2}{72} + \frac{(15 - 24)^2}{24} = \frac{9^2}{72} + \frac{9^2}{24} = 1,125 + 3,375 = 4,5$$

Konklusion. Da teststørrelsen er over 3,84 forkastes hypotesen.

Opgaver 4. Vil du forkaste eller acceptere, hvis resultatet giver 80 grønne og 16 hvide?

Tabel over kritiske værdier for forskellige frihedsgrader, når signifikansniveauet er 5 %

Antal frihedsgrader	1	2	3	4	5
Kritisk værdi	3,84	5,99	7,81	9,49	11,1

Læg mærke til, jo flere frihedsgrader jo større kritisk værdi. Med 5 frihedsgrader skal teststørrelsen være over 11 før hypotesen forkastes. Jo flere frihedsgrader jo flere led i teststørrelsen og jo større værdi af teststørrelsen kan vi tolerere, før vi er kritiske.

Opgave 5. Der kastes 60 gange med en terning.

Antal øjne	1	2	3	4	5	6
Observeret	7	13	17	5	10	8
Forventet	10	10	10	10	10	10

Udregn teststørrelsen og bestem antallet af frihedsgrader. Er terningen fair?

I statistisk sprogbrug hedder det et test. Et χ^2 -test kan ikke udføres på frekvenser (fx som brøk eller %). Tabellen over de kritiske værdier ligger i Excel som vi vil vise nedenfor.



χ^2 -test i Excel

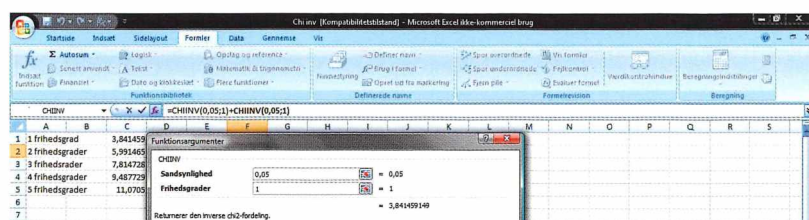
Tre metoder i Excel.

Metode 1. Ved at finde teststørrelsens kritiske værdi (samme metode som ovenfor)

Metode 2. Ved at finde sandsynligheden for at få teststørrelsens værdi.

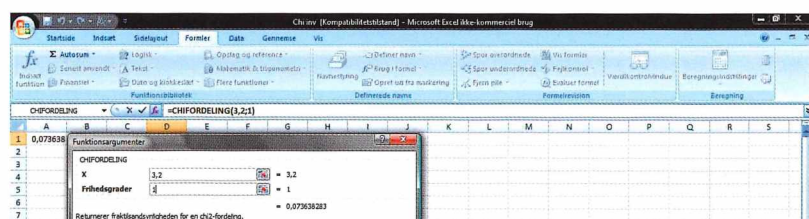
Metode 3. Direkte via et χ^2 -test.

Metode 1. Den kritiske værdi på 3,84 findes i Excel ved at benytte den inverse χ^2 -fordeling (inverse betyder omvendte). Sandsynligheden er signifikansniveauet, der sædvanligvis sættes til 5 % = 0,05. I den nyeste udgave af Excel skal du benytte CHI2.INV.RT



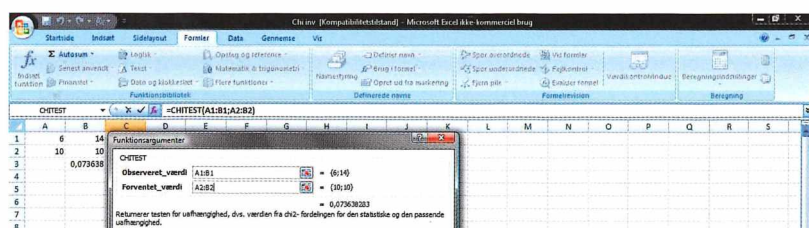
Metode 2. Testsandsynligheden eller p -værdien angiver sandsynligheden for at få en teststørrelse der er større end eller lig 3,2, hvis mønten er fair. I Excel findes p -værdien på 7,4 % via χ^2 -fordelingen..

I den nyeste udgave af Excel skal du benytte CHI2.FORDELING.RT.



Konklusion. Hypotesen accepteres, fordi sandsynligheden for at få teststørrelsen på 3,2 ligger over signifikansniveauet på de 5 %.

Metode 3. Excel kan udregne p -værdien direkte (uden at udregne teststørrelsen) ved at indtaste de observerede og forventede værdier. De observerede værdier markeres og indsættes, tilsvarende med de forventede værdier. I den nyeste udgave af Excel skal du benytte CHI2.TEST.



Konklusion. Hypotesen accepteres, da p -værdien på 7,4 % ligger over signifikansniveauet.

Oversigt over χ^2 -test med metode 1.

- Teststørrelsen udregnes.
- Ved brug af Excel udregnes teststørrelsens kritiske værdi.
- Hvis teststørrelsens værdi er under den kritiske værdi accepteres hypotesen, og hvis teststørrelse er over den kritiske værdi, forkastes hypotesen.

Oversigt over χ^2 -test med metode 2.

- Teststørrelsen udregnes.
- Ved brug af Excel udregnes teststørrelsens sandsynlighed (p -værdi).
- Hvis p -værdien er over 5 % accepteres hypotesen, og hvis p -værdien er under 5 %, forkastes hypotesen (signifikansniveau på 5 %)

Oversigt over χ^2 -test med metode 3.

- De observerede og forventede tal indtastes i Excel.
- Hvis p -værdien er over 5 % accepteres hypotesen, og hvis p -værdien er under 5 %, forkastes hypotesen (signifikansniveau på 5 %)

Fortolkning af en p -værdi på 7,4 %:

I 74 ud af 1000 stikprøver, vil en fair mønt give en teststørrelse på 3,2 eller derover. Det er så ofte, så vi ikke kan afvise, at mønten er fair.

Bemærk. Et χ^2 -test giver en sandsynlighed for, at få stikprøvens resultat, hvis hypotesen er rigtig og ikke sandsynligheden for hypotesens rigtighed på baggrund af stikprøvens resultat.