More frequent events

```
knitr::opts_chunk$set(echo = TRUE)
source("../Rcode/events.R")
```

Define 2 Groups

```
n <- 20

#Group 1
AE.1 <- AE(10,3,20) # time to initial event , duration , gap in between
AE.2 <- AE(20,4,30)
AE.3 <- AE(50,10,40)

AEs.1 <- list(AE.1,AE.2,AE.3)

scores.1 <- c()
for(i in 1:n){
  AEs_.1  <- simulate_events(AEs.1,180)
  scores.1 <- c(scores.1,AEs_.1$Score)
}
scores.1
```

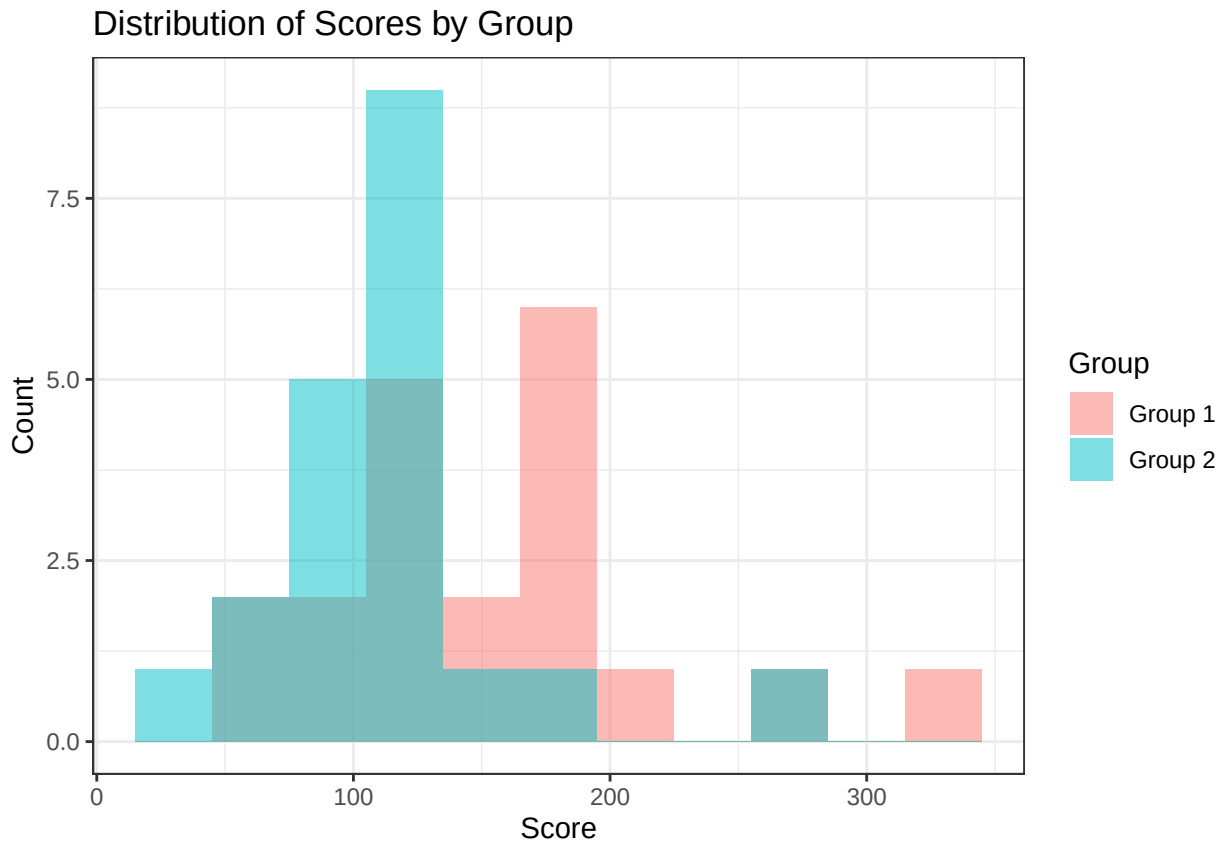
```
##  [1] 108.18904  73.38076 104.93798 181.68604 167.34373 180.17511 261.70060 113.35025 198.36478 187.73
## [12] 158.61278 145.46436 128.04332 328.81521 176.99445  75.57229  65.31855 129.37025 183.34413
```

```
#Group 2
AE.1 <- AE(10,3,40)
AE.2 <- AE(20,4,60)
AE.3 <- AE(50,10,80)

AEs.2 <- list(AE.1,AE.2,AE.3)

scores.2 <- c()
for(i in 1:n){
  AEs_.2  <- simulate_events(AEs.2,180)
scores.2 <- c(scores.2,AEs_.2$Score)
}
scores.2
```

```
##  [1] 133.05469 124.93796 139.15364 113.84422  71.00362 105.17620 131.31994 117.48269 124.64248 264.01
## [12]  33.93933  88.74338  78.57551  76.11886 103.71923 113.78515 181.60199  99.22964  58.81376
```

```r
## put both scores in a df for with the groups as a factor variable
scores <- c(scores.1,scores.2)
groups <- factor(c(rep(1,length(scores.1)),rep(2,length(scores.2))))
score_data <- data.frame(scores,groups)

#score_data
```

```r
# plot the distriubtion of the scores

#install.packages("ggplot2")
library(ggplot2)

ggplot(score_data, aes(x = scores, fill = groups)) +
  geom_histogram(binwidth = 30, position = "identity", alpha =0.5) +
  labs(title = "Distribution of Scores by Group", x = "Score", y = "Count") +
  scale_fill_discrete(name = "Group", labels = c("Group 1", "Group 2")) +
  theme_bw()
```




```r
# First a simple linear regression model
lm_model <- lm(scores ~ groups, data = score_data)

summary(lm_model)
```

```
##
## Call:
```

```
## lm(formula = scores ~ groups, data = score_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -88.42 -38.95  -0.78  23.58 175.07
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   153.74      12.70  12.105 1.31e-14 ***
## groups2       -39.15      17.96  -2.179   0.0356 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.8 on 38 degrees of freedom
## Multiple R-squared:  0.1111, Adjusted R-squared:  0.08772
## F-statistic:  4.75 on 1 and 38 DF,  p-value: 0.03557
```

```
# Fit a Tweedie regression model
#install.packages("tweedie")
#install.packages("statmod")

library(tweedie)
library(statmod)

tweedie_model <- glm(score_data$scores~score_data$groups, family =
           tweedie(var.power=1,
           link.power=0), control = glm.control(maxit = 100))

tweedie_model.2 <- glm(score_data$scores~score_data$groups, family =
           tweedie(var.power=1.5,
           link.power=0), control = glm.control(maxit = 100))

tweedie_model.3 <- glm(score_data$scores~score_data$groups, family =
           tweedie(var.power=2,
           link.power=0), control = glm.control(maxit = 100))

tweedie_model.4 <- glm(score_data$scores~score_data$groups, family =
           tweedie(var.power=1,
           link.power=1), control = glm.control(maxit = 100))
summary(tweedie_model.4)
```

```
##
## Call:
## glm(formula = score_data$scores ~ score_data$groups, family = tweedie(var.power = 1,
##     link.power = 1), control = glm.control(maxit = 100))
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          153.74      13.46  11.425 7.42e-14 ***
## score_data$groups2   -39.15      17.78  -2.202   0.0338 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Tweedie family taken to be 23.55605)
```

```
##
##     Null deviance: 959.28  on 39  degrees of freedom
## Residual deviance: 844.66  on 38  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
```

```
# Fit quantile regression model
#install.packages("quantreg")

library(quantreg)
quantile_regression_model <- rq(score_data$scores~score_data$groups)
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
summary(quantile_regression_model, se="nid") #https://cran.r-project.org/web/packages/quantreg/vignette
```

```
##
## Call: rq(formula = score_data$scores ~ score_data$groups)
##
## tau: [1] 0.5
##
## Coefficients:
##                     Value     Std. Error t value   Pr(>|t|)
## (Intercept)         145.46436  15.13915    9.60849   0.00000
## score_data$groups2 -31.62014  18.51416   -1.70789   0.09582
```

AIC analysis

```
AIC(lm_model)
```

```
## [1] 440.6237
```

```
AICtweedie(tweedie_model,dispersion=1) # simple AIC is NAN
```

```
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 108.189036

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 73.380758

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 104.937977

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 181.686037

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 167.343728

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 180.175109

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 261.700603
```

```
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 113.350254
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 198.364785
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 187.730532
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 106.408847
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 158.612782
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 145.464360
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 128.043321
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 328.815212
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 176.994446
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 75.572287
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 65.318552
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 129.370255
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 183.344132
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 133.054685
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 124.937958
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 139.153639
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 113.844217
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 71.003619
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 105.176197
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 131.319942
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 117.482693
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 124.642477
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 264.016300
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 132.735137
```

```
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 33.939329

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 88.743375

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 78.575511

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 76.118864

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 103.719226

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 113.785148

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 181.601993

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 99.229643

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 58.813759

## [1] Inf
```

```
AICtweedie(tweedie_model.2,dispersion=1)
```

```
## [1] 439.6377
```

```
AICtweedie(tweedie_model.3,dispersion=1)
```

```
## [1] 475.0666
```

```
AICtweedie(tweedie_model.4,dispersion=1)
```

```
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 108.189036

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 73.380758

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 104.937977

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 181.686037

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 167.343728

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 180.175109

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 261.700603

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 113.350254

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 198.364785
```

```
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 187.730532
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 106.408847
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 158.612782
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 145.464360
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 128.043321
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 328.815212
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 176.994446
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 75.572287
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 65.318552
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 129.370255
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 183.344132
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 133.054685
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 124.937958
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 139.153639
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 113.844217
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 71.003619
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 105.176197
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 131.319942
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 117.482693
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 124.642477
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 264.016300
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 132.735137
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 33.939329
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 88.743375
```

```
## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 78.575511

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 76.118864

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 103.719226

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 113.785148

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 181.601993

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 99.229643

## Warning in dpois(x = y/phi, lambda = mu/phi): non-integer x = 58.813759

## [1] Inf
```

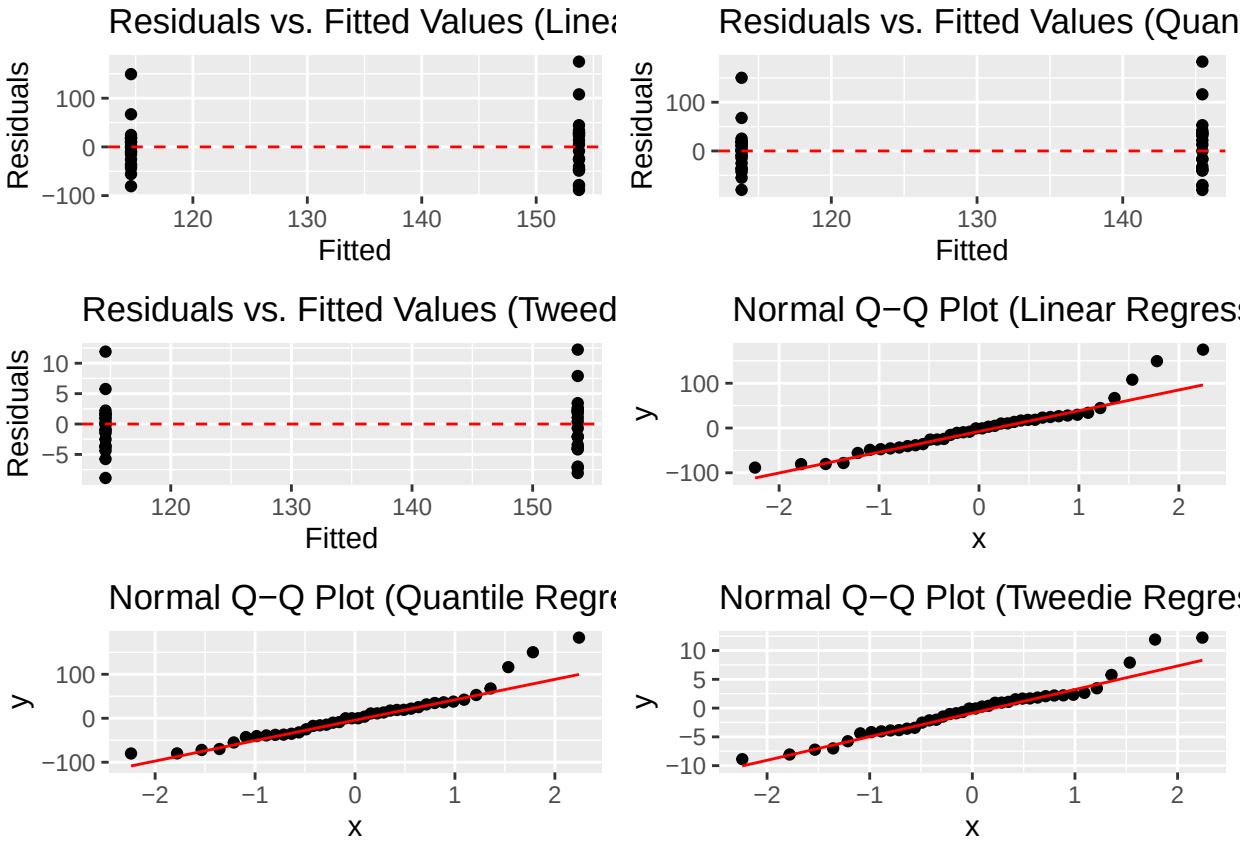
```
AIC(quantile_regression_model)[1]
```

```
## [1] 434.9818
```

residual analysis

```
lm_resid <- resid(lm_model)
quantile_resid <- resid(quantile_regression_model)
tweeie_resid <- resid(tweedie_model)

#install.packages("ggplot2")
#install.packages("gridExtra")

library(ggplot2)
library(gridExtra)

# Create residual vs. fitted value plots
residual_vs_fitted_lm <- ggplot(data = data.frame(Fitted = fitted(lm_model), Residuals = lm_resid), aes
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residuals vs. Fitted Values (Linear Regression)")

residual_vs_fitted_quantile <- ggplot(data = data.frame(Fitted = fitted(quantile_regression_model), Resi
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residuals vs. Fitted Values (Quantile Regression)")

residual_vs_fitted_tweedie <- ggplot(data = data.frame(Fitted = fitted(tweedie_model), Residuals = twee
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residuals vs. Fitted Values (Tweedie Regression)")

# Create normal Q-Q plots
```

```
qq_plot_lm <- ggplot(data = data.frame(Residuals = lm_resid), aes(sample = Residuals)) +
  geom_qq() +
  geom_qq_line(color = "red") +
  labs(title = "Normal Q-Q Plot (Linear Regression)")

qq_plot_quantile <- ggplot(data = data.frame(Residuals = quantile_resid), aes(sample = Residuals)) +
  geom_qq() +
  geom_qq_line(color = "red") +
  labs(title = "Normal Q-Q Plot (Quantile Regression)")

qq_plot_tweedie <- ggplot(data = data.frame(Residuals = tweeie_resid), aes(sample = Residuals)) +
  geom_qq() +
  geom_qq_line(color = "red") +
  labs(title = "Normal Q-Q Plot (Tweedie Regression)")

# Combine the plots
grid.arrange(residual_vs_fitted_lm, residual_vs_fitted_quantile, residual_vs_fitted_tweedie, qq_plot_lm
```




```
#install.packages("lmtest")
library(lmtest)
# Perform a Likelihood Ratio Test
lrt_result <- lrtest(lm_model, tweedie_model.2) #quantile_regression_model not applicable
print("Likelihood Ratio Test:")
```

```
## [1] "Likelihood Ratio Test:"
```

```
print(lrt_result)
```

```
## Likelihood ratio test
##
## Model 1: scores ~ groups
## Model 2: score_data$scores ~ score_data$groups
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1   3 -217.31
## 2   2         -1
```

```
# for nested models only
#deviance_lm <- deviance(lm_model)
#deviance_tweedie <- deviance(tweedie_model.2)

#print("Deviance for Linear Model:")
#print(deviance_lm)
#print("Deviance for Tweedie Model:")
#print(deviance_tweedie)
```

basic tests

```
t_test_result <- t.test(scores ~ groups, data = score_data)
anova_result <- aov(scores ~ groups, data = score_data)
wilcoxon_result <- wilcox.test(scores ~ groups, data = score_data)

"t_test"
```

```
## [1] "t_test"
```

```
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  scores by groups
## t = 2.1794, df = 35.196, p-value = 0.03608
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##   2.689597 75.601332
## sample estimates:
## mean in group 1 mean in group 2
##        153.7402        114.5947
```

```
"wilcoxon"
```

```
## [1] "wilcoxon"
```

```
wilcoxon_result
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  scores by groups
## W = 276, p-value = 0.04018
## alternative hypothesis: true location shift is not equal to 0
```

```
"anova"
```

```
## [1] "anova"
```

```
summary(anova_result)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## groups       1  15324   15324    4.75 0.0356 *
## Residuals   38 122589    3226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## anova with ln
delta = 1
```

```
score_data.ln <- score_data
score_data.ln$scores <- log(score_data.ln$scores+delta)
anova_result.ln <- aov(scores ~ groups, data = score_data.ln)
"anova log"
```

```
## [1] "anova log"
```

```
summary(anova_result.ln)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## groups       1  0.852  0.8516   4.886 0.0332 *
## Residuals   38  6.623  0.1743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P-values

```
lm_p_value <- summary(lm_model)$coefficients["groups2", "Pr(>|t|)"]
tweedy_p_value <- summary(tweedie_model)$coefficients["score_data$groups2", "Pr(>|t|)"]
p_value_groups2_quantile <- summary(quantile_regression_model, se="nid")$coefficients["score_data$groups

# Print model the p-values
output_text <- c(
print(paste(" Linear Regression,", lm_p_value, "\n")),
```

```
print(paste("Tweedie Regression,", tweedy_p_value, "\n")),
print(paste("Quantile Regression,", p_value_groups2_quantile, "\n")),
print(paste("wilcoxon,",wilcoxon_result$p.value, "\n"))
)
```

```
## [1] " Linear Regression, 0.0355651558178036 \n"
## [1] "Tweedie Regression, 0.0344183917215819 \n"
## [1] "Quantile Regression, 0.095815367975765 \n"
## [1] "wilcoxon, 0.0401750848962727 \n"
```

```
output_text
```

```
## [1] " Linear Regression, 0.0355651558178036 \n" "Tweedie Regression, 0.0344183917215819 \n"
## [3] "Quantile Regression, 0.095815367975765 \n" "wilcoxon, 0.0401750848962727 \n"
```

```
file_name <- "frequency_results.csv"

if (!file.exists(file_name)) {
  file.create(file_name)
}

cat(output_text, file = file_name,append=TRUE)
```

Different Runs

[1] “P-values:” [1] “Linear Regression: 0.00246703866214937” [1] “Tweedie Regression: 0.00289991935233989” [1] “Quantile Regression: 0.111913705926138” [1] “wilcoxon: 0.00291451132965859

[1] “P-values:” [1] “Linear Regression: 0.0857024280385969” [1] “Tweedie Regression: 0.0842261384751977” [1] “Quantile Regression: 0.7488317040991” [1] “wilcoxon: 0.253379898841039”

[1] “P-values:” [1] “Linear Regression: 0.0379739381365747” [1] “Tweedie Regression: 0.0343608203137708” [1] “Quantile Regression: 0.345106750248194” [1] “wilcoxon: 0.076266190262418”

“‘