

Varianans analyse tabellen

Kilder	FG	SS	MS	F	P
Behandling	$k-1$	$SSTR$	$SSTR/(k-1)$	$\frac{SSTR}{k-1} / \frac{SSE}{(n-k)}$	$P(\bar{F} \geq F_{obs})$
Feil	$n-k$	SSE	$SSE/(n-k)$		
Total	$n-1$	$SSTOT$	$SSTOT/(n-1)$		

Sammenligning af to-udvals test og F test

$X_1, \dots, X_n \sim N(\mu_1, \sigma^2)$ og uafh.

$Y_1, \dots, Y_m \sim N(\mu_2, \sigma^2)$ og uafh. og uafh. af $X_1, \dots, X_n$

Total gennemsnit $= (n\bar{x} + m\bar{y})/(n+m)$

$$SSTR = n \left(\bar{x} - \frac{n\bar{x} + m\bar{y}}{n+m} \right)^2 + m \left(\bar{y} - \frac{n\bar{x} + m\bar{y}}{n+m} \right)^2$$

$$= n \left[\frac{n(\bar{x} - \bar{y})}{n+m} \right]^2 + m \left[\frac{m(\bar{y} - \bar{x})}{n+m} \right]^2$$

$$= (\bar{x} - \bar{y})^2 \cdot \frac{nm^2 + mn^2}{(n+m)^2} = \frac{nm}{n+m} (\bar{x} - \bar{y})^2$$

$$\frac{SSE}{n+m-2} = \frac{\sum_{i=1}^n (X_i - \bar{x})^2 + \sum_{j=1}^m (Y_j - \bar{y})^2}{n+m-2} = s_p^2$$

$$F = \frac{\frac{nm}{n+m} (\bar{x} - \bar{y})^2}{\frac{SSE}{n+m-2}} = \frac{\frac{nm}{n+m} (\bar{x} - \bar{y})^2}{s_p^2} \sim F_{1, n+m-2}$$

$$= \frac{(\bar{x} - \bar{y})^2}{s_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)} = \left(T_{n+m-2} \right)^2$$

To-sided test

$$\alpha = P(T \leq -t_{\frac{\alpha}{2}, m+m-2} \cup T \geq t_{\frac{\alpha}{2}, m+m-2})$$

$$\Leftrightarrow \alpha = P(T \geq t_{\frac{\alpha}{2}, m+m-2}) = P(T \geq F_{1-\alpha, 1, m+m-2})$$

12.3 Multipli sammenlikninger.

Dersom $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ blir forkasta har vi funnet at det er forskjeller mellom forventede verdier. Men vi vet ikke ~~hvor~~ ^{(altså hvor} ~~store~~ ^{store} ~~den~~ ^{den} skil seg fra de andre. Et nyttig verktøy er parvise sammenlikninger.

LSD-metoden (Minste signifikante forskjell)

Lag konfidensintervall for $\mu_j - \mu_k$

$$C_{jk} = \bar{y}_j - \bar{y}_k \pm t_{\frac{\alpha}{2}, v} s \sqrt{\frac{1}{m_j} + \frac{1}{m_k}}$$

Det er $\binom{k}{2}$ slike intervall. Dersom C_{jk} ikke dekker 0, har vi vist at μ_j og μ_k er forskjellige.

Bonferroni's metode

LSD-metoden har vist seg å være nyttig sammen med T -testene, men en må være klar over følgende.

La $H_0: \mu_1 = \mu_2 = \dots = \mu_k$

La $\bar{C}_{jk}$ være hendelse at C_{jk} dekker 0. $P(\bar{C}_{jk}) = 1 - \alpha$ dersom

H_0 er sann.

Videre $P(\bar{C}_{12} \cap \bar{C}_{13} \cap \dots \cap \bar{C}_{k-1,k}) = (1 - \alpha)^{\binom{k}{2}}$ dersom vi kan anta uavhengighet mellom intervalla og

$$P(\text{forkaste } H_0) = 1 - (1 - \alpha)^{\binom{k}{2}} \approx \binom{k}{2} \alpha \geq \alpha$$

Derfor justeres en ofte signifikansnivået i parvise sammenlikninger.

Bonferroni sin metode foreslår at vi bruker

$\frac{\alpha}{\binom{k}{2}}$ som signifikansnivå α : vi substituerer

$t_{\frac{\alpha}{2}, v}$ med $t_{\frac{\alpha}{\binom{k}{2}}, v}$ i konfidansintervall for μ^0

sammenlikne to og to forventninger. Dette går bra om k ikke er for stor. Men det er ikke realistisk å anta at alle intervaller er uavh.

Taking sin test

Bytt ut $t_{\frac{\alpha}{2}, v}$ med $\frac{Q(\alpha, k, v)}{t_2}$

Tabell A.5 i appendix gir $Q(\alpha, k, v)$. Metoden er eksakt α : $P(\bar{C}_2 \cap \bar{C}_3 \cap \dots \cap \bar{C}_{k-1, n}) = 1 - \alpha$ dersom $\mu_1 = \mu_2 = \dots = \mu_k$.

~~Dersom R er variasjonsbredden til en mengde~~
Definisjon 12.3.1

La $W_1, W_2, \dots, W_k$ være en mengde av uavh. og normalfordelte variabler med forventning μ og varians σ^2 og la R være variasjonsbredden

$$R = \max_i W_i - \min_i W_i$$

La S^2 være estimatoren for variansen

$Q_{\alpha, v} = \frac{R}{S}$ blir kalla Student's test variasjonsbredde.

Boka prövar ä väse teorem 12.3.1. Men!

$$2a \quad W_t = \bar{Y}_{0,t} - \mu$$

$$\{ (\max W_t - \min W_t) < t \} \Leftrightarrow |W_j - W_i| < t, \quad \forall i, j$$

$$\equiv \bigcap_{i,j} \{ |W_j - W_i| < t \}$$

$$P((\max W_t - \min W_t) < c) = 1 - \alpha \Leftrightarrow P\left(\bigcap_{i,j} |W_j - W_i| < c\right) = 1 - \alpha$$

$$\neq P(|W_j - W_i| < c) = 1 - \alpha, \quad \forall i, j \text{ slik som boka brukar.}$$