Biostat 250B HW7

Elvis Cui
Han Cui
Dept. of Biostat
UCLA

1 Show that the statistic for testing equality of means from two normal populations with different variances using sample sizes n_1 and n_2 has a t-distribution with degree of freedom approximately equal to

$$(g_1 + g_2)^2/\{g_1^2/(n_1 - 1) + g_2^2/(n_2 - 1)\},$$

where $g_i = s_i^2/n_i$ i=1, 2and s_1^2 and s_2^2 are the sample variances.

2 Suppose you conduct an experiment with 3 factors A, B, C and the two factors A, B are fixed and C is a random factor. The number of levels for each of the factors are a, b, and c, respectively, with n observations per cell. Find the mean square errors for all the effects, like we did in class, and provide the test statistics for testing all effects in the model and state the rejection rules.

3 (Linear mixed effects model question) Consider the model

$$Y = X\,\beta + Z\,u + e$$

Where x and Z are known matrices, b is an unobservable vector of fixed effects and u is an unobservable vector of random effects with E(u) = 0, Cov(u) = D, and Cov(u, e) =). Let Cov(e) = R.

a. Show that $V = \mathrm{Cov}(Y) = Z\,D\,Z' + R$.
b. Write down the likelihood functions for Y|u and that for u. Then differentiate the joint likelihood function of y and u to obtain the maximum likelihood estimators for β and u (strictly speaking, the estimate of u is the predicted value of the random variable u). Show that the resulting equations to solve for these estimators β^* and u^* are

$$X'\,R^{-1}\,X\,\beta \;+\; X'\,R^{-1}\,Z\,u \;=\; X'\,R^{-1}\,Y \quad \text{and} \quad Z'\,R^{-1}\,X\,\beta + D^{-1} \;+\; Z'\,R^{-1}\,Z\,u \;=\; Z'\,R^{-1}\,y.$$

c. Assuming that all the below indicated inverses exist, use the formula from Biostat 250A

$$V^{-1} = R^{-1} + R^{-1} - R^{-1}\,Z\,(D^{-1} \;+\; Z'\,R^{-1}\,Z)\,Z'\,R^{-1},$$

to show that the sought estimators are

$$b^* \;=\; (X'\,V^{-1}\,X)'\,X'\,V^{-1}\,y \quad \text{and}\quad u^* \;=\; DZ'V^{-1}\,(Y - X\,\beta^*).$$

[This last question requires quite a bit of algebra but is doable based on what we learned from Biostat 250A and I will give guidance on Monday]

Q1:

1 Show that the statistic for testing equality of means from two normal populations with different variances using sample sizes n_1 and n_2 has a t-distribution with degree of freedom approximately equal to

$$(g_1 + g_2)^2/\{g_1^2/(n_1 - 1) + g_2^2/(n_2 - 1)\},$$

where $g_i = s_i^2/n_i$ i=1, 2and s_1^2 and s_2^2 are the sample variances.

Sol. we have $(n_i-1)S_i^2/\sigma^2 \overset{ind}{\sim} \chi^2_{n_i-1}, \; i=1,2$

Lemma (Saterwaite): Let $u_1, \dots, u_k$ be independent χ^2 rvs with $r_1, \dots, r_k$ df, Let $u = \sum_{i=1}^{k} a_i u_i$, ν be a constant

then $\frac{\nu u}{\mathbb{E}u} \sim \chi^2_b$

where $b = \nu = \frac{\nu^2}{(\mathbb{E}u)^2} \sum_{i=1}^{k} (a_i \mathbb{E}u_i)^2 / r_i$

By method of moments,

$$\hat{b} = \left(\sum_{i=1}^{k} a_i u_i\right)^2 \Big/ \left\{\left(\sum_{i=1}^{k} (a_i u_i)^2\right) / r_i\right\}$$

Our test statistic:

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{g_1 + g_2}} \quad (*)$$

$$Var(\bar{X}_1 - \bar{X}_2) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2$$

$$\Rightarrow (\bar{X}_1 - \bar{X}_2) / \sigma\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}} \sim N(0,1)$$

$$(*) = \frac{N(0,1)}{\sqrt{\frac{g_1+g_2}{\sigma^2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}}}$$

Need to find dist. of

$$\sqrt{\frac{1}{\sigma^2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}(g_1 + g_2)}$$

by letting

$k=2$, $a_1 = \frac{1}{n_1(n_1-1)}$, $a_2 = \frac{1}{n_2(n_2-1)}$,

$u_1 = \frac{(n_1-1)n_1}{\sigma^2} g_1$, $u_2 = \frac{n_2(n_2-1)}{\sigma^2} g_2$

$r_1 = n_1 - 1$, $r_2 = n_2 - 1$,

we have

$$\frac{1}{\sigma^2}(g_1 + g_2) \mathrel{\dot\sim} \chi^2_{\hat{b}},$$

$$\hat{b} = (g_1 + g_2)^2 / (g_1^2/n_1 - 1 + g_2^2/n_2 - 1)$$

Also we have

$\mathbb{E}g_1 = \frac{\sigma^2}{n_1}$, $\mathbb{E}g_2 = \frac{\sigma^2}{n_2}$,

$\mathbb{E}u = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

$$\Rightarrow b = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \Big/ \left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right)$$

Thus,

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{g_1 + g_2}} \mathrel{\dot\sim} t_{\hat{b}} \quad \square$$

Q2:

2 Suppose you conduct an experiment with 3 factors A, B, C and the two factors A, B are fixed and C is a random factor. The number of levels for each of the factors are a, b, and c, respectively, with n observations per cell. Find the mean square errors for all the effects, like we did in class, and provide the test statistics for testing all effects in the model and state the rejection rules.

Sol.

	A a F i	B b F j	C c R k	n R l	EMS
τ_i	0	b	c	n	①
β_j	a	0	c	n	②
γ_k	a	b	1	n	③
$(\tau\beta)_{ij}$	0	0	c	n	④
$(\tau\gamma)_{ik}$	1	b	1	n	⑤
$(\beta\gamma)_{jk}$	a	1	1	n	⑥
$(\tau\beta\gamma)_{ijk}$	1	1	1	n	⑦
ε_{ijkl}	1	1	1	1	⑧

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}.$$

① $= bcn \frac{\sum_i \tau_i^2}{a-1} + bn\,\sigma_{\tau\gamma}^2 + \sigma^2$

② $= acn \frac{\sum_i \beta_j^2}{b-1} + an\,\sigma_{\beta\gamma}^2 + \sigma^2$

③ $= abn\,\sigma_\gamma^2 + \sigma^2$

④ $= cn \frac{\sum_i \sum_j (\tau\beta)_{ij}}{(a-1)(b-1)} + n\,\sigma_{\tau\beta\gamma}^2 + \sigma^2$

⑤ $= bn\,\sigma_{\tau\gamma}^2 + n\,\sigma_{\tau\beta\gamma}^2 + \sigma^2$

⑥ $= an\,\sigma_{\beta\gamma}^2 + n\,\sigma_{\tau\beta\gamma}^2 + \sigma^2$

⑦ $= n\,\sigma_{\tau\beta\gamma}^2 + \sigma^2$

⑧ $= \sigma^2$

Test statistic:

$H_0: (\tau\beta)_{ij} = 0 \ \forall i,j \rightarrow$

MS_{AB}/MS_{ABC}

$H_0: \sigma_\gamma^2 = 0 \rightarrow MS_C/MSE$

$H_0: \sigma_{\tau\gamma}^2 = 0 \rightarrow MS_{AC}/MSE$

$H_0: \sigma_{\beta\gamma}^2 = 0 \rightarrow MS_{BC}/MSE$

$H_0: \sigma_{\tau\beta\gamma}^2 = 0 \rightarrow MS_{ABC}/MSE$

$H_0: \tau_i = 0 \ \forall i \rightarrow$

$MS_A/(MS_{AC} - MS_{ABC} + MSE)$

$H_0: \beta_j = 0 \ \forall j \rightarrow$

$MS_B/(MS_{BC} - MS_{ABC} + MSE)$

Q3:

3 (Linear mixed effects model question) Consider the model

$$Y = X\beta + Z u + e$$

Where x and Z are known matrices, b is an unobservable vector of fixed effects and u is an unobservable vector of random effects with E(u) = 0, Cov(u) = D, and Cov(u, e) =). Let Cov(e) = R.

a. Show that $V = \text{Cov}(Y) = Z D Z' + R$.
b. Write down the likelihood functions for Y|u and that for u. Then differentiate the joint likelihood function of y and u to obtain the maximum likelihood estimators for β and u (strictly speaking, the estimate of u is the predicted value of the random variable u). Show that the resulting equations to solve for these estimators β^* and u^* are

$$X' R^{-1} X \beta + X' R^{-1} Z u = X' R^{-1} Y \quad \text{and} \quad Z' R^{-1} X \beta + D^{-1} + Z' R^{-1} Z u = Z' R^{-1} y.$$

c. Assuming that all the below indicated inverses exist, use the formula from Biostat 250A

$$V^{-1} = R^{-1} + R^{-1} - R^{-1} Z (D^{-1} + Z' R^{-1} Z) Z' R^{-1},$$

to show that the sought estimators are

$$b^* = (X' V^{-1} X)' X' V^{-1} y \quad \text{and } u^* = DZ'V^{-1} (Y - X\beta^*).$$

[This last question requires quite a bit of algebra but is doable based on what we learned from Biostat 250A and I will give guidance on Monday]

Sol. (a): $\text{Cov}\, Y = \text{Cov}(X\beta + Zu + e)$

$= \text{Cov}(Zu + e)$ fixed eff.

$= \text{Cov}(Zu) + \text{Cov}(e)$ ind. assump.

$= ZDZ^T + R$

$\Rightarrow V = ZDZ^T + R$

(b): ①: $Y|u \sim \mathcal{N}(X\beta + Zu, R)$

$$f(Y|u) = \frac{1}{(2\pi)^{\frac{n}{2}} |R|^{\frac{1}{2}}} e^{-(Y-\mu)^T R^{-1} (Y-\mu)}$$

$\mu = X\beta + Zu.$

$$f(u) = \frac{1}{(2\pi)^{\frac{n}{2}} |D|^{\frac{1}{2}}} e^{-u^T D^{-1} u}$$

$\Rightarrow L(\beta, u|Y) = f(Y|u) f(u)$

$\propto e^{-(Y-\mu)^T R^{-1} (Y-\mu) - u^T D^{-1} u}$

② $(\hat{\beta}, \hat{u})_{MLE} = \text{argmax} \log L(\beta, u|Y)$

But

$\ell(\beta, u) = \log L(\beta, u|Y) \propto$

$-(Y-\mu)^T R^{-1} (Y-\mu) - u^T D^{-1} u.$

$\mu = X\beta + Zu.$

$\Rightarrow$

$$\frac{\partial \ell}{\partial \beta} = -2X^T R^{-1} (Y-\mu) = 0$$

or, $X^T R^{-1} Y = X^T R^{-1} (X\beta + Zu)$ (1)

$$\frac{\partial \ell}{\partial u} = -2Z^T R^{-1} (Y-\mu) - 2D^{-1} u = 0$$

or, $Z^T R^{-1} Y = Z^T R^{-1} (X\beta + Zu) + D^{-1} u.$ (2)

(C): From (2):

$u^* = (Z^T R^{-1} Z + D^{-1})^{-1} Z^T R^{-1} (Y - X\beta^*)$

substitute this into (1):

$X^T R^{-1} Y = X^T R^{-1} X\beta^* + X^T R^{-1} Z (Z^T R^{-1} Z + D^{-1})^{-1} Z^T R^{-1} (Y - X\beta^*)$

Since

$R^{-1} - R^{-1} Z (D^{-1} + Z R^{-1} Z^T)^{-1} Z^T R^{-1} = V^{-1}$ (△)

we have

$X^T R^{-1} Y = X^T R^{-1} X\beta^* + X^T (R^{-1} - V^{-1})(Y - X\beta^*)$

$\Rightarrow X V^{-1} X \beta^* = X V^{-1} Y$

$\Rightarrow \beta^* = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$

Also,

$(Z^T R^{-1} Z + D^{-1}) D Z^T = Z^T R^{-1} Z D Z^T + Z^T$

$= Z^T R^{-1} (Z^T D Z + R)$

$= Z^T R^{-1} V$

$\Rightarrow D Z^T V^{-1} = (Z R^{-1} Z^T + D^{-1}) Z R^{-1}$

$\Rightarrow u^* = D Z^T V^{-1} (Y - X\beta^*)$ □.