Statistik - guldnoter

Indhold

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 sidet 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 → større chance for at accepterer H_0 selvom den ikke er sand.

Observatorværdi betydning:
Jo højere observator værdi jo større er chancen for at begå en type 1 fejl ved 2 sidet test.

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

Dette ses ved at kigge på formlen for observatorværdien. For en større difference i $\bar{x} - \mu_0$ jo større vil observatorværdien 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 → man vil mindske chancen for at lave en type 1 fejl → sæt lavt signifikansniveau

$$\text{Såfremt } |t_{obs}| \leq \left|t_{n-1;\frac{a}{2}}\right| \text{ fastholdes } H_0$$
$$\text{Såfremt } |t_{obs}| > \left|t_{n-1;\frac{a}{2}}\right| \text{ afvises } H_0$$

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: $t_{n-1;\frac{a}{2}}$

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

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

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

husk her at korrigere 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

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

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

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

6. Beregn og konkluder