STAA 552: Final Exam Part 4

Your Name Here

Honor Code from Part 1 applies here, too.

Epilepsy (Q20 - Q27)

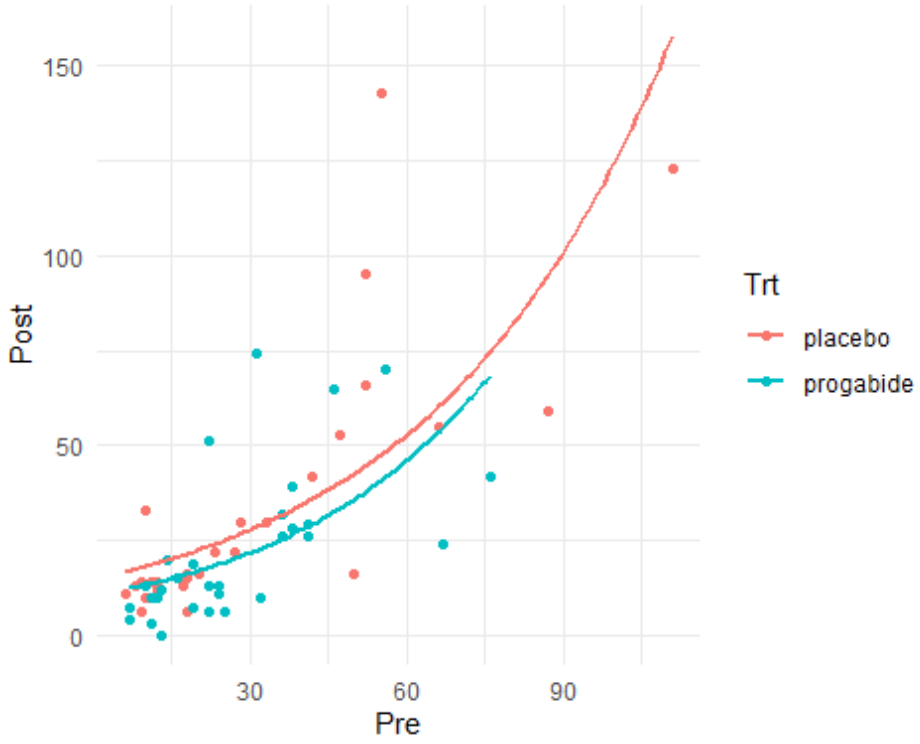
To study the anti-epileptic drug progabide, researchers randomly assigned subjects suffering from epileptic seizures to receive either progabide or placebo. The goal of the study is to evaluate whether progabide is effective in reducing seizures as compared to placebo. We want to control for Age and the number or Pre-treatment seizures. **Post is the response variable for all models.** We will work with n = 58 subjects. The data is available from Canvas as epilepsy.csv.

The data includes the following variables:

- ID: ID number (should NOT be used for model fitting)
- Age: age (in years)
- Trt: placebo or progabide
- Pre: number of epileptic seizures in the 8 weeks prior to start of treatment
- Post: number epileptic seizures in the 8 weeks after start of treatment.

Q20

Create a scatter plot showing number of seizures Post vs Pre treatment with observations color coded by Trt. Overlay separate Poisson regression curves (with log link) for each Trt.



Q21

For **Model 1**, fit a Poisson regression model (with log link), using Post as the response and including Age, Trt and log(Pre) as predictors. Show the coefficients table (including coefficient estimates and Wald test p-values).

```
##
## Call:
## glm(formula = Post ~ Age + Trt + log(Pre), family = poisson,
##     data = epilepsy)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.526008   0.198950  -2.644   0.0082 **
## Age           0.023663   0.003943   6.001 1.96e-09 ***
## Trtprogabide -0.243602   0.052225  -4.664 3.09e-06 ***
## log(Pre)      0.983111   0.037291  26.363  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1284.72  on 57  degrees of freedom
## Residual deviance:  437.44  on 54  degrees of freedom
## AIC: 721.15
```

```
##
## Number of Fisher Scoring iterations: 5
```

Q22

Using **Model 1**, provide a detailed one sentence **interpretation** of the coefficient corresponding to **Trt** in context.

Response

```
## Trtprogabide
##    0.7837997
```

Controlling for age and pre-treatment seizure frequency, patients on progabide have .78 times the post-treatment seizure rate of those on placebo

---

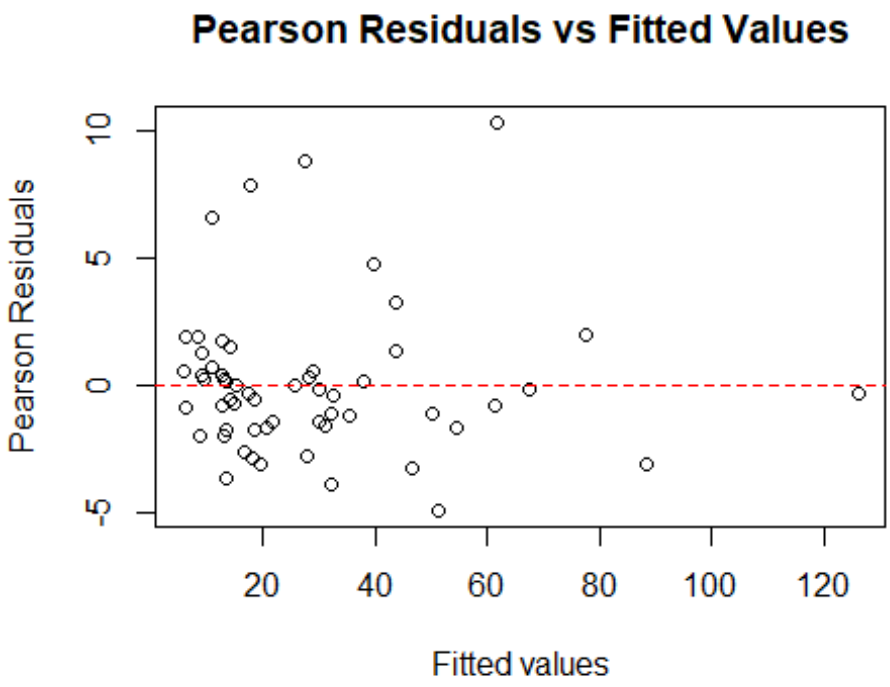
Q23 (2 pts)

Using **Model 1**, briefly discuss whether we should consider simplifying the model by dropping predictors.

Response Both features in model 1 are highly significant so we dont need to simplify the model *****

Q24

Using **Model 1**, do we have evidence of lack of fit (or over-dispersion)? Provide one piece of evidence and briefly discuss. (If multiple pieces of evidence are provided, the first will be graded.)




```
## Proportion of residuals > 2 in absolute value: 0.2586207
```

Discussion: There are more extreme residuals than you would expect.

Q25 (2 pts)

Regardless of your answer to Q24, if we were concerned that there was lack of fit due to model "mis-specification", briefly mention one thing we could try (using available data).

Response We would use a quasipoisson instead of regular poisson *****

Q26

Regardless of your answer to Q24, fit a model similar to Model 1 but allowing for over-dispersion. Show the coefficients table (including coefficient estimates and Wald test p-values). We will call this **Model 2**.

```
##
## Call:
## glm(formula = Post ~ Age + Trt + log(Pre), family = quasipoisson,
##     data = epilepsy)
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.52601    0.59784  -0.880   0.3828
## Age           0.02366    0.01185   1.997   0.0509 .
## Trtprogabide -0.24360    0.15694  -1.552   0.1264
## log(Pre)      0.98311    0.11206   8.773 5.69e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 9.029991)
##
##     Null deviance: 1284.72  on 57  degrees of freedom
## Residual deviance:  437.44  on 54  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Q27 (2 pts)

Based on **Model 2**, what can we conclude about effectiveness of Trt?

Note: Comparing the various epilepsy models, you may find that the conclusion about effectiveness of Trt varies.

Response When we account for over dispersion in the data, we see that the effect is no longer significant. *****

Appendix

```
#Retain this code chunk!!!
library(knitr)
library(tidyverse)
library(ggplot2)
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
#Q20
epilepsy <- read.csv("epilepsy.csv")
epilepsy$Trt <- factor(epilepsy$Trt, levels = c("placebo", "progabide"))


ggplot(epilepsy, aes(x=Pre, y=Post, color=Trt)) +
  geom_point() +
  stat_smooth(method="glm",
              method.args = list(family = poisson(link="log")),
              formula = y ~ x,
              se = FALSE) +
```

```r
  theme_minimal()
#Q21
Model1 <- glm(Post ~ Age + Trt + log(Pre), data = epilepsy, family = poisson)
summary(Model1)
#Q22

exp_coef <- exp(coef(Model1)["Trtprogabide"])
exp_coef
#Q24
pearson_res <- residuals(Model1, type = "pearson")

plot(fitted(Model1), pearson_res,
     xlab = "Fitted values", ylab = "Pearson Residuals",
     main = "Pearson Residuals vs Fitted Values")
abline(h = 0, col = "red", lty = 2)

prop_greater_2 <- sum(abs(pearson_res) > 2) / length(pearson_res)
cat("Proportion of residuals > 2 in absolute value:", prop_greater_2, "\n")
#Q26
Model2 <- glm(Post ~ Age + Trt + log(Pre), data = epilepsy, family =
quasipoisson)
summary(Model2)
```