

Q5

Tejal Kolte

2023-03-13

Set up

```
# Set up
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0     v purrr    1.0.1
## v tibble   3.1.8     v dplyr    1.0.10
## v tidyrr   1.2.1     v stringr  1.5.0
## v readr    2.1.3     vforcats  0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(lmtest)

## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

library(sandwich)

setwd("C:/Users/tejal/OneDrive/Desktop/UW/Data 557/Project")
```

Data Cleaning

```
# Import King County file
accidents_data <- read.csv('US_Accidents_King_County.csv')

# Clean data to use
accidents_data_clean <- accidents_data %>%
  mutate(Amenity=ifelse(Amenity=='False',0,1)) %>%
```

```

mutate(Bump=ifelse(Bump=='False',0,1)) %>%
mutate(Crossing=ifelse(Crossing=='False',0,1)) %>%
mutate(Give_Way=ifelse(Give_Way=='False',0,1)) %>%
mutate(Junction=ifelse(Junction=='False',0,1)) %>%
mutate(No_Exit=ifelse(No_Exit=='False',0,1)) %>%
mutate(Railway=ifelse(Railway=='False',0,1)) %>%
mutate(Roundabout=ifelse(Roundabout=='False',0,1)) %>%
mutate(Stop=ifelse(Stop=='False',0,1)) %>%
mutate(Station=ifelse(Station=='False',0,1)) %>%
mutate(Turning_Loop=ifelse(Turning_Loop=='False',0,1)) %>%
mutate(Traffic_Calming=ifelse(Traffic_Calming=='False',0,1)) %>%
mutate(Traffic_Signal=ifelse(Traffic_Signal=='False',0,1)) %>%
mutate(Sunrise_Sunset=ifelse(Sunrise_Sunset=='Day',0,1)) %>%
mutate(Civil_Twilight=ifelse(Civil_Twilight=='Day',0,1)) %>%
mutate(Nautical_Twilight=ifelse(Nautical_Twilight=='Day',0,1)) %>%
mutate(Astronomical_Twilight=ifelse(Astronomical_Twilight=='Day',0,1))

# Replace 0 with NA so no issues when taking the log
accidents_data_clean$Distance.mi.<- na_if(accidents_data_clean$Distance.mi., 0)

```

Model 1: Full Model

```

### MODEL 1: Full model (possible)
fit_full_dist <- lm(log(Distance.mi.)~Severity+SideR+Temperature.F.+
                     Wind_Chill.F.+Humidity...+Pressure.in.+
                     Visibility.mi.+Wind_Speed.mph.+Amenity+Bump+
                     Crossing+Give_Way+Junction+No_Exit+Railway+Roundabout+Station+
                     Stop+Traffic_Calming+Traffic_Signal+Turning_Loop+
                     Civil_Twilight+Nautical_Twilight+Astronomical_Twilight+
                     Sunrise_Sunset+weekday+hours,data=accidents_data_clean)

coeftest(fit_full_dist, vcov = vcovHC)

##
## t test of coefficients:
##
##                               Estimate Std. Error   t value Pr(>|t|)    
## (Intercept)           14.5469957  1.4148052 10.2820 < 2.2e-16 ***
## Severity            -0.0861803  0.0273471 -3.1514 0.0016307 ** 
## SideR                0.6188766  0.0523956 11.8116 < 2.2e-16 ***
## Temperature.F.      -0.0474350  0.0093369 -5.0804 3.845e-07 ***
## Wind_Chill.F.        0.0385122  0.0080319  4.7949 1.655e-06 ***
## Humidity...          -0.0063260  0.0011489 -5.5059 3.777e-08 ***
## Pressure.in.         -0.5071551  0.0466572 -10.8698 < 2.2e-16 ***
## Visibility.mi.       -0.0203121  0.0062606 -3.2444 0.0011813 ** 
## Wind_Speed.mph.      0.0026952  0.0041084  0.6560 0.5118309  
## Amenity             -0.6322654  0.0716493 -8.8244 < 2.2e-16 ***
## Bump                 0.6228061  1.2791804  0.4869 0.6263563  
## Crossing            -0.6462102  0.0672919 -9.6031 < 2.2e-16 ***
## Give_Way             -0.1751061  0.3102597 -0.5644 0.5725064  
## Junction            -0.0064512  0.0383983 -0.1680 0.8665814 

```

```

## No_Exit           -0.8298431  0.2327169 -3.5659  0.0003646 ***
## Railway          -0.2097295  0.1251708 -1.6755  0.0938629 .
## Roundabout        1.0357609  0.1569028  6.6013  4.315e-11 ***
## Station          -0.3305067  0.0775168 -4.2637  2.032e-05 ***
## Stop              -0.5138853  0.1302116 -3.9465  7.991e-05 ***
## Traffic_Calming -0.3703541  1.2665424 -0.2924  0.7699774
## Traffic_Signal   -0.5169635  0.0752270 -6.8720  6.764e-12 ***
## Civil_Twilight    0.1227869  0.0844038  1.4548  0.1457732
## Nautical_Twilight -0.0507962  0.0727492 -0.6982  0.4850476
## Astronomical_Twilight 0.0792244  0.0470242  1.6848  0.0920712 .
## Sunrise_Sunset    -0.0547493  0.0664914 -0.8234  0.4103011
## weekday           0.0567838  0.0074766  7.5949  3.400e-14 ***
## hours             0.0210338  0.0024371  8.6307 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

# Confidence Interval
`97.5%` <- (coeftest(fit_full_dist, vcov = vcovHC)[,2]*1.96) +
  coeftest(fit_full_dist, vcov = vcovHC)[,1]
`2.5%` <- (coeftest(fit_full_dist, vcov = vcovHC)[,2]*(-1.96)) +
  coeftest(fit_full_dist, vcov = vcovHC)[,1]

cbind(`2.5%`, `97.5%`)

```

	2.5%	97.5%
##		
## (Intercept)	11.773977596	17.320013806
## Severity	-0.139780555	-0.032580026
## SideR	0.516181360	0.721571922
## Temperature.F.	-0.065735385	-0.029134653
## Wind_Chill.F.	0.022769591	0.054254729
## Humidity...	-0.008577959	-0.004074093
## Pressure.in.	-0.598603106	-0.415707076
## Visibility.mi.	-0.032582979	-0.008041291
## Wind_Speed.mph.	-0.005357242	0.010747553
## Amenity	-0.772698100	-0.491832718
## Bump	-1.884387375	3.129999655
## Crossing	-0.778102417	-0.514318051
## Give_Way	-0.783215100	0.433002937
## Junction	-0.081711795	0.068809408
## No_Exit	-1.285968242	-0.373717880
## Railway	-0.455064260	0.035605230
## Roundabout	0.728231442	1.343290327
## Station	-0.482439606	-0.178573715
## Stop	-0.769100088	-0.258670435
## Traffic_Calming	-2.852777209	2.112068920
## Traffic_Signal	-0.664408417	-0.369518502
## Civil_Twilight	-0.042644613	0.288218333
## Nautical_Twilight	-0.193384672	0.091792267
## Astronomical_Twilight	-0.012943059	0.171391884
## Sunrise_Sunset	-0.185072379	0.075573873
## weekday	0.042129659	0.071438002
## hours	0.016257081	0.025810424

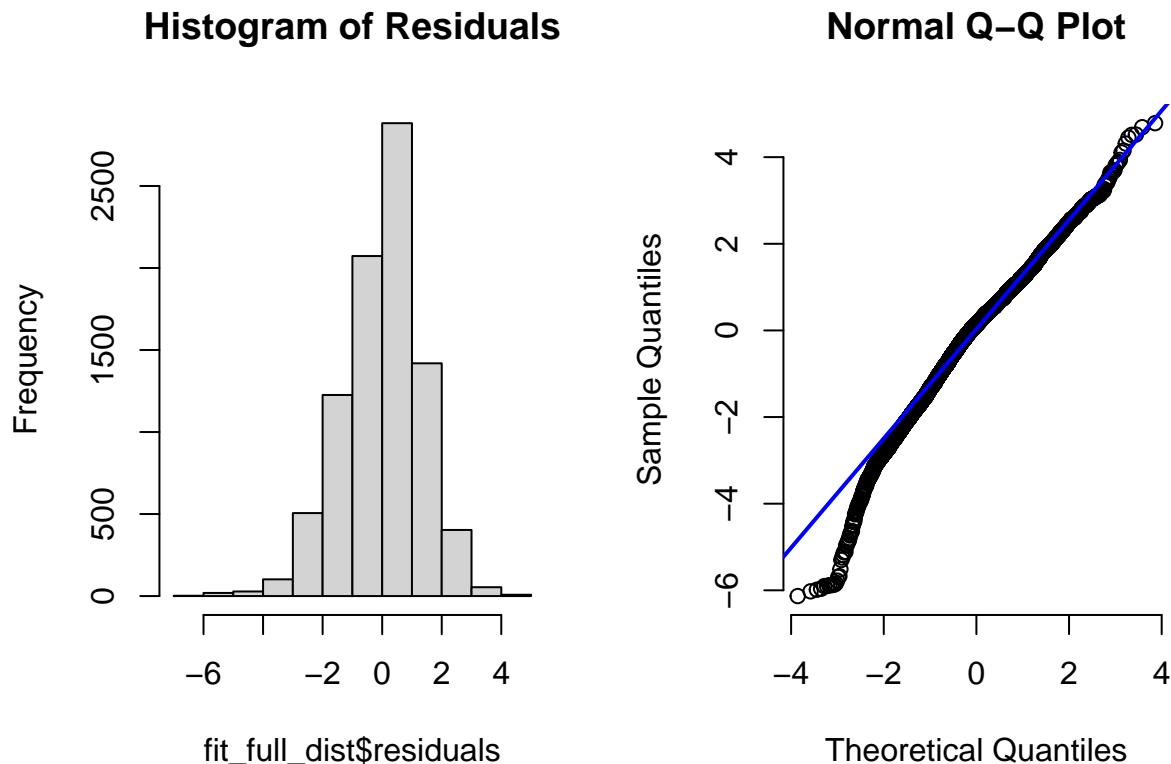
```

par(mfrow=c(1,2))
# Check assumptions

# Normality
hist(fit_full_dist$residuals,main='Histogram of Residuals')

qqnorm(fit_full_dist$residuals, pch = 1, frame = FALSE)
qqline(fit_full_dist$residuals, col = "blue", lwd = 2)

```



```

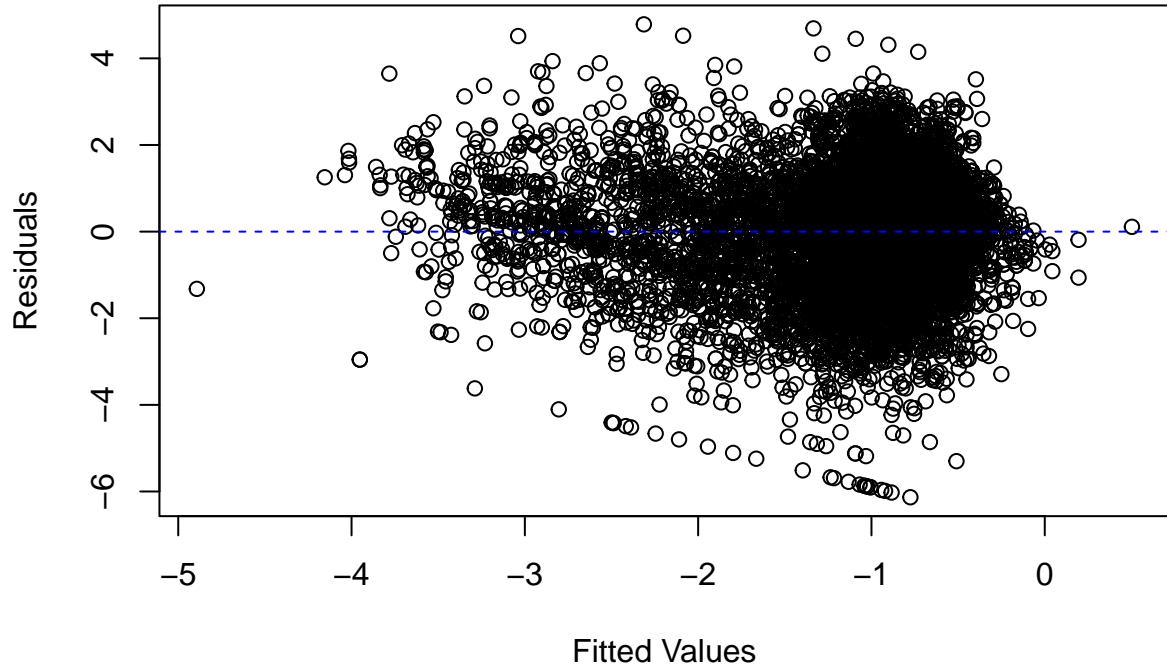
# Constant Variance
# Constant variance assumption is not met

par(mfrow=c(1,1))

plot(fit_full_dist$residuals~fit_full_dist$fitted,type='p',
      xlab='Fitted Values',ylab='Residuals',main='Residuals vs Fitted Values')
abline(h=0,col='blue',lty=2)

```

Residuals vs Fitted Values



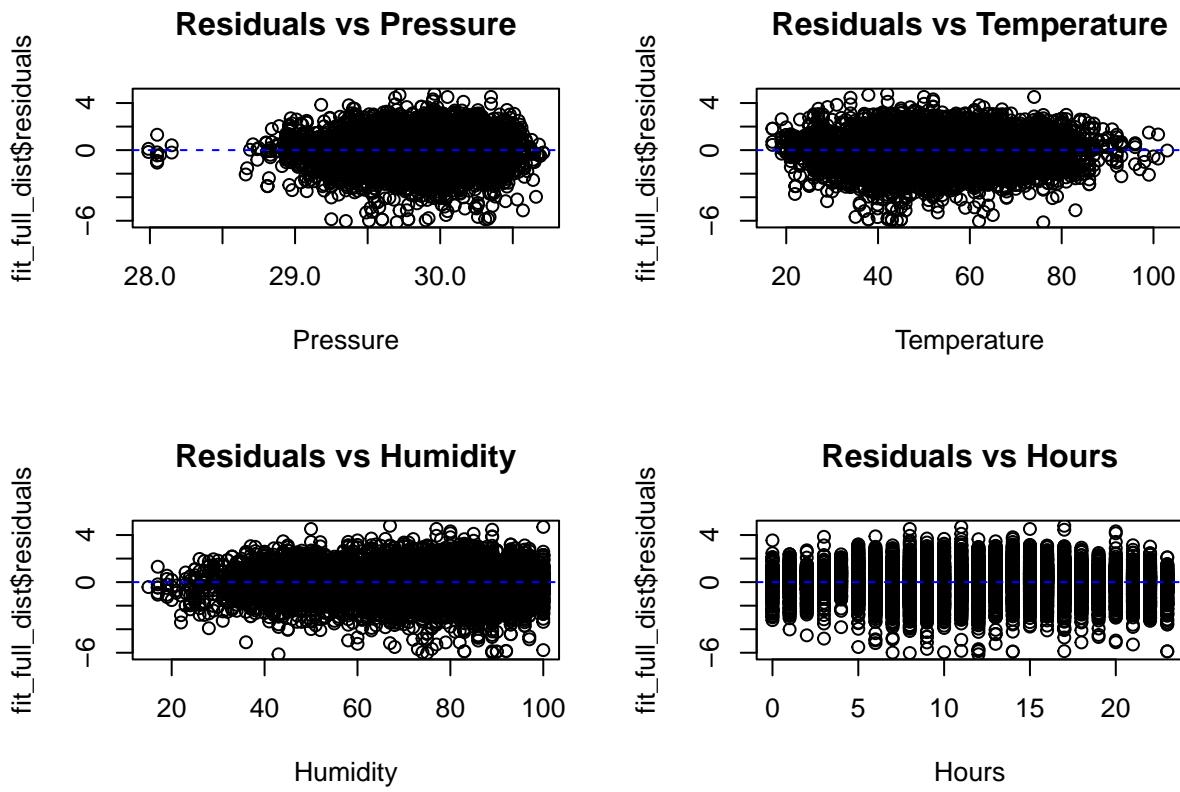
```
# Look at residuals vs some predictors

par(mfrow=c(2,2))
# Pressure
plot(fit_full_dist$residuals~fit_full_dist$model$Pressure.in.,type='p',
      xlab='Pressure',main='Residuals vs Pressure')
abline(h=0,col='blue',lty=2)

# Temperature
plot(fit_full_dist$residuals~fit_full_dist$model$Temperature.F.,type='p',
      xlab='Temperature',main='Residuals vs Temperature')
abline(h=0,col='blue',lty=2)

# Humidity
plot(fit_full_dist$residuals~fit_full_dist$model$Humidity...,type='p',
      xlab='Humidity',main='Residuals vs Humidity')
abline(h=0,col='blue',lty=2)

# Hours
plot(fit_full_dist$residuals~fit_full_dist$model$hours,type='p',
      xlab='Hours',main='Residuals vs Hours')
abline(h=0,col='blue',lty=2)
```



```
# Linearity
# Fitted vs. residuals plot seems to have a mean around 0 so linearity is met
# We will use robust standard errors to deal with the constant variance issues
```

Global F-test

```
# Model with no predictors
fit_red_0 <- lm(`log(Distance.mi.)` ~ 1, data=fit_full_dist$model)

# Global F-test
waldtest(fit_red_0, fit_full_dist, vcov=vcovHC)

## Wald test
##
## Model 1: `log(Distance.mi.)` ~ 1
## Model 2: log(Distance.mi.) ~ Severity + Side + Temperature.F. + Wind_Chill.F. +
##           Humidity... + Pressure.in. + Visibility.mi. + Wind_Speed.mph. +
##           Amenity + Bump + Crossing + Give_Way + Junction + No_Exit +
##           Railway + Roundabout + Station + Stop + Traffic_Calming +
##           Traffic_Signal + Turning_Loop + Civil_Twilight + Nautical_Twilight +
##           Astronomical_Twilight + Sunrise_Sunset + weekday + hours
##   Res.Df Df      F    Pr(>F)
## 1     8722
```

```

## 2    8696 26 81.593 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Model 2: Reduced Model

```

### MODEL 2: Reduced model

fit_reduced_dist <- lm(`log(Distance.mi.)`~Severity+Side+Temperature.F.+
                         Wind_Chill.F.+Humidity...+Pressure.in.+
                         Amenity+Crossing+No_Exit+Roundabout+Station+Stop+
                         Traffic_Signal+Astronomical_Twilight+
                         weekday+hours,data=fit_full_dist$model)

coeftest(fit_reduced_dist, vcov = vcovHC)

##
## t test of coefficients:
##
##                               Estimate Std. Error   t value Pr(>|t|)    
## (Intercept)           14.9877112  1.3833450 10.8344 < 2.2e-16 ***
## Severity              -0.0859940  0.0272362 -3.1573 0.0015976 ** 
## SideR                 0.6137255  0.0522285 11.7508 < 2.2e-16 ***
## Temperature.F.        -0.0424810  0.0074977 -5.6659 1.509e-08 ***
## Wind_Chill.F.         0.0342429  0.0064329  5.3231 1.046e-07 ***
## Humidity...            -0.0049411  0.0010099 -4.8926 1.013e-06 *** 
## Pressure.in.          -0.5320744  0.0451965 -11.7725 < 2.2e-16 ***
## Amenity                0.6338243  0.0715123 -8.8631 < 2.2e-16 ***
## Crossing               -0.6750041  0.0648009 -10.4166 < 2.2e-16 ***
## No_Exit                -0.8586901  0.2368291 -3.6258 0.0002897 *** 
## Roundabout              1.0145311  0.1551057  6.5409 6.460e-11 *** 
## Station                 -0.3322975  0.0771662 -4.3063 1.679e-05 *** 
## Stop                   -0.5042092  0.1290489 -3.9071 9.411e-05 *** 
## Traffic_Signal          -0.5178493  0.0749522 -6.9091 5.222e-12 *** 
## Astronomical_Twilight  0.0754184  0.0305189  2.4712 0.0134849 *  
## weekday                 0.0589827  0.0074406  7.9271 2.520e-15 *** 
## hours                  0.0212880  0.0024115  8.8278 < 2.2e-16 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Confidence Interval

`97.5%` <- (coeftest(fit_reduced_dist, vcov = vcovHC)[,2]*1.96) +
  coeftest(fit_reduced_dist, vcov = vcovHC)[,1]
`2.5%` <- (coeftest(fit_reduced_dist, vcov = vcovHC)[,2]*(-1.96)) +
  coeftest(fit_reduced_dist, vcov = vcovHC)[,1]

cbind(`2.5%`, `97.5%`)

```

```

##                                     2.5%      97.5%
## (Intercept)           12.276355052 17.69906741
## Severity              -0.139376919 -0.03261108
## SideR                 0.511357619  0.71609330
## Temperature.F.        -0.057176400 -0.02778559
## Wind_Chill.F.         0.021634372  0.04685149
## Humidity...            -0.006920463 -0.00296164
## Pressure.in.          -0.620659570 -0.44348915
## Amenity                -0.773988405 -0.49366015
## Crossing               -0.802013983 -0.54799431
## No_Exit                -1.322875034 -0.39450513
## Roundabout              0.710523922  1.31853837
## Station                 -0.483543175 -0.18105180
## Stop                   -0.757145103 -0.25127332
## Traffic_Signal          -0.664755678 -0.37094293
## Astronomical_Twilight   0.015601375  0.13523538
## weekday                 0.044399135  0.07356635
## hours                  0.016561539  0.02601450

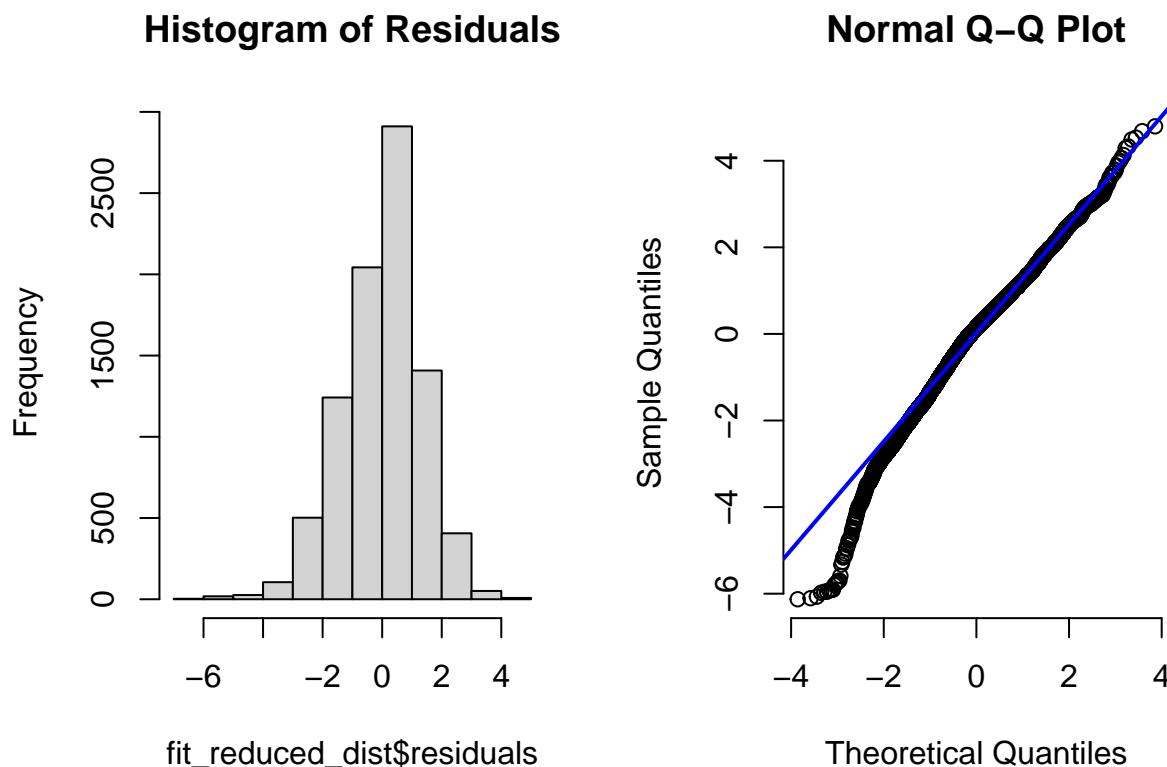
```

```

# Check assumptions (fit_reduced_dist)
par(mfrow=c(1,2))
# Normality
hist(fit_reduced_dist$residuals,main='Histogram of Residuals')

qqnorm(fit_reduced_dist$residuals, pch = 1, frame = FALSE)
qqline(fit_reduced_dist$residuals, col = "blue", lwd = 2)

```

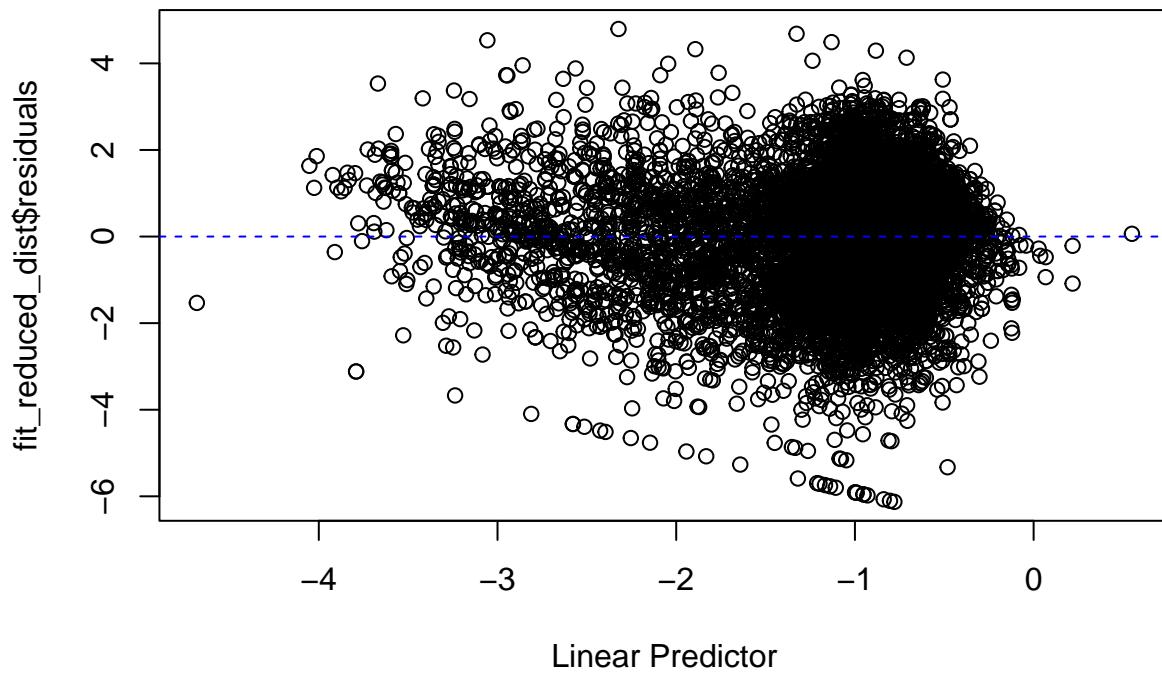


```

# Constant Variance
# Constant variance assumption is not met
par(mfrow=c(1,1))
plot(fit_reduced_dist$residuals~fit_reduced_dist$fitted, type='p',
      xlab='Linear Predictor', main='Residuals vs Fitted Values')
abline(h=0, col='blue', lty=2)

```

Residuals vs Fitted Values



```

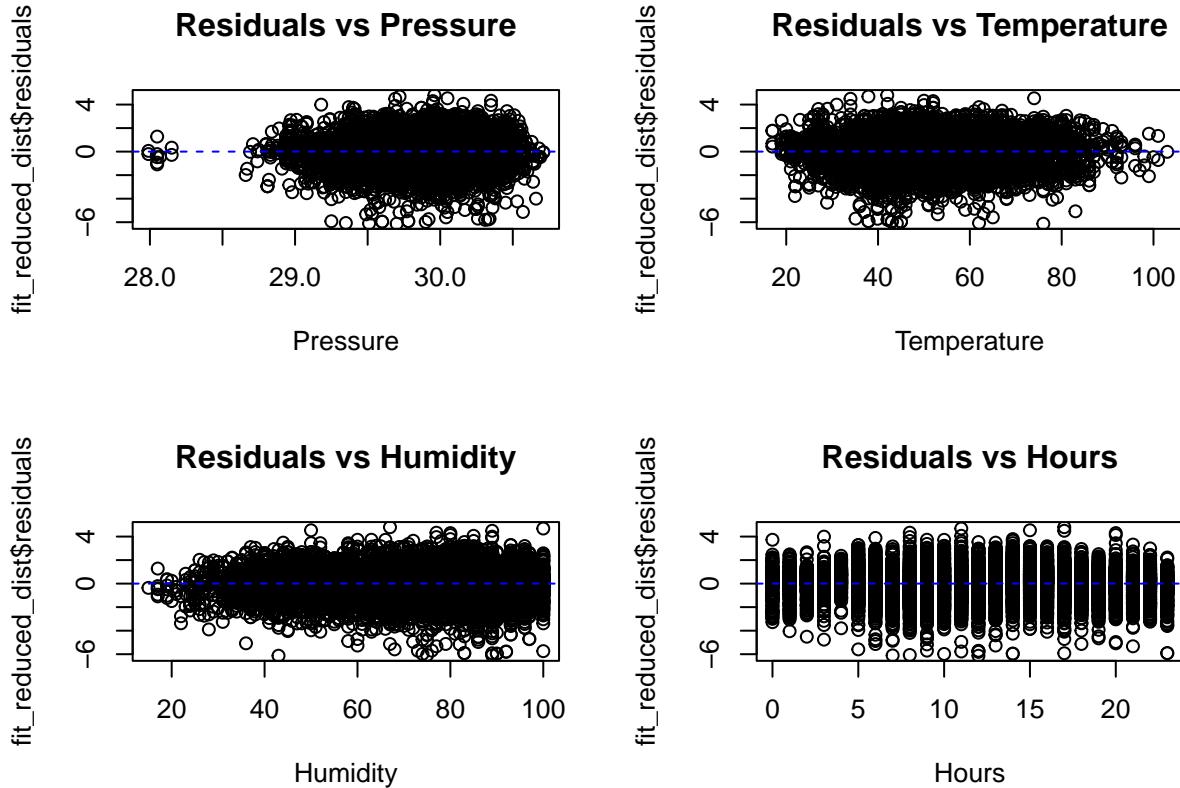
# Look at residuals vs some predictors
par(mfrow=c(2,2))
# Pressure
plot(fit_reduced_dist$residuals~fit_reduced_dist$model$Pressure.in., type='p',
      xlab='Pressure', main='Residuals vs Pressure')
abline(h=0, col='blue', lty=2)

# Temperature
plot(fit_reduced_dist$residuals~fit_reduced_dist$model$Temperature.F., type='p',
      xlab='Temperature', main='Residuals vs Temperature')
abline(h=0, col='blue', lty=2)

# Humidity
plot(fit_reduced_dist$residuals~fit_reduced_dist$model$Humidity..., type='p',
      xlab='Humidity', main='Residuals vs Humidity')
abline(h=0, col='blue', lty=2)

```

```
# Hours
plot(fit_reduced_dist$residuals~fit_reduced_dist$model$hours, type='p',
      xlab='Hours', main='Residuals vs Hours')
abline(h=0, col='blue', lty=2)
```



```
# Linearity
# Fitted vs. residuals plot seems to have a mean around 0 so linearity is met
# We will use robust standard errors to deal with the constant variance issues
```

Model 3: Further Reduced Model

```
### MODEL 3: Further Reduced model

# Does not include any type of road signs or structures
fit_reduced_dist_extra <- lm(`log(Distance.mi.)` ~ Severity + Temperature.F. +
                           Wind_Chill.F. + Humidity... + Pressure.in. +
                           Astronomical_Twilight +
                           weekday + hours, data = fit_reduced_dist$model)

coeftest(fit_reduced_dist_extra, vcov = vcovHC)
```

##

```

## t test of coefficients:
##
##                               Estimate Std. Error   t value Pr(>|t|) 
## (Intercept)           19.3423539  1.4899680 12.9817 < 2.2e-16 ***
## Severity              -0.1492547  0.0277488 -5.3788 7.694e-08 ***
## Temperature.F.        -0.0341624  0.0082019 -4.1652 3.141e-05 ***
## Wind_Chill.F.          0.0244781  0.0070103  3.4918 0.0004822 ***
## Humidity...             -0.0043280  0.0010891 -3.9740 7.123e-05 ***
## Pressure.in.            -0.6586339  0.0487560 -13.5088 < 2.2e-16 ***
## Astronomical_Twilight  0.0357037  0.0345971  1.0320 0.3021073
## weekday                 0.0433385  0.0083106  5.2148 1.882e-07 ***
## hours                   0.0189403  0.0027255  6.9493 3.936e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Confidence Interval
`97.5%` <- (coeftest(fit_reduced_dist_extra, vcov = vcovHC)[,2]*1.96) +
  coeftest(fit_reduced_dist_extra, vcov = vcovHC)[,1]
`2.5%` <- (coeftest(fit_reduced_dist_extra, vcov = vcovHC)[,2]*(-1.96)) +
  coeftest(fit_reduced_dist_extra, vcov = vcovHC)[,1]

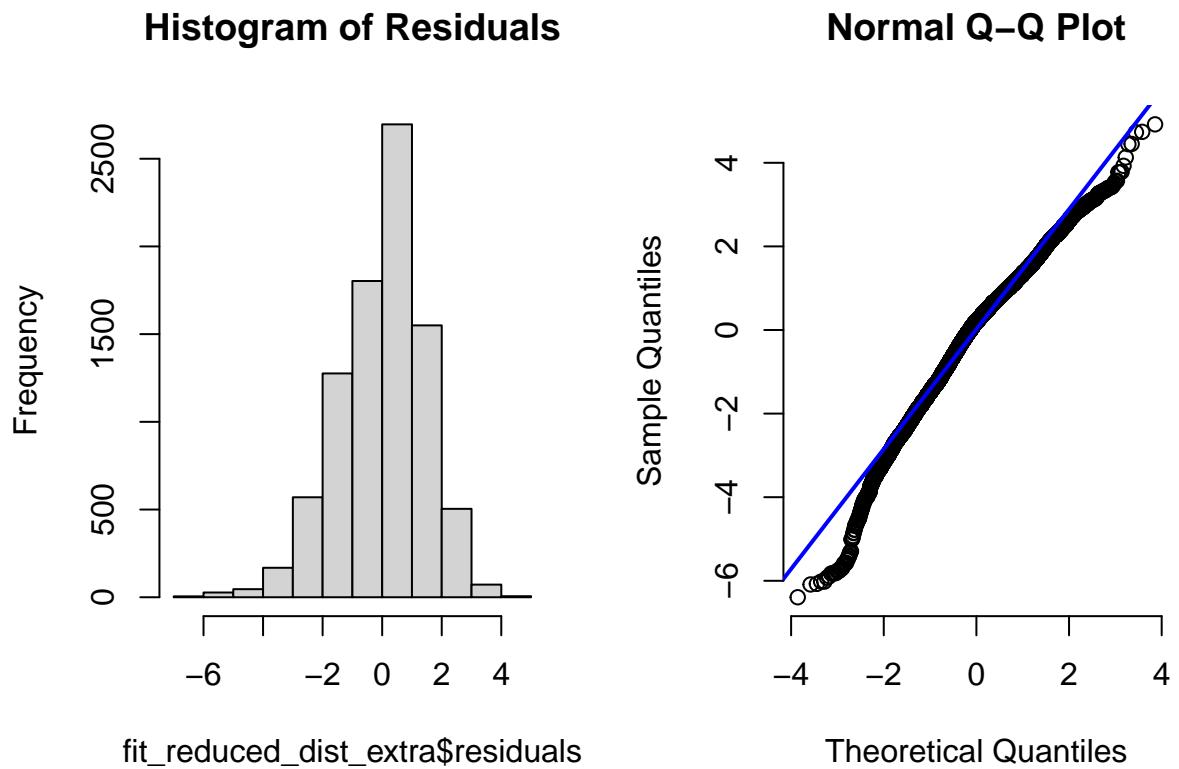
cbind(`2.5%`, `97.5%`)

##                               2.5%      97.5%
## (Intercept)           16.422016659 22.262691127
## Severity              -0.203642339 -0.094867039
## Temperature.F.        -0.050238002 -0.018086702
## Wind_Chill.F.          0.010737980  0.038218180
## Humidity...             -0.006462632 -0.002193452
## Pressure.in.            -0.754195678 -0.563072151
## Astronomical_Twilight -0.032106545  0.103513897
## weekday                  0.027049612  0.059627352
## hours                   0.013598304  0.024282291

# Check assumptions (fit_reduced_dist_extra)
par(mfrow=c(1,2))
# Normality
hist(fit_reduced_dist_extra$residuals, main='Histogram of Residuals')

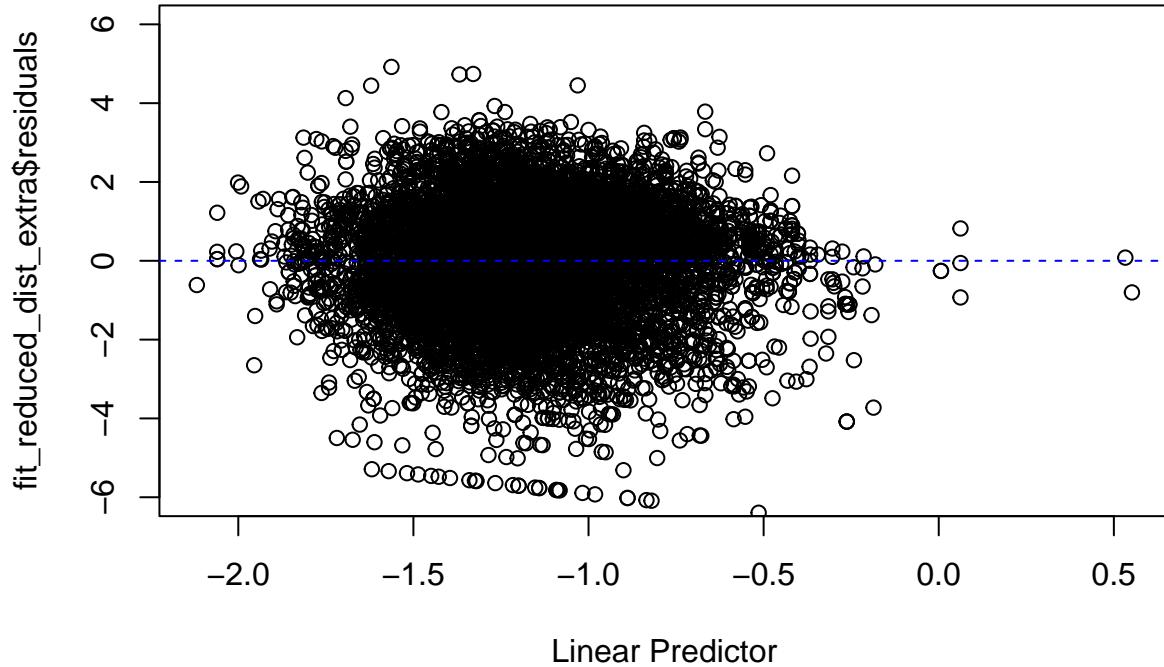
qqnorm(fit_reduced_dist_extra$residuals, pch = 1, frame = FALSE)
qqline(fit_reduced_dist_extra$residuals, col = "blue", lwd = 2)

```



```
par(mfrow=c(1,1))
# Constant Variance
plot(fit_reduced_dist_extra$residuals~fit_reduced_dist_extra$fitted,type='p',
      xlab='Linear Predictor',main='Residuals vs Fitted Values',ylim=c(-6,6))
abline(h=0,col='blue',lty=2)
```

Residuals vs Fitted Values

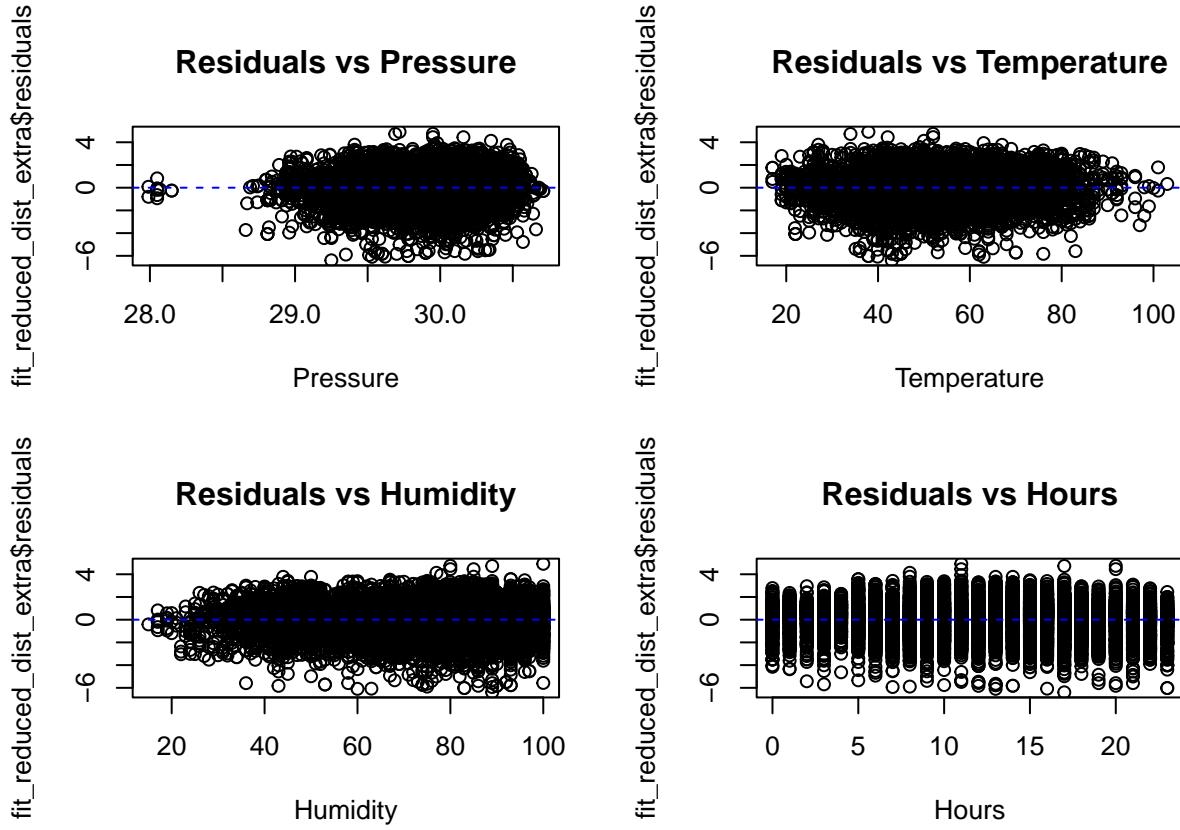


```
# Look at residuals vs some predictors
par(mfrow=c(2,2))
# Pressure
plot(fit_reduced_dist_extra$residuals~fit_reduced_dist_extra$model$Pressure.in.,type='p',
      xlab='Pressure',main='Residuals vs Pressure')
abline(h=0,col='blue',lty=2)

# Temperature
plot(fit_reduced_dist_extra$residuals~fit_reduced_dist_extra$model$Temperature.F.,type='p',
      xlab='Temperature',main='Residuals vs Temperature')
abline(h=0,col='blue',lty=2)

# Humidity
plot(fit_reduced_dist_extra$residuals~fit_reduced_dist_extra$model$Humidity...,type='p',
      xlab='Humidity',main='Residuals vs Humidity')
abline(h=0,col='blue',lty=2)

# Hours
plot(fit_reduced_dist_extra$residuals~fit_reduced_dist_extra$model$hours,type='p',
      xlab='Hours',main='Residuals vs Hours')
abline(h=0,col='blue',lty=2)
```



```
# Linearity
# Fitted vs. residuals plot seems to have a mean around 0 so linearity is met
# We will use robust standard errors to deal with the constant variance issues
```

Multiple F-Tests

```
### Multiple F-tests (robust standard errors)

# Compare full model to reduced model
# Not significant, use reduced model (MODEL 2)
waldtest(fit_reduced_dist, fit_full_dist, vcov=vcovHC)

## Wald test
##
## Model 1: 'log(Distance.mi.)' ~ Severity + Side + Temperature.F. + Wind_Chill.F. +
##           Humidity... + Pressure.in. + Amenity + Crossing + No_Exit +
##           Roundabout + Station + Stop + Traffic_Signal + Astronomical_Twilight +
##           weekday + hours
## Model 2: log(Distance.mi.) ~ Severity + Side + Temperature.F. + Wind_Chill.F. +
##           Humidity... + Pressure.in. + Visibility.mi. + Wind_Speed.mph. +
##           Amenity + Bump + Crossing + Give_Way + Junction + No_Exit +
##           Railway + Roundabout + Station + Stop + Traffic_Calming +
##           Traffic_Signal + Turning_Loop + Civil_Twilight + Nautical_Twilight +
```

```

##      Astronomical_Twilight + Sunrise_Sunset + weekday + hours
##    Res.Df Df      F  Pr(>F)
## 1     8706
## 2     8696 10 1.76 0.06228 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Compare reduced model to further reduced model
# Significant, use reduced model (MODEL 2)
waldtest(fit_reduced_dist_extra, fit_reduced_dist, vcov=vcovHC)

## Wald test
##
## Model 1: 'log(Distance.mi.)' ~ Severity + Temperature.F. + Wind_Chill.F. +
##           Humidity... + Pressure.in. + Astronomical_Twilight + weekday +
##           hours
## Model 2: 'log(Distance.mi.)' ~ Severity + Side + Temperature.F. + Wind_Chill.F. +
##           Humidity... + Pressure.in. + Amenity + Crossing + No_Exit +
##           Roundabout + Station + Stop + Traffic_Signal + Astronomical_Twilight +
##           weekday + hours
##   Res.Df Df      F      Pr(>F)
## 1     8714
## 2     8706  8 202.47 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Adjusted R-squared

```
summary(fit_full_dist) # Adjusted R-squared:  0.1863
```

```

##
## Call:
## lm(formula = log(Distance.mi.) ~ Severity + Side + Temperature.F. +
##     Wind_Chill.F. + Humidity... + Pressure.in. + Visibility.mi. +
##     Wind_Speed.mph. + Amenity + Bump + Crossing + Give_Way +
##     Junction + No_Exit + Railway + Roundabout + Station + Stop +
##     Traffic_Calming + Traffic_Signal + Turning_Loop + Civil_Twilight +
##     Nautical_Twilight + Astronomical_Twilight + Sunrise_Sunset +
##     weekday + hours, data = accidents_data_clean)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -6.1325 -0.8216  0.1380  0.8800  4.7849
##
## Coefficients: (1 not defined because of singularities)
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.546996  1.458563  9.974 < 2e-16 ***
## Severity    -0.086180  0.025704 -3.353 0.000803 ***
## SideR       0.618877  0.051946 11.914 < 2e-16 ***
## Temperature.F. -0.047435  0.009494 -4.996 5.95e-07 ***
## Wind_Chill.F.  0.038512  0.008131  4.736 2.21e-06 ***

```

```

## Humidity...      -0.006326  0.001158  -5.465 4.76e-08 ***
## Pressure.in.   -0.507155  0.047935  -10.580 < 2e-16 ***
## Visibility.mi. -0.020312  0.006343  -3.203 0.001367 **
## Wind_Speed.mph. 0.002695  0.004102   0.657 0.511146
## Amenity        -0.632265  0.067341  -9.389 < 2e-16 ***
## Bump            0.622806  0.822508   0.757 0.448948
## Crossing       -0.646210  0.058335  -11.078 < 2e-16 ***
## Give_Way        -0.175106  0.294782  -0.594 0.552515
## Junction       -0.006451  0.045612  -0.141 0.887527
## No_Exit         -0.829843  0.201111  -4.126 3.72e-05 ***
## Railway         -0.209730  0.110474  -1.898 0.057669 .
## Roundabout      1.035761  0.780260   1.327 0.184393
## Station          -0.330507  0.069153  -4.779 1.79e-06 ***
## Stop             -0.513885  0.115082  -4.465 8.09e-06 ***
## Traffic_Calming -0.370354  0.771572  -0.480 0.631240
## Traffic_Signal  -0.516963  0.063976  -8.081 7.31e-16 ***
## Turning_Loop     NA          NA          NA          NA
## Civil_Twilight   0.122787  0.084289   1.457 0.145227
## Nautical_Twilight -0.050796  0.072649  -0.699 0.484448
## Astronomical_Twilight  0.079224  0.048729   1.626 0.104028
## Sunrise_Sunset    -0.054749  0.066296  -0.826 0.408925
## weekday           0.056784  0.007727   7.349 2.17e-13 ***
## hours             0.021034  0.002554   8.236 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.335 on 8696 degrees of freedom
##   (7180 observations deleted due to missingness)
## Multiple R-squared:  0.1888, Adjusted R-squared:  0.1863
## F-statistic: 77.82 on 26 and 8696 DF,  p-value: < 2.2e-16

```

```
summary(fit_reduced_dist) # Adjusted R-squared: 0.1856
```

```

##
## Call:
## lm(formula = 'log(Distance.mi.)' ~ Severity + Side + Temperature.F. +
##   Wind_Chill.F. + Humidity... + Pressure.in. + Amenity + Crossing +
##   No_Exit + Roundabout + Station + Stop + Traffic_Signal +
##   Astronomical_Twilight + weekday + hours, data = fit_full_dist$model)
##
## Residuals:
##      Min      1Q      Median      3Q      Max 
## -6.1284 -0.8247  0.1376  0.8645  4.7945 
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 14.987711  1.423016 10.532 < 2e-16 ***
## Severity    -0.085994  0.025665 -3.351 0.00081 ***
## SideR        0.613725  0.051810 11.846 < 2e-16 ***
## Temperature.F. -0.042481  0.007807 -5.441 5.44e-08 ***
## Wind_Chill.F.  0.034243  0.006658  5.143 2.76e-07 ***
## Humidity...   -0.004941  0.001057 -4.676 2.98e-06 ***
## Pressure.in. -0.532074  0.046271 -11.499 < 2e-16 ***
## Amenity      -0.633824  0.067245 -9.426 < 2e-16 ***

```

```

## Crossing           -0.675004  0.056976 -11.847 < 2e-16 ***
## No_Exit            -0.858690  0.200459 -4.284 1.86e-05 ***
## Roundabout          1.014531  0.780453  1.300  0.19366
## Station             -0.332297  0.069130 -4.807 1.56e-06 ***
## Stop                -0.504209  0.114762 -4.394 1.13e-05 ***
## Traffic_Signal      -0.517849  0.063930 -8.100 6.23e-16 ***
## Astronomical_Twilight 0.075418  0.032225  2.340  0.01929 *
## weekday              0.058983  0.007695  7.666 1.97e-14 ***
## hours                 0.021288  0.002540  8.381 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.336 on 8706 degrees of freedom
## Multiple R-squared:  0.1871, Adjusted R-squared:  0.1856
## F-statistic: 125.3 on 16 and 8706 DF, p-value: < 2.2e-16

```

```
summary(fit_reduced_dist_extra) # Adjusted R-squared: 0.03513
```

```

##
## Call:
## lm(formula = 'log(Distance.mi.)' ~ Severity + Temperature.F. +
##     Wind_Chill.F. + Humidity... + Pressure.in. + Astronomical_Twilight +
##     weekday + hours, data = fit_reduced_dist$model)
##
## Residuals:
##    Min      1Q  Median      3Q      Max
## -6.3935 -0.9487  0.2030  0.9853  4.9197
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               19.342354   1.538410 12.573 < 2e-16 ***
## Severity                  -0.149255   0.027797 -5.370 8.10e-08 ***
## Temperature.F.             -0.034162   0.008495 -4.022 5.83e-05 ***
## Wind_Chill.F.              0.024478   0.007242  3.380 0.000728 ***
## Humidity...                 -0.004328   0.001150 -3.764 0.000168 ***
## Pressure.in.                -0.658634   0.050127 -13.139 < 2e-16 ***
## Astronomical_Twilight      0.035704   0.035040  1.019 0.308263
## weekday                     0.043338   0.008358  5.185 2.21e-07 ***
## hours                       0.018940   0.002763  6.854 7.67e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.454 on 8714 degrees of freedom
## Multiple R-squared:  0.03602, Adjusted R-squared:  0.03513
## F-statistic: 40.7 on 8 and 8714 DF, p-value: < 2.2e-16

```