1. The following table gives the estimates of the means and the common dispersion matrix of 3 characters X_1, X_2 and X_3 for each of two groups of individuals on 20 and 72 observations from these groups respectively.

	Means		Dispersion Matrix		
Variables	Gr. 1	Gr. 2	X_1	X_2	X_3
X_1	25.80	28.35	4.7350	0.5622	1.4685
X_2	7.81	7.41		0.1413	0.2174
X_3	10.77	10.75			0.5702

Examine whether the mean vectors of the two groups are significantly different.

Solution:

Theoretical Understanding:

Hotelling T^2 statistic is a multivariate generalization of student's t distribution.

Result: If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then $(\boldsymbol{X} - \boldsymbol{\mu_0})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu_0}) \sim \chi^2_{p,\theta}$, where $\theta = (\boldsymbol{\mu} - \boldsymbol{\mu_0})'\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu_0})$.

Proof: $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \Rightarrow \boldsymbol{X} - \boldsymbol{\mu_0} \sim N_p(\boldsymbol{\mu} - \boldsymbol{\mu_0}, \Sigma)$.

Since Σ is p.d., $\exists$ a non-singular matrix B $\ni$

$$B\Sigma B' = I_p$$

$$\Rightarrow \quad \Sigma = (B'B)^{-1}, \text{ i.e., } \Sigma^{-1} = B'B.$$

Put $\boldsymbol{Y} = B(\boldsymbol{X} - \boldsymbol{\mu_0})$. Then, $\boldsymbol{Y} \sim N_p\left(B(\boldsymbol{\mu} - \boldsymbol{\mu_0}), B\Sigma B' = I_p\right)$.

Hence, the components of $\boldsymbol{Y}$, say $Y_1, Y_2, \ldots, Y_p$ are independently distributed with $Y_i \sim N(\gamma_i, 1)$, for $i = 1(1)p$, where γ_i= i-th component of $\boldsymbol{\gamma} = B(\boldsymbol{\mu} - \boldsymbol{\mu_0})$.

Consider a random sample $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_N}$ be a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$. Suppose we want to test $H_0: \boldsymbol{\mu} = \boldsymbol{\mu_0}$ against $H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$.

<u>Case 1:</u> Σ is known.

In this case, $\overline{\boldsymbol{X}} \sim N_p(\boldsymbol{\mu}, \Sigma/N)$. Then,

$$\chi^2 = (\overline{\boldsymbol{X}} - \boldsymbol{\mu_0})'\left(\frac{\Sigma}{N}\right)^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0}) = N(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0})'\Sigma^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0}) \sim \chi^2_{p,\theta},$$

where $\theta = (\boldsymbol{\mu} - \boldsymbol{\mu_0})'\left(\frac{\Sigma}{N}\right)^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu_0}) = N(\boldsymbol{\mu} - \boldsymbol{\mu_0})'\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu_0})$.

$\lambda = 0 \Leftrightarrow \boldsymbol{\mu} = \boldsymbol{\mu_0}$, that is, H_0 holds.

Under H_0, $\chi^2 \sim$ central χ^2_p.

So, we may take the test criterion to be χ^2, and the critical region is $W = \{(\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N})$: obs. $\chi^2 > \chi^2_p(\alpha)\}$.

<u>Case 2:</u> Σ is unknown.

Here it will be natural to replace Σ in the test criterion by its sample unbiased estimator $\frac{A}{n}$, where $n = N - 1$ and $A = \sum_{i=1}^{N}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$.

Then the test criterion is

$T^2 = Nn(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0})'A^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0})$. This test criterion is known as Hotelling T^2 statistic. For $p = 1$, it reduces to t^2, where t is Student's t statistic.

<u>R code :</u>

```
#load mvtnorm package
library(mvtnorm)
# define mean vectors and covariance matrix
mean_gr1 <- c(25.80, 7.81, 10.77)
mean_gr2 <- c(28.35, 7.41, 10.75)
Sigma=cov_matrix <- matrix(c(4.7350, 0.5622, 1.4685,
                     0.5622, 0.1413, 0.2174,
                     1.4685, 0.2174, 0.5702),
                   nrow = 3, ncol = 3)
x = rbind(mean_gr1,mean_gr2)
# calculate Hotelling's T-Squared statistic
mvt=hotelling.test(x,Sigma,
             n = c(20,72))
mvt
```

<u>Output:</u>

```
Test stat:  5789.4
Numerator df:  3
Denominator df:  1
P-value:  0.02897
```

<u>Interpretation:</u>

<u>The p-value is less than 0.05, which means that we reject the null hypothesis of equal mean vectors. This indicates that there is a significant difference between the mean vectors of the two groups. Therefore, we can conclude that at least one of the variables (X1, X2, X3) has a significantly different mean in Group 1 compared to Group 2.</u>