Multivariate Statistical Analysis, Exercise 6, Fall 2011, Prof. S.K. Jeng

November 17, 2011 TA: H.C. Cheng

(Comparing Tow Mean Vector) Assume there are two populations $X_{i1} \sim N_p(\mu_1, \Sigma_1)$, with $i = 1, \ldots, n_1$ and $X_{i2} \sim N_p(\mu_2, \Sigma_2)$, with $j = 1, \ldots, n_2$, where all the variables are independent. The realizations are recorded in the data set as 200×6 matrix where the observations $1 \sim 100$ come from the first population and the other 100 observations are the second one. The number of variate $p = 6$. Now we want to compare the tow mean vector; that is

$$H_1 : \mu_1 = \mu_2, \text{ vs. } H_2 : \mu_1 \neq \mu_2.$$

Please answer the following questions.

(a) We first consider the problem with the same covariance matrix $\Sigma_1 = \Sigma_2 = \Sigma$. Let $\overline{x}_k$, $S_k, k = 1, 2$ and $\delta = \mu_1 - \mu_2$. we have

$$(\overline{x}_1 - \overline{x}_2) \sim N_p\left(\delta, \frac{n_1 + n_2}{n_1 n_2}\Sigma\right)$$
$$n_1 S_1 + n_2 S_2 \sim W_p(\Sigma, n_1 + n_2 - 2).$$

Let $S = (n_1 + n_2)^{-1}(n_1 S_1 + n_2 S_2)$ be the weighted mean of S_1 and S_2. Since the two samples are independent and since S_k is independent of $\overline{x}_k, k = 1, 2$, it follows that S is independent of $(\overline{x}_1 - \overline{x}_2)$. Hence,

$$\frac{(n_1 n_2)(n_1 + n_2 - 2)}{(n_1 + n_2)^2}\{(\overline{x}_1 - \overline{x}_2) - \delta\}^T S^{-1}\{(\overline{x}_1 - \overline{x}_2) - \delta\} \sim T^2(p, n_1 + n_2 - 2),$$

or

$$\{(\overline{x}_1 - \overline{x}_2) - \delta\}^T S^{-1}\{(\overline{x}_1 - \overline{x}_2) - \delta\} \sim \frac{p(n_1 + n_2)^2}{(n_1 + n_2 - p - 1)n_1 n_2}F_{p,n_1+n_2-p-1},$$

where $T^2(p, n)$ is the Hotelling T^2-distribution with p and n degrees of freedom, and $F_{1-\alpha;n,m}$ is the $1 - \alpha$ quantile of the F-distribution with n and m degrees of freedom.

This result can be used to test $H_1 : \delta = 0$ or to constructa confidence region for $\delta \in \mathbb{R}^p$. The rejection region is given by:

$$\frac{n_1 n_2(n_1 + n_2 - p - 1)}{p(n_1 + n_2)^2}(\overline{x}_1 - \overline{x}_2)^T S^{-1}(\overline{x}_1 - \overline{x}_2) \geq F_{1-\alpha;p,n_1+n_2-p-1}.$$

A $(1-\alpha)$ confidence region for δ is given by the ellipsoid centered at $(\overline{x}_1-\overline{x}_2)$

$$\{\delta-(\overline{x}_1-\overline{x}_2)\}^T S^{-1}\{\delta-(\overline{x}_1-\overline{x}_2)\} \leq \frac{p(n_1+n_2)^2}{(n_1+n_2-p-1)(n_1n_2)} F_{1-\alpha;p,n_1+n_2-p-1},$$

and the simultaneous confidence intervals for all linear combinations $a^T\delta$ of the elements δ are given by

$$a^T\delta \in a^T(\overline{x}_1-\overline{x}_2) \pm \sqrt{\frac{p(n_1+n_2)^2}{(n_1+n_2-p-1)(n_1n_2)} F_{1-\alpha;p,n_1+n_2-p-1} a^T S a}.$$

In particular we have at the $(1-\alpha)$ level, for $j=1,\ldots,p$,

$$\delta_j \in (\overline{x}_{1j}-\overline{x}_{2j}) \pm \sqrt{\frac{p(n_1+n_2)^2}{(n_1+n_2-p-1)(n_1n_2)} F_{1-\alpha;p,n_1+n_2-p-1} S_{jj}}.$$

With $\alpha=0.05$, do we reject the hypothesis? Please calculate the confidence intervals for the differences $\delta_j=\mu_{1j}-\mu_{2j}, j=1,\ldots,p$.

(b) For the case of unequal covariance matrices and large samples, we have

$$(\overline{x}_1-\overline{x}_2) \sim N_p\left(\delta, \frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2}\right).$$

Therefore,

$$(\overline{x}_1-\overline{x}_2)^T\left(\frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2}\right)^{-1}(\overline{x}_1-\overline{x}_2) \sim \chi^2_p.$$

Since S_k is a consistent estimator of Σ_k for $k=1,2$, we have

$$(\overline{x}_1-\overline{x}_2)^T\left(\frac{S_1}{n_1}+\frac{S_2}{n_2}\right)^{-1}(\overline{x}_1-\overline{x}_2) \to \chi^2_p \quad \text{in distribution.}$$

Hence the rejection region for $\delta=0$ at the level α will be

$$(\overline{x}_1-\overline{x}_2)^T\left(\frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2}\right)^{-1}(\overline{x}_1-\overline{x}_2) > \chi_{1-\alpha,p},$$

where $\chi^2_{1-\alpha,p}$ is the $1-\alpha$ quantile of the χ^2 distribution with p degrees of freedom.

The $(1-\alpha)$ confidence intervals for $\delta_j, 1,\ldots,p$ are

$$\delta_j \in (\overline{x}_1-\overline{x}_2) \pm \sqrt{\chi^2_{1-\alpha,p}\left(\frac{S_{1,jj}}{n_1}+\frac{S_{2,jj}}{n_2}\right)}.$$

Similarly, Do we reject the hypothesis with $\alpha=0.05$ and what are the confidence interfvals of $\delta_j, j=1,\ldots,5$.