Kapitel 6

MANOVA

6.1 Inferens om väntevärde

6.2 6.2 Parvisa observationer

6.2.1 Univariat

X_{1j} = respons för behandling 1 för j:te observationen
X_{2j} = respons för behandling 2 för j:te observationen
$j = 1, 2, \ldots, n$

Bilda differens $D_j = X_{1j} - X_{2j}$.

6.2.2 Multivariat

$\boldsymbol{X}_{1j}$ = respons för behandling 1 för j:te observationen
$\boldsymbol{X}_{2j}$ = respons för behandling 2 för j:te observationen
$j = 1, 2, \ldots, n$

Bilda differens $\boldsymbol{D}_j = \boldsymbol{X}_{1j} - \boldsymbol{X}_{2j}$. $E(\boldsymbol{D}_j) = \boldsymbol{\delta}$, $Kov(\boldsymbol{D}_j) = \boldsymbol{\Sigma}_D$.

Om $\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_n$ oberoende observationer från $\mathcal{N}_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_D)$ kan hypotesen $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ dvs $H_0 : \boldsymbol{\delta} = \boldsymbol{0}$ prövas med

$$T^2 n = \bar{\boldsymbol{D}}' \boldsymbol{S}_D^{-1} \bar{\boldsymbol{D}}$$

och $\frac{n-p}{(n-1)p} T^2$ är fördelad som $F_{p,n-p}$ om H_0 är sann.

Vi kan också bestämma simultana konfidensintervall för $\mu_{1i} - \mu_{2i}$.

6.2.3 Upprepade mätningar för jämförelse av behandlingar

En responsvariabel med q behandlingar.

X_{ij} =respons på i :tex behandlingen för individ j

$$\boldsymbol{X}_j' = (X_{1j}, X_{2j}, \ldots, X_{qj})$$

j :te observationen. $j = 1, 2, \ldots, n$. $\boldsymbol{\mu} = E(\boldsymbol{X}_j)$. Pröva hypotesen

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_q$$

Denna hypotes kan undersökas med kontraster:

$$H_{01} : \mu_1 - \mu_2 = \mu_1 - \mu_3 = \cdots = \mu_1 - \mu_q = 0$$

$$\begin{pmatrix} \mu_1-\mu_2 \\ \mu_1-\mu_3 \\ \vdots \\ \mu_1-\mu_q \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix} = \boldsymbol{C}_1\boldsymbol{\mu}$$

$$H_{02}: \mu_2-\mu_1 = \mu_3-\mu_2 = \dots = \mu_q-\mu_{q-1} = 0$$

$$\begin{pmatrix} \mu_2-\mu_1 \\ \mu_3-\mu_2 \\ \vdots \\ \mu_q-\mu_{q-1} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix} = \boldsymbol{C}_2\boldsymbol{\mu}$$

kan prövas med

$$T^2 = n(\boldsymbol{C}\bar{\boldsymbol{x}})'(\boldsymbol{CSC}')^{-1}\boldsymbol{C}\bar{\boldsymbol{x}}$$

6.3 6.4 Jämförelse av flera väntevärdesvektorer (envägs MANOVA)

Antag att vi har g oberoende stickprov

$$\begin{array}{cccc} \boldsymbol{X}_{11}, & \boldsymbol{X}_{12}, & \dots, & \boldsymbol{X}_{1n_1} \\ \boldsymbol{X}_{21}, & \boldsymbol{X}_{22}, & \dots, & \boldsymbol{X}_{2n_2} \\ \\ \boldsymbol{X}_{g1}, & \boldsymbol{X}_{g2}, & \dots, & \boldsymbol{X}_{gn_g} \end{array}$$

Varje vektor har dimensionen p.

Antaganden

- Det i:te stickprovet kommer från en population med väntevärdesvektor $\boldsymbol{\mu}_i$.
- Populationerna har alla en och samma kovariansmatris $\boldsymbol{\Sigma}$.
- Varje population är multivariat normalfördelad.

$\boldsymbol{X}_{ij} \in \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$.

Anmärkning: Om $g = 2$ är situationen den vid två stickprov och Hotellings T^2-test.

6.3.1 Univariat

Om $p = 1$, det univariata fallet anbefalles i denna situation en ANOVA test: Skriv $\mu_i = \mu + \tau_i$ där μ anger ett generellt populationsvärde och τ_i "effekten" av behandling i. I en population gäller uppdelningen

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

där ε_{ij} är oberoende $\mathcal{N}(0, \sigma^2)$.

Vi vill testa $H_0: \mu_1 = \mu_2 = \dots = \mu_g$ eller med parametriseringen $\mu_i = \mu + \tau_i$,

$$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$$

Denna hypotes prövas med F-test

$$F = \frac{MS_{\text{mellan}}}{MS_{\text{inom}}} = \frac{SS_{\text{mellan}}/(g-1)}{SS_{\text{inom}}/(n-g)}$$

där totalsumman delats upp

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})' = \sum_{i=1}^{g} n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})' + \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'$$

i

$$SS_{\text{mellan}} = \sum_{i=1}^{g} n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'$$

och

$$SS_{\text{inom}} = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'$$

Om H_0 är sann är F fördelad som $F_{g-1,n-g}$.

6.3.2 Multivariat

Dela upp den p-dimensionella observationen $\boldsymbol{x}_{ij}$ i ett generellt medelvärde $\bar{\boldsymbol{x}}$, en skattad "behandlingseffekt" $\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}$ och en "residual" $\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i$:

$$\boldsymbol{x}_{ij} = \bar{\boldsymbol{x}} + \bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}} + \boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i$$

viket ger

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}})' = \sum_{i=1}^{g} n_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})' + \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'$$

$$\boldsymbol{W} = \sum_{i=1}^{g}(n_i-1)\boldsymbol{S}_i$$

$$\boldsymbol{B} = \sum_{j=1}^{g} n_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})'$$

Test av H_0 : alla $\boldsymbol{\mu}_i$ lika, eller $H_0 : \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_g = \boldsymbol{0}$ ges av

$$\Lambda^* = \frac{|\boldsymbol{W}|}{|\boldsymbol{T}|} = \frac{|\boldsymbol{W}|}{|\boldsymbol{B}+\boldsymbol{W}|}$$

(Wilks lambda). Om H_0 är sann och n stor är

$$-((n+1)-(p+g)/2)\ln\Lambda^*$$

fördelad som $\chi^2(p(g-1))$.

Det existerar också exakta F-test för specialfall.

6.4 6.5 Simultana konfidensintervall för behandlingseffekter

Bonferroniintervall

$$\tau_{ki} - \tau_{li} = \mu_{ki} - \mu_{li} = \bar{x}_{ki} - \bar{x}_{li} \pm t_{n-g}(\alpha/(pg(g-1))\sqrt{\frac{w_{ii}}{n-g}(\frac{1}{n_k}+\frac{1}{n_l}))}$$

$i = 1, 2, \ldots, p,\ k, l = 1, \ldots, g, k < l$

w_{ii} är i:te diagonalelementet i $\boldsymbol{W}$.

6.5 6.6 Profilanalys

p behandlingar till 2 eller flera grupper. Alla responsvariablerna mäts på samma skala (jämförbara enheter).

Profil μ_i som funktion av i.

2 grupper: väntervärdesvektorer $\boldsymbol{\mu}_1$ och $\boldsymbol{\mu}_2$.

3 frågor

1. Är profilerna parallella?
2. Om profilerna är paralella, sammanfaller de då?
3. Om profilerna sammanfaller, har de konstant värde?

1) H_{01} innebär att $\mu_{1i} - \mu_{1,i-1} = \mu_{2i} - \mu_{2,i-1}$, $i = 2, 3, \ldots, p$ eller

$$\boldsymbol{C}\boldsymbol{\mu}_1 = \boldsymbol{C}\boldsymbol{\mu}_2$$

där

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix} : (p-1) \times p$$

Hypotesen kan prövas med en 2-grupps T^2-test

$$T^2 = (\boldsymbol{C}\bar{\boldsymbol{X}}_1 - \boldsymbol{C}\bar{\boldsymbol{X}}_2)'((\frac{1}{n_1} + \frac{1}{n_2})\boldsymbol{CSC}')^{-1}(\boldsymbol{C}\bar{\boldsymbol{X}}_1 - \boldsymbol{C}\bar{\boldsymbol{X}}_2)$$

2) H_{02} innebär att $\sum_{i=1}^{p} \mu_{1i} = \sum_{i=1}^{p} \mu_{2i}$ eller

$$\boldsymbol{1}'\boldsymbol{\mu}_1 = \boldsymbol{1}'\boldsymbol{\mu}_2$$

Vi prövar H_{02} med ett vanligt t-test (t^2-test)

$$t^2 = \boldsymbol{1}'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)((\frac{1}{n_1} + \frac{1}{n_2})\boldsymbol{1}'\boldsymbol{S}\boldsymbol{1})^{-1}\boldsymbol{1}'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$$

3) H_{03} innebär att $\mu_1 = \mu_2 = \cdots = \mu_p$

$$\boldsymbol{C}\boldsymbol{\mu} = \boldsymbol{0}$$

Denna hypotes prövas med T^2-test:

$$T^2 = (n_1 + n_2)(\boldsymbol{C}\bar{\boldsymbol{x}})'(\boldsymbol{CSC}')^{-1}\boldsymbol{C}\bar{\boldsymbol{x}}$$

6.6 6.7 2-sidig MANOVA

6.6.1 Univariat

$$X_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$i = 1, 2, \ldots, g, j = 1, 2, \ldots, b, k = 1, 2, \ldots, n$

g nivåer på faktor 1, b nivåer på faktor 2, n oberoende observationer för varje kombination av nivåer på faktor 1 och 2.

Kvadratsummeuppdelning:

$$SS_{\text{tot(korr)}} = SS_{\text{faktor 1}} + SS_{\text{faktor 2}} + SS_{\text{samspel}} + SS_{\text{residual}}$$

6.6.2 Multivariat

$$\boldsymbol{X}_{ijk} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\varepsilon}_{ijk}$$

$\boldsymbol{\varepsilon}_{ijk} \in \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$.

Uppdelning av matriser med kvadratsummor och korsprodukter (SSP)

$$SSP_{\text{tot(korr)}} = SSP_{\text{faktor 1}} + SSP_{\text{faktor 2}} + SSP_{\text{samspel}} + SSP_{\text{residual}}$$

Test av H_0 : alla $\boldsymbol{\gamma}_{ij} = \mathbf{0}$ sker med

$$\Lambda^* = \frac{|SSP_{\text{residual}}|}{|SSP_{\text{samspel}} + SSP_{\text{residual}}|}$$

Det gäller att

$$-(gb(n-1) - \frac{p+1-(g-1)(b-1)}{2} \ln \Lambda^*$$

är approximativt $\chi^2((g-1)(b-1)p)$ fördelad om H_0 är sann.

Test av H_0 : alla $\boldsymbol{\tau}_i = \mathbf{0}$ sker med

$$\Lambda^* = \frac{|SSP_{\text{residual}}|}{SSP_{\text{faktor 1}} + SSP_{\text{residual}}|}$$

Det gäller att

$$-(gb(n-1) - \frac{p+1-(g-1)}{2} \ln \Lambda^*$$

är approximativt $\chi^2((g-1)p)$ fördelad om H_0 är sann.

Test av H_0 : alla $\boldsymbol{\beta}_j = \mathbf{0}$ sker analogt.