

Statistik - guldnoter

Indhold

Beregning af fejlmargen	3
HUSK! Altid argumenter for ensidet eller tosidet test.	3
To højre skæve fordelinger lagt sammen vil være nogenlunde symmetriske	3
Forudsætninger:	4
Type 1 og type 2 fejl	5
Hypoteseprøvning vedrørende én μ med kendt varians (z-fordeling)	6
Beregning af n	8
Beregning af konfidensinterval	8
Hypoteseprøvning vedrørende én μ med ukendt varians (t-fordeling)	9
Bestemmelse af n, stikprøvestørrelsen	12
Konfidensinterval	12
Hypoteseprøvning vedrørende σ^2 : χ^2 -fordelingen (s. 21, 22, 45, 51)	13
Konfidensinterval	15
Hypoteseprøvning vedrørende p (test for andele) (s. 27, 44 og 52)	16
Bestemmelse af n, stikprøvestørrelsen	18
Konfidensinterval	19
Hypoteseprøvning vedrørende forhold mellem to varianser (s. 25, 44, 51)	20
hvor S^2 er den største af de to varianser	20
Konfidensinterval	22
Hypoteseprøvning vedrørende forsk. mellem to gennemsnit (s. 25, 43, 50)	23
Varianshomogenitetstest	24
hvor S^2_1 er den største af de to varianser	24
Konfidensinterval	28
Hypoteseprøvning vedrørende parvise stikprøver	29
Konfidensinterval parvise stikprøver	31
Hypoteseprøvning vedrørende forskellen mellem to andele (s. 28)	32
Konfidensinterval	35
Forudsætninger:	36
Test:	36
Konfidensinterval:	36
Poissonfordelingen:	36
Binomialfordelingen:	36
Den hypergeometriske fordeling	36

Type 1 og type 2 fejl

	H_0 er sand	H_0 er ikke sand
Accepterer		Type 2 fejl
Forkaster	Type 1 fejl	

Signifikansniveau betydning:

Jo **lavere** signifikansniveau, jo **lavere** er chancen for at begå en type 1 fejl.

Jo **højere** signifikansniveau desto lavere er chancen for en type 2 fejl. Og jo lavere signifikansniveau, jo højere er chancen for at begå en type 2 fejl.

For ved 2 sided test gælder:

- 10% signifikansniveau er acceptområdet $\pm 1,645$
- 5% signifikansniveau er acceptområdet $\pm 1,96$

Derved er der et større acceptområde jo højere signifikansområdet er \rightarrow større chance for at accepterer H_0 selvom den ikke er sand.

Observerværdi betydning:

Jo højere observerværdi jo større er chancen for at begå en type 1 fejl ved 2 sided test.

Jo mere μ afviger fra, hvad der postuleres i H_0 når man gennemfører en to-sided test desto større er risikoen for at begå en type 1 fejl.

Jo mindre μ afviger fra, hvad der postuleres i H_0 når man gennemfører en to-sided test jo større er risikoen for at begå en type 2 fejl.

Dette ses ved at kigge på formlen for observerværdien. For en større difference i $\bar{x} - \mu_0$ jo større vil observerværdien blive og den vil dermed have større tendens til at falde udenfor de kritiske grænser.

Eksempel med forbryder:

Type 1 fejl: Dømmer en uskyldig skyldig

Type 2 fejl: Dømmer en skyldig uskyldig

Her er det værst at begå en type 1 fejl \rightarrow man vil mindske chancen for at lave en type 1 fejl \rightarrow sæt lavt signifikansniveau

Såfremt $|t_{obs}| \leq |t_{n-1; \frac{\alpha}{2}}|$ fastholdes H_0

Såfremt $|t_{obs}| > |t_{n-1; \frac{\alpha}{2}}|$ afvises H_0

Ved kritiske grænser skal der approksimeres til normalfordelingen såfremt frihedsgraderne er over 200. (svarer til at tage uendelig i t-fordelingen, ligesom man gør ved en z-fordeling)

Ensidet test:

Ved øvre test: $t_{n-1; \frac{\alpha}{2}}$

Ved nedre test: $-t_{n-1; \frac{\alpha}{2}} = t_{n-1; 1-\frac{\alpha}{2}}$

Evt. indsæt en grafisk illustration! - HUSK der kun er én grænse her!

Såfremt t_{obs} ligger udenfor den kritiske grænse afvises H_0

Ved kritiske grænser skal der approksimeres til normalfordelingen såfremt frihedsgraderne er over 200. (svarer til at tage uendelig i t-fordelingen, ligesom man gør ved en z-fordeling)

7. Beregning af p-værdi

Tosidet test:

$$2 * P(T_{n-1} \leq -|t_{obs}|) = 2 * P(T_{n-1} > |t_{obs}|)$$

Man finder p-værdien ved at man går ud fra de pågældende frihedsgrader for t_{obs} og finder den værdi i rækken der er tættest på den fundne t_{obs} og aflæser da hvilket signifikansniveau der er tættest på den fundne t_{obs} . Det gør man i tabel 4.

Det er kun ved p-værdien man kan approksimere til normalfordelingen. Dette kan man gøre når frihedsgraderne er over 30. Gør det kun hvis frihedsgraderne IKKE er tabellagt i t-fordelingen.

Ensidet test:

Nedre test:

$$P(T_{n-1} \leq -|t_{obs}|) = P(T_{n-1} > |t_{obs}|)$$

Øvre test:

$$P(T_{n-1} \geq |t_{obs}|) = -P(T_{n-1} > |t_{obs}|)$$

Man finder p-værdien ved at man går ud fra de pågældende frihedsgrader for t_{obs} og finder den værdi i rækken der er tættest på den fundne t_{obs} og aflæser da hvilket signifikansniveau der er tættest på den fundne t_{obs} . Det gør man i tabel 4.

Konfidensinterval

1. Definer stokastisk variabel
2. Opskriv givne oplysninger
3. Valg af signifikansniveau
4. Valg af formel

$$\sigma^2 \in \left[\frac{(n-1) * s^2}{X^2_{n-1; \frac{\alpha}{2}}}; \frac{(n-1) * s^2}{X^2_{n-1; 1-\frac{\alpha}{2}}} \right]$$

husk her at korrigerer antal frihedsgrader både over og under brøkstregen, hvis du vælger et andet antal frihedsgrader ved grænserne, grundet det ikke er tabellagt.

Ved $n-1 > 50$ approksimer over til normalfordeling ved grænser

$$X^2_{n-1; 1-\frac{\alpha}{2}} \sim (n-1) - Z_{\frac{\alpha}{2}} * \sqrt{2 * (n-1)}$$

$$X^2_{n-1; \frac{\alpha}{2}} \sim (n-1) + Z_{\frac{\alpha}{2}} * \sqrt{2 * (n-1)}$$

5. Forudsætninger

- $X \sim NF$ (UCGS)
- STU
- Troværdighed
- Uafhængighed
- $n/N < 5\%$ eller proces + konstanthed

6. Beregn og konkluder