Annie Thwing

Advanced Data Analysis

**Project 0: Dental Data Analysis**

**Introduction**

In this project the effect that a gel treatment had on gum disease was analyzed, specifically its effect on pocket depth and attachment loss. Patients in this study were assigned to one of five different treatment groups. There was a control group that was given no gel at all, and then four groups whose subjects were each given a different level of substance in the gel, the lowest of which had no active ingredient, another control group. This study of 130 patients was measured at a baseline value before treatment began and again after 1 year of applying the gel treatment twice a day. Measurements of pocket depth and attachment loss were averaged across numerous sites throughout the mouth at each of these two time points to provide final measurements for each outcome. The hypothesis of interest is whether or not gel treatment results in lower average pocket depth and attachment loss at one year, H0: u1=u2=u3=u4=u5=0.

**Methods**

First, I analyzed and cleaned the data given in order to see what was going on with it, if there were any particular outliers or patterns to watch out for. From looking at histograms and tables of the variables included in the dataset I looked at how the patients were distributed across treatment groups, how the missing values were distributed, and whether or not adjustments to the model needed to be made in light of either of these analyses.

To begin with, it was easy to see that the subjects were split up evenly throughout each treatment group, with 26 patients in each group. Next, I analyzed how the missing values were distributed. Overall, there was 1 missing value for age, 1 missing value for smoking status and 27 missing values at the one year follow up time point, for attachment loss and pocket depth. As can be seen in Figure 1, these missing values were not evenly distributed across treatment groups, with treatment group 5 ending up with missing values for 10 out of its original 26 patients, while the other treatment groups lost significantly fewer subjects. I did not do anything to adjust for these missing values however, as our investigator said it was not necessary.

Other than these missing values, I did not see anything exceptional in the data to watch out for at this stage. I could see that age, race, sex and smoking status were distributed evenly across the treatment groups. I also knew from the information given that the subjects were randomized to their treatment groups, so because of that I chose not to include age, race, sex or smoking status in the model. Additionally, the number of sites looked to be more or less evenly distributed across the treatment groups, and that knowledge combined with knowing that treatment groups were assigned randomly led to my decision to exclude the number of sites from the final model. Finally, I decided to run one model with the baseline as a covariate and one without and compare AIC values to decide on my final model. The AIC value of the model with the baseline included as a covariate was 12.4, while the AIC value of the model without the baseline as a covariate was 27.99 for the attachment loss model. So I chose my final model to include baseline as a covariate as it had the lowest AIC by a very significant amount. I found the AIC values for the pocket depth model and also chose to also include the pocket depth baseline measurement as a covariate.

The last decision I made was choosing to have the model outcome be the difference between the baseline measurement and the final measurement. I chose this as the outcome as the amount that a measurement increases or decreases is more informative than just a number with no other reference value.

Since the investigator asked for the treatment groups to be analyzed separately, I fit them in the model as factor levels so that they would be easily separated out in the results. Once I decided on this model, I fit it so that I could look at the residual diagnostics and make sure that no assumptions were violated.

The models that I fit were: attach\_fit <- lm(attach\_diff ~ factor(trtgroup),data = dental\_data) and pd\_fit <- lm(pd\_diff ~ factor(trtgroup) + pdbase,data = dental\_data). In checking the residuals, I looked at histograms, and different diagnostic plots of the residuals to check that all of the assumptions were met.

The residual plots for the attachment loss model looked pretty good, there appeared to be two significant outliers that likely messed with them a little bit. As can be seen in Figures 2 and 3, the points labeled 67 and 4 look like they skew otherwise good diagnostic plots. I decided that the diagnostic plots still looked good enough to use the model, the homoscedasticity plot doesn’t show any significant pattern, the qqplot follows the line fairly well and the histogram shows residuals that are normally distributed. I do not think that any outliers needed to be removed, but I think that they are worth pointing out.

The residuals for the pocket depth model looked really good. As can be seen in Figures 4 and 5, no major assumption looks like it is being violated, I did not see a reason to try and transform or do anything to the data. Additionally, a histogram of the residuals showed that normality also holds. From what I saw, the residuals looked good and so I used this model to fit the data and interpret results.

**Results**

The results do not show that the gel treatment significantly lowers average pocket depth and attachment loss at one year when testing for association using the models described above.

Average attachment loss is not significantly associated with any of the five treatment groups. We can see that on average attachment loss is decreased by -0.04202 (-0.19317182 , 0.10913933) for those subjects in treatment group 2 (control with gel) as opposed to the control without gel, which is not a significant association (p = 0.5824). We can see that treatment group 3 is not significantly associated (p = 0.1693) with average attachment loss, it is only increased on average by 0.10403 (-0.04508840 ,0.25315585) units from the no gel control group. We can see that treatment group 4 is not significantly associated (p = 0.0841) with average attachment loss, it is only increased on average by 0.13384(-0.01837368 0.28605180) units. Finally, we can see that treatment group 5 is not significantly associated (p = 0.8515) with average attachment loss, it is only increased on average by -0.01537(-0.17788604 ,0.14714895) units.

While most of the treatments do not look associated with average pocket depth, treatment 3 looks like it could be. We can see that on average pocket depth is increased by 0.157688(0.002441242 , 0.312934121) for those subjects in treatment group 3 as opposed to the control, this is a significant association (p = 0.0466) though not in the direction the researchers were hoping for. Treatment group 2 is not significantly associated (p = 0.6700 ) with a change in pocket depth as pocket depth is increased on average by 0.032851(-0.119662206 , 0.185363346) units. Treatment group 4 is not significantly associated (p = 0.0767) with a change in pocket depth as on average pocket depth is increased by 0.141293 (-0.015441466 , 0.298027214) units for those in treatment group 4 as opposed to the no gel control group. Treatment group 5 is not significantly associated (p = 0.7818) as on average pocket depth is decreased by 0.023368 (-0.190387504 0.143651995) units.

**Conclusions**

In conclusion, given this data I would not say that treatment results in a significantly lower average pocket depth and attachment loss at one year. Additionally, treatment group three may be showing the opposite effect, where attachment loss and pocket depth may be slightly increasing in the low treatment group. From this study, this gel treatment does not appear to significantly improve gum disease, and in a certain dosage may even worsen it. A limitation to this study that should be taken into account is that all of this data was collected at one clinic in the Midwest, from a predominantly white population. Additionally, the large amount of NA values in treatment group 5 could be causing a problem as we do not have that large of a sample size once we realize that we are missing a significant proportion of measurements from the different treatment groups.

**Reproducible Research**

<https://github.com/BIOS6623-UCD/bios6623-athwing/tree/master/Project0>

**Figures and Tables**



Figure



Figure



Figure



Figure



Figure