Stat 439: Homework 7

Due Thur 4/28/22 by 11pm in Gradescope

Your name here

Instructions

You are strongly encouraged to use R Markdown to complete your Homework assignments, starting with this file as a template and Knitting to pdf. Submit your homework to Gradescope as a single pdf file.

Part I

The Skin Cancer Prevention Study was a randomized, double-blind, placebo-controlled clinical trial of beta carotene to prevent non-melanoma skin cancer in high-risk subjects[1] A total of 1805 subjects were randomized to either placebo or 50 mg of beta carotene per day for five years. Subjects were examined once a year and biopsied if a cancer was suspected to determine the number of new skin cancers occurring since the last exam. The outcome variable is a count of the number of new skin cancers per year. The outcome was evaluated on 1683 subjects comprising a total of 7081 measurements. The main objective of the analyses is to compare the effects of beta carotene on skin cancer rates.

Variable descriptions:

- `ID`: Unique identifier for each subject
- `Center`: Medical center (1, 2, 3, or 4)
- `Age`: Age (in years) of each subject at the time of randomization to treatments
- `Skin`: Skin type (1 = burns; 0 = otherwise)
- `Gender`: Sex (1 = male; 0 = female)
- `Exposure`: Count of the number of previous skin cancers
- `Y`: Count of the number of new skin cancers per year
- `Treatment`: Treatment (1 = beta carotene; 0 = placebo)
- `Year`: Year of follow-up

The following code will read the data into R.

```
skin <- read.table("http://www.hsph.harvard.edu/fitzmaur/ala2e/skin.dat",
     header=F)
names(skin) <- c("ID", "Center", "Age", "Skin", "Gender",
                 "Exposure","Y","Treatment","Year")
skin$Treatment <- factor(skin$Treatment, levels=c(0,1),
                         labels=c("Placebo", "BetaCarotene"))
skin$Gender <- factor(skin$Gender, levels=c(0,1),
                      labels=c("Female", "Male"))
skin$Skin <- factor(skin$Skin, levels=c(0,1),
                    labels=c("NoBurns", "Burns"))
```

[1]This problem is taken from Exercise 14.2, p. 436, in *Applied Longitudinal Analysis*, 2nd edition, by Fitzmaurice, Laird, and Ware. The original study was published by Greenberg et al., 1989, 1990; see also Stukel, 1993.

1.

Is the data set in "long" or "wide" format? Explain.

2.

Create a plot that summarizes the relationship between the outcome variable Y, the treatment, and year. Write a few sentences summarizing what you learn from the plot.

3.

Create another plot that you think shows something interesting about the relationship between at least two variables in the data set. Write a few sentences summarizing what you learn from the plot.

4.

Consider the following GLMM model:

```
mod_glmm <- glmer(Y ~ Treatment*Year + (1 + Year | ID), family = poisson,
              data=skin)
summary(mod_glmm)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Y ~ Treatment * Year + (1 + Year | ID)
##    Data: skin
##
##      AIC      BIC   logLik deviance df.resid
##   8330.5   8378.5  -4158.2   8316.5     7074
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5646 -0.2954 -0.2356 -0.2180  3.6384
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 2.9422   1.7153
##         Year        0.1441   0.3797   -0.53
## Number of obs: 7081, groups:  ID, 1683
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -2.37389    0.15470 -15.345   <2e-16 ***
## TreatmentBetaCarotene       0.10643    0.15587   0.683    0.495
## Year                       -0.05165    0.04939  -1.046    0.296
## TreatmentBetaCarotene:Year  0.02406    0.04741   0.508    0.612
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmBC Year
## TrtmntBtCrt -0.540
## Year        -0.830  0.400
## TrtmntBtC:Y  0.419 -0.772 -0.516
```

a. Write out the model being fit to the data — this should be the true model (using β's and b's), not the fitted model. Include all model assumptions.
b. Explain why the researchers should be concerned if the `TreatmentBetaCarotene` coefficient differs significantly from zero. Hint: Think about what this coefficient represents and the study design.
c. Write a sentence interpreting the `TreatmentBetaCarotene:Year` estimated coefficient.
d. What is the estimated standard deviation of the random slopes? Explain what this value is measuring in context.
e. Based on the GLMM output, does it seem like beta carotene has a significant effect on skin cancer rates? Justify your answer.
f. What are the subject-specific coefficients for ID 419238? Interpret this individual's subject-specific `TreatmentBetaCarotene:Year` estimated coefficient.

5.

Consider the following GEE model:

```
mod_gee <- gee(Y ~ Treatment*Year, id = ID, family = poisson,
            corstr = "exchangeable", data=skin, )
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
##                (Intercept)      TreatmentBetaCarotene
##                -1.34810280                 0.05850156
##                       Year TreatmentBetaCarotene:Year
##                -0.01192366                 0.03113123
```

```
summary(mod_gee)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Logarithm
##  Variance to Mean Relation: Poisson
##  Correlation Structure:     Exchangeable
##
## Call:
## gee(formula = Y ~ Treatment * Year, id = ID, data = skin, family = poisson,
##     corstr = "exchangeable")
##
## Summary of Residuals:
##        Min         1Q     Median         3Q        Max
## -0.3189117 -0.2893387 -0.2553602 -0.2552488 21.6912674
##
##
## Coefficients:
##                                Estimate Naive S.E.       Naive z Robust S.E.
## (Intercept)                -1.3656621468 0.12017278 -11.364155159  0.11754855
## TreatmentBetaCarotene       0.0606273286 0.16493640   0.367580033  0.15795715
## Year                        0.0001454531 0.03293302   0.004416632  0.03080627
## TreatmentBetaCarotene:Year  0.0322933015 0.04503166   0.717124443  0.04837198
##                                Robust z
## (Intercept)                -11.617856519
## TreatmentBetaCarotene        0.383821365
```

```
## Year                          0.004721541
## TreatmentBetaCarotene:Year   0.667603435
##
## Estimated Scale Parameter:  2.642347
## Number of Iterations:  3
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 1.0000000 0.3779963 0.3779963 0.3779963 0.3779963
## [2,] 0.3779963 1.0000000 0.3779963 0.3779963 0.3779963
## [3,] 0.3779963 0.3779963 1.0000000 0.3779963 0.3779963
## [4,] 0.3779963 0.3779963 0.3779963 1.0000000 0.3779963
## [5,] 0.3779963 0.3779963 0.3779963 0.3779963 1.0000000
```

a. Write out the model being fit to the data — this should be the true model (using β's), not the fitted model. Include all model assumptions.
b. Write a sentence interpreting the `TreatmentBetaCarotene:Year` estimated coefficient. How does your interpretation differ from your interpretation of the same coefficient in the GLMM?

6.

Fit a generalized linear mixed effects model with a random intercept (no random slope), with linear trends for the log rate over time and allow the slopes to depend on the treatment group, while also controlling for skin type, age, and the count of the number of previous skin cancers. What conclusions do you draw about the effect of beta carotene on the rate of skin cancers after adjusting for skin type, age, and the count of the number of previous skin cancers?

Part II: Cite Sources

Write the sources you used to complete this assignment at the end of your homework submission, adhering to the "Guidance on Citing Sources" bullet points in the collaboration policy section on our course syllabus.