Final Exam for MATH 354 – Data Analysis I
Darbepoetin alfa for Small-Cell Lung Cancer Patients

Delaney Scheiern
Colgate University
Mathematics
dscheiern@colgate.edu

Abstract

Amgen's new drug darbepoetin alfa, intended to extend life in SCLC patients, is analyzed. The analysis models the treatment effects on change in Hemoglobin levels, fatigue, and time until disease progression.

Keywords: Amgen, SCLC, regression, residuals, mediation analysis

1 Introduction

The company Amgen is developing drugs to treat small-cell lung cancer (SCLC), which is typically fatal when in an advanced stage. Amgen's new drug called darbepoetin alfa aims to extend life in SCLC patients by increasing hemoglobin levels. Darbepoetin alfa has already been shown to increase hemoglobin concentrations in anemic patients. The current study aims to evaluate whether darbepoetin alfa, with platinum-containing chemotherapy, extended survival time in SCLC patients. It also aims to evaluate whether darbepoetin alfa improves the FACT-Fatigue subscale scores in these patients. More information about darbepoetin alfa can be found at https://www.aranesp.com/professional/oncology/.

2 Methods

The data given consists of subjects in randomized treatment and control groups. The treatment group received darbepoetin alfa while the placebo group did not. The study consisted of only White subjects, with approximately two-thirds male. The majority of subjects were between 50 and 70 years old.

3 Results

3.1 Models

Below, models are created and analyzed to determine whether the drug helps with change in hemoglobin, FACT score, and PDDY. Additionally, important code is presented in Appendix 5. Data is processed and plots are created using `tidyverse` (Wickham et al., 2019) and `patchwork` (Pedersen, 2020), quantile models are created using `quantreg` (Koenker, 2020), and mediation analysis is conducted using `mediation` (Imai et al., 2010).

3.1.1 Hemoglobin

First we check how the drug influences the change in hemoglobin for patients by considering their treatment group, sex, baseline weight category, baseline hemoglobin, and the interaction between treatment group and sex. The assumptions for a linear model are visually checked in Figure 1.

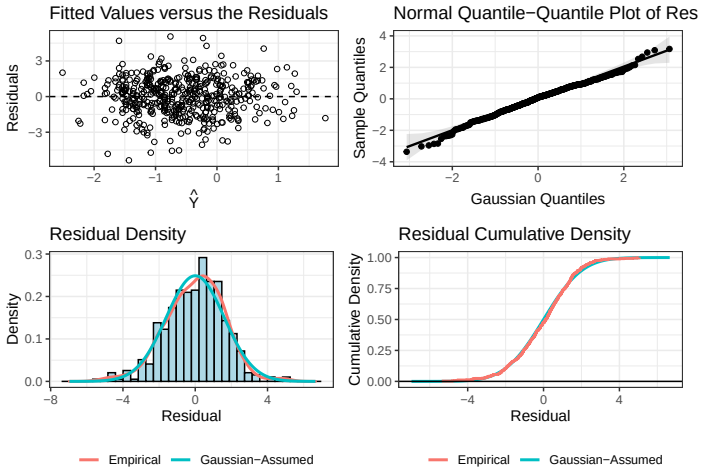

Figure 1: Change in Hemoglobin Model Residuals

Paired with a p-value far below 0.05, the model was found to be significant. Specifically, the treatment group and base hemoglobin levels were both significant. Those that were in the placebo group had a more negative change in hemoglobin levels than those in the treatment group. Those with higher base hemoglobin levels also had a more negative change in hemoglobin levels than those with lower base hemoglobin levels. Thus, Amgen's new drug is helping to maintain hemoglobin levels in patients so it decreases less due to the disease.

3.1.2 FACT Fatigue

Next, this analysis checks how darbepoetin alfa influences the FACT Fatigue score for patients. A regression model using the treatment group, sex, baseline weight category, baseline hemoglobin, change in hemoglobin, and the interaction between treatment group and sex is created. The residual assumptions are checked in Figure 2.

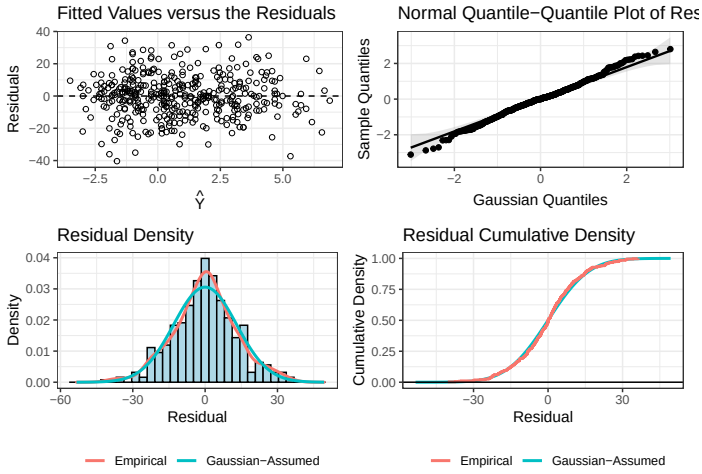

Figure 2: FACT score Model Residuals

Our model overestimates the error variance since the Gaussian curve shows more dispersion than the empirical density estimate, though slight. Since our assumptions do not hold, we cannot analyze this model. Instead, we can choose to fit a quantile regression model, which does not make assumptions about the underlying distribution in the errors. In the quantile regression model, none of the variables are significant. Thus, there is no evidence that Amgen's new drug influences fatigue.

3.1.3 Time until Disease Progression

Next, this analysis checks how darbepoetin alfa influences the time until disease progression for patients. A regression model using the the treatment group, sex, baseline weight category, baseline hemoglobin, change in hemoglobin, and the interaction between treatment group and sex is created. The residual assumptions are checked in Figure 3.

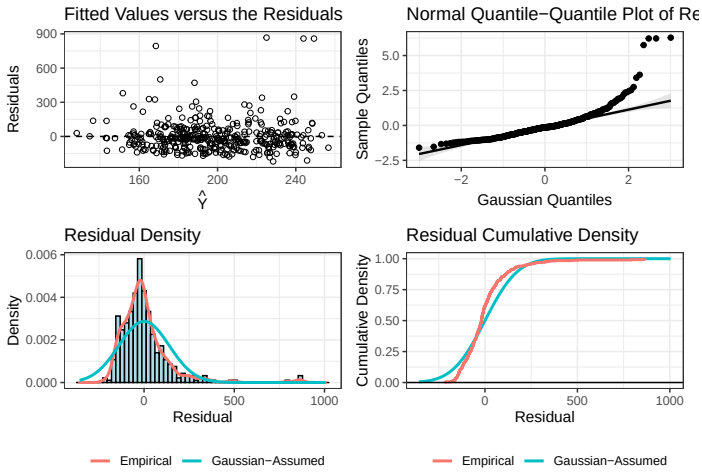

Figure 3: Time Until Disease Progression Model Residuals

The assumptions do not hold to use a least squares model – the residuals are not Gaussian distributed. There also seem to be some unusual points that may be influencing the model. In the residuals versus prediction plot, we see at least four residuals that are far greater than the rest. In the residuals histogram, we see a few points may be causing a stronger right skew. After finding the Cook's distance for each observation with the typical cutoffs for influence, there were not moderately or strongly influential observations. However, using DFFITS with the typical threshold, there were thirteen influential observations. A more thorough analysis of these points should be done to determine if they were recorded in error. Without this extra information, we will continue with the influential observations included. Since our model suffers from heteroskedasticity, we will use a quantile regression model like in Section 3.1.2. None of the variables were statistically significant in this quantile model. Thus, there is no evidence that Amgen's new drug influences time until disease progression.

3.1.4 Mediation Analysis

Next we consider consider mediation analysis, which quantifies the effect of one independent variable "through" another. We analyze whether or not the treatment aids fatigue "through" it's effect on change in hemoglobin levels. Using the `mediation` library, we discover that for female patients, there is an significant increase in the FACT Fatigue "through" the treatment effect on change in hemoglobin levels, shown through a positive and statistically significant Average Causal Mediation Effects (ACME). Also, although not significant, there is a decrease as a direct effect of the treatment. The total effect is not significant, because these effects "cancel" out.

4 Discussion

This analysis shows that darbepoetin alfa maintains patient hemoglobin levels better than chemotherapy alone. There is not evidence that Amgen's new drug improves the FACT Fatigue score or prolongs the time until disease progression. Through mediation analysis, we found that through the treatment effect on change in hemoglobin levels, there is a significant increase in the FACT Fatigue score. Additional research should be done do discover other positive significant effects of darbepoetin alfa. Additional studies with a more representative sample of the population should be conducted as well.

References

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2010). Causal mediation analysis using r. In Vinod, H. D., editor, *Advances in Social Science Research Using R*. Springer-Verlag, New York.

Koenker, R. (2020). *quantreg: Quantile Regression*. R package version 5.74.

Pedersen, T. L. (2020). *patchwork: The Composer of Plots*. R package version 1.0.1.

Wickham, H., Averick, M., and et al (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

5 Appendix

Loaded Data

```
library(tidyverse,quietly = T)
suppressMessages(dat <- read_csv("hemoglobin.csv"))
source("https://cipolli.com/students/code/plotResiduals.R")
library(patchwork)
```

Hemoglobin Model

```
pdf("figure/hemmod.pdf",width=7, height=5)
hem.model <- lm(CHHB ~ TXGROUP + SEX + B_WGTN + B_HGB + TXGROUP*SEX, data=dat)
summary(hem.model)
plotResiduals(hem.model)
dev.off()
```

FACT-Fatigue MLS Model

```
pdf("figure/factmod.pdf",width=7, height=5)
fact.mod <- lm(FATCHG1 ~ TXGROUP + SEX + B_WGTN + B_HGB + TXGROUP*SEX, data=dat)
summary(fact.mod)
plotResiduals(fact.mod)
dev.off()
```

FACT-Fatigue Quantile Model

```
library(quantreg)
mod.f.quantile<-rq(FATCHG1 ~ TXGROUP + SEX + B_WGTN + B_HGB + TXGROUP*SEX, data=dat)
summary(mod.f.quantile,se = "ker")
```

Time until Disease Progression MLS Model & Influential Points

```
library(tidyr)
dat.mod <- dat %>% drop_na(PDDY,TXGROUP,SEX,B_WGTN,B_HGB,CHHB)
pdf("figure/tdpmod.pdf",width=7, height=5)
tdp.mod <- lm(PDDY ~ TXGROUP + SEX + B_WGTN + B_HGB + CHHB + TXGROUP*SEX, data=dat)
summary(tdp.mod)
plotResiduals(tdp.mod)
dev.off()
dat.mod %>% mutate(cook.d=cooks.distance(tdp.mod))%>%
  filter(cook.d>.5 | cook.d>1)
dat.mod %>% mutate(dffits=dffits(tdp.mod))%>%
  filter(abs(dffits)>2*sqrt((6+2)/(386-6-2)))
```

Time until Disease Progression Quantile Model

```
library(quantreg)
mod.pddy.quantile<-rq(PDDY ~ TXGROUP + SEX + B_WGTN + B_HGB + CHHB + TXGROUP*SEX, data=dat)
summary(mod.pddy.quantile,se = "ker")
```

Mediation Analysis

```
dat.med <- dat %>% drop_na(TXGROUP,SEX,B_WGTN,B_HGB,CHHB,FATCHG1)
dat.med$TXGROUP <- factor(dat.med$TXGROUP, levels=c("PLACEBO","NESP"))
library(mediation)
med.fit <- lm(CHHB ~ TXGROUP + SEX + B_WGTN + B_HGB, data=dat.med)
out.fit <- lm(FATCHG1 ~ CHHB + TXGROUP + SEX + B_WGTN + B_HGB, data=dat.med)
med.out <- mediate(med.fit, out.fit, treat = 'TXGROUP', mediator = 'CHHB')
summary(med.out)
```