MH4501		**Name:**	Singh Jasraj
Assignment 2		**Matric:**	U1940558D
April 29, 2023		**Collaborators:**	N.A.

Problem 1

Consider the linear transformation of the sample,

$$y_1, \ldots, y_n, \quad \text{where } y_i = Cx_i$$

Then, we have a transformed sample of size n from a p-variate normal distribution having mean $\mu_y = C\mu$ and covariance $\Sigma_y = C\Sigma C^T$. We denote the sample mean as $\bar{y} = C\bar{x}$ and the sample covariance $S_y = CSC^T$.

a)

$$\begin{aligned} C\bar{x} &= \bar{y} \\ &\sim \mathcal{N}\left(C\mu, \frac{C\Sigma C^T}{n}\right) \end{aligned}$$

b)

$$\begin{aligned} & n\left(C\bar{x} - C\mu\right)^T \left(C\Sigma C^T\right)^{-1} \left(C\bar{x} - C\mu\right) \\ = & n\left(\bar{y} - \mu_y\right)^T \Sigma_y^{-1} \left(\bar{y} - \mu_y\right) \\ \sim & \chi^2(p) \end{aligned}$$

c)

$$\begin{aligned} & n\left(C\bar{x} - C\mu\right)^T \left(CSC^T\right)^{-1} \left(C\bar{x} - C\mu\right) \\ = & n\left(\bar{y} - \mu_y\right)^T S_y^{-1} \left(\bar{y} - \mu_y\right) \\ \sim & \frac{(n-1)\,p}{n-p} F(p, n-p) \end{aligned}$$

d)

$$\begin{aligned} & n\left(C\bar{x} - C\mu\right)^T \left(CSC^T\right)^{-1} \left(C\bar{x} - C\mu\right) \\ \sim & \frac{(n-1)\,p}{n-p} F(p, n-p) \xrightarrow{n\to\infty} \chi^2(p) \end{aligned}$$

e) Considering significance level α for the test,

   i. Using Problem 1 Part c), under H_0,

$$T^2 \sim \frac{(n-1)\,p}{n-p} F(p, n-p), \quad \text{or,} \quad \frac{n-p}{(n-1)\,p} T^2 \sim F(p, n-p)$$

   We reject H_0 if T^2 is too big, i.e., if

$$\frac{n-p}{(n-1)\,p} T^2 > F_\alpha[p, n-p], \quad \text{or,} \quad T^2 > \frac{(n-1)\,p}{n-p} F_\alpha[p, n-p]$$

ii. Using Problem 1 Part d), under H_0,

$$T^2 \sim \chi^2(p)$$

We reject H_0 if T^2 is too big, i.e., if

$$T^2 > \chi^2_\alpha[p]$$

Problem 2

a)

$$\begin{aligned}
n &= 87\\
p &= 3\\
\bar{x} &= \begin{pmatrix}526.586206 & 54.689655 & 25.126436\end{pmatrix}^T\\
S &= \begin{pmatrix}5808.059342 & 597.835204 & 222.029671\\ 597.835204 & 126.053728 & 23.388532\\ 222.029671 & 23.388532 & 23.111734\end{pmatrix}
\end{aligned}$$

b) Using the result from Problem 1 Part e) i., and with $T^2 = n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) = 8.656460$,

$$\frac{n-p}{(n-1)p}T^2 = 2.818382 > 2.713227 = F_\alpha[p, n-p]$$

Thus, we reject H_0 at the 99$\%$ confidence level. If μ_0 is the average over the last 10 years, then the group of students represented by the scores in this dataset is scoring differently because there is less than 5% chance that this sample belongs to a distribution with mean μ_0.

c)

$$\begin{aligned}
\mathcal{R}_{0.95} &= \left\{\mu \in \mathbb{R}^3 : n(\bar{x}-\mu)^T S^{-1}(\bar{x}-\mu) \leq \frac{(n-1)p}{n-p}F_\alpha[p, n-p]\right\}\\
&= \left\{\mu \in \mathbb{R}^3 : (\bar{x}-\mu)^T S^{-1}(\bar{x}-\mu) \leq 0.095787\right\}
\end{aligned}$$

d) The lengths and the directions of axes of the 95% confidence region are given using the eigenvalues and eigenvectors of S.

$$\begin{aligned}
S &= \begin{pmatrix}5808.059342 & 597.835204 & 222.029671\\ 597.835204 & 126.053728 & 23.388532\\ 222.029671 & 23.388532 & 23.111734\end{pmatrix}\\
&= U\Lambda U^T,
\end{aligned}$$

where,

$$\begin{aligned}
U = \begin{pmatrix}u_1 & u_2 & u_3\end{pmatrix} &= \begin{pmatrix}0.993905 & -0.103731 & 0.037307\\ 0.103443 & 0.994589 & 0.009577\\ 0.038099 & 0.005660 & -0.999257\end{pmatrix}\\
\Lambda = \text{diag}_{3\times 3}(\lambda_1, \lambda_2, \lambda_3) &= \begin{pmatrix}5878.791649 & 0 & 0\\ 0 & 63.835099 & 0\\ 0 & 0 & 14.598057\end{pmatrix}
\end{aligned}$$

The directions of axes are given by u_1, u_2, u_3, and the corresponding lengths are given by $\sqrt{k\lambda_1} = 23.729997$, $\sqrt{k\lambda_2} = 2.472768$, $\sqrt{k\lambda_3} = 1.182500$, where

$$k = \frac{(n-1)\,p}{n\,(n-p)} F_{0.05}\,[3,84] = 0.095787,$$

as in Problem 2 Part c).

e) With $\alpha^* := \alpha/p = 0.05/3 = 1/60$, the 95% Bonferroni SCIs are given as,

$$\begin{aligned}
\mathcal{R}_1 &= \left(\bar{x}_1 - t_{\alpha^*/2}\,[n-1]\sqrt{\frac{S_{11}}{n}}, \bar{x}_1 + t_{\alpha^*/2}\,[n-1]\sqrt{\frac{S_{11}}{n}}\right) \\
&= (506.635889, 546.536524) \\
\mathcal{R}_2 &= \left(\bar{x}_2 - t_{\alpha^*/2}\,[n-1]\sqrt{\frac{S_{22}}{n}}, \bar{x}_2 + t_{\alpha^*/2}\,[n-1]\sqrt{\frac{S_{22}}{n}}\right) \\
&= (51.750570, 57.628740) \\
\mathcal{R}_3 &= \left(\bar{x}_3 - t_{\alpha^*/2}\,[n-1]\sqrt{\frac{S_{33}}{n}}, \bar{x}_3 + t_{\alpha^*/2}\,[n-1]\sqrt{\frac{S_{33}}{n}}\right) \\
&= (23.867944, 26.384929)
\end{aligned}$$

Problem 3

a) Using the result from Problem 1 Part e) ii., and with

$$T^2 = n\left((\bar{x}_1, \bar{x}_3) - (\mu_1, \mu_3)\right)^T S_{13}^{-1}\left((\bar{x}_1, \bar{x}_3) - (\mu_1, \mu_3)\right) = 0.063445,$$

where,

$$\begin{aligned}
n &= 87 \\
(\bar{x}_1, \bar{x}_3) &= (526.586206, 25.126436) \\
S_{13} &= \begin{pmatrix} 5808.059342 & 222.029671 \\ 222.029671 & 23.111734 \end{pmatrix},
\end{aligned}$$

we have

$$T^2 = 0.063445 < 9.210340 = \chi^2_{0.01}\,[2]\,,$$

where we get $\chi^2_{0.01}\,[2]$ using a table. Thus, we fail to reject H_0 at 99% confidence level.

b) Using n, $(\bar{x}_1, \bar{x}_3)$, S_{13} and $\chi^2_\alpha\,[2]$ as in Problem 3 Part a),

$$\begin{aligned}
\mathcal{R} &= \left\{(\mu_1, \mu_3) \in \mathbb{R}^2 : n\left((\bar{x}_1, \bar{x}_3) - (\mu_1, \mu_3)\right)^T S_{13}^{-1}\left((\bar{x}_1, \bar{x}_3) - (\mu_1, \mu_3)\right) \le \chi^2_\alpha\,[2]\right\} \\
&= \left\{(\mu_1, \mu_3) \in \mathbb{R}^2 : \left((\bar{x}_1, \bar{x}_3) - (\mu_1, \mu_3)\right)^T S_{13}^{-1}\left((\bar{x}_1, \bar{x}_3) - (\mu_1, \mu_3)\right) \le 0.105865\right\}
\end{aligned}$$

c) With $\alpha^* := \alpha/p = 0.01/2 = 1/200$, the 99% Bonferroni SCIs are given as,

$$\begin{aligned}
\mathcal{R}_1 &= \left(\bar{x}_1 - z_{\alpha^*/2}\sqrt{\frac{S_{11}}{n}}, \bar{x}_1 + z_{\alpha^*/2}\sqrt{\frac{S_{11}}{n}}\right) \\
&= (503.650953, 549.521459) \\
\mathcal{R}_3 &= \left(\bar{x}_3 - z_{\alpha^*/2}\sqrt{\frac{S_{33}}{n}}, \bar{x}_3 + z_{\alpha^*/2}\sqrt{\frac{S_{33}}{n}}\right) \\
&= (23.679650, 26.573222)
\end{aligned}$$

Problem 4

a) We begin by calculating the sample statistics,

$$\begin{aligned}
\bar{x}_1 &= \begin{pmatrix} 12.218611 & 8.112500 & 9.590277 \end{pmatrix}^T \\
\bar{x}_2 &= \begin{pmatrix} 10.105652 & 10.762173 & 18.167826 \end{pmatrix}^T \\
S_1 &= \begin{pmatrix} 23.013360 & 12.366395 & 2.906608 \\ 12.366395 & 17.544110 & 4.773082 \\ 2.906608 & 4.773082 & 13.963334 \end{pmatrix} \\
S_2 &= \begin{pmatrix} 4.362316 & 0.759887 & 2.362099 \\ 0.759887 & 25.851235 & 7.685732 \\ 2.362099 & 7.685732 & 46.654399 \end{pmatrix} \\
S_{\text{pool}} &:= \frac{(n_1 - 1)\, S_1 + (n_2 - 1)\, S_2}{n_1 + n_2 - 2} = \begin{pmatrix} 15.814712 & 7.886690 & 2.696447 \\ 7.886690 & 20.750369 & 5.897262 \\ 2.696447 & 5.897262 & 26.580938 \end{pmatrix}
\end{aligned}$$

Now, using that

$$T^2 := (\bar{x}_1 - \bar{x}_2)^T \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} S_{\text{pool}}^{-1} (\bar{x}_1 - \bar{x}_2) \sim \frac{(n_1 + n_2 - 2)\, p}{n_1 + n_2 - 1 - p} F\left(p, n_1 + n_2 - 1 - p\right),$$

and that

$$T^2 = 50.912786 > 12.930959 = \frac{(n_1 + n_2 - 2)\, p}{n_1 + n_2 - 1 - p} F_\alpha \left[p, n_1 + n_2 - 1 - p\right]$$

Thus, we reject H_0 at the 99% confidence level.

b)

$$\begin{aligned}
\mathcal{R}_{\text{fuel}} &= (\bar{x}_1 - \bar{x}_2)_1 \pm t_{\alpha^*/2}\left[n_1 + n_2 - 2\right]\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{11}} \\
&= (-1.139583, 5.365501) \\
\mathcal{R}_{\text{repair}} &= (\bar{x}_1 - \bar{x}_2)_2 \pm t_{\alpha^*/2}\left[n_1 + n_2 - 2\right]\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{22}} \\
&= (-6.375351, 1.076003) \\
\mathcal{R}_{\text{capital}} &= (\bar{x}_1 - \bar{x}_2)_3 \pm t_{\alpha^*/2}\left[n_1 + n_2 - 2\right]\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{33}} \\
&= (-12.794295, -4.360801)
\end{aligned}$$

where S_{jj} is the j^{th} diagonal entry of S_{pool}. From the intervals, it seems like the capital costs are quite different for the two classes of trucks.

Problem 5

a) First we note that we have the number of parameter, $p = 4$, the number of groups, $G = 3$, the number of samples in each group, $n_1 = n_2 = n_3 = 30$, and the total number of sample, $n = n_1 + n_2 + n_3 = 90$. The group means and the overall mean are calculated to be,

$$\begin{aligned}
\bar{x}_1 &= \begin{pmatrix} 131.37 & 133.60 & 99.17 & 50.53 \end{pmatrix}^T \\
\bar{x}_2 &= \begin{pmatrix} 132.37 & 132.70 & 99.07 & 50.23 \end{pmatrix}^T \\
\bar{x}_3 &= \begin{pmatrix} 134.47 & 133.80 & 96.03 & 50.57 \end{pmatrix}^T \\
\bar{x} &= \begin{pmatrix} 132.73 & 133.37 & 98.09 & 50.44 \end{pmatrix}^T
\end{aligned}$$

Using this, we have

$$B := \sum_{g=1}^{G} n_g \left(\bar{x}_g - \bar{x}\right)\left(\bar{x}_g - \bar{x}\right)^T = \begin{pmatrix} 150.20 & 20.33 & -161.83 & 5.03 \\ 20.30 & 20.60 & -38.73 & 6.43 \\ -161.83 & -38.73 & 190.29 & -10.86 \\ 5.03 & 6.43 & -10.86 & 2.02 \end{pmatrix}$$

$$W := \sum_{g=1}^{G} \sum_{i=1}^{n_g/} \left(x_{gi} - \bar{x}_g\right)\left(x_{gi} - \bar{x}_g\right)^T = \begin{pmatrix} 1785.40 & 172.50 & 128.97 & 289.63 \\ 172.50 & 1924.30 & 178.80 & 171.90 \\ 128.97 & 178.80 & 2153.00 & -1.70 \\ 289.63 & 171.90 & -1.70 & 840.20 \end{pmatrix}$$

$$W + B := \sum_{g=1}^{G} \sum_{i=1}^{n_g} \left(x_{gi} - \bar{x}\right)\left(x_{gi} - \bar{x}\right)^T = \begin{pmatrix} 1935.60 & 192.80 & -32.87 & 294.67 \\ 192.80 & 1944.90 & 140.07 & 178.33 \\ -32.87 & 140.07 & 2343.29 & -12.56 \\ 294.67 & 178.33 & -12.56 & 842.22 \end{pmatrix}$$

$$\Lambda^* := \frac{\det(W)}{\det(W + B)} = \frac{5.669247 \times 10^{12}}{6.829573 \times 10^{12}} = 0.830102$$

Now, with $p \geq 1$ and $G = 3$, we know that,

$$\left(\frac{n-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F\left(2p, 2\left(n-p-2\right)\right),$$

and using the sample at hand, and with $\alpha = 0.05$,

$$\left(\frac{n-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = 2.049068 > 1.993883 = F_\alpha\left[2p, 2\left(n-p-2\right)\right]$$

Hence, we reject H_0 at the 5% level, and conclude that the average skulls sizes were different over the given 3 periods.

b) For all components, $j \in \{1, \ldots, 4\}$, and all pairs of groups, $k < l \in \{1, \ldots, 3\}$, we have the confidence region for the difference in mean component, j, between the groups k and l given by,

$$\left(\bar{x}_{lj} - \bar{x}_{kj}\right) \pm t_{\alpha^*/2}\left[n-G\right]\sqrt{\left(\frac{1}{n_k}+\frac{1}{n_l}\right)\frac{w_{jj}}{n-G}},$$

where $\bar{x}_{kj}$ is the j^{th} component of the mean of the k^{th} group, w_{jj} is the j^{th} diagonal entry of W, and,

$$\alpha^* = \frac{\alpha}{pG\left(G-1\right)/2} = 0.004167$$

i. Component 1 (maximum breadth of skull)

- Group 1 (4000 B.C. period) and Group 2 (3300 B.C. period)

$$\left(\bar{x}_{21} - \bar{x}_{11}\right) \pm t_{\alpha^*/2}\left[n-G\right]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{w_{11}}{n-G}} = \left(-2.442312, 4.442312\right)$$

- Group 1 (4000 B.C. period) and Group 3 (1850 B.C. period)

$$\left(\bar{x}_{31} - \bar{x}_{11}\right) \pm t_{\alpha^*/2}\left[n-G\right]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{11}}{n-G}} = \left(-0.342312, 6.542312\right)$$

- Group 2 (3300 B.C. period) and Group 3 (1850 B.C. period)

$$\left(\bar{x}_{31} - \bar{x}_{21}\right) \pm t_{\alpha^*/2}\left[n-G\right]\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\frac{w_{11}}{n-G}} = \left(-1.342312, 5.542312\right)$$

The maximum breadth of skull seems to vary significantly only between periods 1 and 3 at the 95% confidence level, but not between periods 1 and 2 or between periods 2 and 3.

ii. Component 2 (basibregmatic height of skull)

- Group 1 (4000 B.C. period) and Group 2 (3300 B.C. period)

$$\left(\bar{x}_{22} - \bar{x}_{12}\right) \pm t_{\alpha^*/2}\left[n-G\right]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{w_{22}}{n-G}} = \left(-4.473706, 2.673706\right)$$

- Group 1 (4000 B.C. period) and Group 3 (1850 B.C. period)

$$(\bar{x}_{32} - \bar{x}_{12}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{22}}{n-G}} = (-3.373706, 3.773706)$$

- Group 2 (3300 B.C. period) and Group 3 (1850 B.C. period)

$$(\bar{x}_{32} - \bar{x}_{22}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\frac{w_{22}}{n-G}} = (-2.473706, 4.673706)$$

The basibregmatic height of skull doesn't seem to vary significantly in any pair of periods at the 95% confidence level.

iii. Component 3 (basialveolar length of skull)

- Group 1 (4000 B.C. period) and Group 2 (3300 B.C. period)

$$(\bar{x}_{23} - \bar{x}_{13}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{w_{33}}{n-G}} = (-3.880110, 3.680110)$$

- Group 1 (4000 B.C. period) and Group 3 (1850 B.C. period)

$$(\bar{x}_{33} - \bar{x}_{13}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{33}}{n-G}} = (-6.913443, 0.646776)$$

- Group 2 (3300 B.C. period) and Group 3 (1850 B.C. period)

$$(\bar{x}_{33} - \bar{x}_{23}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\frac{w_{33}}{n-G}} = (-6.813443, 0.746776)$$

The basialveolar length of skull seems to vary significantly between periods 1 and 3 and between periods 2 and 3 at the 95% confidence level, but not between periods 1 and 2.

iv. Component 4 (nasal height of skull)

- Group 1 (4000 B.C. period) and Group 2 (3300 B.C. period)

$$(\bar{x}_{24} - \bar{x}_{14}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{w_{44}}{n-G}} = (-2.661423, 2.061423)$$

- Group 1 (4000 B.C. period) and Group 3 (1850 B.C. period)

$$(\bar{x}_{34} - \bar{x}_{14}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{44}}{n-G}} = (-2.328089, 2.394756)$$

- Group 2 (3300 B.C. period) and Group 3 (1850 B.C. period)

$$(\bar{x}_{34} - \bar{x}_{24}) \pm t_{\alpha^*/2}\,[n-G]\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\frac{w_{44}}{n-G}} = (-2.028089, 2.694756)$$

The nasal height of skull doesn't seem to vary significantly in any pair of periods at the 95% confidence level.