**Multivariate Statistical Analysis Mid Term Solution**

**November 9, 2012**

**4 pages, 14 problems, 100 points**

1. Eigenvalues: 9, 1

Corresponding normalized eigenvectors: $\left[\begin{matrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{matrix}\right]'$ and $\left[\begin{matrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{matrix}\right]^{'}$ . (7%)

----------------------------------------------------------------------------------------------

1. $\left[\begin{matrix}5&4\\4&5\end{matrix}\right]=9\left[\begin{matrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{matrix}\right]\left[\begin{matrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{matrix}\right]+\left[\begin{matrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\end{matrix}\right]\left[\begin{matrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\end{matrix}\right]=\left[\begin{matrix}\frac{9}{2}&\frac{9}{2}\\\frac{9}{2}&\frac{9}{2}\end{matrix}\right]+\left[\begin{matrix}\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}\end{matrix}\right]$. (7%)

----------------------------------------------------------------------------------------------

1. $\max\_{x\ne 0}\frac{x'Sx}{x'x}=9$, $\min\_{x\ne 0}\frac{x'Sx}{x'x}=1$. (7%)

----------------------------------------------------------------------------------------------

1. (1)When at least one deviation vector lies in the (hyper) plane formed by any one linear combinations of the others

In other words, the columns of the matrix of deviations are linearly dependent

(2) When the sample size is not larger than the number of variables.

 (7%)

----------------------------------------------------------------------------------------------

1. $\hat{μ}=\overbar{x}=\left[\begin{matrix}4\\6\end{matrix}\right],$ $\hat{Σ}=\frac{1}{4} \left\{\left[\begin{matrix}3-4\\6-6\end{matrix}\right]\left[\begin{matrix}3-4&6-6\end{matrix}\right]+\left[\begin{matrix}4-4\\4-6\end{matrix}\right]\left[\begin{matrix}4-4&4-6\end{matrix}\right]+\left[\begin{matrix}5-4\\7-6\end{matrix}\right]\left[\begin{matrix}5-4&7-6\end{matrix}\right]+\left[\begin{matrix}4-4\\7-6\end{matrix}\right]\left[\begin{matrix}4-4&7-6\end{matrix}\right]\right\}=\frac{1}{4}\left[\begin{matrix}2&1\\1&6\end{matrix}\right]$. (7%)

----------------------------------------------------------------------------------------------

1. . (7%)

----------------------------------------------------------------------------------------------

1. . (7%)

----------------------------------------------------------------------------------------------

1. . (7%)

----------------------------------------------------------------------------------------------

1. . (7%)

----------------------------------------------------------------------------------------------

1. one-at-a-time confidence intervals: each interval consider only one dimension at a time, don’t care the other dimensions.

 simultaneous confidence intervals: all linear combinations of the unknown parameters.

Thus  simultaneous confidence intervals are with more strict constraints, and need longer intervals to satisfy them . (7%)

----------------------------------------------------------------------------------------------

1. Initial estimate , , , ,

 .

 Partition , , and predict

 

 

 

 

 

 ,

 

. (7%)

----------------------------------------------------------------------------------------------

1. Eq. (6-21), p. 285, Textbook

 .

Eq. (6-24), p. 286, Textbook

 , larger than the critical value . Thus the hypothesis is rejected at 10% significance level. (7%)

----------------------------------------------------------------------------------------------

Follow the equation in p. 291,

1. . (7%)

----------------------------------------------------------------------------------------------

1. Assume that the pair of paired samples are with *n* data, mean difference $\overbar{d}$, and standard deviation of difference *sd*. The confidence interval by the paired comparison is $\overbar{d}\pm t\_{n-1}(\frac{α}{2})\frac{s\_{d}}{\sqrt{n}}$. While if we use the unpaired method to the same paired samples, the confidence interval will be $\overbar{d}\pm t\_{n+n-2}\left(\frac{α}{2}\right)\sqrt{\left(\frac{1}{n}+\frac{1}{n}\right)}s\_{pooled}$

Since $s\_{pooled}=\frac{n-1}{n+n-2}s\_{d}+\frac{n-1}{n+n-2}s\_{d}=s\_{d},$ the confidence interval using unpaired method can be reduced to

$$\overbar{d}\pm t\_{n+n-2}\left(\frac{α}{2}\right)\sqrt{\left(\frac{1}{n}+\frac{1}{n}\right)}s\_{pooled}=\overbar{d}\pm t\_{2n-2}\left(\frac{α}{2}\right)\sqrt{\frac{2}{n}}s\_{d} $$

The lengths of the confidence intervals for the paired and the unpaired methods will be $2t\_{n-1}(\frac{α}{2})\frac{s\_{d}}{\sqrt{n}}$ and $2t\_{2n-2}(\frac{α}{2})\frac{\sqrt{2}s\_{d}}{\sqrt{n}}$. Because for *n* not too small, $t\_{n-1}(\frac{α}{2})$ does not larger than $\sqrt{2}t\_{2n-2}(α)$, $2t\_{n-1}(\frac{α}{2})\frac{s\_{d}}{\sqrt{n}}$ is smaller than $2t\_{2n-2}(\frac{α}{2})\frac{\sqrt{2}s\_{d}}{\sqrt{n}}$. Thus the paired method is more accurate. (9%)