

Hypoteseprøvning

Definition:

- Et udsagn ift. en population, som skal be- eller afkræftes.
- Man vil udtale sig om forhold i populationen, baseret på data fra en stikprøve.
- Tilfældig udtrækning --> Stikprøve

Begreber

n = Antal

$\bar{x}$ = Gennemsnit

s = Standardafvigelse

X = Antal specielle

T_s = Test-størrelse

Vigtige "basics"

Venstre sided test: Når udsagnet af test er mindre end/lavere

Højre sided test: Når udsagnet af test er mere end/over

Dobbeltsided test: Når udsagnet af test er lig med/forskellige fra

Venstre sided test

$$H_0: P \geq P_0$$

$$H_1: P < P_0$$

Højre sided test

$$H_0: P \leq P_0$$

$$H_1: P > P_0$$

Dobbeltsided test

$$H_0: P = P_0$$

$$H_1: P \neq P_0$$

Eksempel på spørgsmål

- Er den sande andel der stemmer på Socialdemokratiet mindre end 30 %? (Test på 5 %)

Konklusioner / Regler

- Hvis sandsynligheden for at H_0 er korrekt er under testniveauet på f.eks. 5 %, så afvises H_0 og dermed er det H_1 der er rigtigt.

Test

σ_0 er den tærskelværdi, der skal testes og T_s er teststørrelsen.

- $T_s = (n-1) \cdot \frac{s^2}{\sigma_0^2}$

T_s følger en χ^2 -fordeling med $fg = n - 1$ frihedsgrader.

	H_0	H_1	H_0 forkastes hvis	p-værdi
Venstresidet	$\sigma \geq \sigma_0$	$\sigma < \sigma_0$	$T_s < c$	$P(\chi^2 < T_s)$ A
Højresidet	$\sigma \leq \sigma_0$	$\sigma > \sigma_0$	$T_s > c$	$P(\chi^2 > T_s)$ B
Dobbeltsidet	$\sigma = \sigma_0$	$\sigma < \sigma_0$ eller $\sigma > \sigma_0$	$T_s < c_1$ eller $T_s > c_2$	$2 \cdot \min(A, B)$

c er den kritiske værdi i χ^2 -fordelingen

Opgave gennemgået i online-time

$NB: \sqrt{\frac{s^2}{c_2} \cdot (n-1)}$ ØB: $\sqrt{\frac{s^2}{c_1} \cdot (n-1)}$
 $NB: \sqrt{\frac{2,3^2}{2,6/19} \cdot (15-1)} = 1,68$
 $ØB: \sqrt{\frac{2,3^2}{5,629} \cdot (15-1)} = 3,63$
 $\sigma \in [1,68; 3,63]$ m. 95% sikkerhed

Typeopgave: Venstresidet
 Problem: Er skedep for venstresidet lavere end 5 min - test på 5% Niveau.
 Stikprøve: $n = 15$
 $S = 2,3$

$H_0: \sigma \geq 5$ } venstresidet
 $H_1: \sigma < 5$ } Sidedt.

Beregn $T_s: T_s = \frac{s^2}{\sigma_0^2} \cdot (n-1) = \frac{2,3^2}{5^2} \cdot (15-1)$
 $T_s = 2,96$

Find c eller p-værdi i Excel eller 3D en χ^2 .
 Testen følger en χ^2 -fordeling!

Da T_s ligger i 5% område
 $\Rightarrow H_0$ afvises og H_1 Accept.

p-værdi = 0,0009
 us sign. niv = 0,05 } H_1 ✓

Hypotese: $\sigma \geq 5$ (mindre end!)
 Venstresidet (mindre end!)
 $H_0: \sigma \geq 5$
 $H_1: \sigma < 5$

Konkl. Den samlede skedep er lavere end 5.
 Testet på 5% Niv.

Visuel vurdering af forudsætningerne

Observationerne (x, y) tegnes som **punkter i et koordinatsystem**.
Herved får man en 'sværm' af punkter, der giver et umiddelbart visuelt indtryk af den sammenhæng, der er mellem x og y.
Man kan vurdere om en lineær regressionsmodel er velegnet til at beskrive sammenhæng mellem x og y ved at vurdere om modellens forudsætninger er opfyldt:

- 1** Punkterne skal centrere sig omkring en (tænkt) midterlinie, Dvs. skal ligge i en opadgående eller nedadgående lineær trend.
- 2** Punkternes spredningen i op og nedadgående retning i forhold til en midterlinie, skal være konstant uanset x-værdien.
- 3** Punkterne må ikke danne mønster, der må ikke være systematik i afvigelserne.
- 4** Der er flest punkter ved midterlinien med 'udtynding' i op og nedadgående retning.

Beregninger / Formler

$$\bullet \quad b_1 = \frac{SS_{xy}}{SS_x} = \frac{SAP_{xy}}{SAK_x}$$

regressionsliniens hældning

$$\bullet \quad b_0 = \bar{y} - b_1 \cdot \bar{x}$$

regressionsliniens skæring med y-aksen

$$\bullet \quad s = \hat{\sigma} = \sqrt{\frac{SSE}{n-2}}$$

'Spredningen' omkring regressionslinien ('op' / 'ned').

SSE er summen af afvigelseskvadraterne for residualerne (e)

$$\bullet \quad \hat{y} = \hat{E}(y) = b_0 + b_1 \cdot x$$

Den estimerede regressionslinie