

# P8157 Homework 1 yz4184

Yunlin Zhou

2022-09-27

```
tlc = read.table("./TLC.dat") %>%
  janitor::clean_names() %>%
  rename(id = v1,
         assignment = v2,
         week0 = v3,
         week1 = v4,
         week4 = v5,
         week6 = v6)%>%
  mutate(assignment = as.factor(assignment))%>%
  as.data.frame()

zerbe2 = read.table("./ZERBE2.DAT") %>%
  janitor::clean_names() %>%
  rename(assignment = v1,
         id = v2,
         "0h" = v3,
         "0.5h" = v4,
         "1h" = v5,
         "1.5h" = v6,
         "2h" = v7,
         "3h" = v8)
```

## Part A

1

(a)  $H_0: ABC = D = 0$

```
matrix(c(1, -1), nrow = 1, ncol = 2, byrow = TRUE)
```

**Matrix A**

```
##      [,1] [,2]
## [1,]    1  -1
```

**Matrix B** Matrix B is a 2\*6 matrix consisting of group means at each time point.

```
diag(6)
```

Matrix C

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

(b)  $H_0: ABC = D = 0$

```
matrix(c(1, -1), nrow = 1, ncol = 2, byrow = TRUE)
```

Matrix A

```
##      [,1] [,2]
## [1,]    1   -1
```

**Matrix B** Matrix B is a 2\*6 matrix consisting of group means at each time point.

```
matrix(c(-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), nrow = 6, ncol = 5, byrow = TRUE)
```

Matrix C

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

(c)  $H_0: ABC = D = 0$

```
matrix(c(1, -1), nrow = 1, ncol = 2, byrow = TRUE)
```

**Matrix A**

```
##      [,1] [,2]  
## [1,]    1  -1
```

**Matrix B** Matrix B is a 2\*2 matrix consisting of group means at 2 and 3 hours time point.

```
matrix(c(-1, 1), nrow = 2, ncol = 1, byrow = TRUE)
```

**Matrix C**

```
##      [,1]  
## [1,]   -1  
## [2,]    1
```

## Part B

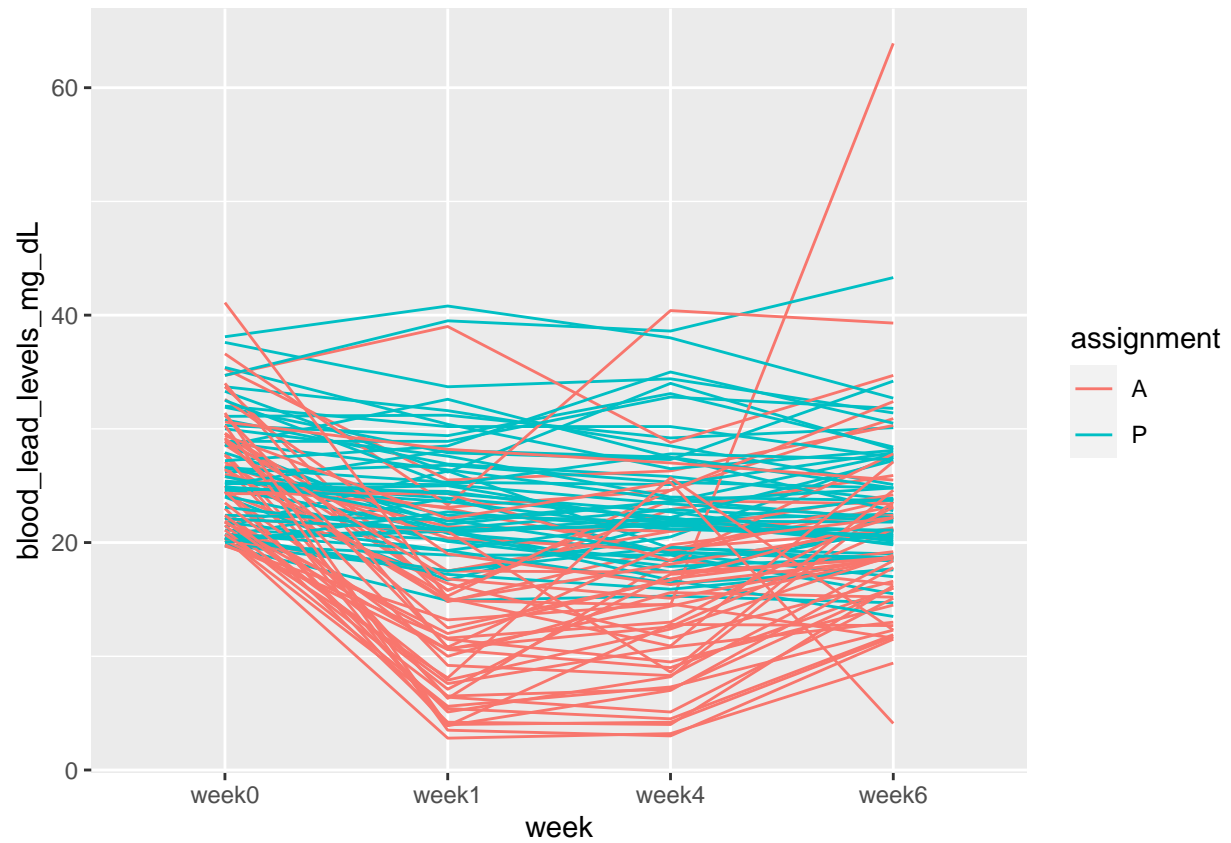
1

**EDA**

```
tlc_df = tlc %>%  
  pivot_longer(cols = week0:week6,  
               names_to = "week",  
               values_to = "blood_lead_levels_mg_dL") %>%  
  mutate(week = as.factor(week))
```

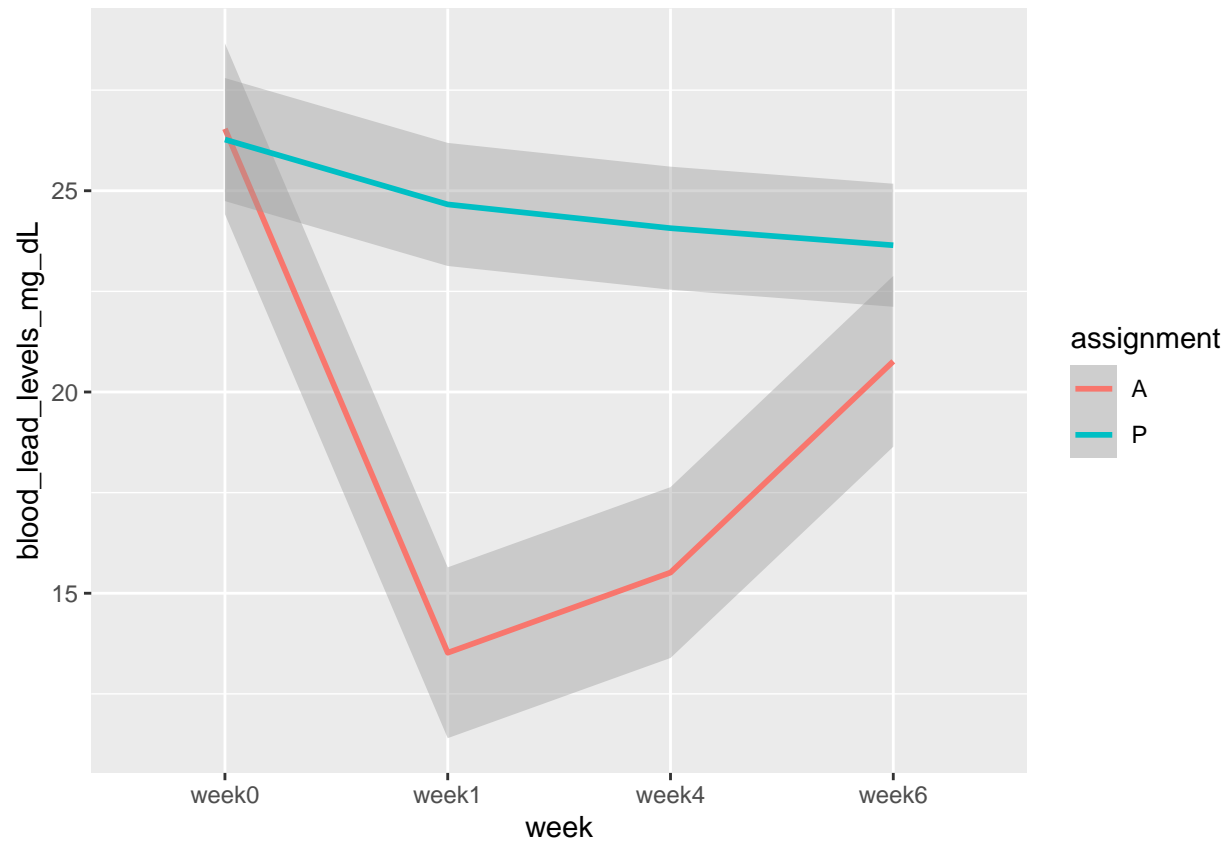
```
tlc_p1 = ggplot(data = tlc_df, aes(x = week,  
                                   y = blood_lead_levels_mg_dL,  
                                   col = assignment,  
                                   group = id)) + geom_line()
```

```
tlc_p1
```



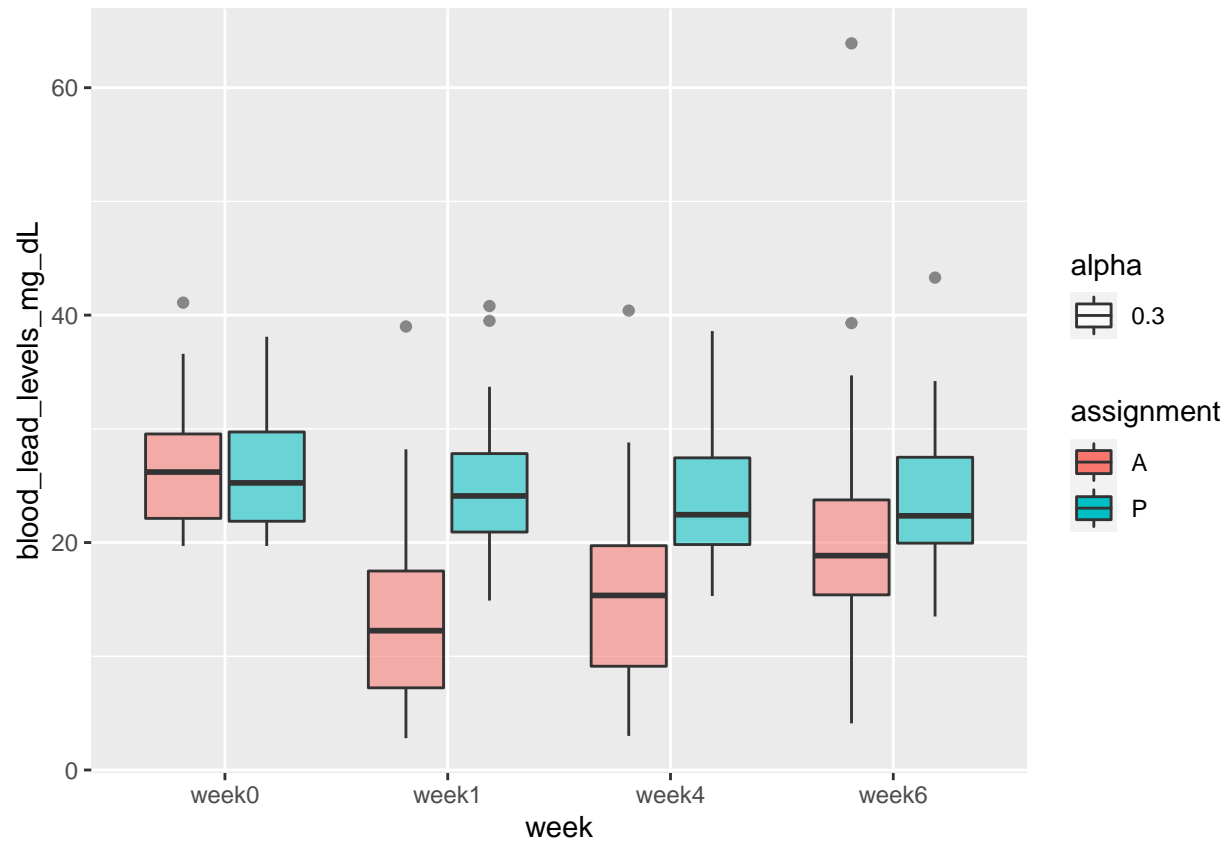
```
tlc_p2 = ggplot(data = tlc_df, aes(x = week,  
  y = blood_lead_levels_mg_dL,  
  col = assignment,  
  group = assignment)) + geom_smooth()
```

```
tlc_p2
```



```
tlc_p3 = ggplot(data = tlc_df, aes(x = week,  
  y = blood_lead_levels_mg_dL,  
  fill = assignment,  
  alpha = 0.3))+geom_boxplot()
```

```
tlc_p3
```



### Hotelling's T2 test

```
tlc_test1 = hotelling.test(~assignment, data = tlc[, -1])
tlc_test1
```

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

```
tlc2 = t(tlc[, -c(1:2)])
cmat = matrix(c(-1, 1, 0, 0, 0, -1, 1, 0, 0, 0, -1, 1), ncol = 4, byrow = TRUE)
tlc3 = cmat %*% tlc2
tlc3 = t(tlc3)
tlc3 = cbind(tlc[, 2], tlc3) %>%
  data.frame()
tlc_test2 = hotelling.test(~X1, data = tlc3)
tlc_test2
```

```
## Test stat: 107.79
## Numerator df: 3
## Denominator df: 96
## P-value: 1.998e-15
```

## 2

### Hotelling's T2 test

```
z_test1 <- hotelling.test(~assignment, data = zerbe2[, -2])
z_test1
```

(a)

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

```
zerb_1 = t(zerbe2[, -c(1:2)])
cmat_z = matrix(c(-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), nrow = 6, ncol = 5, byrow = TRUE)%>%
  t()

zerb_2 = cmat_z %*% zerb_1
zerb_2 = t(zerb_2)
zerb_2 = cbind(zerbe2[, 1], zerb_2)%>%
  data.frame()
test2 = hotelling.test(~X1, data = zerb_2)
test2
```

(b)

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

```
zerb_3 = t(zerbe2[, -c(1:6)])
cmat_z2 <- matrix(c(-1, 1), nrow = 2, ncol = 1, byrow = TRUE) %>% t()

zerb_4 = cmat_z2 %*% zerb_3
zerb_4 = t(zerb_4)
zerb_4 = cbind(zerbe2[, 1], zerb_4)%>%
  data.frame()
```

```
test3 = hotelling.test(~X1, data = zerb_4)
test3
```

(c)

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

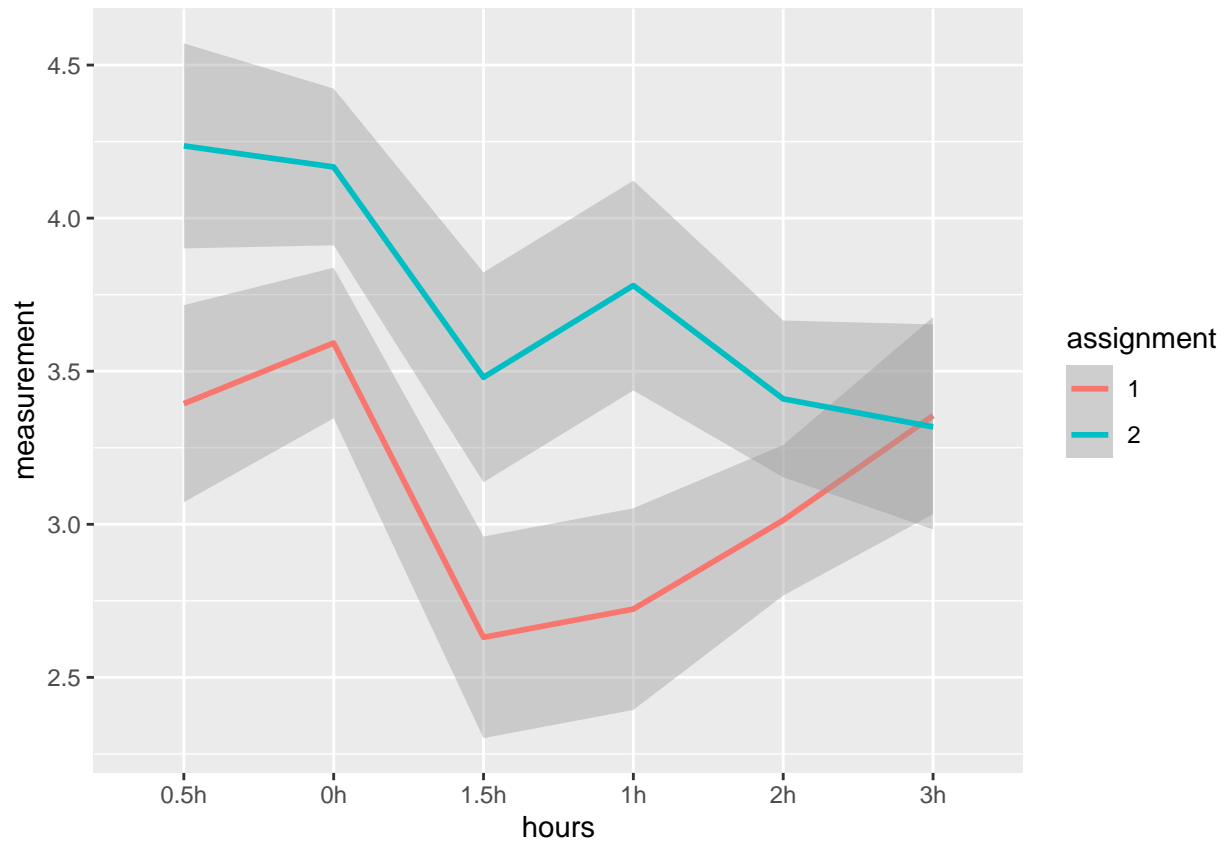
## EDA

```
z_df = zerbe2%>%
  pivot_longer(
    cols = "0h":"3h",
    names_to = "hours",
    values_to = "measurement"
  ) %>%
  mutate(hours = as.factor(hours),
         assignment = as.factor(assignment),
         id = as.factor(id))%>%
  as.data.frame()
```

```
z_p1 = ggplot(data = z_df, aes(x = hours,
                               y = measurement,
                               col = assignment,
                               group = assignment)) + geom_smooth()

z_p1
```

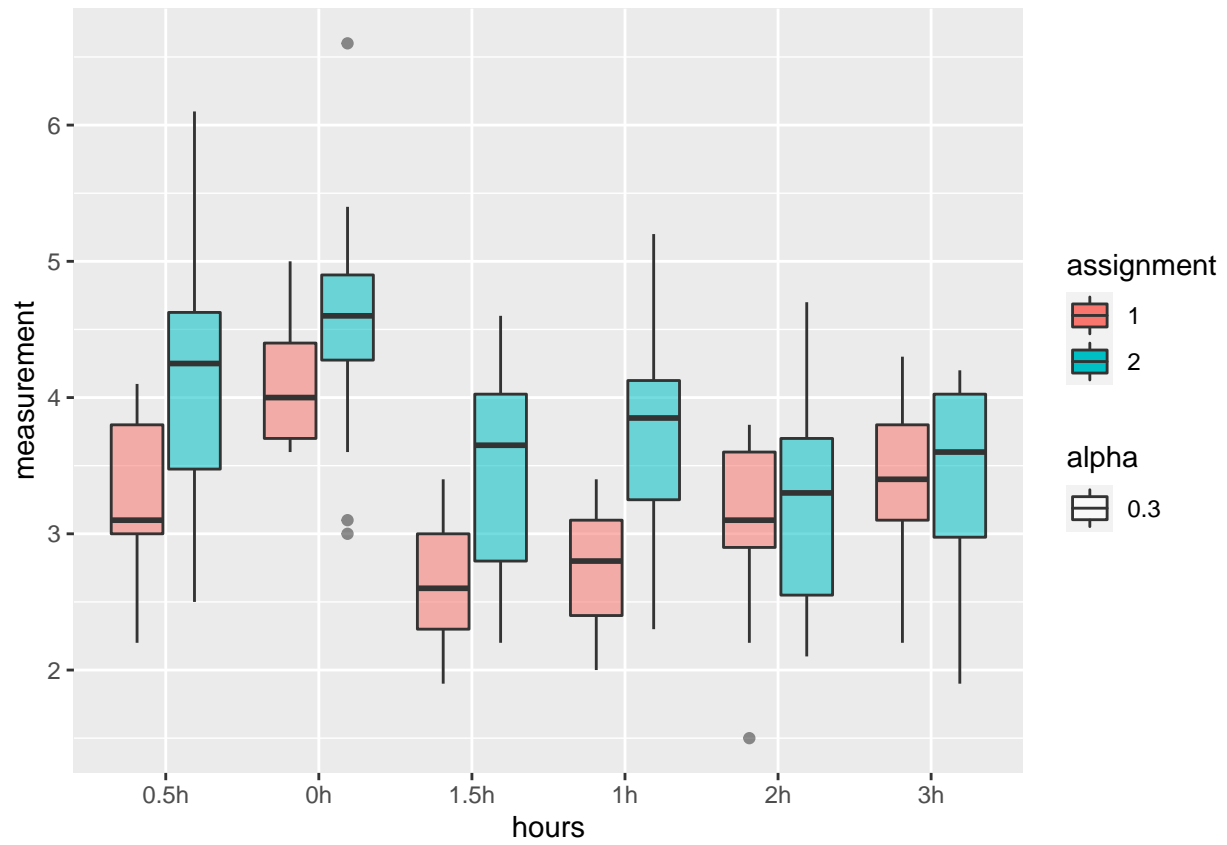




We can conclude that the 2 groups are not parallel.

```
z_p2 = ggplot(data = z_df, aes(x = hours,
                                y = measurement,
                                fill = assignment,
                                alpha = 0.3)) + geom_boxplot()

z_p2
```



We can conclude that in control group, Plasma inorganic phosphate measurements are lower all the time.