

# p8157\_hw1\_jsg2145

Jared Garfinkel

9/25/2020

## Part A

### Problem 1

B is a 2 x 6 matrix of group means.

#### Part a

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

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

```
C = diag(6)
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
```

#### Part b

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

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

```
C = 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), ncol = 5, byrow = TRUE)
```

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
```

## Part c

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

A

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

```
C = matrix(c(0,0,0,0,1,-1))
```

C

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

## Part B

### Problem 1

Exploratory Data Analysis of blood lead levels.

```
dat = read.table("./data/TLC.dat") %>%
  janitor::clean_names() %>%
  rename("id" = v1,
         "group" = v2,
         "week0" = v3,
         "week1" = v4,
         "week4" = v5,
```

```

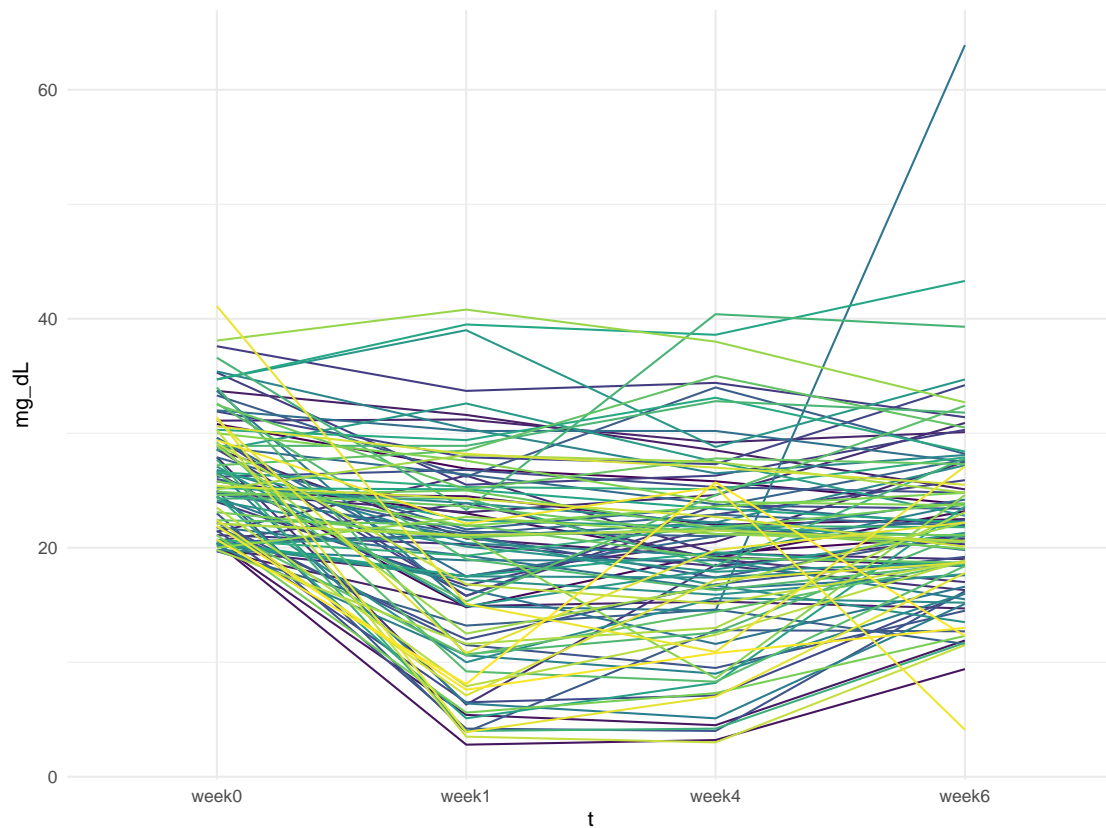
    "week6" = v6) %>%
mutate(group = recode(group,
    "P" = 0,
    "A" = 1))

```

```

dat %>%
  pivot_longer(cols = week0:week6, names_to = "t", values_to = "mg_dL") %>%
  ggplot(aes(x = t, y = mg_dL, group = id)) +
  geom_path(aes(color = id)) +
  theme(legend.position = "none")

```

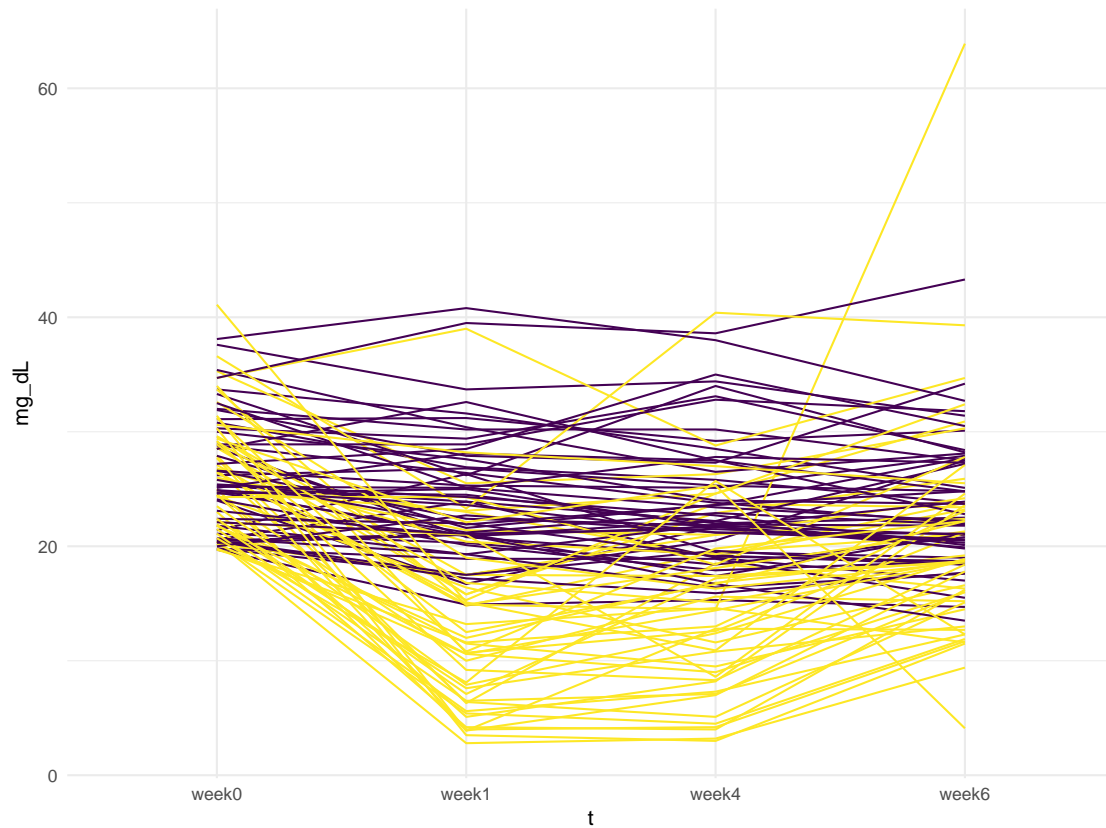


It appears that those in one group have a response to treatment after the first week, but that the effects begin to wear off by week 6.

```

dat %>%
  pivot_longer(cols = week0:week6, names_to = "t", values_to = "mg_dL") %>%
  ggplot(aes(x = t, y = mg_dL, group = id)) +
  geom_path(aes(color = group)) +
  theme(legend.position = "none")

```



## Problem 2

```
df = read.table("./data/ZERBE2.DAT") %>%
  janitor::clean_names() %>%
  rename("group" = v1,
         "id" = v2,
         "t0" = v3,
         "t0.5" = v4,
         "t1" = v5,
         "t1.5" = v6,
         "t2" = v7,
         "t3" = v8)
```

### Part a

Test the group means.

```
test.ph = hotelling.test(~group, data = df[, -2])
test.ph
```

```
## Test stat: 8.5531
## Numerator df: 6
```

```
## Denominator df: 26
## P-value: 0.00003495
```

Since the p-value is 0.000035 we can reject the null. There is evidence of a difference of means among groups.

## Part b

Test the parallelism.

```
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

df2 = t(df[, -c(1:2)])
df2
# 6 x 33

df3 = cmat %*% df2
df3 = t(df3)
df3 = cbind(df[, 1], df3) %>%
  data.frame()
```

```
test.contrast2 = hotelling.test(.-X1, data = df3)
test.contrast2
```

```
## Test stat: 8.1805
## Numerator df: 5
## Denominator df: 27
## P-value: 0.00008344
```

Since the p-value is 0.0000834 we may reject the null indicating that there is a difference in the slopes between time points among groups.

## Part c

Test the difference in group means between the last two time points.

```
df_final = df %>%
  select(t2, t3) %>%
  t()

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

df_final2 = cmat_final %*% df_final %>%
  t()

df_final3 = cbind(df[, 1], df_final2) %>%
  data.frame()
```

```
test.contrast3 = hotelling.test(.-X1, data = df_final3)
test.contrast3
```

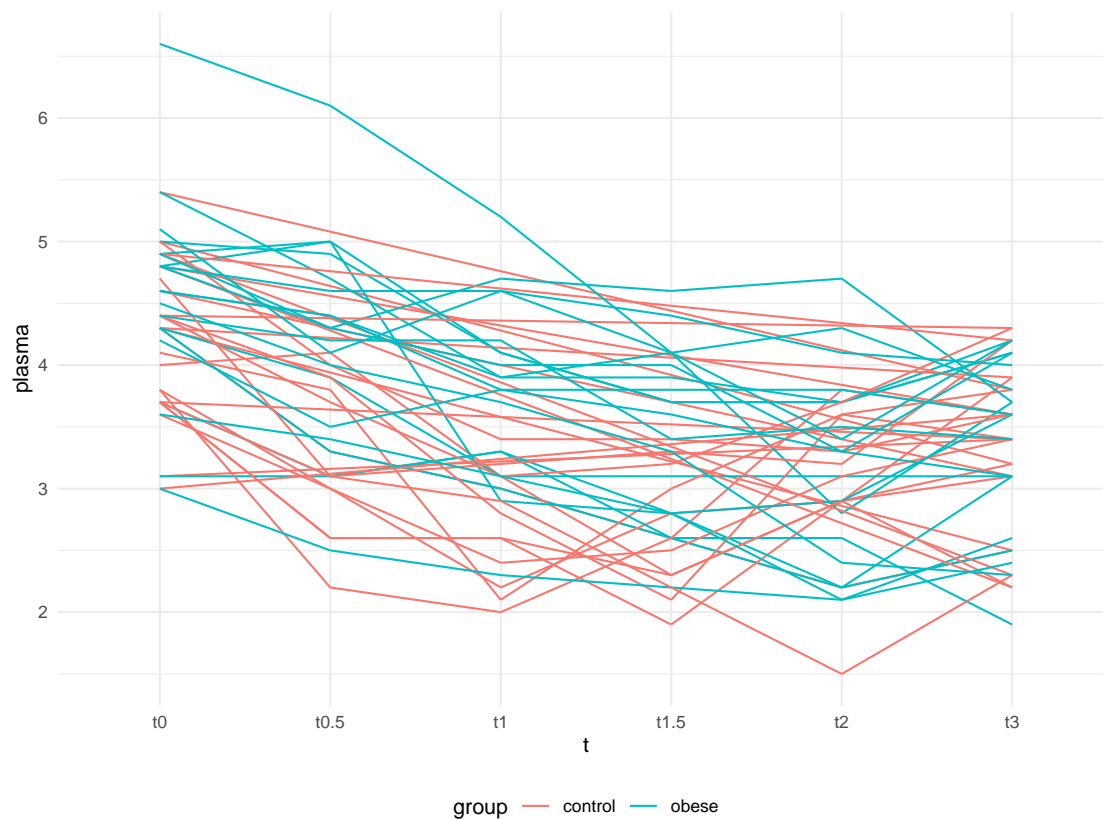
```
## Test stat: 0.41711
## Numerator df: 1
## Denominator df: 31
## P-value: 0.5231
```

The pvalue is 0.5231339, indicating we do not reject the null, and there is not evidence of a difference in the time points between groups for time  $t = 2$  and time  $t = 3$ .

## Exploratory Data Analysis

Exploratory data analysis does not appear to show much of a difference between groups at first.

```
df %>%
  mutate(group = factor(group, labels = c("control" = 1, "obese" = 2))) %>%
  pivot_longer(cols = t0:t3, names_to = "t", values_to = "plasma") %>%
  ggplot(aes(x = t, y = plasma, group = id)) +
  geom_path(aes(color = group)) +
  scale_color_discrete(labels = c("control", "obese"))
```



In a side-by-side plot, it appears that perhaps both groups decrease in plasma phosphate levels over time. The control group may have more a “U” shape, where the phosphate levels decrease and then increase. On

the other hand, it appears to take longer for those in the obese group to reach a valley in the phosphate levels.

As with the test for parallelism, one may test the slopes between individuals timepoints. Furthermore, it may be of interest to determine at which time point there is a difference in group means.

```
df %>%
  mutate(group = factor(group, labels = c("control" = 1, "obese" = 2))) %>%
  pivot_longer(cols = t0:t3, names_to = "t", values_to = "plasma") %>%
  ggplot(aes(x = t, y = plasma, group = id)) +
  geom_path(aes(color = group)) +
  scale_color_discrete(labels = c("control", "obese")) +
  facet_grid(.~group)
```

