Exam I 206641

1. Given

$$S=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} \text{ and } S=\begin{bmatrix}1&-\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&1&-\frac{1}{2}\\-\frac{1}{2}&-\frac{1}{2}&1\end{bmatrix}$$

a) Calculate the total sample variance for each S. Compute the result.

Total sample variance of $S=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$ = $s_{11}+s_{22}+s_{33}$ = 1 + 1 + 1 = 3

Total sample variance of $S=\begin{bmatrix}1&-\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&1&-\frac{1}{2}\\-\frac{1}{2}&-\frac{1}{2}&1\end{bmatrix}$ = $s_{11}+s_{22}+s_{33}$ = 1 + 1 + 1 = 3

Total sample variance for each S is same values = 3 because the total sample variance criterion pays no attention to the orientation (correlation structure) of the residual vectors.

b) Calculate the generalized sample variance for each S, compare the result.

Generalized sample variance of $S=\begin{vmatrix}1&0&0\\0&1&0\\0&0&1\end{vmatrix}$ = 1

Generalized sample variance of $S=\begin{vmatrix}1&-\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&1&-\frac{1}{2}\\-\frac{1}{2}&-\frac{1}{2}&1\end{vmatrix}$ = 0

Generalized sample variance สำหรับแต่ละค่าของ S ได้ 1 และ 0 ตามลำดับ ทั้งนี้เนื่องจากการคำนวณ Generalized sample variance ได้พิจารณาเกี่ยวกับ ความสัมพันธ์ของแต่ละตัวแปรด้วย ซึ่ง

ในกรณีนี้ $S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ไม่มีค่า Covariance ระหว่างตัวแปรทั้ง 3 ดังนั้นทำให้ เกิด large volume ซึ่งเป็นไปในทางเดียวกันกับ large generalized variance ซึ่งในกรณีนี้มีค่าเป็น 1

ส่วนในกรณี $S = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}$ เราจะเห็นว่ามีค่า Covariance ระหว่างตัวแปรทั้ง 3 แสดงว่าตัวแปรทั้งสามเกิดความสัมพันธ์กัน แต่ในกรณีนี้ยังได้ค่า Generalized sample variance = 0 แสดงว่าไม่เกิด volume of deviation vectors (Geometry: At least one deviation vector lies in the (hyper) plane formed by all linear combination of the other and the three dimension volume is Zero) แสดงว่าเกิด Extreme degeneracy ขึ้นแน่ๆ ซึ่งต้องมีอย่างต่ำที่สุด 1 ตัวแปร(column of the matrix) ที่ควรนำออกจากการพิจารณา

Comment on the discrepancies, if any, found between Part a and b.

ใน Part a และ b ผลที่ได้แตกต่างกันทั้งนี้เนื่องจากใน Part a การหา Total sample variance เป็นเพียงการคำนวณอย่างง่ายๆ และมีความสะดวกโดยการรวมกันของ Variance แต่ละตัวแปรทำให้ทราบผลรวมของ sum of the squared lengths of the p deviation vectors $d_1 = (y_1 - \bar{x}_1 1), ..., d_p = (y_p - \bar{x}_p 1)$, Divided by $n-1$ แต่มันไม่ได้พิจารณาในส่วน Covariance ซึ่งเป็นตัวบ่งบอกความสัมพันธ์ของแต่ละตัวแปร ใน Part b การหา Generalized sample variance นั้นพิจารณาถึงขนาดความสัมพันธ์กันของแต่ละตัวแปรรวมไปถึงสามารถพิจาณาการเกิด Extreme degeneracy ด้วย ทำให้สามารถนำไปลดขนาดของตัวแปร (column of the matrix) ที่ไม่จำเป็นออกไปได้

2. Let X have covariance matrix

$$\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$$

a) **Determine ρ and $\mathrm{V}^{1/2}$**

ใช้สูตร $\rho_{ik} = \dfrac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$ และ $\rho_{11} = 1, \rho_{22} = 1, \rho_{33} = 1$

$$\rho_{12} = \frac{-2}{\sqrt{25}\sqrt{4}}, \qquad \rho_{13} = \frac{4}{\sqrt{25}\sqrt{9}}, \qquad \rho_{23} = \frac{1}{\sqrt{4}\sqrt{9}}$$

$$\rho = \begin{bmatrix} 1 & -0.2 & 0.267 \\ -0.2 & 1 & 0.167 \\ 0.267 & 0.167 & 1 \end{bmatrix}$$

$$\mathrm{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 \\ 0 & 0 & \sqrt{\sigma_{33}} \end{bmatrix}$$

$$\mathrm{V}^{1/2} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

b) **Multiply your matrices to check the relation $\mathrm{V}^{1/2}\rho\mathrm{V}^{1/2} = \Sigma$.**

$$\mathrm{V}^{1/2}\rho\mathrm{V}^{1/2} = \Sigma$$

Check :

$$\mathrm{V}^{1/2}\rho\mathrm{V}^{1/2} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & -0.2 & 0.267 \\ -0.2 & 1 & 0.167 \\ 0.267 & 0.167 & 1 \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\mathrm{V}^{1/2}\rho\mathrm{V}^{1/2} = \begin{bmatrix} 5 & -1 & 1.33 \\ -0.4 & 2 & 0.33 \\ 0.8 & 0.5 & 3 \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\mathrm{V}^{1/2}\rho\mathrm{V}^{1/2} = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix} = \Sigma$$

3. Exercise 6.6 p. 333

6.6. Use the data for treatments 2 and 3 in Exercise 6.8.

Treatment 2: $\begin{bmatrix}3\\3\end{bmatrix},\begin{bmatrix}1\\6\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}$

Treatment 3: $\begin{bmatrix}2\\3\end{bmatrix},\begin{bmatrix}5\\1\end{bmatrix},\begin{bmatrix}3\\1\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}$

a) Calculate S_{pooled}

$$\bar{x}_2 = \begin{bmatrix}2\\4\end{bmatrix} \qquad S_2 = \begin{bmatrix}1 & -1.5\\-1.5 & 3\end{bmatrix}$$

$$\bar{x}_3 = \begin{bmatrix}3\\2\end{bmatrix} \qquad S_3 = \begin{bmatrix}2 & -1.33\\-1.33 & 1.33\end{bmatrix}$$

$$S_{pooled} = \frac{n_2-1}{n_2+n_3-2}S_2 + \frac{n_3-1}{n_2+n_3-2}S_3$$

$$S_{pooled} = \frac{3-1}{3+4-2}\begin{bmatrix}1 & -1.5\\-1.5 & 3\end{bmatrix} + \frac{4-1}{3+4-2}\begin{bmatrix}2 & -1.33\\-1.33 & 1.33\end{bmatrix}$$

$$S_{pooled} = \begin{bmatrix}1.6 & -1.398\\-1.398 & 1.998\end{bmatrix}$$

b) Teat $H_o : \mu_2 - \mu_3 = 0$ employing a two-sample approach with $\alpha = 0.01$

1. $H_o : \mu_2 - \mu_3 = 0$
   $H_1 : \mu_2 - \mu_3 \neq 0$

2. $\alpha = 0.01$

3. $$T^2 = (\bar{x}_2 - \bar{x}_3 - 0)'\left[\left(\frac{1}{n_2}+\frac{1}{n_3}\right)S_{pooled}\right]^{-1}(\bar{x}_2 - \bar{x}_3 - 0)$$

$$= \left(\begin{bmatrix}2\\4\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix}\right)'\left[\left(\frac{1}{3}+\frac{1}{4}\right)\begin{bmatrix}1.6 & -1.398\\-1.398 & 1.998\end{bmatrix}\right]^{-1}\left(\begin{bmatrix}2\\4\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix}\right)$$

$$= \left(\begin{bmatrix}-1\\2\end{bmatrix}\right)'\left[\begin{bmatrix}0.933 & -.816\\-.816 & 1.166\end{bmatrix}\right]^{-1}\left(\begin{bmatrix}-1\\2\end{bmatrix}\right)$$

$$= \begin{bmatrix}-1 & 2\end{bmatrix}\begin{bmatrix}2.757 & 1.929\\1.929 & 2.208\end{bmatrix}\begin{bmatrix}-1\\2\end{bmatrix}$$

$$T^2 = 3.872$$

4. $c^2 = \frac{5(2)}{4} F_{2,4}(.01) = \frac{10}{4}(18) = 45$

5. $T^2 = 3.872 < 45$

พวกเรา Not Reject $H_o : \mu_2 - \mu_3 = 0$ ที่ระดับนัยสำคัญ 0.01 แสดงว่ามีค่าเฉลี่ยของ Responses ทั้งสอง ไม่แตกต่างกันในแต่ละ Treatment หรือ Treatment ไม่มีอิทธิพลต่อค่า Responses

c) Construct 99% simultaneous confidence intervals for the differences

$\mu_{2i} - \mu_{3i}, i = 1,2$

เนื่องจากสนใจเฉพาะ $\mu_{21} - \mu_{31}$ และ $\mu_{22} - \mu_{32}$ ดังนั้นจะใช้ Bonferroni 99% simultaneous confidence intervals เพื่อ 2 population mean differences ดังนี้

$$\mu_{1i} - \mu_{2i} : (\overline{x}_{2i} - \overline{x}_{3i}) \pm c\sqrt{\left(\frac{1}{n_2} + \frac{1}{n_3}\right) s_{ii,pooled}} \quad , i = 1,2 \qquad \text{โดย } c = t_{n_2+n_3-2,(\alpha/2p)}$$

$$\mu_{1i} - \mu_{2i} : (\overline{x}_{2i} - \overline{x}_{3i}) \pm t_{n_2+n_3-2,(\alpha/2p)}\sqrt{\left(\frac{1}{n_2} + \frac{1}{n_3}\right) s_{ii,pooled}} \quad , i = 1,2$$

หา $\mu_{21} - \mu_{31} : (2-3) \pm 4.773\sqrt{\left(\frac{1}{3} + \frac{1}{4}\right)1.6}$

$$-5.611 \le \mu_{21} - \mu_{31} \le 3.611$$

หา $\mu_{22} - \mu_{32} : (4-2) \pm 4.773\sqrt{\left(\frac{1}{3} + \frac{1}{4}\right)1.998}$

$$-3.15285 \le \mu_{22} - \mu_{32} \le 7.152851$$

Note: ในกรณีสนใจ **<u>All</u>** possible linear combinations of the differences in the mean vectors. ให้ใช้ตามสูตรนี้ $\mu_{1i} - \mu_{2i} : (\overline{x}_{2i} - \overline{x}_{3i}) \pm c\sqrt{\left(\frac{1}{n_2} + \frac{1}{n_3}\right) s_{ii,pooled}} \quad , i = 1,2$

โดย $c = \sqrt{[(n_2 + n_3 - 2)p/(n_2 + n_3 - p - 1)]F_{p,n_2+n_3-p-1}(\alpha)}$

ซึ่งสามารถนำไปหาเฉพาะ $\mu_{21} - \mu_{31}$ และ $\mu_{22} - \mu_{32}$ ได้ แต่ได้ช่วงความเชื่อมั่นที่กว้างมากกว่าแบบ Bonfferroni และนำไปสรุปผลกับสมมติฐานที่ได้ทดสอบแล้วใน Part b สู้การหาช่วงความเชื่อมั่นแบบ Bonfferroni ไม่ได้ ในกรณีนี้ถ้าลองคำนวณจะได้ผลลัพธ์ดังนี้

$-7.48074 \le \mu_{21} - \mu_{31} \le 5.480741$ และ $-5.24206 \le \mu_{22} - \mu_{32} \le 9.242065$

4. Exercise 6.7 p. 333

6.7. Using the summary statistics for the electricity-demand data given in Example 6.4, compute T^2 and test the hypothesis $H_o: \mu_1 - \mu_2 = 0$, assuming that $\Sigma_1 = \Sigma_2$. Set $\alpha = .05$. Also, determine the linear combination of mean components most responsible for the rejection of H_o.

1. $H_o: \mu_1 - \mu_2 = 0$

   $H_o: \mu_1 - \mu_2 \neq 0$

2. $\alpha = .05$

3. $T^2 = (\bar{x}_1 - \bar{x}_2 - 0)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_{pooled} \right]^{-1} (\bar{x}_1 - \bar{x}_2 - 0)$

$$= \left(\begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix} - \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix} \right)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \begin{bmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{bmatrix} \right]^{-1} \left(\begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix} - \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix} \right)$$

$T^2 = 16.066$

4. $c^2 = \frac{98(2)}{97} F_{2,97}(.05) = 6.26$

5. $T^2 = 16.066 > 6.26$

   พวกเรา Reject $H_o: \mu_1 - \mu_2 = 0$ ที่ระดับนัยสำคัญ 0.05 แสดงว่ามีค่าเฉลี่ยอย่างน้อย 1 คู่แตกต่างกัน ดังนั้นจะใช้ Bonferroni 95% simultaneous confidence intervals เพื่อ 2 population mean differences เป็น

$$\mu_{1i} - \mu_{2i} : (\bar{x}_{1i} - \bar{x}_{2i}) \pm t_{n_1+n_2-2,(\alpha/2p)} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) s_{ii,pooled}}$$

หา $\mu_{11} - \mu_{21} : (204.4 - 130.0) \pm 2.276362 \sqrt{\left(\frac{1}{45} + \frac{1}{55} \right) 10963.7}$

$$26.489 \le \mu_{11} - \mu_{21} \le 122.311$$

หา $\mu_{12}-\mu_{22} : (556.6-355.0) \pm 2.276362\sqrt{\left(\frac{1}{45}+\frac{1}{55}\right)63661.3}$

$$86.151 \le \mu_{12}-\mu_{22} \le 317.049$$

We conclude that there is a difference in electrical consumption between those with air-conditioning and those without. This difference is evident in both onpeak and off-peak consumption

5. Find the maximum likelihood estimates of the 2 x 1 mean vector μ , and the 2 x 2 covariance matrix Σ based on the random sample

$$X = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

จาก Result 4.11. $\hat{\mu} = \overline{X}$ and $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)\left(X_j - \overline{X}\right)' = \frac{(n-1)}{n}S$

เป็น the maximum likelihood estimators of μ และ Σ โดยถ้าพิจารณาจากค่าที่ได้สำรวจมา $\overline{x}$ และ $\frac{1}{n}\sum_{j=1}^{n}\left(x_j - \overline{x}\right)\left(x_j - \overline{x}\right)'$ จะเรียกว่า the maximum likelihood estimates of μ และ Σ

$$\overline{x} = \begin{bmatrix} \dfrac{3+4+5+4}{4} \\ \dfrac{6+4+7+7}{4} \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

$$\frac{(n-1)}{n}S = \frac{3}{4}\begin{bmatrix} 0.667 & 0.333 \\ 0.333 & 2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 1.5 \end{bmatrix}$$

ดังนั้น the maximum likelihood estimates of μ และ Σ คือ $\hat{\mu} = \overline{x} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ และ

$\hat{\Sigma} = \frac{(n-1)}{n}S = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 1.5 \end{bmatrix}$ ตามลำดับ

6. Exercise 3.15 p. 148

Repeat Exercise 3.14 using the data matrix

$$X = \begin{bmatrix} 1 & 4 & 3 \\ 6 & 2 & 6 \\ 8 & 3 & 3 \end{bmatrix}$$

And the linear combinations

$$b'X = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

And

$$c'bX = \begin{bmatrix} 1 & 2 & -3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

We have $n = 3$ **observations on** $p = 3$ **variables** X_1, X_2 and X_3

a) **Evaluate the sample means, variances, and covariance of $b'X$ and $c'X$ from first principles. That is, calculate the observed values of $b'X$ and $c'X$, and then use the sample mean, variance, and covariance formulas.**

$b'x_1 = 1 + 4 + 3 = 8$

$b'x_2 = 6 + 2 + 6 = 14$

$b'x_3 = 8 + 3 + 3 = 14$

Sample mean = $\dfrac{8+14+14}{3} = 12$

Sample variance = $\dfrac{(8-12)^2 + (14-12)^2 + (14-12)^2}{3-1} = 12$

In a similar manner, $c'X$

$c'x_1 = 1(1) + 4(2) + 3(-3) = 0$

$c'x_2 = 6(1) + 2(2) + 6(-3) = -8$

$c'x_3 = 8(1) + 3(2) + 3(-3) = 5$

Sample mean = $\dfrac{0-8+5}{3} = -1$

Sample variance = $\dfrac{(0+1)^2 + (-8+1)^2 + (5+1)^2}{3-1} = 43$

Sample covariance

$= \dfrac{(8-12)(0+1) + (14-12)(-8+1) + (14-12)(5+1)}{3-1} = -3$

b) Calculate the sample means, variances, and covariance of $b'X$ and $c'X$ using (3-36). Compare the results in (a) and (b)

สมการ (3-36) sample mean of $b'X = b'\overline{x}$

sample mean of $c'X = c'\overline{x}$

$$\overline{x} = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}$$

$$b'\overline{x} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = 5 + 3 + 4 = 12$$ (check $b'X = 12$)

$$c'\overline{x} \begin{bmatrix} 1 & 2 & -3 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = 5 + 6 - 12 = -1$$ (check $c'X = -1$)

$$\left.\begin{array}{l} \therefore \text{ sample mean of } b'X = b'\overline{x} \\ \text{sample mean of } c'X = c'\overline{x} \end{array}\right\}$$ นั้นคือผลการคำนวณ a) และ b) มีค่า sample mean เท่ากัน

สมการ (3-36) sample variance of $b'X = b'Sb$

sample variance of $c'X = c'Sc$

$$S = \begin{bmatrix} 13 & -2.5 & 1.5 \\ -2.5 & 1 & -1.5 \\ 1.5 & -1.5 & 3 \end{bmatrix}$$

$$b'Sb = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 13 & -2.5 & 1.5 \\ -2.5 & 1 & -1.5 \\ 1.5 & -1.5 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 12 & -3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 12$$ (check $b'X = 12$)

$$c'Sc = \begin{bmatrix} 1 & 2 & -3 \end{bmatrix} \begin{bmatrix} 13 & -2.5 & 1.5 \\ -2.5 & 1 & -1.5 \\ 1.5 & -1.5 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} 3.5 & 4 & -10.5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = 43 \qquad (\text{check } c'X = 43)$$

$\therefore$ sample variance of $b'X = b'Sb$ / sample variance of $c'X = c'Sc$ } นั้นคือผลการคำนวณ a) และ b) มีค่า sample variance เท่ากัน

สมการ (3-36) sample covariance of $b'X$ and $c'X = b'Sc$

$$b'Sc = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 13 & -2.5 & 1.5 \\ -2.5 & 1 & -1.5 \\ 1.5 & -1.5 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} 12 & -3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = -3 \qquad (\text{check } b'X \text{ and } c'X = -3)$$

$\therefore$ sample covariance of $b'X$ and $c'X = b'Sc$ } นั้นคือผลการคำนวณ a) และ b) มีค่า sample covariance เท่ากัน

7. Use the sample covariance obtained in Example 3.7 to verify equation (3-29) and (3-30), which state that $R = D^{-1/2}SD^{-1/2}$ and $D^{-1/2}RD^{-1/2} = S$

$$S = \begin{bmatrix} 252.04 & -68.43 \\ -68.43 & 123.67 \end{bmatrix}$$

พิสูจน์ $R = D^{-1/2}SD^{-1/2}$

$$D^{1/2} = \begin{bmatrix} \sqrt{252.04} & 0 \\ 0 & \sqrt{123.67} \end{bmatrix}$$

$$D^{-1/2} = \begin{bmatrix} 1/\sqrt{252.04} & 0 \\ 0 & 1/\sqrt{123.67} \end{bmatrix}$$

$$D^{-1/2}SD^{-1/2} = \begin{bmatrix} 1/\sqrt{252.04} & 0 \\ 0 & 1/\sqrt{123.67} \end{bmatrix}\begin{bmatrix} 252.04 & -68.43 \\ -68.43 & 123.67 \end{bmatrix}\begin{bmatrix} 1/\sqrt{252.04} & 0 \\ 0 & 1/\sqrt{123.67} \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & -0.3876 \\ -0.3876 & 1 \end{bmatrix}$$

พิสูจน์ $S = D^{-1/2}RD^{-1/2}$

$$D^{1/2}RD^{1/2} = \begin{bmatrix} \sqrt{252.04} & 0 \\ 0 & \sqrt{123.67} \end{bmatrix}\begin{bmatrix} 1 & -0.3876 \\ -0.3876 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{252.04} & 0 \\ 0 & \sqrt{123.67} \end{bmatrix}$$

$S = \begin{bmatrix} 252.04 & -68.43 \\ -68.43 & 123.67 \end{bmatrix} = S$ จาก Example 3.7

8. Exercise 5.20 p. 268

5.20. A wildlife ecologist measured x_1 = tail length (in millimeters) and x_2 = wing length (in millimeters) for a sample of $n = 45$ female hook-billed kites. These data are displayed in Table 5.12. Using the data in the table,

ชุดข้อมูล	x1	x2	ชุดข้อมูล	x1	x2	ชุดข้อมูล	x1	x2
1	191	284	16	186	266	31	173	271
2	197	285	17	197	285	32	194	280
3	208	288	18	201	295	33	198	300
4	180	273	19	190	282	34	180	272
5	180	275	20	209	305	35	190	292
6	188	280	21	187	285	36	191	286
7	210	283	22	207	297	37	196	285
8	196	288	23	178	268	38	207	286
9	191	271	24	202	271	39	209	303
10	179	257	25	205	285	40	179	261
11	208	289	26	190	280	41	186	262
12	202	285	27	189	277	42	174	245
13	200	272	28	211	310	43	181	250
14	192	282	29	216	305	44	189	262
15	199	280	30	189	274	45	188	258

a) **Find and sketch the 95% confidence ellipse for the population means μ_1 and μ_2. Suppose it is known that $\mu_1 = 190$ mm and $\mu_1 = 275$ mm for male hook-billed kites. Are these plausible values for the mean tail length and mean wing length for the female birds? Explain.**

$\overline{x} = \begin{bmatrix} 193.62 \\ 279.78 \end{bmatrix}$

$S = \begin{bmatrix} 120.69 & 122.35 \\ 122.35 & 208.54 \end{bmatrix}$

$\lambda_i = \begin{bmatrix} 294.61 \\ 34.626 \end{bmatrix}$

$\lambda_1 = 294.61 \qquad e_1 = \begin{bmatrix} 0.5754 \\ 0.8179 \end{bmatrix}$

$\lambda_2 = 34.626 \qquad e_2 = \begin{bmatrix} -0.818 \\ 0.5754 \end{bmatrix}$

$\pm\sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} =$ 17.164 x 0.3823 = 6.5626

$\pm\sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}e_1 = \begin{bmatrix} 3.7761 \\ 5.3676 \end{bmatrix}$

$\pm\sqrt{\lambda_2}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} =$ 5.8844 x 0.3823 = 2.2499

$\pm\sqrt{\lambda_2}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}e_2 = \begin{bmatrix} -1.84 \\ 1.2946 \end{bmatrix}$

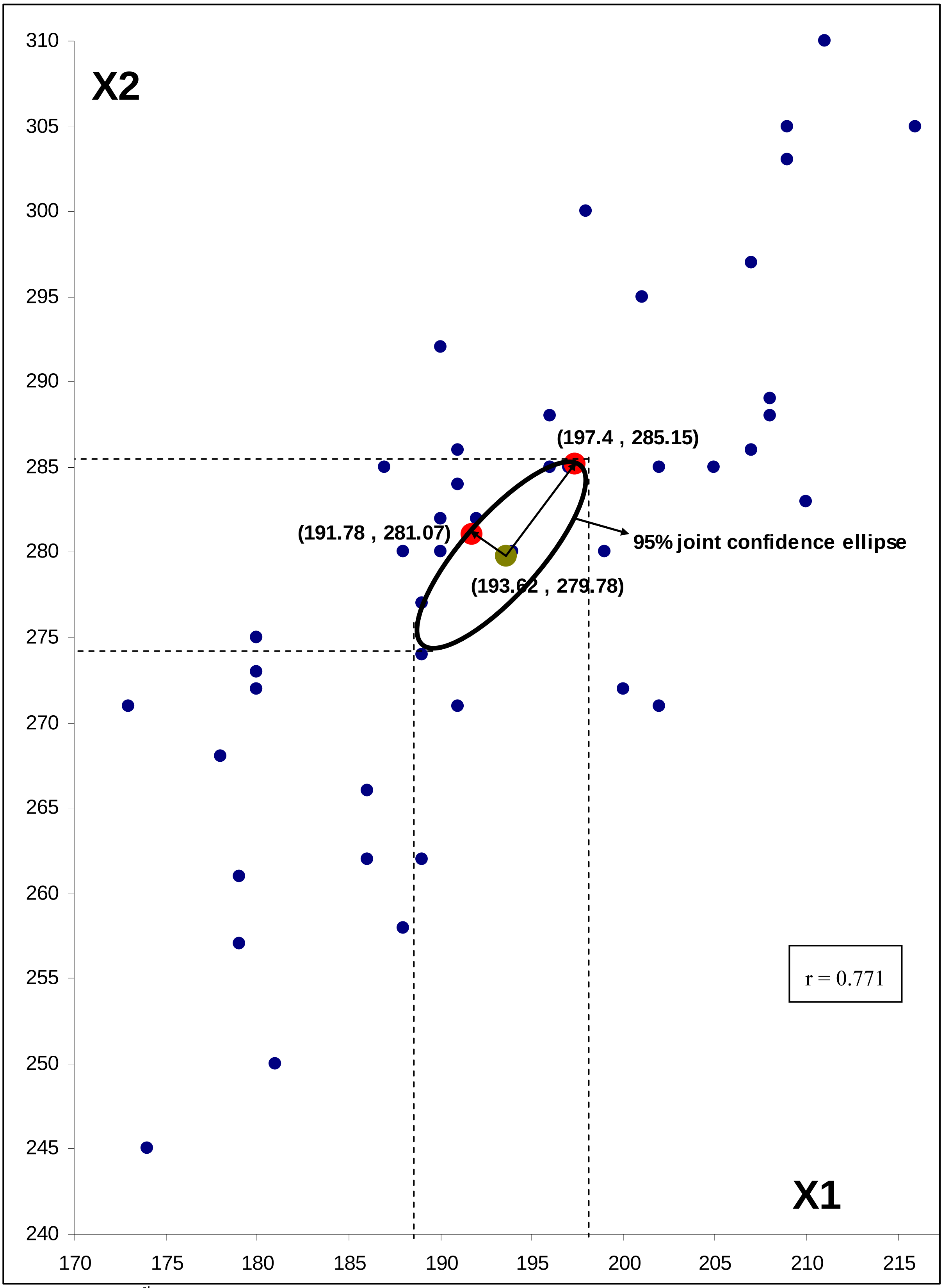
Note: เนื่องจากตัวอย่างมีขนาดใหญ่ สามารถใช้ Chi-Square คำนวณได้ซึ่งจะได้ค่าผลลัพธ์ใกล้เคียงกันดังนี้

$\pm\sqrt{\lambda_1}\sqrt{\frac{\chi_p^2}{n}(\alpha)} =$ 17.164 x 0.3648 = 6.2622

$\pm\sqrt{\lambda_1}\sqrt{\frac{\chi_p^2}{n}(\alpha)}e_1 = \begin{bmatrix} 3.6033 \\ 5.1219 \end{bmatrix}$

$\pm\sqrt{\lambda_2}\sqrt{\frac{\chi_p^2}{n}(\alpha)} =$ 5.8844 x 0.3648 = 2.1469

$\pm\sqrt{\lambda_2}\sqrt{\frac{\chi_p^2}{n}(\alpha)}e_2 = \begin{bmatrix} -1.756 \\ 1.2353 \end{bmatrix}$



Note: จากภาพนี้ Ellipsoid ของรูปมีขนาดเล็ก ซึ่งอาจเป็นไปได้ว่าข้อมูลไม่มีการแจกแจงแบบ bivariate normal โดยต้องทำการทดสอบ หรือ เกิด Outlier จำนวนมากกับข้อมูลชุดนี้

1. $H_0 : \mu' = [190, 275]$
$H_1 : \mu' \neq [190, 275]$

2. $\alpha = 0.05$

3. $T^2 = n(\bar{x} - \mu_o)' S^{-1} (\bar{x} - \mu_o)$

$$n(\bar{x} - \mu_o)' S^{-1} (\bar{x} - \mu_o) = 45 \begin{pmatrix} 3.6222 & 4.7778 \end{pmatrix} \begin{pmatrix} 0.0204 & -0.012 \\ -0.012 & 0.0118 \end{pmatrix} \begin{pmatrix} 3.6222 \\ 4.7778 \end{pmatrix}$$

$$45 \begin{pmatrix} 0.0167 & 0.0131 \end{pmatrix} \begin{pmatrix} 3.6222 \\ 4.7778 \end{pmatrix}$$

$$45 \quad x \quad 0.1232$$

$$n(\bar{x} - \mu_o)' S^{-1} (\bar{x} - \mu_o) = 5.5431$$

$T^2 = 5.5431$

4. $$\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) = \frac{(45-1)2}{(45-2)} F_{2,45-2}(0.05) = 6.5785$$

Note: ถ้าใช้ Chi-Square $\chi^2_{p(0.05)} = 5.99$

5. $$T^2 = 5.5431 < \frac{(45-1)2}{(45-2)} F_{2,45-2}(0.05) = 6.5785$$

We not reject H_o at the 5% level of significance. So, $\mu' = [190, 275]$ plausible value of mean tail length = 190 and plausible value of mean wing length = 275 for the female birds.

b) Construct the simultaneous 95% T^2 -intervals for μ_1 and μ_2 and for μ_1 and μ_2. Compare the two sets of intervals. What advantage, if any, do the T^2-intervals have over the Bonferroni intervals?

The simultaneous 95% T^2 -intervals

เมื่อ $\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha) = \frac{(45-1)2}{(45-2)}F_{2,45-2}(0.05) = 6.5785$

$$\mu_1 = \overline{x}_1 \pm \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \qquad = 193.62 \pm \sqrt{6.5785}\sqrt{\frac{120.69}{45}}$$

$$189.4217 \le \mu_1 \le 197.8227$$

$$\mu_2 = \overline{x}_2 \pm \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \qquad = 279.78 \pm \sqrt{6.5785}\sqrt{\frac{208.54}{45}}$$

$$274.2564 \le \mu_2 \le 285.2992$$

The 95% Bonferroni intervals

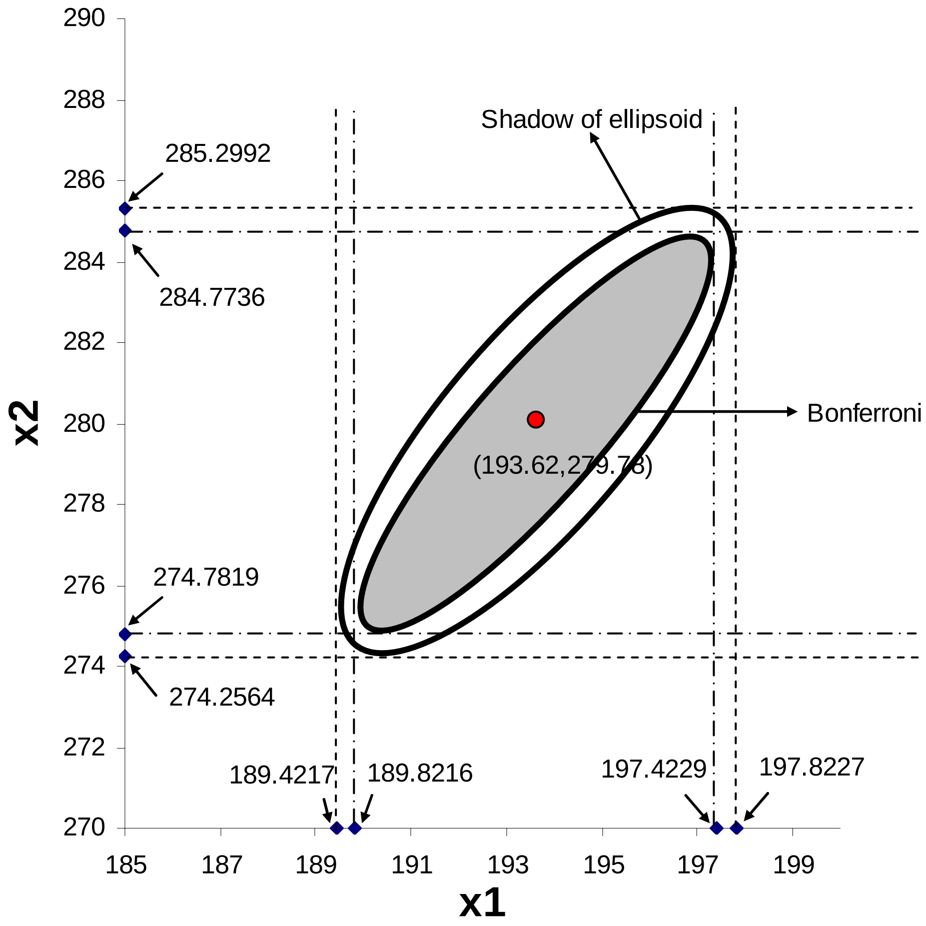
เมื่อ $t_{n-1,(\alpha/2p)} = t_{45-1,(0.05/2x2)} = 2.320711$

$$\mu_1 = \overline{x}_1 \pm t_{n-1,(\alpha/2p)}\sqrt{\frac{s_{11}}{n}} \qquad = 193.62 \pm 2.320711\sqrt{\frac{120.69}{45}}$$

$$189.8216 \le \mu_1 \le 197.4229$$

$$\mu_2 = \overline{x}_2 \pm t_{n-1,(\alpha/2p)}\sqrt{\frac{s_{22}}{n}} \qquad = 279.78 \pm 2.320711\sqrt{\frac{208.54}{45}}$$

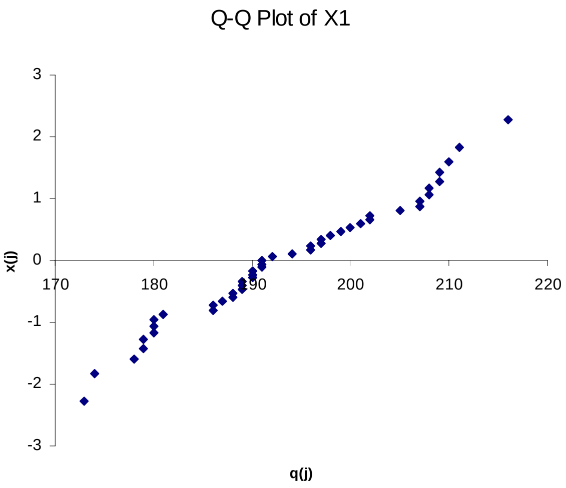
$$274.7819 \le \mu_2 \le 284.7736$$

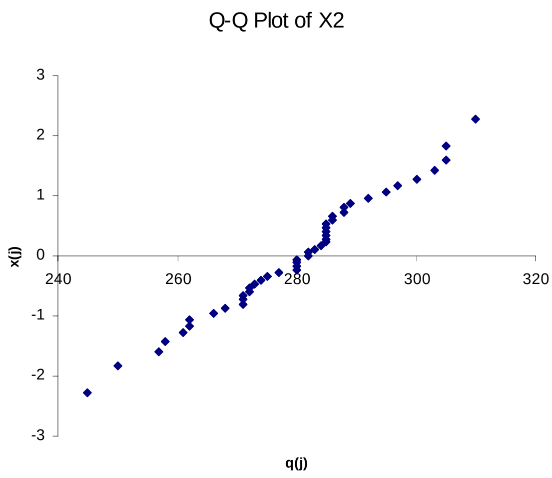


สรุปได้ว่า 95% Bonferroni Simultaneous confidence interval for component means จะมีช่วงความเชื่อมั่นแคบกว่าการใช้ 95% T^2 Simultaneous confidence interval for component means ซึ่งการใช้ T^2 Simultaneous confidence interval ในการหาช่วงความเชื่อมั่นเพื่อการทดสอบสมมติฐานอาจเกิดความผิดพลาดในการทดสอบได้ ทั้งนี้เนื่องจาก T^2 มีช่วงความเชื่อมั่นที่กว้างเกินไป **แต่อย่างไรก็ตามทั้ง T^2 และ Bonferroni สามารถ Hold ความเชื่อมั่นที่** $(1-\alpha)$

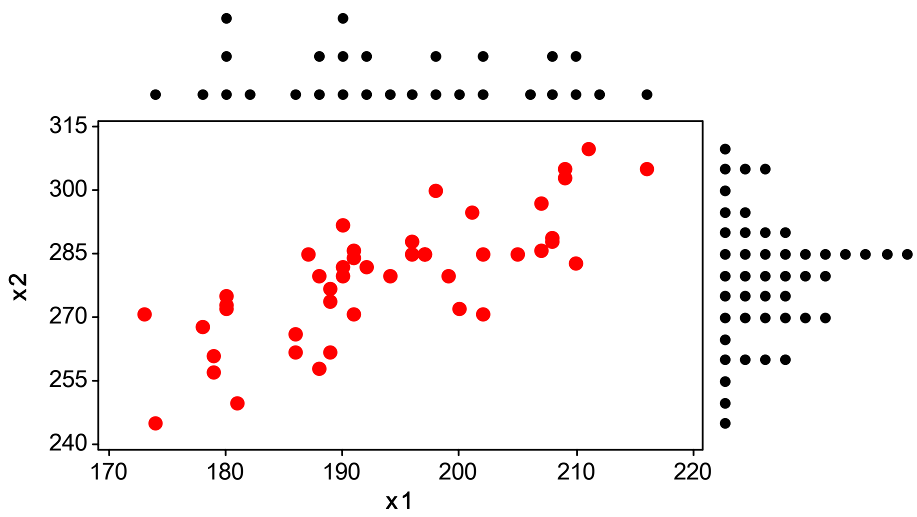
Note: ในกรณีใช้ตัวสถิติ Large Sample Inferences ก็ได้ผลลัพธ์ใกล้เคียงกันกับผลลัพธ์ข้างต้น

c) Is the bivariate normal distribution a viable population model? Explain with reference to Q-Q plots and a scatter diagram.





Scatter and Dot plot Diagram



x_1 = tail length (in millimeters)

$$r_Q = \frac{\sum_{j=1}^{n}(x_{(j)}-\overline{x})(q_{(j)}-\overline{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)}-\overline{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)}-\overline{q})^2}} = 0.9874$$

จากตารางที่ 4.2 ได้ค่า Critical points for the Q-Q plot ที่ $\alpha = 0.05$ n = 45 คือ .9749

$r_Q = 0.9874 > 0.9749$

แสดงว่าเราไม่ปฏิเสธสมมติฐาน นั้นคือ ข้อมูลมีการแจกแจงแบบปกติที่ระดับนัยสำคัญ 0.05

x_2 = wing length (in millimeters)

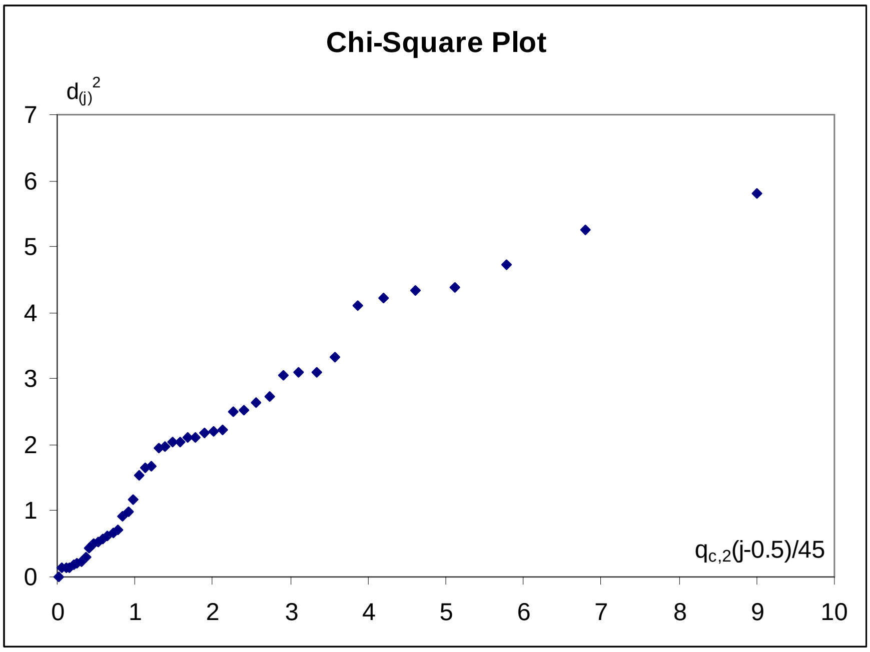
$$r_Q = \frac{\sum_{j=1}^{n}(x_{(j)}-\overline{x})(q_{(j)}-\overline{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)}-\overline{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)}-\overline{q})^2}} = 0.9913$$

จากตารางที่ 4.2 ได้ค่า Critical points for the Q-Q plot ที่ $\alpha = 0.05$ n = 45 คือ .9749

$r_Q = 0.9913 > 0.9749$

แสดงว่าเราไม่ปฏิเสธสมมติฐาน นั้นคือ ข้อมูลมีการแจกแจงแบบปกติที่ระดับนัยสำคัญ 0.05

การตรวจสอบ bivariate normal อยู่ในบทที่ 4 หน้า 183 ได้ผลดังนี้



พบว่าข้อมูลไม่อยู่บน slope 1 จำนวนมาก ดังนั้นไม่น่าจะ Approximately bivariate normal ได้ และข้อมูลอาจประกอบไปด้วยข้อมูลที่เป็น Outliers จำนวนมาก

และเมื่อพิจารณา r_Q = 0.96224 จากตารางที่ 4.2 ได้ค่า Critical points for the Q-Q plot ที่ $\alpha = 0.05$ n = 45 คือ .9749

$r_Q = 0.96224 < 0.9749$

แสดงว่าเราปฏิเสธสมมติฐาน นั้นคือ ข้อมูลไม่มีการแจกแจงแบบปกติที่ระดับนัยสำคัญ 0.05

9. The quantities, S and S^{-1} are given in Example 5.3 p. 221 for the transformed microwave-radiation data. Conduct a test of the null hypothesis $H_0 : \mu' = [0.55, 0.60]$ at the $\alpha = 0.05$ level of significance.

$$\bar{X} = \begin{bmatrix} .564 \\ .603 \end{bmatrix} \quad S = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix} \quad S^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix}$$

1. $H_0 : \mu' = [0.55, 0.60]$
   $H_1 : \mu' \neq [0.55, 0.60]$

2. $\alpha = 0.05$

3. $T^2 = n(\bar{x} - \mu_o)' S^{-1} (\bar{x} - \mu_o)$

$$n(\bar{x} - \mu_o)' S^{-1} (\bar{x} - \mu_o) = 42 \begin{bmatrix} 0.014 & 0.003 \end{bmatrix} \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix} \begin{bmatrix} 0.014 \\ 0.003 \end{bmatrix}$$

$$42 \begin{bmatrix} 2.3521 & -1.687 \end{bmatrix} \begin{bmatrix} 0.014 \\ 0.003 \end{bmatrix}$$

$$42 \quad x \quad 0.0279$$

$$n(\bar{x} - \mu_o)' S^{-1} (\bar{x} - \mu_o) = 1.1705$$

$T^2 = 1.1705$

4. $\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) = \frac{(42-1)2}{(42-2)} F_{2,42-2}(0.05) = 6.62$

5. $T^2 = 1.1705 < \frac{(42-1)2}{(42-2)} F_{2,42-2}(0.05) = 6.62$

We not reject H_o at the 5% level of significance. So, $\mu' = [0.55, 0.60]$ plausible value of mean $x_1 = \sqrt[4]{\text{measured radiation with door closed}} = 0.55$ and plausible value of mean $x_2 = \sqrt[4]{\text{measured radiation with door open}} = 0.60$.

10. Let X be distributed $N_3(\mu,\Sigma)$, where $\mu' = [1,-1,2]$ and

$$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

a) X_1 and X_2

Since X_1 and X_2 have covariance $\sigma_{12} = 0$. They are independent.

and distribution of X_1 and X_2 is $N_2\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}\right)$

b) X_1 and X_3

Since X_1 and X_3 have covariance $\sigma_{13} = -1$. They are not independent.

and distribution of X_1 and X_3 is $N_2\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}\right)$

c) X_2 and X_3

Since X_2 and X_3 have covariance $\sigma_{23} = 0$. They are independent.

and distribution of X_2 and X_3 is $N_2\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}\right)$

d) (X_1, X_3) and X_2

$$\Sigma = \left[\begin{array}{cc|c} 4 & -1 & 0 \\ -1 & 2 & 0 \\ \hline 0 & 0 & 5 \end{array}\right] = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

$$X = \left[\begin{array}{c} x_1 \\ x_3 \\ \hline X_2 \end{array}\right] = \left[\begin{array}{c} X_1 \\ \hline X_2 \end{array}\right] \qquad \mu = \left[\begin{array}{c} \mu_1 \\ \mu_3 \\ \hline \mu_2 \end{array}\right] = \left[\begin{array}{c} 1 \\ 2 \\ \hline -1 \end{array}\right]$$

Since $\underset{(q_1x1)}{X_1}$ and $\underset{(q_2x1)}{X_2}$ have covariance $\Sigma_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, (X_1, X_3) and X_2 are independent. and distribution of $\underset{(q_1x1)}{X_1}$ and $\underset{(q_2x1)}{X_2}$ is $N_{q_1+q_2}\left(\left[\begin{array}{c} 1 \\ 2 \\ \hline -1 \end{array}\right], \left[\begin{array}{cc|c} 4 & -1 & 0 \\ -1 & 2 & 0 \\ \hline 0 & 0 & 5 \end{array}\right]\right)$

e) X_1 **and** $X_1 + 3X_2 - 2X_3$

ทำ linear combination of the components of a normal random vector

$$\begin{bmatrix} X_1 \\ X_1 + 3X_2 - 2X_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix} = A$$

$$A\Sigma A' = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ 0 & -2 \end{bmatrix}$$

$$A\Sigma A' = \begin{bmatrix} 4 & 6 \\ 6 & 61 \end{bmatrix}$$

$$A\mu = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -6 \end{bmatrix}$$

Since $\underset{(q_1 x 1)}{X_1}$ and $\underset{(q_2 x 1)}{X_2} = X_1 + 3X_2 - 2X_3$ have covariance = 6. They are not independent. and distribution of $\underset{(q_1 x 1)}{X_1}$ and $\underset{(q_2 x 1)}{X_2}$ is $N_{q_1+q_2}\left(\begin{bmatrix} 1 \\ -6 \end{bmatrix}, \begin{bmatrix} 4 & 6 \\ 6 & 61 \end{bmatrix} \right)$