HW1_answer

Guojing Wu / UNI: gw2383

9/23/2019

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 T^2 statistic. Set up suitable model clearly stating the assumptions. Suggest appropriate contrast matrices for the hypothesis tests proposed by the investigators.

Set up the model as: $Y = XB + E$, and we assume:

- $y_{s_i} \sim MVN(\mu_{s_i}, \Sigma)$
- each group has its own μ_{s_i}
- Σ is the same across groups

The null hypothesis has the general form as $ABC = D$, where

$$B = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} & \mu_{16} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & \mu_{25} & \mu_{26} \end{pmatrix}$$

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

$$A = I_2$$
$$C = \begin{pmatrix} I_5 \\ -1_5' \end{pmatrix}$$
$$D = \mathbf{0_{2\times 5}}$$

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

$$A = \begin{pmatrix} 1 & -1 \end{pmatrix}$$
$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$
$$D = \mathbf{0_{1\times 5}}$$

(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.

We need to test whether $\mu_{15} - \mu_{25} = \mu_{16} - \mu_{26}$

$$A = \begin{pmatrix} 1 & -1 \end{pmatrix}$$
$$C = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$
$$D = \begin{pmatrix} 0 \end{pmatrix}$$

Part B

1. Exposure to lead can produce cognitive impairment, especially among young children and infants. Interventions known as chelation treatments can help a child to excrete the lead that has been ingested. A chelating agent known as Succimer can be administered orally leading to urinary excretion of lead, unlike previous treatments which required injections and hospitalization. The Treatment of Lead- Exposed Children (TLC) trial was a placebo-controlled, randomized study of succimer (a chelating agent) in children with blood lead levels of 20-44 micrograms/dL. These data (TLC.dat) consist of subject id, assignment to chelation treatment with succimer or placebo and four repeated measurements of blood lead levels obtained at baseline (or week 0), week 1, week 4, and week 6 on 100 children who were randomly assigned to chelation treatment with succimer or placebo.

From Fig.1 to Fig.3, we can hardly tell if the blood lead level is decreasing through time in treatment group. But there exists difference between treatment and control group.

From Fig.4 and Fig.5, we can see that there is strong correlation between weeks close in time, and week effect differs between treament and placebo group.

Figure1. spaghetti plot for TLC data

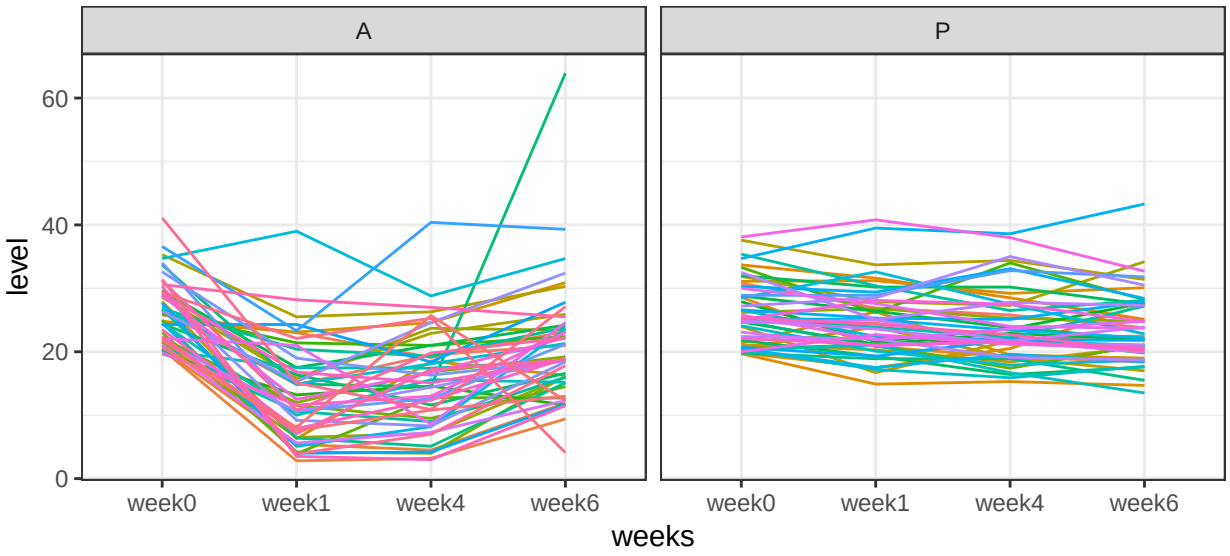


Figure2. boxplot for TLC data mean and s.d.

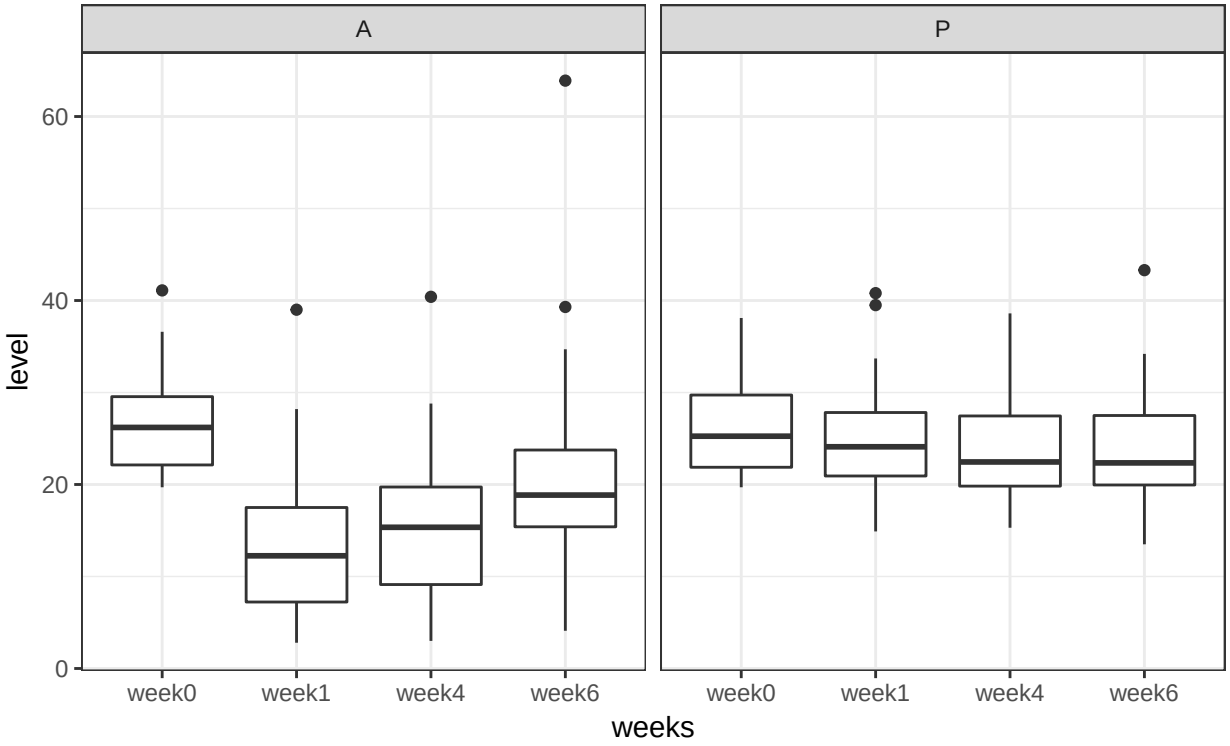


Figure3. spaghetti plot for standardized TLC data

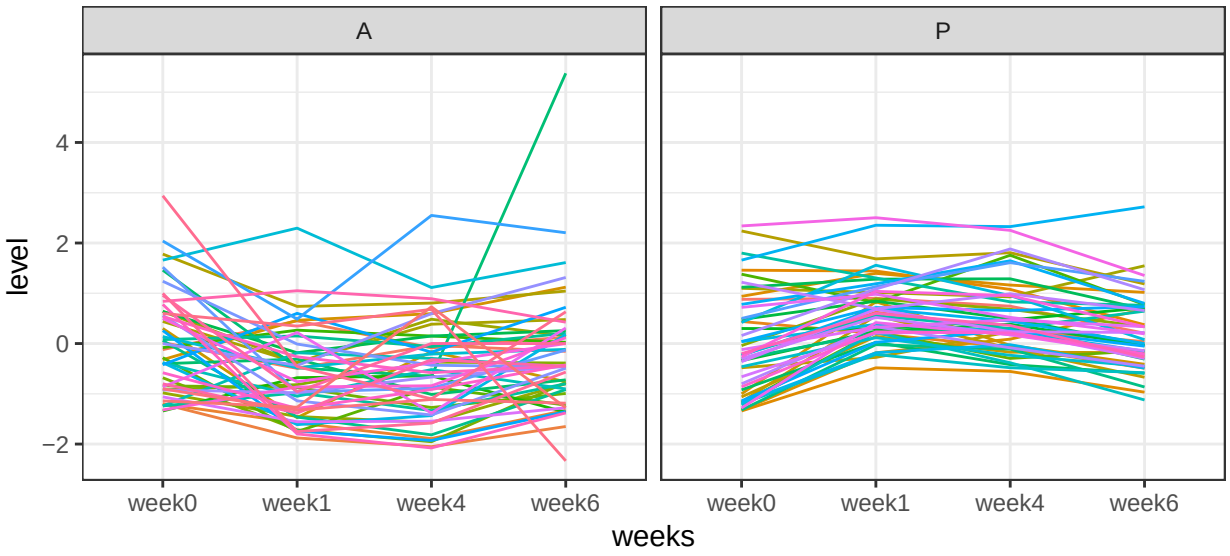


Figure4. median polish and ACF for TLC data

```
1: 1187.3
2: 1160
Final: 1159.9
```

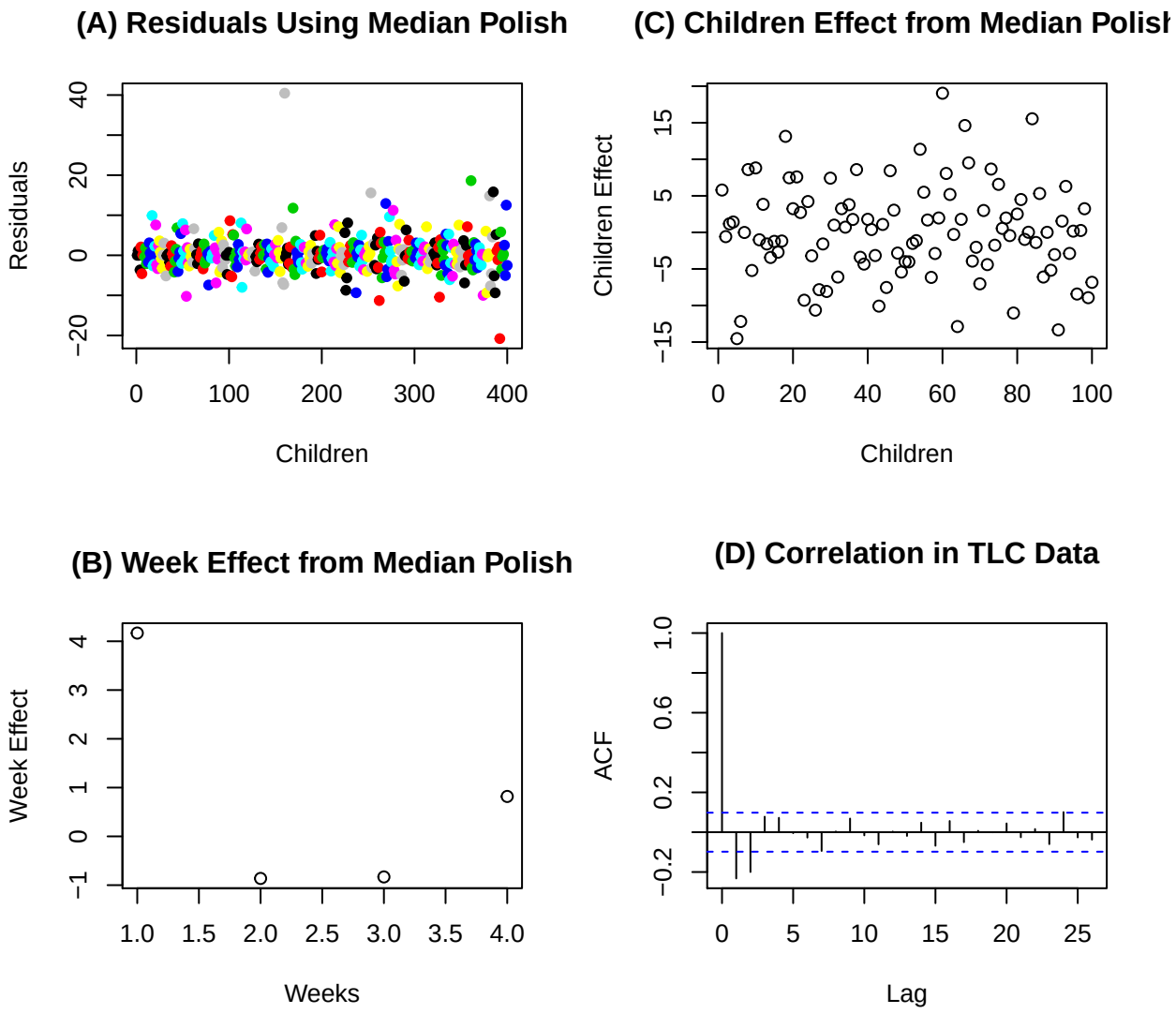
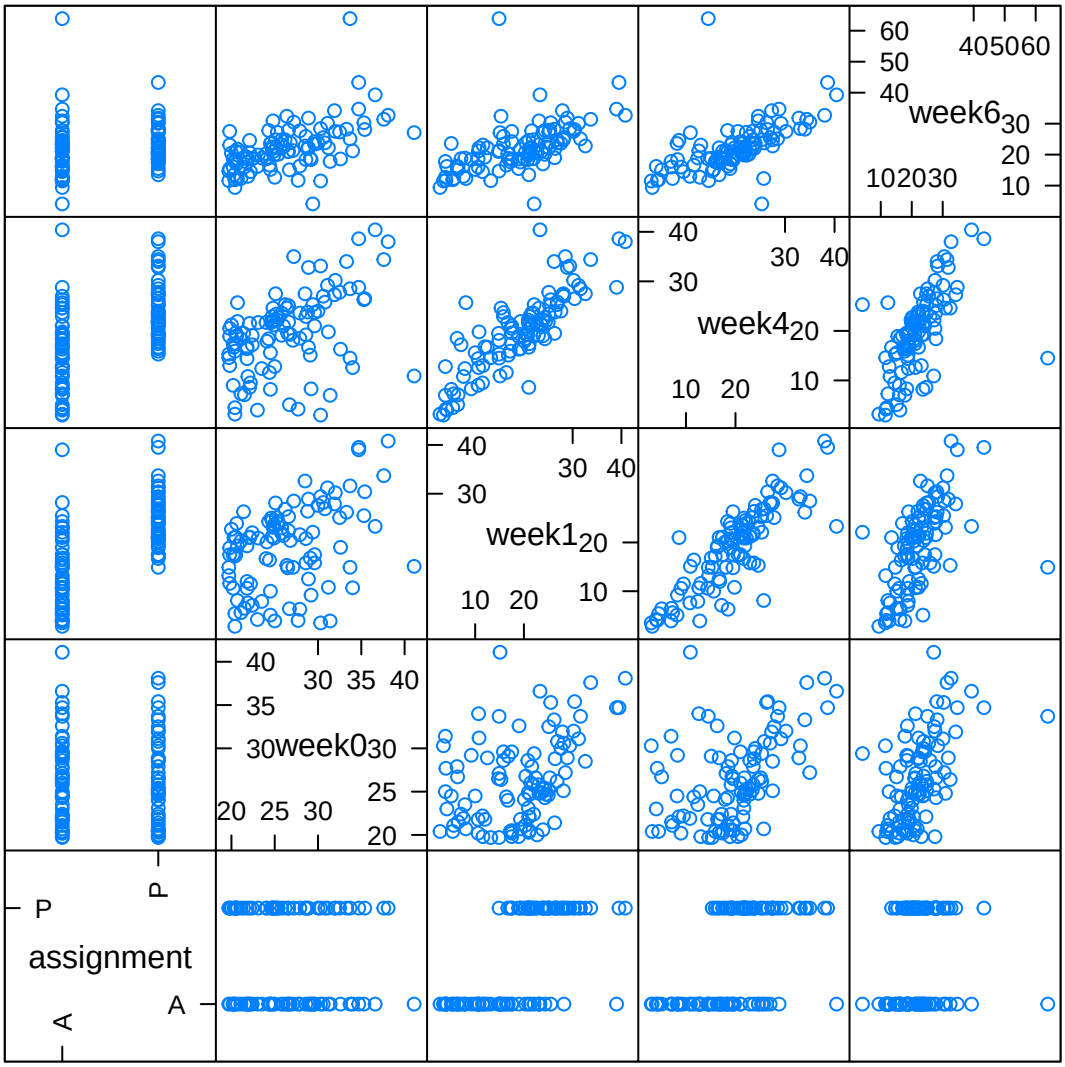


Figure5. SPLOM for TLC data



2. The data (dental.dat) are from a study of dental growth measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure were obtained on 11 girls and 16 boys at ages 8, 10, 12, and 14. The variables consist of ID, Gender, and responses at ages 8,10,12 and 14.

From Fig.6 to Fig.8, we can hardly tell if the distance is increasing through time. But there exists difference between male and femal.

From Fig.9 and Fig.10, we can see that there is strong correlation between weeks close in time, and distance increase as age effect increase.

Figure6. spaghetti plot for dental data

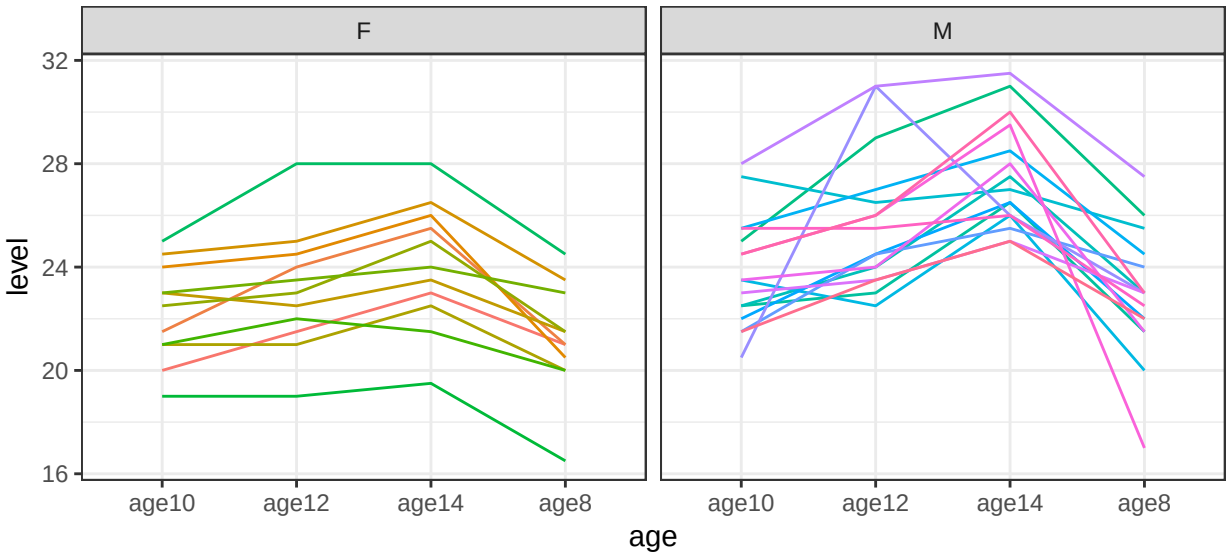


Figure7. boxplot for dental data mean and s.d.

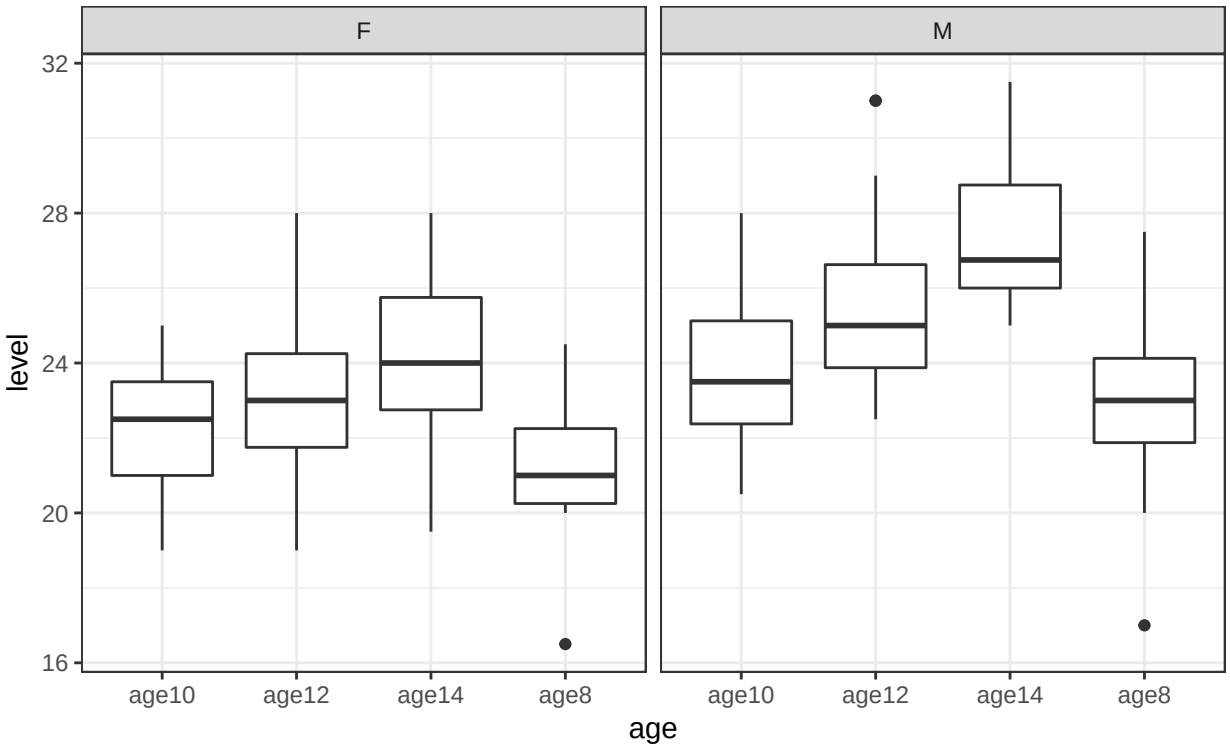


Figure8. spaghetti plot for standardized dental data

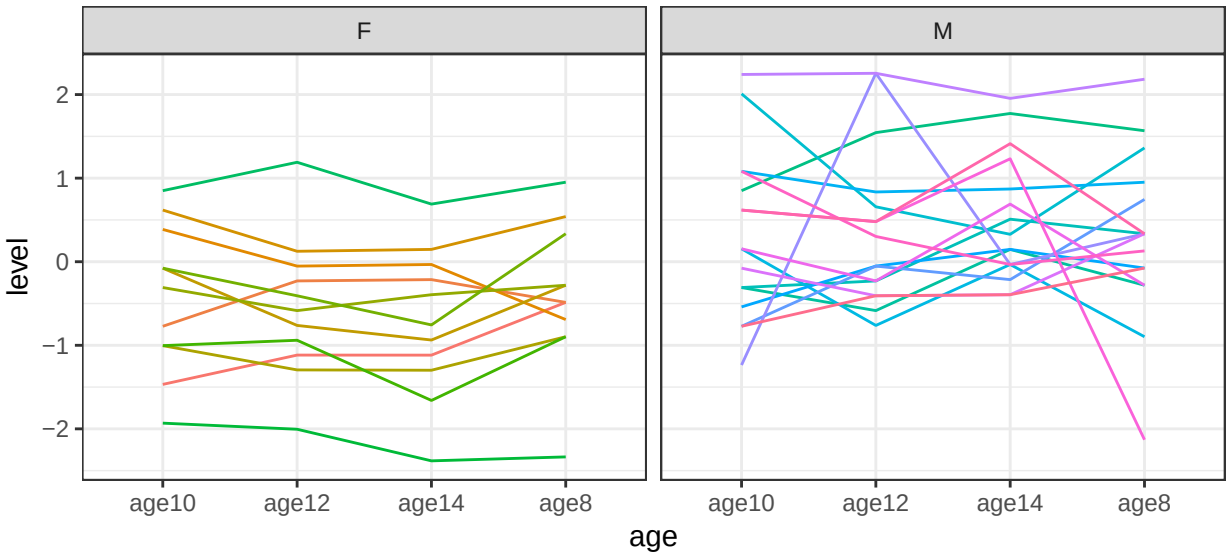


Figure9. median polish and ACF for dental data

```
1: 90.5
2: 87
Final: 87
```

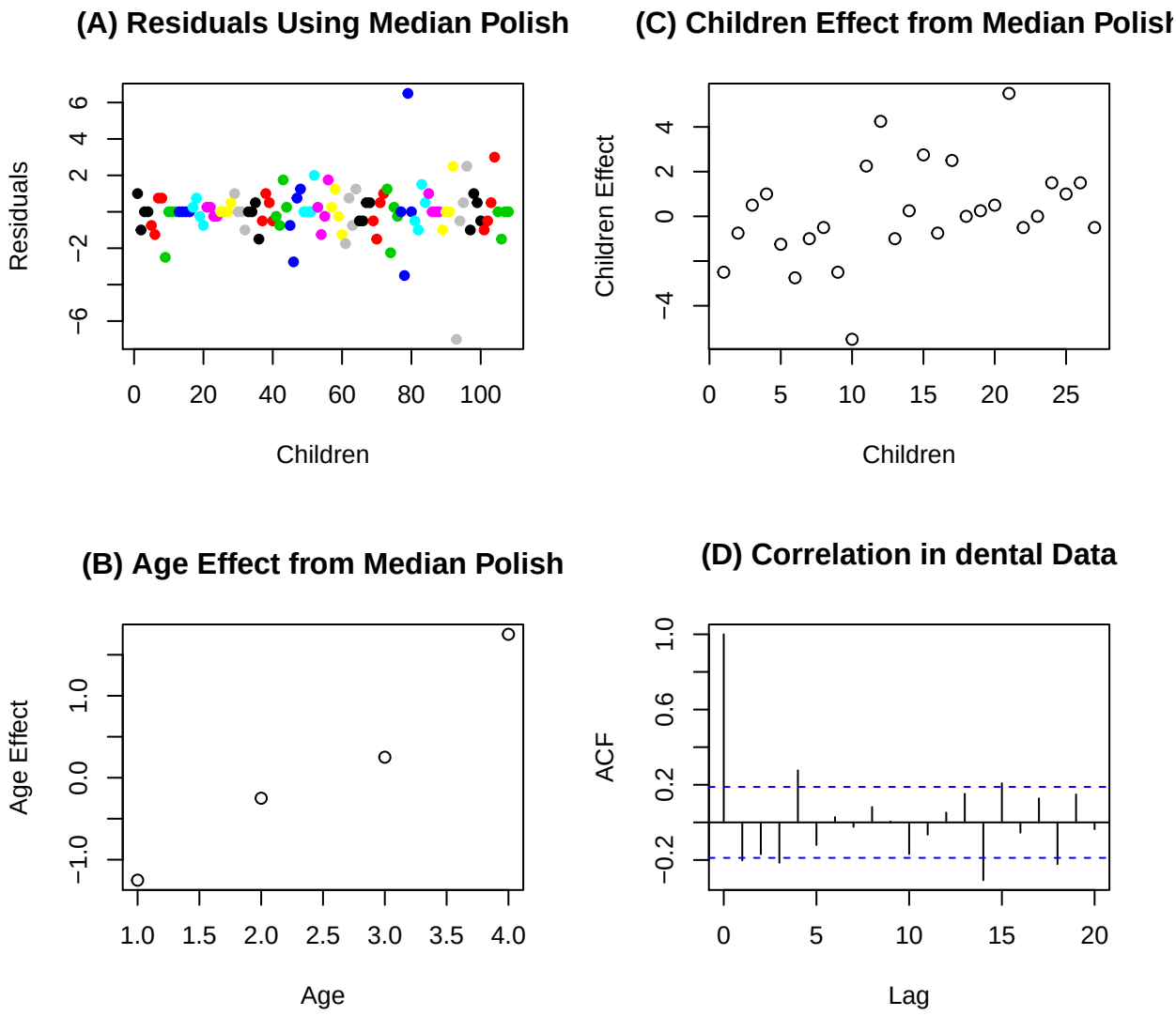
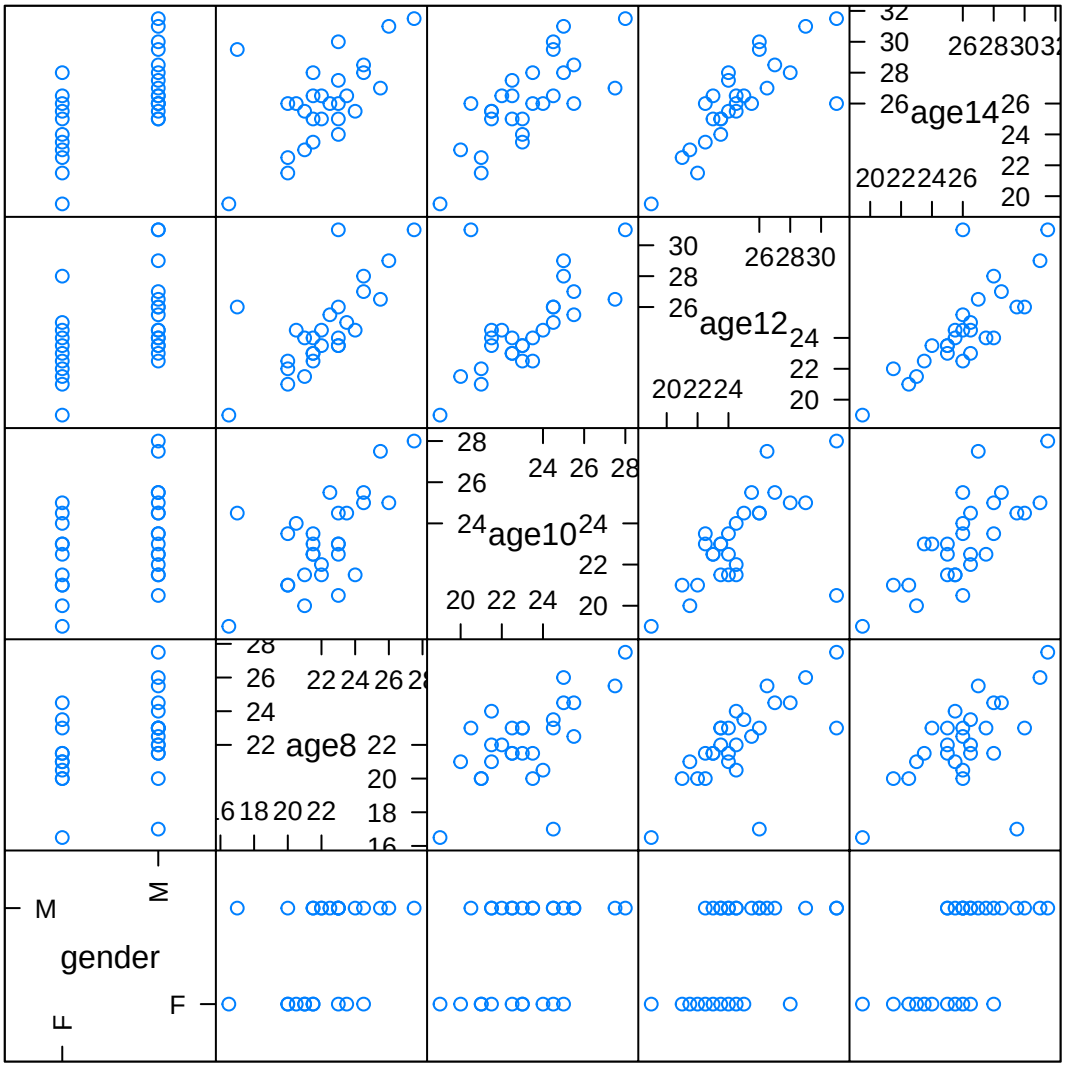


Figure10. SPLOM for dental data



3. The data for the Plasma inorganic phosphate measurements from 13 control and 20 obese patients describe in question 2 of part A are available in the file ZERBE2.dat. Carry out the Hotelling's T^2 test to test the hypotheses proposed above in part A. Perform EDA and provide any insights available from exploration. The data has the following information:

- Column 1 : Group (control=1,Obese=2)
- Column 2 : Subject id
- Columns 2-8 : Plasma inorganic phosphate measurements 0, 0.5, 1, 1.5, 2, and 3 hours after an oral glucose challenge

Figure11. boxplot for plasma data mean and s.d.

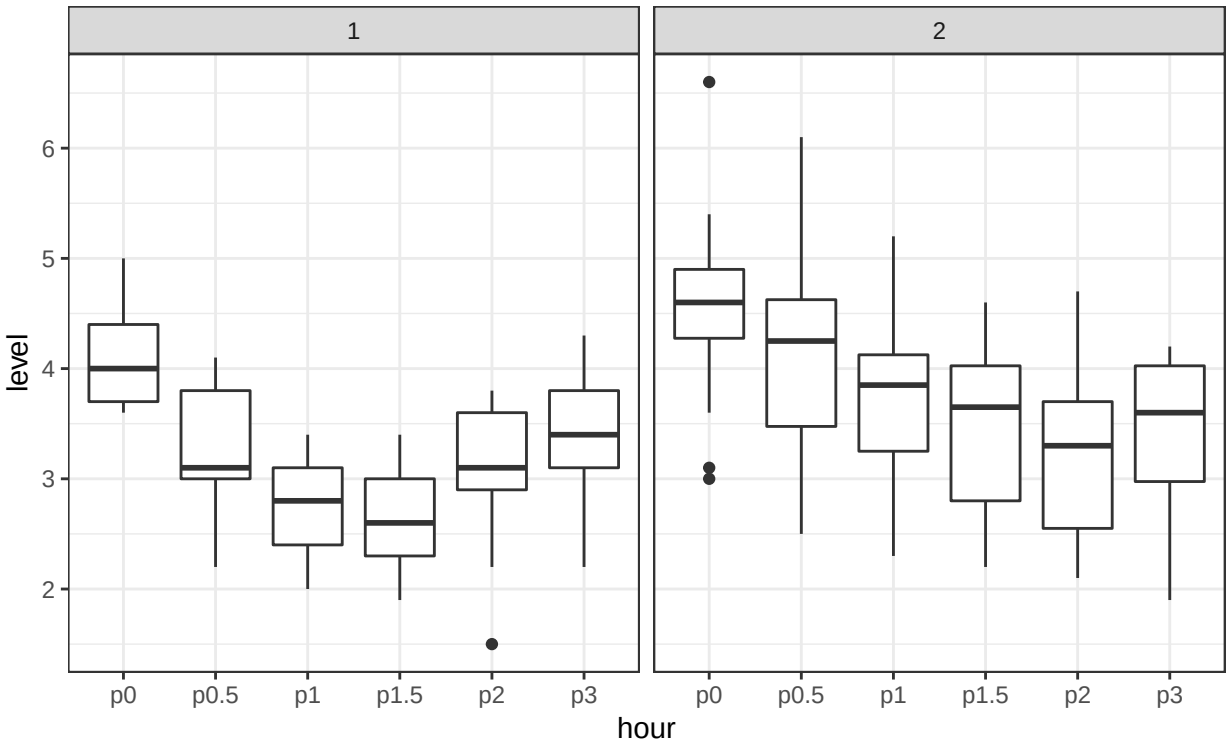


Figure12. spaghetti plot for standardized plasma data

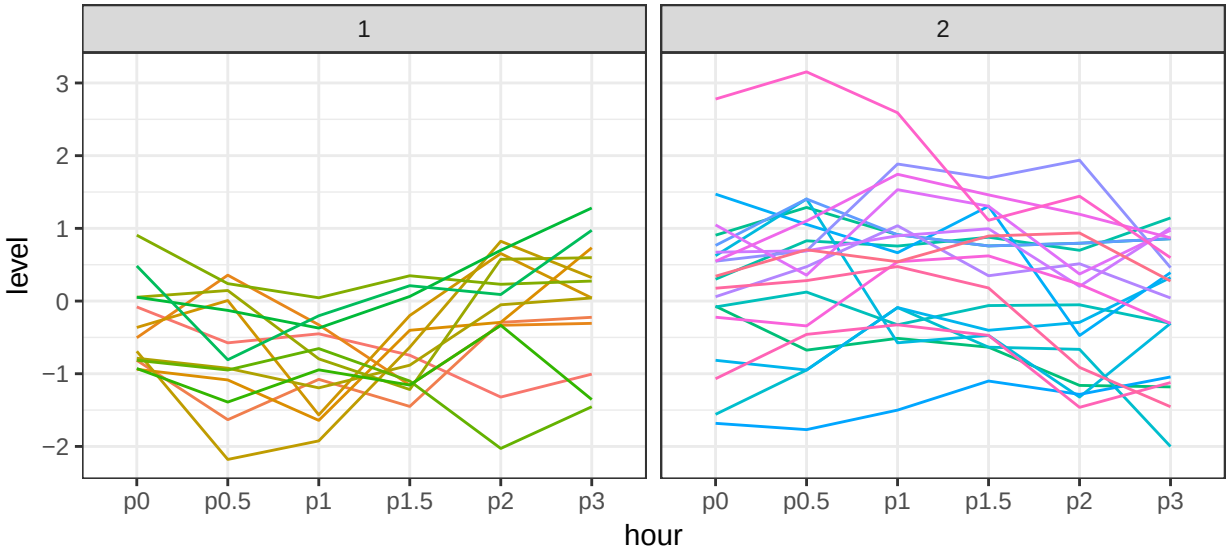


Figure13. median polish and ACF for plasma data

```
1: 64
2: 60.5
Final: 60.05
```

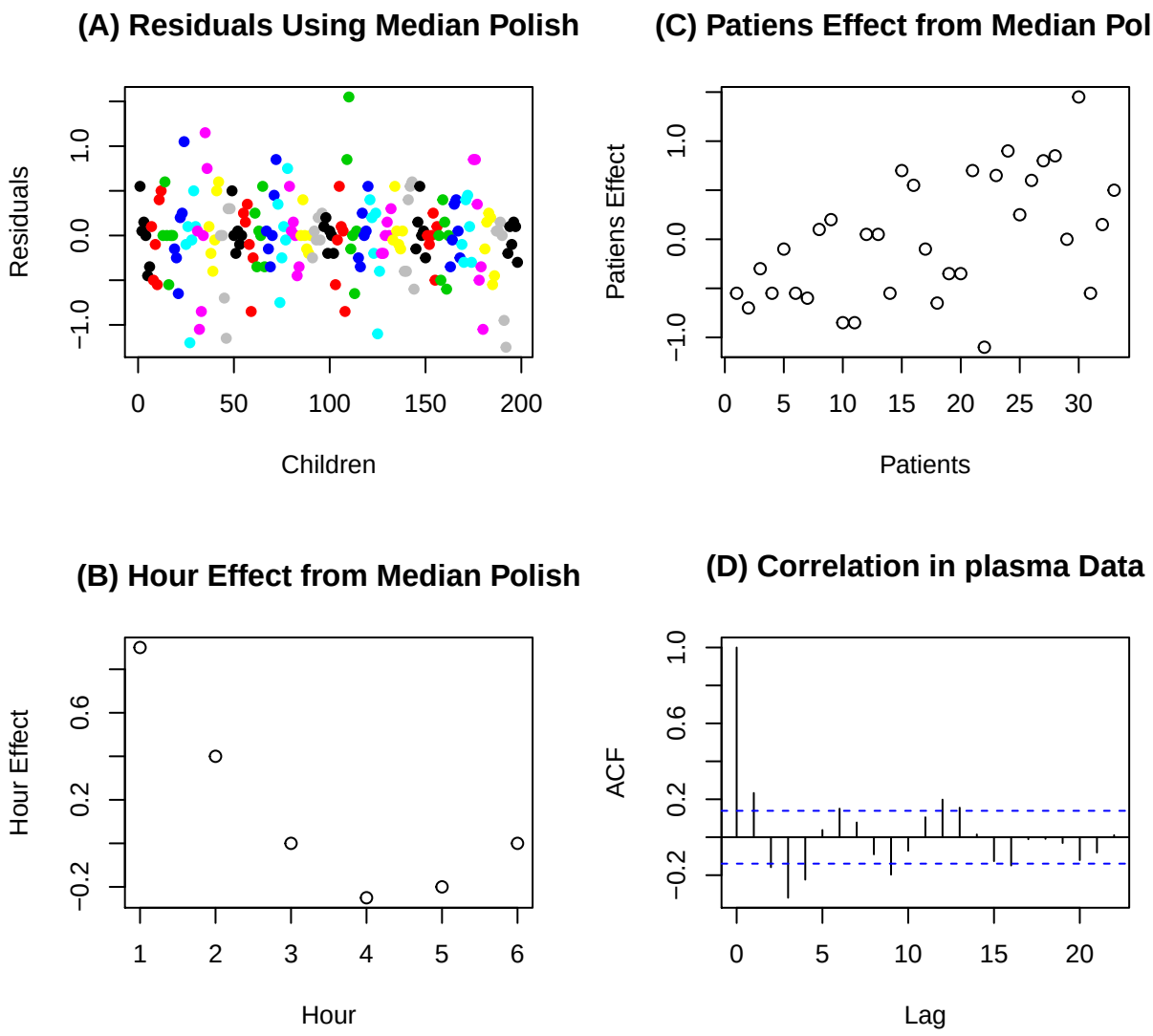
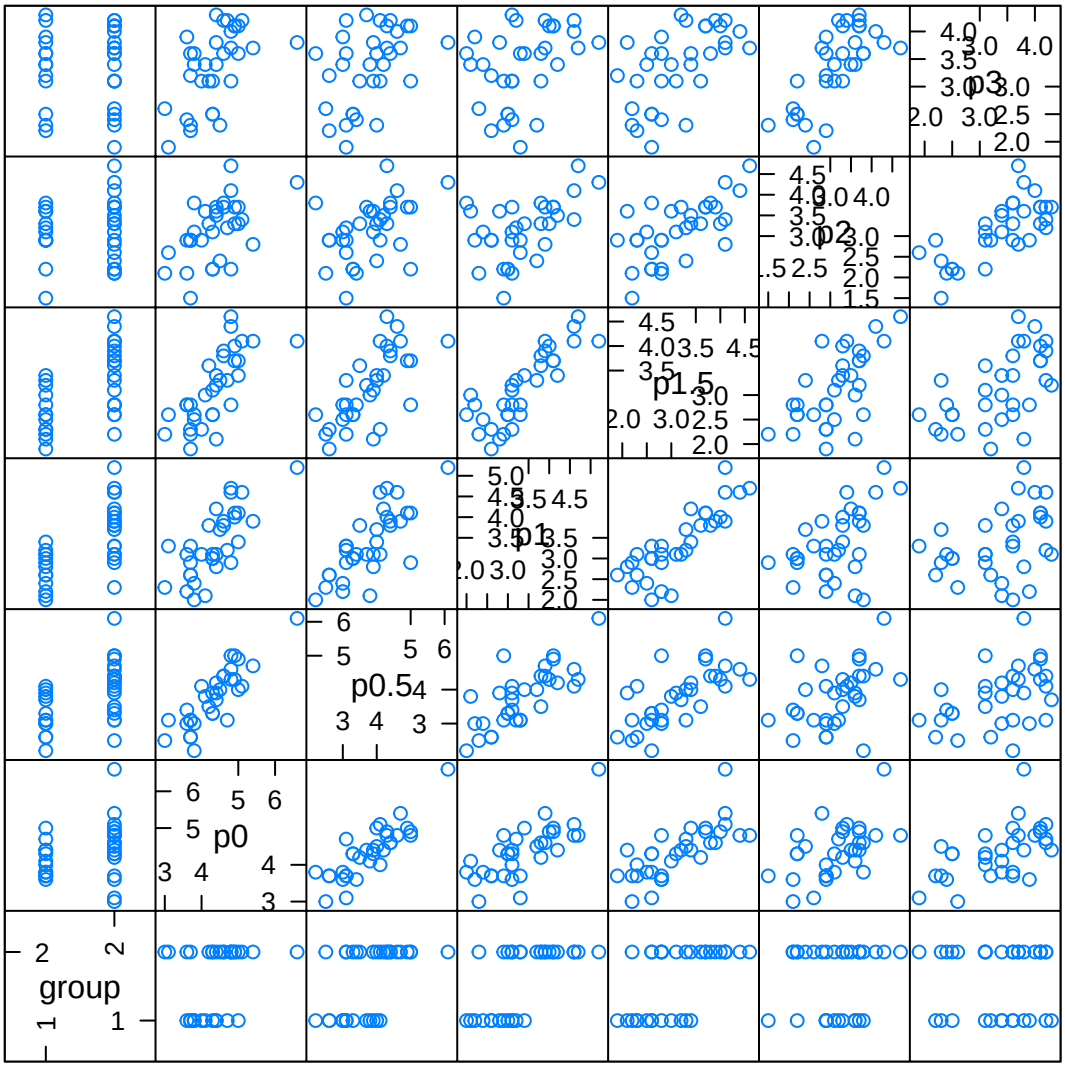


Figure14. SPLOM for plasma data



Perfrom Hotelling's T^2 test, here we assume that:

- each patients data follow MVN
- number of time points (p) < number of subjects (n)
- $\Sigma_1 = \Sigma_2$

(a) we assume non parallelism, so we test two groups seperately by using one-sample Hotelling's T^2 test:

- for control group, we calculate the pvalue = $0 < 0.05$, so we reject the null hypothesis and conclude that there is not enough evidence to support that control group means are the same at all six measurement times
- for obese group, we calculate the pvalue = $0 < 0.05$, so we reject the null hypothesis and conclude that there is not enough evidence to support that obese group means are the same at all six measurement times

(b) By using two-sample Hotelling's T^2 test, we calculate the pvalue = $0 < 0.05$, so we reject the null hypothesis and conclude that there is not enough evidence to support that the profiles in the two groups are parallel

(c) By using two-sample Hotelling's T^2 test, we calculate the pvalue = $0.523 > 0.05$, so we fail to reject the null hypothesis and conclude that there is enough evidence to support that the differences in means at

2 and 3 hours after an oral glucose challenge are different between the control and obese patients

Appendix

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE, comment = "")
library(tidyverse)
library(lattice) # for xyplot and SPLOM
options(knitr.table.format = "latex")
theme_set(theme_bw())
# EDA for TLC data
TLC = read.table("TLC.dat", header = F) %>%
  rename(id = V1, assignment = V2, week0 = V3, week1 = V4, week4 = V5, week6 = V6) %>%
  mutate(id = as.factor(id)) %>%
  as.tibble()

# scatterplot
TLC %>% gather(week0:week6, key = "weeks", value = "level") %>%
  ggplot(aes(x = weeks, y = level, group = id, color = id)) +
  geom_line() +
  facet_grid(. ~ assignment) +
  theme(legend.position = "none")

# boxplot about mean and sd
TLC %>% gather(week0:week6, key = "weeks", value = "level") %>%
  ggplot(aes(x = weeks, y = level, group = weeks)) +
  geom_boxplot() +
  facet_grid(. ~ assignment) +
  theme(legend.position = "none")

# standardized time plot
TLC2 <- sweep(TLC[,c(3:6)], 2, apply(TLC[,c(3:6)], 2, mean))
sd1 <- apply(TLC[,c(3:6)], 2, sd)
TLC2 <- sweep(TLC2, 2, sd1, FUN = "/")
TLC2 <- cbind(TLC[,c(1:2)], TLC2)
TLC2 %>% gather(week0:week6, key = "weeks", value = "level") %>%
  ggplot(aes(x = weeks, y = level, group = id, color = id)) +
  geom_line() +
  facet_grid(. ~ assignment) +
  theme(legend.position = "none")

# median polish
junk1 <- medpolish(TLC[,c(3:6)])
res <- junk1$res
cols <- rep(1:100, rep(4, 100))
par(mfcol = c(2, 2))
plot(as.vector(t(res)), col = cols, pch = 19, cex = 0.8,
     xlab = "Children", ylab = "Residuals", main = "(A) Residuals Using Median Polish")
plot(junk1$col ,xlab = "Weeks", ylab = "Week Effect", main = "(B) Week Effect from Median Polish")
plot(junk1$row, xlab = "Children", ylab = "Children Effect", main = "(C) Children Effect from Median Po
acf(as.vector(t(res)), xlab = "Lag", main = "(D) Correlation in TLC Data")

# SPLOM
```

```r
splom(TLC[2:6], xlab = NULL)

# load dental data
dent = read.table("dental.dat", header = F) %>%
  rename(id = V1, gender = V2, age8 = V3, age10 = V4, age12 = V5, age14 = V6) %>%
  mutate(id = as.factor(id)) %>%
  as.tibble()

# scatterplot
dent %>% gather(age8:age14, key = "age", value = "level") %>% ggplot(aes(x = age, y = level, group = id
facet_grid(. ~ gender) +
theme(legend.position = "none")

# boxplot about mean and sd
dent %>% gather(age8:age14, key = "age", value = "level") %>%
  ggplot(aes(x = age, y = level, group = age)) +
  geom_boxplot() +
  facet_grid(. ~ gender) +
  theme(legend.position = "none")

# standardized time plot
dent2 <- sweep(dent[,c(3:6)], 2, apply(dent[,c(3:6)], 2, mean))
sd2 <- apply(dent[,c(3:6)], 2, sd)
dent2 <- sweep(dent2, 2, sd2, FUN = "/")
dent2 <- cbind(dent[,c(1:2)], dent2)
dent2 %>% gather(age8:age14, key = "age", value = "level") %>%
  ggplot(aes(x = age, y = level, group = id, color = id)) +
  geom_line() +
  facet_grid(. ~ gender) +
  theme(legend.position = "none")

# median polish
junk2 <- medpolish(dent[,c(3:6)])
res2 <- junk2$res
cols2 <- rep(1:27, rep(4, 27))
par(mfcol = c(2, 2))
plot(as.vector(t(res2)), col = cols2, pch = 19, cex = 0.8,
     xlab = "Children", ylab = "Residuals", main = "(A) Residuals Using Median Polish")
plot(junk2$col ,xlab = "Age", ylab = "Age Effect", main = "(B) Age Effect from Median Polish")
plot(junk2$row, xlab = "Children", ylab = "Children Effect", main = "(C) Children Effect from Median Po
acf(as.vector(t(res2)), xlab = "Lag", main = "(D) Correlation in dental Data")

# SPLOM
splom(dent[2:6], xlab = NULL)

# load dental data
plasma = read.table("ZERBE2.DAT", header = F) %>%
  rename(group = V1, id = V2, p0 = V3, p0.5 = V4, p1 = V5, p1.5 = V6, p2 = V7, p3 = V8) %>%
  mutate(id = c(1:33),
         id = as.factor(id),
         group = as.factor(group)) %>%
  as.tibble()
```

```r
# boxplot about mean and sd
plasma %>% gather(p0:p3, key = "hour", value = "level") %>%
  ggplot(aes(x = hour, y = level, group = hour)) +
  geom_boxplot() +
  facet_grid(. ~ group) +
  theme(legend.position = "none")

# standardized time plot
plasma2 <- sweep(plasma[,c(3:8)], 2, apply(plasma[,c(3:8)], 2, mean))
sd3 <- apply(plasma[,c(3:6)], 2, sd)
plasma2 <- sweep(plasma2, 2, sd3, FUN = "/")
plasma2 <- cbind(plasma[,c(1:2)], plasma2)
plasma2 %>% gather(p0:p3, key = "hour", value = "level") %>%
  ggplot(aes(x = hour, y = level, group = id, color = id)) +
  geom_line() +
  facet_grid(. ~ group) +
  theme(legend.position = "none")

# median polish
junk3 <- medpolish(plasma[,c(3:8)])
res3 <- junk3$res
cols3 <- rep(1:33, rep(6, 33))
par(mfcol = c(2, 2))
plot(as.vector(t(res3)), col = cols3, pch = 19, cex = 0.8,
     xlab = "Children", ylab = "Residuals", main = "(A) Residuals Using Median Polish")
plot(junk3$col ,xlab = "Hour", ylab = "Hour Effect", main = "(B) Hour Effect from Median Polish")
plot(junk3$row, xlab = "Patients", ylab = "Patiens Effect", main = "(C) Patiens Effect from Median Polis
acf(as.vector(t(res3)), xlab = "Lag", main = "(D) Correlation in plasma Data")

# SPLOM
splom(plasma[c(1,3:8)], xlab = NULL)

# Hotelling T test
library(ICSNP)
library(mvtnorm)

# for (a)
x11 = plasma %>%
  filter(group == 1) %>%
  select(p0:p3)
x12 = plasma %>%
  filter(group == 2) %>%
  select(p0:p3)
C1 = rbind(diag(5), rep(-1,5))
H11 = HotellingsT2(as.matrix(x11) %*% C1)

H12 = HotellingsT2(as.matrix(x12) %*% C1)

# for (b)
x21 = plasma %>%
  filter(group == 1) %>%
  select(p0:p3) %>%
  as.matrix()
```

```r
x22 = plasma %>%
  filter(group == 2) %>%
  select(p0:p3) %>%
  as.matrix()
C2 = matrix(c(c(1,-1,0,0,0,0),
              c(0,1,-1,0,0,0),
              c(0,0,1,-1,0,0),
              c(0,0,0,1,-1,0),
              c(0,0,0,0,1,-1)), nrow = 6)
H2 = HotellingsT2(rbind(x21 %*% C2, x22 %*% C2) ~ plasma$group)

# for (c)
x31 = plasma %>%
  filter(group == 1) %>%
  select(p0:p3) %>%
  as.matrix()
x32 = plasma %>%
  filter(group == 2) %>%
  select(p0:p3) %>%
  as.matrix()
C3 = matrix(c(0,0,0,0,1,-1), nrow = 6)
H3 = HotellingsT2(rbind(x31 %*% C3, x32 %*% C3) ~ plasma$group)
```