**Spring 2023**

**BIOS 755**

**Homework 5 Solutions**

(a) (15 points) Consider a Poisson-generalized linear mixed model with random intercepts for the subject-specific log rate of skin cancers (logE(Yij jbi)) with time, treatment, and a time by treatment interaction. Find the best way to include time and the

time by treatment interaction.

The best way to determine how to include time and the time\*treatment interaction is to look at the plot of the log transformed means over time. The code below creates this plot.

**proc** **sort** data=skin;

by Treatment year;

**run**;

\*Calculate the mean by week;

**proc** **means** mean data=skin noprint;

by Treatment year;

var Y;

output out = MN\_TRT\_dat mean = mn\_TRT\_Y;

**run**;

\*transform the means to the log scale;

**data** MN\_TRT\_dat;

set MN\_TRT\_dat;

l\_mean = log(mn\_TRT\_Y);

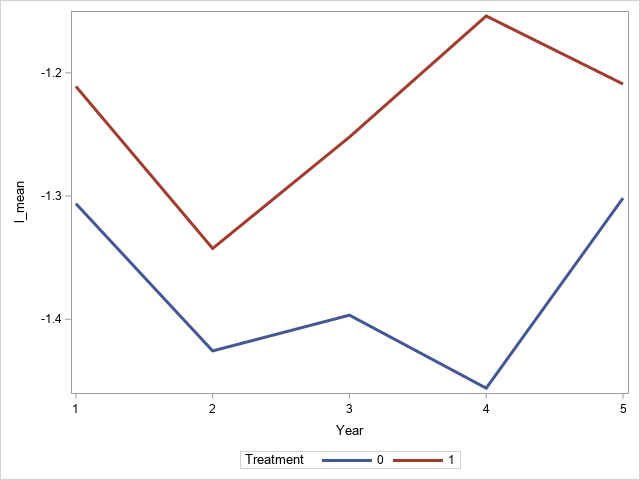
**run**;

\*Let's look at the mean (on the log scale) by TRT group;

**Proc** **SGplot** data = MN\_TRT\_dat;

series x=year y=mn\_TRT\_Y / group =Treatment LineAttrs= (pattern=**1** thickness=**3**);

**run**;



The differences here are small (note the y-axis) so I would probably go with using a linear version of year. However, I think a profile or quadratic are also acceptable versions.

Below is the code to run a glmm with 20 quadrature points. Note that the offset term is not necessary since the data are given as “new skin cancers per year”.

**proc** **glimmix** data=skin method=quad(QPOINTS=**20**);

class ID Treatment (ref='0');

model Y = year Treatment year\*Treatment/d=poisson link=log solution;

random intercept/subject=ID G;

run;

**quit**;

(PROC GLIMMIX) output

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Solutions for Fixed Effects** | | | | | | |
| **Effect** | **treatment** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  | -2.3783 | 0.1095 | 1681 | -21.71 | <.0001 |
| **Year** |  | -0.00010 | 0.02617 | 5396 | -0.00 | 0.9968 |
| **Treatment** | **1** | 0.07928 | 0.1383 | 5396 | 0.57 | 0.5664 |
| **Year\*Treatment** | **1** | 0.03477 | 0.03590 | 5396 | 0.97 | 0.3329 |

The estimated variance of the random intercept

|  |  |  |  |
| --- | --- | --- | --- |
| **Covariance Parameter Estimates** | | | |
| **Cov Parm** | **Subject** | **Estimate** | **Standard Error** |
| **Intercept** | **id** | 2.1452 | 0.1632 |

(b) (5 points) Regardless of significance, give an interpretation for the impact of time on the outcome.

B1= -0.00010

Interpretation: For each subject in the placebo group we estimate that an additional year will result in the mean number of skin cancers decreasing by a factor of exp(-0.00010)=0.9999 (basically no change).

(c) (5 points) Regardless of significance, give an interpretation for time by treatment interaction.

B3 = 0.03477

Interpretation: For each additional year, a subject in the treatment group has a mean number of skin cancers that increase by 3.5% (exp(0.03477)=1.035) more than the same subject’s mean number of skin cancers in the control group.

(d) (5 points) From these results, what conclusions do you draw about the effect of beta and why?

The treatment does not have significant effect on the number of skin cancer over time as evidenced by the non-significant effect of the treatment\*time interaction.

(e) (5 points) Carry out a marginal model analysis using GEE, using the same covariates as in (a). Fit this model with exchangeable and unstructured covariance matrices, then choose the best according to QIC.

marginal model: UN has the smallest QIC hence the best for the data in hand

Exchangeable covariance structure

|  |  |
| --- | --- |
| **GEE Fit Criteria** | |
| QIC | 3356.1836 |
| QICu | 3352.3145 |

Unstructured covariance structure

|  |  |
| --- | --- |
| **GEE Fit Criteria** | |
| QIC | 3319.5735 |
| QICu | 3316.5052 |

The unstructured model fits the data the best and will be used going forward.

(f) (5 points) Regardless of significance, give an interpretation for the impact of time on the outcome.

Effect of beta-carotine from a marginal model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis Of GEE Parameter Estimates** | | | | | | | |
| **Empirical Standard Error Estimates** | | | | | | | |
| **Parameter** |  | **Estimate** | **Standard Error** | **95% Confidence Limits** | | **Z** | **Pr > |Z|** |
| **Intercept** |  | -1.3597 | 0.1201 | -1.5952 | -1.1242 | -11.32 | <.0001 |
| **year** |  | -0.0037 | 0.0323 | -0.0671 | 0.0597 | -0.12 | 0.9080 |
| **treatment** |  | 0.0490 | 0.1636 | -0.2717 | 0.3697 | 0.30 | 0.7648 |
| **year\*treatment** |  | 0.0327 | 0.0492 | -0.0637 | 0.1291 | 0.66 | 0.5062 |

Comment: treatment not significant over time.

The standard errors of the parameter estimates in GLMM are smaller than those from GEE. In GLMM some variability is explained by the random intercepts as opposed to the population level model that averages out the random effects. All in all, both models lead to the same conclusion on the effect of the treatment over time.

B1= -0.00011

Interpretation: Subjects in the placebo group have their mean number of skin cancers decrease by a factor of exp(-0.0037)=0.996 (basically no change) for each additional year.

(g) (5 points) Regardless of significance, give an interpretation for time by treatment interaction.

B3 = 0.0327

Interpretation: For each additional year, the subjects in the treatment group have a mean number of skin cancers increases 3.3% (exp(0.0327)= 1.033) more than the mean number of skin cancers of subjects in the control group.

(h) (5 points) For the GEE analysis, what do you **conclude** about the effect of beta carotene on skin cancers and why? Comment on the difference (or lack of) with the

random effect model.

The treatment does not have significant effect on the number of skin cancer over time as evidenced by the non-significant effect of the treatment\*time interaction.

(i) (5 points) Which approach do you think is more appropriate, the GEE or GLMM?

Here I think the GLMM is more appropriate since the study is a randomized controlled trial aimed to determine if the treatment leads a **subject** having better skin outcomes. The interest is not on the population level, so GEE is not appropriate here.

(j) (5 points) Repeat the random effect model analysis in (a) while appropriately adjusting for skin type and age. What conclusions do you draw about effect of beta

carotene on skin cancers and why?

Adjusted Poisson random intercept model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Solutions for Fixed Effects** | | | | | | |
| **Effect** | **treatment** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  | -4.1375 | 0.3176 | 1678 | -13.03 | <.0001 |
| **year** |  | 0.001093 | 0.02609 | 5396 | 0.04 | 0.9666 |
| **treatment** |  | 0.02768 | 0.1289 | 5396 | 0.21 | 0.8300 |
| **age** |  | 0.01851 | 0.004588 | 5396 | 4.03 | <.0001 |
| **skin** |  | 0.3293 | 0.08685 | 5396 | 3.79 | 0.0002 |
| **exposure** |  | 0.1866 | 0.01045 | 5396 | 17.86 | <.0001 |
| **year\*treatment** |  | 0.03644 | 0.03578 | 5396 | 1.02 | 0.3086 |

Still treatment is not significant over time after adjusting for age, skin type and previous exposure to skin cancers.

(k) (5 points) Does the model in (a) or (j) fit the data better? Give some quantification as to how you made your choice.

All model fit statistics (AIC/BIC/etc.) show that the adjusted model fits the data better than the unadjusted model. This is probably due to the inclusion of age, skin, and exposure, all of which are strongly related to the outcome.

**For the rest of the questions, we will slightly change the outcome variable.**

**A colleague recommends that the** Y **variable be modeled as a** 0=1 **outcome**

**instead of a count. That is, create a new** Y **variable, say** Y \_**, where** Y \_ =

**min**(Y; 1)**.**

(l) (15 points) Rerun the model in (a) using Y \_ with and a logistic instead of a Poisson model. Find the best way to include time and the time by treatment interaction.

Similar to (a), the best way to determine how to include time and the time\*treatment interaction is to look at the plot of the logit transformed means over time. The code below creates this plot.

**data** skin01;

set skin;

Y\_ast = Y;

if Y gt **0** then Y\_ast = **1**;

**run**;

**proc** **sort** data=skin01;

by Treatment year;

**run**;

\*Calculate the mean by week;

**proc** **means** mean data=skin01 noprint;

by Treatment year;

var Y\_ast;

output out = MN\_TRT\_dat mean = mn\_TRT\_Y\_ast;

**run**;

\*transform the means to the logit scale;

**data** MN\_TRT\_dat;

set MN\_TRT\_dat;

l\_mean = log(mn\_TRT\_Y\_ast/(**1**-mn\_TRT\_Y\_ast));

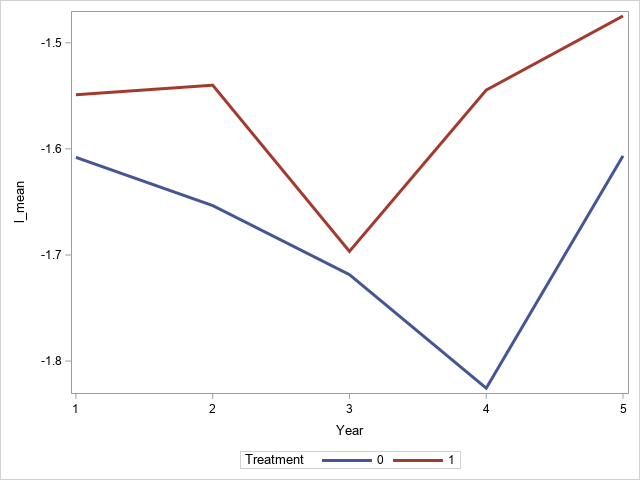
**run**;

\*Let's look at the mean (on the logit scale) by TRT group;

**Proc** **SGplot** data = MN\_TRT\_dat;

series x=year y=l\_mean / group =Treatment LineAttrs= (pattern=**1** thickness=**3**);

**run**;



Here, I think a quadratic version is probably the best. Profile analysis would also be acceptable.

With the new outcome and a logistic model (with quadratic time), the parameter estimates from a model with 20 quadrature nodes are:

| **Solutions for Fixed Effects** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **Treatment** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  | -1.9906 | 0.2524 | 1681 | -7.89 | <.0001 |
| **Year** |  | -0.2716 | 0.1941 | 5394 | -1.40 | 0.1618 |
| **Year\*Year** |  | 0.04376 | 0.03334 | 5394 | 1.31 | 0.1894 |
| **Treatment** | **1** | 0.03017 | 0.3455 | 5394 | 0.09 | 0.9304 |
| **Treatment** | **0** | 0 | . | . | . | . |
| **Year\*Treatment** | **1** | 0.05142 | 0.2693 | 5394 | 0.19 | 0.8486 |
| **Year\*Treatment** | **0** | 0 | . | . | . | . |
| **Year\*Year\*Treatment** | **1** | -0.00140 | 0.04630 | 5394 | -0.03 | 0.9759 |
| **Year\*Year\*Treatment** | **0** | 0 | . | . | . | . |

(m) (5 points) Regardless of significance, give an interpretation for the impact of time on the outcome.

Main effect of year = -0.2716

quadratic effect of year = 0. 04376

To interpret the effect of year, we have to say what year we are referring to.

Interpretation for year 1: For each subject in the placebo group we estimate that an additional year at year 1 will result in their odds of skin cancer decreasing by a factor of exp(-0.2716+0.04376)=0.796 (or by 20.4%).

Or

Interpretation for year 5: For each subject in the placebo group we estimate that an additional year at year 5 will result in their odds of skin cancer decreasing by a factor of exp(-0.2716+5\*0.4376)=0.949 (or by 5.1%).

(n) (5 points) Regardless of significance, give an interpretation for time by treatment interaction.

Interaction with year = 0.05142

Interaction with year\*year = -0.0014

To interpret the interaction, we have to say what year we are referring to.

Interpretation of interaction at year 1: For each additional year at year 1, a subject in the treatment group has their odds of skin cancer increased by a factor of 1.051 (exp(0.05142-0.0014)=1.051, or by 5.1%) versus the same subject in the control group.

Or

Interpretation of interaction at year 5: For each additional year at year 5, a subject in the treatment group has their odds of skin cancer increased by a factor of 1.0454 (exp(0.05142-5\*0.0014)=1.0454, or by 4.5%) versus the same subject in the control group.

(o) (5 points) From these results, what **conclusions** do you draw about the effect of beta caroteneon skin cancers and why?

The treatment does not have significant effect on the number of skin cancer over time as evidenced by the non-significant effect of either interaction term.

(p) (5 points) Comment on the differences between the logistic and Poisson GLMM approaches. Which approach would you recommend?

The overall conclusion remains the same between logistic and Poisson approaches.

The argument for the Poisson: I would use the Poisson approach since it does not “throw away data”. That is, the logistic approach treats all values that are greater than or equal to one the same the Poisson approach can differentiate these values from one another. This will give the Poisson greater power.

The argument for the logistic: I would use the logistic approach since it has interpretation that is more clinically meaningful. That is, we are interested in the subjects not having skin cancer (at all). So, that is what we should be modeling. We don’t care (as much) if we reduce the number of skin cancers per year, just the overall probability of getting **any** skin cancer..