HW1

Anna Ma

2022-09-25

Part A

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.

Groups: S=2

Time points: t=6

Number of patients: N $= n_c + n_o = 13 + 20 = 33$

Multivariate Model Set Up:

$$Y_{N\times t} = X_{N\times S}B_{S\times t} + E_{N\times t}$$
$$Y_{33\times 6} = X_{33\times 2}B_{2\times 6} + E_{33\times 6}$$

The investigators intend to test the following hypotheses using Hotelling's statistic :

(a) To test the null hypothesis that the group means are the same at all six measurement times. Test of difference among the groups $H_0 : \mu_c = \mu_o$

$$H_0 = A_{(S-1)\times S}B_{S\times t}C_{t\times 1} = D_{(S-1)\times 1}$$
$$A_{1\times 2}B_{2\times 6}C_{6\times 1} = D_{1\times 1}$$

1.Assume parallelism is reasonable

$H_0 : \mu_{11} + \mu_{12} + \mu_{13} = \mu_{31} + \mu_{32} + \mu_{33}; \mu_{21} + \mu_{22} + \mu_{23} = \mu_{31} + \mu_{32} + \mu_{33}$

$$A_{1\times 2} = [I, -1] = [1, -1]$$
$$B_{2\times 6} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} & \mu_{16} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & \mu_{25} & \mu_{26} \end{bmatrix}$$
$$C_{6\times 1} = j_6 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$
$$D_{1\times 1} = 0$$

Therefore,

$$A \times B \times C = D$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} & \mu_{16} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & \mu_{25} & \mu_{26} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

2.Without assuming parallelism

$H_0 : \mu_{11} = \mu_{31}; \mu_{21} = \mu_{31}...$

$$C_{6\times6} = I_6 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$A_{1\times2} = [I, -1] = [1, -1]$$

$$D_{1\times6} = 0_{1\times6}$$

Therefore,

$$A \times B \times C = D$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} & \mu_{16} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & \mu_{25} & \mu_{26} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) To test whether the profiles in the two groups are parallel. $H_0 : ABC = D$

$H_0 : \mu_{12} - \mu_{11} = \mu_{22} - \mu_{21}; \mu_{22} - \mu_{21} = \mu_{32} - \mu_{31}...$

$$A_{(S-1)\times S} = A_{1\times2} = [I_1, -1_1] = [1, -1]$$

$$B_{S\times t} = B_{2\times6}$$

$$C_{t\times(t-1)} = C_{6\times5} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

$$D_{(s-1)\times(t-1)} = 0_{1\times5}$$

Therefore,

$$A \times B \times C = D$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} & \mu_{16} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & \mu_{25} & \mu_{26} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(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. This is equivalent to test for parallelism between control and treatment group at time 2 and 3. Means at 2 and 3 hours correspond to μ_{15}, μ_{16} for the control group, and μ_{25}, μ_{26} for the treatment group.

$H_0 : ABC = D$

$H_0 : \mu_{16} - \mu_{15} = \mu_{26} - \mu_{25}$

$$A_{(S-1)\times S} = A_{1\times 2} = [I_1, -1_1] = [1, -1]$$
$$B_{S\times t} = B_{2\times 2} = \begin{bmatrix} \mu_{15} & \mu_{16} \\ \mu_{25} & \mu_{26} \end{bmatrix}$$
$$C_{t\times(t-1)} = C_{2\times 1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$D_{(s-1)\times(t-1)} = 0_{1\times 1} = 0$$

Therefore,

$$A \times B \times C = D$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{15} & \mu_{16} \\ \mu_{25} & \mu_{26} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0$$

Part B

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

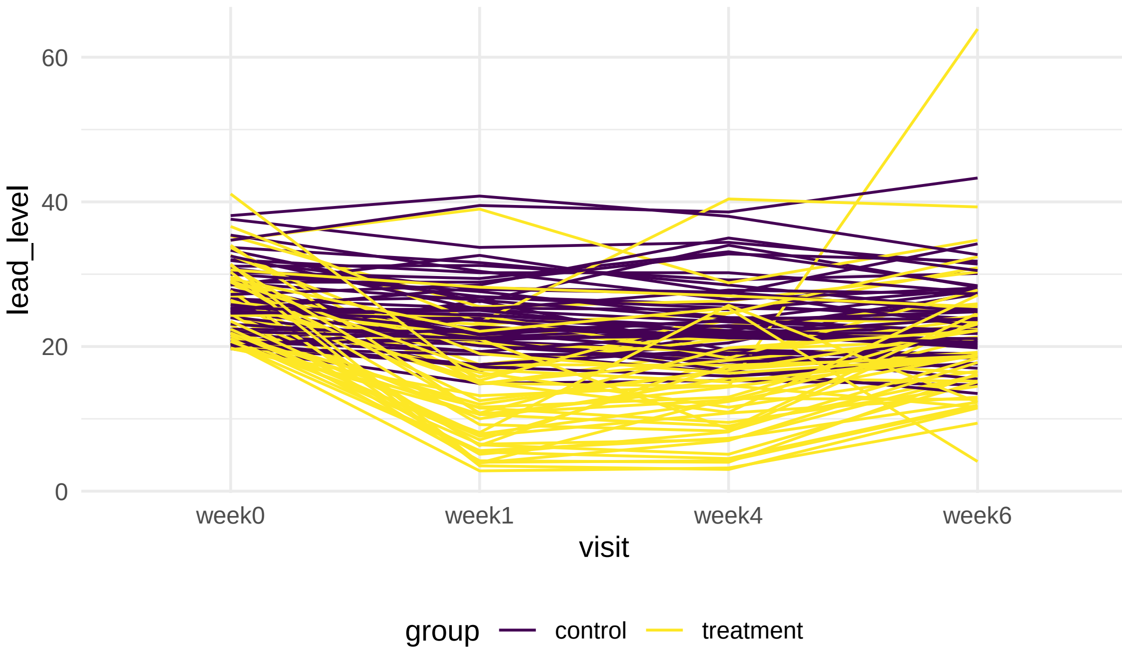
```
library(dplyr)
library(ggplot2)
knitr::opts_chunk$set(
  fig.width = 6,
  fig.asp = .6,
  out.width = "90%"
)
```

```r
theme_set(theme_minimal() + theme(legend.position = "bottom"))

options(
  ggplot2.continuous.colour = "viridis",
  ggplot2.continuous.fill = "viridis"
)

scale_colour_discrete = scale_colour_viridis_d
scale_fill_discrete = scale_fill_viridis_d
```

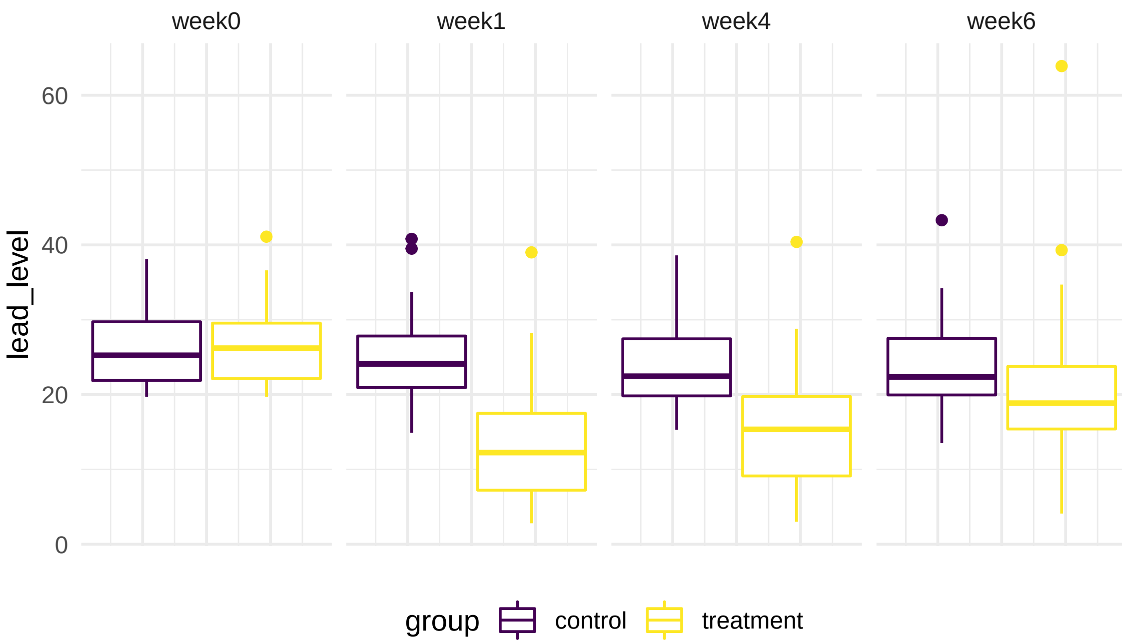
1. Treatment of Lead-Exposed Children Exploratory Data Analysis (EDA), to get insights into the data to eventually perform an appropriate longitudinal data analysis.

```r
tlc_df = read.table("TLC.dat") %>%
  rename("id" = V1,"group" = V2,"week0" = V3,
         "week1" = V4,"week4" = V5,"week6" = V6) %>%
  mutate(group = recode(group, "A" = "treatment", "P" = "control")) %>%
  pivot_longer(
    week0:week6,
    names_to = "visit",
    values_to = "lead_level")
```

```r
tlc_df  %>%
  ggplot(aes(x = visit, y = lead_level, group = id, color = group)) + geom_line()
```




```r
tlc_df  %>%
  ggplot(aes(y = lead_level, color = group))+
          geom_boxplot()+
          facet_grid(~visit)+
          theme(axis.text.x=element_blank(),
          axis.ticks.x=element_blank())
```



From the spaghetti plot and the side-by-side boxplot, we can see that at baseline week 0, the control and treatment group had similar blood lead level. On week 1, the treatment group had a drop in their blood lead level while as that of the control group remained similar to baseline. However, as time progress to week 4 and 6, the treatment effect seem to diminish as the treatment group experiences a bounce back in their blood lead level such that the lead level increases in week 4 and continue to increase in week 6.

```
library(Hotelling)
```

2. Plasma inorganic phosphate measurements

```
## Loading required package: corpcor

##
## Attaching package: 'Hotelling'

## The following object is masked from 'package:dplyr':
##
##     summarise
```

a: group means are the same at all six measurement times

$H_0 : \mu_c = \mu_o$

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose
```

```
zerb = fread("ZERBE2.DAT") %>%
  rename("group" = V1, "id" = V2, "t0" = V3, "t0.5" = V4,
         "t1"=V5, "t1.5" = V6,"t2" = V7,"t3" = V8)

zerb_h1 = zerb[,-2]

test_a= hotelling.test(.~group, data = zerb_h1)

test_a
```

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

The p-value is 0.000034495, which is less than 0.05, therefore, we can reject the null hypothesis at a significant level of 0.05 and conclude that the group means are not the same at all 6 measurement times.

b: Test for parallelism

H_0: $\mu_{12} - \mu_{11} = \mu_{22} - \mu_{21}; \mu_{22} - \mu_{21} = \mu_{32} - \mu_{31}; \ldots$

```
t_zerb = t(zerb[,-c(1:2)])
cmat_b = 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)


zerb_h2a = cmat_b %*% t_zerb
zerb_h2b = t(zerb_h2a)
zerb_h2 = cbind(zerb[,1], zerb_h2b)
test2 = hotelling.test(.~group, data = zerb_h2)
test2
```

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

The p-value is 0.00008344, which is less than 0.05. Therefore, we reject the null hypothesis and conclude that the two groups are not parallel.

c: Test parallelism at hour 2 and 3

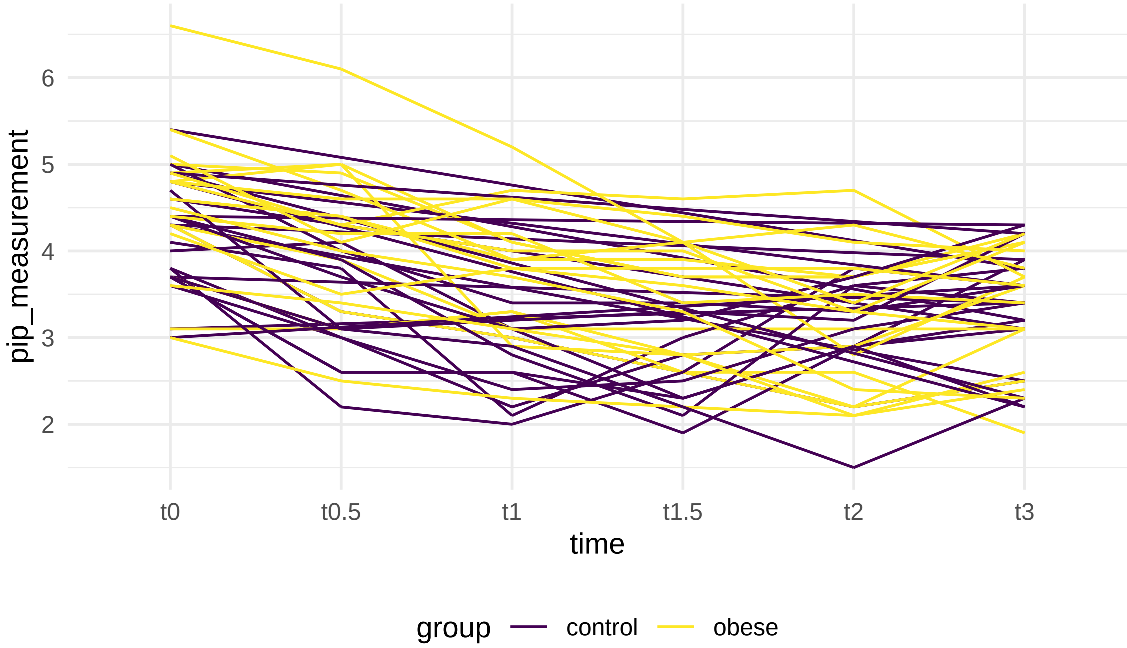
```
test_c = hotelling.test(V5~group, data = zerb_h2)
test_c
```

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

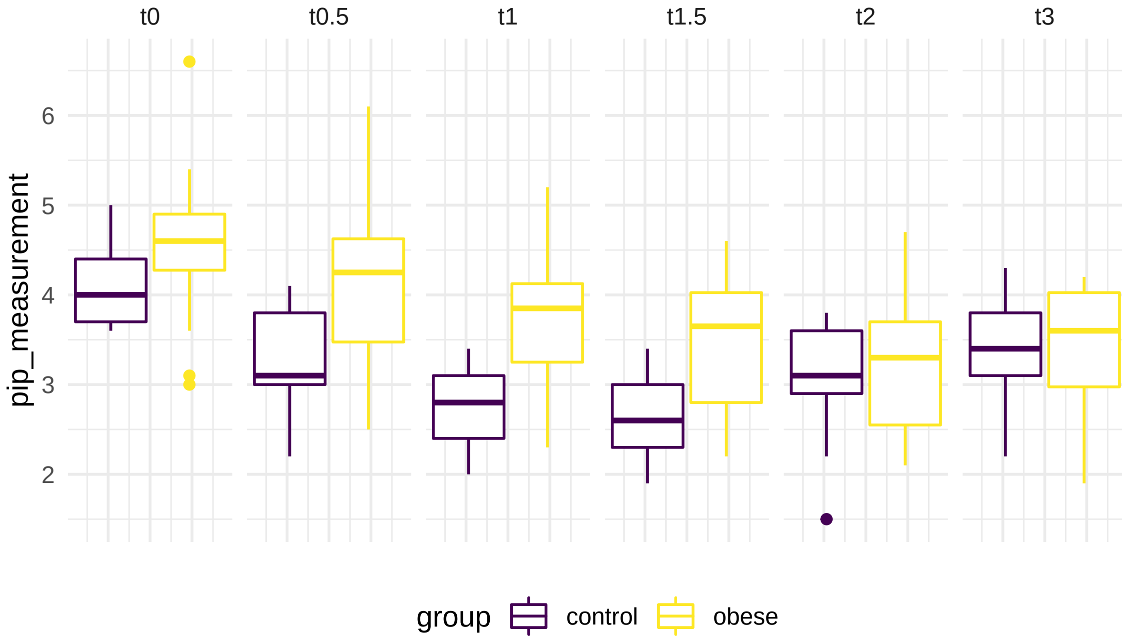
The p-value is 0.5231, which is greater than 0.05. Therefore, we fail to reject the null hypothesis and conclude that the difference in means at 2 and 3 hours after an oral glucose challenge are not different between the control and obese patients.

EDA

```
zerb_plot = zerb%>%
  pivot_longer(
    "t0":"t3",
    names_to = "time",
    values_to = "pip_measurement") %>%
  mutate(group = recode(group, "1" = "control", "2" = "obese")) %>% mutate(group = factor(group)) %>% di


zerb_plot  %>%
  ggplot(aes(x = time, y = pip_measurement, group = id)) + geom_path(aes(color = group))
```




```
zerb_plot  %>%
  ggplot(aes(y = pip_measurement, color = group))+
          geom_boxplot()+
          facet_grid(~time)+
          theme(axis.text.x=element_blank(),
          axis.ticks.x=element_blank())
```



From the spaghetti plot and the side by side boxplot, we can see that both the control and obese group has an overall decreasing plasma inorganic phosphate measurements. However, the rate of decrease occurred faster in the control group, where as the measurement in the obese group changes only slightly between measurement time. Moreover, the phosphate level bounces back in the control group, where it start to increase after the 1.5 hour; and the phosphate level for the obese group starts to increase after the 2 hour as well.