PHP2516: Applied Longitudinal Data Analysis

Homework 3

Antonella Basso

March 16, 2022

Data

The "**amenorrhea**" dataset pertains to a longitudinal clinical trial of contracepting women comparing the effectiveness of two dosages (100mg or 150mg) of depot-medroxyprogesterone acetate (DMPA) (Machin et al., 1988), an injectable contraceptive. A total of $N = 1,151$ women kept a menstrual diary according to whether they had experienced amenorrhea in four successive three-month intervals (binary outcome). This clinical trial had a substantial dropout. The "amenorrhea" dataset has information on the following variables:

- **id**: Person ID
- **dose**: Treatment group ($0 =$ Low (100mg), $1 =$ High (150mg))
- **time**: Time point $(1, 2, 3, 4)$, of the four consecutive 3-month injection intervals
- **status**: Amenorrhea status ($0 =$ no amenorrhea, $1 =$ amenorrhea)

The primary objective of this study is to determine changes in the risk of amenorrhea over the course of the study (12 months), and the influence of DMPA dosage on changes in a woman's risk of amenorrhea. Of note, the treatment covariate (dose) is time invariant.

Question 1: Exploratory Data Analysis (EDA)

Provide a table with appropriate descriptive statistics and a plot for summarizing and presenting the information collected in this study, in regards to the primary objective.

Solution

Overview of Data:

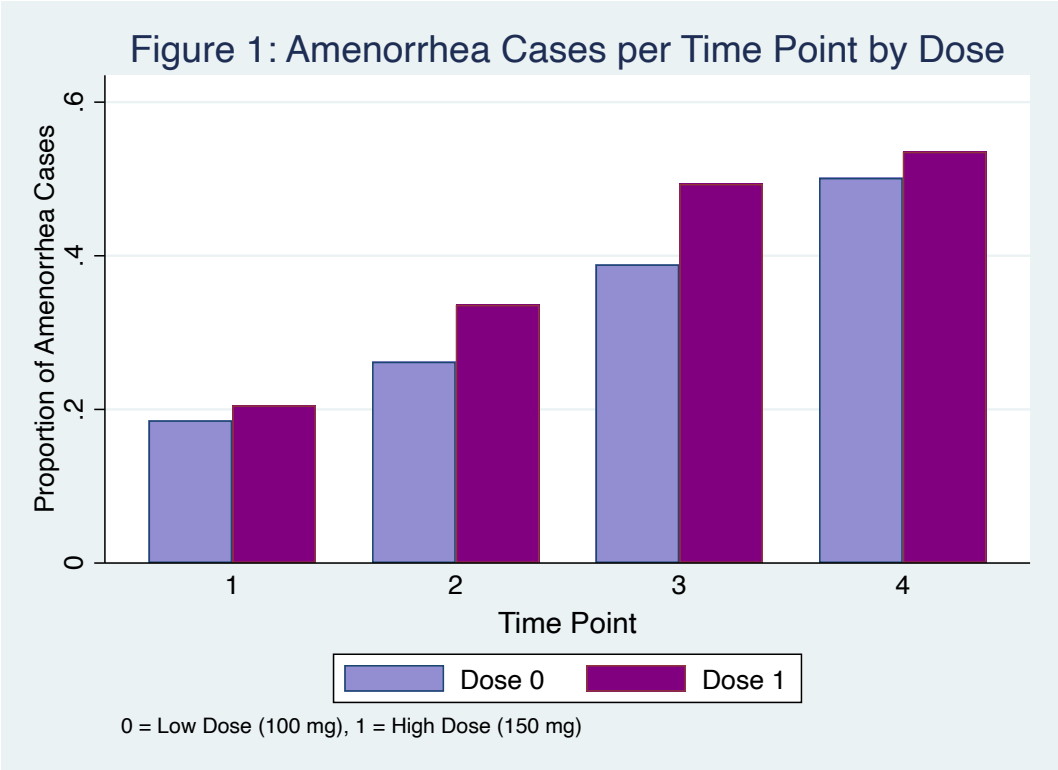
- 4,604 total observations for 1,151 individuals
- 988 missing observations (for amenorrhea `status`)
- 1,231 documented amenorrhea cases (572 for dose=0, 659 for dose=1)
- 2,385 documented non-amenorrhea cases
- 633 unique individuals that experienced amenorrhea at least once
- 518 unique individuals that did not experience amenorrhea at all

The information displayed in Table 1 below, obtained in STATA, gives the total number of cases recorded at each time point, as well as the proportion of individuals in the study that experienced amenorrhea at each time point (one of four consecutive 3-month intervals) for low and high doses of DMPA, which are denoted as $0 =$ Low (100mg) and $1 =$ High (150mg), respectively. It is important to note that this information, which may be visualized in Figure 1 below, is purely representative of total (population-level) documented cases of amenorrhea at each time point and does not reflect the amount of missing data present nor the occurrence of amenorrhea on an individual basis. To get a better sense for how frequently and how significantly individuals are affected by amenorrhea at given doses of DMPA, Tables 2-5 were created. Specifically, for the 276, 174, 125, and 58 individuals that experienced amenorrhea exactly once, twice, three times, and four times,

respectively (out of the 633 total individuals who experienced amenorrhea), these tables show the time point(s) at which they experienced amenorrhea for their corresponding dose levels.

Table 1: Proportions of Cases by Dose & Count per Time Point

Time Point	Dose 0	Dose 1	Total Cases
1	0.1858	0.2052	225
2	0.2621	0.3361	285
3	0.3888	0.4936	351
4	0.5014	0.5354	370



Aside from Figure 1, which conveys the information provided in Table 1, Figures 2 and 3 below provide means for visualizing some of the information given by Tables 2-5. Specifically, Figure 2 displays the information from Table 2, and hence, gives the number of individuals that experienced amenorrhea, for their first and only time, in the first, second, third, and fourth time points (3-month intervals in the 12-month study), showing each treatment group's response over time (in this particular category). Additionally, Figure 3 gives the distribution of the difference in the number of individuals between high and low dose treatment groups for all amenorrhea occurrence patterns. That is, the distribution of the value obtained from subtracting the "Dose 0" column from the "Dose 1" column in Tables 2-5, such that its distribution reflects the spread of how much greater counts in the high dose group were in contrast to those in the low dose group.

Table 2: Counts of Cases at Each Time Point for Individuals Who Experienced Amenorrhea Once

First Case	Dose 0	Dose 1	Total
1	34	41	75
2	22	31	53
3	27	36	63
4	49	36	85

Table 3: Counts of Cases at Each Time Point for Individuals Who Experienced Amenorrhea Twice

First Case	Second Case	Dose 0	Dose 1	Total
1	2	12	17	29
1	3	2	3	5
1	4	7	6	13
2	3	9	13	22
2	4	8	12	20
3	4	41	44	85

Table 4: Counts of Cases at Each Time Point for Individuals Who Experienced Amenorrhea Three Times

First Case	Second Case	Third Case	Dose 0	Dose 1	Total
1	2	3	8	8	16
1	2	4	4	3	7
1	3	4	10	12	22
2	3	4	32	48	80

Table 5: Counts of Cases at Each Time Point for Individuals Who Experienced Amenorrhea Four Times

First Case	Second Case	Third Case	Fourth Case	Dose 0	Dose 1	Total
1	2	3	4	30	28	58

Figure 2: Count of Amenorrhea Cases per Time Point by Dose

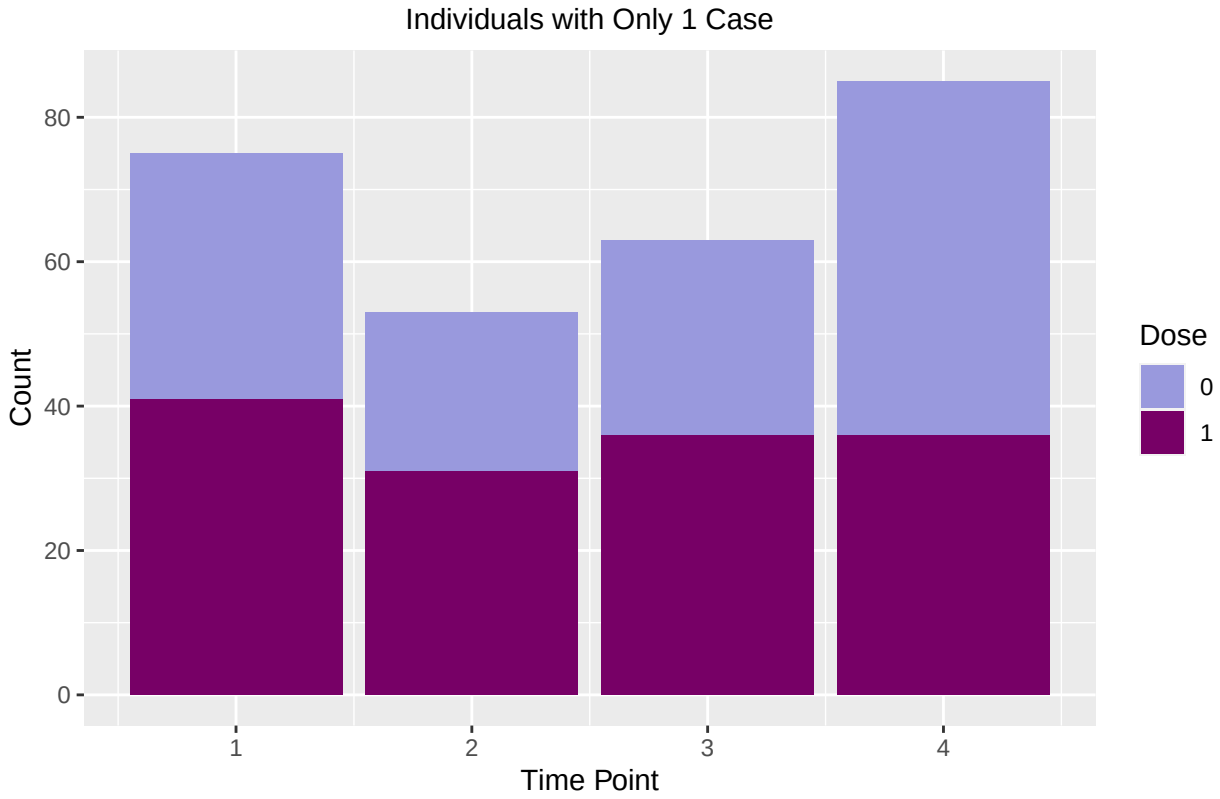
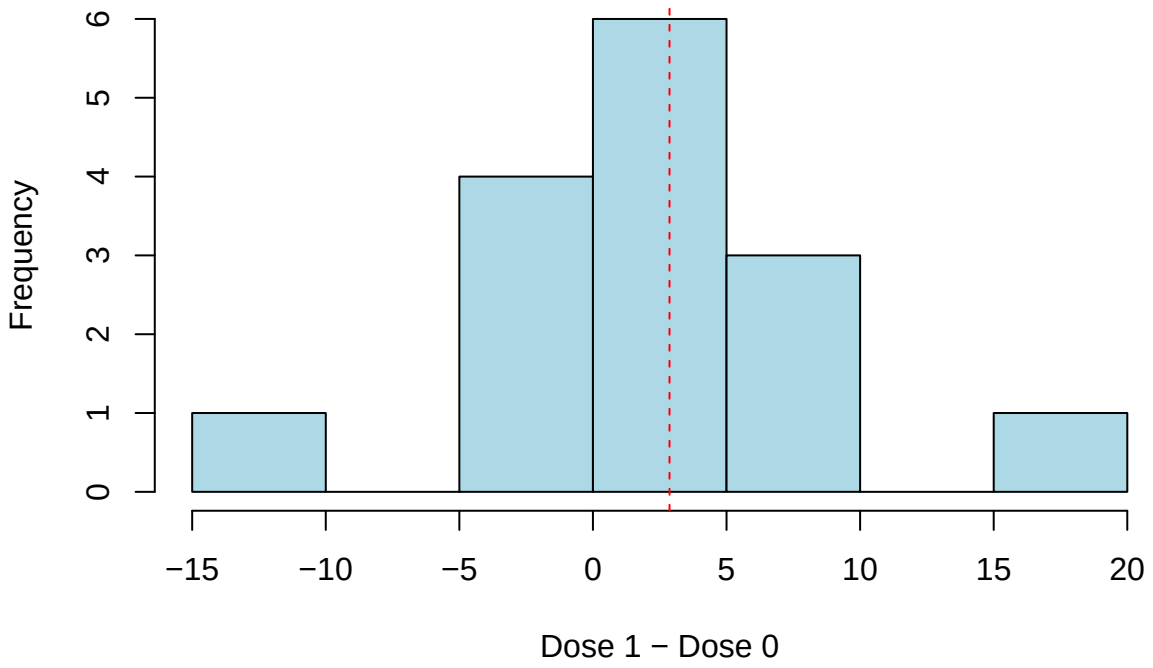


Figure 3: Distribution of Difference in Dose Counts



Question 2:

Comment on the results from both descriptive statistics and plots. What is your conclusions regarding the DMPA dosage on changes in a woman's risk of amenorrhea over time?

Solution

Based on the information presented above, it appears that an individuals's risk of amenorrhea increases over time in a similar fashion for both DMPA dose groups. Both the population-based information (Table 1 and Figure 1) and individual-level information (Tables 2-5 and Figures 2-3) presented above paint a picture for how the frequency of amenorrhea cases changes from time point to time point. We see more clearly in Figure 1, that there is a growing trend in cases over time that remains relatively constant between treatment groups, despite counts in the high dose group being slightly larger at each time interval. What we see in Tables 2-5 and Figure 2 that is not captured here however, is that (for both groups) there is an initial spike in cases that rapidly drops and subsequently follows a more steady increase that reaches a maximum at the last time point. Moreover, we see that, based on the data, individuals may be more likely to experience amenorrhea once than they are to exprience it twice; more likely to experience it twice than they are three times; and more likely to experience it three times than they are all four times. And, of those more likely to experience it once, they (based on the observations) may be most likely to experience it between the 8-12-month mark (fourth time point). Specifically, we see in Figure 2 that most who experienced amenorrhea only once experienced it later on in the study, irrespective of their corresponding DMPA dosage (as with those who experienced it more times). This indicates that increase in amenorrhea cases, and hence, an individual's risk of amenorrhea, might be more so due to an effect of time on the response as opposed to an increase in dose. Still however, it is clear that the number of amenorrhea cases is generally slightly greater in the high dose treatment group than in the low dose treatment group, which could indicate an increased risk of amenorrhea for those recieving 150mg of DMPA (in comparison to those recieving 100mg). For this reason, Figure 3 was used to gauge the significance of this difference. Looking at the distribution of count difference between groups, this value appears to be normally distributed around a mean of ≈ 2.87 with a standard deviation of ≈ 6.5, meaning that, given the data, a higher dose of DMPA may not necessarily bring about a change in the response. Thus, although counts in the low dose group appear slightly more turbulent, given the EDA from above, it seems that the risk of experiencing amenorrhea for individuals taking DMPA increases with time. Yet, it is not entirely clear whether this risk is exacerbated by an increase in DMPA dosage. That is, if a higher dose of DMPA does in fact have an effect on the outcome, then this data (on surface-level) doesn't convey that (at best it hints at a possible minor difference).

Question 3:

Fit an appropriate model that expresses changes in the primary outcome as a function of `time`, `time`^{2}, `dose` and the two-way interactions of `dose` with `time` and `time`^{2}. Include time as a continuous variable. Try the following:

a) A marginal model

b) A mixed-effects model assuming a random intercept and a random slope

Solution

The marginal model (a) below gives the (population) mean response (log-odds of experiencing amenorrhea) over time as a function of `time`, `time`^{2}, `dose` and the two-way interactions of `dose` with `time` and `time`^{2}. This model was fit in STATA using a GEE estimation approach and reflects statistically significant coefficients for all terms. Moreover, the mixed effects model (b) below, fit in the same way, gives the subject-specific responses over time, assumming a random intercept and a random slope, which was the best fitting (according to BIC comparisons).

Let:

- Y_i be the outcome of interest, amenorrhea status, for the i^{th} individual in the study, such that $\mathrm{E}[Y_i] = \mu_i$ is the corresponding probability of experiencing amenorrhea ($P(Y_i = 1)$) and $\text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$)
- X_i be the binary predictor for dose, where $X_i = 0$ indicates a low dose of 100 mg (reference group), and $X_i = 1$ indicates a high dose of 150 mg
- T_i be the continuous time covariate, in the range $(1, 4)$, denoting the 3-month interval (of four) in the study
- T_i^2 be the continuous squared time covariate

a) **Marginal Model**:

```
xtgee status i.dose##c.time i.dose##c.time2, family(binomial) link(logit)
corr(unstr) robust
```

$$\text{logit}(\mathrm{E}[Y_j]) = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot T_i + \beta_3 \cdot T_i^2 + \beta_4 \cdot (X_i * T_i) + \beta_5 \cdot (X_i * T_i^2)$$

$$\text{logit}(\mathrm{E}[Y_j]) = -2.0013 - 0.4614 \cdot X_i + 0.5088 \cdot T_i + 0.0022 \cdot T_i^2 + 0.7045 \cdot (X_i * T_i) - 0.1346 \cdot (X_i * T_i^2)$$

b) **Mixed-Effects Model**:

```
xtmelogit status i.dose##c.time i.dose##c.time2 || id: time
```

$$\text{logit}(\mathrm{E}[Y_i]) = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot T_i + \beta_3 \cdot T_i^2 + \beta_4 \cdot (X_i * T_i) + \beta_5 \cdot (X_i * T_i^2) + b_0 + b_1 \cdot T_i$$

$$\text{logit}(\mathrm{E}[Y_i]) = -2.7939 - 0.8719 \cdot X_i + 0.3185 \cdot T_i + 0.1048 \cdot T_i^2 + 1.2694 \cdot (X_i * T_i) - 0.2317 \cdot (X_i * T_i^2) + b_{0i} + b_{1i} \cdot T_i$$

Where $\vec{b_i} = [b_{0i}, b_{1i}]^{\mathrm{T}}$ is the random effects part of the model representing subject-specific intercepts and slopes (with respect to time). That is, the i^{th} individual's deviation from the population mean response in terms of the baseline and change with respect to time, which have standard deviations of ≈ 1.8719 and ≈ 0.6333, respectively.

Question 4:

Interpret the regression coefficients of the models in Question 3. (**HINT**: You may juxtapose meaningful quantities calculated from each model in a table.)

Solution

Tables 6-8 below give, for each model above (marginal model (a) & mixed-effects model (b)), the corresponding coefficients and exponentiated coefficients; the effect of `dose` on the response in terms of an odds-ratio (OR) of Dose=1/Dose=0; as well as the effect of `time` on the response expressed as the odds-ratio (OR) of a 1-unit increase in time. It is worth noting that given the interaction terms in these models, as well as the inclusion of a `time`^{2} term (resulting ia second degree polynomial), the model coefficients in Table 6 are not decipherable in this basic form (even when exponentiated). For this reason, Tables 7 and 8, give more interpretable values for which we can make reasonable claims about the relative effects of DMPA dose and time on the primary outcome of interest (risk of amenorrhea). Given the model's structure and use of a logit (log of odds) link function on the response, the OR's for the effect of dose (shown in Table 7), are calculated using the $M_{a1} - M_{a0}$ and $M_{b1} - M_{b0}$ equations below, which are obtained as follows. Let M_{a1} and M_{a0} be the marginal models (a) for $X_i = 1$ and $X_i = 0$, respectively. Similarly, let M_{b1} and M_{b0} be the mixed-effects

models (b) for $X_i = 1$ and $X_i = 0$. Then the log of the OR's for each model for the high dose group compared to the low dose group at time point T_i are given by:

$$\log\left(\frac{\text{Odds}|X_i = 1}{\text{Odds}|X_i = 0}\right) = M_{a1} - M_{a0} = \beta_1 + \beta_4 \cdot T_i + \beta_5 \cdot T_i^2 = -0.4614 + 0.7045 \cdot T_i - 0.1346 \cdot T_i^2$$

$$\log\left(\frac{\text{Odds}|X_i = 1}{\text{Odds}|X_i = 0}\right) = M_{b1} - M_{b0} = \beta_1 + \beta_4 \cdot T_i + \beta_5 \cdot T_i^2 = -0.8719 + 1.2694 \cdot T_i - 0.2317 \cdot T_i^2$$

Exponentiating these values at each time point yields the desired OR's, which may be observed in Table 7. That is,

$$\exp(M_{a1} - M_{a0}) = \frac{\text{Odds}|X_i = 1}{\text{Odds}|X_i = 0}, \;\; \exp(M_{b1} - M_{b0}) = \frac{\text{Odds}|X_i = 1}{\text{Odds}|X_i = 0}$$

Similarly, the effect of time on the outcome, expressed as the OR's of a 1-unit increase in time (shown in Table 8), are calculated using the $M_{a,T_i+1} - M_{a,T_i}$ and $M_{b,T_i+1} - M_{b,T_i}$ equations below, which are obtained as follows. Let M_{a,T_i+1} and M_{a,T_i} be the marginal models (a) for times $T_i + 1$ and T_i, respectively. Similarly, let M_{b,T_i+1} and M_{b,T_i} be the mixed-effects models (b) for $T_i + 1$ and T_i. Then the log of the OR's for each model for the "next" time point compared to the "current" time point for DMPA dosage X_i are given by:

$$\log\left(\frac{\text{Odds}|T_i + 1}{\text{Odds}|T_i}\right) = M_{a,T_i+1} - M_{a,T_i} = \beta_2 + \beta_3 \cdot (2T_i + 1) + \beta_4 \cdot X_i + \beta_5 \cdot X_i \cdot (2T_i + 1)$$

$$= 0.5088 + 0.0022 \cdot (2T_i + 1) + 0.7045 \cdot X_i - 0.1346 \cdot X_i \cdot (2T_i + 1)$$

$$\log\left(\frac{\text{Odds}|T_i + 1}{\text{Odds}|T_i}\right) = M_{b,T_i+1} - M_{b,T_i} = \beta_2 + \beta_3 \cdot (2T_i + 1) + \beta_4 \cdot X_i + \beta_5 \cdot X_i \cdot (2T_i + 1)$$

$$= 0.3185 + 0.1048 \cdot (2T_i + 1) + 1.2694 \cdot X_i - 0.2317 \cdot X_i \cdot (2T_i + 1)$$

As before, exponentiating these values for each dose level yields the desired OR's, which may be observed in Table 8. That is,

$$\exp(M_{a,T_i+1} - M_{a,T_i}) = \frac{\text{Odds}|T_i + 1}{\text{Odds}|T_i}, \;\; \exp(M_{b,T_i+1} - M_{b,T_i}) = \frac{\text{Odds}|T_i + 1}{\text{Odds}|T_i}$$

Table 6: Marginal & Mixed-Effects Model Coefficients

Terms	Marginal	Mixed-Effects	exp(Marginal)	exp(Mixed-Effects)
Constant	-2.0013	-2.7939	0.1351595	0.0611821
Dose (X_i)	-0.4614	-0.8719	0.6304005	0.4181563
Time (T_i)	0.5088	0.3185	1.6632940	1.3750636
Time^2 (T_i^2)	0.0022	0.1048	1.0022024	1.1104885
Dose$*$Time ($X_i * T_i$)	0.7045	1.2694	2.0228350	3.5587167
Dose$*$Time^2 ($X_i * T_i^2$)	-0.1346	-0.2317	0.8740655	0.7931840

Table 7, specifically, gives the OR's of a high dose (100 mg) to a low dose (150 mg) of DMPA at each time point (3-month interval) in the study for each of the models. That is, it gives each model's predicted odds of experiencing amenorrhea for those in the high dose group compared to those for the low dose group (reference group), in the form of a quotient (Dose=1/Dose=0). Thus, each value represents the relative difference between the two (how much greater odds are for the high dose group than they are for the low dose group) for given time point. For example, the marginal OR of 1.114605 at time point 1, indicates that the odds of experiencing ammenorrhea within the first 3-month period is $\approx 11.46\%$ higher for individuals receiving the

high dose of DMPA than it is for those receiving the low dose. That is to say, individuals in the high dose group have an $\approx 11.46\%$ elevated risk of experiencing ammenorrhea at the first time point (compared to those in the low dose group).

Table 7: Dose Effect on Response (Odds-Ratio: Dose=1/Dose=0)

Time Point	Marginal OR	Mixed-Effects OR
1	1.114605	1.180337
2	1.505613	2.096145
3	1.553794	2.341988
4	1.225073	1.646250

Unlike the odds-ratios calculated for the effects of dose on the response, those shown in Table 8 below expresses the change in the odds of experiencing ammenorrhea for a single unit increase in time at each time point for each DMPA dosage group. That is, for example, the odds of experiencing ammenorrhea in the "next" 3-month interval for those in the low dose group is $\approx 67.44\%$ higher compared to their "current" odds at time point 1. And conversely, the risk of ammenorrhea in the second time point is roughly 2.26 times greater than it is at time point 1 for those in the high dose group. Noticeably, both Tables 7 and 8 give us a general sense for how the relative effects of the covariates on the response change over time, and hence provide a promising means for addressing the research question.

Table 8: Time Effect on Response (Odds-Ratio: 1-unit increase)

Time Point	Marginal OR, Dose=0	Marginal OR, Dose=1	Mixed-Effects OR, Dose=0	Mixed-Effects OR, Dose=1
1	1.674442	2.261843	1.883064	3.344108
2	1.681826	1.735647	2.322165	2.594516
3	1.689242	1.331865	2.863659	2.012947
4	1.696691	1.022019	3.531420	1.561739

Question 5:

What is your conclusion about the influence of DMPA dosage on changes in a woman's risk of amenorrhea over time, based on the results from the:

a) Marginal model?

b) Mixed-effects model assuming a random intercept and a random slope?

Solution

Given the OR's displayed in Tables 7 and 8, which correspond to the marginal and mixed-effects models' estimates, we see that the odds of experiencing amenorrhea is always greater for those in the high dose group compared to the low dose group. However, as can be observed in Table 7, the OR's rate of change is almost always positive. That is, in both models, the OR's increase pretty significantly from time points 1 to 2, and then very slightly from time points 2 to 3, before dropping to almost the same value as time point 1 in time point 4. More than showing us that, given the data, an increase in DMPA dosage is correlated with an increased risk of experiencing amenorrhea, Table 7 gives us an idea for how the magnitude of this risk changes over time. Evidently, the passage of time does appear to exacerbate these odds (even if by a small amount), up until the last 3-month period, where the odds are still greater for individuals receiving the high dose of DMPA, but not as great as they were at time points 2 and 3. Moreover, looking at the effect of time on the response (Table 8), we notice specifically that for those in the low dose group, the OR is almost constant (although slightly increasing) over time (more so in the mixed-effects model). That is, for

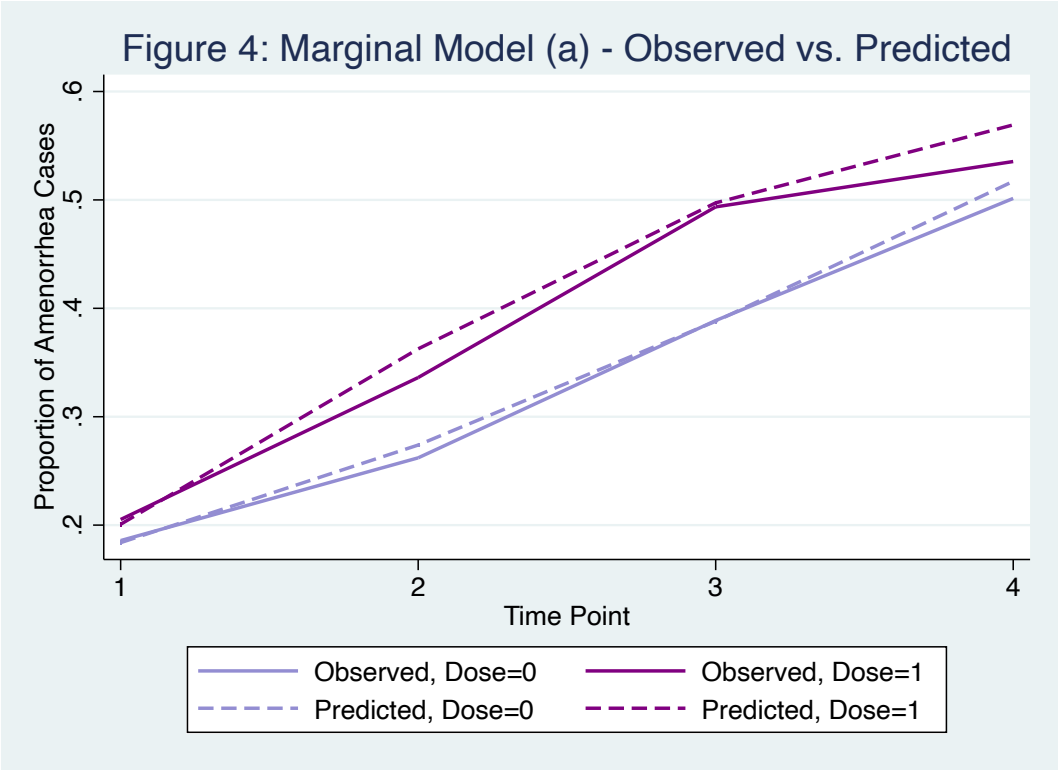
this group, the risk of amenorrhea at the "next" time point is always $\approx 69\%$ larger than it is at the "current" time point, according to the marginal model. On the other hand, as shown in Table 8, there is an almost opposite rate of change in OR's over time for individuals in the high dose group. Specifically, while the OR's given by the mixed-effects model are practically flipped for the high dose group relative to the low dose group, the marginal model shows a more drastic decrease in the high dose group's OR's compared to the increase observed for the low dose group. Still however, the mixed-effects model gives overall more drastic OR values than the marginal model. All that said, given the OR's provided by both models, the data seems to reflect an overall increased risk of amenorrhea for individuals in the high dose group, while the passage of time more greatly impacts those in the low dose group. That is, although receiving a higher dose of DMPA might elevate the odds of amenorrhea, receiving a lower dose of DMPA means that one's risk will increase more rapidly with respect to time.

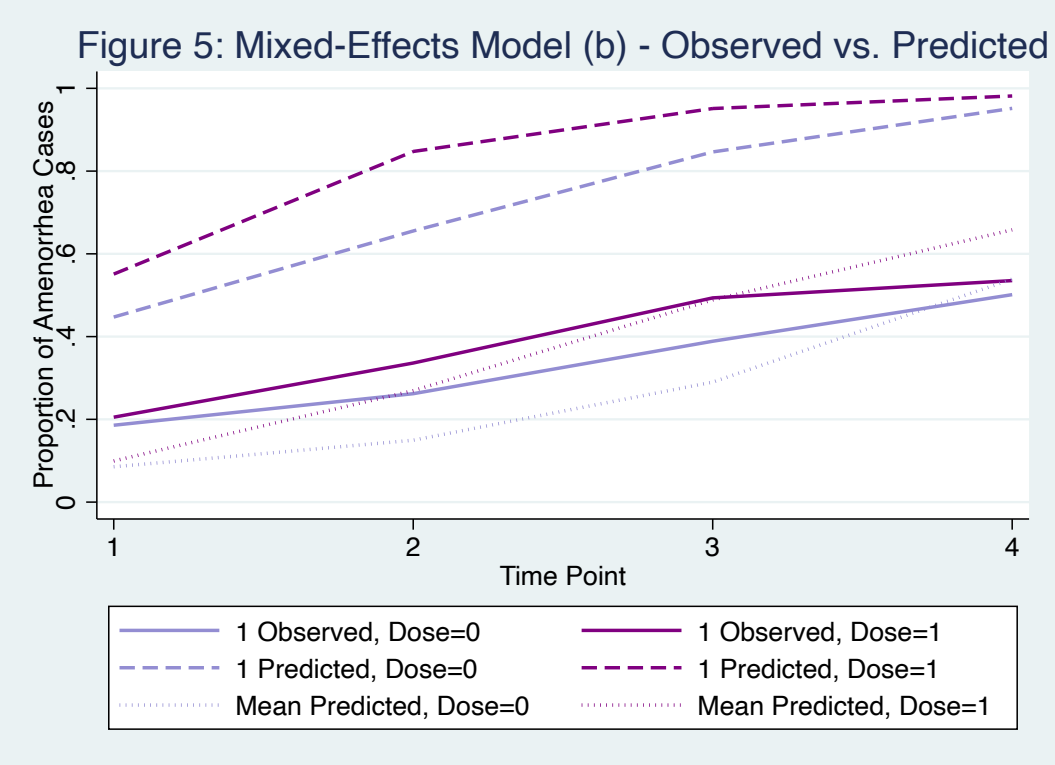
Question 6:

Create one plot for each model in Question 3, to compare observed versus predicted values.

Solution

The following plots of observed and predicted values for marginal and mixed-effects models, given by Figures 4 and 5 below, were generated using STATA.





Question 7:

Which modeling approach would you select for analyzing the data in this study? Explain.

Solution

Not only would a marginal model approach be inappropriate and potentially misleading given the substantial amount of missing data, but since the primary objective of the study is to identify changes in the risk of amenorrhea over time, as well as to determine the influence of DMPA dosage on such changes for individuals, it follows that a mixed-effects model would be better suited to address these research aims. Even if a marginal model provides a better fit to the data available (as seen above), the fact that this data may not be truly representative of the population, in tandem with the fact that we are primarily interested in subject-specific changes in amenorrhea risk, tells us that a mixed-effects model (which sheds light on these presummed between- and within-subject variations in the data) is the more fitting and relevant of the two approaches.

Code

```
#installing tidyverse
library(tidyverse)

#importing "amenorrhea" data
amenorrhea <- read.csv("/Users/antonellabasso/Desktop/PHP2516/DATA/amenorrhea.csv")
head(amenorrhea)
```

```
############################## Descriptive Statistics ##############################

# length of dataset (number of observations)
dim(amenorrhea) #4604

# number of individuals in study
length(unique(amenorrhea$id)) #1151

# number of NA values in the status column
sum(is.na(amenorrhea$status)) #988

# number of (documented) amenorrhea cases
sum(na.omit(amenorrhea$status==1)) #1231

# number of (documented) non-amenorrhea cases
dim(amenorrhea)[1]-sum(is.na(amenorrhea$status))-sum(na.omit(amenorrhea$status==1)) #2385

# df of amenorrhea cases
amenorrhea_cases <- na.omit(amenorrhea[which(amenorrhea$status==1), ])
#head(amenorrhea_cases)

# unique individuals that experienced amenorrhea
length(unique(amenorrhea_cases$id)) #633

# number of amenorrhea cases at each time point
table(amenorrhea_cases$time)

# number of amenorrhea cases for each dose
table(amenorrhea_cases$dose) #D0=572 D1=659

# individuals that experienced amenorrhea 1,2,3, and 4 times (counts)
n_occur <- data.frame(table(amenorrhea_cases$id))
nrow(n_occur[n_occur$Freq==1, ]) #276
nrow(n_occur[n_occur$Freq==2, ]) #174
nrow(n_occur[n_occur$Freq==3, ]) #125
nrow(n_occur[n_occur$Freq==4, ]) #58
```

```
####################################### Table 1 ########################################

# STATA Table
time_points <- c(1, 2, 3, 4)
dose_0_props <- c(0.1857639, 0.2620545, 0.3887531, 0.5013850)
dose_1_props <- c(0.2052174, 0.3361345, 0.4935733, 0.5354108)
tot_cases <- c(225, 285, 351, 370)

prop_stata <- data.frame(tps=time_points, d0=dose_0_props,
```

```
                        d1=dose_1_props, tc=tot_cases)
colnames(prop_stata) <- c("Time Point", "Dose 0", "Dose 1", "Total Cases")

knitr::kable(prop_stata, format="markdown",
             caption="Proportions of Cases by Dose & Count per Time Point",
             digits=4)
```

```
################################### Tables 2-5 ###################################

# Numbers of individuals that experienced amenorrhea at each time point
#(or combination of time points) for those who experienced it
#1, 2, 3, or all 4 times by dose

# Table 2: Time point counts for 1 amenorrhea case by dose
times1 <- c()
doses1 <- c()
for (i in n_occur[n_occur$Freq==1, ]$Var1){
  t <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$time
  d <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$dose
  times1 <- c(times1, t)
  doses1 <- c(doses1, d)
}
times_1 <- data.frame(times1, doses1) %>%
  group_by_all() %>%
  summarise(Count=n()) %>%
  pivot_wider(names_from=doses1, values_from=Count) %>%
  rename("First Case"=times1, "Dose_0"="0", "Dose_1"="1") %>%
  mutate(Total=Dose_0+Dose_1) %>%
  rename("Dose 0"="Dose_0", "Dose 1"="Dose_1")

# Table 3: Time point (combination) counts for 2 amenorrhea cases by dose
times2 <- c()
for (i in n_occur[n_occur$Freq==2, ]$Var1){
  t <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$time
  d <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$dose
  fs <- c(t[1], t[2], d[1])
  times2 <- c(times2, fs)
}
times2_m <- matrix(times2, ncol=3, byrow=TRUE)
times_2 <- as.data.frame(times2_m) %>%
  group_by_all() %>%
  summarise(Count=n()) %>%
  pivot_wider(names_from=V3, values_from=Count) %>%
  rename("First Case"=V1, "Second Case"=V2, "Dose_0"="0", "Dose_1"="1") %>%
  mutate(Total=Dose_0+Dose_1) %>%
  rename("Dose 0"="Dose_0", "Dose 1"="Dose_1")

# Table 4: Time point (combination) counts for 3 amenorrhea cases by dose
times3 <- c()
for (i in n_occur[n_occur$Freq==3, ]$Var1){
  t <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$time
  d <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$dose
  fst <- c(t[1], t[2], t[3], d[1])
  times3 <- c(times3, fst)
```

```
}
times3_m <- matrix(times3, ncol=4, byrow=TRUE)
times_3 <- as.data.frame(times3_m) %>%
  group_by_all() %>%
  summarise(Count=n()) %>%
  pivot_wider(names_from=V4, values_from=Count) %>%
  rename("First Case"=V1, "Second Case"=V2, "Third Case"=V3,
         "Dose_0"="0", "Dose_1"="1") %>%
  mutate(Total=Dose_0+Dose_1) %>%
  rename("Dose 0"="Dose_0", "Dose 1"="Dose_1")

# Table 5: Time point count for 4 amenorrhea cases by dose
doses4 <- c()
for (i in n_occur[n_occur$Freq==4, ]$Var1){
    d <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$dose
    fstf <- c(d[1])
    doses4 <- c(doses4, fstf)
}
times4 <- c(1, 2, 3, 4, length(doses4)-sum(doses4), sum(doses4), length(doses4))
times4_cols <- c("First Case", "Second Case", "Third Case", "Fourth Case",
                 "Dose 0", "Dose 1", "Total")
times_4 <- data.frame(times4, times4_cols) %>%
  pivot_wider(names_from=times4_cols, values_from=times4)
```

```
################################### Figure 2 ########################################

# Bar Graph - Count of Amenorrhea Cases per Time Point by Dose
# Individuals with Only 1 Case

times12 <- c()
doses12 <- c()
for (i in n_occur[n_occur$Freq==1, ]$Var1){
  t <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$time
  d <- amenorrhea_cases[which(na.omit(amenorrhea_cases$id==i)), ]$dose
  times12 <- c(times12, t)
  doses12 <- c(doses12, d)
}
times_12 <- data.frame(times12, doses12) %>%
  group_by_all() %>%
  summarise(Count=n())

ggplot(times_12, aes(fill=as.character(doses12), y=Count, x=times12)) +
  geom_bar(position="stack", stat="identity") +
  labs(title="Figure 2: Count of Amenorrhea Cases per Time Point by Dose",
       subtitle="Individuals with Only 1 Case",
       y="Count",
       x="Time Point",
       fill="Dose") +
  theme(plot.title=element_text(face="bold"),
        plot.subtitle=element_text(hjust=0.5)) +
  scale_fill_manual(values=c("#9999DD", "#770066"))
```

```
################################### Figure 3 ########################################
```

```r
# Distribution of Difference in Dose Counts (Dose 1 - Dose 0)
diffs <- as_vector(c(times_1[3]-times_1[2], times_2[4]-times_2[3],
           times_3[5]-times_3[4], times_4[6]-times_4[5]))
hist(diffs,
     main="Figure 3: Distribution of Difference in Dose Counts",
     xlab="Dose 1 - Dose 0",
     col="lightblue") #482482
abline(v=mean(diffs), lty=2, col="red") # 2.87
# sd(diffs) # 6.49
```

```r
############################ Coefficient Interpretations ############################

# Table 6
# Marginal & Mixed-Effects Models Coefficients and exp(Coefficients)
terms <- c("Constant", "Dose ($X_i$)", "Time ($T_i$)", "Time$^2$ ($T_i^2$)",
           "Dose$*$Time ($X_i*T_i$)",
           "Dose$*$Time$^2$ ($X_i*T_i^2$)")
marginal1 <- c(-2.0013, -0.4614, 0.5088, 0.0022, 0.7045, -0.1346)
mixed_effects1 <- c(-2.7939, -0.8719, 0.3185, 0.1048, 1.2694, -0.2317)
coeffs <- data.frame(terms=terms,
                     marginal1=marginal1, mixed_effects1=mixed_effects1,
                     exp_marginal1=exp(marginal1),
                     exp_mixed_effects1=exp(mixed_effects1)) %>%
  rename("Terms"=terms,
         "Marginal"=marginal1, "Mixed-Effects"=mixed_effects1,
         "exp( Marginal )"=exp_marginal1,
         "exp( Mixed-Effects )"=exp_mixed_effects1)

#################################### General ####################################

# Marginal Model Function - General
marg_fun <- function(x, t, t2=t^2){
  y <- -2.0013+(-0.4614*x)+(0.50888*t)+(0.0022*t2)+(0.7045*x*t)+(-0.1346*x*t2)
  prob <- exp(y)/(1+exp(y))
  return(y)
  #return(exp(y))
  #return(prob)
  #return(logit(prob))
  #return(log(y/(1-y)))
}

# Mixed-Effects Model Function - General
me_fun <- function(x, t, t2=t^2){
  y <- -2.7939+(-0.8719*x)+(0.3185*t)+(0.1048*t2)+(1.2694*x*t)+(-0.2317*x*t2)
  return(y)
}

t <- seq(1, 4, 0.1)

# Marginal Graphs
y_0 <- marg_fun(0, t)
y_1 <- marg_fun(1, t)
# plot(t, y_0, type='l')
# plot(t, y_1, type='l')
```

```r
# Mixed-Effects Graphs
y_0b <- me_fun(0, t)
y_1b <- me_fun(1, t)
# plot(t, y_0b, type='l')
# plot(t, y_1b, type='l')

# Response/Outcome Interpretations - General
time_points <- seq(1, 4, 1)
my_0 <- marg_fun(0, time_points) # marginal model, dose=0, all time points
my_1 <- marg_fun(1, time_points) # marginal model, dose=1, all time points
mey_0 <- me_fun(0, time_points) # mixed-effects model, dose=0, all time points
mey_1 <- me_fun(1, time_points) # mixed-effects model, dose=1, all time points

exp_y <- data.frame(t=time_points,
                    my_0=my_0, my_1=my_1,
                    mey_0=mey_0, mey_1=mey_1) %>%
  mutate(m_lor=(my_1/my_0), me_lor=(mey_1/mey_0),
         m_or=exp(m_lor), me_or=exp(me_lor)) %>%
  rename("Time Point"=t,
         "Marginal, Dose=0"=my_0, "Marginal, Dose=1"=my_1,
         "Mixed_Effects, Dose=0"=mey_0, "Mixed-Effects, Dose=1"=mey_1,
         "Marginal log(OR)"=m_lor, "Mixed-Effects log(OR)"=me_lor,
         "Marginal OR"=m_or, "Mixed-Effects OR"=me_or)

################################### Dose Effect #####################################

# Marginal Model Function for Dose Effect (only terms with Dose)
marg_fun_dose <- function(x=1, t, t2=t^2){
  y <- (-0.4614*x)+(0.7045*x*t)+(-0.1346*x*t2)
  return(y)
}

# Mixed-Effects Model Function for Dose Effect (only terms with Dose)
me_fun_dose <- function(x=1, t, t2=t^2){
  y <- (-0.8719*x)+(1.2694*x*t)+(-0.2317*x*t2)
  return(y)
}

# Log-OR for Dose Effect - Dose=1/Dose=0
my_dose <- marg_fun_dose(t=time_points) # marginal
mey_dose <- me_fun_dose(t=time_points) # mixed-effects

# Table 7: Dose Effect OR Interpretations
OR_dose <- data.frame(t=time_points,
                      #my_dose=my_dose, mey_dose=mey_dose,
                      exp_my_dose=exp(my_dose), exp_mey_dose=exp(mey_dose)) %>%
  rename("Time Point"=t,
         #"Marginal log(OR)"=my_dose, "Mixed-Effects log(OR)"=mey_dose,
         "Marginal OR"=exp_my_dose, "Mixed-Effects OR"=exp_mey_dose)

################################### Time Effect #####################################

# Marginal Model Function for Time Effect (only terms with Time)
```

```r
marg_fun_time <- function(x, t, t2=t^2){
   y <- (0.50888*t)+(0.0022*t2)+(0.7045*x*t)+(-0.1346*x*t2)
   y2 <- (0.50888*(t+1))+(0.0022*((t+1)^2))+(0.7045*x*(t+1))+(-0.1346*x*((t+1)^2))
  return(y2-y)
}

# Mixed-Effects Model Function for Time Effect (f(t+1)-f(t))
me_fun_time <- function(x, t, t2=t^2){
  y <- (0.3185*t)+(0.1048*t2)+(1.2694*x*t)+(-0.2317*x*t2)
  y2 <- (0.3185*(t+1))+(0.1048*((t+1)^2))+(1.2694*x*(t+1))+(-0.2317*x*((t+1)^2))
  return(y2-y)
}

# Log-OR for Time Effect - (1 unit increase in Time)
my_time_0 <- marg_fun_time(0, time_points) # marginal model, dose=0
my_time_1 <- marg_fun_time(1, time_points) # marginal model, dose=1
mey_time_0 <- me_fun_time(0, time_points) # mixed-effects model, dose=0
mey_time_1 <- me_fun_time(1, time_points) # mixed-effects model, dose=1

# Table 8: Time Effect OR Interpretations
OR_time <- data.frame(t=time_points,
                      #my_time_0=my_time_0, my_time_1=my_time_1,
                      #mey_time_0=mey_time_0, mey_time_1=mey_time_1,
                      exp_my_t0=exp(my_time_0), exp_my_t1=exp(my_time_1),
                      exp_mey_t0=exp(mey_time_0), exp_mey_t1=exp(mey_time_1)) %>%
  rename("Time Point"=t,
         #"Marginal log(OR), Dose=0"=my_time_0,
         #"Marginal log(OR), Dose=1"=my_time_1,
         #"Mixed-Effects log(OR), Dose=0"=mey_time_0,
         #"Mixed-Effects log(OR), Dose=1"=mey_time_1,
         "Marginal OR, Dose=0"=exp_my_t0,
         "Marginal OR, Dose=1"=exp_my_t1,
         "Mixed-Effects OR, Dose=0"=exp_mey_t0,
         "Mixed-Effects OR, Dose=1"=exp_mey_t1)
```