# Abstract

The purpose of this study was to examine the influence of various factors on time to adoption for cats at animal shelters. A Cox proportional hazards model was used to assess the relationship between stray status, primary color, and gender and time to adoption for N = 57,404 adoptable cats based on the Louisville Animal Services Intake and Outcome open dataset. Tan color and male gender were associated with shorter time to adoption. Being a stray, unknown gender, and the combination of being black and male was associated with longer time to adoption. Black color itself was not associated with time to adoption.

# Introduction

Across the globe, many cats are taken in by animal shelters each year. Cats represent a large portion of animals at shelters offered for adoption. Due to limited resources, many shelters become stretched to overcapacity and are forced to euthanize cats to make room. Higher rates of adoption can free up space for shelters to house more cats, preventing euthanization and saving lives. An understanding of the relationship between cats’ demographic characteristics and how they affect time to adoption may inform shelters of what groups are at risk for longer shelter stays. Such information might allow shelter personnel to channel more effort into promoting adoption for the cats who need help the most.

While potential cat owners take into account a myriad of considerations before adopting a cat, evidence suggests that demographic and aesthetic characteristics play a large role in their decisions.

Stray cats have been found to have longer lengths of stay in animal shelters. In their analysis of five years of data from the Guelph Humane Society, Janke et al. (2017) observed that compared to cats that were surrendered by their owners, stray cat had an 11% lower hazard rate of adoption, suggesting that stray cats have a longer time to adoption than non-strays.

A study by Weiss et al. in 2012 analyzing responses from an adopter survey found that appearance was in the top three reasons people chose their pets. Interviewees were reported to prefer kittens to cats mainly for appearance. Given the importance people place on aesthetics, it is unsurprising that color has been found to influence cat adoption outcomes. It has been shown that black cats experience longer time to adoption than cats of other colors. A 2013 analysis by Kogan, Schoenfeld-Tacher, and Hellyer of data from two Colorado animal shelters showed that black cats take anywhere from 2 to 6 more days to be adopted than cats of other colors and that black kittens take anywhere from 2 to 4 more days to be adopted than kittens of other colors. The authors mention that negative attitudes persist toward black cats due to their historical association with witches and other superstitions as well as a common belief that black cats are not as photogenic or aesthetically pleasing as other cats. Conversely, other research has shown that people prefer lighter-colored cats. One study from a New York no-kill animal shelter found that young, light colored cats had a shorter time to adoption than older cats with darker-colored coats (Brown & Morgan, 2014).

Research has also revealed that male cats typically have a shorter time to adoption than female cats. Janke et al. (2017) found that the hazard ratio of adoption for male cats vs. female cats was 1.2, indicating male cats had a 20% higher rate of adoption than female cats. Brown and Morgan (2014), Lepper et al. (2002), and Vojtkovská, Voslářová, and Večerek, (2019) have also documented an adoption preference for male cats.

The aim of this study is to investigate which of stray status, color, and gender are related to time to adoption for shelter cats and how these factors influence time to adoption. Based on the literature above, it is expected that at all timepoints, strays will have a lower rate of adoption than non-strays, black cats will have a lower rate of adoption than non-black cats, lighter-colored cats will have a higher rate of adoption than darker-colored cats, and male cats will have a higher rate of adoption than female cats.

# Methods

***Dataset***

The Animal Service Intake and Outcome data provided by the city of Louisville, Kentucky was used for this study. This open data set, updated daily, provides a record of all animals received and transferred by the Louisville Department of Animal Services and their outcomes. The dataset analyzed in this study was downloaded from the Louisville Open Data website on April 20, 2020.

***Study sample***

The study sample consisted of 57,404 cats. Only cats assumed to be eligible for adoption were included in the study, i.e., cats with an outcome type of adoption, euthanization, transfer, or other (e.g., missing). All cats designated for spay, trap, neuter, or release were excluded as they were not part of the adoptable population. Cats with unknown outcomes were also excluded from the study as it was unclear whether they were eligible for adoption at any point.

Table 1 below details the descriptive statistics for this study.

*Table 1: Descriptive statistics*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **%** | **N** | **Total** | **Mean** | **SE** | **Median** | **Med 95%LCL** | **Med 95% UCL** | **Min** | **Max** |
| Stray |  |  |  |  |  |  |  |  |  |  |
| *Stray* | 69.4 | 17,593 |  |  |  |  |  |  |  |  |
| *Non-stray* | 30.6 | 39,811 |  |  |  |  |  |  |  |  |
|  |  |  | 57,404 |  |  |  |  |  |  |  |
| Primary Color |  |  |  |  |  |  |  |  |  |  |
| *Black* | 25.6 | 14,701 |  |  |  |  |  |  |  |  |
| *Brown* | 1.0 | 557 |  |  |  |  |  |  |  |  |
| *Gray* | 13.4 | 7,693 |  |  |  |  |  |  |  |  |
| *Tan* | 1.4 | 805 |  |  |  |  |  |  |  |  |
| *White* | 40.6 | 23,335 |  |  |  |  |  |  |  |  |
| *Other* | 18.0 | 10,313 |  |  |  |  |  |  |  |  |
|  |  |  | 57,404 |  |  |  |  |  |  |  |
| Gender |  |  |  |  |  |  |  |  |  |  |
| *Female* | 37.3 | 21,404 |  |  |  |  |  |  |  |  |
| *Male* | 30.6 | 17,586 |  |  |  |  |  |  |  |  |
| *Unknown* | 32.1 | 18,414 |  |  |  |  |  |  |  |  |
|  |  |  | 57,404 |  |  |  |  |  |  |  |
| Adopted |  |  |  |  |  |  |  |  |  |  |
| *Adopted* | 12.1 | 6,930 |  |  |  |  |  |  |  |  |
| *Not adopted* | 87.9 | 50,474 | 57,404 |  |  |  |  |  |  |  |
| Time to adoption (days) |  |  |  | 103 | 3 | 58 | 57 | 59 | 0 | 1173 |

***Statistical Analysis***

The outcome variable for this study was time to adoption in days. Time to adoption was calculated by subtracting the intake date from the outcome date provided in the data.

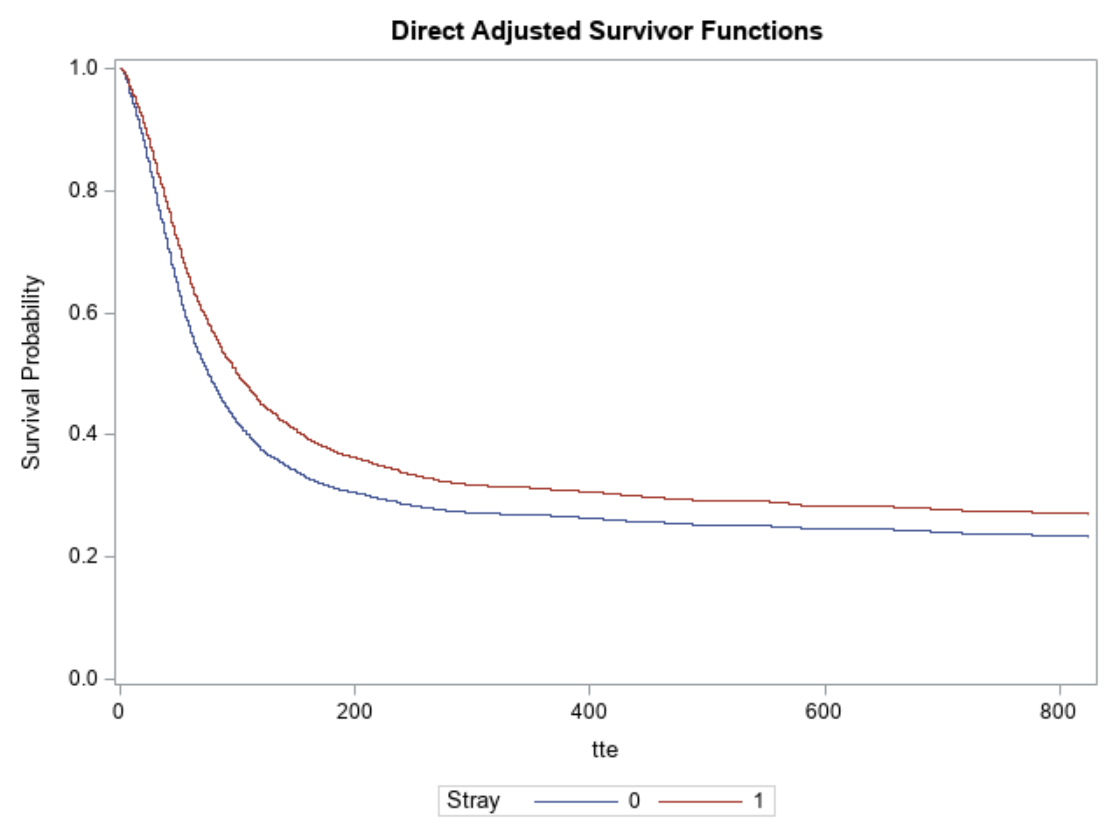
All significant, confounding, and relevant variables were included as predictors in the model. Predictor variables included stray status, primary color, and gender. While age was originally considered for analysis, it was excluded due to having over 90% missing data. Other variables in the dataset were also excluded due to complexity and large amounts of missing data. Please see Appendix 1: Data cleaning for more information on variable selection and reparameterization.

Many animals entered the shelter multiple times. Only the first case for each cat was included in the analysis. Cats with missing intake or outcome dates were excluded.

A Cox Proportional Hazard model was fit to the data to assess the relationship between time to adoption and stray status, primary color, and gender. Potential interactions between the predictor variables were also accounted for by interaction terms. Significance was assessed at the level. Cases for which the outcome type was not adoption (i.e., euthanization, transfer, other) were right-censored. In total, 50,474 cases and 87.9 % of the data were right-censored. Adjusted survival curves were generated for each predictor variable. Cox residuals, deviance residuals, dfbetas, and Schoenfield residuals were then used to assess model fit, potential outliers, influential points, and proportionality respectively. Please see Appendix 2: Diagnostics for details. Analyses were performed using SAS 9.4 and SAS Studio 3.8 as well as R version 3.6.0 and R Studio 1.2.1335.

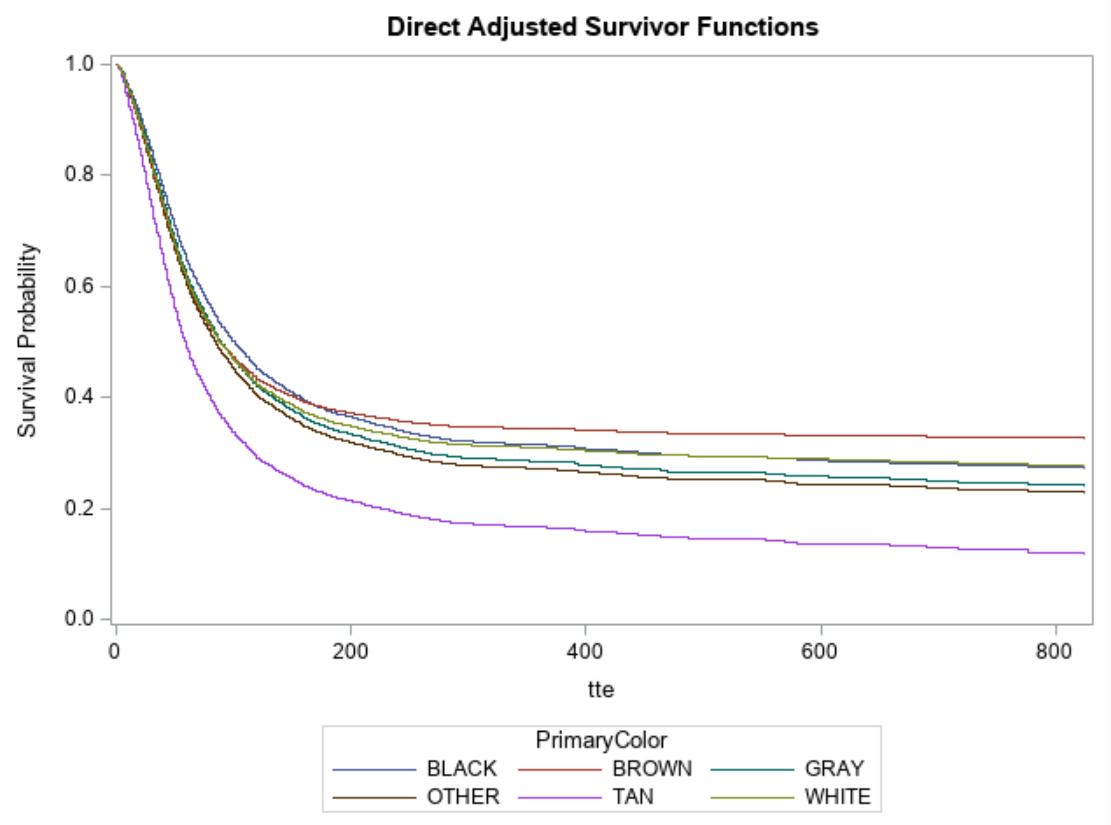
# Results

*Figure 1: Adjusted survival curves for stray status*



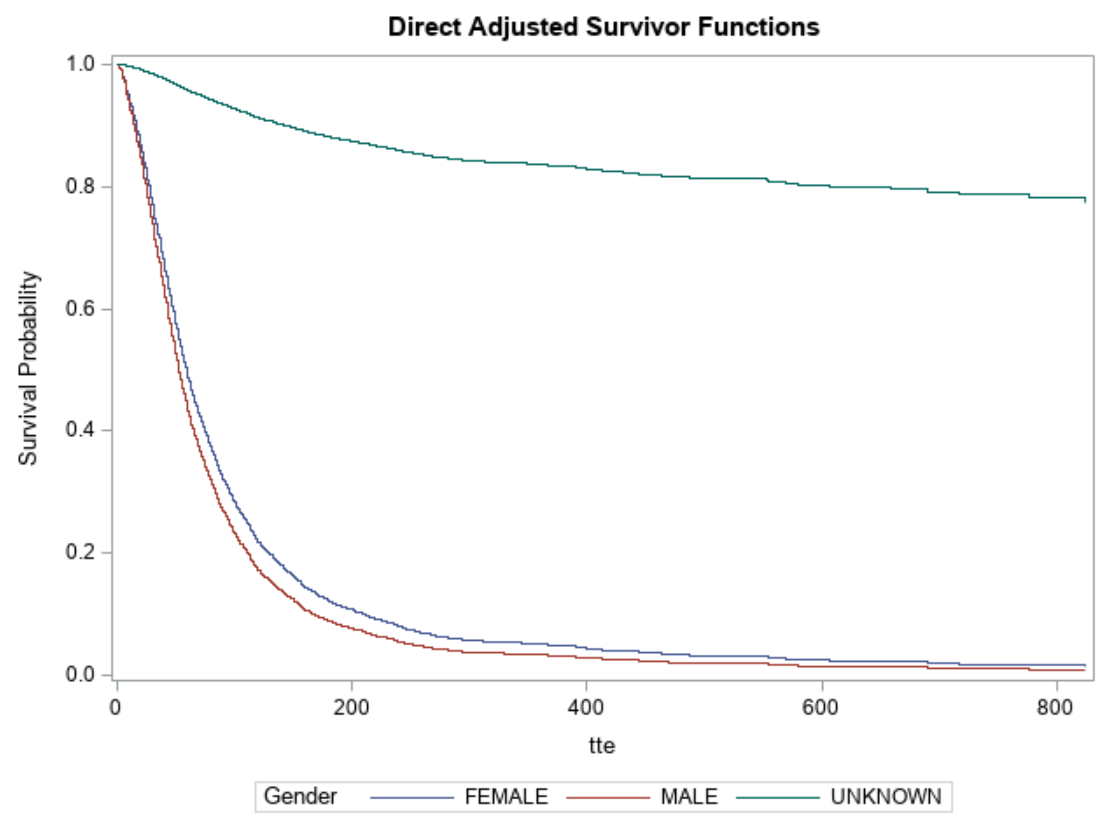
Strays consistently appear to have a higher probability of not being adopted across all time points. The curves are parallel, indicating that the proportional hazards assumption is fulfilled.

*Figure 2: Adjusted survival curves for primary color*



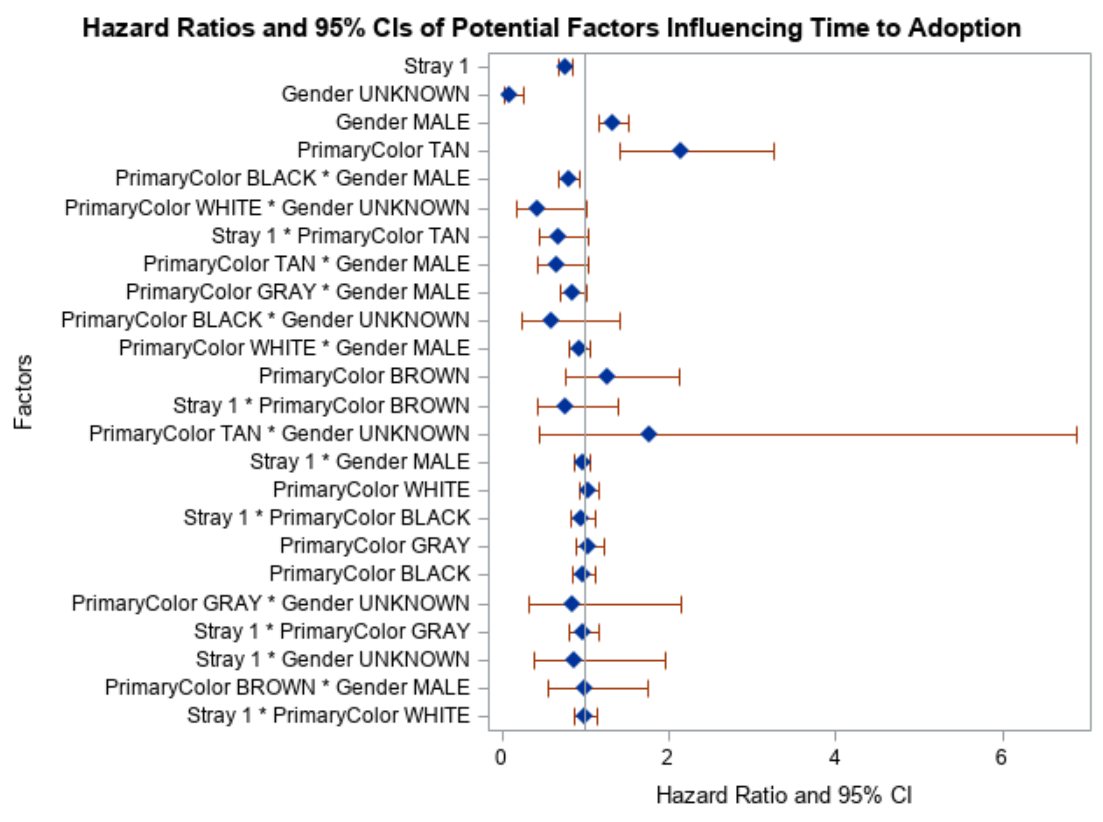
The adjusted survival curves show that tan cats are overwhelmingly the most likely to be adopted across all times. In decreasing order of probability of being adopted, other colored, gray, white, and black cats follow tan cats, but their survival curves are close together, suggesting that their probabilities of being adopted are closer in value. Brown cats are the least likely to be adopted. There appears to be some overlap between the curves, challenging the assumption of proportionality.

*Figure 3: Adjusted survival curves for gender*



Male cats have the highest probability of being adopted, followed closely by female cats. Both male and female cats are much more likely to be adopted than cats with unknown gender, whose probability of not being adopted remains above 80% for most of the time. The curves appear to be relatively parallel, so the proportional hazards assumption seems to be satisfied.

*Figure 4: Forest plot of hazard ratios of potential factors influencing time to cat adoption with 95% confidence intervals*



*Table 2: Estimates of hazard ratios (HR) and 95% confidence intervals for factors of interest*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **LogHR** | **StdErr** | **ChiSq** | **p-value** | **HR** | **95% lower CL** | **95% upper CL** |
| Stray | 1 | -0.27 | 0.06 | 21.92 | <0.001\* | 0.77 | 0.69 | 0.86 |
| Tan | 1 | 0.77 | 0.21 | 13.09 | <0.001\* | 2.15 | 1.42 | 3.26 |
| Black | 1 | -0.03 | 0.07 | 0.14 | 0.71 | 0.98 | 0.85 | 1.11 |
| Male | 1 | 0.29 | 0.07 | 18.33 | <0.001\* | 1.34 | 1.17 | 1.53 |
| Unknown | 1 | -2.30 | 0.50 | 20.89 | <0.001\* | 0.10 | 0.04 | 0.27 |
| Black male | 1 | -0.22 | 0.08 | 7.44 | 0.006\* | 0.80 | 0.69 | 0.94 |

*\*significant at*

An illustration of the hazard ratios and 95% confidence intervals for each variable can be seen in Figure 4. Table 2 provides estimates of hazard ratios and 95% confidence intervals for variables of interest. For estimates of hazard ratios and confidence intervals for all variables and interactions in the study, please refer to Appendix 3: Additional detail.

Being a stray, being tan, being of male or unknown gender, and being black and male all had significant effects on time to adoption. Stray cats had a 23% lower rate of adoption at any time compared to non-strays (HR = 0.77, 95% CI = [0.69-0.86]). Tan cats had 2.15 times the rate of adoption at any time compared to non-tan cats (HR = 2.15, 95% CI = [1.42, 3.26]). Male cats had 1.34 times the rate of adoption at any time compared to female cats (HR = 1.34, 95% CI = [1.17, 1.53]). Cats of unknown gender experienced a 90% lower rate of adoption than female cats (HR = 0.10, 95% CI = [0.04, 0.27]).

Being a black cat had no significant effect on time to adoption (HR = 0.98, 95%CI = [0.85, 1.11]).

There was a significant interaction of being black and male. Compared to black female non-stray cats (baselines), black male cats had a 20% decreased rate of adoption at any time (HR = 0.80, 95% CI = [0.69, 0.94].

# Discussion

Many of the findings of this study overlap with the literature and confirm the hypotheses that at all timepoints, strays will have a lower rate of adoption than non-strays, lighter-colored cats will have a higher rate of adoption than darker-colored cats, and male cats will have a higher rate of adoption than female cats. The large hazard ratio of adoption for tan cats vs. cats of other colors is consistent with the findings by Brown and Morgan (2014) that lighter colored cats experience shorter times to adoption. The finding of longer times to adoption for stray cats follows the same pattern as the research from Janke et al. (2017). Shorter times to adoption for male cats also corroborate the research from Janke et al. (2017) Brown and Morgan (2014), Lepper et al. (2002), and Vojtkovská, Voslářová, and Večerek (2019).

However, it was interesting that there was no main effect of black itself. The negative impact of having a black coat on cat adoption outcomes is well-documented. This result is in direct contrast to Kogan, Schoenfeld-Tacher, and Hellyer’s (2013) finding that black cats have longer time to adoption than cats of other colors, regardless of gender or age. Lepper, Kass, and Hart (2002) also found that black cats had 0.52 of the odds of being adopted compared to other cats. On the other hand, Janke et al. (2017), who used a Cox Proportional Hazards model to examine the hazard ratios for other cats vs. black cats, found that other cats had 1.09 times the hazard rate of adoption compared to black cats but that this difference was not significant. One possibility is that previous studies have increased awareness for the plight of black cats and the findings of this study reflect a consequent reduction in their time to adoption. Another possibility is that much of the large bulk of censored data was comprised of black cats who were at the shelter for a long time but experienced other outcomes such as euthanization.

The finding that black male cats have longer times to adoption was also surprising as being male in general was associated with a shorter time to adoption. One theory is that the demographic that prefers black cats may also have a greater preference for female cats compared to the general population of potential adopters.

The significant main effect of unknown gender more likely reflected a failure to exclude part of the non-target population from the analysis than an actual effect on time to adoption. As the other results and previous research suggest, gender affects people’s decision on whether to adopt an animal and how long an animal stays in the shelter. If an animal is not gendered in the data, it seems likely that they were not eligible for the adoption process at any point.

## *Limitations*

Several limitations should be noted in the data and the analysis. Only the top five animal colors were analyzed, and all other colors were categorized as “Other”. There may have been some relevant color-related effects missed due to this generalization. While “Other” was used as the reference category for the hazard ratios, it might have been more appropriate or interesting to select a different reference color. It was unclear just based on the proportions of cats of each color what the best reference category would be.

Overlap between curves was observed during examination of the adjusted survival curves for primary color, which may suggest a potential violation of the proportional hazards assumptions for this variable. It was difficult to determine from the Schoenfield residual plots if the amount of deviation from a straight line was enough to suggest significant non-proportionality. Corrections for non-proportionality of color were attempted by adding an interaction effect between primary color and adoption as suggested in Borucka (2014). However, this attempt appeared to induce non-proportionality for gender. Another method for controlling for non-proportionality might have been stratification by color. However, the number of variables in the dataset was already limited. Furthermore, the purpose of this research was to examine factors affecting time to adoption for all adoptable shelter cats rather than trends affecting time to adoption by color, and only being tan and the interaction between black and male were significant in the original model. Thus, it seemed more appropriate to use the original model with a broader interpretation. A different type of survival model may have been a more appropriate way to control for this potential violation.

Examination of the deviance residuals as well as the dfbeta plots and tables for each of the factors and interactions indicated the presence of many potential influential outliers. Removing potential influential outliers indicated a >10% change in the estimates for log hazard ratio. Due to the large number of potential influential outliers as well as the possibility that potential outliers may be relevant and informative to the relationship between the factors of interest and shelter outcomes, the outliers were ultimately not removed from the analysis. However, the results should be interpreted with caution.

While the factors selected for examination in this study seem to be important, they certainly do not reflect the scope of all factors that potential cat owners take when deciding to adopt a shelter cat. Within this dataset itself, analyzing the effects of breed and secondary color might have captured more of the decision-making process behind cat adoption. Other considerations not covered by this data include friendliness and disposition (Gourkow, 2001), coat and pattern (Brown &Morgan, 2014), activity (Fantuzzi, Miller, & Weiss, 2010) and socialization (Brown & Stephen, 2014). The factors in the current study may tell part of the story about time to adoption, but other variables outside its scope might be better predictors or act in conjunction with the current variables to explain time to adoption for shelter cats.

# Conclusion

Factors associated with shorter time to adoption for shelter cats include being tan and being male. Factors associated with longer time to adoption include being a stray, having unknown gender, and being black and male. No significant association was found between time to adoption and being a black cat. More research is required to determine what other factors might influence time to adoption and how those factors interact with the characteristics investigated during this study.

# References

Borucka, J. (2014). Extensions of Cox model for non-proportional hazards purpose. *Ekonometria*, (45),

85-101.

Brown, W. P., & Morgan, K. T. (2015). Age, breed designation, coat color, and coat pattern influenced

the length of stay of cats at a no-kill shelter. *Journal of applied animal welfare science*, 18(2),

169-180.

Brown, W. P., & Stephan, V. L. (2020). The influence of degree of socialization and age on length of stay

of shelter cats. *Journal of applied animal welfare science*, 1-8.

Bursac, Z. (2010). *Creating forest plots from pre-computed data using PROC*

*SGPLOT and graph template language*. [Poster]. SAS Global Forum 2010.

<https://support.sas.com/resources/papers/proceedings10/195-2010.pdf>

Fantuzzi, J. M., Miller, K. A., & Weiss, E. (2010). Factors relevant to adoption of cats in an animal

shelter*. Journal of applied animal welfare science*, 13(2), 174-179.

Gourkow, N. (2001). *Factors affecting the welfare and adoption rate of cats in an animal*

*shelter* (Doctoral dissertation, University of British Columbia).

Janke, N., Berke, O., Flockhart, T., Bateman, S., & Coe, J. B. (2017). Risk factors affecting length of stay

of cats in an animal shelter: A case study at the Guelph Humane Society, 2011–2016. *Preventive veterinary medicine*, 148, 44-48.

Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2013). Cats in animal shelters: Exploring the

common perception that black cats take longer to adopt. *The open veterinary science*

*journal*, *7*(1).

Kubesova, K., Voslarova, E., Vecerek, V., & Vucinic, M. (2017). Investigating some of the factors that

affect the selection of shelter cats by adopters in the Czech Republic. *Anthrozoös* , 30, 623–633.

Lepper, M., Kass, P.H., & Hart, L.A. (2002). Prediction of adoption versus euthanasia among dogs and

cats in a California animal shelter. *Journal of applied animal welfare science*. 5, 29–42

Louisville Department of Animal Services (2020). Animal service intake and outcome (Version 20 April

2020). [Data file]. <https://data.louisvilleky.gov/dataset/animal-service-intake-and-outcome>

Vojtkovská, V., Voslářová, E., & Večerek, V. (2019). Comparison of outcome data for shelter dogs and

cats in the Czech Republic. *Animals*, 9(9), 595.

Weiss, E., Miller, K., Mohan-Gibbons, H., & Vela, C. (2012). Why did you choose this pet?: Adopters

and pet selection preferences in five animal shelters in the United States. *Animals*, *2*(2), 144-159.

# Appendix 1: Data cleaning

#packages

library(tidyr)

library(dplyr)

library(lubridate)

library(psych)

#change working directory

setwd("C:/Users/meg/Desktop/Courses/Spring 2020 Courses/Survival Analysis")

#read in the data

animals <- read.csv("animals.csv")

#start with N = 150,842

#get rid of unnecessary columns

#these are the columns relevant for this analysis

cols <- c("AnimalID", "AnimalType","IntakeDate","IntakeType","PrimaryColor","Gender","OutcomeDate","OutcomeType")

animals <- animals%>%

select(cols)

##################################################

#animal ID

#convert to character

animals$AnimalID <- as.character(animals$AnimalID)

# #get rid of the A at the beginning of each ID

# #function that gets rid of A in a string and returns resulting string

# idfun <- function(x){

# id <- gsub("A","",x)

# return(id)

# }

# animals$AnimalID <- sapply(animals$AnimalID, idfun)

# #convert to numeric

# animals$AnimalID <- as.numeric(animals$AnimalID)

###################################################

#animal type

animals$AnimalType <- as.character(animals$AnimalType)

#only interested in dogs and cats

animals <- animals%>%

filter(AnimalType == "CAT"|AnimalType == "DOG")

#N = 145,886

#use dogs as reference category

animals$dog <- ifelse(animals$AnimalType == 'DOG',1,0)

####################################################

#intake date and time

#will need to work with dates, split date and time down the middle

animals <- animals %>%

separate(IntakeDate, c("IntakeDate","IntakeTime"), sep = " ")

#don't think I'll do anything with time

#months/days should suffice

#convert to date type with lubridate package

animals$IntakeDate <- ymd(animals$IntakeDate)

#####################################################

#intake type

animals$IntakeType <- as.character(animals$IntakeType)

#add variable for stray

animals$Stray <- ifelse(animals$IntakeType == "STRAY", 1, 0)

#####################################################

#primary color

animals$PrimaryColor <- as.character(animals$PrimaryColor)

#take only the first color

animals$PrimaryColor <- ifelse(grepl (' ', animals$PrimaryColor),

unlist(strsplit(animals$PrimaryColor, ' '))[1],

animals$PrimaryColor)

#recode chocolate, mahogany, and brindle as brown

animals$PrimaryColor[animals$PrimaryColor == "CHOCOLATE"|animals$PrimaryColor == "MAHOGANY"|

animals$PrimaryColor == "BRINDLE"] <- "BROWN"

#recode fawn and beige and cream as tan

animals$PrimaryColor[animals$PrimaryColor == "FAWN"|animals$PrimaryColor == "BEIGE"|

animals$PrimaryColor == "CREAM"] <- "TAN"

#summary of animal colors

colorcount <- animals%>%

group\_by(PrimaryColor)%>%

summarize(count = n())%>%

arrange(desc(count))

#top 5 are white, black, brown, gray, and tan

#consolidate all others into a category "other"

animals$PrimaryColor[!(animals$PrimaryColor%in%c("WHITE","BLACK","BROWN","GRAY","TAN"))] <- "OTHER"

table(animals$PrimaryColor)

# result of cleaning

# BLACK BROWN GRAY OTHER TAN WHITE

# 41280 16621 10781 23250 9598 44356

#####################################################

#gender

animals$Gender <- as.character(animals$Gender)

#not going to pay attention to spayed/neutered - just male, female, and unknown

#unknown becomes NA

animals$Gender[animals$Gender == "LITTER"|animals$Gender == "UNKNOWN"] <- "UNKNOWN"

animals$Gender[animals$Gender == "NEUTERED MALE"] <- "MALE"

animals$Gender[animals$Gender == "SPAYED FEMALE"] <- "FEMALE"

######################################################

#outcome date and time

#will need to work with dates, split date and time down the middle

animals$OutcomeDate <- as.character(animals$OutcomeDate)

animals <- animals %>%

separate(OutcomeDate, c("OutcomeDate","OutcomeTime"), sep = " ")

#got warning for missing pieces in 198 rows - will have to drop

#these cases at it isn't reasonable to go back and examine each case individually

#convert outcome to date type

animals$OutcomeDate <- ymd(animals$OutcomeDate)

#198 cases of NA

########################################################

#outcome type

animals$OutcomeType <- as.character(animals$OutcomeType)

outcomecount <- animals%>%

group\_by(OutcomeType)%>%

summarize(count = n())%>%

arrange(desc(count))

#we want to examine what features are related to time to adoption

#so it makes sense that we would only want to look at animals for

#whom adoption/euthanization is a possibility

#there is unfortunately no data dictionary, so this represents my best interpretation of the data present

#I had to make sense earlier of how to combine categories, which is why there is all this work

#for just adopted vs. non-adopted

#don't count "no show" because that's basically missing data -> NA

#418 " " -> NA

animals$OutcomeType[animals$OutcomeType == ""|animals$OutcomeType == "NO SHOW"] <- NA

#can't say whether these are part of target population or not so will omit

#relocate/transport can go under "transfer"

animals$OutcomeType[animals$OutcomeType == "RELOCATE"|animals$OutcomeType == "TRANSPORT"] <- "TRANSFER"

#other = et process, indefinite, lost exp, missing, missing ex (members of the adopt/euth process with

#miscellaneous outcomes)

animals$OutcomeType[animals$OutcomeType == "ET PROCESS"|animals$OutcomeType == "INDEFINITE"|

animals$OutcomeType == "LOST EXP"|animals$OutcomeType == "MISSING"|

animals$OutcomeType == "MISSING EX"] <- "OTHER"

animals$OutcomeType[animals$OutcomeType == "ET PROCESS"] <- "OTHER"

animals$OutcomeType[animals$OutcomeType == "INDEFINITE"] <- "OTHER"

#snr, tnr, rtf (spay/trap/neuter release/return to field), release, released probably shouldn't be counted

#because those animals were not going to be up for adoption/euthanization

#exclude after doing the dates

animals <- animals%>%

filter(!(OutcomeType%in%c("RTF","RTO","SNR","TNR","RELEASE","RELEASED")))

#N = 123,135

addmargins(table(animals$OutcomeType))

# ADOPTION DIED DISPOSAL EUTH FOSTER OTHER TRANSFER Sum

# 25117 2455 1700 68358 4574 1359 19107 122670

#N = 123,135

#465 NAs as expected

########################################################

#create time to event variable

#definte interval between intake date and outcome date

tte\_int <- animals$IntakeDate %--% animals$OutcomeDate

#time to event variable in months

animals$tte <- as.duration(tte\_int)/ddays(1)

#filter out time to events < 0 because they don't make sense

animals <- animals%>%

filter(tte >= 0)

#N = 122,897 - loss of 238 to data entry issues

#check tte variable

summary(animals$tte)

#the maximum value of 5849 seems feasible unless there is a typo on intake/outcome date

#checked and the dates look like they were intentional

#one dog was adopted after 16 years

#########################################################

#create censoring variable for adoption

animals$adopted <- ifelse(animals$OutcomeType=="ADOPTION",1,0)

#########################################################

#check number of missing values in each field

print(sapply(animals, function(x){sum(is.na(x))}))

#filter out NAs for outcome type/adopted

animals <- animals%>%drop\_na(adopted)

#N = 122,621 (loss of 276 as expected)

#number of unique IDs

length(unique(animals$AnimalID))

#only 114,589 - 8032 duplicates

#handling shleter repeat animals

#subset the animals who have duplicate IDS

#take first instance of each ID

animalsf <- animals%>%

group\_by(AnimalID)%>%

arrange(OutcomeDate)%>%

slice(1)%>%

ungroup()

View(animalsf)

#N = 114,589 - matches number of unique IDS

#################################################################################################

#Dataset

#N = 114,589

#drop unnecessary variables

cols2 <- c("AnimalID", "AnimalType", "IntakeDate", "IntakeType",

"PrimaryColor", "Gender", "OutcomeDate",

"OutcomeType", "tte", "adopted", "Stray", "dog")

animalsf<- animalsf %>%

select(cols2)

print(sapply(animalsf,function(x){sum(is.na(x))}))

#get rid of NAs just in case

animalsf<- drop\_na(animalsf)

#N = 114,589

#turns out that I'll need ID number to generate the deviance residuals

#get rid of the A at the beginning of each ID

#function that gets rid of A in a string and returns resulting string

idfun <- function(x){

id <- gsub("A","",x)

return(id)

}

animalsf$AnimalID<-sapply(animalsf$AnimalID, idfun)

#convert to numeric

animalsf$AnimalID <- as.numeric(animalsf$AnimalID)

#save out the datafile to a csv

write.csv(animalsf,file = "animalsfinal.csv")

#################################################################################################

#Table 1 Statistics

#subset only cats

cats <- animalsf %>%

filter(dog == 0)

print("Table 1 Stats")

#stray

print("Stray")

stray\_t <- table(cats$Stray)

addmargins(stray\_t)

prop.table(stray\_t)

#color

print("Color")

color\_t <- table(cats$PrimaryColor)

addmargins(color\_t)

prop.table(color\_t)

#gender - female = 1, male = 0

print("Gender")

gen\_t <- table(cats$Gender)

addmargins(gen\_t)

prop.table(gen\_t)

#outcome type

print("Outcome Type")

otype\_t <- table(cats$OutcomeType)

addmargins(otype\_t)

prop.table(otype\_t)

#adopted

print("Adopted")

adopt\_t <- table(cats$adopted)

addmargins(adopt\_t)

prop.table(adopt\_t)

#time to adoption

#had trouble with processing time using formedian estimates with SAS as in class, so decided to #use R for this part since it was Table 1

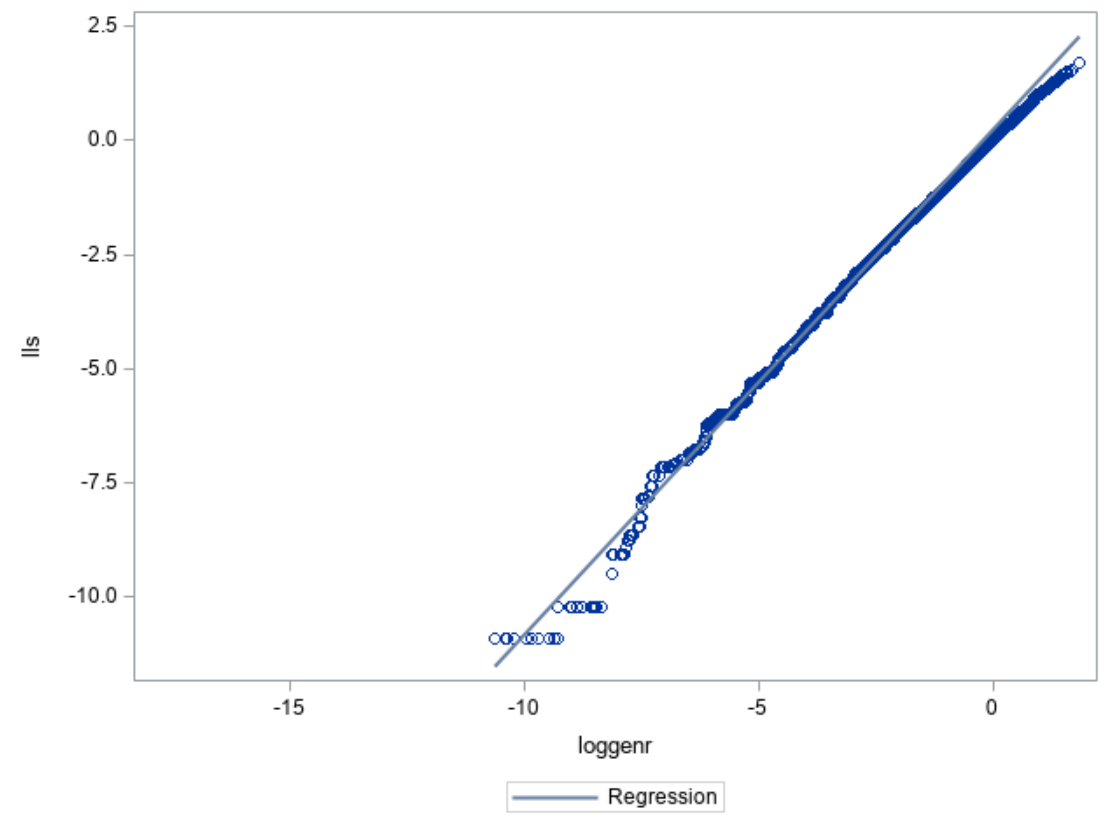
print("Time to adoption")

print(survfit(Surv(tte, adopted) ~ 1, data = cats), print.rmean = T)

summary(cats$tte)

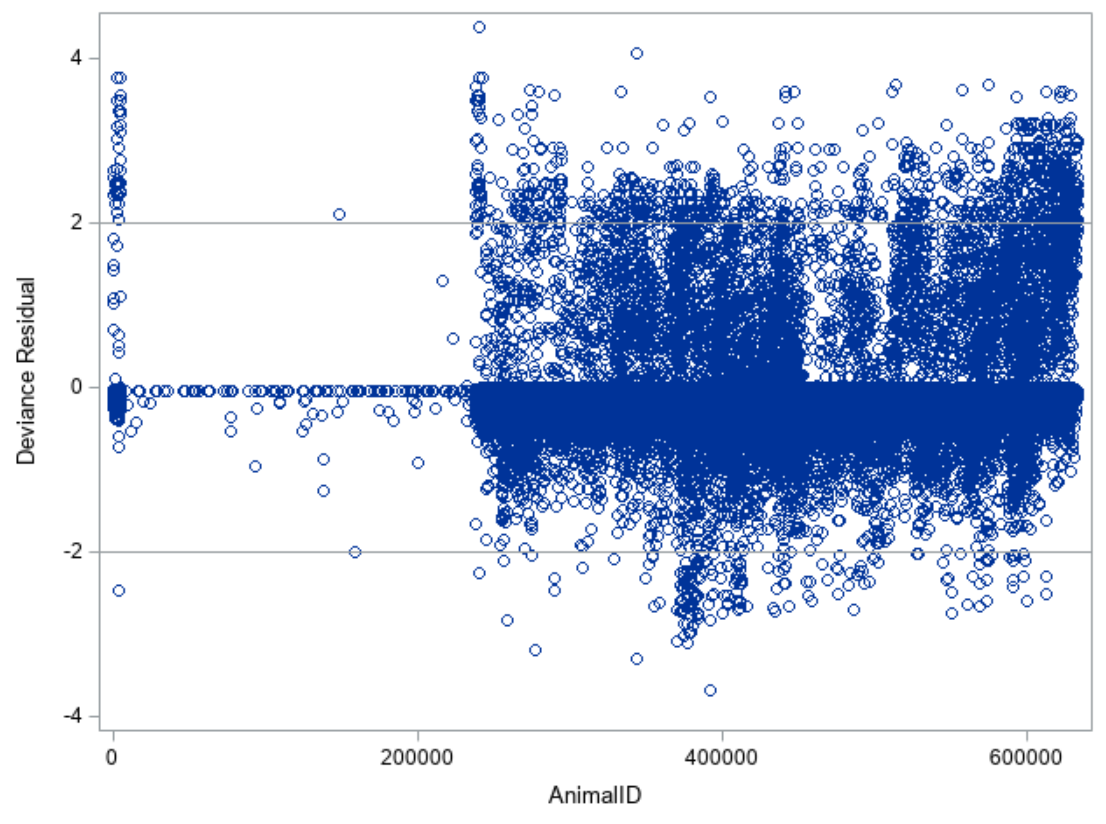
# Appendix 2: Diagnostics

## *Cox-Snell Residuals*



Aside from some deviation in the lower tail and upper tail, the Cox-Snell residuals seem to follow a straight 45-degree line well, indicating a good model fit.

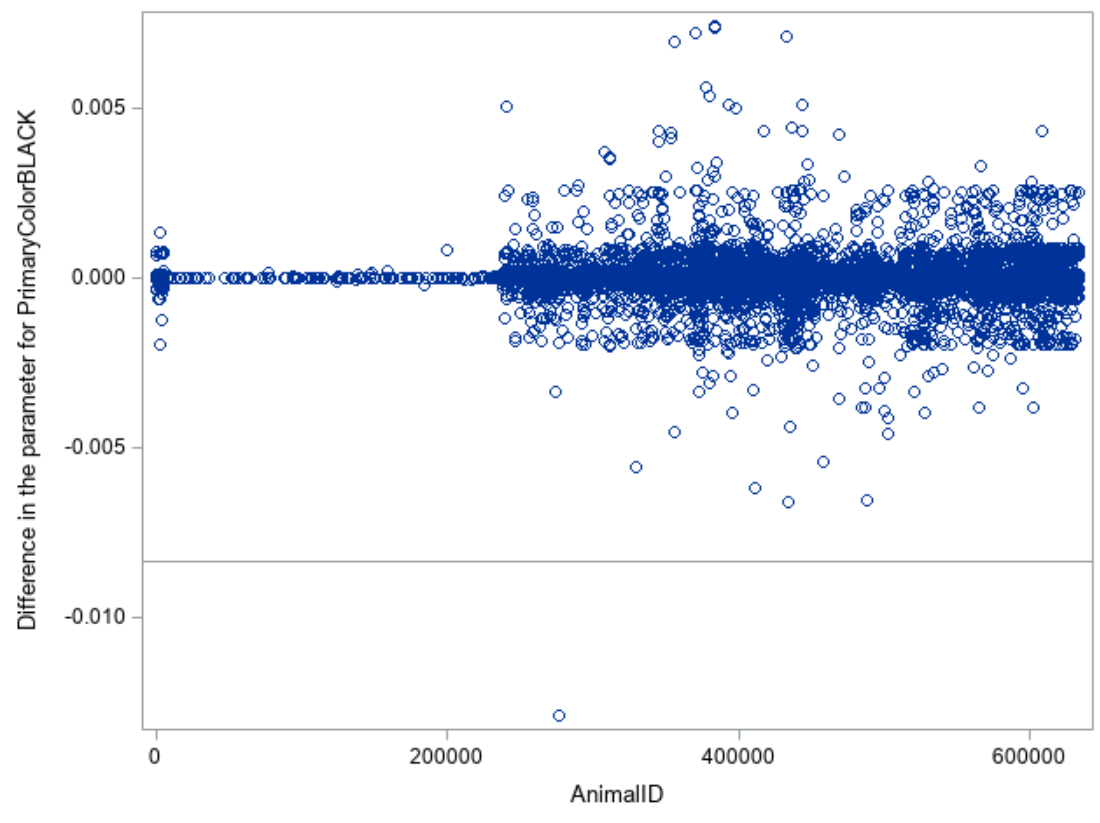
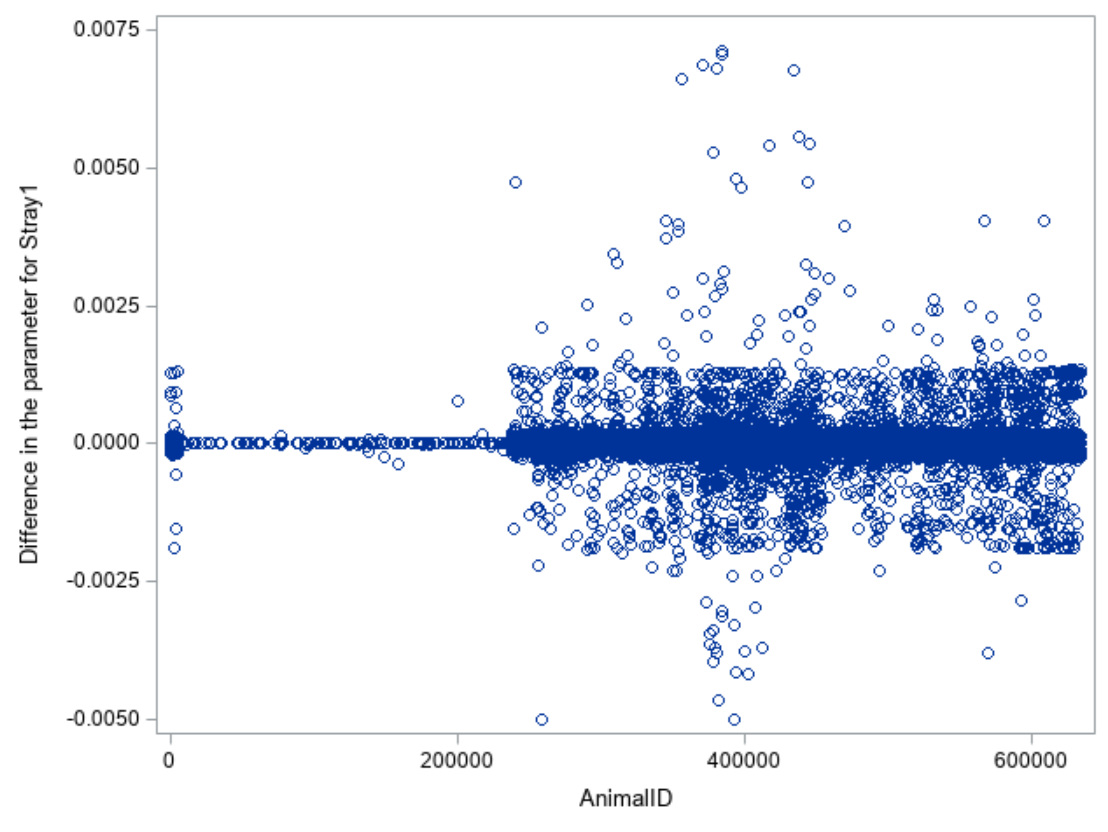
## *Deviance Residuals*

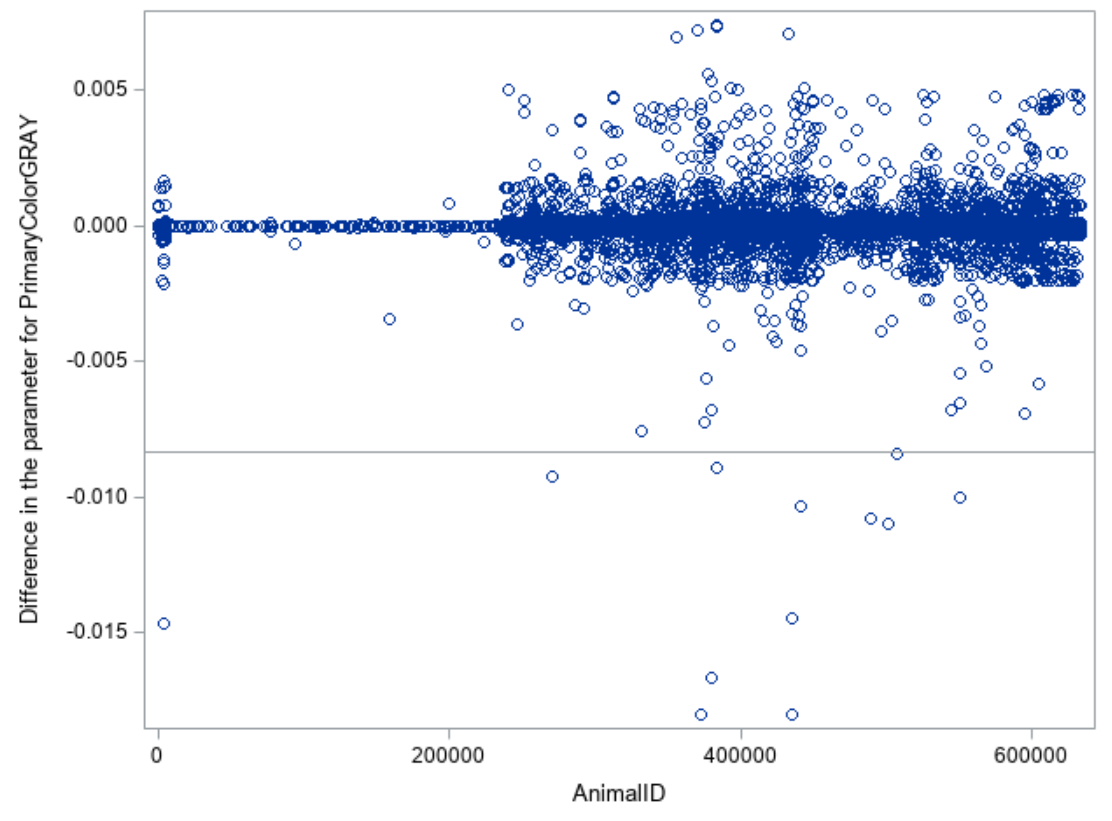
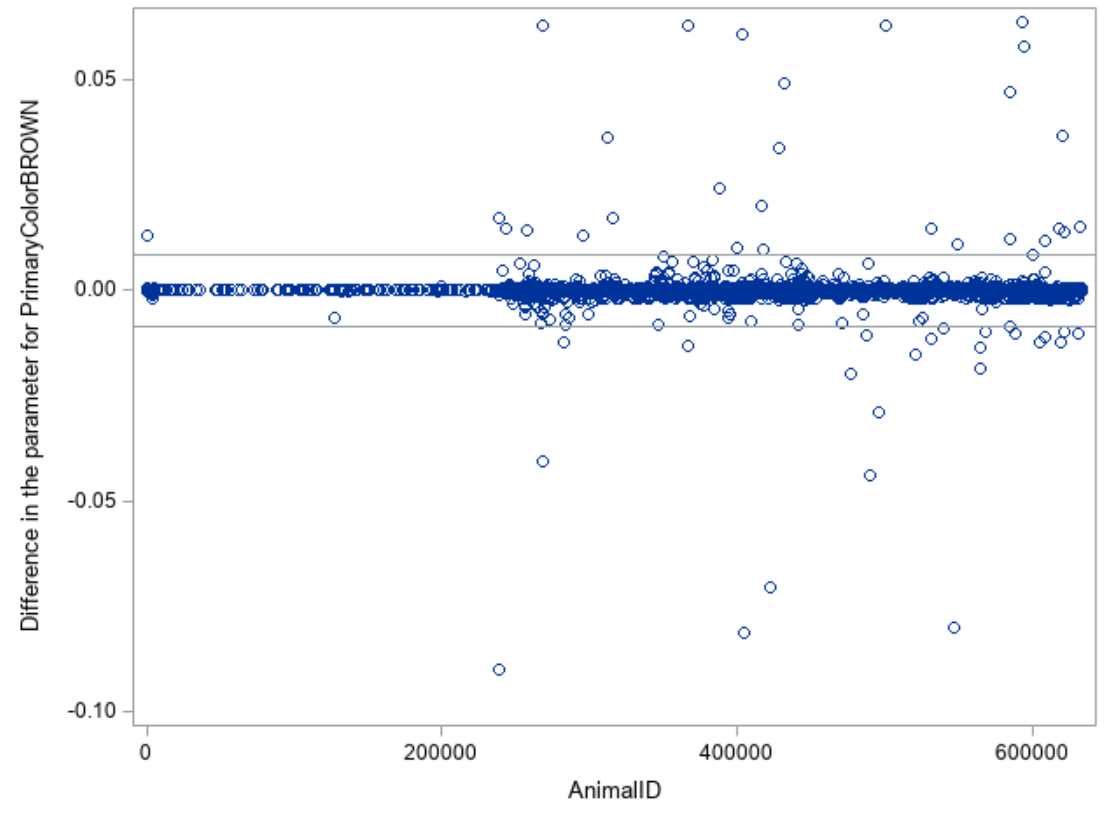


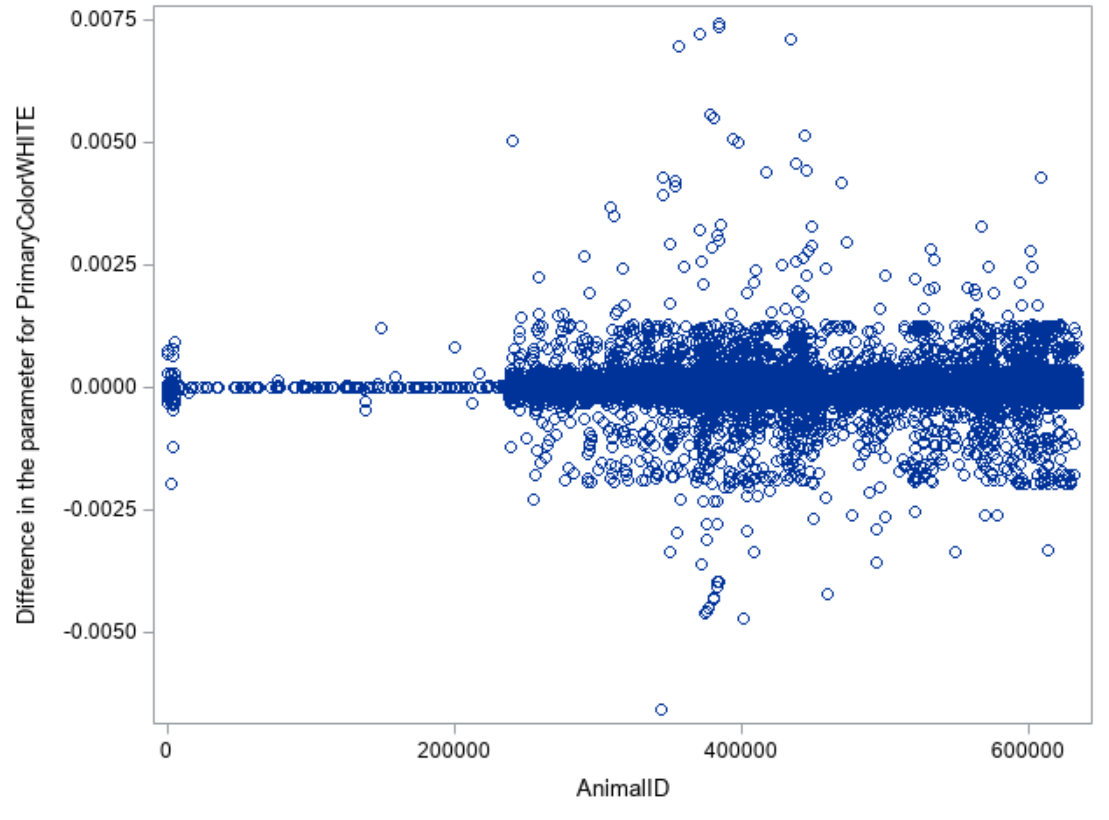
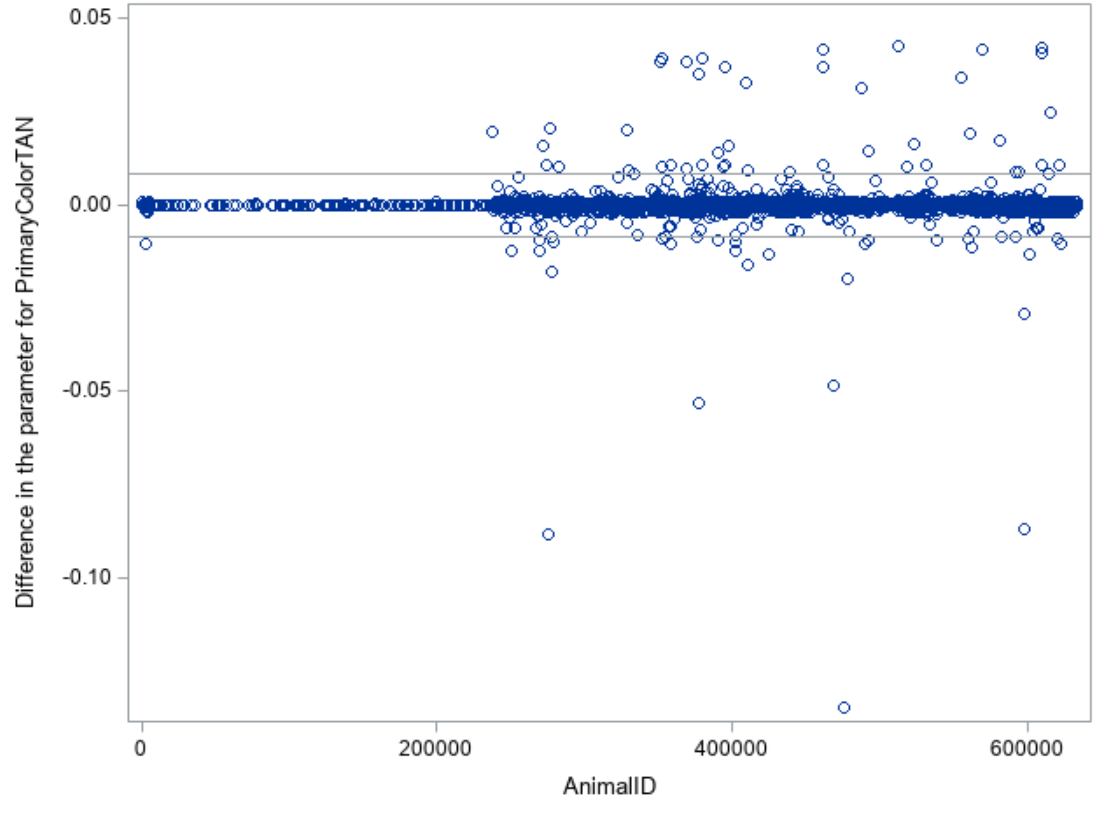
Many of the deviance residuals fall outside the range of [-2, 2]. There appears to be a great number of outliers.

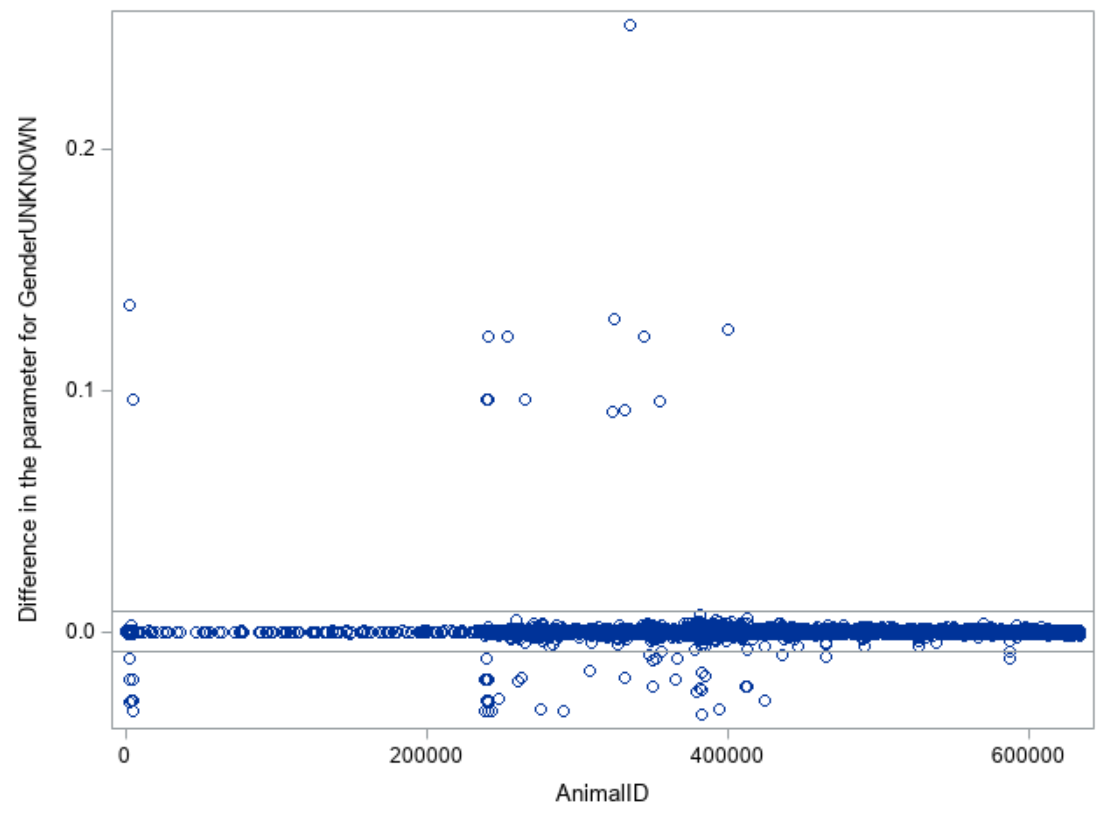
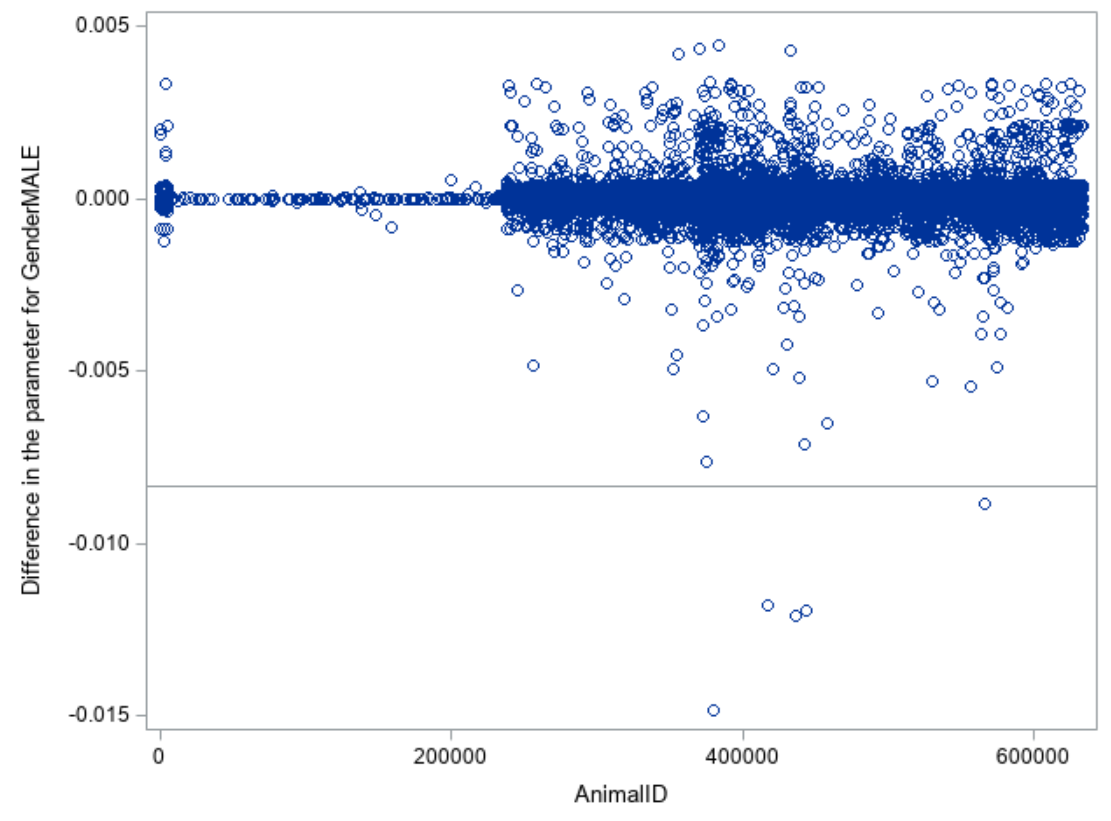
## *dfbetas*

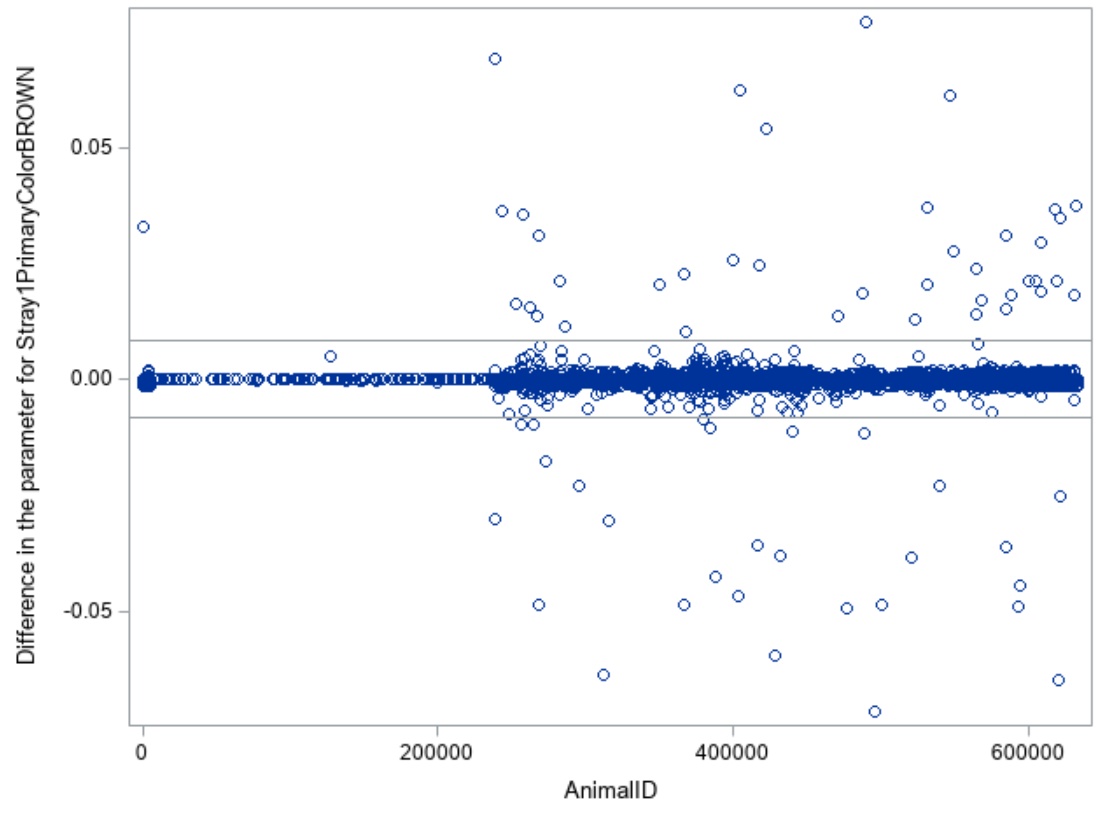
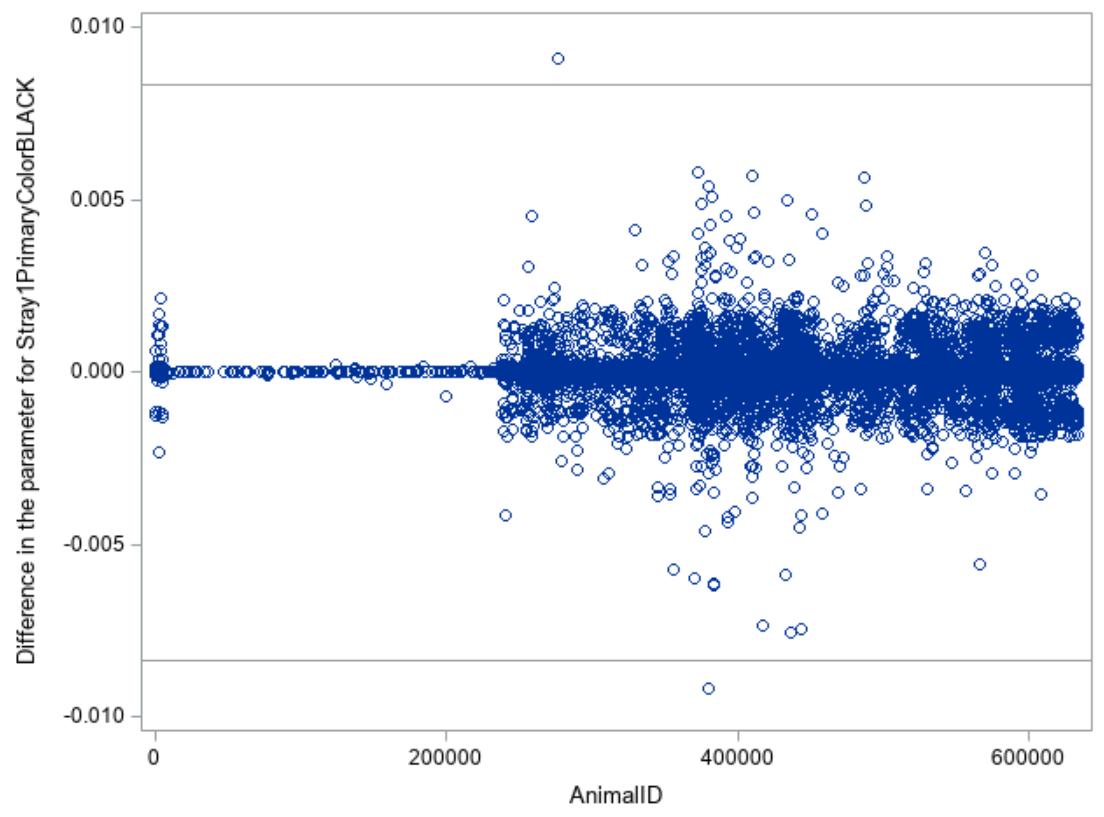
For each variable, including interactions, we examine the graph for points that lie outside . As there are 25 different main effects and interactions, it would be cumbersome to comment on each one individually. The main takeaway from the graphs is that several of the main effects and interactions have many dfbeta points that fall outside [-0.008, 0.008], so there appears to be a great number of potential influential outliers. Due to the vast number and their potential relevance to time to adoption, I decided to leave the outliers in and have chosen not to display the individual table of most influential outliers for each variable here.

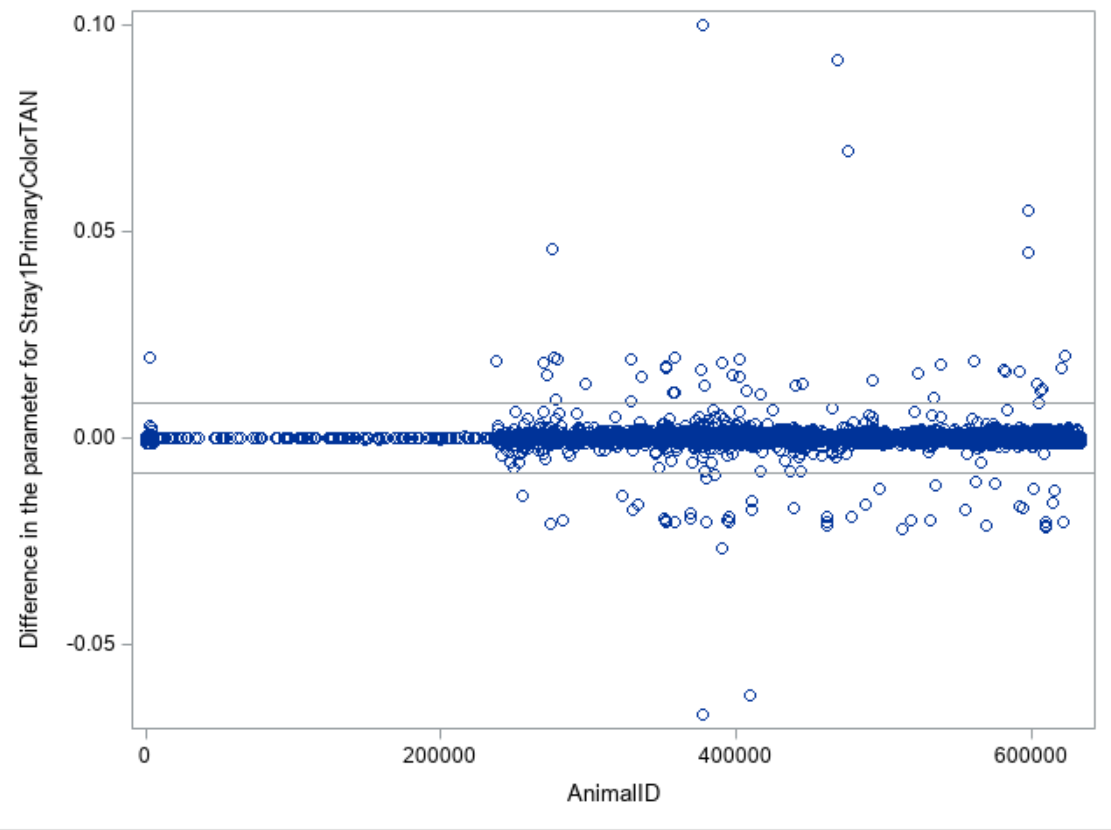
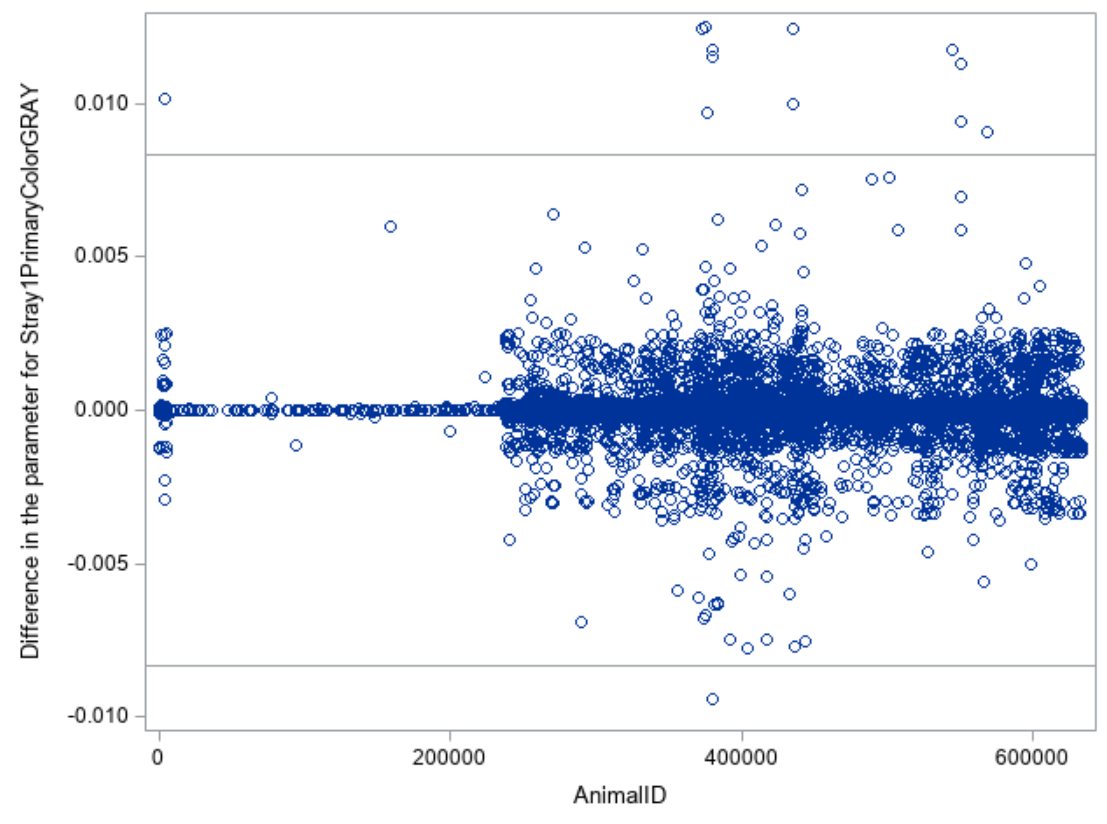


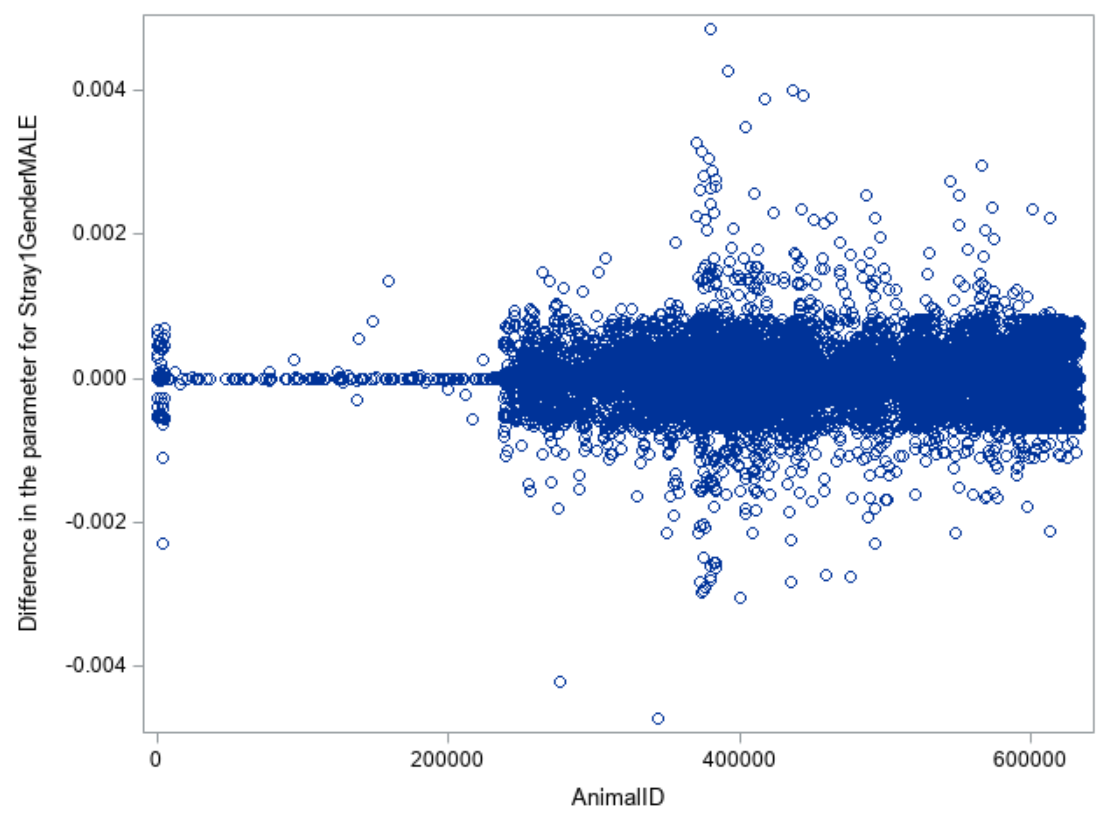
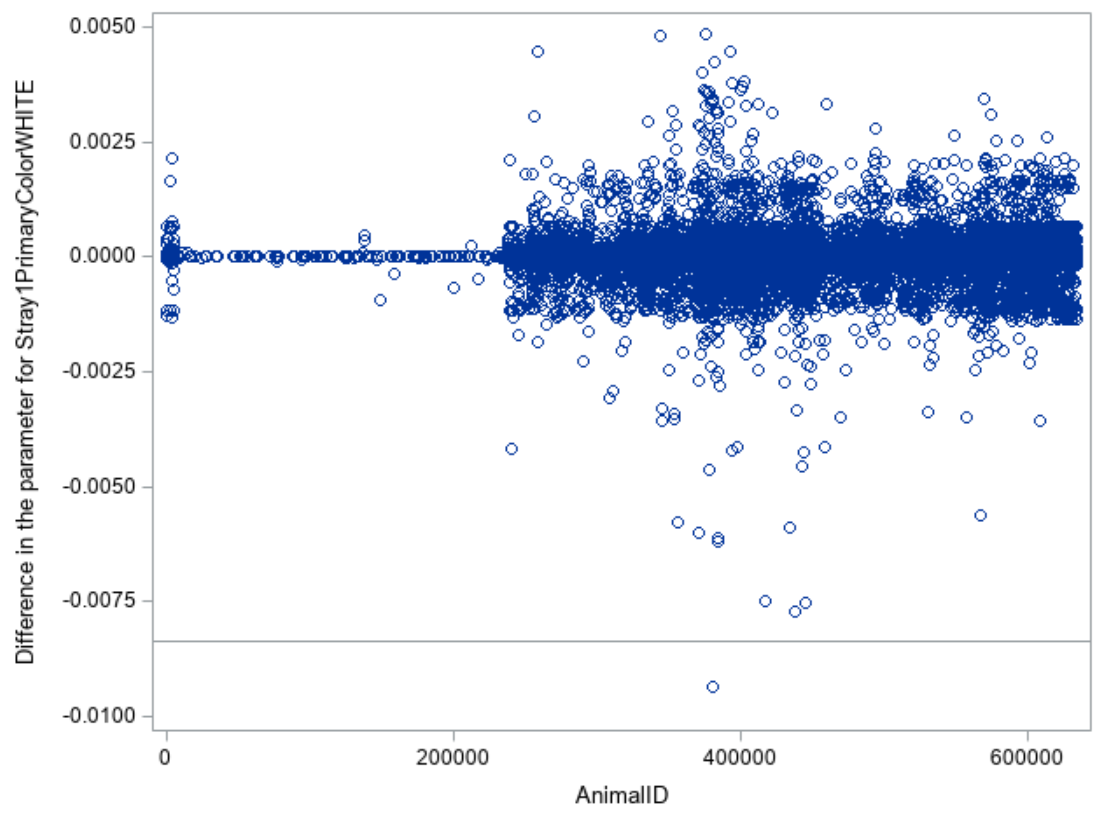


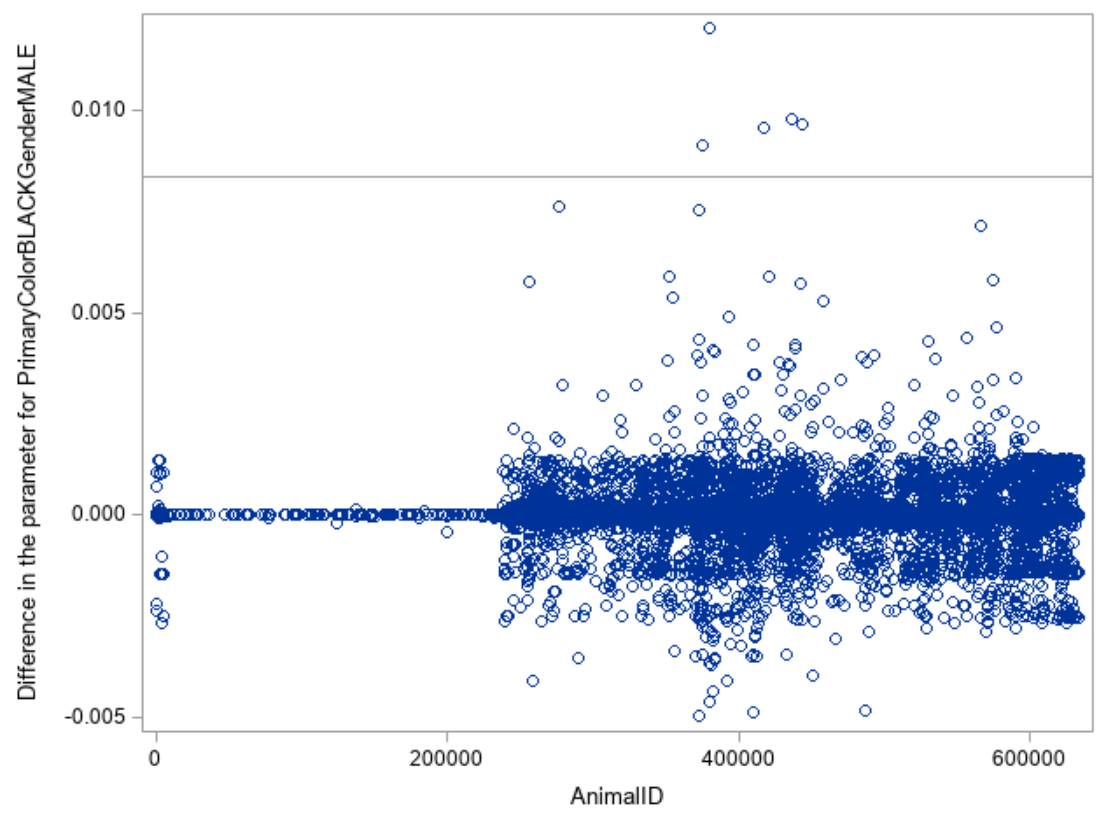
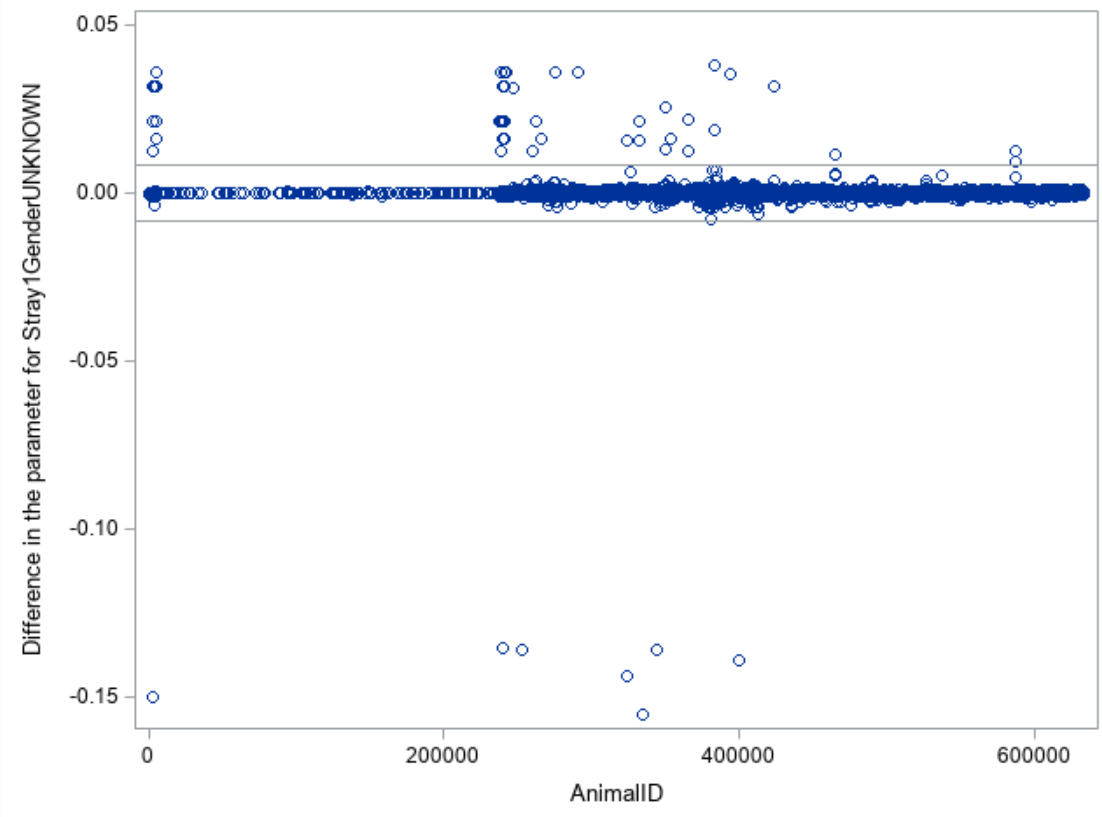


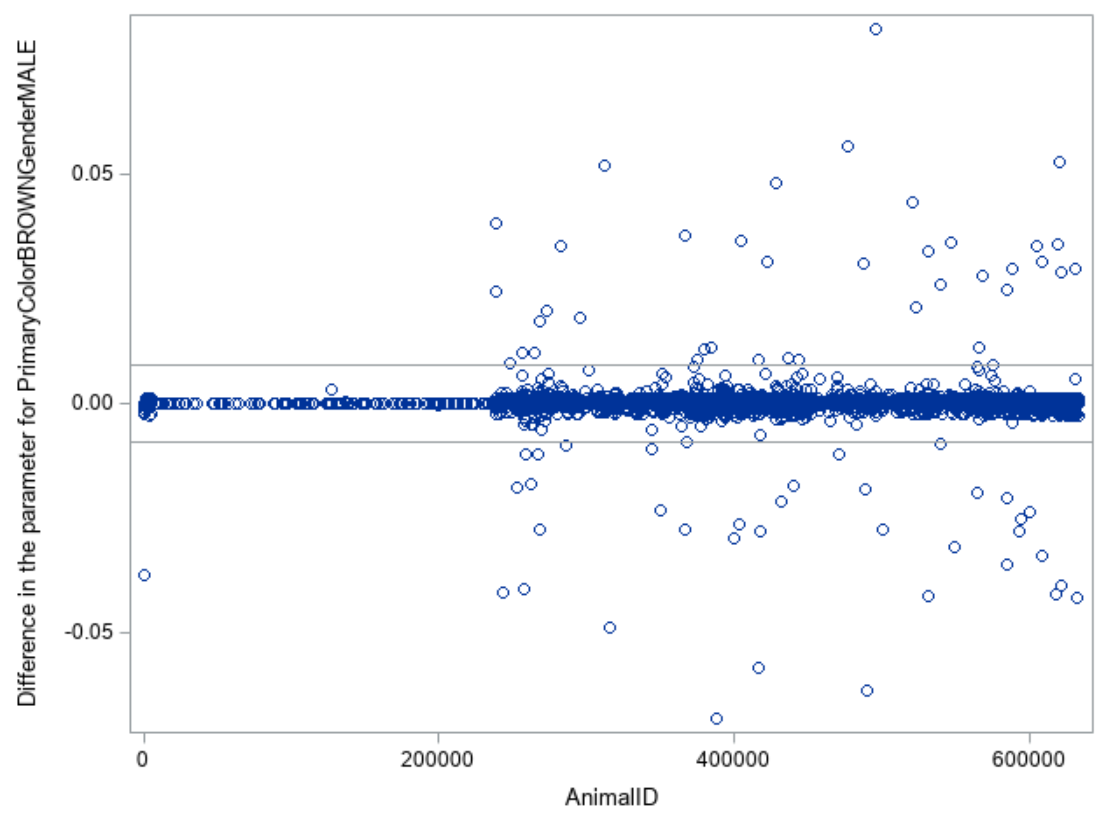
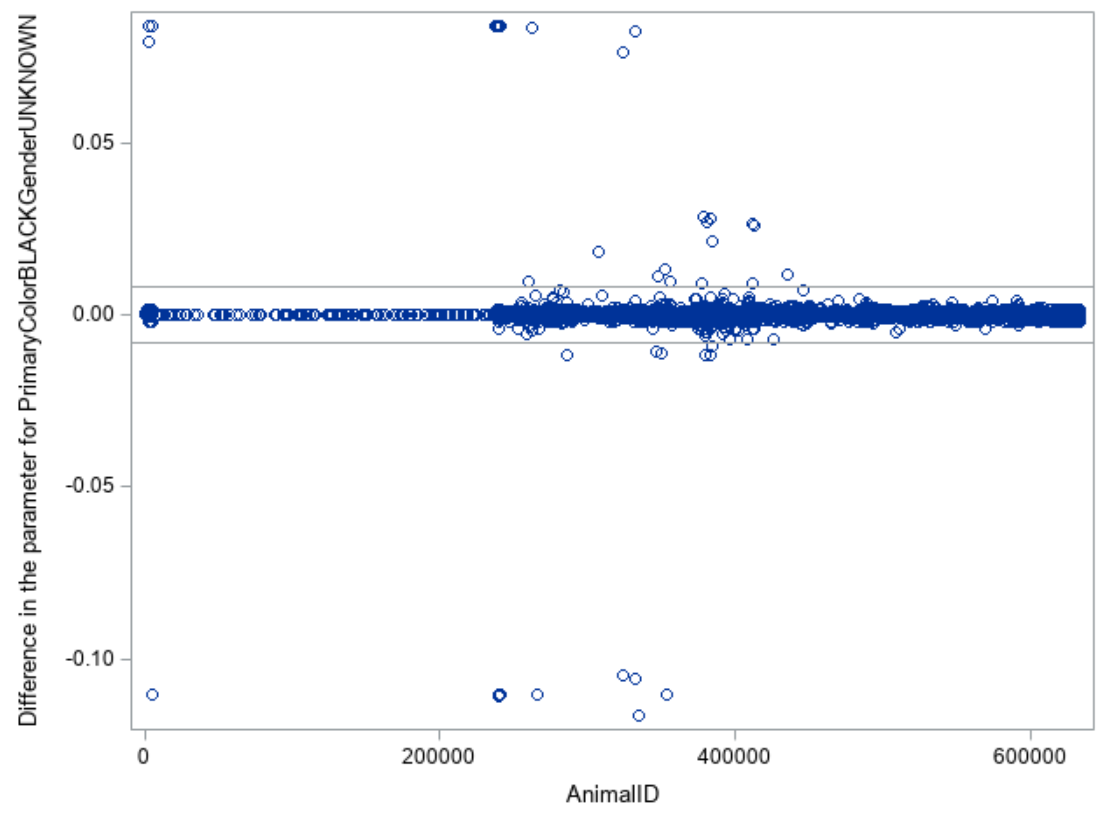


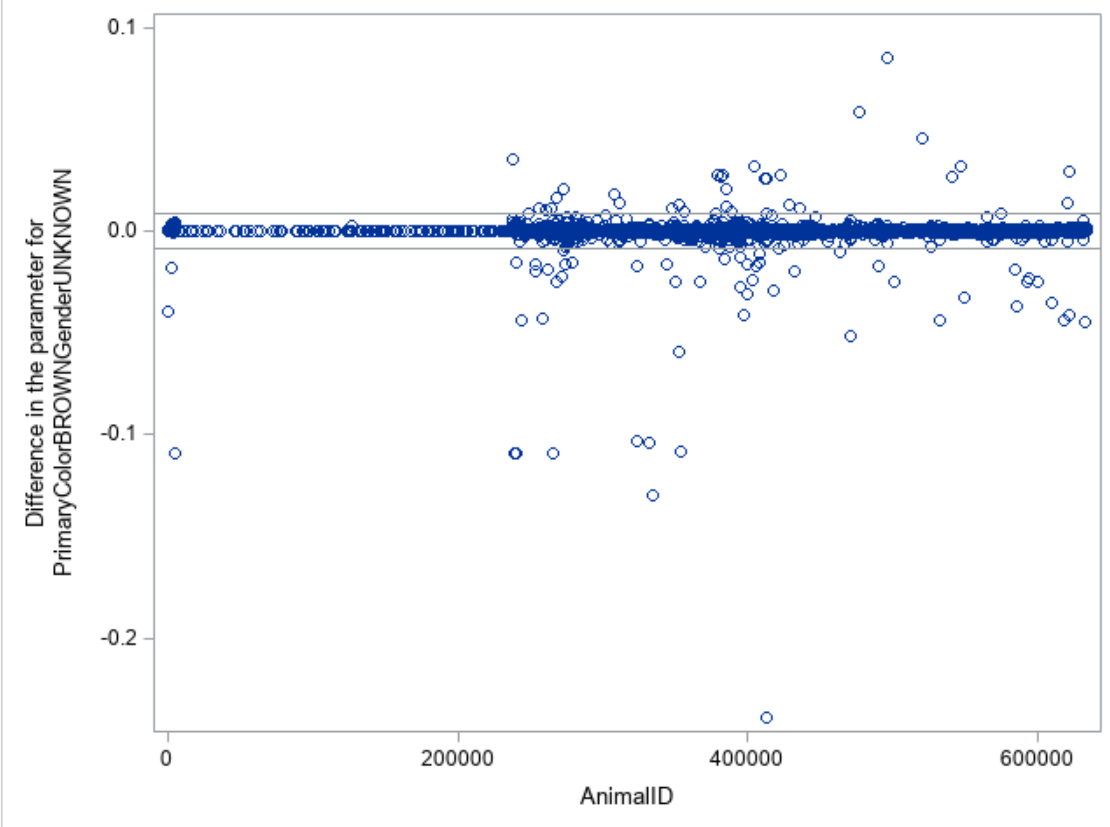


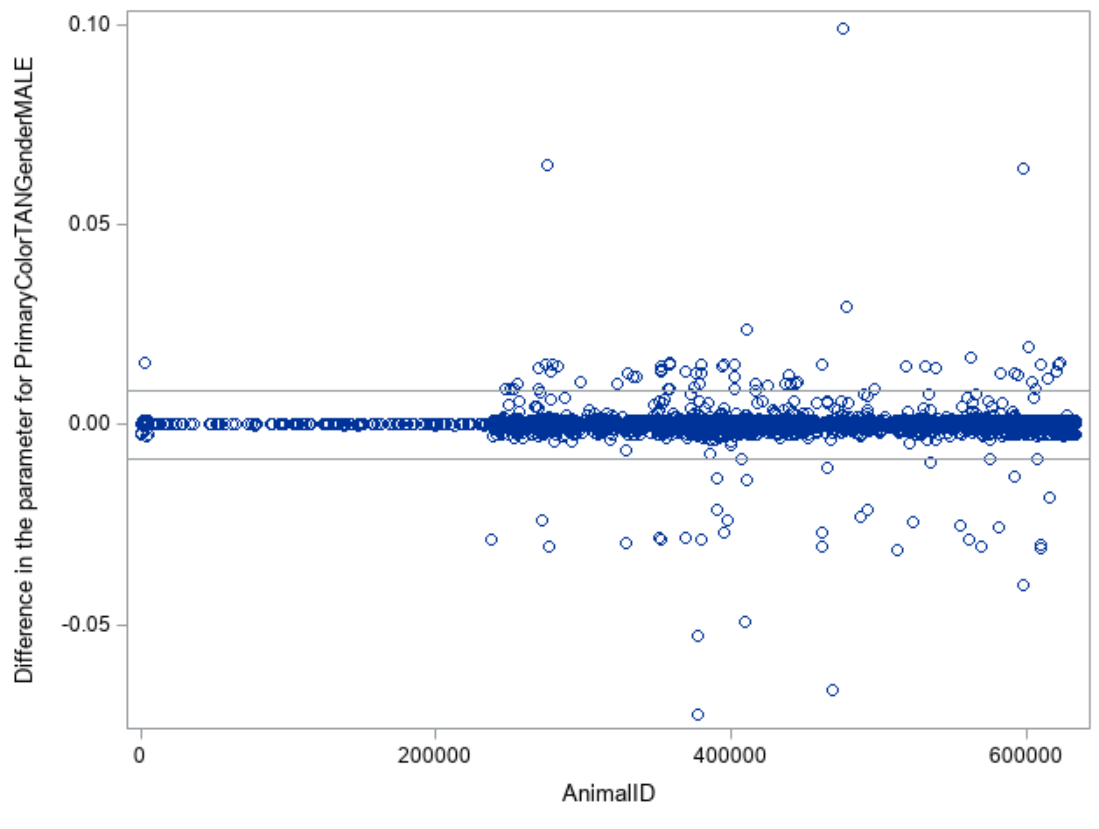
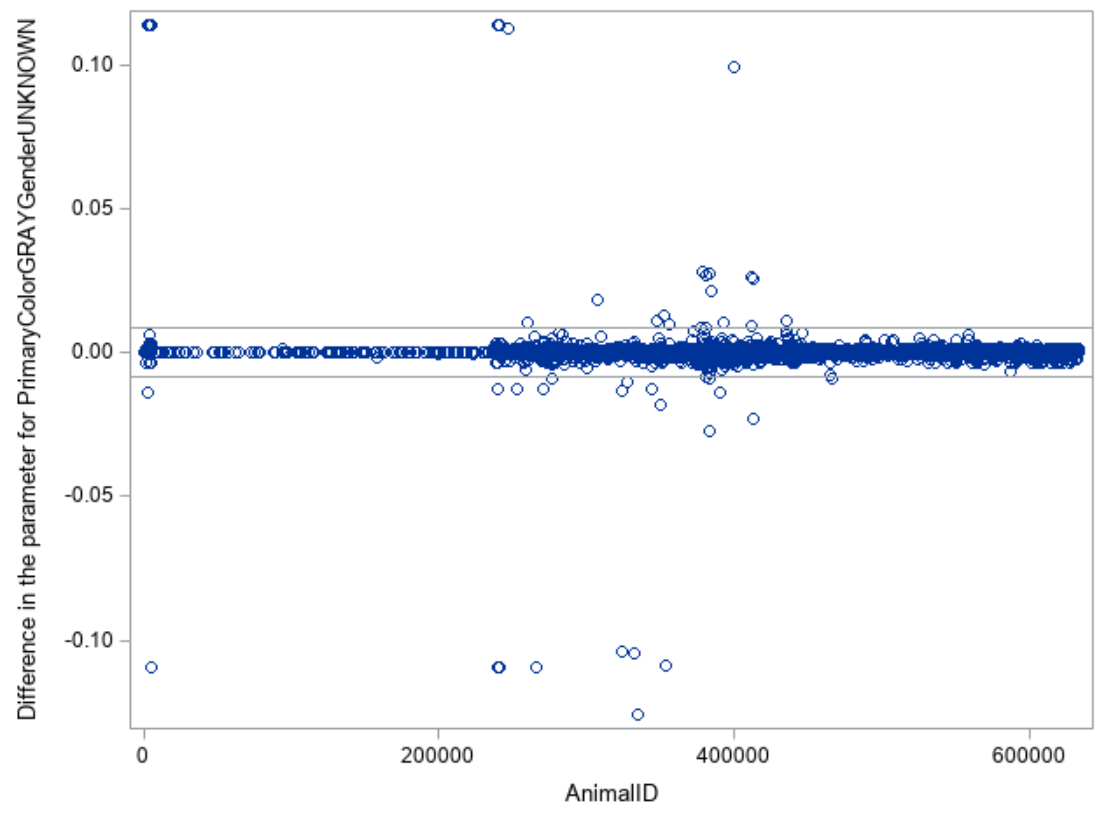


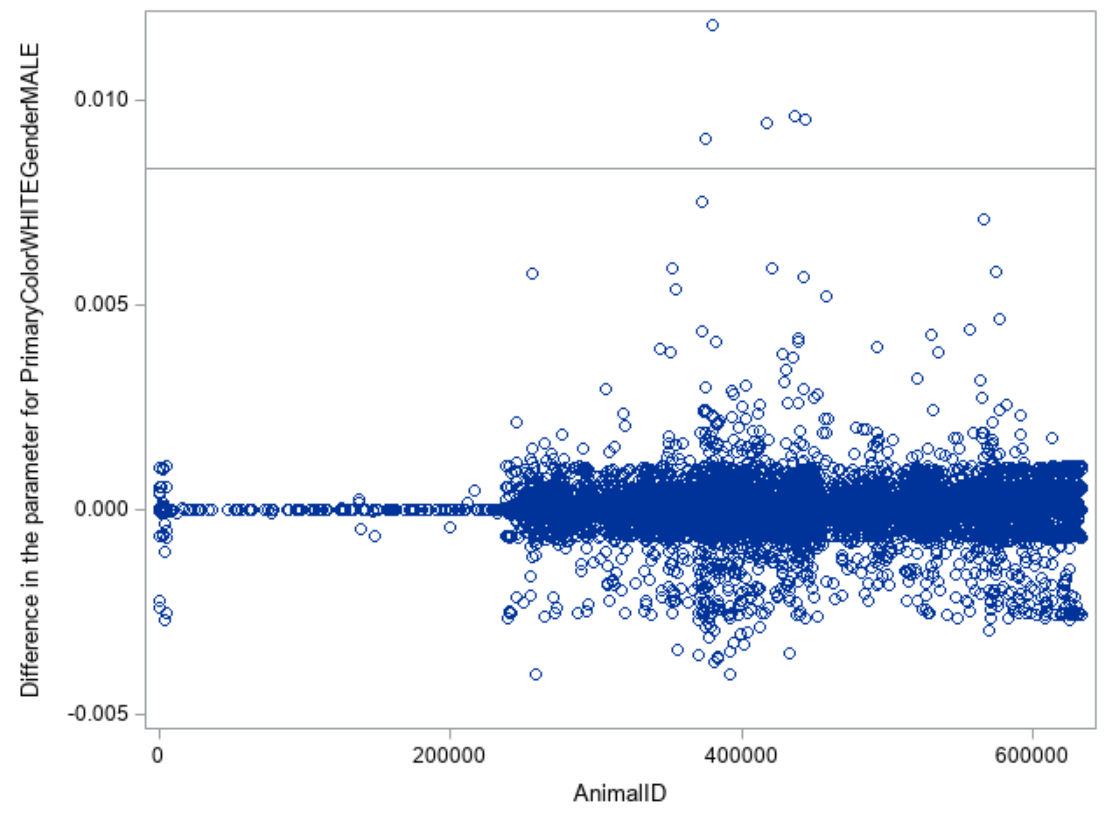
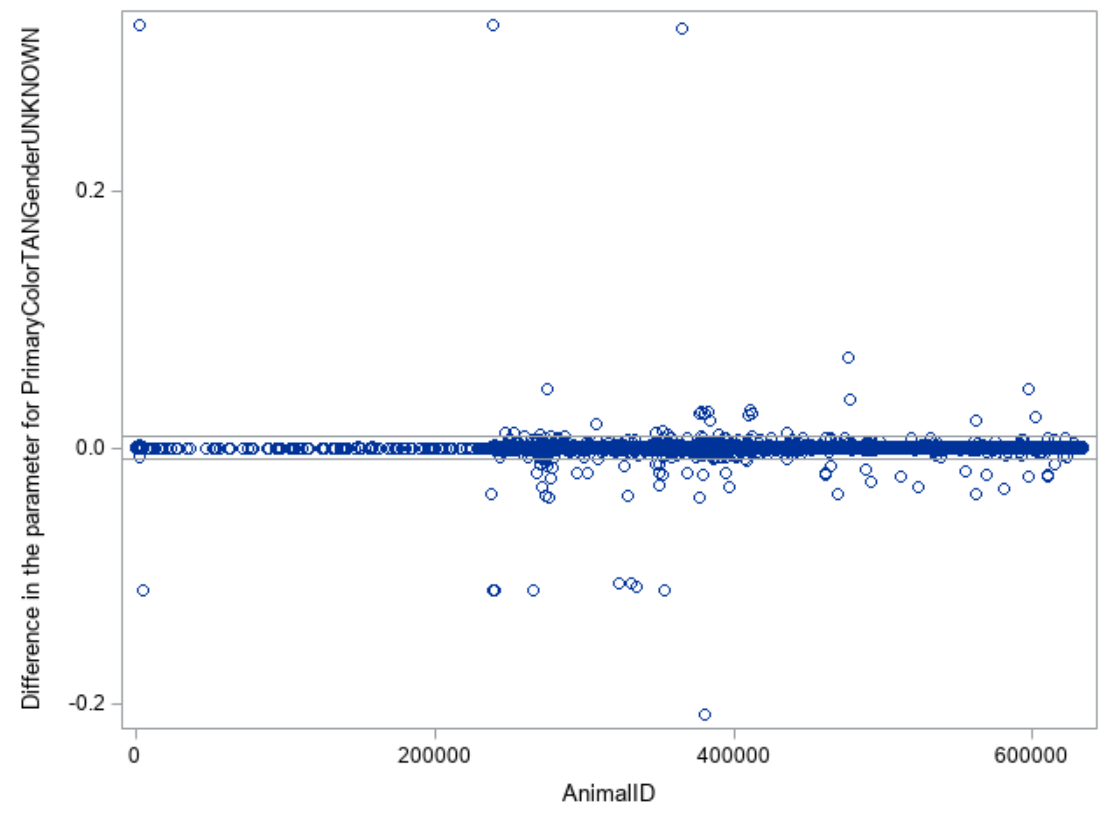


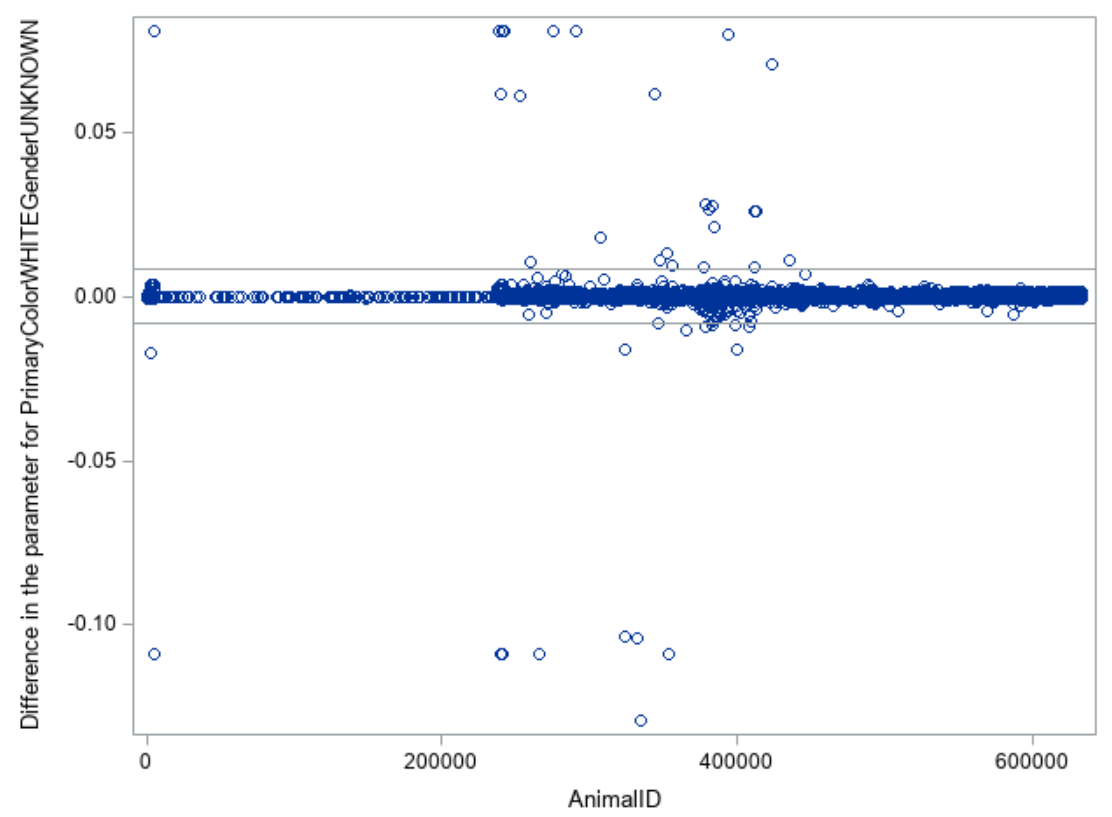






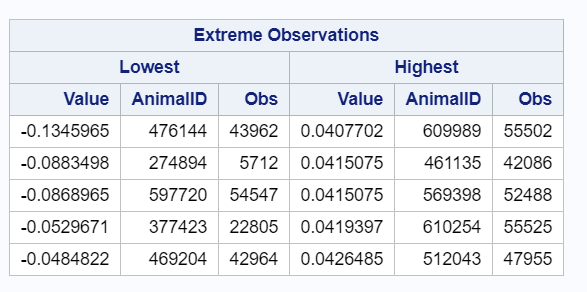






### *Extreme Observations for Primary Color = Tan*

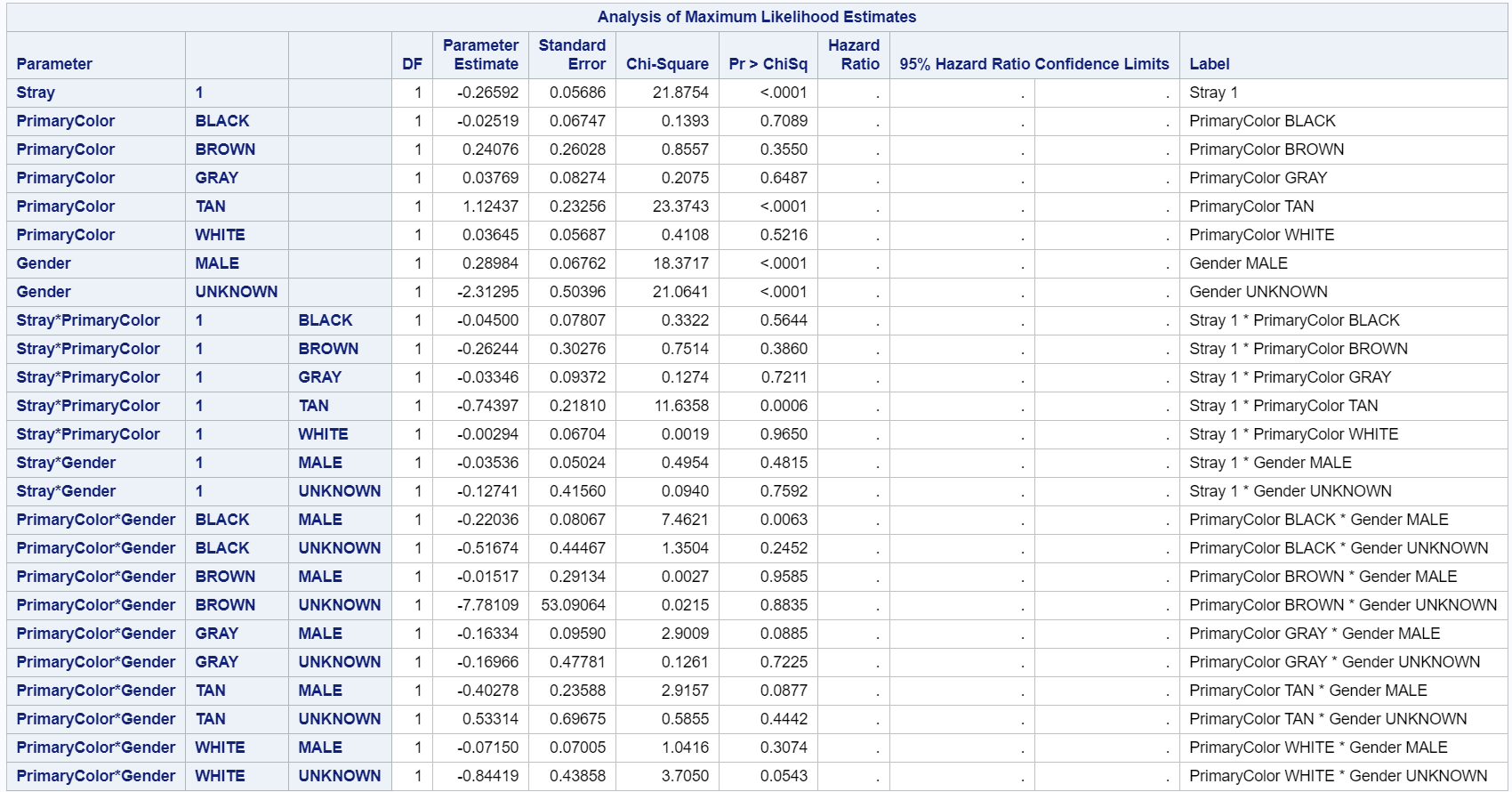
I did attempt to compare estimates between the original dataset and a dataset with a few of the influential outliers removed using one variable to see if caution should be taken in interpreting the analysis.



*Comparison of estimates*

Original estimates

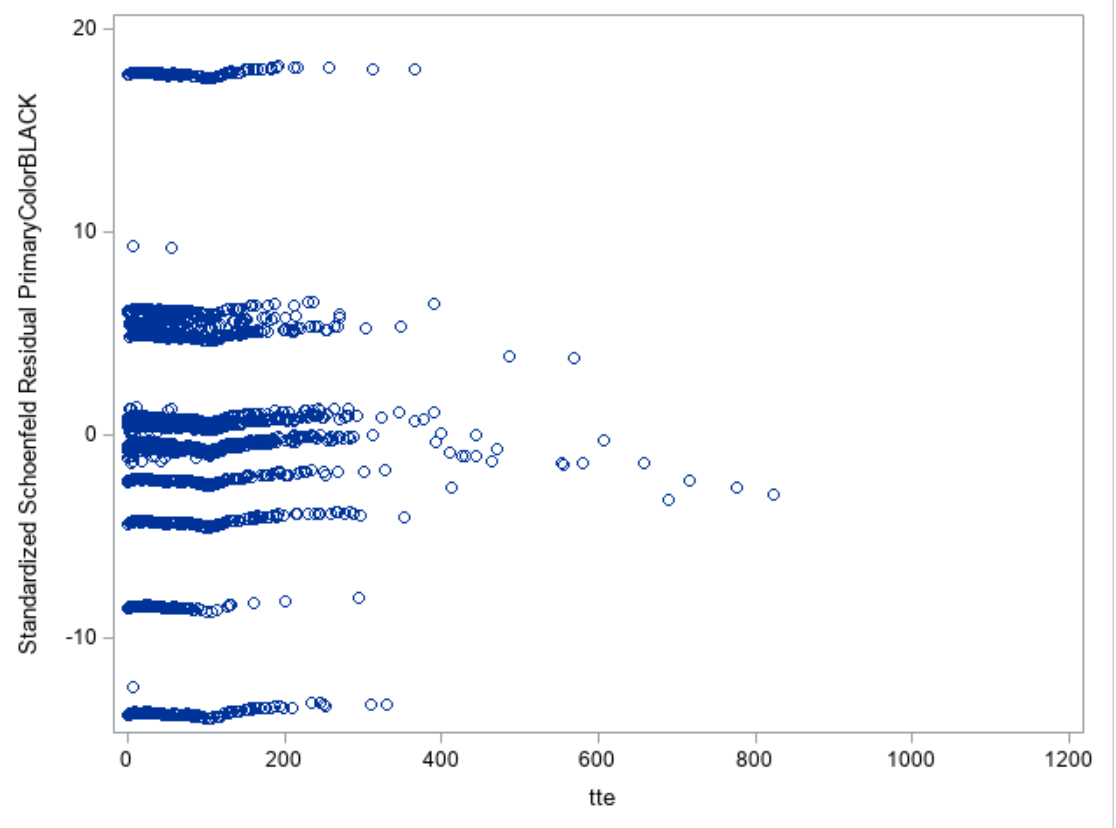
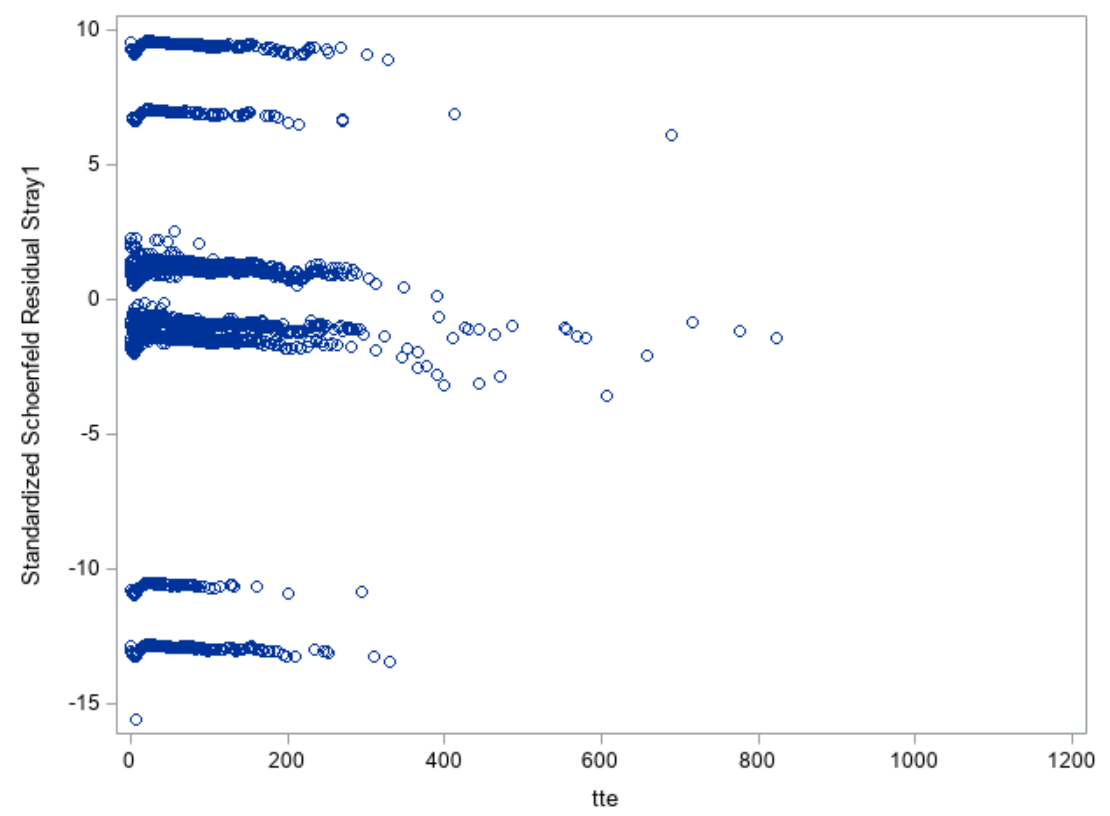


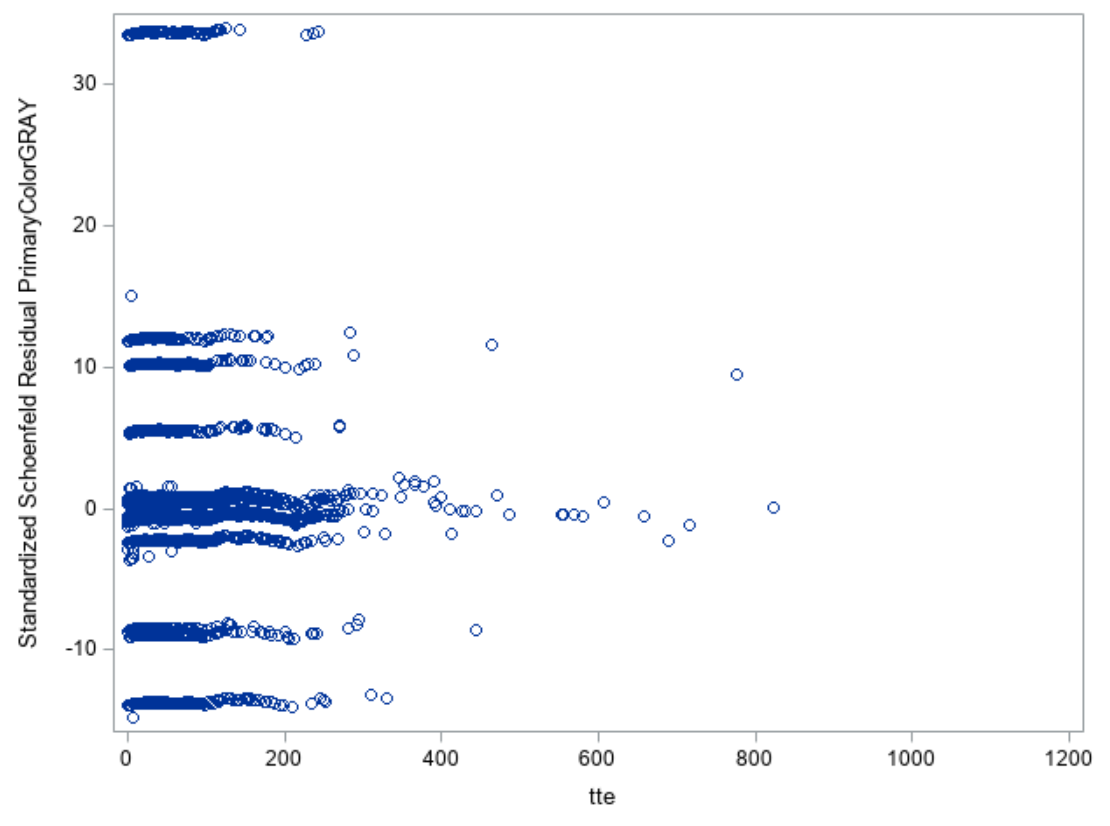
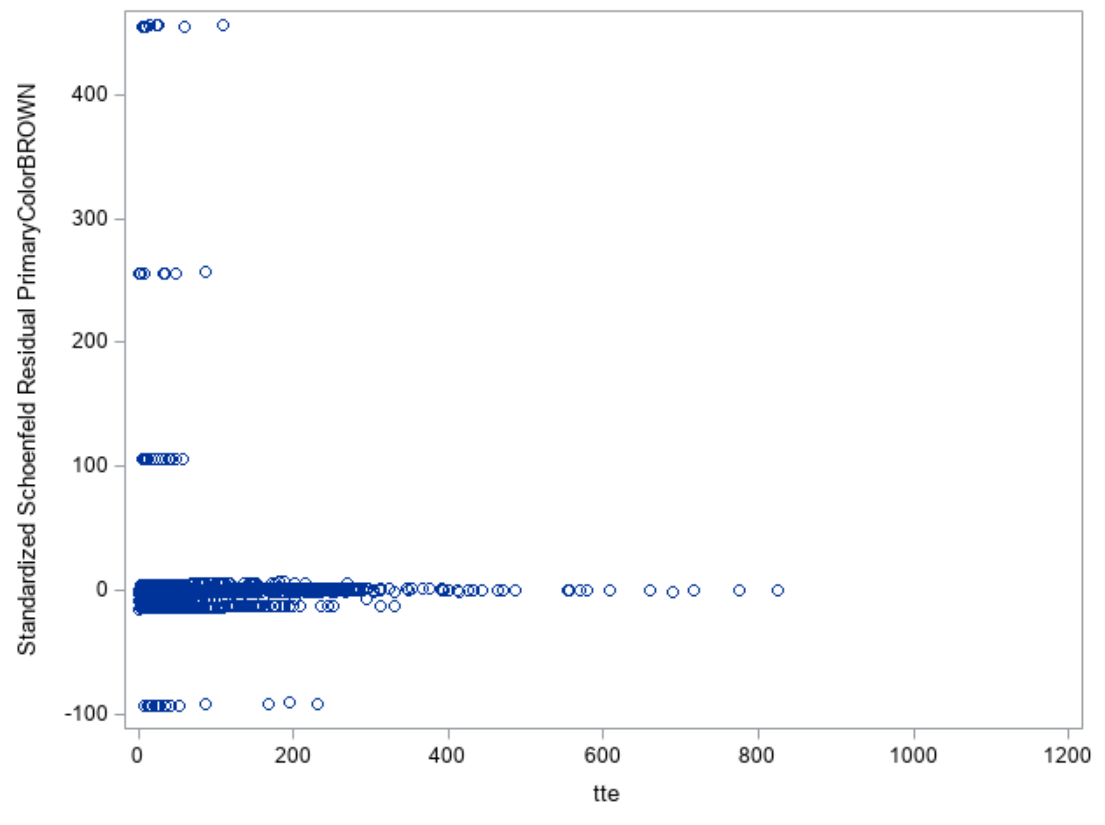


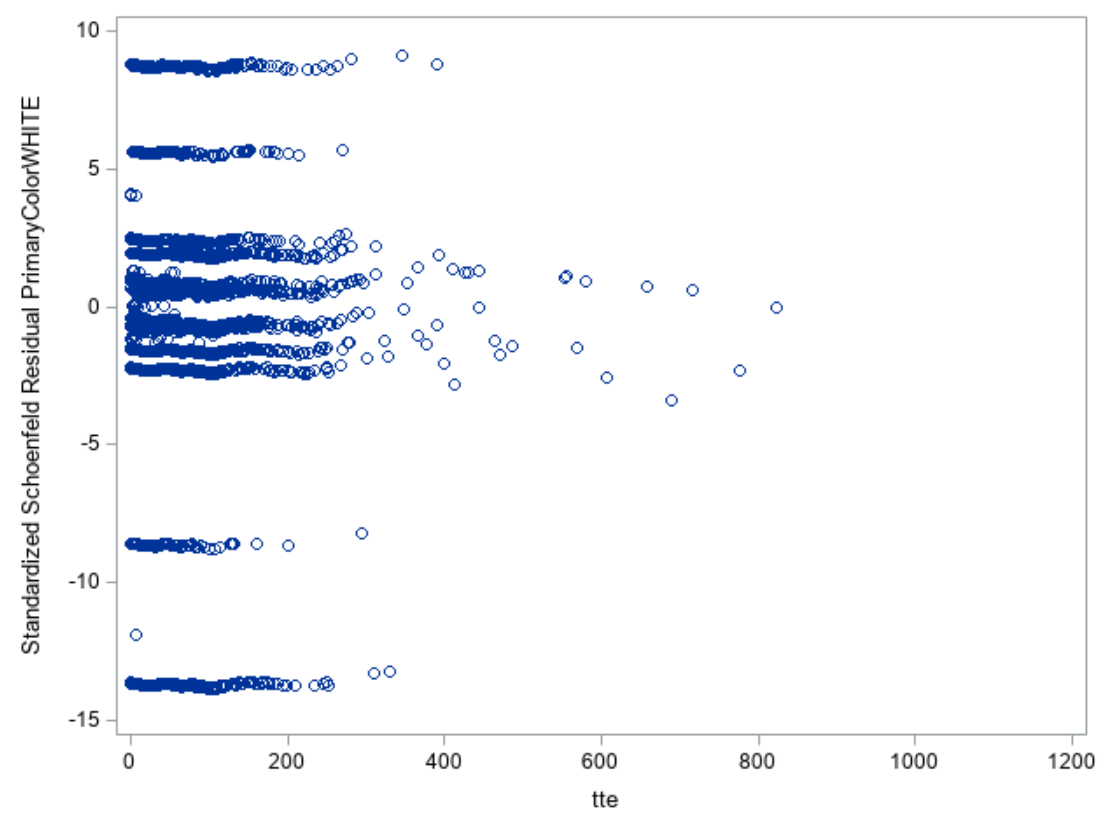
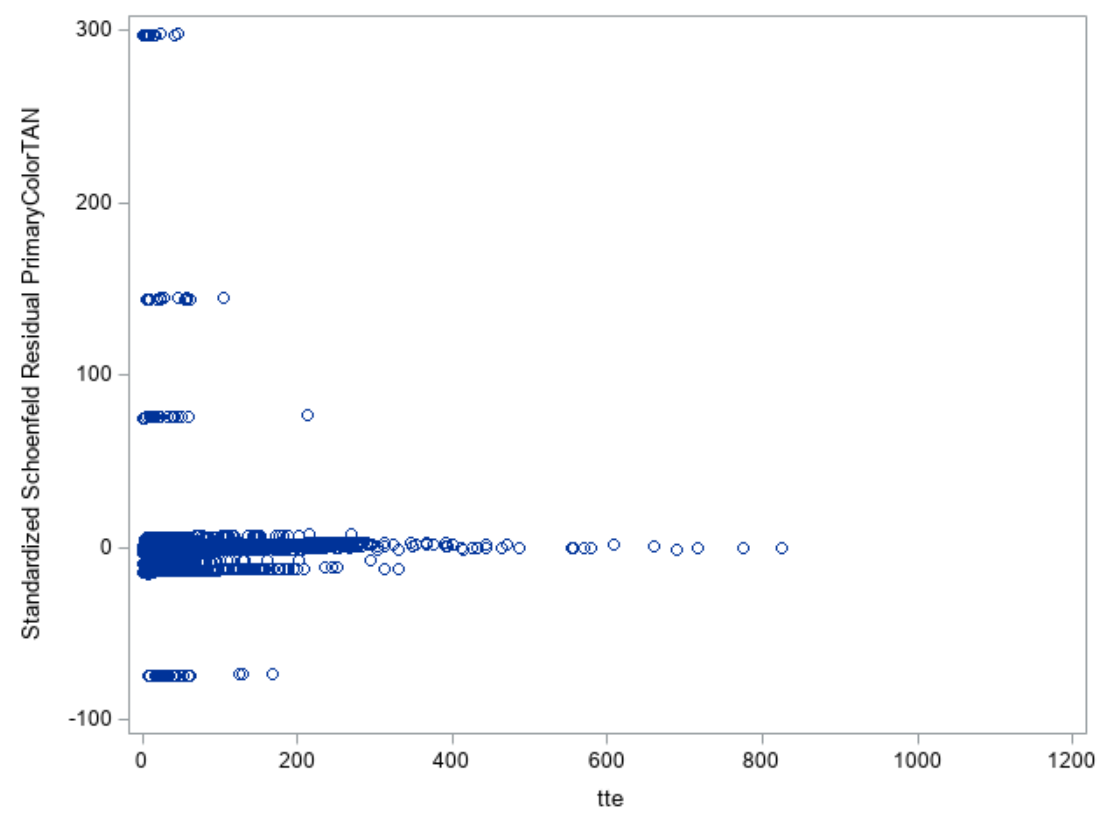
A visual inspection revealed that the estimates for log hazard ratio of the main effect of tan (original log HR = 0.77, new log HR = 1.12) and the stray x tan interaction (original log HR = -0.39, new log HR = -0.74) changed by more than 10%. All other estimates did not have change more than 10%. This suggests that the results of this analysis should be interpreted with caution.

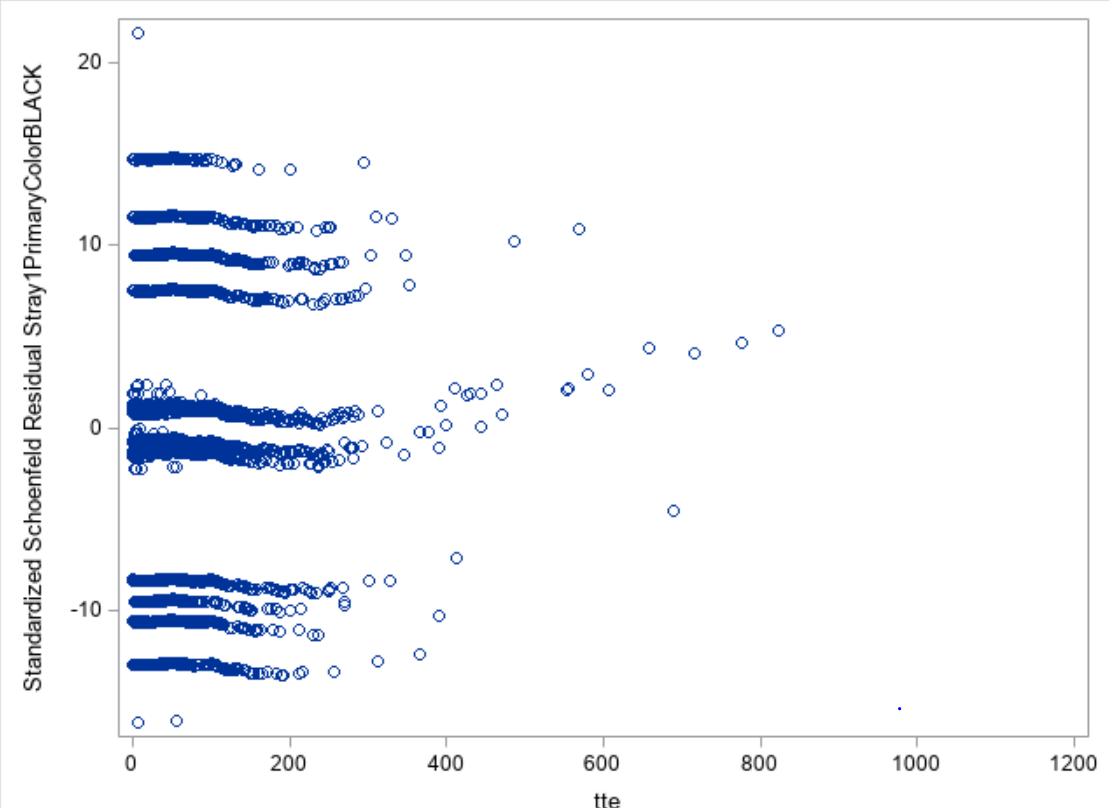
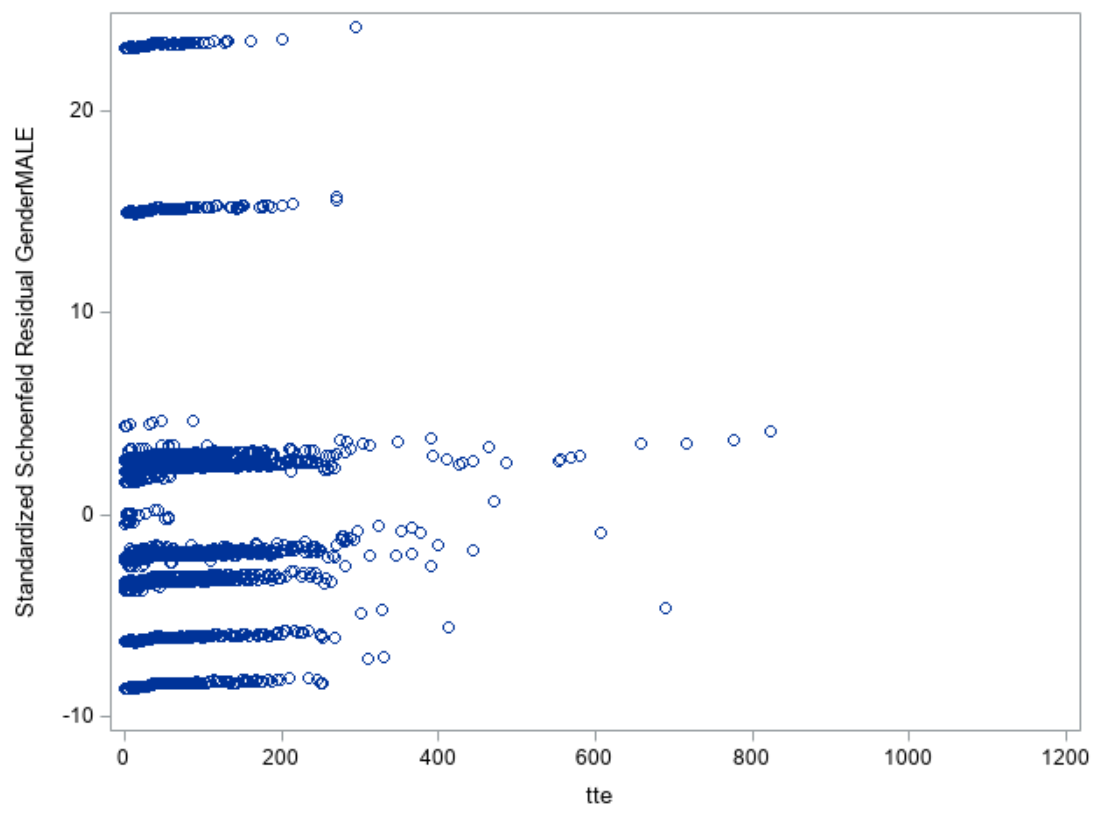
## *Schoenfield Residuals*

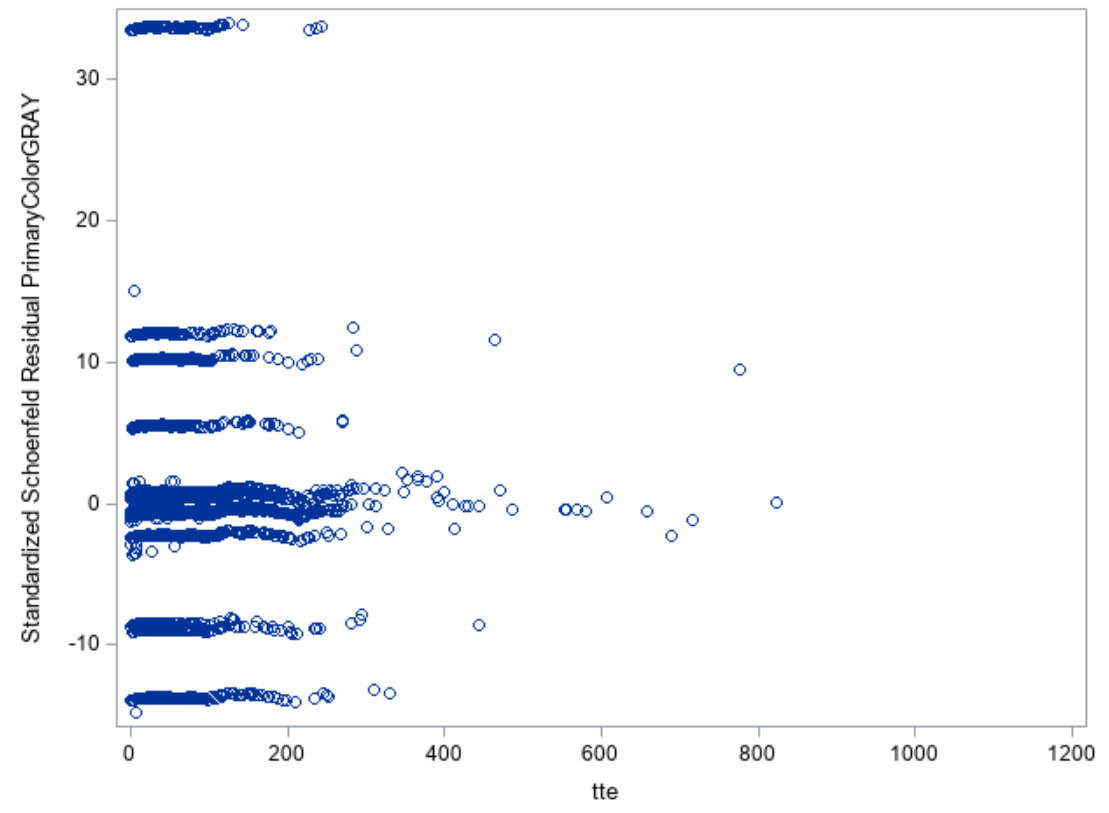
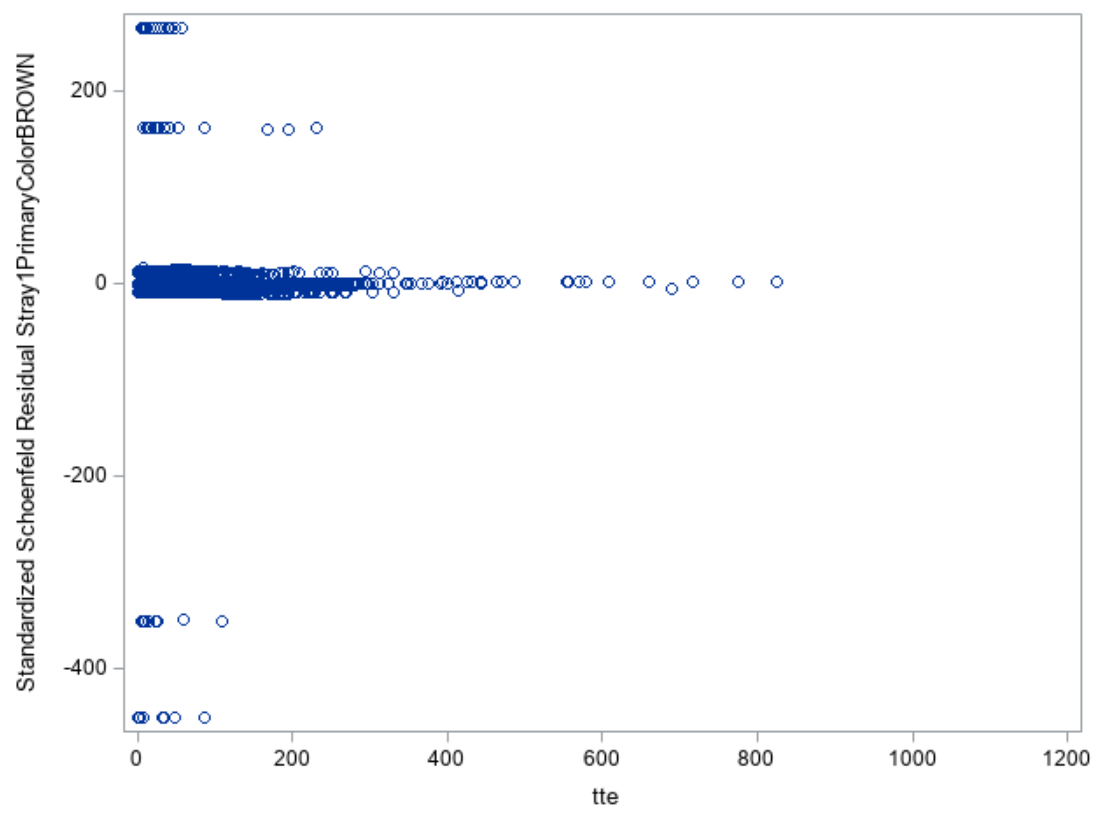
Schoenfield residuals are used to determine if the hazards for each categorical variable remain proportional over time. Ideally, the plot of Schoenfield residuals should be completely horizontal, showing no trend over time. The plots for each variable are included below. It seemed like most of the variables seem to have lines that are relatively horizontal minus some points on the right tail. However, it is not obvious from the Schoenfield residual graphs how much deviation from horizontal is necessary to indicate significant non-proportional hazards, so these should be interpreted with caution.

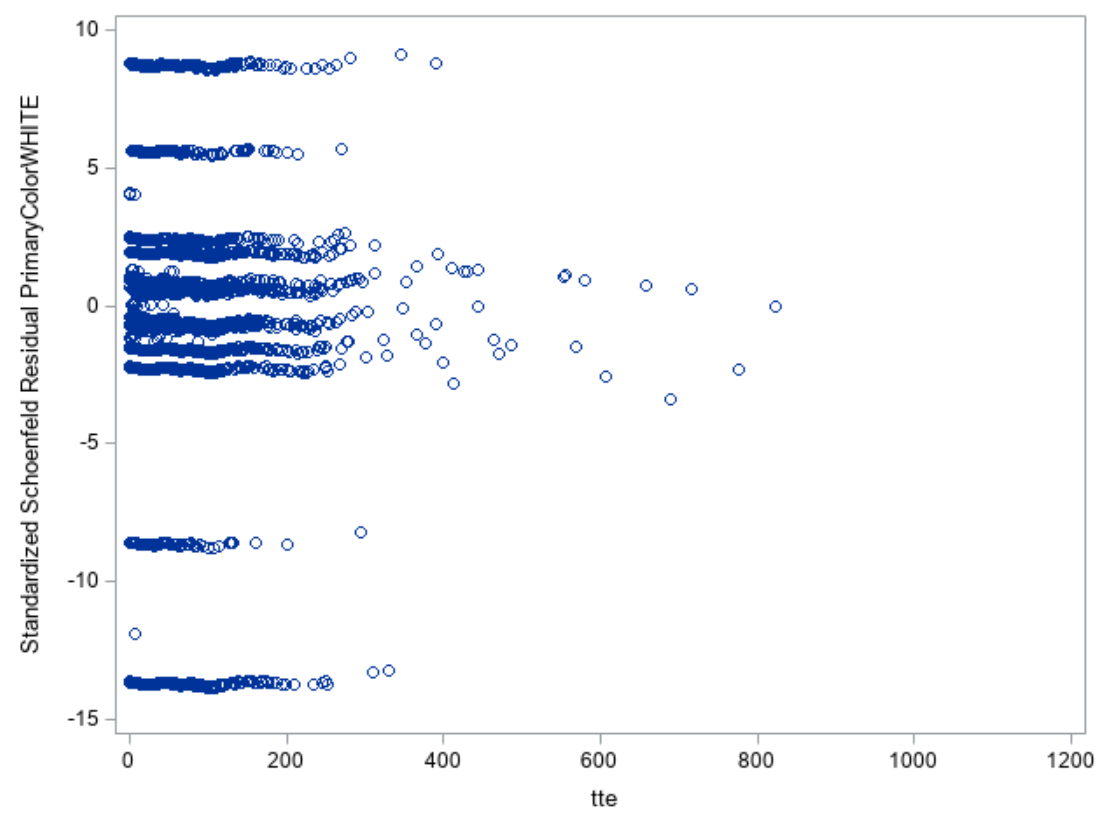
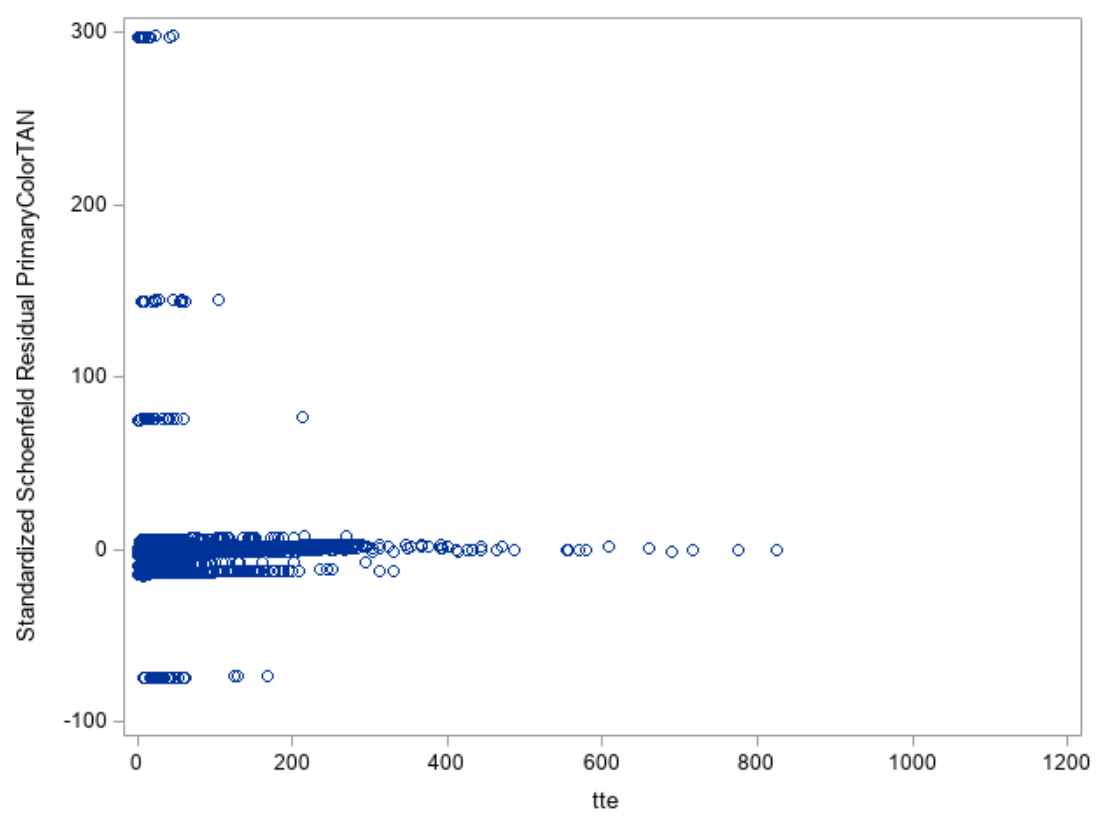


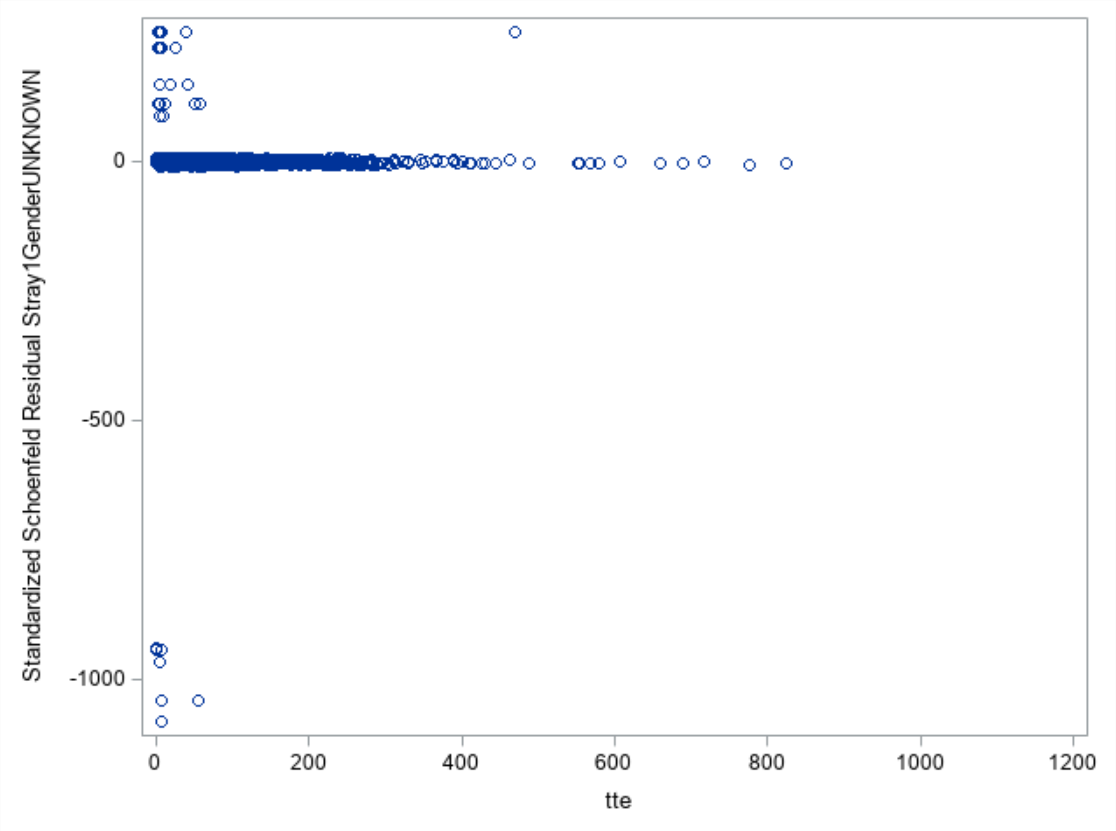
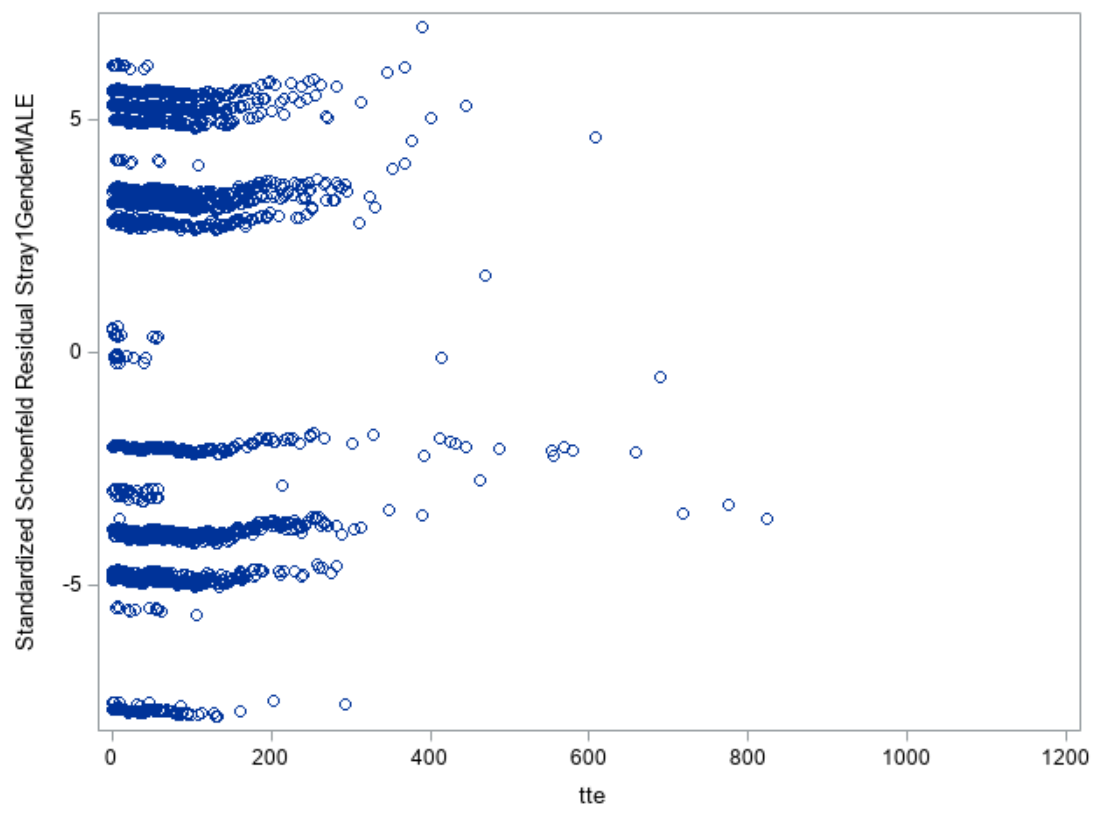


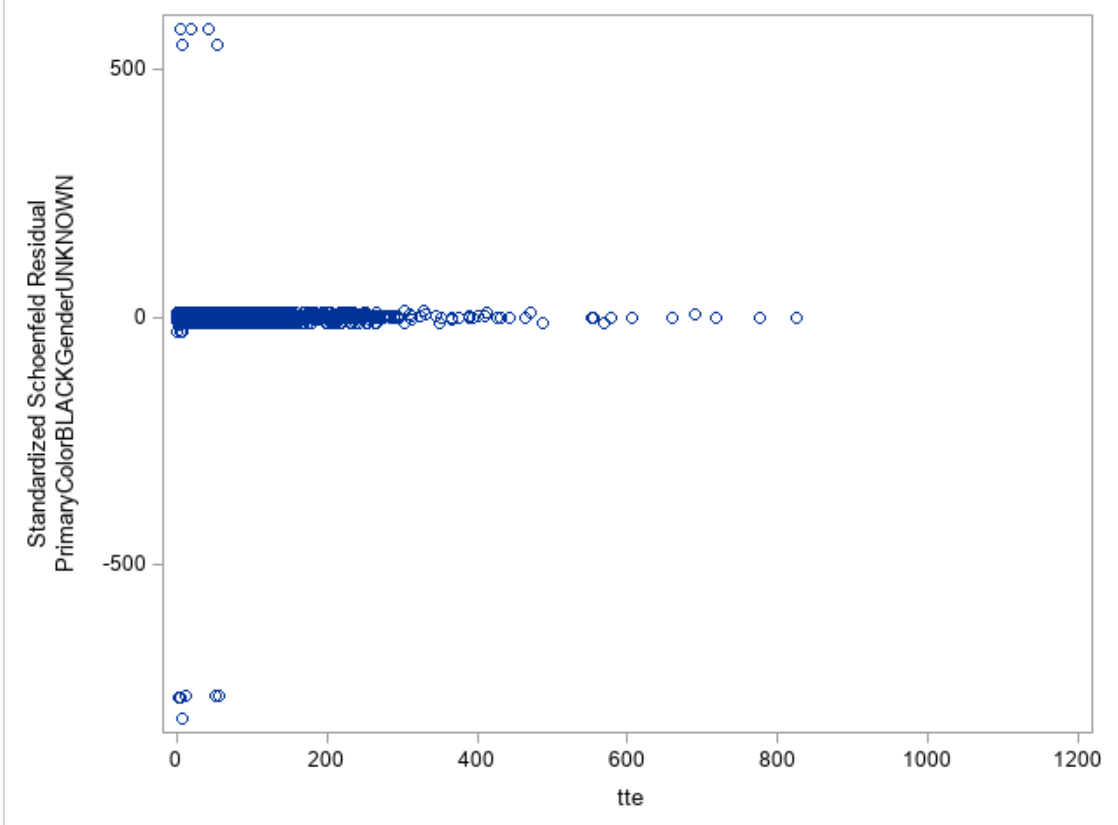
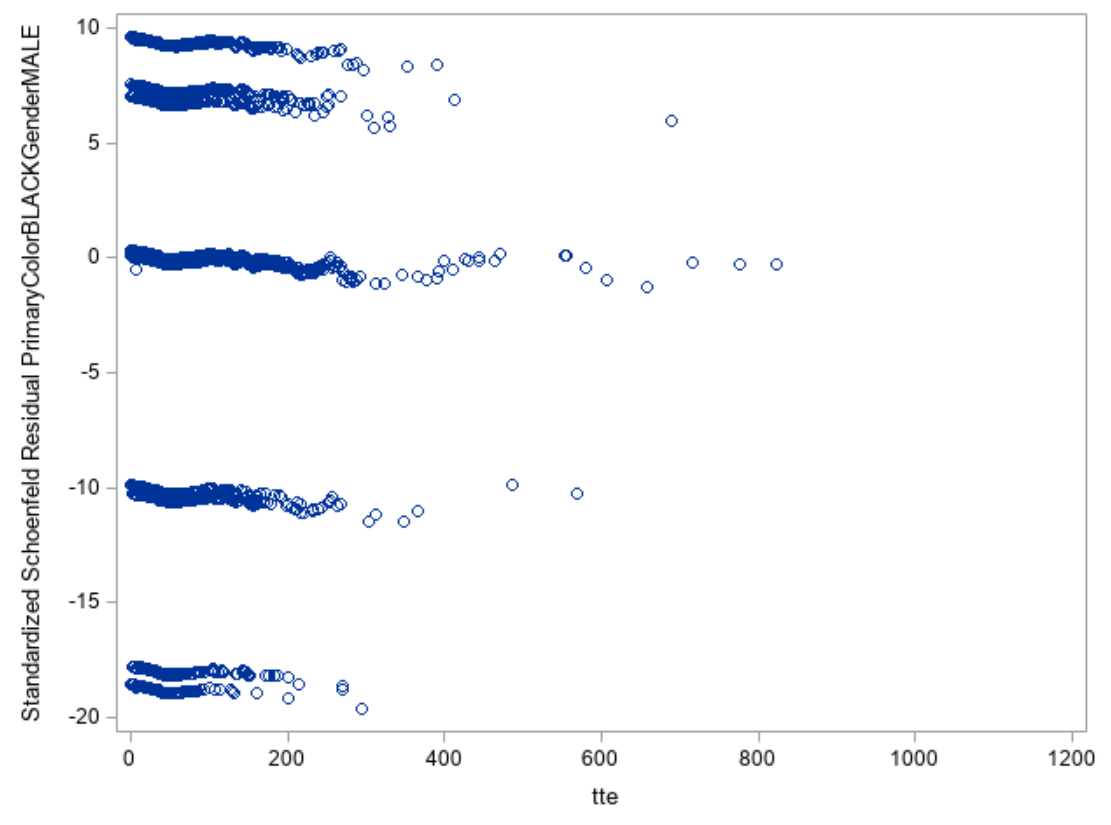


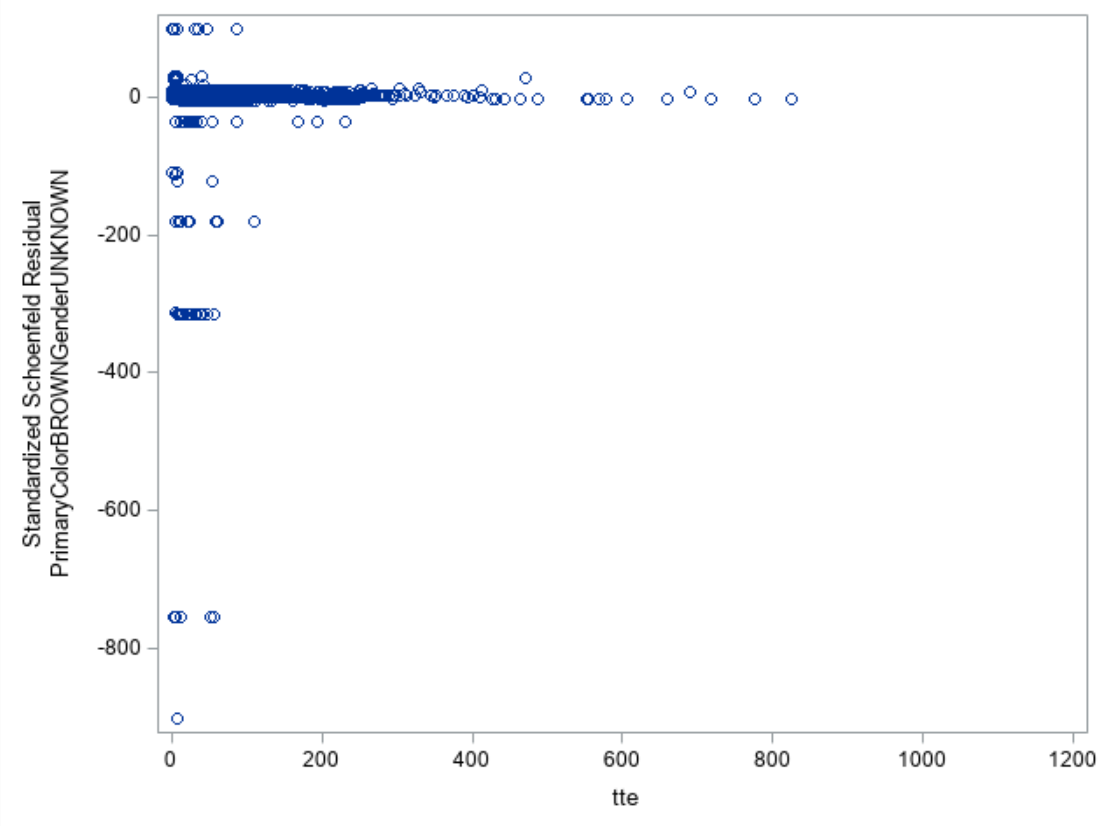
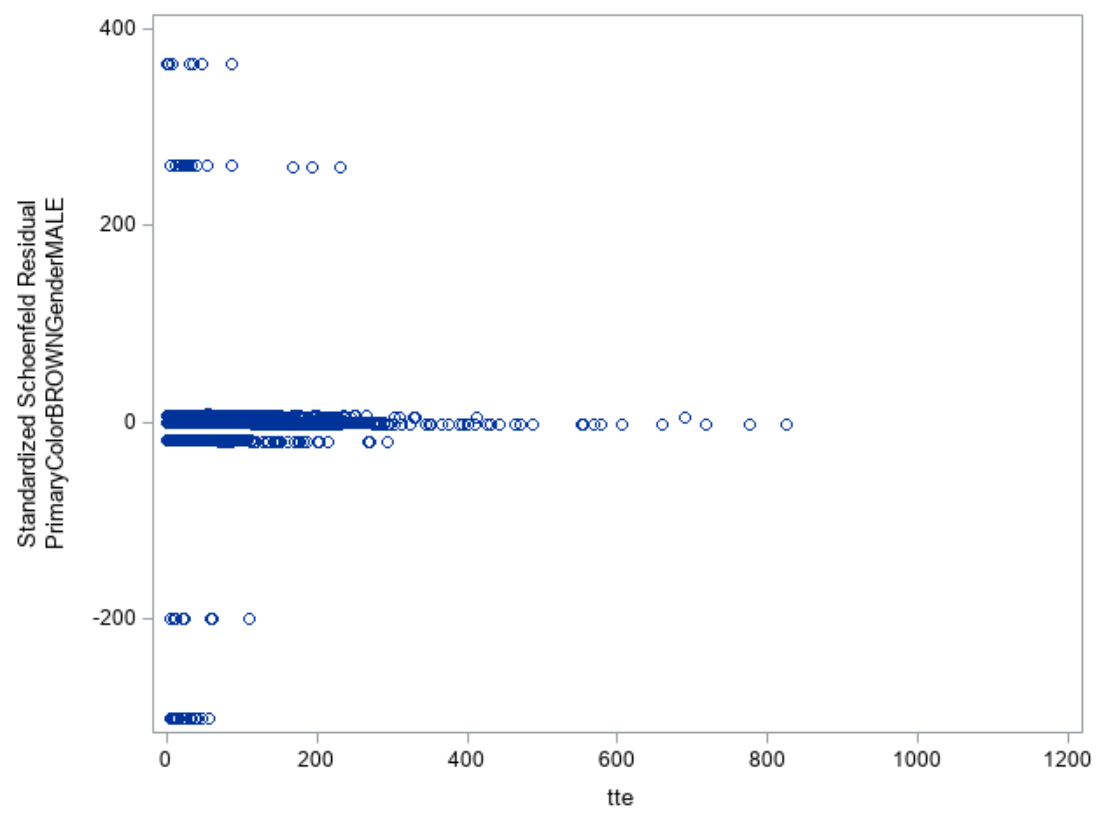


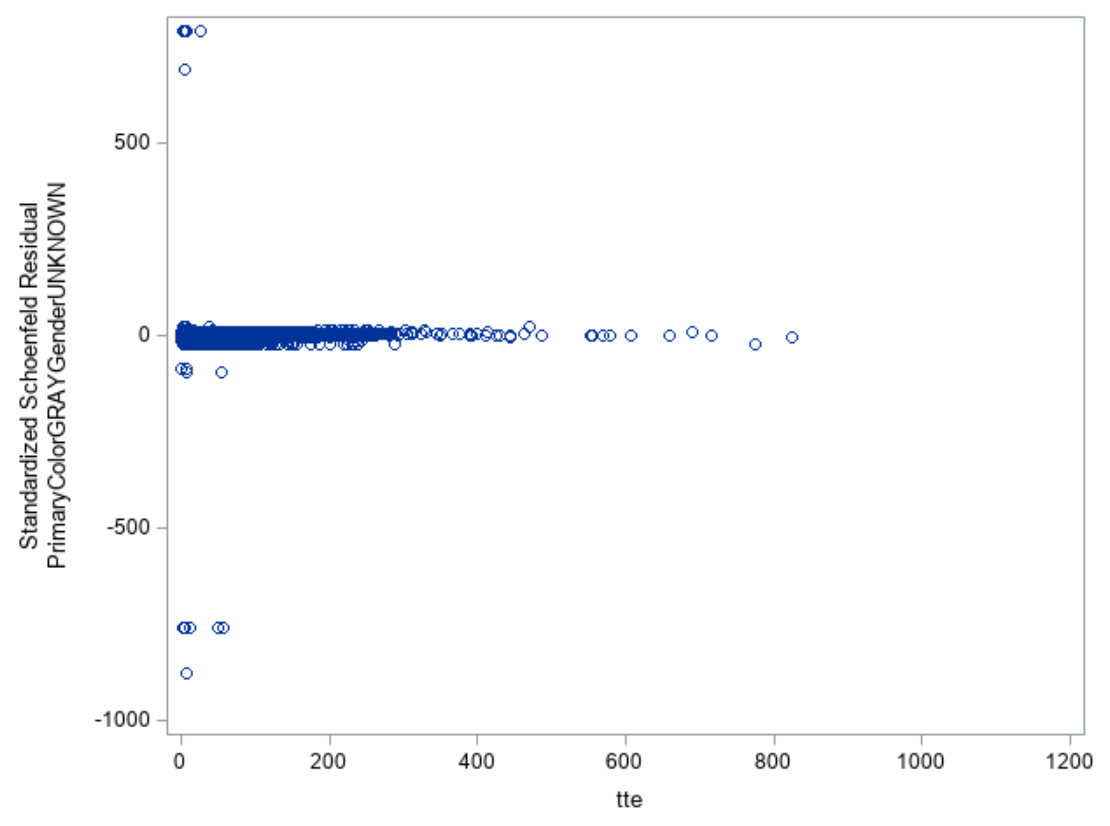
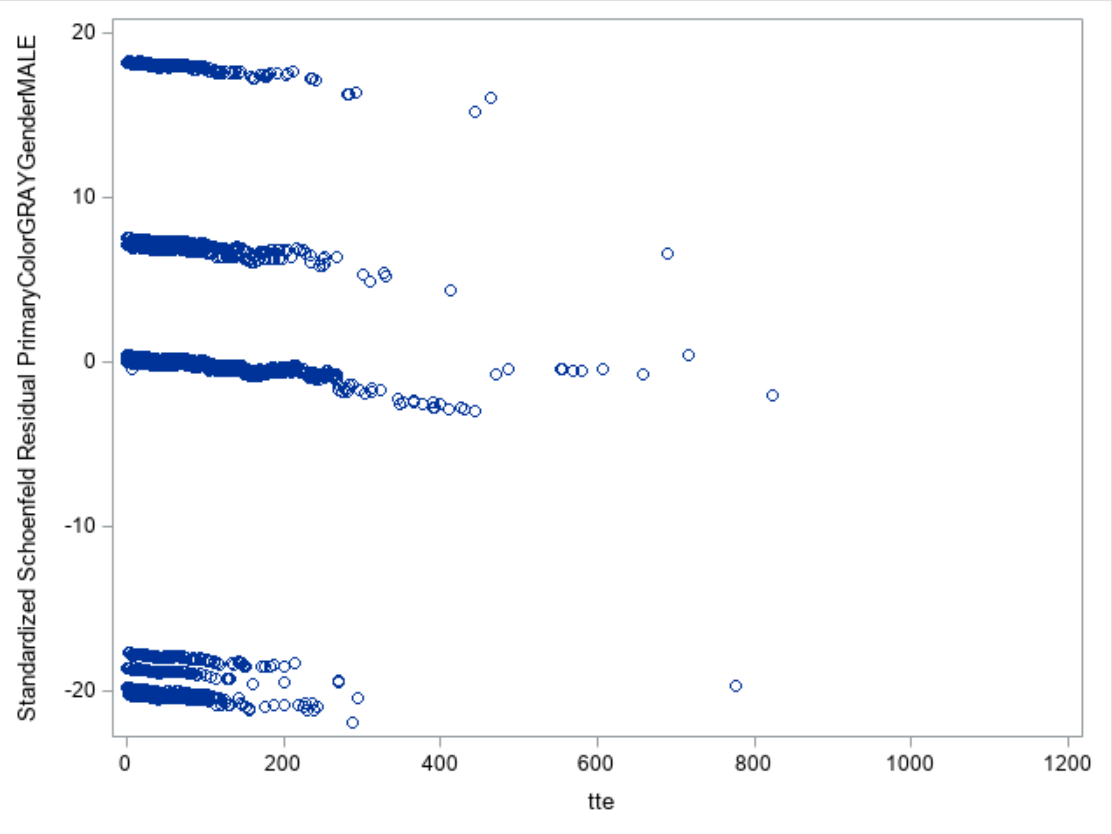


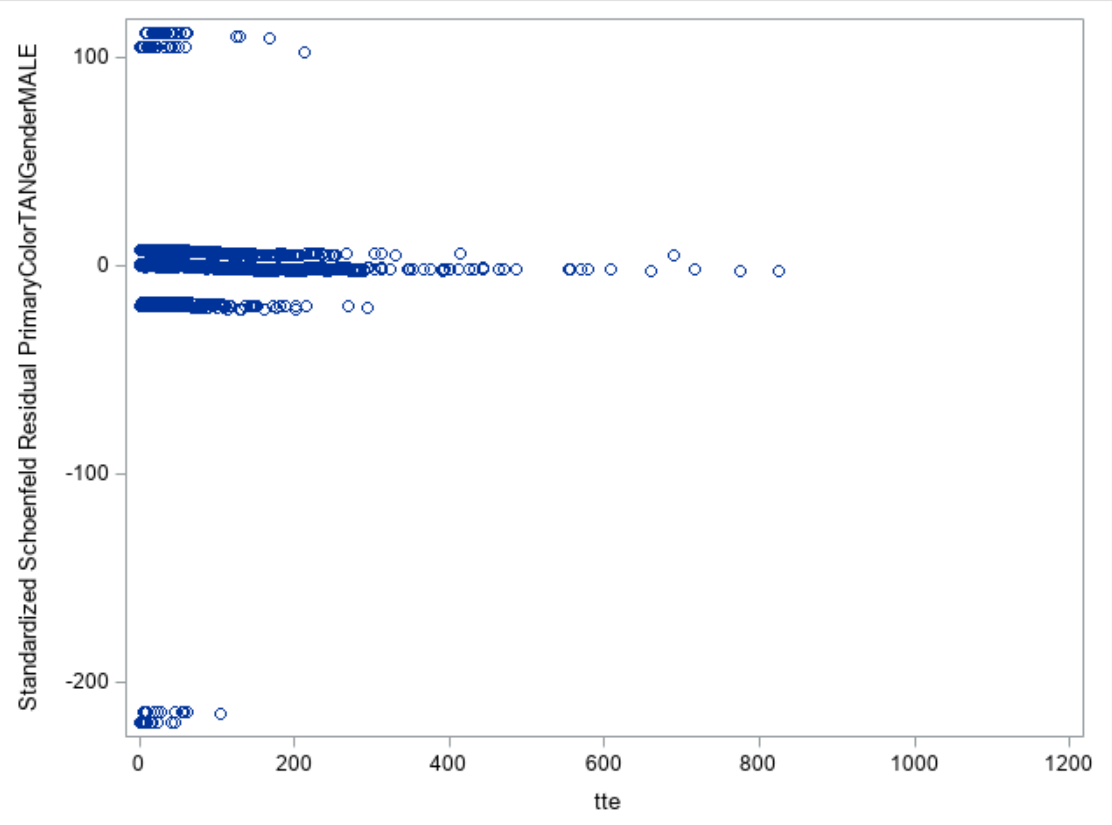


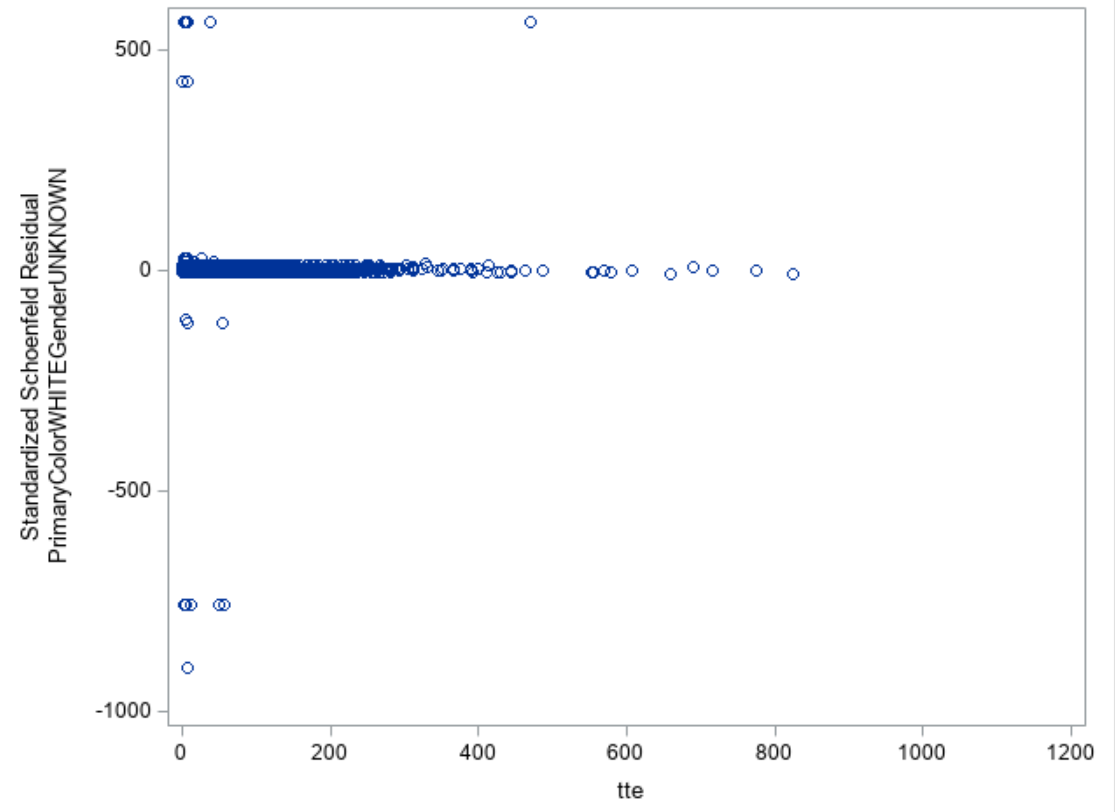
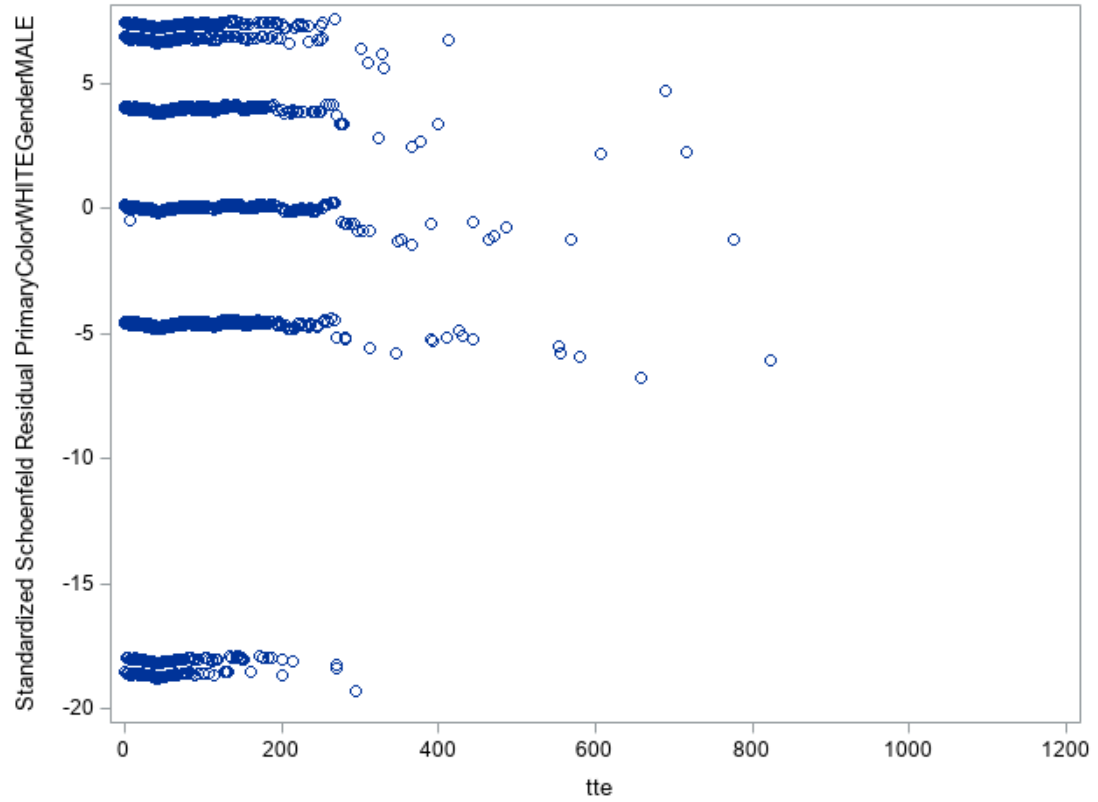












# Appendix 3: Additional detail

## *Hazard ratios and 95% confidence intervals for all variables and interactions*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **LogHR** | **StdErr** | **ChiSq** | **p-value** | **HR** | **95% lower CL** | **95% upper CL** |
| Stray | 1 | -0.27 | 0.06 | 21.92 | 0.00 | 0.77 | 0.69 | 0.86 |
| Black | 1 | -0.03 | 0.07 | 0.14 | 0.71 | 0.98 | 0.85 | 1.11 |
| Brown | 1 | 0.24 | 0.26 | 0.85 | 0.36 | 1.27 | 0.76 | 2.12 |
| Gray | 1 | 0.04 | 0.08 | 0.20 | 0.65 | 1.04 | 0.88 | 1.22 |
| Tan | 1 | 0.77 | 0.21 | 13.09 | 0.00 | 2.15 | 1.42 | 3.26 |
| White | 1 | 0.04 | 0.06 | 0.41 | 0.52 | 1.04 | 0.93 | 1.16 |
| Male | 1 | 0.29 | 0.07 | 18.33 | 0.00 | 1.34 | 1.17 | 1.53 |
| Unknown | 1 | -2.30 | 0.50 | 20.89 | 0.00 | 0.10 | 0.04 | 0.27 |
| Stray\*Black | 1 | -0.04 | 0.08 | 0.33 | 0.57 | 0.96 | 0.82 | 1.11 |
| Stray\*Brown | 1 | -0.26 | 0.30 | 0.75 | 0.39 | 0.77 | 0.43 | 1.39 |
| Stray\*Gray | 1 | -0.03 | 0.09 | 0.13 | 0.72 | 0.97 | 0.80 | 1.16 |
| Stray\*Tan | 1 | -0.39 | 0.21 | 3.38 | 0.07 | 0.68 | 0.45 | 1.03 |
| Stray\*White | 1 | 0.00 | 0.07 | 0.00 | 0.97 | 1.00 | 0.87 | 1.14 |
| Stray\*Male | 1 | -0.04 | 0.05 | 0.49 | 0.48 | 0.97 | 0.87 | 1.07 |
| Stray\*Unknown | 1 | -0.14 | 0.42 | 0.12 | 0.73 | 0.87 | 0.38 | 1.96 |
| Black\*Male | 1 | -0.22 | 0.08 | 7.44 | 0.01 | 0.80 | 0.69 | 0.94 |
| Black\*Unknown | 1 | -0.52 | 0.44 | 1.36 | 0.24 | 0.60 | 0.25 | 1.42 |
| Brown\*Male | 1 | -0.01 | 0.29 | 0.00 | 0.96 | 0.99 | 0.56 | 1.74 |
| Gray\*Male | 1 | -0.16 | 0.10 | 2.88 | 0.09 | 0.85 | 0.70 | 1.03 |
| Gray\*Unknown | 1 | -0.17 | 0.48 | 0.13 | 0.72 | 0.84 | 0.33 | 2.15 |
| Tan\*Male | 1 | -0.40 | 0.23 | 3.20 | 0.07 | 0.67 | 0.43 | 1.04 |
| Tan\*Unknown | 1 | 0.57 | 0.70 | 0.66 | 0.42 | 1.76 | 0.45 | 6.89 |
| White\*Male | 1 | -0.07 | 0.07 | 1.04 | 0.31 | 0.93 | 0.81 | 1.07 |
| White\*Unknown | 1 | -0.85 | 0.44 | 3.73 | 0.05 | 0.43 | 0.18 | 1.01 |

## *Output for final Cox PH model*

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.CATS |
| **Dependent Variable** | tte |
| **Censoring Variable** | adopted |
| **Censoring Value(s)** | 0 |
| **Ties Handling** | BRESLOW |

|  |  |
| --- | --- |
| **Number of Observations Read Number of Observations Used** | 57404 57404 |

| **Class Level Information** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Class** | **Value** | **Design Variables** | | | | |
| **Stray** | **0** | 0 |  |  |  |  |
|  | **1** | 1 |  |  |  |  |
|  |  |  |  |  |  |  |
| **PrimaryColor** | **BLACK** | 1 | 0 | 0 | 0 | 0 |
|  | **BROWN** | 0 | 1 | 0 | 0 | 0 |
|  | **GRAY** | 0 | 0 | 1 | 0 | 0 |
|  | **OTHER** | 0 | 0 | 0 | 0 | 0 |
|  | **TAN** | 0 | 0 | 0 | 1 | 0 |
|  | **WHITE** | 0 | 0 | 0 | 0 | 1 |
|  |  |  |  |  |  |  |
| **Gender** | **FEMALE** | 0 | 0 |  |  |  |
|  | **MALE** | 1 | 0 |  |  |  |
|  | **UNKNOWN** | 0 | 1 |  |  |  |

| **Summary of the Number of Event and Censored Values** | | | |
| --- | --- | --- | --- |
| **Total** | **Event** | **Censored** | **Percent Censored** |
| 57404 | 6930 | 50474 | 87.93 |

| **Convergence Status** |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Without Covariates** | **With Covariates** |
| **-2 LOG L** | 118139.18 | 116551.10 |
| **AIC** | 118139.18 | 116601.10 |
| **SBC** | 118139.18 | 116772.19 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 1588.0766 | 25 | <.0001 |
| **Score** | 1119.9892 | 25 | <.0001 |
| **Wald** | 654.7009 | 25 | <.0001 |

| **Joint Tests** | | | |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Stray** | 1 | 21.9211 | <.0001 |
| **PrimaryColor** | 5 | 15.0182 | 0.0103 |
| **Gender** | 2 | 41.1475 | <.0001 |
| **Stray\*PrimaryColor** | 5 | 4.5413 | 0.4744 |
| **Stray\*Gender** | 2 | 0.5812 | 0.7478 |
| **PrimaryColor\*Gender** | 10 | 19.4827 | 0.0345 |

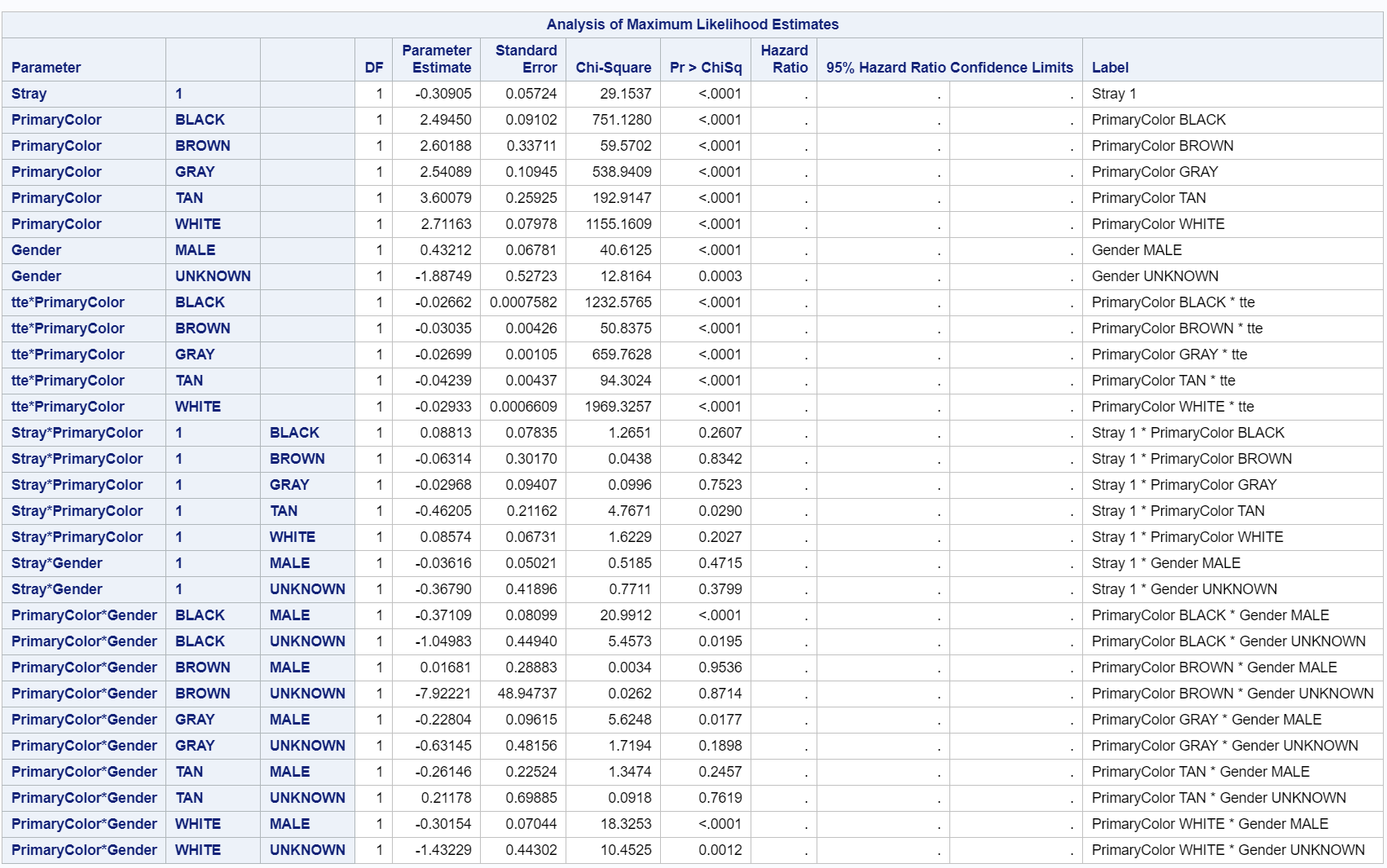
The tests of the global null hypothesis were all significant (p < 0.001), so at least one of the variables affected time to adoption. Significant joint tests for stray, primary color, gender, and primarycolor\*gender indicate that for each variable, the effect of at least one category is different from that of the other categories. See table in section above for the parameter estimates, hazard ratios, and 95% confidence intervals.

## *Attempts at controlling for potential violation of proportional hazards assumption for primary color*

Examination of the adjusted survival curves for primary color suggested that the proportional hazards assumption was violated for this variable. It was difficult to determine from the Schoenfield residual plots for each main effect of color (see Figures 1-3 for adjusted survival curves and Appendix 2 for Schoenfiled plots) if the amount of deviation from a straight line was enough to suggest significant non-proportionality.

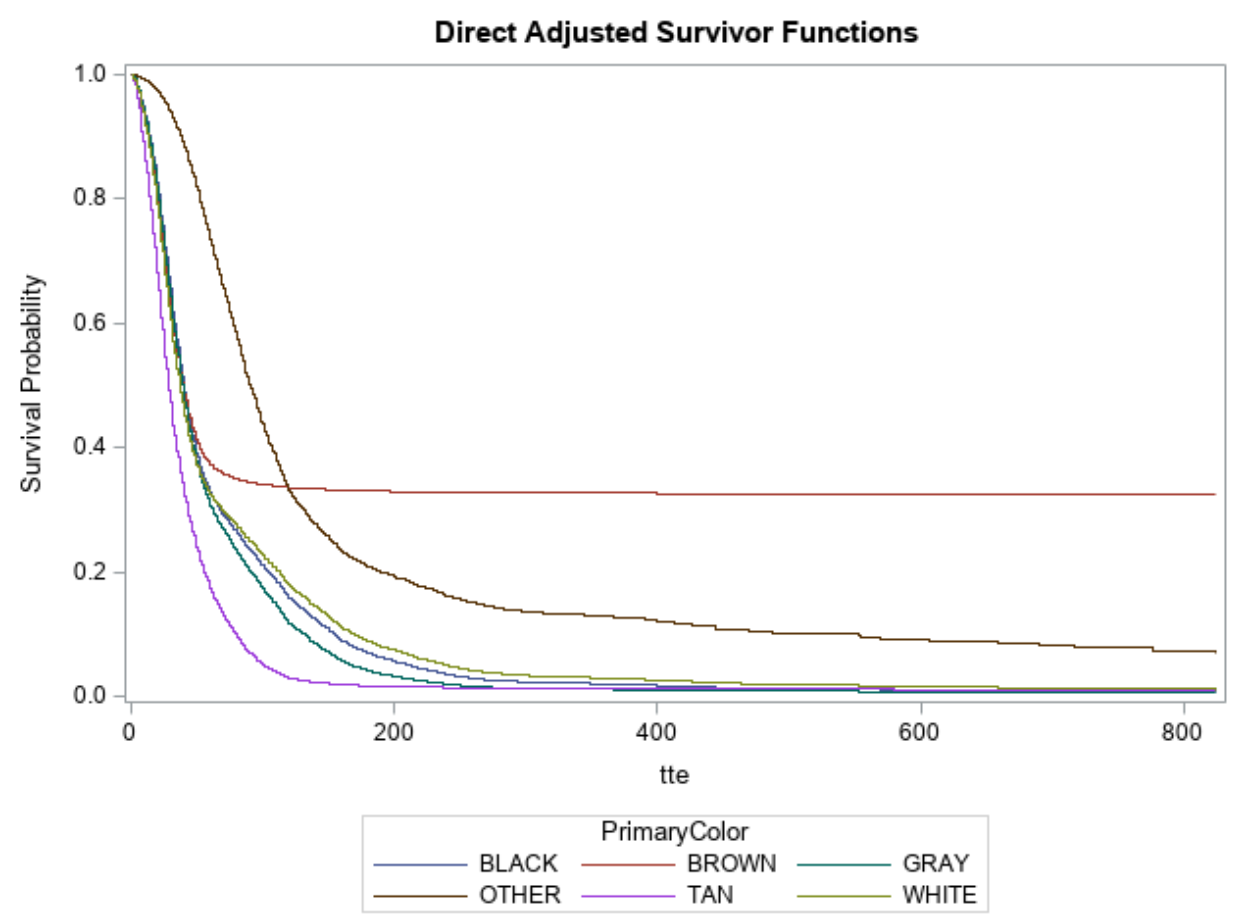
First, I attempted to adjust for the violation by adding an interaction term between color and time.

*Model estimates after adding color\*time interaction*



All time by color interactions were significant, indicating violation of the assumption of proportional hazards. I generated adjusted survival curves for stray and gender to see how adding the interactions affected these variables.

*Survival curves for primary color after adding color\*time interaction*



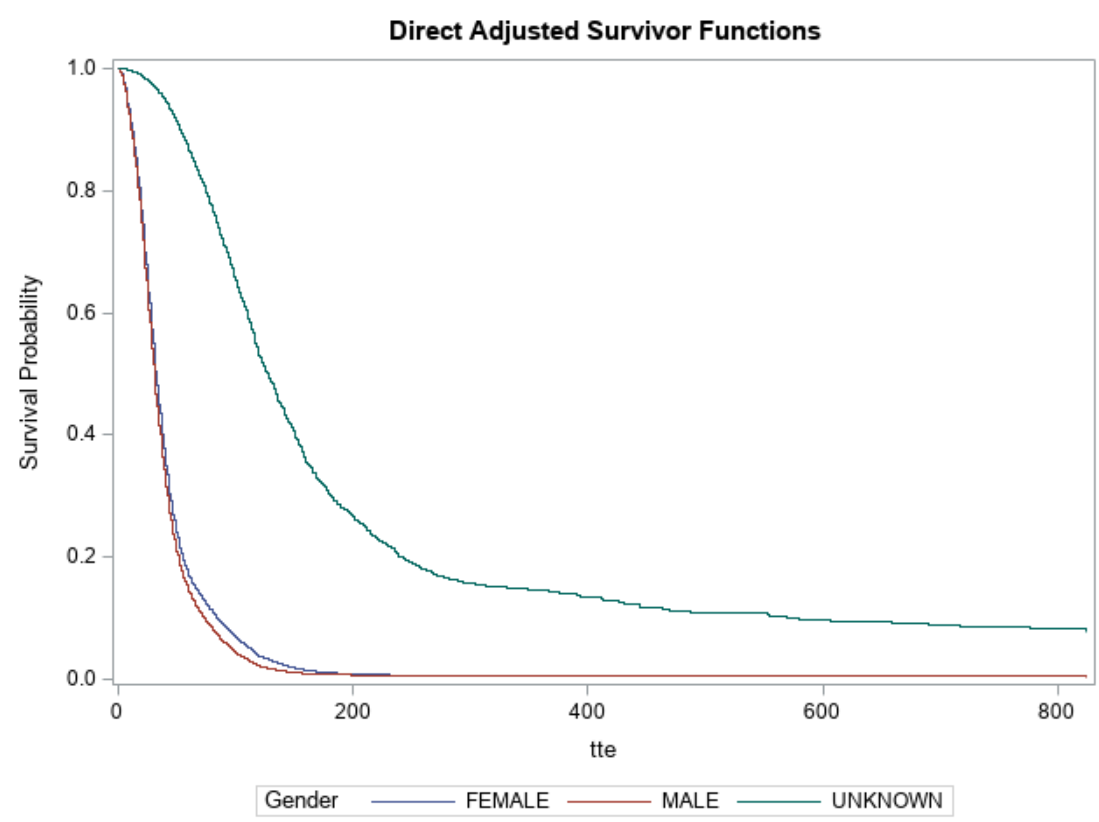
While the adjusted survival curves looked more proportional, there was still obvious overlap between them. I decided to see what effect adding color\*time interactions had on assumptions for the rest of the variables.

*Survival curves for stray after adding color\*time interaction*



Curves for stray remained parallel, so no violation observed.

*Survival curves for gender after adding color\*time interaction*



Female and male appeared to overlap, so proportionality seemed to be violated. I tried to correct this by adding another interaction term for gender by time.

*Survival curves for gender after adding both color\*time and gender\*time interactions*



Adding the gender\*time interaction didn’t seem to help with proportionality since male and female were still overlapping despite the proportional hazards assumption holding in the original model (see Figure 3). Since there weren’t many variables in the data, interpreting the data based on the effects of stray and gender for each cat color seemed limiting and inappropriate as tan and the interaction between black and being male being the only significant color-related effects in the original model. I ultimately decided it was best to just go with the original model and discuss this problem in the limitations.

# Appendix 4: SAS Analysis Code

\*read in the data;

proc import datafile = "C:\Users\meg\Desktop\Courses\Spring 2020 Courses\Survival Analysis\animalsfinal.csv"

out = animals

dbms = csv

replace;

getnames = yes;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Model fitting\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*just add in the all the variables because there aren't many of them;

\*cats analysis;

\*get subset with cats;

data cats;

set animals;

if dog = 0;

run;

\*see Table 1 stats for estimates of survival mean and median using R;

\*I attempted to use the method we learned in class with the formedian dataset;

\*But ran into processing issues because my computer is a Microsoft Surface Pro;

\*Model 1 - all factors and interactions - this is what I ran the diagnostics on;

\*cox p-h cats;

ods graphics on;

ods output ParameterEstimates = overall;

ods output HazardRatios = hrtable;

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = stray;

\*tried doing the hazard ratios this way but it was too busy - calculated manually after for \*easier plotting;

\*hazardratio stray;

\*hazardratio primarycolor;

\*hazardratio gender;

\*variables for diagnostics;

\*apparently I have to make residual variables for every combination of the factor interactions;

OUTPUT OUT=outp

survival=survival XBETA=xb LOGSURV=snell

RESMART=mart

RESDEV=dev

DFBETA=dfbetstray dfbetblack dfbetbrown dfbetgray dfbettan dfbetwhite dfbetmale dfbetunknown dfbetstrayblack

dfbetstraybrown dfbetstraygray dfbetstraytan dfbetstraywhite dfbetstraymale dfbetstrayunk dfbetblackmale

dfbetblackunk dfbetbrownmale dfbetbrownunk dfbetgraymale dfbetgrayunk dfbettanmale dfbettanunk dfbetwhitemale

dfbetwhiteunk

WTRESSCH =resschstray resschblack resschbrown resschgray resschtan resschwhite resschmale resschunknown

resschstrayblack resschstraybrown resschstraygray resschstraytan resschstraywhite resschstraymale

resschstrayunk resschblackmale resschblackunk resschbrownmale resschbrownunk resschgraymale resschgrayunk

resschtanmale resschtanunk resschwhitemale resschwhiteunk

LMAX=lmax

RESSCO=ressco;

run;

ods graphics off;

\*adjusted survival curves for color;

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = primarycolor;

run;

\*adjusted survival curves for gender;

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = gender;

run;

\*add the hazard ratios to the parameter estimate table for the forest plot;

data forestdata;

set overall;

HazardRatio = exp(Estimate);

HRLowerCL = exp(Estimate - 1.96\*StdErr);

HRUpperCL = exp(Estimate + 1.96\*StdErr);

run;

\*get rid of the anomalous brown\*unknown interaction that blew up;

\*doesn't seem to have converged;

data forest;

set forestdata;

if hruppercl >1000 then delete;

run;

\*sort by descending p-value supposedly - all significant p-values end up on top though;

proc sort data = forest;

by descending probchisq;

run;

\*forest plot of results;

\*modified code from this paper: https://support.sas.com/resources/papers/proceedings10/195-2010.pdf;

title "Hazard Ratios and 95% CIs of Potential Factors Influencing Time to Adoption";

proc sgplot data=forest;

scatter x=HazardRatio y= Label/ xerrorlower=hrlowercl

xerrorupper=hruppercl

markerattrs=HazardRatio

(symbol=DiamondFilled size=8);

refline 1 / axis=x;

xaxis label="Hazard Ratio and 95% CI " min=0;

yaxis label="Factors";

run;

title;

\*write forest data to a file that can be edited for the manuscript;

proc export

data=forestdata

dbms=xlsx

outfile="C:\Users\meg\Desktop\Courses\Spring 2020 Courses\Survival Analysis\hazardratiosfinal.xlsx"

replace;

run;

\*only select the significant ones;

data forestdata;

set forestdata;

if probchisq > 0.05 then delete;

run;

\*write forest data to a file that can be edited for the manuscript;

proc export

data=forestdata

dbms=xlsx

outfile="C:\Users\meg\Desktop\Courses\Spring 2020 Courses\Survival Analysis\table2sigfinal.xlsx"

replace;

run;

/\*Attempted additional Cox PH models based on model diagnostics\*/

\*take out top outliers when applicable and see how much the estimates change;

\*if they don't by >10%, then we can say it's not a matter of concern;

\*demonstrate this with primary color = tan b/c this is a significant effect that has 8 influential points;

\*not able to do this for all 25 terms due to time constraints - so I will just do this for primary color

\*tan;

\*delete based on IDs (see output);

data tandata;

set cats;

if animalid in (476144 274894 597720 377423 469204 609989 461135 569398 610524 512043) then delete;

run;

\*rerun proc phreg;

\*compare estimates in output to estimates from first regression and see if any have changed by >10% ;

proc phreg data = tandata plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = primarycolor;

run;

\*interpret results with caution - >10% change in the estimates for tan;

\*but not going to use this as the model as I'd probably have to keep taking outliers out;

/\*schoenfield residuals and adjusted curves for colors looked a little suspect

for gray and white - add time interactions to see if that changes things\*/

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender primarycolor\*tte stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = primarycolor;

run;

\*check how adding time\*color affects assumptions for other variables;

\*adjusted survival curves for stray;

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender primarycolor\*tte stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = stray;

run;

\*adjusted survival curves for gender;

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender primarycolor\*tte stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = gender;

run;

\*gender now looks to be non-PH -- add a time interaction to see if that helps;

\*new adjusted survival curves for gender;

proc phreg data = cats plots(overlay) = survival;

\*set reference group to 0;

class stray(ref = '0') primarycolor(ref = 'OTHER') gender(ref = 'FEMALE');

model tte\*adopted(0) = stray primarycolor gender primarycolor\*tte gender\*tte stray\*primarycolor stray\*gender primarycolor\*gender/ties = breslow rl;

baseline covariates = cats out=overall2 survival = \_all\_/diradj group = gender;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Cox-Snell Residuals\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*testing for overall fit of the model\*\*\*\*\*\*\*\*\*/

data outp;

set outp;

genres=-snell;

run;

proc lifetest data=outp outsurv=survres noprint;

time genres\*adopted(0);

run;

data survres;

set survres;

lls=log(-log(survival));

loggenr=log(genres);

run;

proc sgplot data=survres;

reg y=lls x=loggenr;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Deviance Residuals\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*testing for outliers\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

proc sgplot data=outp;

scatter y=dev x=animalid;

refline -2 2 /axis=y;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*DFBETA Residuals (score residuals)\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*testing for influential points\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 2/sqrt(57404)=0.008347548\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DFBETAs are the difference between the estimated regression coefficient

using all observations and that without the i-th individual.

This can be useful for assessing the influence of an individual.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

proc sgplot data=outp;

scatter y=dfbetstray x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetblack x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetbrown x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetgray x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbettan x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y =dfbetwhite x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetmale x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetunknown x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstrayblack x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstraybrown x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstraygray x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstraytan x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstraywhite x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstraymale x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetstrayunk x=animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetblackmale x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetblackunk x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetbrownmale x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetbrownunk x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetgraymale x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetgrayunk x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbettanmale x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbettanunk x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetwhitemale x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc sgplot data=outp;

scatter y=dfbetwhiteunk x= animalid;

refline -0.008347548 0.008347548 /axis=y;

run;

proc univariate data=outp;

var dfbetstray dfbetblack dfbetbrown dfbetgray dfbettan dfbetwhite dfbetmale dfbetunknown dfbetstrayblack

dfbetstraybrown dfbetstraygray dfbetstraytan dfbetstraywhite dfbetstraymale dfbetstrayunk dfbetblackmale

dfbetblackunk dfbetbrownmale dfbetbrownunk dfbetgraymale dfbetgrayunk dfbettanmale dfbettanunk dfbetwhitemale

dfbetwhiteunk;

id animalid;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Weighted Schoenfeld Residuals to assess time trends and lack of proportionality.\*\*\*\*\*\*\*\*\*\*/

proc sgplot data=outp;

scatter y=resschstray x=tte;

run;

proc sgplot data=outp;

scatter y=resschblack x=tte;

run;

proc sgplot data=outp;

scatter y=resschbrown x=tte;

run;

proc sgplot data=outp;

scatter y=resschgray x=tte;

run;

proc sgplot data=outp;

scatter y=resschtan x=tte;

run;

proc sgplot data=outp;

scatter y=resschwhite x=tte;

run;

proc sgplot data=outp;

scatter y=resschmale x=tte;

run;

proc sgplot data=outp;

scatter y=resschunk x=tte;

run;

proc sgplot data=outp;

scatter y=resschstrayblack x=tte;

run;

proc sgplot data=outp;

scatter y=resschstraybrown x=tte;

run;

proc sgplot data=outp;

scatter y=resschgray x=tte;

run;

proc sgplot data=outp;

scatter y=resschtan x=tte;

run;

proc sgplot data=outp;

scatter y=resschwhite x=tte;

run;

proc sgplot data=outp;

scatter y=resschstraymale x=tte;

run;

proc sgplot data=outp;

scatter y=resschstrayunk x=tte;

run;

proc sgplot data=outp;

scatter y =resschblackmale x=tte;

run;

proc sgplot data=outp;

scatter y =resschblackunk x=tte;

run;

proc sgplot data=outp;

scatter y =resschbrownmale x=tte;

run;

proc sgplot data=outp;

scatter y =resschbrownunk x=tte;

run;

proc sgplot data=outp;

scatter y =resschgraymale x=tte;

run;

proc sgplot data=outp;

scatter y =resschgrayunk x=tte;

run;

proc sgplot data=outp;

scatter y =resschtanmale x=tte;

run;

proc sgplot data=outp;

scatter y =resschtanunk x=tte;

run;

proc sgplot data=outp;

scatter y =resschwhitemale x=tte;

run;

proc sgplot data=outp;

scatter y =resschwhiteunk x=tte;

run;