

9.6 p -værdi

[INFO](#)[DEL](#)

Resultatet af et binomialtest kan også afgøres ud fra den såkaldte p -værdi.

I et binomialtest er det teststørrelsen, der afgør testet: Er teststørrelsen blandt de mest sandsynlige udfald omkring fordelings middelværdi, så accepteres h_0 . Er omvendt teststørrelsen blandt de ekstreme udfald i den ene eller begge af fordelings haler, så forkastes h_0 .

Sandsynligheden for, med en ny stikprøve, at få en ny teststørrelse x_1 , der er mindst lige så ekstrem som den aktuelle teststørrelse x_0 , givet at h_0 er sand, kaldes testets p -værdi.

Med den definition fås et mål for, hvor atypisk stikprøven og dermed den observerede teststørrelse x_0 er, når h_0 antages at være sand: Jo mindre p -værdi, jo mere atypisk teststørrelse.

Tester vi på 5% signifikansniveau, og får vi en p -værdi, der er mindre end eller lig med 5%, så er teststørrelsen blandt de 5% mest ekstreme udfald, og h_0 skal derfor afvises. Er p -værdien omvendt over 5%, så er teststørrelsen ikke blandt de 5% mest ekstreme udfald, og der er derfor ikke grund til at afvise h_0 .

Med kendskab til p -værdien kan testet altså afgøres:

HYPOTHESETEST

[DEL](#)

For et binomialtest udført på 5%-signifikansniveau gælder:

- hvis p -værdien er mindre end eller lig med 5% afvises h_0 .
- hvis p -værdien er større end 5% accepteres h_0 .

Beregningen af testets p -værdi udføres på lidt forskellige måder afhængig af, om testet er ensidet eller tosidet.

EKSEMPEL 14

[DEL](#)

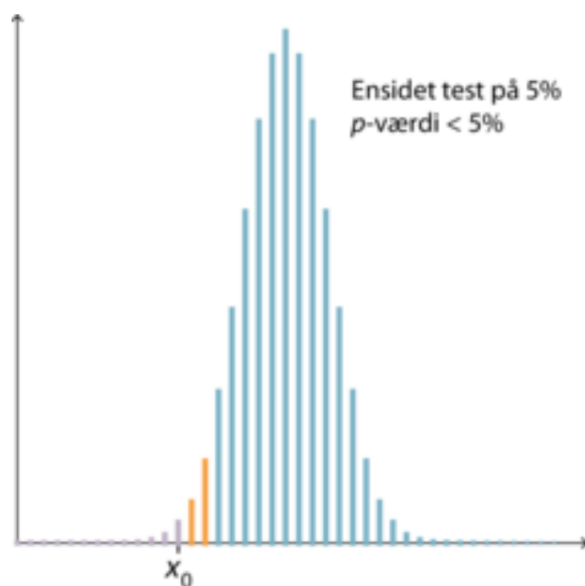
Vi ser igen på eksempel 11, hvor vi med et venstresidet test undersøgte om kun tre gevinster ud af 50 lodder giver grund til at mistænke snyd, når gevinstsandsynligheden er oplyst at være mindst 15%.

Kun udfaldene 0, 1 og 2 gevinster ud af 50 lodder er mere ekstreme end det udfald, vi faktisk fik. Derfor gælder

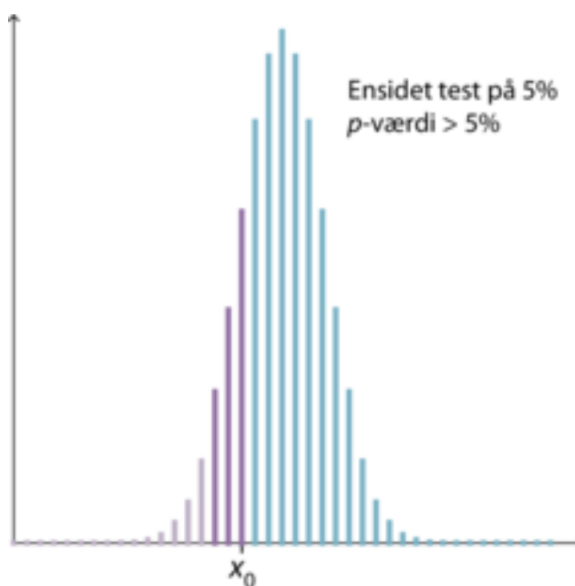
$$p\text{-værdi} = P(X \leq 3) = 4,6\%.$$

Udregningen viser, at p -værdien er under 5%. Hypotesen om en gevinstsandsynlighed på mindst 15% skal derfor afvises. Det var også konklusionen i eksempel 11.

Da sandsynligheden for et udfald i den kritiske mængde pr. definition er højst 5%, (når der testes på 5% signifikansniveau), er det klart, at konklusionen i et hvilket som helt ensidet test bliver den samme, hvad enten man afgør testet på baggrund af p -værdien, eller den kritiske mængde. Dette forhold er illustreret på figur 6 og 7.



Figur 6



Figur 7

Beregning af p -værdien er mere kompliceret, når testet er tosidet, og som vi skal se, er der heller ikke en fælles standard for, hvordan det skal gøres.

EKSEMPEL 15 DEL

Vi vender tilbage til eksempel 8, hvor et tosidet test skulle afgøre, om 11 personer med blodtype B i en stikprøve på 60 personer, var grund nok til at mistænkte en anden blodtype B andel i befolkningen end 10%.

Alle udfald over 11 er mindst lige så ekstreme som det aktuelle udfald $x_0 = 11$. Da testet er tosidet, er der ligeledes mulighed for ekstreme udfald i fordelings modsatte hale.

Hvis vi regner med, at de ekstreme udfald tilsammen i hver af de to haler har ca. samme sandsynlighed, så har vi tilnærmelsen

$$p\text{-værdi} \approx 2 \cdot P(X \geq 11) = 2 \cdot 3,4\% = 6,8\%.$$

Resultatet er over 5%, og hypotesen h_0 skal derfor ikke forkastes, når vi lægger denne tilnærmede p -værdi til grund. Konklusionen er den samme som i eksempel 8, hvor testet blev afgjort ud fra den kritiske mængde.

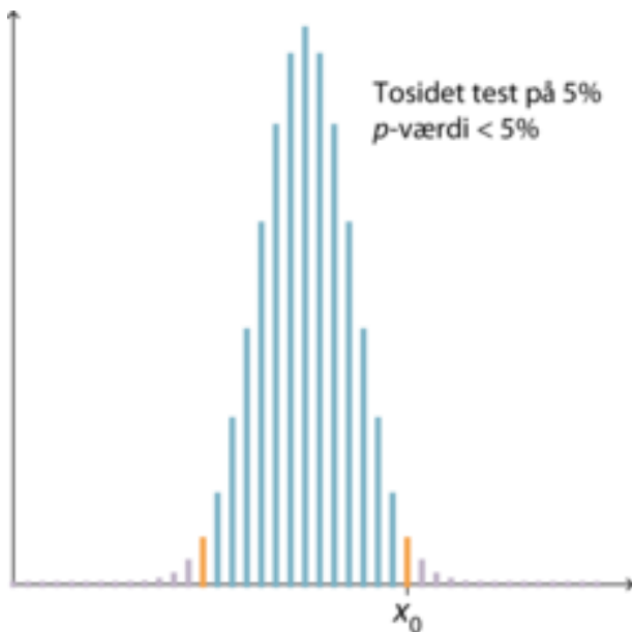
Beregning af den eksakte p -værdi kræver en udpegning af de ekstreme udfald i fordelings anden hale. Der er imidlertid ikke enighed om, hvad vi præcis skal forstå ved et 'ekstremt udfald'. Nogle fortolker 'ekstremt udfald' som et udfald, der har en sandsynlighed, der er mindre end eller lig med sandsynligheden for teststørrelsen. Andre fortolker 'ekstremt udfald' som et udfald, der har en afstand til middelværdien, der er mindst lige så stor som teststørrelsens afstand til middelværdien. I nogle test giver det samme resultatet - i andre test forskellige resultater.

I dette eksempel finder man, uanset fortolkningen, 0 og 1 som de eneste ekstreme udfald i fordelings 'anden' hale. Den eksakte p -værdi efter begge fortolkninger er derfor

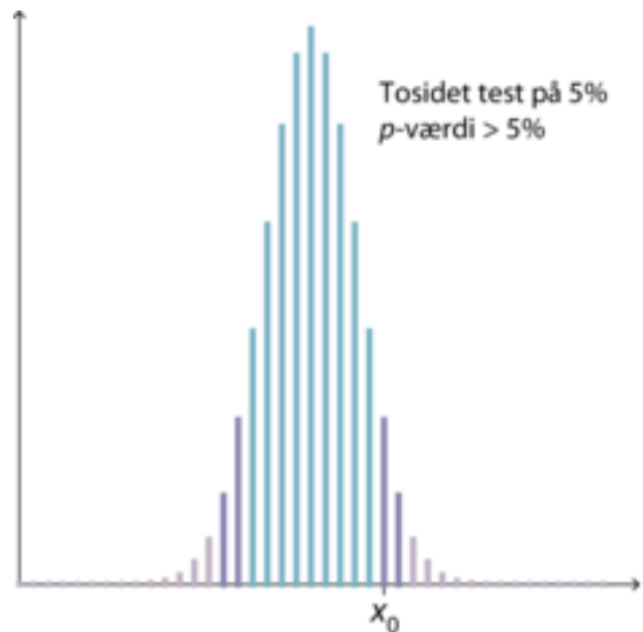
$$p\text{-værdi} = P(X \geq 11) + P(X = 0) + P(X = 1) = 4,8\%.$$

Værdien er under 5 %, og nulhypotesen skal derfor forkastes, når vi lægger denne p -værdi til grund. Vi ser, at resultatet af testet er afhængig af, om der slutes ud fra den eksakte p -værdi eller den kritiske mængde.

Eksempel 15 viser nogle af de problemer, der kan være ved at konkludere på et tosidet binomialtest. Særligt fordelings form spiller her en rolle. Hvis binomialfordelingen er symmetrisk, dvs. sandsynlighedsparameteren er 0,5 (eller tæt på 0,5), giver de forskellige fortolkninger og metoder samme konklusion. Det er illustreret på figur 8 og 9.



Figur 8



Figur 9

Er fordelingen derimod skæv, dvs. hvis sandsynlighedsparameteren enten er tæt på 0 eller tæt på 1, sådan som det er tilfældet i eksempel 15, så kan man få modstridende konklusioner. Tydelighed omkring definitioner og metoder er derfor i sådanne tilfælde særlig vigtig.

I eksempel 15 gav den tilnærmede p -værdi og den kritiske mængde samme konklusion på testet. Det gælder generelt uanset fordelings form. Mange CAS-programmer kan udføre et binomialtest med en kommando, og her er konklusionen altid i overensstemmelse med den konklusion, man får ud fra den kritiske mængde.