1. Let, $X \sim N_3 \left(\mu = \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix}, \Sigma = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \right)$

(a) Let, $Y = \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$

So, $Y \sim N_2 \left(\mu_Y = \begin{pmatrix} -3 \\ 2 \end{pmatrix}, \Sigma_Y = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \right)$

$\square$

(b) Now, we need to find the distribution of $Y|X_2 = x_2$,

we know, $Y \sim N_2 \left(\mu_Y = \begin{pmatrix} -3 \\ 2 \end{pmatrix}, \Sigma_Y = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \right)$ and $X_2 \sim N(1, 2)$

So, $\begin{pmatrix} Y \\ X_2 \end{pmatrix} \sim N_3 \left(\begin{bmatrix} \mu_Y \\ 1 \end{bmatrix}, \Sigma_{new} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix} \right)$

where $\Sigma_{11} = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix}$, $\Sigma_{12} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, Σ_{21}, $\Sigma_{22} = 2$

Using the result of conditional multivariate normal distribution,
we get,

$Y|X_2 = x_2 \sim N_2 \left(\mu_Y + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - 1), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$

$\equiv N_2 \left(\begin{pmatrix} -3 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \left(\frac{x_2 - 1}{2} \right), \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} - \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \right)$

$\equiv N_2 \left(\begin{pmatrix} -3 + \frac{x_2-1}{2} \\ 2 + \frac{x_2-1}{2} \end{pmatrix}, \Sigma_{Y|X_2} = \begin{pmatrix} 3.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix} \right)$

$\equiv N_2 \left(\mu_{Y|X_2} = \begin{pmatrix} \frac{x_2-7}{2} \\ \frac{x_2+3}{2} \end{pmatrix}, \Sigma_{Y|X_2} = \begin{pmatrix} 3.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix} \right)$

$\square$

(c) Similarly,

$$X_2 \mid \underset{\sim}{Y} = \underset{\sim}{t} \sim N\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\underset{\sim}{y} - \mu_Y), \ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

$$\equiv N\left(1 + (1\ \ 1)\begin{pmatrix}4 & 0\\0 & 3\end{pmatrix}^{-1}\left[\begin{pmatrix}x_1\\x_3\end{pmatrix} - \begin{pmatrix}-3\\2\end{pmatrix}\right], \ 2 - (1\ 1)\begin{pmatrix}4 & 0\\0 & 3\end{pmatrix}^{-1}\begin{pmatrix}1\\1\end{pmatrix}\right)$$

$$\equiv N\left(1 + (0.25\ \ 0.33)\begin{pmatrix}x_1 + 3\\x_3 - 2\end{pmatrix}, \ 1.42\right)$$

$$\equiv N\left(1 + (0.25x_1 + 0.75 + 0.33x_3 - 0.66), \ 1.42\right)$$

$$\equiv N\left((0.25x_1 + 0.33x_3 + 1.09), \ 1.42\right)$$

<div align="center">□</div>

(d) Now, $z = X_1 + 3X_2$

Let, $Z = \underset{\sim}{b}'\underset{\sim}{X} = (1\ \ 3\ \ 0)\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix}$

So, $Z \sim N\left(\underset{\sim}{b}'\underset{\sim}{\mu}, \ \underset{\sim}{b}'\Sigma\underset{\sim}{b}\right)$,

where, $\underset{\sim}{b}'\underset{\sim}{\mu} = (1\ \ 3\ \ 0)\begin{pmatrix}-3\\1\\2\end{pmatrix} = 0$

$$\underset{\sim}{b}'\Sigma\underset{\sim}{b}' = (1\ \ 3\ \ 0)\begin{pmatrix}4 & 1 & 0\\1 & 2 & 1\\0 & 1 & 3\end{pmatrix}\begin{pmatrix}1\\3\\0\end{pmatrix} = 28$$

Thus, $Z \sim N(0, 28)$

<div align="center">□.</div>

2. Let Y be the concentration of a certain substance in liver tissue after the administration of a drug at 30 days (indicated by Y_2) and 60 days (indicated by Y_3) and Y_1 indicating the first administration.

Now, $\bar{y} = \begin{pmatrix} 14 \\ 24 \\ 22 \end{pmatrix}$ and $S = \begin{pmatrix} 30 & 10 & 14 \\ 10 & 15 & 4 \\ 14 & 4 & 37 \end{pmatrix}$

(a) Here, we want to test :

$H_0 : \mu_1 = \mu_2 = \mu_3$ vs $H_1 :$ not H_0.

So, we use the test statistic :

$F_{stat} = \dfrac{u_A - u}{(u_A - 1)u} T^2$, where, $T^2 = u\,(c\,\bar{y})'\,(c\,S\,c')^{-1}\,(c\,\bar{y})$

$u_A = 25$, $c = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$, $r = \text{rank}(c)$
$= 2$.

So, $T^2 = 25 \cdot \left[\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 14 \\ 24 \\ 22 \end{pmatrix} \right]' \left[\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 30 & 10 & 14 \\ 10 & 15 & 4 \\ 14 & 4 & 37 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \right]^{-1}$

$\left[\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 14 \\ 24 \\ 22 \end{pmatrix} \right]$

$= 111.43$

Thus, $F_{stat} = \dfrac{(25-2)}{(25-1)\,2}\, 11.43 = 53.39$

The critical region is $F < 0.025$ and $F > 4.349$. So, as $F_{stat} > 4349$ we reject the null hypothesis at 5% level of significance and conclude that the mean concentrations of the substance at different times after the administration of the drug are not same.

$\square$

(b) Let $\underline{z}$ denote the results for drug B.

So, $\underline{z} = \begin{pmatrix} 15 \\ 20 \\ 24 \end{pmatrix}$ and $W = \begin{pmatrix} 26 & 12 & 10 \\ 12 & 17 & 8 \\ 10 & 8 & 43 \end{pmatrix}$

(i) Let the population covariance-variance matrices for drugs A and B be the same.

So, an estimate of the pooled sample variance is:

$$S_{pooled} = \frac{(n_A - 1)S + (n_B - 1)W}{(n_A + n_B - 2)} \quad , \text{where}, \ n_B = 17$$

$$= \frac{24}{40} \begin{pmatrix} 30 & 10 & 14 \\ 10 & 15 & 4 \\ 14 & 4 & 37 \end{pmatrix} + \frac{16}{40} \begin{pmatrix} 26 & 12 & 10 \\ 12 & 17 & 8 \\ 10 & 8 & 43 \end{pmatrix}$$

$$= \begin{pmatrix} 28.4 & 10.8 & 12.4 \\ 10.8 & 15.8 & 5.6 \\ 12.4 & 5.6 & 39.4 \end{pmatrix}$$

$\quad\quad\quad\quad\quad\quad\quad D$

(ii) To perform Bartlett's test, we test $H_0 : \Sigma_A = \Sigma_B$ vs $H_1 :$ not H_0.

First, we find $M = (N - k) \log |S| - \sum_{i=1}^{K} (n_i - 1) \log |S_i|$,

where, $N = \sum n_i = 42, \ k = 2_B$

So, $M = -120.09$.

Next, $c^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(K-1)} \left[\sum_{i=1}^{K} \left(\frac{1}{n_i - 1} \right) - \frac{1}{N-k} \right]$, where, $p = 3$

$\quad\quad\quad = 0.91$

The test statistic is : $Me^{-1} = -109.28$

Now, the critical region is greater than $X^2_{1-\alpha;(K-1)\frac{P(P+1)}{2}}$

$$= X^2_{0.95; 6} = 12.59$$

As, $Me^{-1} \not> 12.59$, we fail to reject Ho at 5% level of significance and thus conclude that the population variance-covariances matrices are not same.

☐.

(iii) The Mahalanobis distance is estimated as :

$$d^2(\bar{y},\bar{z}) = (\bar{y}-\bar{z})' S^{-1}_{pooled} (\bar{y}-\bar{z})$$

$$= \begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix}' \begin{pmatrix} 28.4 & 10.8 & 12.4 \\ 10.8 & 15.8 & 5.6 \\ 12.4 & 5.6 & 39.4 \end{pmatrix} \begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix}$$

$$= 1.78.$$

☐

(iv) To test Ho : $\underset{\sim}{\mu_1} = \underset{\sim}{\mu_2}$ vs H_1 : not Ho.

The test statistic is $T^2 = \dfrac{n_A n_B}{N} d^2(\bar{y},\bar{z})$

$$= \dfrac{25 \times 17}{42} \times 1.78 = 18.01$$

The critical value is greater than $\dfrac{(n_A+n_B-2)P}{n_A+n_B-P-1} F_{1-\alpha; P, n_A+n_B-P-1} = 0.12$

As $T^2 > 0.12$, we reject H_0 at 5% level of significance and conclude that the means for the two drugs are ~~equal~~ unequal.

(iv) Here, we test: $c(\underline{\mu}_1 - \underline{\mu}_2) = \underline{0}$ vs H_1: not H_0,

where, $c = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$, $r = rank(c) = 2$.

The test statistic is $\dfrac{n_A + n_B - r - 1}{(n_A + n_B - 2) r} T^2$, where $T^2 = \dfrac{n_A n_B}{n_A + n_B} (\bar{y} - \bar{z})' c'$

$$(c s c')^{-1} c (\bar{y} - \bar{z})$$

$$= \frac{25 + 17 - 2 - 1}{(25 + 17 - 2) \times 2} \times 14.27 \qquad = 14.27$$

$$= 6.96$$

The critical value is $f_{1-\alpha; \, r, \, n_A + n_B - r - 1} = 0.05$

Now, as $6.96 > 0.05$, we reject H_0 at 5% level of significance and conclude that there is presence of drug by time interaction.

□

3. $n_1 = 16$, $n_2 = 11$, $p = 4$.

(a) $\underline{\mu}_1$ and $\underline{\mu}_2$ represents the mean vectors of the groups of boys and girls respectively,

Here, we test $H_0: \underline{\mu}_1 = \underline{\mu}_2$ vs H_1: not H_0

$T^2 = 16.5$

The critical region is greater than $\dfrac{(n_1 + n_2 - 2) p}{n_1 + n_2 - p - 1} f_{1-\alpha; \, p, \, n_1 + n_2 - p - 1}$

$$= \frac{25 \times 4}{22} \qquad f_{0.05; \, 4, 22} = 0.77$$

As $T^2 > 0.77$, we reject H_0 at 5% level of significance and conclude that mean profiles of the boys and girls are not same. □

(b) To test: $H_0: c(\underset{\sim}{\mu_1} - \underset{\sim}{\mu_2}) = \underset{\sim}{0}$ vs H_1: not H_0

where, $c = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & & \end{pmatrix}$

where, $c = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$, rank$(c) = 3 = r$ (say)

The test statistic is

$$\frac{n_1 + n_2 - r - 1}{(n_1 + n_2 - 2)r} T^2$$

$$= \frac{16 + 11 - 3 - 1}{(16 + 11 - 2)3} \, 8.8$$

$$= 2.70 .$$

The critical value is greater than $F_{1-\alpha; \, r, \, n_1 + n_2 - r - 1} = 0.12$

As, $2.70 > 0.12$, we reject H_0 at 5% level of significance and thus conclude that there is presence of group by time interaction.

(c) As we reject the null hypothesis in part (b), we might further do post-hoc tests such as pairwise comparison or even perform profile analysis to understand the nature of the interaction.

PROBLEM 4.

Part (a)

```r
library(ggplot2)
library(nlme)
library(dplyr)
library(reshape2)
library(profileR)
data_4 = read.table("C:/Users/Jayaditya Nath/Documents/PothoffRoy1964.dat")

mean_profiles_4 = aggregate(. ~ data_4$V2, data = data_4[, -c(1, 2)], FUN = mean)

# Plot mean profiles
plot(seq(8,14,2), mean_profiles_4[1, -1], type = 'l', col = 1, ylim = c(min(mean_profiles_4[-1]), max(me
lines(seq(8,14,2), mean_profiles_4[2, -1], col = 2)
legend("bottomright", legend = c("F","M"), col = 1:2, lty = 1,cex = 0.5)
```

Mean Profiles for different ages
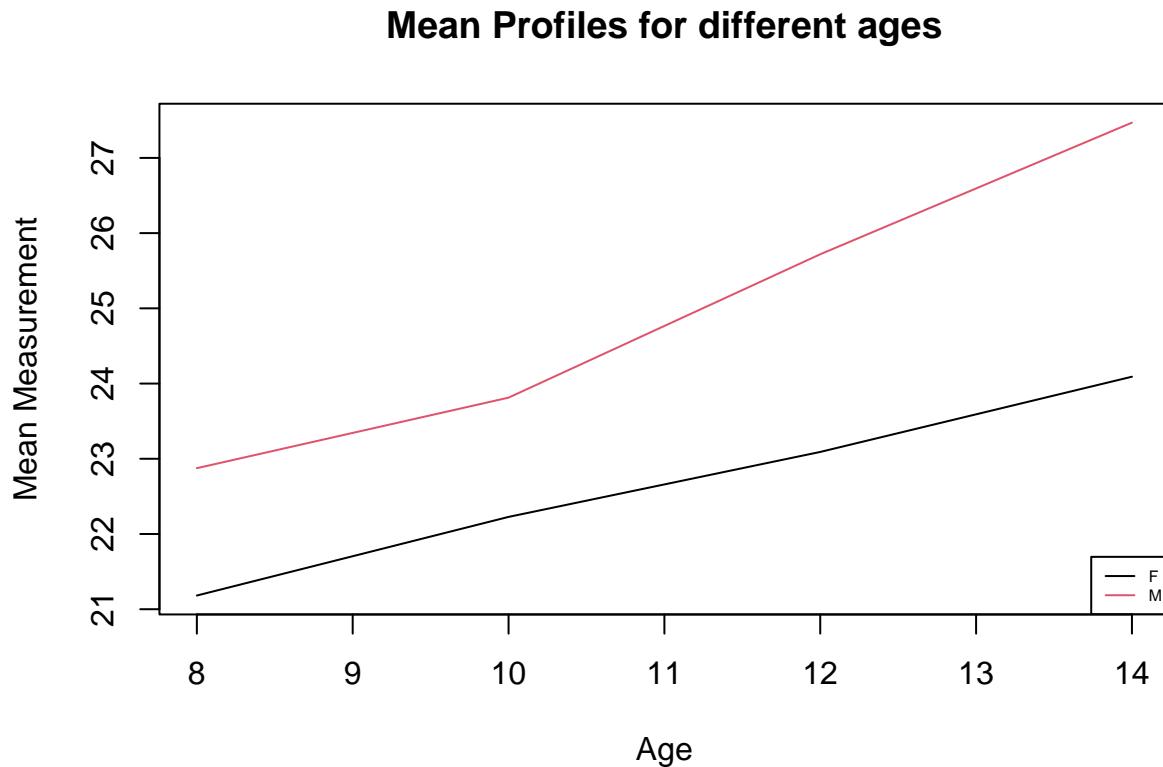
From the profile plots, it is clearly visible that there is a difference between the dental measurements of males and females across different ages. The profiles are not at all similar to each other and males tend to have greater dental measurements compared to females at different ages.

Part (b)

```r
# Performing MANOVA
anova(lm(cbind(V3,V4,V5,V6)~V2,data = data_4))
```

1

```
## Analysis of Variance Table
##
##             Df Pillai approx F num Df den Df  Pr(>F)
## (Intercept)  1 0.9943    960.17      4     22 < 2e-16 ***
## V2           1 0.3977      3.63      4     22 0.02034 *
## Residuals   25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, we are interested in testing :

$H_0 : C(\mu_{males} - \mu_{females}) = \tilde{0}$ vs H_a : not H_0

The Pillai's test statistic for performing MANOVA is :

$$V = \sum_{i=1}^{p} \frac{\lambda_i}{1 + \lambda_i}$$

, where λ_i is the number of non-zero eigen values of the product of the inverse of the within-group variance-covariance matrix with the between-group variance-covariance matrix.

The p-value for the associated test is 0.02.

Thus, we are able to reject the null hypothesis at 5% level of significance and conclude that the dental measurements of femlaes and males vary across time.

Part (c)

```
# Melt the data from wide to long format
names(data_4) = c("ID", "Gender", "age_8", "age_10", "age_12", "age_14")
data_4_long = reshape(data_4,varying = list(c("age_8", "age_10", "age_12", "age_14")),v.names = "Measure

# Fit a linear mixed model with compound symmetry covariance structure
lmm_results = lme(fixed = Measurement ~ Gender*Age,
                  random = ~ 1 | ID,
                  data = data_4_long,
                  correlation = corCompSymm(form = ~ 1 | ID))
summary(lmm_results)
```

```
## Linear mixed-effects model fit by REML
##   Data: data_4_long
##        AIC      BIC    logLik
##   447.7572 466.268 -216.8786
##
## Random effects:
##  Formula: ~1 | ID
##         (Intercept) Residual
## StdDev:    1.816214 1.386382
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | ID
##  Parameter estimate(s):
## Rho
##   0
## Fixed effects:  Measurement ~ Gender * Age
##                 Value Std.Error DF   t-value p-value
```

```
## (Intercept) 17.372727 1.1835071 79 14.679023  0.0000
## GenderM      -1.032102 1.5374208 25 -0.671321  0.5082
## Age           0.479545 0.0934698 79  5.130483  0.0000
## GenderM:Age   0.304830 0.1214209 79  2.510520  0.0141
##  Correlation:
##           (Intr) GendrM Age
## GenderM     -0.770
## Age         -0.869  0.669
## GenderM:Age  0.669 -0.869 -0.770
##
## Standardized Within-Group Residuals:
##        Min          Q1         Med          Q3         Max
## -3.59804400 -0.45461690  0.01578365  0.50244658  3.68620792
##
## Number of Observations: 108
## Number of Groups: 27
```

From the linear mixed model with compound symmetry variance-covariance structure, we can see that the main effect **Gender** is not significant, though **Age** seems to be quite significant. The interaction term is also significant in the model. I feel that somehow the MANOVA and the LMM model results are at agreement with each other in terms of the fact that the interaction between age and gender is significant, however, they disagree on the fact that **Gender** itself is significant or not.

PROBLEM 5.

```r
# Define the data
data_5 = data.frame(
  Reinforcement_Schedule = rep(1:3, each = 5),
  Rat = rep(1:5, times = 3),  Cond_1 = c(29, 24, 31, 41, 30, 25, 20, 35, 35, 26, 10, 9, 7, 8, 11),
  Cond_2 = c(20, 15, 19, 11, 20, 17, 12, 16, 8, 18, 18, 10, 18, 19, 20),
  Cond_3 = c(21, 10, 10, 15, 27, 19, 8, 9, 14, 18, 16, 18, 19, 20, 17),
  Cond_4 = c(18, 8, 31, 42, 53, 17, 8, 28, 40, 51, 14, 11, 12, 5, 6))


# Calculate mean time for each condition within each reinforcement schedule
mean_profiles = aggregate(. ~ data_5$Reinforcement_Schedule, data = data_5[, -c(1, 2)], FUN = mean)

# Plot mean profiles
plot(1:4, mean_profiles[1, -1], type = 'l', col = 1, ylim = c(0, max(mean_profiles[-1])), xlab = "Condi
lines(1:4, mean_profiles[2, -1], col = 2)
lines(1:4, mean_profiles[3, -1], col = 3)
legend("bottomright", legend = c("Sc 1", "Sc 2", "Sc 3"), col = 1:3, lty = 1,cex = 0.5)
```

Mean Profiles for Reinforcement Schedules
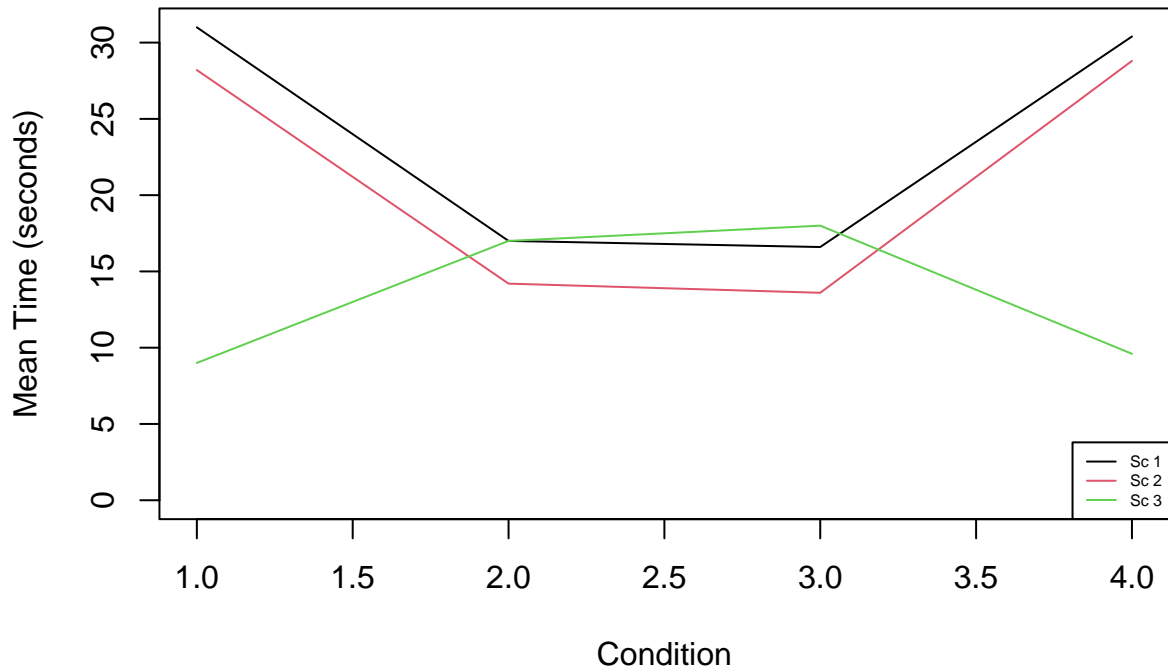
From the profiles plot, it seems that the reinforcement schedules 1,2 and 3 vary vastly among themselves across different conditions although there might be some visible similarity among the schedules 1 and 2 over different conditions.

Part (b)

```
# Reshape data into long format
data_5_long = reshape(data_5,varying = list(c("Cond_1","Cond_2","Cond_3","Cond_4")),v.names = "Time",tir

# Perform one-way ANOVA
anova_res_5 = aov(formula = Time ~ Reinforcement_Schedule + Condition, data = data_5_long)
summary(anova_res_5)
```

```
##                         Df Sum Sq Mean Sq F value  Pr(>F)
## Reinforcement_Schedule  1   1071  1071.2  10.382 0.00211 **
## Condition               1      0     0.3   0.003 0.95938
## Residuals              57   5881   103.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By performing an Analysis of Variance, we can infer that the reinforcement schedule has a significant overall effect. The associated p-value is approximately 0.0021.

Part (c)

```
# Perform profile analysis comparing schedules 1 and 2 across conditions
mod_5_12 = pbg(data = as.matrix(subset(data_5[,3:6],data_5$Reinforcement_Schedule%in%c(1,2))),group = as
summary(mod_5_12)
```

```
## Call:
## pbg(data = as.matrix(subset(data_5[, 3:6], data_5$Reinforcement_Schedule %in%
##      c(1, 2))), group = as.matrix(subset(data_5[, "Reinforcement_Schedule"],
##      data_5$Reinforcement_Schedule %in% c(1, 2))), original.names = T)
##
## Hypothesis Tests:
## $'Ho: Profiles are parallel'
##    Multivariate.Test   Statistic     Approx.F num.df den.df   p.value
## 1             Wilks 0.995637701 0.008762825      3      6 0.9987461
## 2            Pillai 0.004362299 0.008762825      3      6 0.9987461
## 3  Hotelling-Lawley 0.004381412 0.008762825      3      6 0.9987461
## 4               Roy 0.004381412 0.008762825      3      6 0.9987461
##
## $'Ho: Profiles have equal levels'
##            Df Sum Sq Mean Sq F value Pr(>F)
## group       1   16.3   16.26   0.394  0.548
## Residuals   8  330.3   41.29
##
## $'Ho: Profiles are flat'
##           F df1 df2    p-value
## 1 7.314892   3   6 0.01982651
```

On the basis of profile analysis, we can conclude at 5% level of significance on the basis of p-value that the schedules 1 and 2 behave similarly across different conditions and evidently they are parallel to each other.

```
# Perform profile analysis comparing schedules 1,2 with 3 across conditions
data_5_grouped = data_5 %>% mutate(Reinforcement_Schedule= case_when(Reinforcement_Schedule==1~1.2,
                                                                     Reinforcement_Schedule==2~1.2,
                                                                     Reinforcement_Schedule==3~3))

mod_512_3 = pbg(data = as.matrix(data_5_grouped[,3:6]),group = as.matrix(data_5_grouped$Reinforcement_Sc
summary(mod_512_3)
```

```
## Call:
## pbg(data = as.matrix(data_5_grouped[, 3:6]), group = as.matrix(data_5_grouped$Reinforcement_Schedule
##      original.names = T)
##
## Hypothesis Tests:
## $'Ho: Profiles are parallel'
##    Multivariate.Test Statistic Approx.F num.df den.df     p.value
## 1             Wilks 0.2519382 10.88717      3     11 0.001275056
## 2            Pillai 0.7480618 10.88717      3     11 0.001275056
## 3  Hotelling-Lawley 2.9692270 10.88717      3     11 0.001275056
## 4               Roy 2.9692270 10.88717      3     11 0.001275056
##
## $'Ho: Profiles have equal levels'
##            Df Sum Sq Mean Sq F value  Pr(>F)
## group       1  274.5  274.52   10.19 0.00708 **
```

```
## Residuals    13  350.3    26.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $'Ho: Profiles are flat'
##            F df1 df2    p-value
## 1 4.154268   3  11 0.03395195
```

After grouping the schedules 1 and 2 and comparing with the schedule 3, it is evident at 5% level of significance that the profiles for the combined groups behave significantly different from the profile of the third schedule and thus are not parallel to each other.