2019 Problem 2

Emily Shives

May 17, 2020

Problem 2

Part a

Descriptive statistics

```
kable(pilot %>% group_by(group) %>% summarize(mean=mean(totchol)), caption = "Mean by Group") %
>%
  kable_styling(full_width = F)
```

Mean by Group

group	mean
C	212.0
E	206.7

```
#kable(pilot %>% group_by(id) %>% summarize(mean=mean(totchol)), caption = "Mean by ID")%>%
  #kable_styling(full_width = F)
kable(pilot %>% group_by(weeks) %>% summarize(mean=mean(totchol)), caption = "Mean by Weeks")%>%
  kable_styling(full_width = F)
```

Mean by Weeks

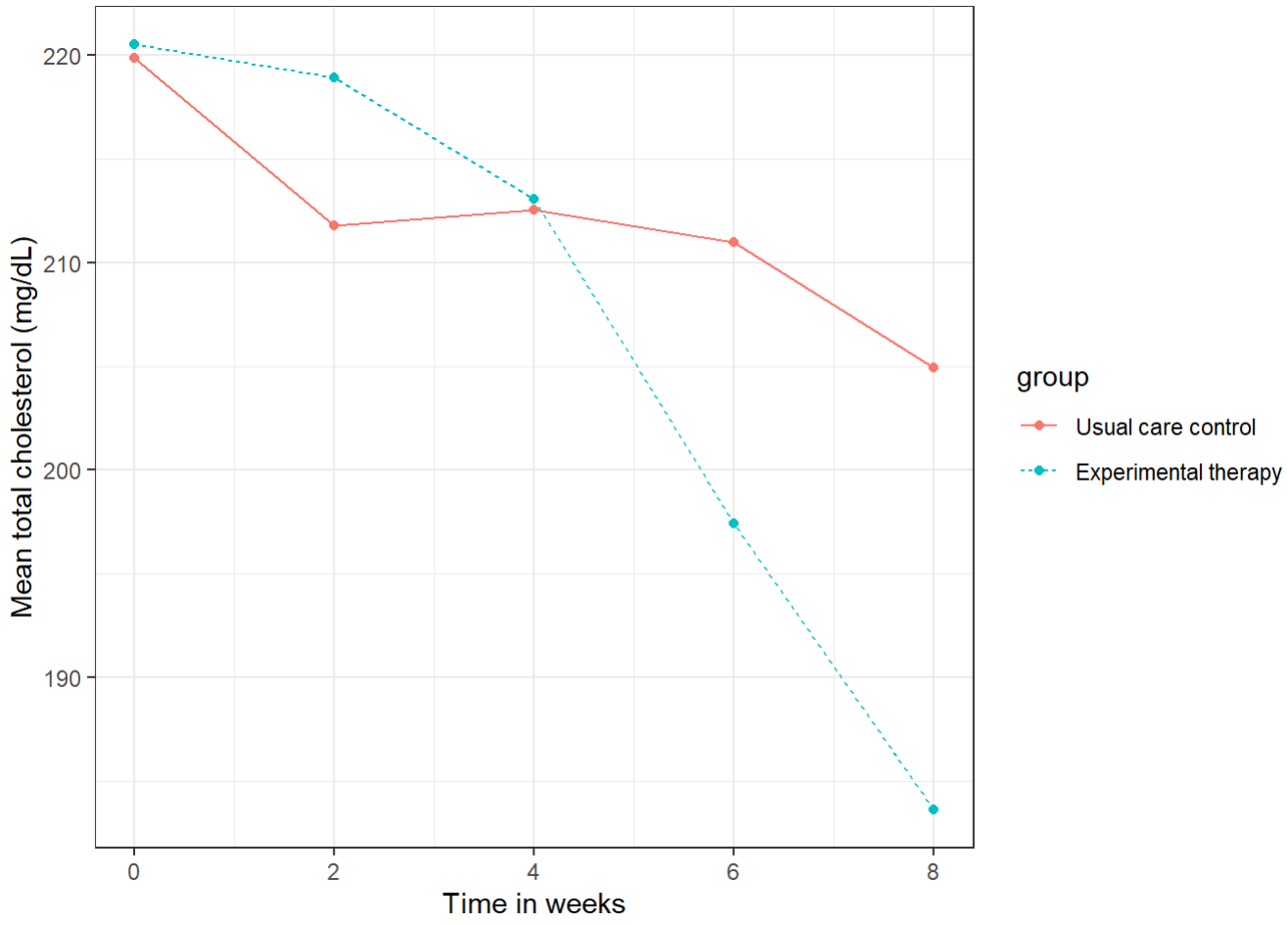
weeks	mean
0	220.2
2	215.4
4	212.8
6	204.2
8	194.3

```
T <- pilot %>% group_by(group,weeks) %>% summarize(mean=mean(totchol))
levels(T$group) <- c("Usual care control","Experimental therapy")
#kable(pilot %>% group_by(group,weeks) %>% summarize(mean=mean(totchol)), caption = "Mean by Weeks and group")%>%
 # kable_styling(full_width = F)
kable(T %>% spread(weeks,mean), caption = "Mean by Weeks and group")%>%
  kable_styling(full_width = F)
```

Mean by Weeks and group

group	0	2	4	6	8
Usual care control	219.9	211.8	212.5	211.0	204.9
Experimental therapy	220.5	218.9	213.1	197.4	183.6

```
ggplot(data = T, aes(x=weeks, y=mean, lty=group, color=group))+
  geom_point()+
  theme_bw()+
  labs(x="Time in weeks",y="Mean total cholesterol (mg/dL)")+
  geom_line()
```



From the pilot study data, $\hat{\sigma}_b^2 = 569.52$ and $\hat{\sigma}_w^2 = 264.16$. Under the null, $\hat{\sigma}_b^2 = 564.68$ and $\hat{\sigma}_w^2 = 288.35$.

```
df_init1 <- data.frame(params=c("$\\beta_0$","$\\beta_1$","$\\beta_2$","$\\beta_3$","$\\sigma^2_
b$","$\\sigma^2_w$"),
                       Estimate=c(218.16, 7.63, -1.53, -3.24, 569.52, 264.16),
                       StdError=c(6.97, 9.85, 0.66, 0.94, "", ""))
kable(df_init1, col.names = c("Parameter", "Estimate", "Standard Error"), caption = "Parameter a
nd standard error estimates from linear mixed model for total cholesterol levels with treatment,
time (in weeks), and their interaction in the model in pilot study", align="c") %>%
  kable_styling(latex_options = "hold_position", full_width = F)
```

Parameter and standard error estimates from linear mixed model for total cholesterol levels with treatment, time (in weeks), and their interaction in the model in pilot study

Parameter	Estimate	Standard Error
β_0	218.16	6.97
β_1	7.63	9.85
β_2	-1.53	0.66
β_3	-3.24	0.94
σ^2_b	569.52	
σ^2_w	264.16	

```
df_init2 <- data.frame(params=c("$\\gamma_0$","$\\gamma_1$","$\\gamma_2$","$\\gamma_3$","$\\rho
$"),
                       Estimate=c(1.63, 0.15, -0.20, -0.13, 0.462),
                       StdError=c(0.58, 0.82, 0.07, 0.10, ""))
kable(df_init2, col.names = c("Parameter", "Estimate", "Standard Error"), caption = "Parameter a
nd model-based standard error estimates from marginal logistic regression model for elevated tot
al cholesterol levels with treatment, time (in weeks), and their interaction in the model in pil
ot study", align="c") %>%
  kable_styling(latex_options = "hold_position", full_width = F)
```

Parameter and model-based standard error estimates from marginal logistic regression model for elevated total cholesterol levels with treatment, time (in weeks), and their interaction in the model in pilot study

Parameter	Estimate	Standard Error
γ_0	1.630	0.58
γ_1	0.150	0.82

Parameter	Estimate	Standard Error
γ_2	-0.200	0.07
γ_3	-0.130	0.1
ρ	0.462	

Part b

Using the formula (from page 588 of Fitzmaurice textbook)

$$\sqrt{N} = \frac{(z_{1-\alpha} + z_{1-\gamma})\sqrt{\sigma_w^2}}{0.5\delta\sqrt{\sum_{j=1}^{n}(t_j - \bar{t})^2}}$$

where z_d is the dth quantile of the standard normal distribution and δ is the difference between the parameter of interest under the null hypothesis and under the alternative hypothesis.

```
weeks <- c(0,2,4,6,8)
sum_of_tsqs <- sum((weeks-mean(weeks))^2)
power <- 0.85
alpha <- 0.025 #one-sided
zpower <- qnorm(power)
zalpha <- qnorm(1-alpha)
parameter_diff <- 0-(-3)
sigma2w <- 264.16
lambda0 <- sqrt(sigma2w)/(0.5*sqrt(sum_of_tsqs))
lambda1 <- lambda0
sqrtn <- (zalpha*lambda0+zpower*lambda1)/parameter_diff
n <- (sqrtn)^2
ceil(n)
```

```
## [1] 27
```

Assumes equal variance in both treatment groups.

Part c

My biggest question in this part is how to generate appropraite values for the gamma's and rho?

```
pilot <- pilot %>% mutate(di_totchol = as.numeric(totchol>=200))
pilot %>% group_by(group,weeks) %>% summarize(mean=mean(di_totchol))
```

```
## # A tibble: 10 x 3
## # Groups:   group [2]
##    group weeks  mean
##    <fct> <int> <dbl>
##  1 C         0 0.867
##  2 C         2 0.8
##  3 C         4 0.667
##  4 C         6 0.533
##  5 C         8 0.6
##  6 E         0 0.867
##  7 E         2 0.733
##  8 E         4 0.667
##  9 E         6 0.4
## 10 E         8 0.333
```

I used gamma0-2 based on the pilot study values? then gamma3 as the hypothesized value and correlation structure from rho of pilot study data with constant correlation?

```
power <- 0.85
alpha <- 0.025 #one-sided
zpower <- qnorm(power)
zalpha <- qnorm(1-alpha)
parameter_diff <- 0-(-0.15)
lambda0 <- 0.4055#0.4326
lambda1 <- 0.4079#0.4283
sqrtn <- (zalpha*lambda0+zpower*lambda1)/parameter_diff
n <- (sqrtn)^2
ceil(n)
```

```
## [1] 66
```

Part d

Use the models given in the problem and hypothesis tests from the problem (the marginal model for the dichotomous outcome assuming an . An available data analysis will be the primary analysis. Missing data can be filled in using multiple imputation using MCMC, assuming that the data is missing at random (or missing completely at random). The results using MI can be compared to the primary analysis and similar results provide support for the validity of the primary analysis. Missing at random is plausible, since this allows the missingness to depend on the observed data, which means that the previous results may have an impact on the missingness. Side effects of the drug are a likely candidate for a reason for missingness, so …