BYLAE A

Metode om presiese vertrouensinterval vir δ, δ_D *en* ψ te bepaal.

1. Twee onafhanklike groepe:

Dit geld dat indien x_1 en x_2 metings is uit normaalverdeelde populasies, dat

$T_v(nsp) = \dfrac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{1/n_1 + 1/n_2}}$ 'n nie-sentrale t-verdeling het met $v = n_1 + n_2 - 2$ vryheidsgrade en nie-sentraliteitsparameter $nsp = \delta\sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$, met $\delta = (\mu_1 - \mu_2)/\sigma$.

Die onderste en boonste grense nsp_O en nsp_B word nou só gekies dat

$$P\left(T_v(nsp_O) \geq \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}}\right) = \alpha/2$$

en

$$P\left(T_v(nsp_B) \leq \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}}\right) = \alpha/2 \quad ,$$

wat meebring dat

$$P(nsp_O \leq nsp \leq nsp_B) = 1\text{-}\alpha \quad ,$$

sodat (nsp_O, nsp_B) die presiese $100(1\text{-}\alpha)\%\ VI$ is van nsp.

Die SAS-stelsel (SAS Institute Inc., 2003) se TNONCT(x,v,P)- funksie gee die nie-sentraliteitsparameter se waarde waarvoor $P(T_v(nsp) \leq x) = P$ en kan gebruik word om nsp_O en nsp_B te bepaal. VI grense vir δ is dan

$$\delta_O = \sqrt{\frac{n_1 + n_2}{n_1 n_2}}\, nsp_O \quad \text{en} \quad \delta_B = \sqrt{\frac{n_1 + n_2}{n_1 n_2}}\, nsp_B \quad . \tag{A1}$$

Die SAS-program **VI_delta** (op die webblad van die handleiding beskikbaar) gebruik hierdie metode.

2. Twee afhanklike groepe:

Om vir δ_D 'n presiese VI te bepaal, neem nou

$$T_v\left(nsp\right)=\frac{\overline{x}_D-\mu_D}{s_D/\sqrt{n}} \quad , \quad v=n\text{-}1 \quad \text{en} \quad nsp=\delta_D\sqrt{n} \quad ,$$

sodat

$$\delta_{DO}=nsp_O/\sqrt{n} \quad \text{en} \quad \delta_{DO}=nsp_B/\sqrt{n} \quad . \qquad \text{(A2)}$$

Die SAS-program **VI_delta_D** (op die webblad van die handleiding beskikbaar) gebruik hierdie metode.

3. Kontraste

Indien 'n kontras van gemiddeldes $\psi=\sum_{i=1}^{k}c_i\mu_i$ beraam word met $\hat{\psi}=\sum_{i=1}^{k}c_i\overline{x}_i$, waar $\overline{x}_i$ die gemiddeld van 'n ewekansige steekproef uit 'n normaalpopulasie met gemiddelde μ_i en SA σ , geld dat $t_{\hat{\psi}}=\hat{\psi}/s_{\hat{\psi}}$ ($s_{\hat{\psi}}$ die standaardfout is van $\hat{\psi}$ soos in (6.30) gedefinieer), 'n nie-sentrale t-verdeling besit met $n_m\text{-}m$ vryheidsgrade en nie-sentraliteitsparameter nsp_ψ (soos in 6.41) gedefinieer). Hier is n_m die totale aantal waarnemings van die m steekproewe wat by $\hat{\psi}$ betrokke is. Die presiese $\left(1-\alpha\right)100\%\ VI$ vir nsp_ψ word netsoos vir dié van nsp in Bylae A.2 hierbo verkry waar $T_v\left(nsp\right)$ nou $t_{\hat{\psi}}$ is, $v=n_m\text{-}m$, en $nsp=nsp_\psi$.

Die SAS-programme *VI_delta_kontras* en *VI_delta_kontras_D* (op die webblad van die handleiding beskikbaar) gebruik hierdie metode.

BYLAE B

Beraming en vertrouensinterval van Mahalanobis D (kyk Steyn & Ellis, 2009)

Vir twee populasies met groottes N_A, N_B en $N = N_A + N_B$, geld

$$D^2 = \frac{N^2}{N_A N_B} \frac{1-\Lambda}{\Lambda} \qquad \text{(B1)}$$

Verder, volgens Johnson & Kotz (1985: 177), geld dat indien die twee populasies m-veranderlik normaal verdeel is dat:

$\frac{1-\hat{\Lambda}}{\hat{\Lambda}} = T^2/(n-2) = \hat{U}^{(1)}$ benaderd verdeel volgens $c(\gamma)F$, waar F 'n sentrale $F_{a(\gamma),b(\gamma)}$ -verdeling besit. Neem $\gamma = \frac{n_A n_B}{n} D^2$ en $n = n_A + n_B$, asook

$$g = (1+2\gamma/m)/(1+\gamma/m),$$

$$h = (1+\gamma/m)^2/(1+2\gamma/m),$$

$$\ell = n-p-3,$$

En definieer dan

$a(\gamma) = mh$, en

$b(\gamma) = 4 + (mh+2)/(B-1)$, waar

$$B = \frac{(\ell+h)(\ell+m)}{(\ell-2)(\ell+1)}$$

en $c(\gamma) = mgh(b-2)/(b\ell)$.

Verder geld $E\left((1-\hat{\Lambda})/\hat{\Lambda}\right) \doteq \frac{m+\gamma}{n-m-3}$, sodat $\hat{\gamma} \doteq (n-m-3)\frac{1-\hat{\Lambda}}{\hat{\Lambda}} - m$, (B2) en (7.7) in Hoofstuk 7 volg.

Die waardes γ_{L1} en γ_{U1} kan op so 'n wyse gekies word, dat

$$P\left(F_{a(\gamma_{L1}),b(\gamma_{L1})} \geq \hat{U}^{(1)} / c(\gamma_{L1})\right) = \alpha/2$$

en

$$P\left(F_{a(\gamma_{U1}),b(\gamma_{U1})} \leq \hat{U}^{(1)} / c(\gamma_{U1})\right) = \alpha/2, \text{ , sodat}$$

$P(\gamma_O < \gamma < \gamma_B) = 1-\alpha$, wat die $(1-\alpha)100\%\ VI$ gee vir γ.

Vir D is die benaderde VI dus:

$$\left(\sqrt{\frac{n}{n_A n_B}\gamma_O}\ ;\ \sqrt{\frac{n}{n_A n_B}\gamma_B}\right) \qquad \text{(B3)}$$

Dieselfde SAS-program as in Bylae C.1 hieronder, nl. *VI_zeta_kwadr1* (op die webblad van die handleiding beskikbaar) gebruik hierdie metode. Kyk ook Zou (2007), waar 'n effens ander benadering gevolg word om die nie-sentrale F te verkry.

Dit kan gebeur dat geen γ_O gevind kan word nie omdat $(1-\hat{\Lambda})/c\hat{\Lambda}$ te klein is. In só 'n geval word die ondergrens in (B3) as nul geneem.

BYLAE C

Beraming en vertrouensinterval van ζ^2 -indeks gebaseer op die Hotelling-Lawley-Statistiek by m veranderlike MANOVA (kyk Steyn & Ellis, 2009)

Vir k populasies wat elk m veranderlikes het, word die indeks ζ^2 gedefinieer as

$$\zeta^2 = \frac{U^{(s)}}{s + U^{(s)}} \quad ,$$

met $U^{(s)} = \sum_{i=1}^{s} \lambda_i = spoor\left(\sum_\mu \sum^{-1}\right)$, waar $\sum_\mu$ die meerveranderlike matriks, analoog aan σ_μ^2, (die tussen-populasie variansie) en $\sum$ die meerveranderlike matriks analoog aan σ^2 (binne-populasie variansie). Die λ-waardes gee die karakteristieke wortels van die matriks $\sum_\mu \sum^{-1}$ terwyl $s = min(k-1, m)$.

Die Hotelling-Lawley-statistiek $\hat{U}^{(s)}$ word gebruik by MANOVA om gelykheid van k populasies se gemiddelde vektore $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k$ te toets en word gedefinieer as

$$\hat{U}^{(s)} = \sum_{i=1}^{s} \hat{\lambda}_i = spoor\left(\boldsymbol{HE}^{-1}\right) \quad ,$$

waar ***H*** en ***E*** die ‘tussen’ en ‘binne’ steekproef som van kwadrate matrikse is, analoog aan die tussen- en binne steekproef som van kwadrate in die eenveranderlike geval.

C.1 Benaderde beramer en vertrouensinterval:

Volgens Betz (1987: 3172) geld dat indien die k populasies m-veranderlik normaal verdeel is, dat $\hat{U}^{(s)}/c(\gamma)$ benaderd 'n $F_{a(\gamma),b(\gamma)}$- verdeling besit, waar γ gegee word deur:

$$\gamma = nU^{(s)} \quad \text{en}$$

$$a(\gamma) = mh \quad \text{met} \quad h = \frac{(k-1+\gamma/m)^2}{k\text{-}1+2\gamma/m} \quad ,$$

en

$$b(\gamma) = 4+(mh+2)/(B-1) \quad \text{met} \quad B = \frac{(\ell+h)(\ell+m)}{(\ell\text{-}2)(\ell+1)} \quad ,$$

waar

$$\ell = n-k-m-1 \ .$$

Verder is

$$c(\gamma) = mgh(b-2)/(b\ell) \quad \text{met} \quad g = \frac{k-1+2\gamma/m}{k-1+\gamma/m} \ .$$

Hierdie benadering met 'n F-verdeling is volgens Betz reeds goed vir klein steekproewe.

Dit geld nou benaderd dat

$$E\left(\hat{U}^{(s)}\right) \doteq \frac{c(\gamma)b(\gamma)}{b-2} = \frac{mgh}{\ell} = \frac{m(k-1)+\gamma}{n\text{-}k\text{-}m\text{-}1} \quad ,$$

sodat 'n benaderd-onsydige beramer vir γ hieruit volg as

$$\hat{\gamma}_1 = (n-k-m-1)\hat{U}^{(s)} - m(k-1). \qquad \text{(C1)}$$

Omdat $\zeta^2 = \dfrac{\gamma}{ns+\gamma}$, is 'n benaderd-onsydige beramer daarvoor:

$$\hat{\zeta}_1^2 = \frac{(n\text{-}k\text{-}m\text{-}1)\hat{U}^{(s)} - m(k-1)}{ns+(n\text{-}k\text{-}m\text{-}1)\hat{U}^{(s)} - m(k-1)} \ . \qquad \text{(C2)}$$

Soos in Bylae B kan γ_{O1} en γ_{B1} só gekies word dat

$$P\left(F_{a(\gamma_{O1}),b(\gamma_{O1})} \geq \hat{U}^{(s)} / c(\gamma_{O1})\right) = \alpha/2$$

en

$$P\left(F_{a(\gamma_{B1}),b(\gamma_{B1})} \leq \hat{U}^{(s)} / c(\gamma_{B1})\right) = \alpha/2, \quad \text{sodat}$$

$P(\gamma_{O1} < \gamma < \gamma_{B1}) = 1-\alpha$, wat 'n benaderde $(1-\alpha)100\%\ VI$ vir $\gamma = nU^{(s)}$ gee.

Benaderd geld dus dat vir ζ^2 die VI:

$$\left(\frac{\gamma_{O1}}{ns+\gamma_{O1}};\frac{\gamma_{B1}}{ns+\gamma_{B1}}\right) \quad . \tag{C3}$$

Die SAS-program *VI_zeta_kwadr1* (op die webblad van die handleiding beskikbaar) gebruik hierdie metode.

Vir $\hat{U}^{(s)} / c$ te klein kan $\hat{\zeta}_1^2$ negatief wees en die ondergrens γ_O nie gevind word nie. Dan neem ons $\hat{\zeta}_1^2$ = 0 en die ondergrens vir (C3) as 0.

C.2 Asimptotiese beramer en vertrouensinterval:

Volgens Seber (1984: 39) geld asimptoties as $n \to \infty$, dat $(n-k)\hat{U}^{(s)}$ 'n nie-sentrale Chi-kwadraatverdeling X_v^2' besit met nie-sentraliteitsparameter $\gamma = nU^{(s)}$ en vryheidsgrade $v = (k-1)m$.

Hieruit volg die asimptoties-onsydige beramer vir ζ^2 as volg:

$E\left((n-k)\hat{U}^{(s)}\right) \doteq m(k-1)+\gamma$, sodat

$\hat{\gamma}_2 = (n-k)\hat{U}^{(s)} - m(k-1)$

'n onsydige beramer is vir γ en vir ζ^2:

$$\hat{\zeta}_2^2 = \frac{(n\text{-}k)\hat{U}^{(s)} - m(k-1)}{ns + (n\text{-}k)\hat{U}^{(s)} - m(k-1)} \tag{C4}$$

Soos in C.1 hierbo, kan γ_{02} en γ_{B2} só gekies word dat

$$P\left(X_v^{2}{}'\left(\gamma_{O2}\right) \geq (n-k)\hat{U}^{(s)}\right) = \alpha/2$$

en

$$P\left(X_v^{2}{}'\left(\gamma_{B2}\right) \leq (n-k)\hat{U}^{(s)}\right) = \alpha/2$$, sodat

$(\gamma_{O2}\ ,\ \gamma_{B2})$ 'n asimptotiese $(1-\alpha)100\%\ VI$ is vir γ.

Die SAS-stelsel (SAS Institute Inc., 2003) se CNONCT (x,v,P)- funksie gee die nie-sentraliteitsparameter se waarde waarvoor $P\left(X_v^{2}{}'(nsp) \leq x\right) = P$ en kan gebruik word om γ_{02} en γ_{B2} te bepaal.

Benaderd as n groot is, geld dus dat

$$\left(\frac{\gamma_{O2}}{ns+\gamma_{O2}};\frac{\gamma_{B2}}{ns+\gamma_{B2}}\right)$$

die VI is vir ζ^2.

Die SAS-program *VI_zeta_kwadr2* (op die webblad van die handleiding beskikbaar) gebruik hierdie metode.

Ook hier vir $\hat{U}^{(s)}$ te klein kan $\hat{\zeta}_2^2$ negatief wees en die ondergrens γ_{O2} nie gevind word nie. Neem dan γ_{02} = 0 en $\hat{\zeta}_2^2$ = 0.