STAT 8250 - Homework # 3 - Solution

5.1 a) $\underline{\bar{x}} = \begin{pmatrix} 6 \\ 10 \end{pmatrix}$ $n=4$, $p=2$, $\underline{S} = \begin{pmatrix} 8 & -10/3 \\ -10/3 & 2 \end{pmatrix}$

(see ex5-1.sas & .lst)

$$T^2 = n(\underline{\bar{x}} - \mu_0)' \underline{S}^{-1} (\underline{\bar{x}} - \mu_0)$$

$$= 4\begin{pmatrix} 6-7 & 10-11 \end{pmatrix} \begin{pmatrix} 8 & -10/3 \\ -10/3 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 6-7 \\ 10-11 \end{pmatrix} = 13.64$$

b) $T^2 \sim \frac{(n-1)p}{n-p} F(p, n-p) = 3F(2,2)$

(i.e., $\frac{1}{3}T^2 \sim F(2,2)$)

c) Using $\alpha = .05$ $F_{.05}(2,2) = 19.00$

Since $T^2 = 13.64 < 3(19) = 57$ we do not reject H_0

Equivalently, we can compute the p-value for T^2 (see ex5-1.lst, p.2) which gives $p = .1803$

Since $p = .1803 > \alpha = .05$ we fail to reject H_0.

Obtaining & reporting the p-value is preferrable to simply reporting the outcome of the test since it says something about the strength of the evidence against H_0.

5.4 a) The axes of the 90% confidence ellipsoid for μ are

$$\begin{pmatrix} .0508 \\ .998 \\ -.0291 \end{pmatrix}, \begin{pmatrix} -.574 \\ .0530 \\ .817 \end{pmatrix}, \begin{pmatrix} .817 \\ -.0249 \\ .575 \end{pmatrix}$$

and the ellipsoid extends along these axes 9.051, 1.361, and .729 units, respectively.
See ex5-4.sas and ex5-4.lst for these results.

b) See ex5-4.sas and ex5-4.lst. Q-Q plots for each of the three variables are given on pp. 2, 4, and 6 of ex5-4.lst. Scatterplots appear on pp. 8, 9, 10. The Q-Q plots all look good (linear), the r_Q values are all high, and the scatterplots are all (roughly!) elliptical. Therefore, the trivariate normal assumption appears to be justified.

5.5 $T^2 = n(\bar{x}-\underline{\mu}_0)' S^{-1} (\underline{\bar{x}}-\underline{\mu}_0) =$

$$= 42 \left(\begin{pmatrix} .564 \\ .603 \end{pmatrix} - \begin{pmatrix} .55 \\ .60 \end{pmatrix} \right)' \begin{pmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{pmatrix} \left(\begin{pmatrix} .564 \\ .603 \end{pmatrix} - \begin{pmatrix} .55 \\ .60 \end{pmatrix} \right)$$

$$= 1.17$$

$F_{.05}(2,40) = 3.23$ so the critical value for T^2 is

$$\frac{2(41)}{40}(3.23) = 6.62$$

Since $T^2 < 6.62$ we do not reject $H_0: \underline{\mu} = \begin{pmatrix} .55 \\ .60 \end{pmatrix}$ at significance level .05. The result is consistent with Figure 5.1 because $\begin{pmatrix} .55 \\ .60 \end{pmatrix}$ is inside the 95% confidence ellipse.

5.7 See ex5-7.sas and ex5-7.lst. The simultaneous 95% confidence intervals are

	Bonferroni	Max-t
μ_1	(3.64, 5.64)	(3.40, 5.88)
μ_2	(37.10, 53.70)	(35.05, 55.75)
μ_3	(8.85, 11.08)	(8.57, 11.36)

Bonferroni intervals are narrower.

5.9a) See ex5-9.sas and ex5-9.lst.
Based on all of the given data ($n=9$) we obtain

$\bar{\underline{x}} = \begin{pmatrix} 5.186 \\ 16.07 \end{pmatrix}$, $S^{-1} = \begin{pmatrix} .0508 & -.0276 \\ -.0276 & .0169 \end{pmatrix}$, $\frac{p(n-1)}{n(n-p)} F_{.10}(p, n-p) = \frac{7.446}{n} = \frac{7.446}{9} = .827$

(see p. 3 of ex5-9.lst).
Therefore, our 90% confidence ellipse for $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ is given by

$$\left\{ \underline{\mu} : \left(\begin{pmatrix} 5.186 \\ 16.07 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right)' \begin{pmatrix} .0508 & -.0276 \\ -.0276 & .0169 \end{pmatrix} \left(\begin{pmatrix} 5.186 \\ 16.07 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right) \le .827 \right\}$$

$$= \left\{ \underline{\mu} : .0508(5.186-\mu_1)^2 + 2(-.0276)(5.186-\mu_1)(16.07-\mu_2) + (16.07-\mu_2)^2(.0169) \le .827 \right\}$$

The boundary of this set is the ellipse given by the equation $.0508(5.186-\mu_1)^2 + 2(-.0276)(5.186-\mu_1)(16.07-\mu_2) + (16.07-\mu_2)^2(.0169) = .827$

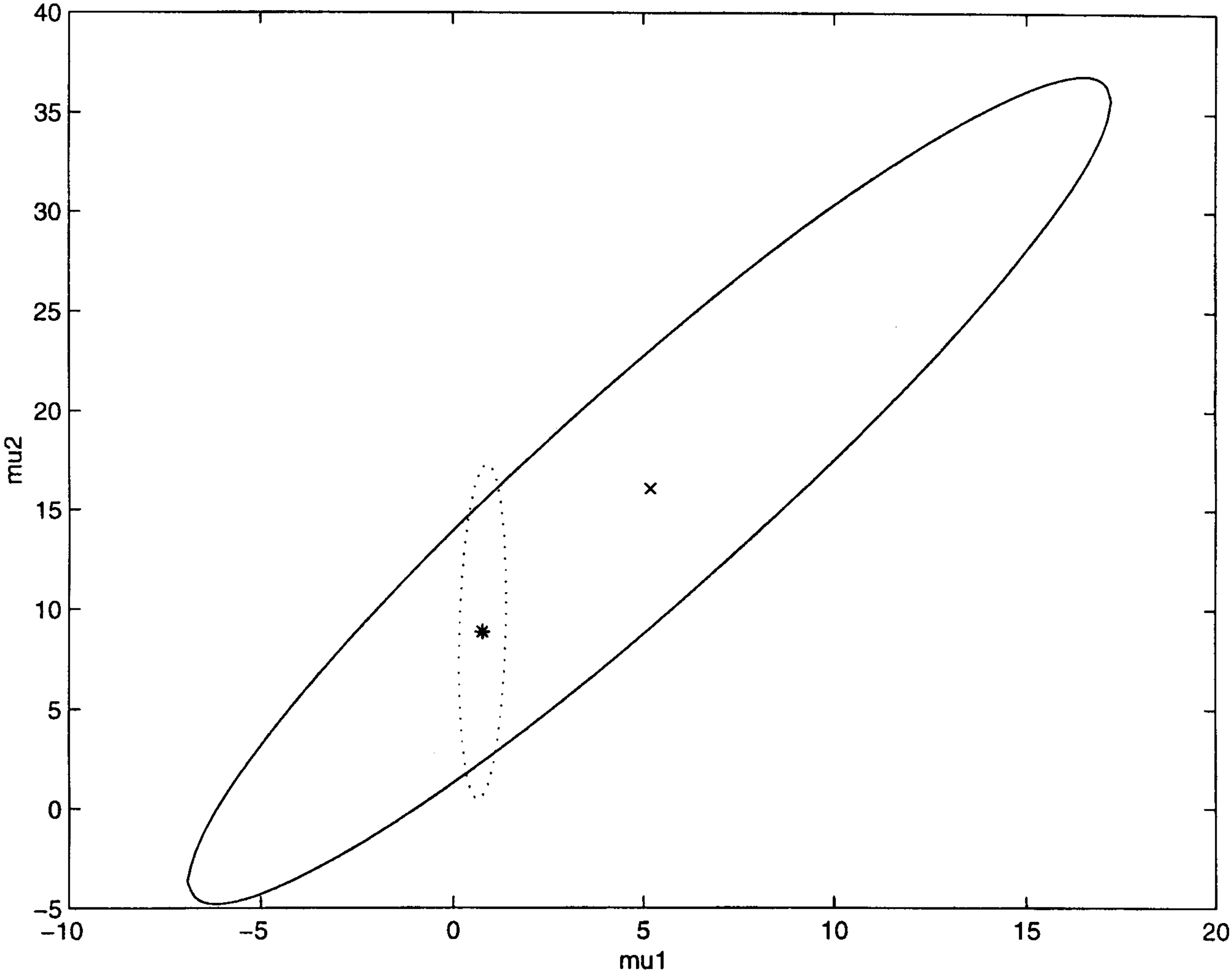
If we solve this equation for μ_2 we get

$$\mu_2 = 7.601 + 1.633\mu_1 \pm 59.172\sqrt{.0114 + .00100\mu_1 - .0000968\mu_1^2}$$

Plugging in values of μ_1 from -6.853 to 17.184 (those values of μ_1 that make the quantity under the $\sqrt{\ }$-sign above positive) we get values of μ_2 that map out the ellipse centered at $(5.186, 16.07)$ and drawn with a solid line on the following page.

(5)

90% Confidence ellipses with (solid) and without (dotted) point #2



More generally, for $\underline{\bar{X}} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix}$, $S^{-1} = \begin{pmatrix} S^{11} & S^{12} \\ S^{12} & S^{22} \end{pmatrix}$ the equation of the ellipse given by $(\underline{\bar{x}} - \underline{\mu})' S^{-1} (\underline{\bar{x}} - \underline{\mu}) = c$ solves as

$$(*) \quad \mu_2 = \frac{S^{12}(\bar{X}_1 - \mu_1) + S^{22}\bar{X}_2 \pm \sqrt{(\bar{X}_1 - \mu_1)^2\left((S^{12})^2 - S^{11}S^{22}\right) + S^{22}c}}{S^{22}}$$

For part (d), we remove point # 2 and recompute $\bar{x}$, S^{-1}, etc, based on a sample size of $n = 8$. The new values are $\underline{\bar{X}} = \begin{pmatrix} .768 \\ 8.869 \end{pmatrix}$, $S^{-1} = \begin{pmatrix} 2.752 & -.0406 \\ -.0406 & .0149 \end{pmatrix}$ (see p. 8 of ex5-9.lst). Using $(*)$ and these new values of $\underline{\bar{X}}$, $\underline{S}^{-1}$ and n, we obtain the 90% confidence region:

$$\left\{ \underline{\mu} : 2.752(.768 - \mu_1)^2 + 2(-.0406)(.768 - \mu_1)(8.869 - \mu_2) + (8.869 - \mu_2)^2 (.0149) \le \underbrace{\frac{8.081}{8}}_{1.010} \right\}$$

Boundary ellipse given by

$$\mu_2 = 6.776 + 2.725\mu_1 \pm 67.114\sqrt{-.00816 + .0605\mu_1 - .0394\mu_1^2}$$

Plugging in values of μ_1 between .149 and 1.386 we obtain the 90% confidence ellipse centered at $* = (.768, 8.869)$ and drawn with a dotted line on page 5.

5.9 b) See ex5-9.sas and ex5-9.lst.
The max-t intervals are (see p.3 & p.8 of ex5-9.lst):

	Including point #2	Omitting point #2
μ_1:	$(-6.881, 17.252)$	$(.149, 1.386)$
μ_2:	$(-4.827, 36.967)$	$(.468, 17.269)$

Based on all of the data $\mu_2 = 10$ is consistent with the data ($\mu_2=10$ lies within the max-t interval).
$\underline{\mu} = (.30, .10)'$ is within the confidence ellipse and gives a non-significant value of T^2 ($T^2 = 10.964$, $p = .0488$).

Same is true based on all of the data except point #2 ($T^2 = 11.323$, $p = .0558$) if we use $\alpha = .10$ as we have to form confidence region, etc.

c) These data have an obvious outlier (point #2) that makes the Q-Q plots (pp.4-5 of ex5-9.lst) nonlinear and makes the scatterplot non-elliptical (p. 1 of ex5-9.lst). After removing this point, the scatterplot (p. 6) and the Q-Q plots (pp. 9-10) look better, but the bivariate normality assumption still looks suspect. It is ~~probably~~ appropriate to consider transforming one or both of the original variables and re-running the analysis.

d) Removing point #2 has a drastic effect on the size & shape of the confidence sets.

5.19 The Bonferroni & max-t intervals are given on p 2 of ex 5-19.lst The program to generate these intervals is ex 5-19.sas Notice that the Bonferroni intervals are narrower.

6.3 See ex6-3.sas and ex6-3.lst. Omitting sample 8 from our paired T^2 analysis of the effluent data yields $T^2 = 11.447$, $p = .0375$ and 95% Bonferroni intervals, $(-21.918, -2.082)$ for μ_1 and $(-3.356, 20.556)$ for μ_2.

The 95% joint confidence ellipse for $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ is given by

$$\left\{ \underline{\mu} : \begin{pmatrix} -12 - \mu_1 \\ 8.6 - \mu_2 \end{pmatrix}' \begin{pmatrix} 136.444 & -52.444 \\ -52.444 & 198.267 \end{pmatrix}^{-1} \begin{pmatrix} -12 - \mu_1 \\ 8.6 - \mu_2 \end{pmatrix} \le \frac{10.033}{10} \right\}$$

(oops. I used $\underline{\mu}$. To be consistent with the book I should have ~~used~~ used $\underline{\delta} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$ instead. It doesn't matter Either way, we're talking about the population mean difference vector.) ~~The~~

The confidence set results are consistent with the hypothesis tests. ~~[illegible]~~ In terms of conclusions, the only thing that has changed from the analysis of the complete data set* is that 0 is no longer in the Bonferroni interval for μ_1

Extra Problem: See hwk3extra.sas and .lst.

~~The test statistic~~ For these data,

$$R = \begin{pmatrix} 1.0 & .885 & .905 & .883 \\ & 1.0 & .826 & .769 \\ & \text{Symmetric} & 1.0 & .923 \\ & & & 1.0 \end{pmatrix}$$

and $|R| = .00507$. So, our test statistic is

$$-n \log |R| = -28 \log(.00507) = 147.959$$ which we compare to the $\chi^2\left(\frac{p(p-1)}{2}\right) = \chi^2(6)$ distribution.

This gives a p-value of nearly 0, and we reject $H_0: P = I$ at any reasonable significance level. The four variables are not independent.