

P8157 HW1

Shihui Zhu

P8157 - Analysis of Longitudinal Data Homework - 1

Part A

1. Plasma inorganic phosphate measurements obtained from 13 control and 20 obese patients 0, 0.5, 1, 1.5, 2, and 3 hours after an oral glucose challenge. The investigators intend to test the following hypotheses using Hotelling's T2 statistic:

- (a) To test the null hypothesis that the group means are the same at all six measurement times.

$H_0 : \mathbf{ABC} = \mathbf{D}$. The group means are the same at all six measurement times, without assuming parallelism:

$$\begin{aligned}\mu_{11} &= \mu_{21} \\ \mu_{12} &= \mu_{22} \\ &\dots \\ \mu_{16} &= \mu_{26}\end{aligned}$$

$H_1 : \mathbf{ABC} \neq \mathbf{D}$. The group means are NOT the same for all points in the six measurement times.

The matrices set-up:

$$\mathbf{B}_{2,6} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{16} \\ \mu_{21} & \mu_{22} & \dots & \mu_{26} \end{bmatrix}$$

$$\mathbf{C}_{6,6} = \mathbf{I}_6 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\begin{aligned}\mathbf{A}_{1,2} &= [\mathbf{I}_1, -\mathbf{1}_1] = [1 \ -1] \\ \mathbf{D} &= \mathbf{0}_{1,6}\end{aligned}$$

- (b) To test whether the profiles in the two groups are parallel.

$H_0 : \mathbf{ABC} = \mathbf{D}$. Profiles in the two groups are parallel i.e. no group-time interaction, the slopes between each time points are identical between the two groups:

$$\begin{aligned}
\mu_{12} - \mu_{11} &= \mu_{22} - \mu_{21} \\
\mu_{13} - \mu_{12} &= \mu_{23} - \mu_{22} \\
&\dots \\
\mu_{16} - \mu_{15} &= \mu_{26} - \mu_{25}
\end{aligned}$$

$H_1 : \mathbf{ABC} \neq \mathbf{D}$. Profiles in the two groups are NOT parallel.

The matrices set-up:

$$\mathbf{B}_{2,6} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{16} \\ \mu_{21} & \mu_{22} & \dots & \mu_{26} \end{bmatrix}$$

$$\mathbf{C}_{6,5} = \begin{bmatrix} -1 & 0 & \dots & 0 \\ 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\begin{aligned}
\mathbf{A}_{1,2} &= [\mathbf{I}_1, -\mathbf{1}_1] = [1 \ -1] \\
\mathbf{D} &= \mathbf{0}_{1,5}
\end{aligned}$$

- (c) To test whether the differences in means at 2 and 3 hours after an oral glucose challenge are different between the control and obese patients.

$H_0 : \mathbf{ABC} = \mathbf{D}$. The differences in means at 2 and 3 hours after an oral glucose challenge are the same between the control and obese patients, without assuming parallelism:

$$\mu_{16} - \mu_{15} = \mu_{26} - \mu_{25}$$

$H_1 : \mathbf{ABC} \neq \mathbf{D}$. The differences in means at 2 and 3 hours after an oral glucose challenge are different between the control and obese patients.

The matrices set-up:

$$\mathbf{B}_{2,6} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{16} \\ \mu_{21} & \mu_{22} & \dots & \mu_{26} \end{bmatrix}$$

$$\mathbf{C}_{6,1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ -1 \\ 1 \end{bmatrix}$$

$$\begin{aligned}
\mathbf{A}_{1,2} &= [\mathbf{I}_1, -\mathbf{1}_1] = [1 \ -1] \\
\mathbf{D} &= \mathbf{0}
\end{aligned}$$

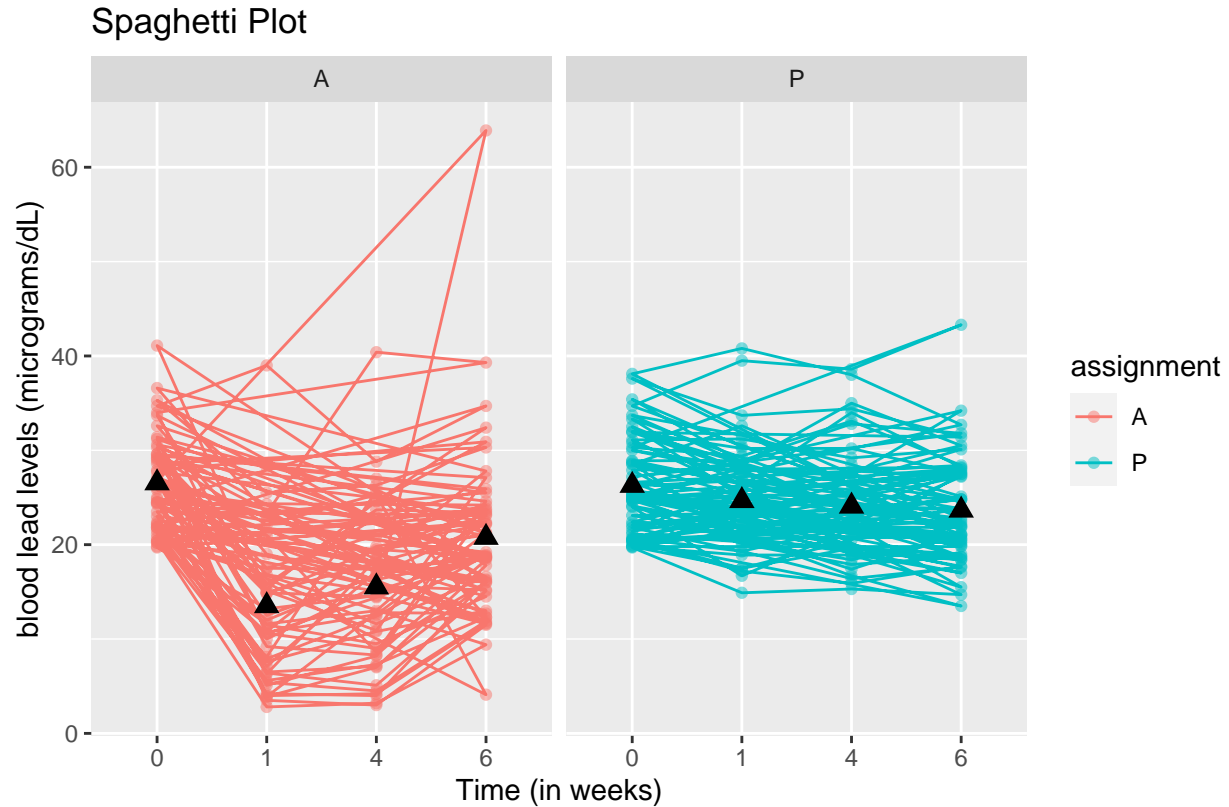
Part B

1. EDA of the Treatment of Lead-Exposed Children (TLC) trial

```
data.TLC <- fread("TLC.dat")
colnames(data.TLC) <- c("subject_id", "assignment", "week0", "week1", "week4", "week6")
head(data.TLC)
```

```
##      subject_id assignment week0 week1 week4 week6
## 1:             1          P  30.8  26.9  25.8  23.8
## 2:             2          A  26.5  14.8  19.5  21.0
## 3:             3          A  25.8  23.0  19.1  23.2
## 4:             4          P  24.7  24.5  22.0  22.5
## 5:             5          A  20.4   2.8   3.2   9.4
## 6:             6          A  20.4   5.4   4.5  11.9
```

```
data.TLC %>%
  pivot_longer(cols = starts_with("week"),
               names_to = "week",
               names_prefix = "week",
               values_to = "bll") %>%
  ggplot(aes(x = week, y = bll, group = assignment, color = assignment)) +
  geom_point(alpha = .5) +
  geom_path() +
  labs(
    title = "Spaghetti Plot",
    x = "Time (in weeks)",
    y = "blood lead levels (micrograms/dL)",
    caption = "Data from the treatment of Lead- Exposed Children (TLC) trial"
  ) +
  stat_summary(aes(group = 1), geom = "point", fun = mean,
               shape = 17, size = 3, color = "black") +
  facet_grid(.~assignment)
```



The triangle in the above plot indicates the group mean at a specific time point. From the spaghetti plot above we found that the profiles among the two groups (A = intervention with succimer, P = Placebo) are not parallel. However, the two groups tend to have similar blood lead level at the baseline. The group treated with succimer tends to have a lower blood lead level compared with the placebo group. The blood lead level of the group treated with succimer tend to have a big drop after 1 week, and it slowly inclines in the following measurements.

We can perform a two-sample Hotelling's T-squared test to check whether there is difference between the group means of the two treatment groups at all 4 measurement time points.

H_0 : The group means of blood lead levels of the two treatment groups are identical at all 4 measurement time points.

$$\begin{aligned}\mu_{11} &= \mu_{21} \\ \mu_{12} &= \mu_{22} \\ \mu_{13} &= \mu_{23} \mu_{14} = \mu_{24}\end{aligned}$$

H_1 : The group means of blood lead levels of the two treatment groups are different at some time points.

```
test14 = hotelling.test(~assignment, data=data.TLC[,-1])
test14
```

```
## Test stat: 112.01
## Numerator df: 4
## Denominator df: 95
## P-value: 4.996e-15
```

The test statistics is 112.01 and the p-value is smaller than 0.05. Therefore we reject the null hypothesis and conclude that the group means of blood lead levels of the two treatment groups are not the same across all time points under significance level of .05.

2. Plasma inorganic phosphate measurements

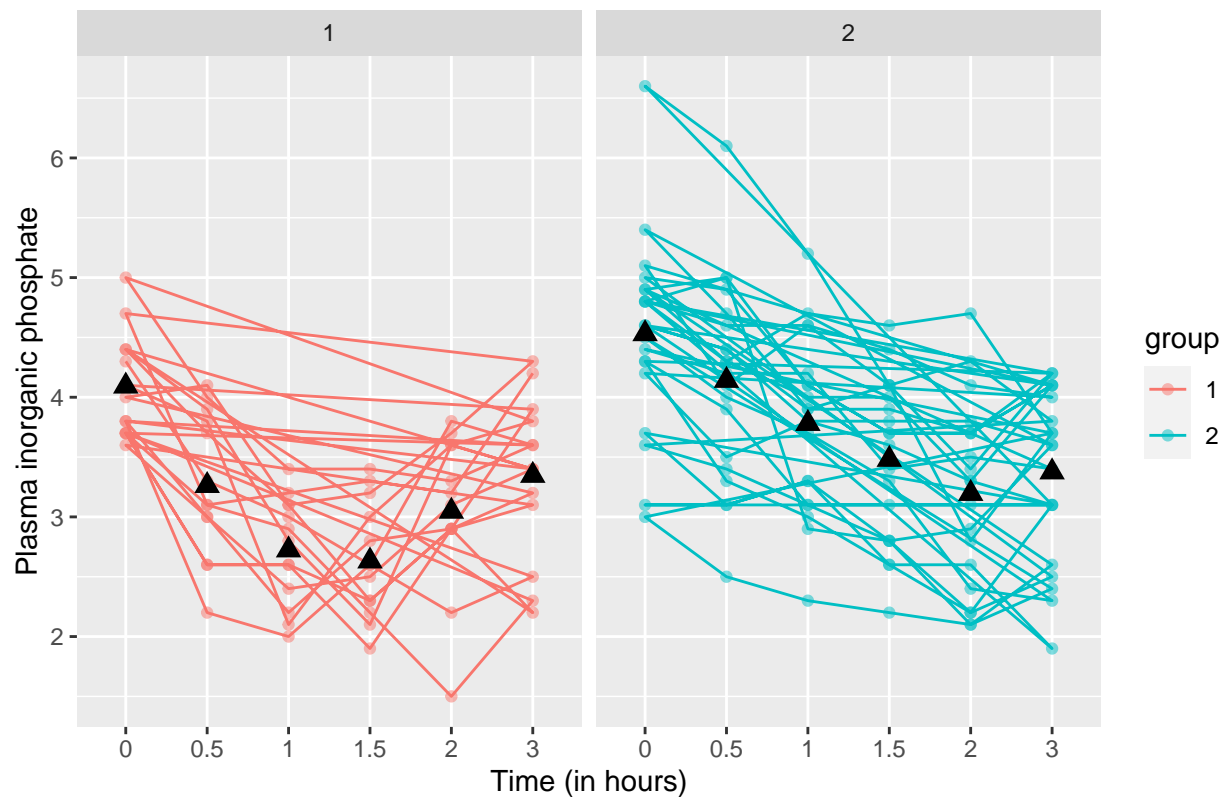
```
data.zer <- fread("ZERBE2.DAT")
colnames(data.zer) <- c("group", "subject_id", "hour0", "hour0.5", "hour1", "hour1.5", "hour2", "hour3")
head(data.zer)
```

EDA

```
##      group subject_id hour0 hour0.5 hour1 hour1.5 hour2 hour3
## 1:      1           1  4.3    3.3   3.0    2.6   2.2   2.5
## 2:      1           2  3.7    2.6   2.6    1.9   2.9   3.2
## 3:      1           3  4.0    4.1   3.1    2.3   2.9   3.1
## 4:      1           4  3.6    3.0   2.2    2.8   2.9   3.9
## 5:      1           5  4.1    3.8   2.1    3.0   3.6   3.4
## 6:      1           6  3.8    2.2   2.0    2.6   3.8   3.6
```

```
data.zer %>%
  mutate(group = as.factor(group)) %>%
  pivot_longer(cols = starts_with("hour"),
               names_to = "hours",
               names_prefix = "hour",
               values_to = "pip") %>%
  ggplot(aes(x = hours, y = pip, group = group, color = group)) +
  geom_point(alpha = .5) +
  geom_path() +
  labs(
    title = "Spaghetti Plot",
    x = "Time (in hours)",
    y = "Plasma inorganic phosphate"
  ) +
  stat_summary(aes(group = 1), geom = "point", fun = mean,
               shape = 17, size = 3, color = "black") +
  facet_grid(.~group)
```

Spaghetti Plot



The triangle in the above plot indicates the group mean at a specific time point. From the spaghetti plot above we found that the profiles among the two groups (1 = control, 2 = obese) are not parallel. The obese group tends to have a higher level of plasma inorganic phosphate at the baseline, but the two group tends to have the same plasma inorganic phosphate after 2-3 hours.

(a) To test the null hypothesis that the group means are the same at all six measurement times.

```
test.a = hotelling.test(~group, data=data.zer[, -2])
test.a
```

```
## Test stat: 61.187
## Numerator df: 6
## Denominator df: 26
## P-value: 3.495e-05
```

The p-value is very small. Therefore we reject the null hypothesis and conclude that the group means are not the same for all points in the six measurement times at a significance level of .05.

(b) To test whether the profiles in the two groups are parallel

```
data.zer2 <- t(data.zer[, -c(1:2)])
head(data.zer2)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## hour0  4.3  3.7  4.0  3.6  4.1  3.8  3.8  4.4  5.0  3.7  3.7  4.4  4.7
```

```
## hour0.5  3.3  2.6  4.1  3.0  3.8  2.2  3.0  3.9  4.0  3.1  2.6  3.7  3.1
## hour1    3.0  2.6  3.1  2.2  2.1  2.0  2.4  2.8  3.4  2.9  2.6  3.1  3.2
## hour1.5  2.6  1.9  2.3  2.8  3.0  2.6  2.5  2.1  3.4  2.2  2.3  3.2  3.3
## hour2    2.2  2.9  2.9  2.9  3.6  3.8  3.1  3.6  3.3  1.5  2.9  3.7  3.2
## hour3    2.5  3.2  3.1  3.9  3.4  3.6  3.4  3.8  3.6  2.3  2.2  4.3  4.2
##          [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## hour0      4.3   5.0   4.6   4.3   3.1   4.8   3.7   5.4   3.0   4.9   4.8   4.4
## hour0.5    3.3   4.9   4.4   3.9   3.1   5.0   3.1   4.7   2.5   5.0   4.3   4.2
## hour1      3.0   4.1   3.9   3.1   3.3   2.9   3.3   3.9   2.3   4.1   4.7   4.2
## hour1.5    2.6   3.7   3.9   3.1   2.6   2.8   2.8   4.1   2.2   3.7   4.6   3.4
## hour2      2.2   3.7   3.7   3.1   2.6   2.2   2.9   2.8   2.1   3.7   4.7   3.5
## hour3      2.5   4.1   4.2   3.1   1.9   3.1   3.6   3.7   2.6   4.1   3.7   3.4
##          [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33]
## hour0      4.9   5.1   4.8   4.2   6.6   3.6   4.5   4.6
## hour0.5    4.3   4.1   4.6   3.5   6.1   3.4   4.0   4.4
## hour1      4.0   4.6   4.6   3.8   5.2   3.1   3.7   3.8
## hour1.5    4.0   4.1   4.4   3.6   4.1   2.8   3.3   3.8
## hour2      3.3   3.4   4.1   3.3   4.3   2.1   2.4   3.8
## hour3      4.1   4.2   4.0   3.1   3.8   2.4   2.3   3.6
```

```
# contrast matrix
cmat <- matrix(c(-1,1,0,0,0,0,
                 0,-1,1,0,0,0,
                 0,0,-1,1,0,0,
                 0,0,0,-1,1,0,
                 0,0,0,0,-1,1), ncol = 6, byrow = TRUE)

cmat
```

```
##          [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]     -1     1     0     0     0     0
## [2,]      0    -1     1     0     0     0
## [3,]      0     0    -1     1     0     0
## [4,]      0     0     0    -1     1     0
## [5,]      0     0     0     0    -1     1
```

```
data.zer3 <- cmat %*% data.zer2
data.zer3 <- t(data.zer3)
data.zer3 <- cbind(data.zer[,1],data.zer3)

head(data.zer3)
```

```
##    group  V1  V2  V3  V4  V5
## 1:     1 -1.0 -0.3 -0.4 -0.4  0.3
## 2:     1 -1.1  0.0 -0.7  1.0  0.3
## 3:     1  0.1 -1.0 -0.8  0.6  0.2
## 4:     1 -0.6 -0.8  0.6  0.1  1.0
## 5:     1 -0.3 -1.7  0.9  0.6 -0.2
## 6:     1 -1.6 -0.2  0.6  1.2 -0.2
```

```
test.b <- hotelling.test(~group, data = data.zer3)

test.b
```

```
## Test stat: 46.962
## Numerator df: 5
## Denominator df: 27
## P-value: 8.344e-05
```

The p-value is very small. Therefore we reject the null hypothesis and conclude that the profiles are not parallel for the two groups at a significance level of .05.

- (c) To test whether the differences in means at 2 and 3 hours after an oral glucose challenge are different between the control and obese patients.

```
# contrast matrix
cmat.c <- t(c(0,0,0,0,-1,1))
cmat.c
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    0    0    0    0   -1    1
```

```
data.zer4 <- cmat.c %*% data.zer2
data.zer4 <- t(data.zer4)
data.zer4 <- cbind(data.zer[,1],data.zer4)

head(data.zer4)
```

```
##      group  V1
## 1:      1 0.3
## 2:      1 0.3
## 3:      1 0.2
## 4:      1 1.0
## 5:      1 -0.2
## 6:      1 -0.2
```

```
test.c <- hotelling.test(~group, data = data.zer4)

test.c
```

```
## Test stat: 0.41711
## Numerator df: 1
## Denominator df: 31
## P-value: 0.5231
```

The p-value is $0.5231 > 0.05$. Therefore we fail to reject the null hypothesis and conclude that there is not enough evidence showing that the differences in means at 2 and 3 hours after an oral glucose challenge are different between the control and obese patients.