STAT 8250 Hwk #4 – Solutions

6.5 $T^2 = n(\underset{\sim}{C}\bar{\underset{\sim}{x}})'(\underset{\sim}{C}\underset{\sim}{S}\underset{\sim}{C}')^{-1}\underset{\sim}{C}\bar{\underset{\sim}{x}}$

$$= 40\left(\begin{pmatrix}1 & -1 & 0\\ 0 & 1 & -1\end{pmatrix}\begin{pmatrix}46.1\\57.3\\50.4\end{pmatrix}\right)'\left(\begin{pmatrix}1 & -1 & 0\\ 0 & 1 & -1\end{pmatrix}\begin{pmatrix}101.3 & 63.0 & 71.0\\ 63.0 & 80.2 & 55.6\\ 71.0 & 55.6 & 97.4\end{pmatrix}\begin{pmatrix}1 & 0\\ -1 & 1\\ 0 & -1\end{pmatrix}\right)^{-1}$$

$$\times \begin{pmatrix}1 & -1 & 0\\ 0 & 1 & -1\end{pmatrix}\begin{pmatrix}46.1\\57.3\\50.4\end{pmatrix} = 90.49$$

(See ex6-5.sas + ex6-5.lst)

The critical value for $\alpha = .05$ of this test statistic is

$$\frac{(n-1)(q-1)}{n-q+1}F_{.05}(q-1, n-q+1) = \frac{(40-1)(3-1)}{40-3+1}F_{.05}(3-1, 40-3+1)$$

$$= 6.66$$

So we reject $H_0: \mu_1 = \mu_2 = \mu_3$ and conclude that the severity measures do not all have the same population mean.

p value is $p = 1.25 \times 10^{-10}$

b) $100(1-\alpha)\%$ "Max-t simultaneous confidence intervals ~~of a given~~ for a contrast among the elements of $\underset{\sim}{\mu}$ of the form $\underset{\sim}{c}'\underset{\sim}{\mu}$ are given by

$$\underset{\sim}{c}'\bar{\underset{\sim}{x}} \pm \sqrt{\frac{(n-1)(q-1)}{n-q+1}F_{\alpha}(q-1, n-q+1)}\sqrt{\frac{\underset{\sim}{c}'\underset{\sim}{S}\underset{\sim}{c}}{n}}$$

6.5 b) (continued) Notice that $\mu_1 - \mu_2 = 1\mu_1 + (-1)\mu_2 + 0\mu_3$ is of this form with $\underline{c} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $\mu_1 - \mu_3$ is of this form with $\underline{c} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\mu_2 - \mu_3$ is of this form with $\underline{c} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$. Therefore, our max-t intervals are

$$(1\ -1\ 0)\begin{pmatrix} 46.1 \\ 57.3 \\ 50.4 \end{pmatrix} \pm \sqrt{6.66}\ \sqrt{\frac{(1\ -1\ 0)\, S \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}}{40}}$$

$= (-14.24, -8.16)$ for $\mu_1 - \mu_2$, and (similarly)

$(3.57, 10.23)$ for $\mu_2 - \mu_3$ and

$(-7.37, -1.23)$ for $\mu_1 - \mu_3$

Notice, none of these intervals cover 0 so all 3 severity measures have significantly diff't means from one-another.

6.17 The "Hint" given here seems to be something of a red herring. See ex6-17.sas and ex6-17.lst. In this program, the vast majority of the code is devoted to checking the m'variate normality assumption for each group (male & female). ~~e~~ I computed ~~the~~ ~~2~~ linear combinations $\hat{\underline{e}}_1'\underline{x}$, $\hat{\underline{e}}_3'\underline{x}$ based on the 1st and last eigenvectors of $\underline{S}$ within each group separately and then checked the univariate normality of these linear combinations, length, width and height within each sample. All 10 of the resulting random variables appear to be sufficiently normal (no Shapiro-Wilk statistics are significant, all Q-Q plots look good, and all Filliben correlation coefficients are high). Therefore, at least to justify normality, there's no reason to take log transformations. However, since the text recommends it, I computed tests and confidence intervals both for the original scale of the data and the log-scale.

I believe that the reason that log-transformations are recommended is that length, width and height have a multiplicative relationship with total shell volume. The log transformation transforms this relationship to an additive one. Nevertheless, I don't feel that its particularly justified in this problem

6.17a) See pp. 43-44 of ~~&~~ ex6-17.lst. On the original scale we get

$T^2 = 72.38 \quad p = 3.967\text{e-}9$

$\Rightarrow p < .05 \Rightarrow$ we reject H_0: Mean length, width, ht Vector same across genders.

On the log-scale we obtain the same conclusion because

$T^2 = 85.05, \; p = 4.36 \times 10^{-10} < .05$

b.) Based on the original scale the linear comb. of length, width + ht most responsible for rejecting H_0 is

$$-.27 \text{ length} - .14 \text{ width} + 1.25 \text{ height}$$

which seems to be a contrast between ~~&~~ size in the horizontal plane with height.

Based on the log scale the ~~as~~ linear comb is

$$-43.73 \log(\text{length}) - 8.71 \log(\text{width}) + 67.55 \log(\text{height})$$

c) On the original scale

Female Mean - Male Mean	Maxt	Bonferroni
length	(7.93, 37.41)	(10.34, 34.99)
width	(5.26, 23.33)	(6.74, 21.84)
ht	(6.04, 16.62)	(6.91, 15.75)

6.17 c) On the log-scale

Female mean - Male Mean	Max-t	Bonferroni
length	(.058, .29)	(.077, .27)
width	(.054, .24)	(.069, .22)
height	(.13, .35)	(.15, .33)

In either case, the Bonferroni intervals are narrower.

6.22 See ex6-22.sas + ex6-22.lst

$$S_1 = \begin{pmatrix} .144 & .00930 \\ & .0111 \end{pmatrix}, \quad S_2 = \begin{pmatrix} .0985 & .0412 \\ & .0391 \end{pmatrix}, \quad S_3 = \begin{pmatrix} .104 & .0476 \\ & .0754 \end{pmatrix}$$

Although S_1 ~~[illegible]~~ looks to be a bit different from S_2, S_3, the difference is not great and with a large sample size and a balanced design, should not effect the inferences based on a normal theory, constant var-cov, MANOVA.

MANOVA Table

Source of Variation	SSCP	df	Λ^*	F	P
Iris Varieties	$\begin{pmatrix} 11.34 & -22.93 \\ -22.93 & 80.41 \end{pmatrix}$	2	.0383	299.94	0
Error	$\begin{pmatrix} 16.96 & 4.81 \\ 4.81 & 6.16 \end{pmatrix}$	147			
Total	$\begin{pmatrix} 28.31 & -18.12 \\ -18.12 & 86.57 \end{pmatrix}$	149			

6.22 (continued) Using Bonferroni, we get the following intervals:

Iris Type Comparison	Variable	Bonferroni Interval
I vs II	Sepal Width	(.476, .840)
I vs III	Sepal Width	(.272, .636)
II vs III	Sepal Width	(-.386, -.0223)
I vs II	Petal Width	(-1.19, -.97)
I vs III	Petal Width	(-1.889, -1.671)
II vs III	Petal Width	(-.809, -.591)

6.26 See ex6-22.sas and ex6-22.lst. ex6-22.sas is highly annotated and contains the details of how I approached the analysis of this data set. The gist of the analysis is that the data set has a couple of possible outliers and a fairly high degree of non-m'variate normality. So, I analyzed the data both with and without the outliers removed and, because transformations were not helpful in satisfying the normality assumption, I analyzed the data on the original scale using large-sample version of Hotelling's T^2.

6.26 Relevant statistics appear on pp. 129-130 of ex6-22.lst. Either with or without the outliers, T^2 is highly significant and we reject the null hypothesis of equal means for the two populations of insects.
In either analysis, the linear combination of the random variables most responsible for rejecting H_0 seems to be a contrast between (X_3, X_5, X_6) and (X_4, X_7). Although the m'variate normality assumption does not appear to hold for this data set, the non-normality is not extreme, especially when outliers are removed (the chi-square Q-Q plot ~~[illegible]~~ looks pretty good). This, combined with the moderately large sample size and the use of large-sample inferential methods, should provide for reasonably accurate inferences. Since the H_0 was overwhelmingly rejected, it is nearly certain that the qualitative results would not have change even had we precisely accounted for the non-normality ~~in our~~ to ~~[illegible]~~ obtain an exact p-value or an optimal test.

Please read through ex6-22.sas carefully and look through the results in ex6-22.lst. This provides ~~the~~ the most complete illustration of a full m'variate analysis of a somewhat difficult data set that I have provided you with. It may be useful when you do your project.

Hwk#4 - Extra Problem: This problem is very similar to lab #4 See hwk4extra.sas and hwk4extra.lst.

a) $T^2 = 55.71$ $p = .0346$ so at $\alpha = .05$, we reject $H_0: \mu_1 = \mu_2 = \cdots = \mu_5$ (the mean response is the same over the 5 time periods).

b.) The L.R. test statistic for testing the H-F conditions is .08203 with p-value = .1818, so we don't reject the hypothesis that the H-F conditions are satisfied. This result justifies using the RCB analysis of variance to test for a time effect, which should be a more powerful analysis.

c) The appropriate test statistic is $F = \dfrac{MS_{Time}}{MS_E} = 4.15$ ($p = .0092$)

Since $p < .05$ we reject with a smaller p-value than we did in part (a).