

Question 1

Joseph Froelicher

4/30/2021

Part A

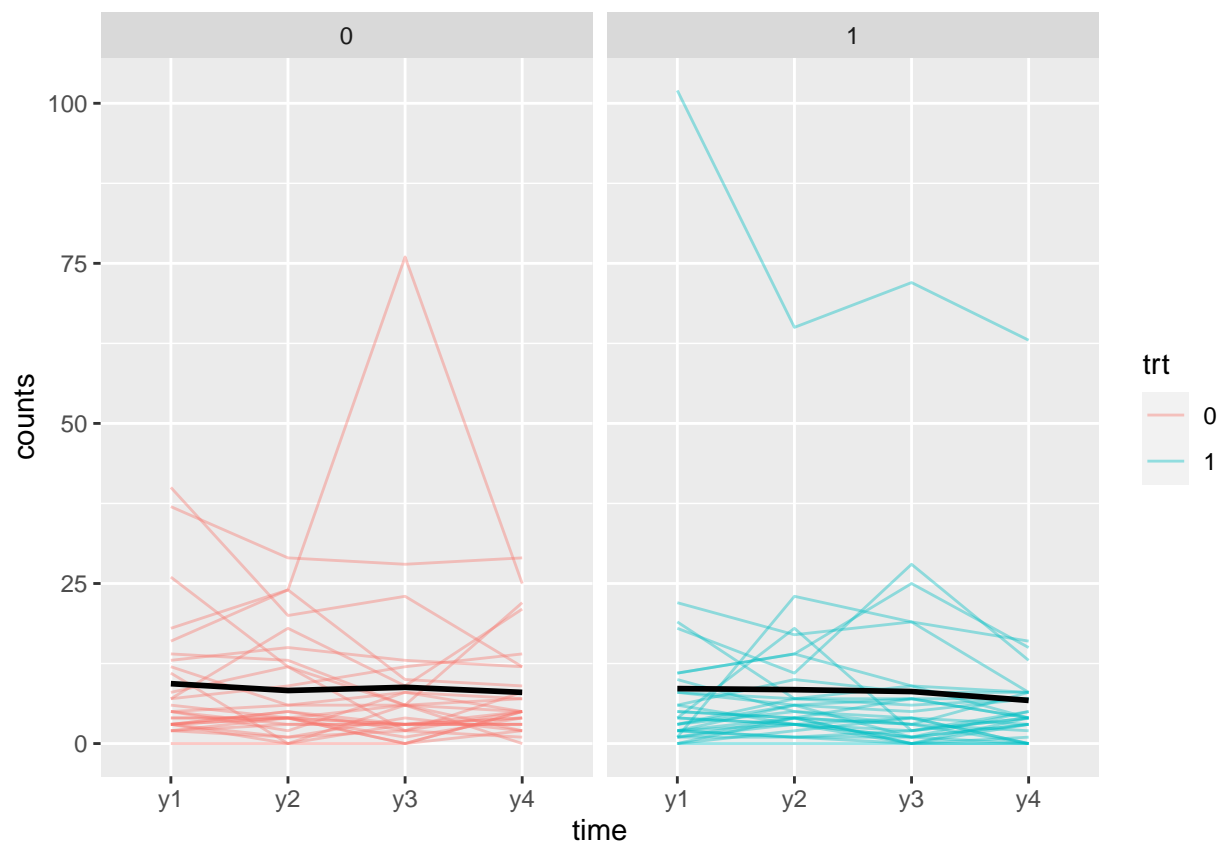
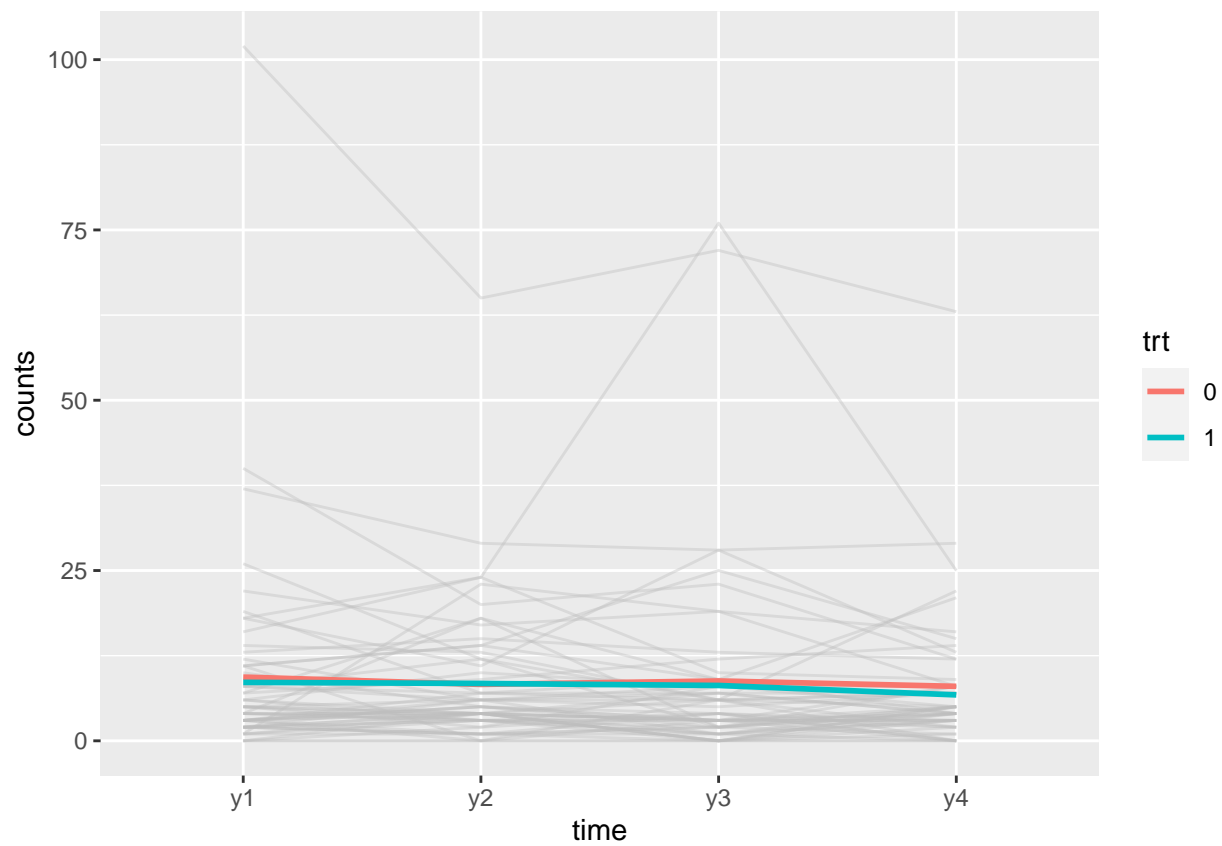
```
data(seizure)
seizure$id = 1:dim(seizure)[1]
data = pivot_longer(data = seizure, cols = c("y1", "y2", "y3", "y4"), names_to = "time", values_to = "c")
data$trt = factor(data$trt)

if ((dim(data)[1] / 4) == dim(seizure)[1]) {
  print("The data is balanced.")
} else {
  print("The data is unbalanced.")
}
```

```
## [1] "The data is balanced."
```

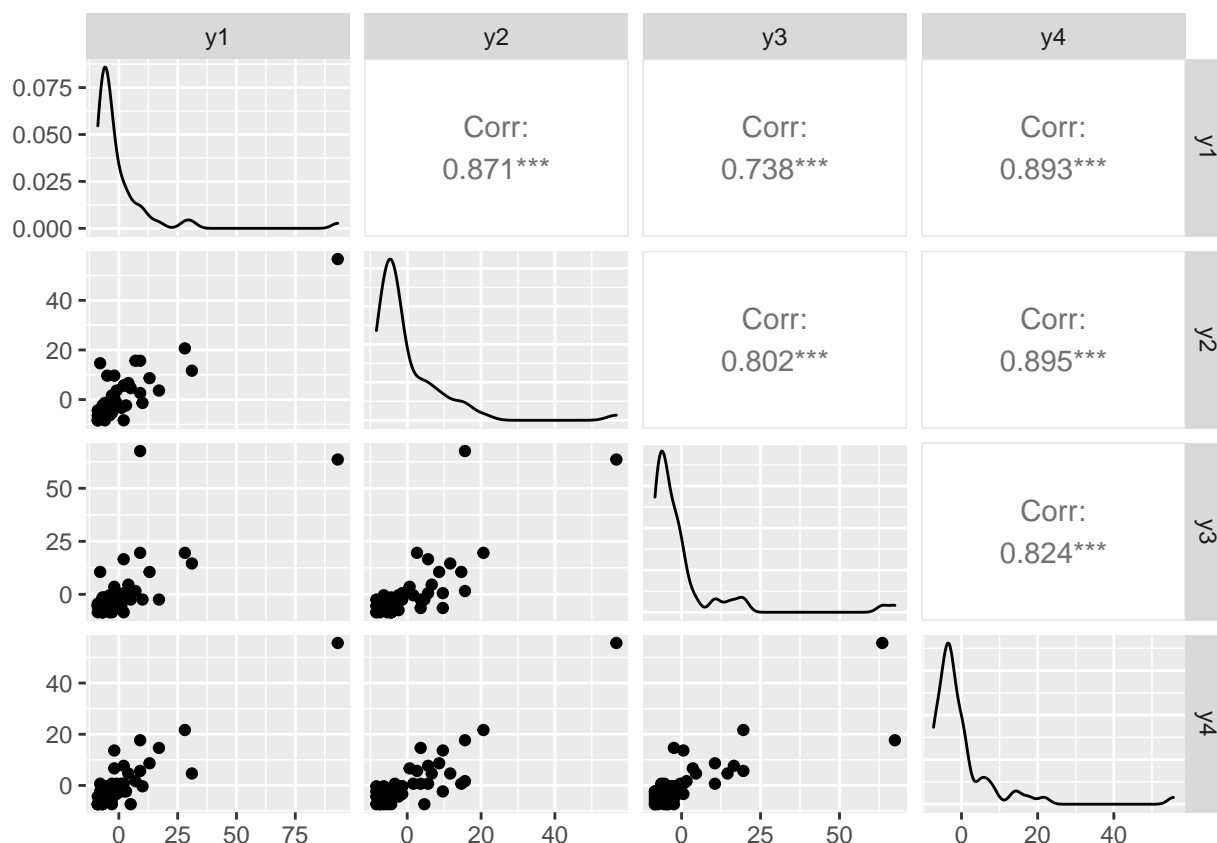
We can say the data is balanced because there are the same number of time points ($n = 4$) for each patient id ($N = 59$).

Part B



As we can see from the plots, there is not much to say about the mean structure of our outcome over time, or the relationships between the outcome and covariates. If anything we would say that for both groups counts of seizures are decreasing slightly over time, but this is visual speculation and we would want to test for this to be certain. There is not much to say that there is a difference between the means of either of the treatment groups either. Lastly, we should note that there are a few outlying patients that are suffering significantly more seizures than the mean number of seizures at each time point. The outlier in treatment group 1 is showing a very interesting case, and it may be of interest to examine the story behind that patient. They are experiencing a very high number of seizures, and a very obvious decrease in seizures over time. A question of interest would then be, how does treatment affect those patients with high numbers of seizures?

Part C



Based on the plot above, the values of correlations between time points is quite similar, and does not follow any pattern based on time points. We would likely not want to assume that the correlations are different based on these plots, thus I would choose exchangeable correlation as our working correlation structure. This seems reasonable because we have well balanced data, and the correlation between any given two covariates residuals is roughly equal. Furthermore, we very clearly should not assume independence between timepoints based on the high values of correlations between time points.

Appendix

```
knitr::opts_chunk$set(include = TRUE, echo = TRUE, warning = FALSE)

library(geepack)
library(GGally)
library(ggplot2)
library(tidyr)
library(tidyverse)

data(seizure)
seizure$id = 1:dim(seizure)[1]
data = pivot_longer(data = seizure, cols = c("y1", "y2", "y3", "y4"), names_to = "time", values_to = "counts")
data$trt = factor(data$trt)

if ((dim(data)[1] / 4) == dim(seizure)[1]) {
  print("The data is balanced.")
} else {
  print("The data is unbalanced.")
}

data %>%
  ggplot(aes(time, counts)) +
  geom_line(alpha = 0.4, aes(group = id), colour = "gray") +
  stat_smooth(aes(group = trt, color = trt), method = loess, geom = "line", fun.y = mean, size = 1)

data %>%
  ggplot(aes(time, counts)) +
  geom_line(alpha = 0.4, aes(group = id, color = trt)) +
  stat_smooth(aes(group = trt), method = loess, geom = "line", fun.y = mean, size = 1) +
  facet_wrap(~ trt)

seizure_residuals = data %>%
  group_by(time) %>%
  mutate(mean = mean(counts)) %>%
  ungroup() %>%
  mutate(residuals = counts - mean) %>%
  select(id, time, residuals)

seizure_residuals_wide = seizure_residuals %>%
  pivot_wider(
    names_from = time,
    values_from = residuals
  ) %>%
  select(-id)

ggpairs(seizure_residuals_wide)
```