Face Book

Erik Pak

June 12, 2024

Libraries

```
library(psych)
library(corrplot)
library(QuantPsyc)
library(car)
library(ggplot2)
library(ggpubr)
```

Import text file

Column reference:
- PTLike <- Page Total Likes
- Type
- Category (1,2,3 - product / action / inspiration)
- PosMon <- Post Month
- PosWkDay <- Post Weekday
- PosHr <- Post Hour
- Paid
- LPTReach <- Lifetime Post Total Reach
- LPTImpr <- Lifetime Post Total Impressions
- LEngUser <- Lifetime Engaged Users
- LPConsumer <- Lifetime Post Consumers
- LPConsump <- Lifetime Post Consumption
- LPIPepLkPage <- Lifetime Post Impressions by people who have liked your Page
- LPRchPepLKPage <- Lifetime Post reach by people who like your Page
- LPepLkEngPos <- Lifetime People who have liked your Page and engaged with your post
- comment
- like
- share
- TotalInterac <- Total Interactions

```
# set working directory
#setwd("/Users/sir/Desktop/DePaul/RScript/DSC423/Project")

# read the csv file
myd <- read.csv("dataset_Facebook.csv",sep = ";", header = T)

# new columns name - refer to columns reference for full description
colnames(myd) <- c('PTLike','Type','Category','PosMon','PosWkDay','PosHr','Paid',
                   'LPTReach','LPTImpr','LEngUser','LPConsumer','LPConsump',
                   'LPIPepLkPage','LPRchPepLKPage','LPepLkEngPos',
                   'comment','like','share','TotalInterac')

# display dataframe
str(myd)
```

```
## 'data.frame':     494 obs. of  19 variables:
##  $ PTLike         : int  139441 139441 139441 139441 139441 139441 139441 139441 13944
1 139441 ...
##  $ Type           : chr  "Photo" "Status" "Photo" "Photo" ...
##  $ Category       : int  2 2 3 2 2 2 3 3 2 3 ...
##  $ PosMon         : int  12 12 12 12 12 12 12 12 12 12 ...
##  $ PosWkDay       : int  4 3 3 2 2 1 1 7 7 6 ...
##  $ PosHr          : int  3 10 3 10 3 9 3 9 3 10 ...
##  $ Paid           : int  0 0 0 1 0 0 1 1 0 0 ...
##  $ LPTReach       : int  2752 10460 2413 50128 7244 10472 11692 13720 11844 4694 ...
##  $ LPTImpr        : int  5091 19057 4373 87991 13594 20849 19479 24137 22538 8668 ...
##  $ LEngUser       : int  178 1457 177 2211 671 1191 481 537 1530 280 ...
##  $ LPConsumer     : int  109 1361 113 790 410 1073 265 232 1407 183 ...
##  $ LPConsump      : int  159 1674 154 1119 580 1389 364 305 1692 250 ...
##  $ LPIPepLkPage   : int  3078 11710 2812 61027 6228 16034 15432 19728 15220 4309 ...
##  $ LPRchPepLKPage : int  1640 6112 1503 32048 3200 7852 9328 11056 7912 2324 ...
##  $ LPepLkEngPos   : int  119 1108 132 1386 396 1016 379 422 1250 199 ...
##  $ comment        : int  4 5 0 58 19 1 3 0 0 3 ...
##  $ like           : int  79 130 66 1572 325 152 249 325 161 113 ...
##  $ share          : int  17 29 14 147 49 33 27 14 31 26 ...
##  $ TotalInterac   : int  100 164 80 1777 393 186 279 339 192 142 ...
```

Data Cleaning / Wrangling

```r
# remove null/na values
myd <- na.omit(myd)

# removing features used for evaluating post impact
# and other not required variables
myd <- myd[,-c(8:10)]
myd <- myd[,-c(9:12)]
myd <- myd[,-c(9:11)]

# makes a copy of every variable in myd
attach(myd)

# create dummy variables
# Type (Photos,Status,Video,Link)
# category Factor: {action, product, inspiration }
myd$typeP=(Type=="Photo")*1
myd$typeS=(Type=="Status")*1
myd$typeV=(Type=="Video")*1
myd$category1=(Category==1)*1
myd$category2=(Category==2)*1

# remove a copy of every variable in myd
detach(myd)

# remove the column for which we created dummy variables
# also removing comment,like,share since we have total interaction
mydata <- myd[,-c(2:3)]
```

Explore Data

```r
# describe the distribution of Life Time post consumers
describe(mydata$LPConsumer)
```

```
##    vars   n   mean     sd median trimmed    mad min   max range skew kurtosis
## X1    1 490 812.15 886.13  559.5  652.43 390.67  23 11328 11305 5.01    44.18
##        se
## X1 40.03
```

```r
# describe the distribution of Life Time post consumers
describe(mydata$PTLike)
```

```
##    vars   n    mean       sd median  trimmed     mad   min    max range  skew
## X1    1 490 123173.8 16183.82 129600 125483.8 12122.48 81370 139441 58071 -0.97
##    kurtosis     se
## X1    -0.29 731.11
```

Histograsms & Scatter Plots

```r
# plot a histogram on the Y variable
plot_hist_lpcf <- ggplot(mydata, aes(x=LPConsumer)) +
                geom_histogram(bins = 30, color="black", fill="lightblue") +
                geom_vline(aes(xintercept=mean(LPConsumer)), col="darkblue") +
                labs(title = "Lifetime Post Consumers \n Histogram",
                    x = "Lifetime Post Consumers", y = 'Frequency') +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5)) +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10)) +
                theme(axis.text.x = element_text(angle = -45, hjust = .1))

# since the histogram looks exponential we try log of Y variable
plot_hist_lpcl <- ggplot(mydata, aes(x=log(LPConsumer))) +
                geom_histogram(bins = 30, color="black", fill="lightblue") +
                geom_vline(aes(xintercept=mean(log(LPConsumer))), col="darkblue") +
                labs(title = "Lifetime \n Post Consumers \n Log Histogram",
                    x = "Log Lifetime Post Consumers", y = 'Frequency') +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5)) +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10)) +
                theme(axis.text.x = element_text(angle = -45, hjust = .1))

# combine histogram plots
hist_com_plot <- ggarrange(plot_hist_lpcf, plot_hist_lpcl,
                labels = c("Fig A", "Fig B"),
                font.label = list(size = 9, color = "blue"))
# plot all
hist_com_plot
```
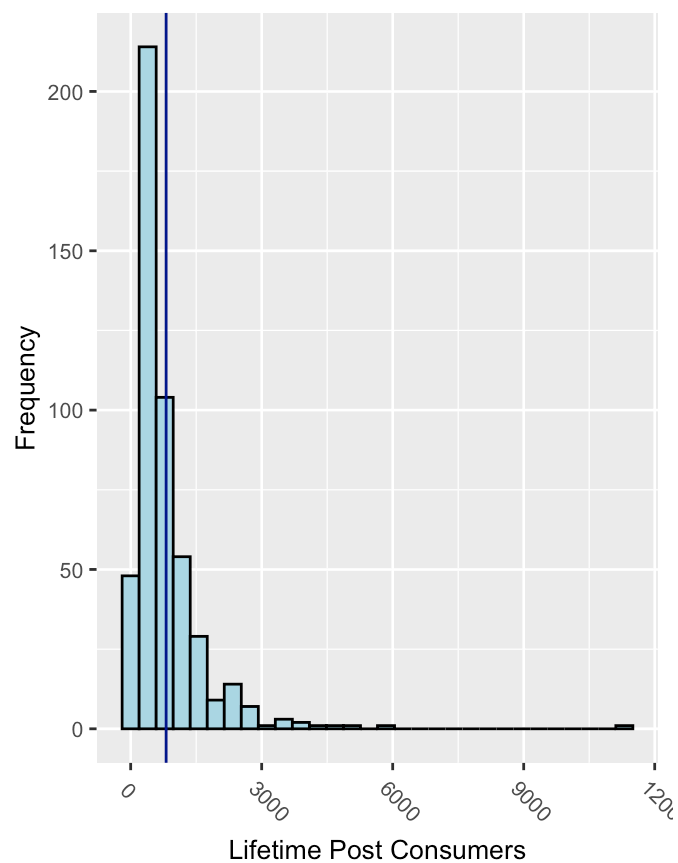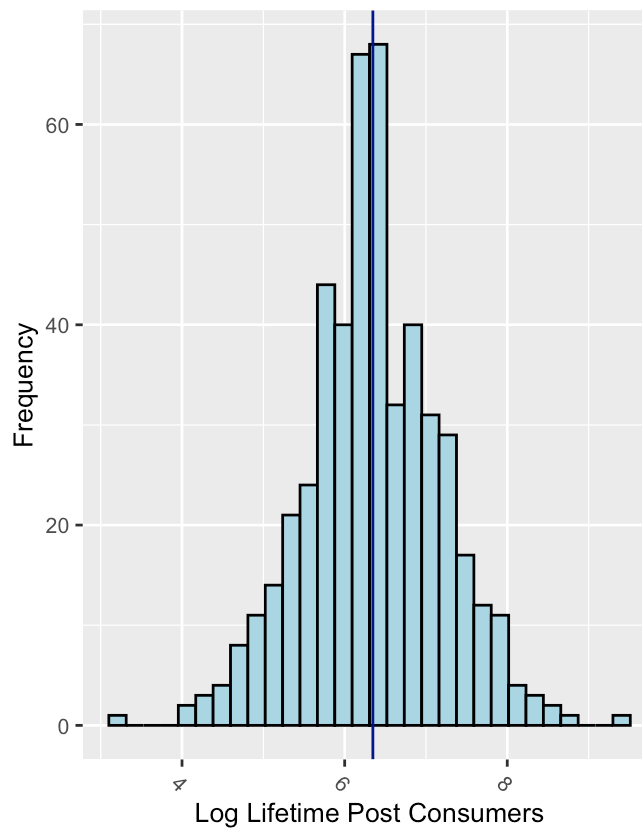
Fig A
Lifetime Post Consumers Histogram

Fig B
Lifetime Post Consumers Log Histogram

```r
# scatter plots for life time post consumers vs independent variables
plot_like_scatter <- ggplot(mydata, aes(x = (PTLike), y = log(LPConsumer))) +
             geom_point() +
             labs(title="Life Time Post Consumer \n vs \n Page Total Likes",
             x="Page Total Likes", y = "Lifetime Post Consumers") +
             # move the title text to the middle
             theme(plot.title=element_text(hjust=0.5)) +
             theme(text = element_text(size = 10))  +
             theme(axis.title = element_text(size = 10))

plot_scatter <- ggplot(mydata, aes(x = (TotalInterac), y = log(LPConsumer))) +
             geom_point() +
             labs(title="Life Time Post Consumer \n vs \n Total Interactions",
             x="Total Interactions", y = "Lifetime Post Consumers") +
             # move the title text to the middle
             theme(plot.title=element_text(hjust=0.5)) +
             theme(text = element_text(size = 10))  +
             theme(axis.title = element_text(size = 10))

plot_log_scatter <- ggplot(mydata, aes(x = log(TotalInterac), y = log(LPConsumer))) +
             geom_point() +
             labs(title="Life Time Post Consumer \n vs \n Log Total Interactions",
             x="Log Total Interactions", y = "Lifetime Post Consumers") +
             # move the title text to the middle
             theme(plot.title=element_text(hjust=0.5)) +
             theme(text = element_text(size = 10))  +
             theme(axis.title = element_text(size = 10))

# combine scatter plots
scatter_com_plot <- ggarrange(plot_like_scatter, plot_scatter, plot_log_scatter,
                 labels = c("Fig A", "Fig B", "Fig C"),
                 font.label = list(size = 9, color = "blue"))

# plot combined scatter
scatter_com_plot
```
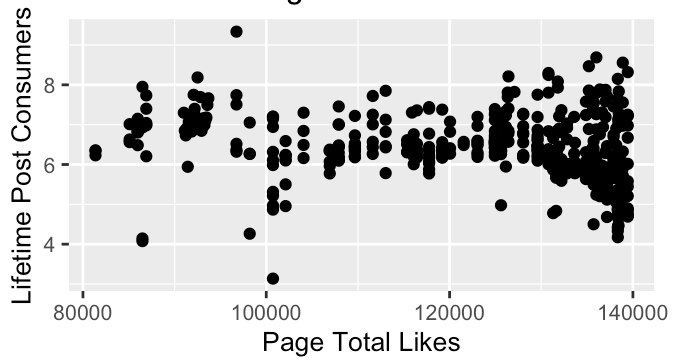
Fig A

Life Time Post Consumer
vs
Page Total Likes

Fig B

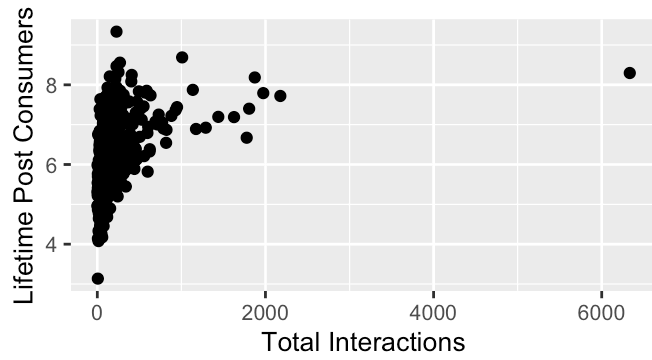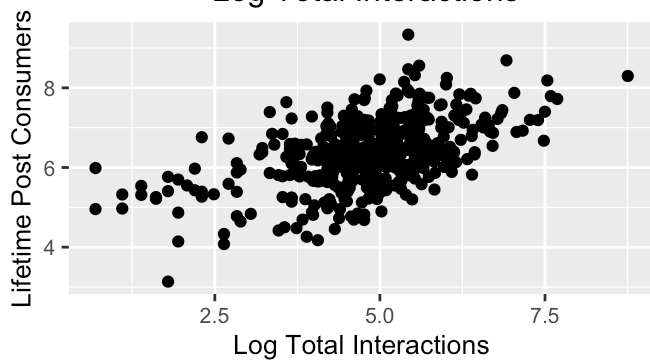Life Time Post Consumer
vs
Total Interactions

Fig C

Life Time Post Consumer
vs
Log Total Interactions

```r
# Box plot Paid
p_box_paid <- ggplot(mydata, aes(x=as.factor(Paid), y=log(LPConsumer), fill=Paid)) +
              geom_boxplot(alpha=0.8) + labs(x="Paid", y = "Life Time Post Comsumer") +
              theme(legend.position="none") + theme(axis.title = element_text(size = 1
0))

# we need to do something about this box plot - too busy
p_box_hour <- ggplot(mydata, aes(x=as.factor(PosHr), y=log(LPConsumer), fill=PosHr)) +
              geom_boxplot(alpha=0.8) + labs(x="Post Hour", y = "Life Time Post Comsume
r") +
              theme(legend.position="none") + theme(axis.title = element_text(size = 1
0))

# maybe we need to do something about this box plot - bit busy
p_box_month <- ggplot(mydata, aes(x=as.factor(PosMon), y=log(LPConsumer), fill=PosMon))
+
         geom_boxplot(alpha=0.8) + labs(x="Post Month", y = "Life Time Post Comsumer")
+
         theme(legend.position="none") + theme(axis.title = element_text(size = 10))

# week day post
p_box_weekd <- ggplot(mydata, aes(x=as.factor(PosWkDay), y=log(LPConsumer),
         fill=PosWkDay)) + geom_boxplot(alpha=0.8) + labs(x="Post Weekday",
         y = "Life Time Post Comsumer") + theme(legend.position="none") +
         theme(axis.title = element_text(size = 10))

# combine histogram plots
box_com_plot <- ggarrange(p_box_paid, p_box_hour, p_box_month, p_box_weekd,
              labels = c("Fig A", "Fig B", "Fig C", "Fig D"),
              font.label = list(size = 9, color = "blue"))



# plot all
box_com_plot
```
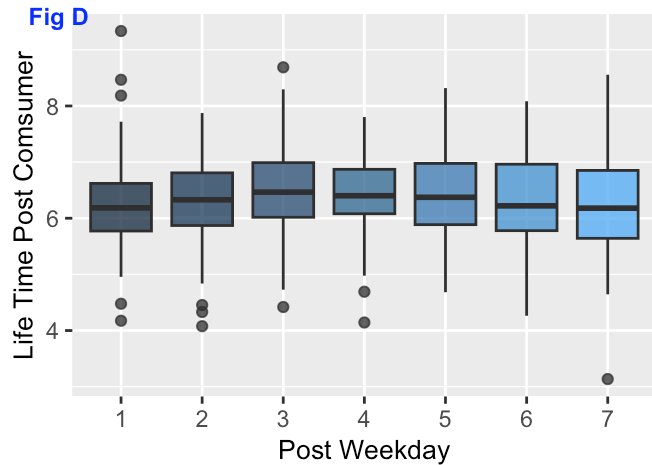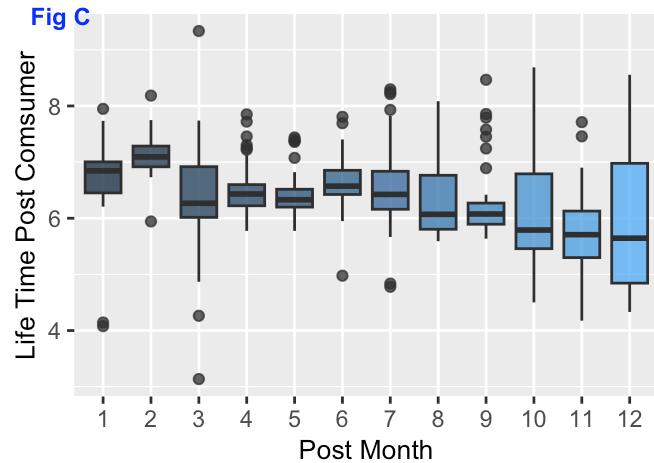
Fig A

Fig B

Fig C

Fig D
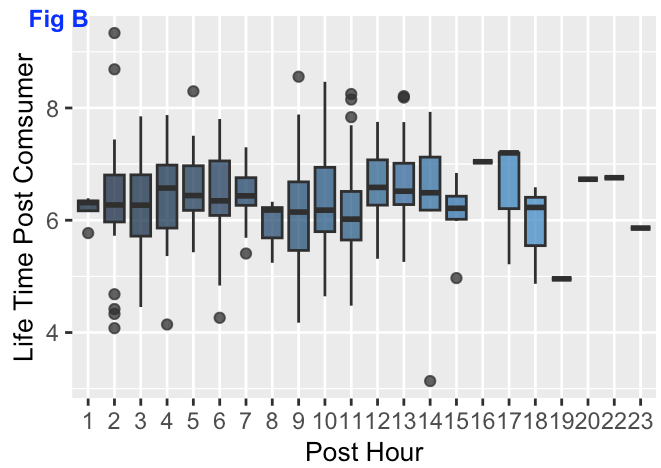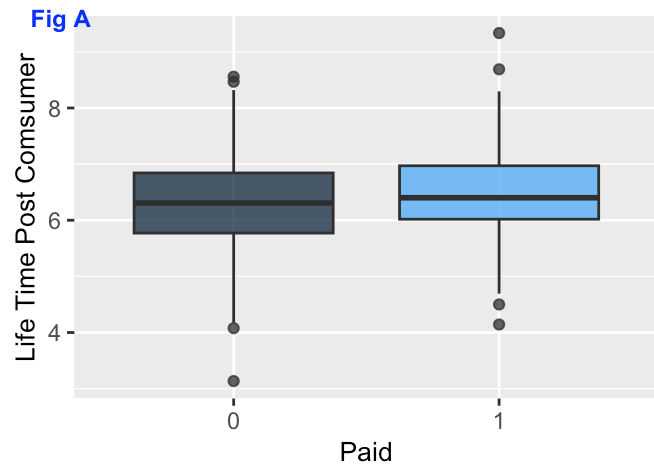
```
# correlation of the dataset
mydata_cor <- cor(mydata[, names(mydata) %in% c("PTLike", "LPConsumer", "TotalIntera
c")],
                 method = "pearson")

corrplot(mydata_cor,type="lower", method = 'number', addCoef.col = 'brown',
         number.cex = 0.8, tl.cex = 0.8)
```
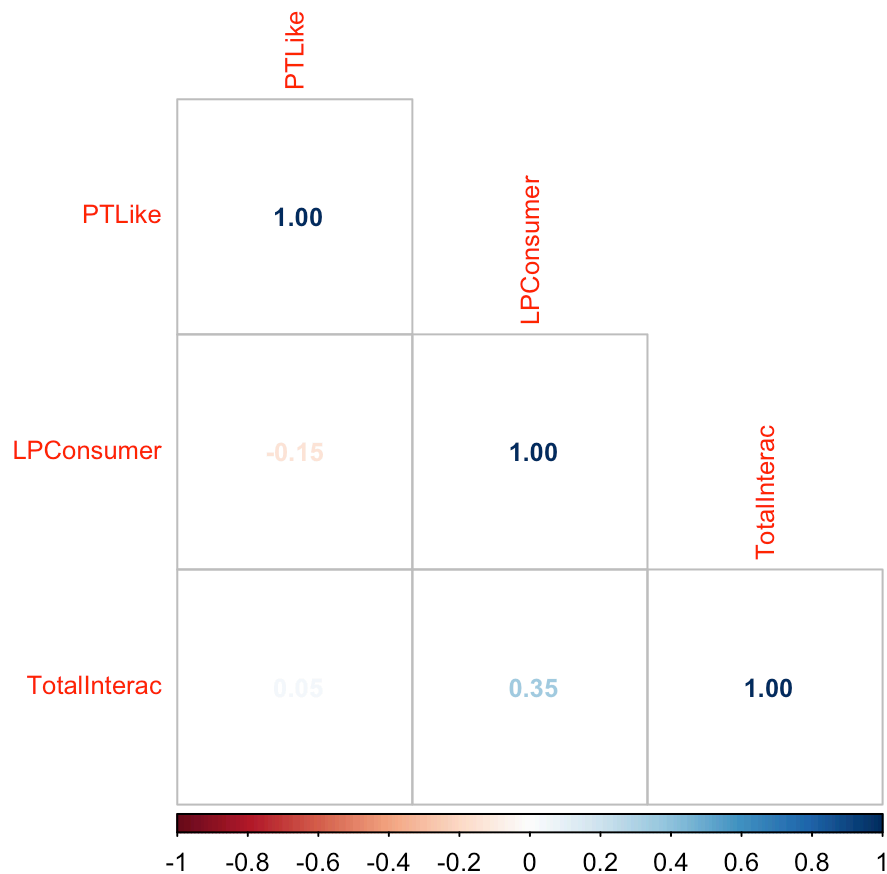
```
# display correlation values
mydata_cor
```

```
##                 PTLike LPConsumer TotalInterac
## PTLike       1.00000000 -0.1496334   0.04831435
## LPConsumer  -0.14963338  1.0000000   0.34941125
## TotalInterac 0.04831435  0.3494112   1.00000000
```

Model Building

```
# model including all relevant variables
fit_full_1 <- lm(log(LPConsumer) ~ PTLike + PosHr + Paid + log(TotalInterac) + typeP + t
ypeS +
                 typeV + category1 + category2, data=mydata)

# summary of full mode
summary(fit_full_1)
```

```
## 
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + PosHr + Paid + log(TotalInterac) +
##     typeP + typeS + typeV + category1 + category2, data = mydata)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.21410 -0.26494 -0.00665  0.26933  1.96768
## 
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.591e+00  2.407e-01  23.223  < 2e-16 ***
## PTLike            -2.104e-05  1.533e-06 -13.725  < 2e-16 ***
## PosHr              3.325e-03  5.573e-03   0.597    0.551
## Paid               9.047e-02  5.308e-02   1.704    0.089 .
## log(TotalInterac)  4.119e-01  2.321e-02  17.748  < 2e-16 ***
## typeP              1.035e+00  1.186e-01   8.729  < 2e-16 ***
## typeS              2.202e+00  1.483e-01  14.849  < 2e-16 ***
## typeV              1.579e+00  2.310e-01   6.836 2.47e-11 ***
## category1          4.481e-01  6.010e-02   7.457 4.18e-13 ***
## category2          9.520e-02  6.761e-02   1.408    0.160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.5185 on 480 degrees of freedom
## Multiple R-squared:  0.6186, Adjusted R-squared:  0.6114
## F-statistic: 86.48 on 9 and 480 DF,  p-value: < 2.2e-16
```

Variable Selection Stepwise: Backward

```
# stepwise variable selection on the full result also gave same result
step(fit_full_1, direction = "backward", trace = FALSE)
```

```
## 
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + log(TotalInterac) +
##     typeP + typeS + typeV + category1, data = mydata)
## 
## Coefficients:
##       (Intercept)             PTLike               Paid   log(TotalInterac)
##         5.626e+00         -2.079e-05          8.762e-02           4.094e-01
##             typeP              typeS              typeV          category1
##         1.043e+00          2.253e+00          1.588e+00           4.143e-01
```

```
# using stepwise variable selction
fit_full_2 <- lm(formula = log(LPConsumer) ~ PTLike + Paid + log(TotalInterac) +
                  typeP + typeS + typeV + category1, data = mydata)

# summary of full model
summary(fit_full_2)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + log(TotalInterac) +
##     typeP + typeS + typeV + category1, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3149 -0.2722 -0.0012  0.2691  1.9652
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.626e+00  2.326e-01  24.190  < 2e-16 ***
## PTLike          -2.079e-05  1.488e-06 -13.973  < 2e-16 ***
## Paid             8.762e-02  5.298e-02   1.654   0.0988 .
## log(TotalInterac) 4.094e-01  2.311e-02  17.717  < 2e-16 ***
## typeP            1.043e+00  1.170e-01   8.920  < 2e-16 ***
## typeS            2.253e+00  1.441e-01  15.632  < 2e-16 ***
## typeV            1.588e+00  2.297e-01   6.911 1.53e-11 ***
## category1        4.143e-01  5.310e-02   7.803 3.79e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5186 on 482 degrees of freedom
## Multiple R-squared:  0.6167, Adjusted R-squared:  0.6112
## F-statistic: 110.8 on 7 and 482 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_2)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                    Df  Sum Sq Mean Sq F value    Pr(>F)
## PTLike              1  25.023  25.023  93.028 < 2.2e-16 ***
## Paid                1   4.580   4.580  17.026 4.343e-05 ***
## log(TotalInterac)   1  96.231  96.231 357.756 < 2.2e-16 ***
## typeP               1  12.324  12.324  45.816 3.784e-11 ***
## typeS               1  39.378  39.378 146.394 < 2.2e-16 ***
## typeV               1  14.718  14.718  54.718 6.223e-13 ***
## category1           1  16.376  16.376  60.880 3.794e-14 ***
## Residuals         482 129.650   0.269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# removing variable due to t-value Paid
fit_full_3 <- lm(formula = log(LPConsumer) ~ PTLike + log(TotalInterac) + typeP +
                 typeS + typeV + category1, data = mydata)

# summary of full model
summary(fit_full_3)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + log(TotalInterac) + typeP +
##     typeS + typeV + category1, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32981 -0.26168  0.00091  0.27016  2.00912
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.620e+00  2.330e-01  24.124  < 2e-16 ***
## PTLike           -2.077e-05  1.490e-06 -13.935  < 2e-16 ***
## log(TotalInterac) 4.149e-01  2.291e-02  18.115  < 2e-16 ***
## typeP             1.042e+00  1.172e-01   8.894  < 2e-16 ***
## typeS             2.247e+00  1.443e-01  15.568  < 2e-16 ***
## typeV             1.605e+00  2.299e-01   6.979 9.85e-12 ***
## category1         4.202e-01  5.307e-02   7.918 1.67e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5196 on 483 degrees of freedom
## Multiple R-squared:  0.6146, Adjusted R-squared:  0.6098
## F-statistic: 128.4 on 6 and 483 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_3)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                    Df  Sum Sq Mean Sq F value    Pr(>F)
## PTLike              1  25.023  25.023  92.695 < 2.2e-16 ***
## log(TotalInterac)   1 100.100 100.100 370.808 < 2.2e-16 ***
## typeP               1  12.269  12.269  45.448 4.483e-11 ***
## typeS               1  38.396  38.396 142.232 < 2.2e-16 ***
## typeV               1  15.182  15.182  56.238 3.100e-13 ***
## category1           1  16.924  16.924  62.694 1.671e-14 ***
## Residuals         483 130.386   0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# check multicollinearity
vif(fit_full_3)
```

```
##             PTLike log(TotalInterac)              typeP              typeS
##           1.053647          1.150277           3.194659           3.154352
##             typeV          category1
##           1.351221          1.245845
```

```
# 95% confidence interval of the fitted model
confint(fit_full_3, level=0.95)
```

```
##                             2.5 %        97.5 %
## (Intercept)          5.162099e+00  6.077553e+00
## PTLike              -2.369447e-05 -1.783819e-05
## log(TotalInterac)    3.699380e-01  4.599527e-01
## typeP                8.118799e-01  1.272303e+00
## typeS                1.963538e+00  2.530790e+00
## typeV                1.152899e+00  2.056441e+00
## category1            3.159508e-01  5.245183e-01
```

Influential Points and Outliers

```
# plot of deleted studentized residuals vs hat values
student_hat <- data.frame(rstudent = rstudent(fit_full_3), hatvalues =  hatvalues(fit_fu
ll_3))

# plot rstudent vs hatvalues
student_hat_plot <- ggplot(student_hat, aes(x = hatvalues, y = rstudent)) + geom_point()
+

                # Change line type and color
                geom_hline(yintercept=3, linetype="dashed", color = "red") +
                geom_hline(yintercept=-3, linetype="dashed", color = "red")

# plot
student_hat_plot
```
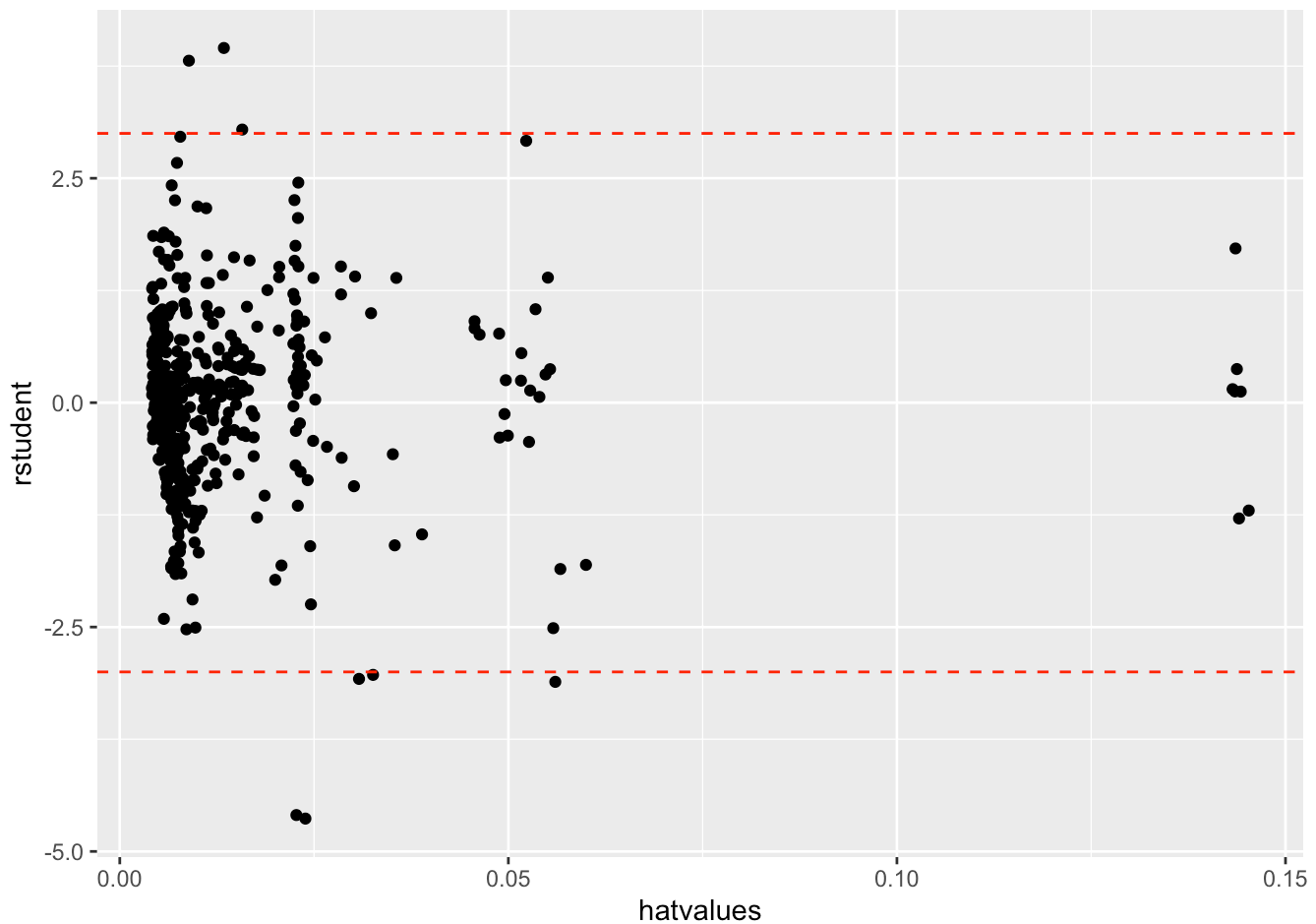
```
# outliers |standardized residuals| > 3
std_residual = data.frame(residual = rstandard(fit_full_3))

# display |standardized residuals| > 3
filter(std_residual, abs(residual) > 3)
```

```
##      residual
## 19  -4.538681
## 42   3.757572
## 83  -4.501761
## 128  3.017122
## 229 -3.050876
## 232 -3.007404
## 426 -3.082844
## 441  3.893078
```

```
# historgram for outliers
ggplot(std_residual, aes(x = residual)) +
   geom_histogram(bins=30, color="black", fill="lightblue") +
   labs(title = "Lifetime Post Consumers \n Standarized Residual", x = "Standarized Resid
ual",
       y = 'Frequency') +
     geom_vline(xintercept = 3, linetype="dotted",
             color = "red") +
     geom_vline(xintercept = -3, linetype="dotted",
             color = "red") +
     # move the title text to the middle
     theme(plot.title=element_text(hjust=0.5)) +
     theme(text = element_text(size = 10))  +
     theme(axis.title = element_text(size = 10))
```


Lifetime Post Consumers
Standarized Residual

```
# print out only observations that may be influential
summary(influence.measures(fit_full_3))
```

```
## Potentially influential observations of
##     lm(formula = log(LPConsumer) ~ PTLike + log(TotalInterac) + typeP +      typeS + ty
peV + category1, data = mydata) :
##
##         dfb.1_ dfb.PTLk dfb.l(TI dfb.typP dfb.typS dfb.typV dfb.ctg1 dffit
## 19   -0.02   -0.08     0.17     0.00    -0.37    -0.02     0.08   -0.72_*
## 22    0.04    0.04    -0.03    -0.11    -0.10    -0.06     0.00    0.13
## 29    0.00    0.00     0.00     0.00     0.00     0.05     0.00    0.06
## 38   -0.05    0.04     0.05    -0.01     0.18    -0.01     0.00    0.38_*
## 41    0.03    0.07     0.07    -0.23    -0.19    -0.13     0.03    0.25
## 42   -0.11    0.17    -0.14     0.08     0.04     0.00     0.16    0.36
## 43   -0.01   -0.01     0.00     0.03     0.02     0.01     0.00   -0.03
## 45    0.01    0.02     0.00    -0.05    -0.04    -0.03     0.00    0.06
## 47   -0.02   -0.02     0.01     0.08     0.06     0.04     0.00   -0.08
## 49    0.01    0.01    -0.01    -0.03    -0.02    -0.01     0.00    0.03
## 50    0.02   -0.12     0.12    -0.01     0.04    -0.01     0.14   -0.23
## 52    0.16   -0.11    -0.12    -0.04    -0.01     0.03    -0.17   -0.25
## 55   -0.02   -0.01     0.06     0.00     0.00    -0.43     0.02   -0.50_*
## 56    0.14   -0.10    -0.09    -0.03    -0.01     0.03    -0.14   -0.21
## 71    0.00    0.00     0.00     0.00     0.00     0.04     0.00    0.05
## 74   -0.01   -0.01     0.05     0.00     0.00    -0.46     0.01   -0.53_*
## 83    0.00   -0.06     0.08     0.00    -0.36    -0.01     0.05   -0.70_*
## 85    0.01    0.01     0.01    -0.05    -0.04    -0.03     0.01    0.06
## 128   0.00    0.13    -0.29     0.09     0.05     0.03     0.07    0.39_*
## 133   0.05    0.05     0.00    -0.16    -0.13    -0.09     0.00    0.17
## 137   0.10    0.16     0.18    -0.63    -0.53    -0.36     0.08    0.69_*
## 141  -0.08    0.10    -0.05     0.06     0.03    -0.01     0.12    0.23
## 146  -0.03   -0.02     0.01     0.08     0.07     0.04     0.00   -0.09
## 180  -0.01    0.00     0.01     0.00     0.00     0.13     0.00    0.15
## 229  -0.14    0.00     0.29    -0.02    -0.27    -0.05     0.11   -0.55_*
## 232  -0.15    0.00     0.31    -0.03    -0.27    -0.05     0.12   -0.56_*
## 240   0.00   -0.02     0.04     0.00     0.00     0.60     0.01    0.70_*
## 273   0.07    0.03    -0.17     0.02    -0.02     0.02    -0.14    0.22
## 274   0.00    0.00     0.00     0.00     0.00     0.04     0.00    0.05
## 276  -0.08    0.02     0.09     0.04     0.03    -0.02     0.15    0.19
## 280  -0.01    0.03    -0.08     0.06     0.04     0.01     0.10    0.20
## 286   0.03   -0.02     0.02    -0.05    -0.04     0.00    -0.12   -0.18
## 311   0.01    0.02    -0.14     0.08     0.06     0.02     0.11    0.26
## 341   0.09    0.00     0.01    -0.19    -0.15    -0.10     0.01    0.20
## 369   0.09   -0.01     0.00    -0.17    -0.14    -0.09     0.01    0.18
## 400   0.11   -0.04     0.14    -0.30    -0.23    -0.16     0.06    0.34
## 405   0.09   -0.02     0.00    -0.15    -0.12    -0.08     0.01    0.17
## 418  -0.06    0.03     0.09    -0.02    -0.02    -0.02     0.00   -0.11
## 421   0.05   -0.01     0.00    -0.07    -0.06    -0.03    -0.03    0.08
## 426  -0.55    0.14     0.29     0.60     0.47     0.27     0.06   -0.76_*
## 434  -0.32    0.09     0.06     0.41     0.34     0.17     0.18   -0.45_*
## 441   0.09   -0.30     0.19     0.10     0.10     0.00     0.28    0.46_*
## 465  -0.06    0.03    -0.02     0.09     0.07     0.05    -0.01   -0.10
## 472   0.01   -0.01     0.00    -0.01    -0.01    -0.01     0.00    0.02
## 476  -0.37    0.16     0.16     0.34     0.26     0.15     0.03   -0.46_*
## 480  -0.48    0.22     0.14     0.49     0.37     0.22     0.02   -0.61_*
## 487   0.06   -0.03     0.02    -0.08    -0.06    -0.04     0.01    0.09
```

```
##        cov.r   cook.d hat
## 19     0.77_*   0.07    0.02
## 22     1.07_*   0.00    0.05_*
## 29     1.18_*   0.00    0.14_*
## 38     0.95_*   0.02    0.02
## 41     1.06_*   0.01    0.05_*
## 42     0.83_*   0.02    0.01
## 43     1.07_*   0.00    0.05_*
## 45     1.07_*   0.00    0.05_*
## 47     1.07_*   0.00    0.05_*
## 49     1.07_*   0.00    0.05_*
## 50     0.93_*   0.01    0.01
## 52     0.94_*   0.01    0.01
## 55     1.16_*   0.04    0.15_*
## 56     0.96_*   0.01    0.01
## 71     1.19_*   0.00    0.14_*
## 74     1.16_*   0.04    0.14_*
## 83     0.77_*   0.07    0.02
## 85     1.07_*   0.00    0.05_*
## 128    0.90_*   0.02    0.02
## 133    1.06_*   0.00    0.05_*
## 137    0.95_*   0.07    0.05_*
## 141    0.92_*   0.01    0.01
## 146    1.06_*   0.00    0.05_*
## 180    1.18_*   0.00    0.14_*
## 229    0.91_*   0.04    0.03
## 232    0.92_*   0.04    0.03
## 240    1.14_*   0.07    0.14_*
## 273    0.96_*   0.01    0.01
## 274    1.18_*   0.00    0.14_*
## 276    0.95_*   0.01    0.01
## 280    0.94_*   0.01    0.01
## 286    0.94_*   0.00    0.01
## 311    0.90_*   0.01    0.01
## 341    1.05_*   0.01    0.05_*
## 369    1.05_*   0.00    0.05_*
## 400    1.04_*   0.02    0.06_*
## 405    1.06_*   0.00    0.05_*
## 418    1.05_*   0.00    0.04
## 421    1.07_*   0.00    0.05_*
## 426    0.94_*   0.08    0.06_*
## 434    1.02     0.03    0.06_*
## 441    0.82_*   0.03    0.01
## 465    1.07_*   0.00    0.05_*
## 472    1.07_*   0.00    0.05_*
## 476    1.03     0.03    0.06_*
## 480    0.98     0.05    0.06_*
## 487    1.07_*   0.00    0.06_*
```

```
# influential Plot
influencePlot(fit_full_3, scale=5, xlab="Hat-Values", ylab="Studentized Residuals",
              fill.col=carPalette()[2], fill.alpha=0.5, id=TRUE)
```
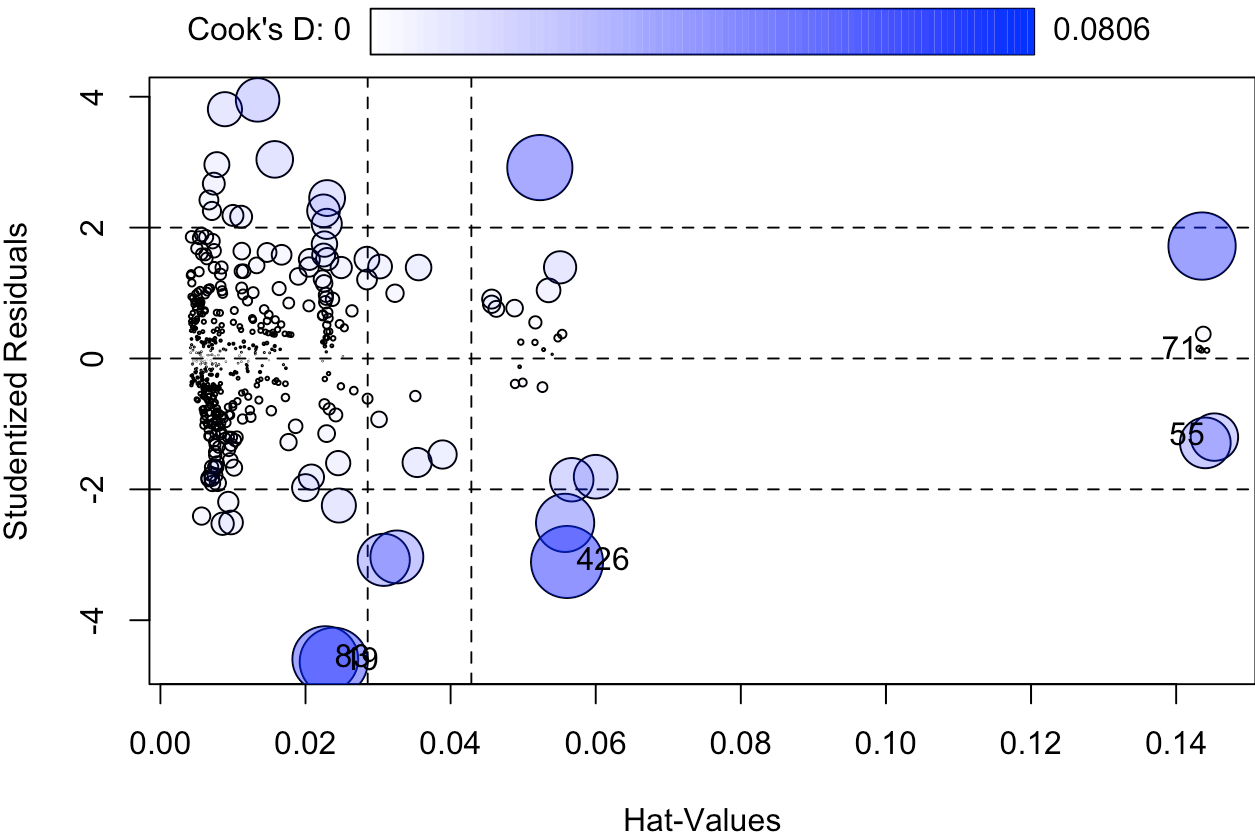


```
##        StudRes        Hat        CookD
## 19   -4.633873  0.02388823  0.0720187742
## 55   -1.201781  0.14527074  0.0350350462
## 71    0.123237  0.14424148  0.0003664455
## 83   -4.594521  0.02272519  0.0673221092
## 426  -3.110404  0.05604218  0.0806060069
```

First Final Model

```
# standardized beta coefficient
lm.beta(fit_full_3)
```

```
##          PTLike log(TotalInterac)          typeP          typeS
##       -0.4040712        0.5488464      0.4490896      0.7810622
##           typeV        category1
##        0.2291780        0.2496601
```

```r
# residual vs fitted Model
residual_plot <-  ggplot(fit_full_3, aes(x = .fitted, y = .resid)) +
                  geom_point() +
                  geom_hline(yintercept = 0, col = "red") +
                  labs(title="Fitted",
                    x = "Fitted", y = "Residual") +
                  # move the title text to the middle
                  theme(plot.title=element_text(hjust=0.5))  +
                  theme(text = element_text(size = 10))  +
                  theme(axis.title = element_text(size = 10))


# Page Total Likes vs residuals
PTLike_plot <-  ggplot(mydata, aes(x = PTLike,
                y = rstandard(fit_full_3))) + geom_point() +
                geom_hline(yintercept = 0, col = "red") +
                labs(title="Page Total Likes",
                x = "Page Total Likes", y = "Residual") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10))


#create Q-Q plot
qq_plot <-  ggplot(fit_full_3, aes(sample=rstandard(fit_full_3))) +
                stat_qq(size=1.5, color='blue') +
                stat_qq_line(col = "red") +
                labs(title="Q-Q Plot",
                  x = "Theoretical Quantiles", y = "Sample Quantiles") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10))


# combine all plots
fit_final_plot <- ggarrange(residual_plot, PTLike_plot, qq_plot,
                  labels = c("Fig A", "Fig B", "Fig C"),
                  font.label = list(size = 9, color = "blue"))
# plot all
fit_final_plot
```

Fig A

Fitted
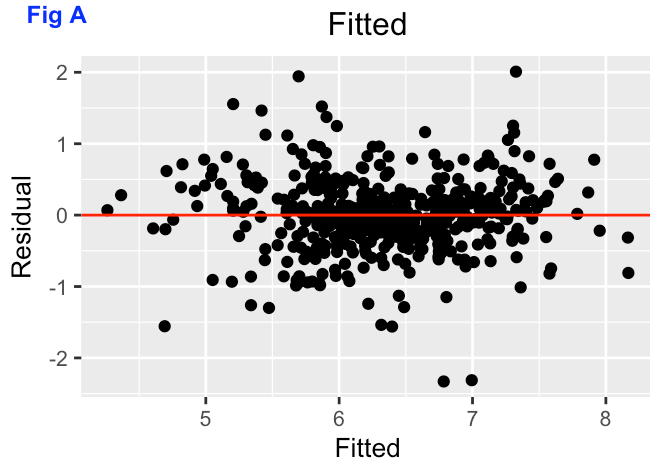
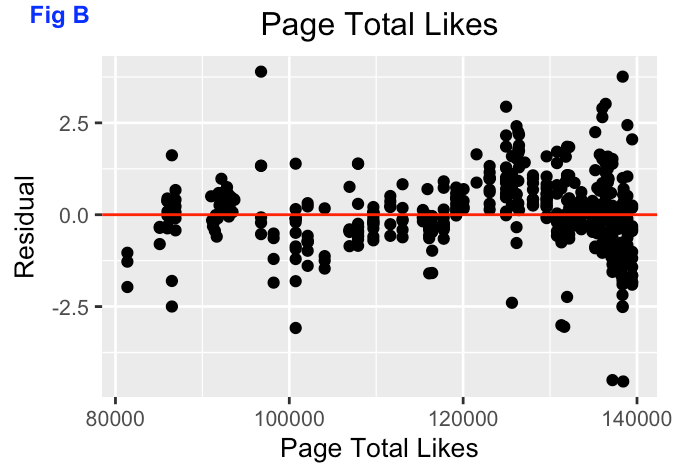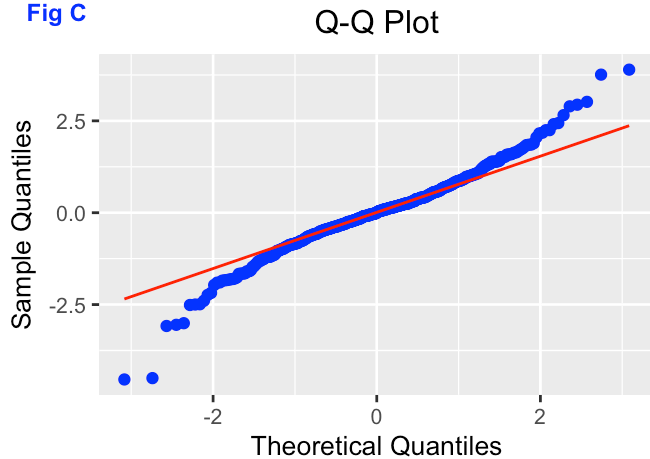

Fig B

Page Total Likes

Fig C

Q-Q Plot

Interaction Model Building

```
# model interaction of Paid with other independent variables
fit_inter_Paid_full_1 <- lm(log(LPConsumer) ~ Paid*(PTLike  + typeP + typeS + log(TotalI
nterac) +
                typeV + category1) , data=mydata)

# summary Report
summary(fit_inter_Paid_full_1)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ Paid * (PTLike + typeP + typeS +
##     log(TotalInterac) + typeV + category1), data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.33456 -0.26515 -0.00041  0.26074  1.97198
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            5.553e+00  2.714e-01  20.458  < 2e-16 ***
## Paid                   3.288e-01  5.412e-01   0.608    0.544
## PTLike                -2.055e-05  1.729e-06 -11.887  < 2e-16 ***
## typeP                  1.005e+00  1.376e-01   7.305 1.18e-12 ***
## typeS                  2.248e+00  1.686e-01  13.339  < 2e-16 ***
## log(TotalInterac)      4.243e-01  2.703e-02  15.699  < 2e-16 ***
## typeV                  1.453e+00  3.322e-01   4.373 1.51e-05 ***
## category1              4.324e-01  6.394e-02   6.764 3.96e-11 ***
## Paid:PTLike           -7.248e-07  3.472e-06  -0.209    0.835
## Paid:typeP             1.640e-01  2.676e-01   0.613    0.540
## Paid:typeS             2.629e-02  3.342e-01   0.079    0.937
## Paid:log(TotalInterac) -5.682e-02  5.369e-02  -1.058    0.290
## Paid:typeV             3.465e-01  4.795e-01   0.723    0.470
## Paid:category1        -4.493e-02  1.165e-01  -0.386    0.700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5207 on 476 degrees of freedom
## Multiple R-squared:  0.6184, Adjusted R-squared:  0.608
## F-statistic: 59.34 on 13 and 476 DF,  p-value: < 2.2e-16
```

Interaction variable selection

```
# stepwise variable selection on the full interaction model
step(fit_inter_Paid_full_1, direction = "backward", trace = FALSE)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ Paid + PTLike + typeP + typeS +
##     log(TotalInterac) + typeV + category1, data = mydata)
##
## Coefficients:
##       (Intercept)               Paid              PTLike               typeP
##         5.626e+00           8.762e-02          -2.079e-05           1.043e+00
##             typeS   log(TotalInterac)               typeV           category1
##         2.253e+00           4.094e-01           1.588e+00           4.143e-01
```

```
# selecting the model from stepwise backward variable selection process
fit_inter_Paid_2 <-lm(formula = log(LPConsumer) ~ Paid + PTLike + typeP + typeS +
    log(TotalInterac) + typeV + category1, data = mydata)

# after stepwise fit model
summary(fit_inter_Paid_2)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ Paid + PTLike + typeP + typeS +
##       log(TotalInterac) + typeV + category1, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3149 -0.2722 -0.0012  0.2691  1.9652
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.626e+00  2.326e-01  24.190  < 2e-16 ***
## Paid              8.762e-02  5.298e-02   1.654   0.0988 .
## PTLike           -2.079e-05  1.488e-06 -13.973  < 2e-16 ***
## typeP             1.043e+00  1.170e-01   8.920  < 2e-16 ***
## typeS             2.253e+00  1.441e-01  15.632  < 2e-16 ***
## log(TotalInterac) 4.094e-01  2.311e-02  17.717  < 2e-16 ***
## typeV             1.588e+00  2.297e-01   6.911 1.53e-11 ***
## category1         4.143e-01  5.310e-02   7.803 3.79e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5186 on 482 degrees of freedom
## Multiple R-squared:  0.6167, Adjusted R-squared:  0.6112
## F-statistic: 110.8 on 7 and 482 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_inter_Paid_2)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                   Df  Sum Sq Mean Sq F value    Pr(>F)
## Paid               1   4.313   4.313  16.033 7.208e-05 ***
## PTLike             1  25.290  25.290  94.021 < 2.2e-16 ***
## typeP              1  11.680  11.680  43.422 1.159e-10 ***
## typeS              1  56.578  56.578 210.340 < 2.2e-16 ***
## log(TotalInterac)  1  79.674  79.674 296.204 < 2.2e-16 ***
## typeV              1  14.718  14.718  54.718 6.223e-13 ***
## category1          1  16.376  16.376  60.880 3.794e-14 ***
## Residuals        482 129.650   0.269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# remove Paid variable selection process
fit_inter_Paid_3 <-lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS +
    log(TotalInterac) + typeV + category1, data = mydata)

# after stepwise fit model
summary(fit_inter_Paid_3)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInterac) +
##       typeV + category1, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32981 -0.26168  0.00091  0.27016  2.00912
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.620e+00  2.330e-01  24.124  < 2e-16 ***
## PTLike            -2.077e-05  1.490e-06 -13.935  < 2e-16 ***
## typeP              1.042e+00  1.172e-01   8.894  < 2e-16 ***
## typeS              2.247e+00  1.443e-01  15.568  < 2e-16 ***
## log(TotalInterac)  4.149e-01  2.291e-02  18.115  < 2e-16 ***
## typeV              1.605e+00  2.299e-01   6.979 9.85e-12 ***
## category1          4.202e-01  5.307e-02   7.918 1.67e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5196 on 483 degrees of freedom
## Multiple R-squared:  0.6146, Adjusted R-squared:  0.6098
## F-statistic: 128.4 on 6 and 483 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_inter_Paid_3)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                    Df  Sum Sq Mean Sq F value    Pr(>F)
## PTLike              1  25.023  25.023  92.695 < 2.2e-16 ***
## typeP               1  11.526  11.526  42.698 1.624e-10 ***
## typeS               1  54.651  54.651 202.448 < 2.2e-16 ***
## log(TotalInterac)   1  84.587  84.587 313.343 < 2.2e-16 ***
## typeV               1  15.182  15.182  56.238 3.100e-13 ***
## category1           1  16.924  16.924  62.694 1.671e-14 ***
## Residuals         483 130.386   0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 95% confidence interval of the fitted model
confint(fit_inter_Paid_3, level=0.95)
```

```
##                          2.5 %        97.5 %
## (Intercept)        5.162099e+00  6.077553e+00
## PTLike            -2.369447e-05 -1.783819e-05
## typeP              8.118799e-01  1.272303e+00
## typeS              1.963538e+00  2.530790e+00
## log(TotalInterac)  3.699380e-01  4.599527e-01
## typeV              1.152899e+00  2.056441e+00
## category1          3.159508e-01  5.245183e-01
```

```
# influential Plot
influencePlot(fit_inter_Paid_3)
```
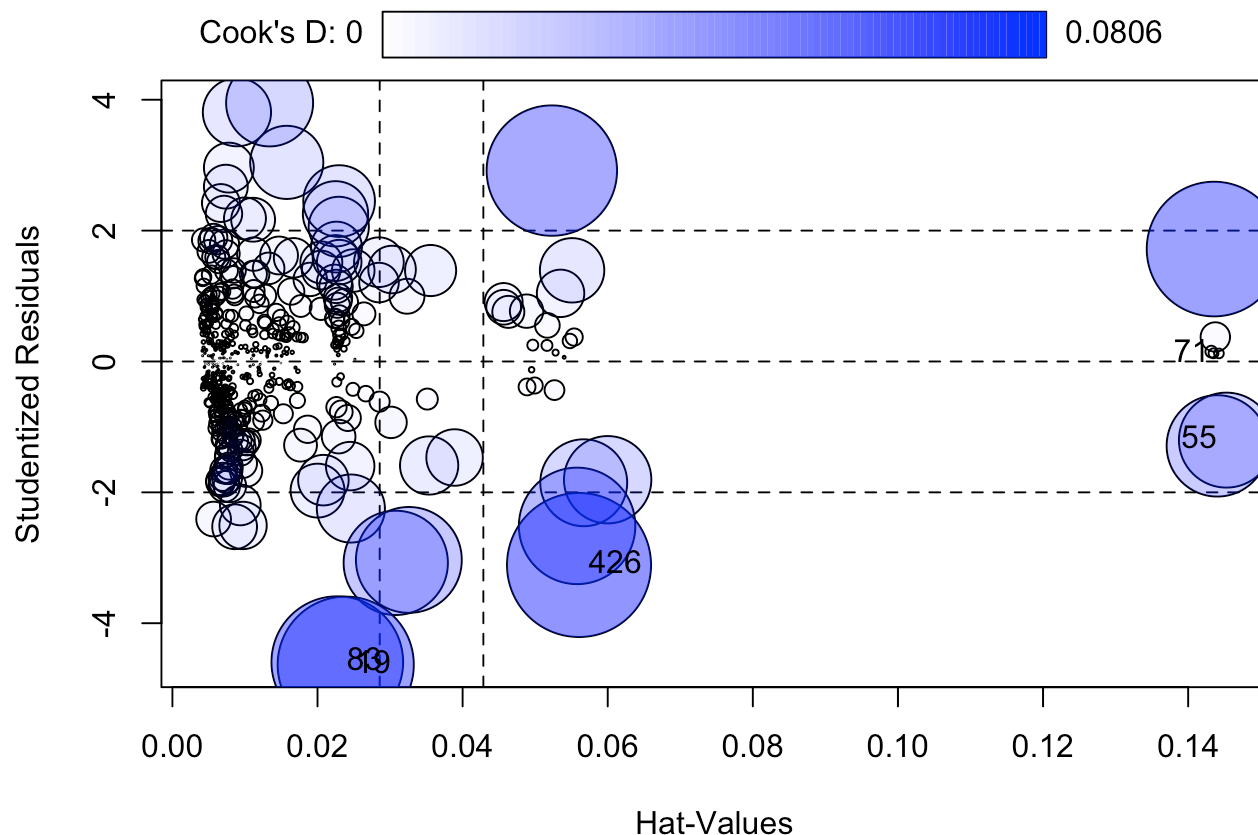
```
##        StudRes        Hat         CookD
## 19  -4.633873  0.02388823  0.0720187742
## 55  -1.201781  0.14527074  0.0350350462
## 71   0.123237  0.14424148  0.0003664455
## 83  -4.594521  0.02272519  0.0673221092
## 426 -3.110404  0.05604218  0.0806060069
```

Influential Points and Outliers

```
# plot of deleted studentized residuals vs hat values
student_hat <- data.frame(rstudent = rstudent(fit_inter_Paid_3), hatvalues =
                          hatvalues(fit_inter_Paid_3))

# plot rstudent vs hatvalues
student_hat_plot <- ggplot(student_hat, aes(x = hatvalues, y = rstudent)) + geom_point()
+
                # Change line type and color
                geom_hline(yintercept=3, linetype="dashed", color = "red") +
                geom_hline(yintercept=-3, linetype="dashed", color = "red")

# plot
student_hat_plot
```
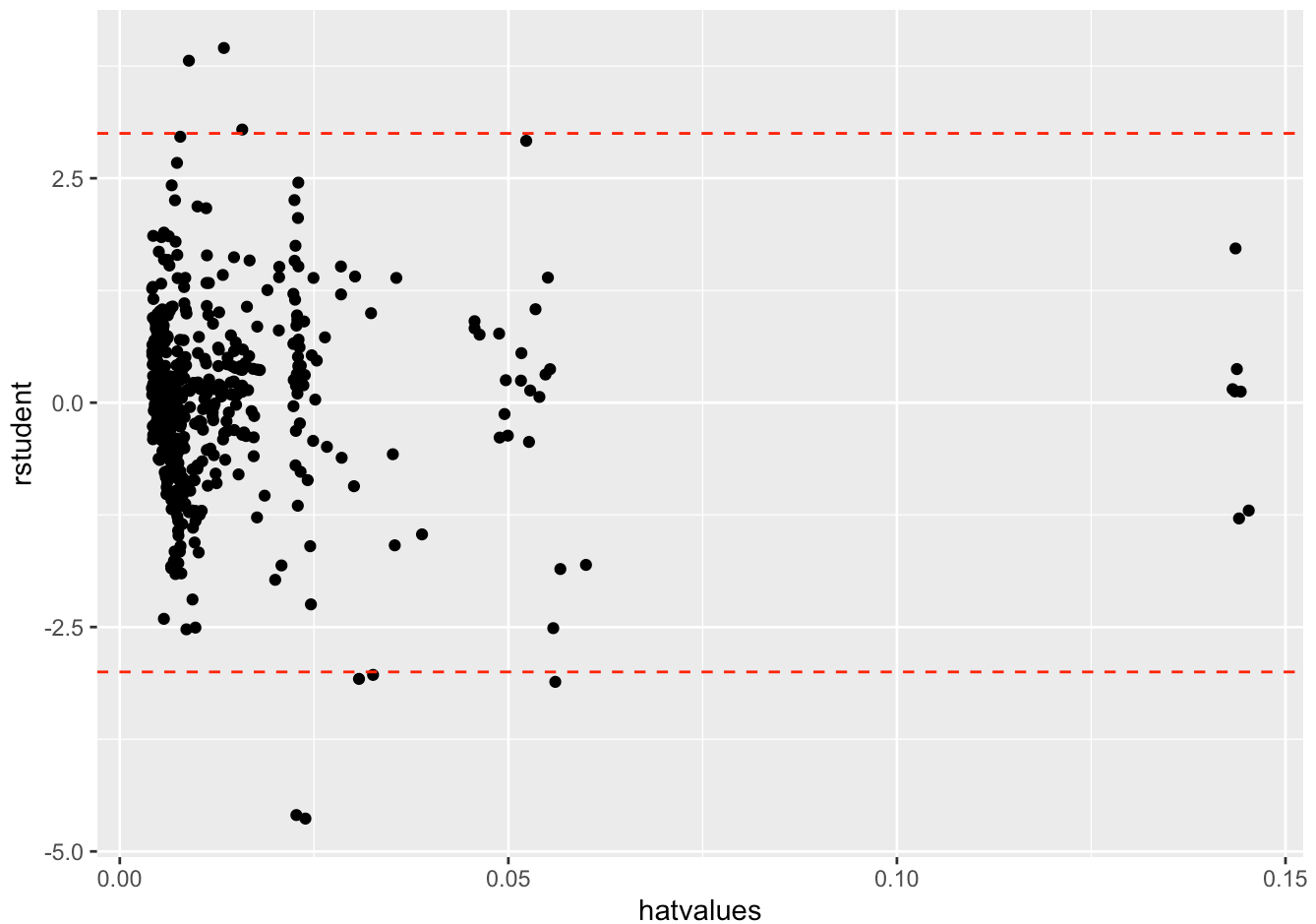
```r
# outliers |standardized residuals| > 3
std_residual = data.frame(residual = rstandard(fit_inter_Paid_3))

# display |standardized residuals| > 3
filter(std_residual, abs(residual) > 3)
```

```
##      residual
## 19  -4.538681
## 42   3.757572
## 83  -4.501761
## 128  3.017122
## 229 -3.050876
## 232 -3.007404
## 426 -3.082844
## 441  3.893078
```

```
# historgram for outliers
ggplot(std_residual, aes(x = residual)) +
  geom_histogram(bins=30, color="black", fill="lightblue") +
  labs(title = "Lifetime Post Consumers \n Standarized Residual", x = "Standarized Resid
ual",
      y = 'Frequency') +
    geom_vline(xintercept = 3, linetype="dotted",
              color = "red") +
    geom_vline(xintercept = -3, linetype="dotted",
              color = "red") +
    # move the title text to the middle
    theme(plot.title=element_text(hjust=0.5)) +
    theme(text = element_text(size = 10))  +
    theme(axis.title = element_text(size = 10))
```


Lifetime Post Consumers
Standarized Residual

```
# print out only observations that may be influential
summary(influence.measures(fit_inter_Paid_3))
```

```
## Potentially influential observations of
##   lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInterac) +      ty
peV + category1, data = mydata) :
##
##        dfb.1_ dfb.PTLk dfb.typP dfb.typS dfb.l(TI dfb.typV dfb.ctg1 dffit
## 19   -0.02  -0.08     0.00    -0.37     0.17    -0.02     0.08   -0.72_*
## 22    0.04   0.04    -0.11    -0.10    -0.03    -0.06     0.00    0.13
## 29    0.00   0.00     0.00     0.00     0.00     0.05     0.00    0.06
## 38   -0.05   0.04    -0.01     0.18     0.05    -0.01     0.00    0.38_*
## 41    0.03   0.07    -0.23    -0.19     0.07    -0.13     0.03    0.25
## 42   -0.11   0.17     0.08     0.04    -0.14     0.00     0.16    0.36
## 43   -0.01  -0.01     0.03     0.02     0.00     0.01     0.00   -0.03
## 45    0.01   0.02    -0.05    -0.04     0.00    -0.03     0.00    0.06
## 47   -0.02  -0.02     0.08     0.06     0.01     0.04     0.00   -0.08
## 49    0.01   0.01    -0.03    -0.02    -0.01    -0.01     0.00    0.03
## 50    0.02  -0.12    -0.01     0.04     0.12    -0.01     0.14   -0.23
## 52    0.16  -0.11    -0.04    -0.01    -0.12     0.03    -0.17   -0.25
## 55   -0.02  -0.01     0.00     0.00     0.06    -0.43     0.02   -0.50_*
## 56    0.14  -0.10    -0.03    -0.01    -0.09     0.03    -0.14   -0.21
## 71    0.00   0.00     0.00     0.00     0.00     0.04     0.00    0.05
## 74   -0.01  -0.01     0.00     0.00     0.05    -0.46     0.01   -0.53_*
## 83    0.00  -0.06     0.00    -0.36     0.08    -0.01     0.05   -0.70_*
## 85    0.01   0.01    -0.05    -0.04     0.01    -0.03     0.01    0.06
## 128   0.00   0.13     0.09     0.05    -0.29     0.03     0.07    0.39_*
## 133   0.05   0.05    -0.16    -0.13     0.00    -0.09     0.00    0.17
## 137   0.10   0.16    -0.63    -0.53     0.18    -0.36     0.08    0.69_*
## 141  -0.08   0.10     0.06     0.03    -0.05    -0.01     0.12    0.23
## 146  -0.03  -0.02     0.08     0.07     0.01     0.04     0.00   -0.09
## 180  -0.01   0.00     0.00     0.00     0.01     0.13     0.00    0.15
## 229  -0.14   0.00    -0.02    -0.27     0.29    -0.05     0.11   -0.55_*
## 232  -0.15   0.00    -0.03    -0.27     0.31    -0.05     0.12   -0.56_*
## 240   0.00  -0.02     0.00     0.00     0.04     0.60     0.01    0.70_*
## 273   0.07   0.03     0.02    -0.02    -0.17     0.02    -0.14    0.22
## 274   0.00   0.00     0.00     0.00     0.00     0.04     0.00    0.05
## 276  -0.08   0.02     0.04     0.03     0.09    -0.02     0.15    0.19
## 280  -0.01   0.03     0.06     0.04    -0.08     0.01     0.10    0.20
## 286   0.03  -0.02    -0.05    -0.04     0.02     0.00    -0.12   -0.18
## 311   0.01   0.02     0.08     0.06    -0.14     0.02     0.11    0.26
## 341   0.09   0.00    -0.19    -0.15     0.01    -0.10     0.01    0.20
## 369   0.09  -0.01    -0.17    -0.14     0.00    -0.09     0.01    0.18
## 400   0.11  -0.04    -0.30    -0.23     0.14    -0.16     0.06    0.34
## 405   0.09  -0.02    -0.15    -0.12     0.00    -0.08     0.01    0.17
## 418  -0.06   0.03    -0.02    -0.02     0.09    -0.02     0.00   -0.11
## 421   0.05  -0.01    -0.07    -0.06     0.00    -0.03    -0.03    0.08
## 426  -0.55   0.14     0.60     0.47     0.29     0.27     0.06   -0.76_*
## 434  -0.32   0.09     0.41     0.34     0.06     0.17     0.18   -0.45_*
## 441   0.09  -0.30     0.10     0.10     0.19     0.00     0.28    0.46_*
## 465  -0.06   0.03     0.09     0.07    -0.02     0.05    -0.01   -0.10
## 472   0.01  -0.01    -0.01    -0.01     0.00    -0.01     0.00    0.02
## 476  -0.37   0.16     0.34     0.26     0.16     0.15     0.03   -0.46_*
## 480  -0.48   0.22     0.49     0.37     0.14     0.22     0.02   -0.61_*
## 487   0.06  -0.03    -0.08    -0.06     0.02    -0.04     0.01    0.09
```

```
##       cov.r  cook.d hat
## 19    0.77_*  0.07   0.02
## 22    1.07_*  0.00   0.05_*
## 29    1.18_*  0.00   0.14_*
## 38    0.95_*  0.02   0.02
## 41    1.06_*  0.01   0.05_*
## 42    0.83_*  0.02   0.01
## 43    1.07_*  0.00   0.05_*
## 45    1.07_*  0.00   0.05_*
## 47    1.07_*  0.00   0.05_*
## 49    1.07_*  0.00   0.05_*
## 50    0.93_*  0.01   0.01
## 52    0.94_*  0.01   0.01
## 55    1.16_*  0.04   0.15_*
## 56    0.96_*  0.01   0.01
## 71    1.19_*  0.00   0.14_*
## 74    1.16_*  0.04   0.14_*
## 83    0.77_*  0.07   0.02
## 85    1.07_*  0.00   0.05_*
## 128   0.90_*  0.02   0.02
## 133   1.06_*  0.00   0.05_*
## 137   0.95_*  0.07   0.05_*
## 141   0.92_*  0.01   0.01
## 146   1.06_*  0.00   0.05_*
## 180   1.18_*  0.00   0.14_*
## 229   0.91_*  0.04   0.03
## 232   0.92_*  0.04   0.03
## 240   1.14_*  0.07   0.14_*
## 273   0.96_*  0.01   0.01
## 274   1.18_*  0.00   0.14_*
## 276   0.95_*  0.01   0.01
## 280   0.94_*  0.01   0.01
## 286   0.94_*  0.00   0.01
## 311   0.90_*  0.01   0.01
## 341   1.05_*  0.01   0.05_*
## 369   1.05_*  0.00   0.05_*
## 400   1.04_*  0.02   0.06_*
## 405   1.06_*  0.00   0.05_*
## 418   1.05_*  0.00   0.04
## 421   1.07_*  0.00   0.05_*
## 426   0.94_*  0.08   0.06_*
## 434   1.02    0.03   0.06_*
## 441   0.82_*  0.03   0.01
## 465   1.07_*  0.00   0.05_*
## 472   1.07_*  0.00   0.05_*
## 476   1.03    0.03   0.06_*
## 480   0.98    0.05   0.06_*
## 487   1.07_*  0.00   0.06_*
```

```
# influential Plot
influencePlot(fit_inter_Paid_3, scale=5, xlab="Hat-Values", ylab="Studentized Residual
s",
            fill.col=carPalette()[2], fill.alpha=0.5, id=TRUE)
```
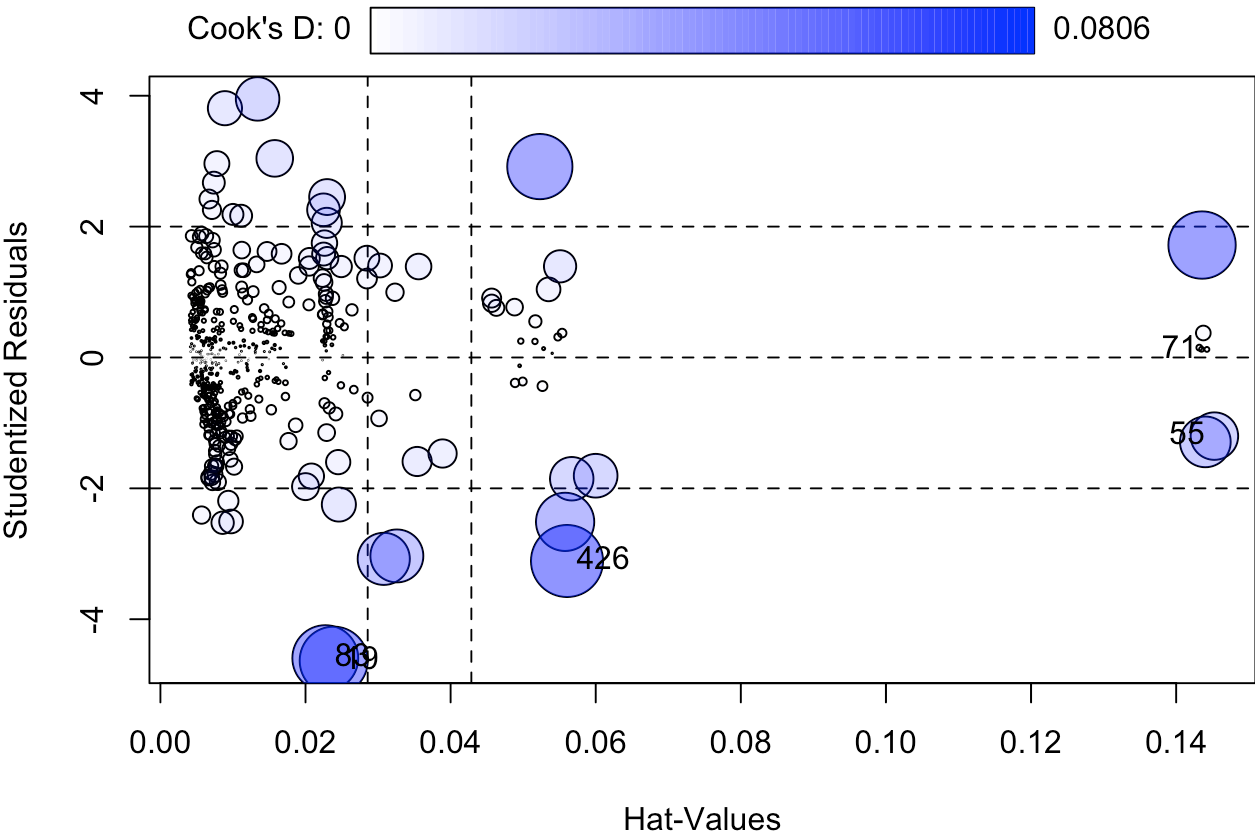


```
##       StudRes        Hat        CookD
## 19  -4.633873 0.02388823 0.0720187742
## 55  -1.201781 0.14527074 0.0350350462
## 71   0.123237 0.14424148 0.0003664455
## 83  -4.594521 0.02272519 0.0673221092
## 426 -3.110404 0.05604218 0.0806060069
```

Model Interation Validation

```r
# residual vs fitted Model
residual_plot <-  ggplot(fit_inter_Paid_3, aes(x = .fitted, y = .resid)) +
                geom_point() +
                geom_hline(yintercept = 0, col = "red") +
                labs(title="Fitted",
                  x = "Fitted", y = "Residual") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10))


# Page Total Likes vs residuals
PTLike_Paid_plot <-  ggplot(mydata, aes(x = PTLike, y =rstandard(fit_inter_Paid_3))) +
                geom_point() +
                geom_hline(yintercept = 0, col = "red") +
                labs(title="Page Total Likes",
                x = "Page Total Likes", y = "Residual") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))   +
                theme(axis.title = element_text(size = 10))


#create Q-Q plot
qq_plot <-  ggplot(fit_inter_Paid_3, aes(sample=rstandard(fit_inter_Paid_3))) +
                stat_qq(size=1.5, color='blue') +
                stat_qq_line(col = "red") +
                labs(title="Q-Q Plot",
                  x = "Theoretical Quantiles", y = "Sample Quantiles") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10))


# combine all plots
fit_inter_Paid_plot <- ggarrange(residual_plot, PTLike_Paid_plot, qq_plot,
                labels = c("Fig A", "Fig B", "Fig C"),
                font.label = list(size = 9, color = "blue"))
# plot all
fit_inter_Paid_plot
```

Fig A

Fitted

Fig B

Page Total Likes

Fig C

Q-Q Plot
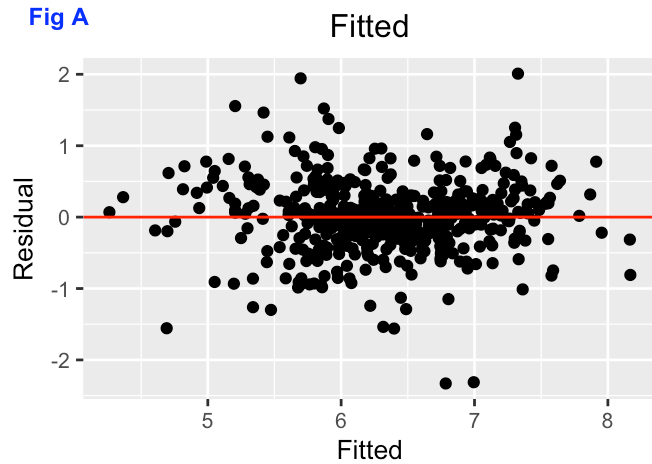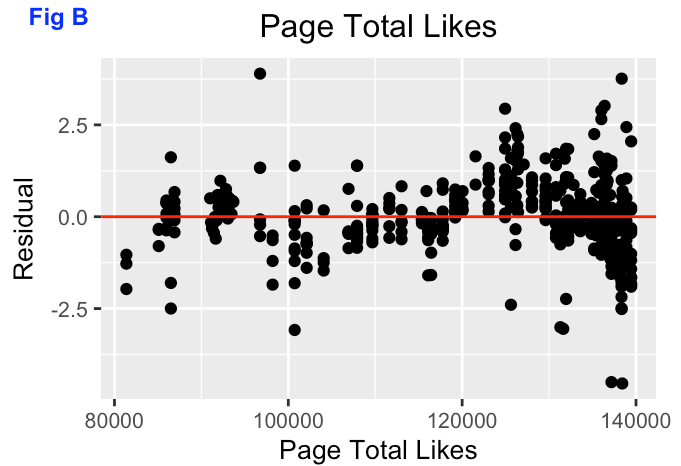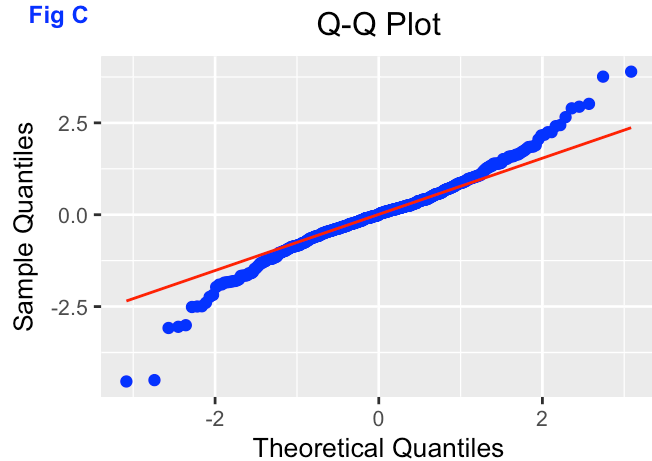

Full Interaction Model

```
# full interaction model
fit_full_interaction_1 <- lm(log(LPConsumer) ~ (PTLike + Paid + typeP + typeS + log(Tota
lInterac) +
                        typeV + category1)^2 , data=mydata)

# summary results
summary(fit_full_interaction_1)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ (PTLike + Paid + typeP + typeS +
##     log(TotalInterac) + typeV + category1)^2, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23684 -0.26875 -0.02323  0.25290  1.98642
##
## Coefficients: (4 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                8.817e-01  1.046e+00   0.843   0.3997
## PTLike                     5.128e-07  8.635e-06   0.059   0.9527
## Paid                       3.606e-01  5.179e-01   0.696   0.4867
## typeP                      4.417e+00  7.787e-01   5.672 2.48e-08 ***
## typeS                      1.649e+00  1.622e+00   1.017   0.3099
## log(TotalInterac)          1.226e+00  1.881e-01   6.521 1.83e-10 ***
## typeV                      7.242e+00  7.441e+00   0.973   0.3309
## category1                  7.862e-01  5.268e-01   1.492   0.1363
## PTLike:Paid               -3.029e-07  3.306e-06  -0.092   0.9270
## PTLike:typeP              -1.075e-05  5.803e-06  -1.853   0.0644 .
## PTLike:typeS              -7.493e-07  1.237e-05  -0.061   0.9517
## PTLike:log(TotalInterac)  -2.779e-06  1.397e-06  -1.990   0.0472 *
## PTLike:typeV              -5.457e-05  4.986e-05  -1.094   0.2744
## PTLike:category1           2.641e-06  3.122e-06   0.846   0.3981
## Paid:typeP                 3.217e-01  2.633e-01   1.222   0.2224
## Paid:typeS                -1.934e-02  3.295e-01  -0.059   0.9532
## Paid:log(TotalInterac)    -1.026e-01  5.215e-02  -1.968   0.0497 *
## Paid:typeV                 6.519e-01  4.836e-01   1.348   0.1783
## Paid:category1            -1.827e-02  1.095e-01  -0.167   0.8676
## typeP:typeS                      NA         NA      NA       NA
## typeP:log(TotalInterac)   -4.627e-01  1.033e-01  -4.480 9.40e-06 ***
## typeP:typeV                      NA         NA      NA       NA
## typeP:category1           -3.673e-01  3.875e-01  -0.948   0.3437
## typeS:log(TotalInterac)    8.329e-02  1.470e-01   0.567   0.5712
## typeS:typeV                      NA         NA      NA       NA
## typeS:category1           -3.837e-01  5.037e-01  -0.762   0.4466
## log(TotalInterac):typeV    1.195e-01  3.115e-01   0.384   0.7014
## log(TotalInterac):category1 -7.309e-02 4.859e-02  -1.504   0.1332
## typeV:category1                  NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4843 on 465 degrees of freedom
## Multiple R-squared:  0.6776, Adjusted R-squared:  0.6609
## F-statistic: 40.71 on 24 and 465 DF,  p-value: < 2.2e-16
```

Full Interaction Variable Selection

```
# stepwise variable selection on the full interaction model
step(fit_full_interaction_1, direction = "backward", trace = FALSE)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
##     log(TotalInterac) + typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##     Paid:typeP + Paid:log(TotalInterac) + typeP:log(TotalInterac) +
##     log(TotalInterac):category1, data = mydata)
##
## Coefficients:
##               (Intercept)                        PTLike
##                 5.892e-01                     3.839e-06
##                      Paid                         typeP
##                 3.570e-01                     4.418e+00
##                     typeS             log(TotalInterac)
##                 1.565e+00                     1.306e+00
##                     typeV                     category1
##                 8.458e-01                     7.803e-01
##             PTLike:typeP     PTLike:log(TotalInterac)
##                -1.188e-05                    -3.026e-06
##               Paid:typeP        Paid:log(TotalInterac)
##                 2.488e-01                    -9.604e-02
##   typeP:log(TotalInterac)  log(TotalInterac):category1
##                -5.080e-01                    -8.096e-02
```

```r
# selecting the model from stepwise backward variable selection process
fit_full_interaction_2 <- lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
                      log(TotalInterac) + typeV + category1 + PTLike:typeP +
                      PTLike:log(TotalInterac) + Paid:typeP + Paid:log(TotalInterac)
+
                      typeP:log(TotalInterac) + log(TotalInterac):category1, data =
mydata)

# after stepwise fit model
summary(fit_full_interaction_2)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
##     log(TotalInterac) + typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##     Paid:typeP + Paid:log(TotalInterac) + typeP:log(TotalInterac) +
##     log(TotalInterac):category1, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24014 -0.27223 -0.02708  0.25462  1.93762
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   5.892e-01  8.509e-01   0.692 0.489009
## PTLike                        3.839e-06  6.724e-06   0.571 0.568351
## Paid                          3.570e-01  2.676e-01   1.334 0.182804
## typeP                         4.418e+00  5.955e-01   7.420 5.44e-13 ***
## typeS                         1.565e+00  1.630e-01   9.601  < 2e-16 ***
## log(TotalInterac)             1.306e+00  1.697e-01   7.697 8.12e-14 ***
## typeV                         8.458e-01  2.438e-01   3.470 0.000568 ***
## category1                     7.803e-01  2.255e-01   3.459 0.000590 ***
## PTLike:typeP                 -1.188e-05  4.657e-06  -2.550 0.011084 *
## PTLike:log(TotalInterac)     -3.026e-06  1.252e-06  -2.417 0.016036 *
## Paid:typeP                    2.488e-01  1.409e-01   1.766 0.078080 .
## Paid:log(TotalInterac)       -9.604e-02  4.840e-02  -1.984 0.047805 *
## typeP:log(TotalInterac)      -5.080e-01  6.795e-02  -7.476 3.72e-13 ***
## log(TotalInterac):category1  -8.096e-02  4.531e-02  -1.787 0.074638 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4812 on 476 degrees of freedom
## Multiple R-squared:  0.6741, Adjusted R-squared:  0.6652
## F-statistic: 75.75 on 13 and 476 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_interaction_2)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                              Df  Sum Sq Mean Sq  F value     Pr(>F)
## PTLike                        1  25.023  25.023 108.0559  < 2.2e-16 ***
## Paid                          1   4.580   4.580  19.7767 1.084e-05 ***
## typeP                         1  11.680  11.680  50.4371 4.508e-12 ***
## typeS                         1  56.578  56.578 244.3191  < 2.2e-16 ***
## log(TotalInterac)             1  79.674  79.674 344.0539  < 2.2e-16 ***
## typeV                         1  14.718  14.718  63.5571 1.165e-14 ***
## category1                     1  16.376  16.376  70.7151 4.835e-16 ***
## PTLike:typeP                  1   3.476   3.476  15.0094   0.000122 ***
## PTLike:log(TotalInterac)      1   1.700   1.700   7.3392   0.006990 **
## Paid:typeP                    1   0.212   0.212   0.9166   0.338859
## Paid:log(TotalInterac)        1   0.145   0.145   0.6281   0.428440
## typeP:log(TotalInterac)       1  13.149  13.149  56.7796 2.477e-13 ***
## log(TotalInterac):category1   1   0.739   0.739   3.1920   0.074638 .
## Residuals                   476 110.229   0.232
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# remove interaction variable due to t-value: Paid:typeP
fit_full_interaction_3 <- lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
                        log(TotalInterac) + typeV + category1 + PTLike:typeP +
                        PTLike:log(TotalInterac) + Paid:log(TotalInterac) +
                        typeP:log(TotalInterac) + log(TotalInterac):category1, data =
mydata)

# summary model display
summary(fit_full_interaction_3)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
##     log(TotalInterac) + typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##     Paid:log(TotalInterac) + typeP:log(TotalInterac) + log(TotalInterac):category1,
##     data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.19466 -0.27598 -0.03091  0.25279  1.95763
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  6.076e-01  8.527e-01   0.713  0.47646
## PTLike                       3.657e-06  6.738e-06   0.543  0.58753
## Paid                         5.479e-01  2.453e-01   2.234  0.02596 *
## typeP                        4.386e+00  5.965e-01   7.352 8.55e-13 ***
## typeS                        1.591e+00  1.626e-01   9.783  < 2e-16 ***
## log(TotalInterac)            1.301e+00  1.700e-01   7.650 1.12e-13 ***
## typeV                        8.000e-01  2.429e-01   3.293  0.00106 **
## category1                    7.544e-01  2.256e-01   3.344  0.00089 ***
## PTLike:typeP                -1.140e-05  4.659e-06  -2.446  0.01482 *
## PTLike:log(TotalInterac)    -3.088e-06  1.254e-06  -2.461  0.01419 *
## Paid:log(TotalInterac)      -9.154e-02  4.844e-02  -1.890  0.05942 .
## typeP:log(TotalInterac)     -4.975e-01  6.784e-02  -7.333 9.70e-13 ***
## log(TotalInterac):category1 -7.565e-02  4.531e-02  -1.670  0.09566 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4823 on 477 degrees of freedom
## Multiple R-squared:  0.672,  Adjusted R-squared:  0.6638
## F-statistic: 81.44 on 12 and 477 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_interaction_3)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                            Df  Sum Sq Mean Sq  F value    Pr(>F)
## PTLike                      1  25.023  25.023 107.5782 < 2.2e-16 ***
## Paid                        1   4.580   4.580  19.6893 1.133e-05 ***
## typeP                       1  11.680  11.680  50.2141 4.983e-12 ***
## typeS                       1  56.578  56.578 243.2391 < 2.2e-16 ***
## log(TotalInterac)           1  79.674  79.674 342.5331 < 2.2e-16 ***
## typeV                       1  14.718  14.718  63.2762 1.316e-14 ***
## category1                   1  16.376  16.376  70.4025 5.525e-16 ***
## PTLike:typeP                1   3.476   3.476  14.9430 0.0001262 ***
## PTLike:log(TotalInterac)    1   1.700   1.700   7.3068 0.0071147 **
## Paid:log(TotalInterac)      1   0.135   0.135   0.5824 0.4457642
## typeP:log(TotalInterac)     1  12.740  12.740  54.7705 6.164e-13 ***
## log(TotalInterac):category1 1   0.648   0.648   2.7874 0.0956625 .
## Residuals                 477 110.951   0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# remove interaction variable due to t-value: log(TotalInterac):category1
fit_full_interaction_4 <- lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
                    log(TotalInterac) + typeV + category1 + PTLike:typeP +
                    PTLike:log(TotalInterac) + Paid:log(TotalInterac) +
                    typeP:log(TotalInterac), data = mydata)


# summary model display
summary(fit_full_interaction_4)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
##     log(TotalInterac) + typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##     Paid:log(TotalInterac) + typeP:log(TotalInterac), data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20930 -0.27061 -0.02066  0.25627  1.92765
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                9.382e-01  8.310e-01   1.129  0.25943
## PTLike                     3.085e-06  6.742e-06   0.458  0.64747
## Paid                       4.514e-01  2.388e-01   1.890  0.05936 .
## typeP                      4.416e+00  5.974e-01   7.393 6.47e-13 ***
## typeS                      1.572e+00  1.626e-01   9.673  < 2e-16 ***
## log(TotalInterac)          1.232e+00  1.653e-01   7.454 4.29e-13 ***
## typeV                      7.486e-01  2.414e-01   3.101  0.00204 **
## category1                  3.870e-01  4.977e-02   7.777 4.62e-14 ***
## PTLike:typeP              -1.158e-05  4.667e-06  -2.482  0.01339 *
## PTLike:log(TotalInterac)  -2.924e-06  1.253e-06  -2.334  0.02003 *
## Paid:log(TotalInterac)    -7.240e-02  4.716e-02  -1.535  0.12537
## typeP:log(TotalInterac)   -5.017e-01  6.792e-02  -7.387 6.75e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4832 on 478 degrees of freedom
## Multiple R-squared:  0.6701, Adjusted R-squared:  0.6625
## F-statistic: 88.26 on 11 and 478 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_interaction_4)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                            Df  Sum Sq Mean Sq  F value     Pr(>F)
## PTLike                      1  25.023  25.023 107.1774 < 2.2e-16 ***
## Paid                        1   4.580   4.580  19.6160 1.175e-05 ***
## typeP                       1  11.680  11.680  50.0271 5.419e-12 ***
## typeS                       1  56.578  56.578 242.3329 < 2.2e-16 ***
## log(TotalInterac)           1  79.674  79.674 341.2570 < 2.2e-16 ***
## typeV                       1  14.718  14.718  63.0405 1.457e-14 ***
## category1                   1  16.376  16.376  70.1402 6.175e-16 ***
## PTLike:typeP                1   3.476   3.476  14.8874 0.0001298 ***
## PTLike:log(TotalInterac)    1   1.700   1.700   7.2796 0.0072211 **
## Paid:log(TotalInterac)      1   0.135   0.135   0.5802 0.4466115
## typeP:log(TotalInterac)     1  12.740  12.740  54.5664 6.746e-13 ***
## Residuals                 478 111.600   0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# remove interaction variable due to t-value: Paid:log(TotalInterac)
fit_full_interaction_5 <- lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
                    log(TotalInterac) + typeV + category1 + PTLike:typeP +
                    PTLike:log(TotalInterac) + typeP:log(TotalInterac), data = myd
ata)

# summary model display
summary(fit_full_interaction_5)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + Paid + typeP + typeS +
##     log(TotalInterac) + typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##     typeP:log(TotalInterac), data = mydata)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -2.21209 -0.27864 -0.01915  0.25804  1.90509
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              9.913e-01  8.314e-01   1.192  0.23373
## PTLike                   3.774e-06  6.737e-06   0.560  0.57561
## Paid                     9.264e-02  4.946e-02   1.873  0.06165 .
## typeP                    4.376e+00  5.976e-01   7.322 1.04e-12 ***
## typeS                    1.568e+00  1.628e-01   9.633  < 2e-16 ***
## log(TotalInterac)        1.219e+00  1.653e-01   7.375 7.27e-13 ***
## typeV                    7.452e-01  2.417e-01   3.083  0.00217 **
## category1                3.829e-01  4.977e-02   7.694 8.17e-14 ***
## PTLike:typeP            -1.172e-05  4.672e-06  -2.509  0.01243 *
## PTLike:log(TotalInterac) -3.046e-06  1.252e-06  -2.432  0.01537 *
## typeP:log(TotalInterac) -4.907e-01  6.764e-02  -7.255 1.62e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4839 on 479 degrees of freedom
## Multiple R-squared:  0.6685, Adjusted R-squared:  0.6615
## F-statistic: 96.58 on 10 and 479 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_interaction_5)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                          Df  Sum Sq Mean Sq F value    Pr(>F)
## PTLike                    1  25.023  25.023 106.875 < 2.2e-16 ***
## Paid                      1   4.580   4.580  19.561 1.207e-05 ***
## typeP                     1  11.680  11.680  49.886 5.770e-12 ***
## typeS                     1  56.578  56.578 241.648 < 2.2e-16 ***
## log(TotalInterac)         1  79.674  79.674 340.293 < 2.2e-16 ***
## typeV                     1  14.718  14.718  62.862 1.573e-14 ***
## category1                 1  16.376  16.376  69.942 6.710e-16 ***
## PTLike:typeP              1   3.476   3.476  14.845 0.0001326 ***
## PTLike:log(TotalInterac)  1   1.700   1.700   7.259 0.0073025 **
## typeP:log(TotalInterac)   1  12.325  12.325  52.640 1.624e-12 ***
## Residuals               479 112.150   0.234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# remove interaction variable due to t-value: Paid:log(TotalInterac)
fit_full_interaction_6 <- lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS +
                       log(TotalInterac) + typeV + category1 + PTLike:typeP +
                       PTLike:log(TotalInterac) + typeP:log(TotalInterac), data = myd
ata)

# summary model display
summary(fit_full_interaction_6)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInterac) +
##     typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##     typeP:log(TotalInterac), data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23186 -0.27005 -0.01573  0.26218  1.96499
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              9.930e-01  8.336e-01   1.191  0.23414
## PTLike                   3.807e-06  6.754e-06   0.564  0.57328
## typeP                    4.392e+00  5.991e-01   7.330 9.84e-13 ***
## typeS                    1.561e+00  1.632e-01   9.568  < 2e-16 ***
## log(TotalInterac)        1.218e+00  1.658e-01   7.349 8.65e-13 ***
## typeV                    7.625e-01  2.422e-01   3.148  0.00174 **
## category1                3.893e-01  4.978e-02   7.821 3.35e-14 ***
## PTLike:typeP            -1.194e-05  4.683e-06  -2.551  0.01106 *
## PTLike:log(TotalInterac) -3.006e-06  1.255e-06  -2.395  0.01703 *
## typeP:log(TotalInterac) -4.884e-01  6.780e-02  -7.203 2.29e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4851 on 480 degrees of freedom
## Multiple R-squared:  0.666,  Adjusted R-squared:  0.6598
## F-statistic: 106.4 on 9 and 480 DF,  p-value: < 2.2e-16
```

```
# analysis of variance
anova(fit_full_interaction_6)
```

```
## Analysis of Variance Table
##
## Response: log(LPConsumer)
##                          Df  Sum Sq Mean Sq  F value    Pr(>F)
## PTLike                    1  25.023  25.023 106.3189 < 2.2e-16 ***
## typeP                     1  11.526  11.526  48.9739 8.772e-12 ***
## typeS                     1  54.651  54.651 232.2032 < 2.2e-16 ***
## log(TotalInterac)         1  84.587  84.587 359.3985 < 2.2e-16 ***
## typeV                     1  15.182  15.182  64.5043 7.506e-15 ***
## category1                 1  16.924  16.924  71.9087 2.803e-16 ***
## PTLike:typeP              1   3.540   3.540  15.0421 0.0001198 ***
## PTLike:log(TotalInterac)  1   1.661   1.661   7.0589 0.0081499 **
## typeP:log(TotalInterac)   1  12.213  12.213  51.8900 2.286e-12 ***
## Residuals               480 112.972   0.235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 95% confidence interval of the fitted model
confint(fit_full_interaction_6, level=0.95)
```

```
##                                2.5 %        97.5 %
## (Intercept)              -6.449017e-01  2.630957e+00
## PTLike                   -9.464860e-06  1.707850e-05
## typeP                     3.214333e+00  5.568783e+00
## typeS                     1.240404e+00  1.881561e+00
## log(TotalInterac)         8.925037e-01  1.543929e+00
## typeV                     2.866074e-01  1.238406e+00
## category1                 2.915339e-01  4.871637e-01
## PTLike:typeP             -2.114688e-05 -2.743537e-06
## PTLike:log(TotalInterac) -5.472391e-06 -5.392596e-07
## typeP:log(TotalInterac)  -6.216273e-01 -3.551799e-01
```

```
# influential Plot
influencePlot(fit_full_interaction_6)
```
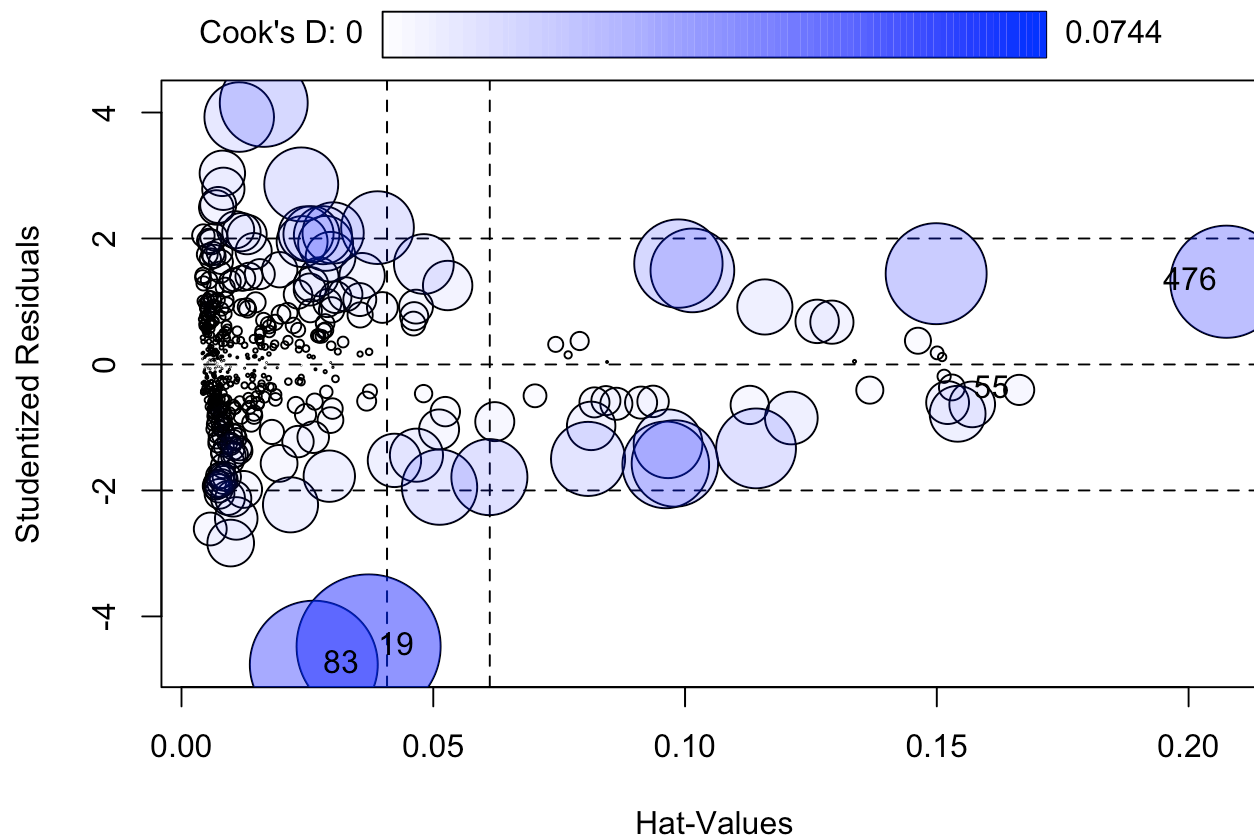
```
##         StudRes        Hat        CookD
## 19  -4.4764820 0.03715739 0.074382439
## 55  -0.4039633 0.16635168 0.003262015
## 83  -4.7664314 0.02627265 0.058645401
## 476  1.3135268 0.20750861 0.045109048
```

Full Interaction Influential Points and Outliers

```
# plot of deleted studentized residuals vs hat values
student_hat <- data.frame(rstudent = rstudent(fit_full_interaction_6), hatvalues =
                          hatvalues(fit_full_interaction_6))

# plot rstudent vs hatvalues
student_hat_plot <- ggplot(student_hat, aes(x = hatvalues, y = rstudent)) + geom_point()
+
                # Change line type and color
                geom_hline(yintercept=3, linetype="dashed", color = "red") +
                geom_hline(yintercept=-3, linetype="dashed", color = "red")

# plot
student_hat_plot
```
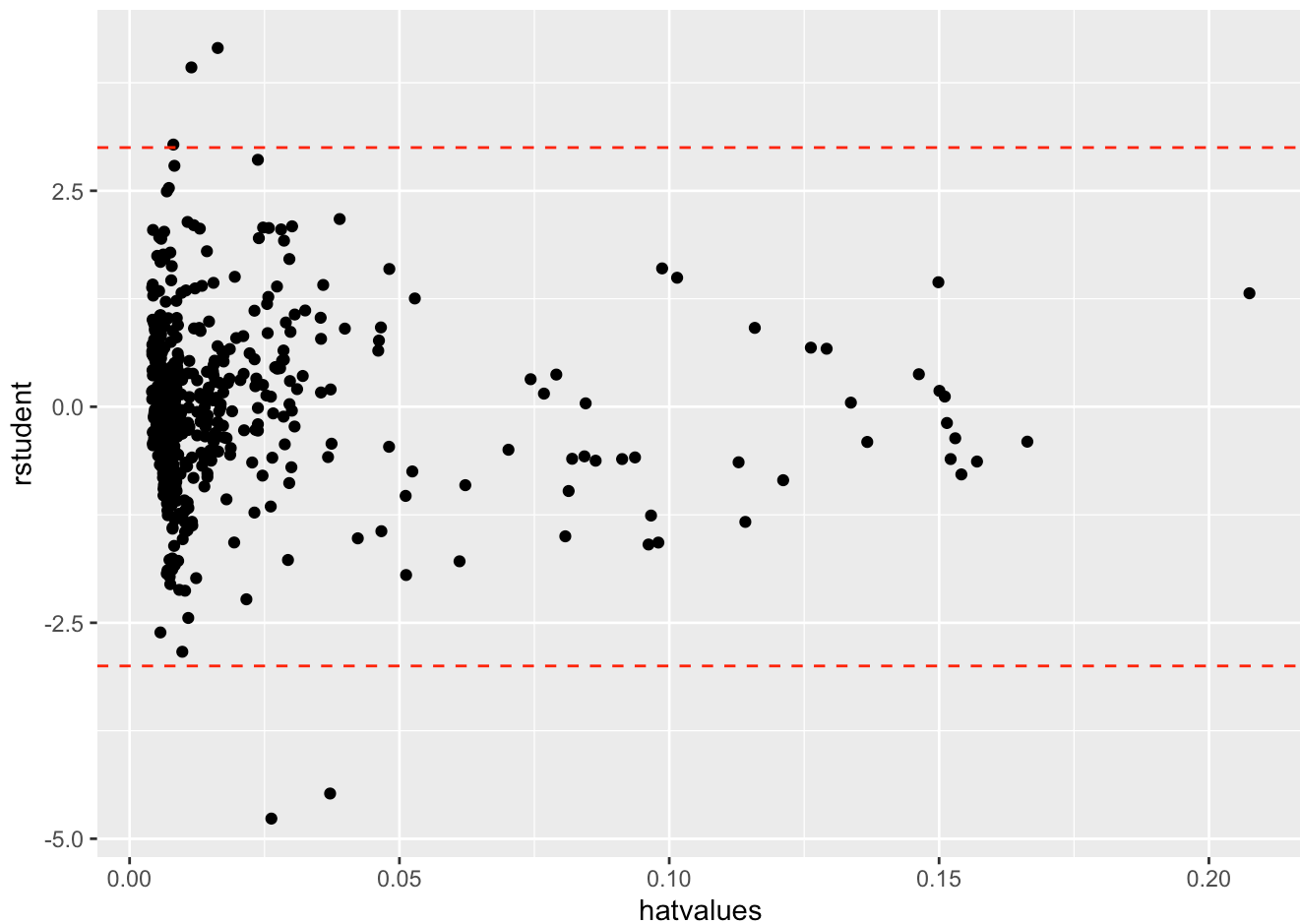
```
# outliers |standardized residuals| > 3
std_residual = data.frame(residual = rstandard(fit_full_interaction_6))

# display |standardized residuals| > 3
filter(std_residual, abs(residual) > 3)
```

```
##       residual
## 19  -4.390260
## 42   3.869897
## 83  -4.662123
## 311  3.008932
## 441  4.083869
```

```
# historgram for outliers
ggplot(std_residual, aes(x = residual)) +
  geom_histogram(bins=30, color="black", fill="lightblue") +
  labs(title = "Lifetime Post Consumers \n Standarized Residual", x = "Standarized Resid
ual",
      y = 'Frequency') +
    geom_vline(xintercept = 3, linetype="dotted",
              color = "red") +
    geom_vline(xintercept = -3, linetype="dotted",
              color = "red") +
    # move the title text to the middle
    theme(plot.title=element_text(hjust=0.5)) +
    theme(text = element_text(size = 10))  +
    theme(axis.title = element_text(size = 10))
```

Lifetime Post Consumers
Standarized Residual
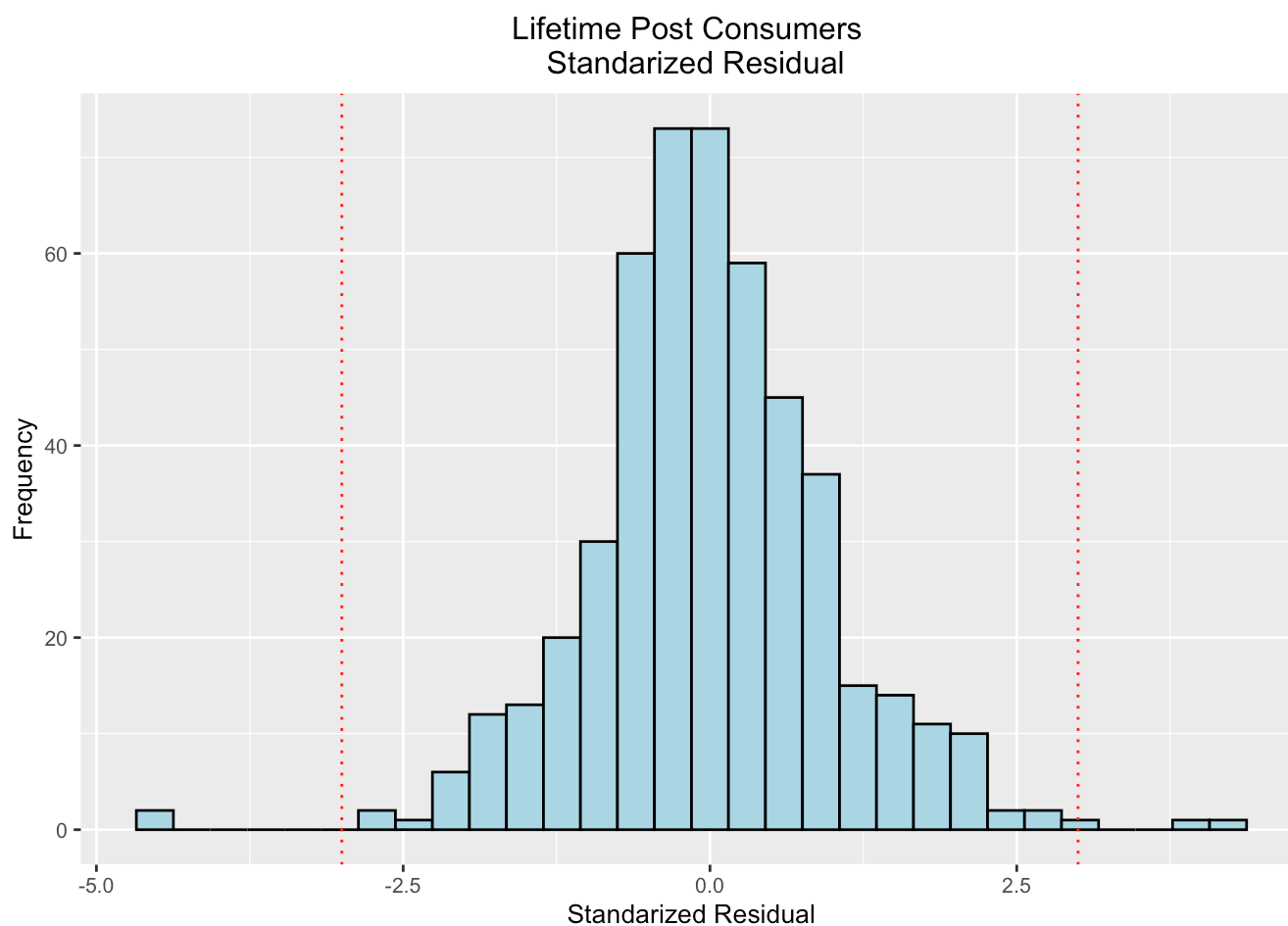

```
# print out only observations that may be influential
summary(influence.measures(fit_full_interaction_6))
```

```
## Potentially influential observations of
##   lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInterac) +      ty
peV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +      typeP:log(TotalIntera
c), data = mydata) :
##
##      dfb.1_ dfb.PTLk dfb.typP dfb.typS dfb.l(TI dfb.typV dfb.ctg1 dfb.PTL:P
## 19    0.10  -0.27    -0.02    -0.39     0.10    -0.11     0.04      0.25
## 22   -0.10   0.18     0.06    -0.15    -0.01    -0.09     0.01     -0.18
## 26    0.01   0.01    -0.05     0.01     0.00     0.03     0.00      0.02
## 29    0.00   0.01     0.01     0.00    -0.01     0.12     0.00     -0.02
## 41    0.07  -0.06    -0.13     0.19    -0.01     0.15    -0.01      0.12
## 42   -0.15   0.14    -0.05     0.05     0.17     0.01     0.16      0.09
## 43    0.06  -0.09    -0.07     0.13     0.00     0.09    -0.01      0.12
## 45    0.06  -0.08    -0.08     0.13     0.00     0.10    -0.01      0.11
## 47    0.06  -0.10    -0.06     0.11     0.00     0.07    -0.01      0.11
## 49   -0.07   0.14     0.04    -0.10    -0.01    -0.06     0.01     -0.14
## 50    0.05  -0.06     0.04     0.03    -0.07    -0.01     0.16     -0.06
## 52   -0.04   0.05     0.03    -0.04     0.06     0.01    -0.17     -0.02
## 55    0.00  -0.02     0.01    -0.02     0.01    -0.15     0.00      0.02
## 71    0.00   0.02    -0.04     0.04     0.00    -0.16     0.00      0.01
## 74    0.00  -0.03     0.01    -0.02     0.02    -0.27     0.00      0.03
## 83    0.08  -0.16    -0.08    -0.32     0.07    -0.02     0.04      0.20
## 85    0.12  -0.11    -0.22     0.35    -0.03     0.27    -0.02      0.22
## 128  -0.18   0.18    -0.02     0.03     0.22     0.02     0.06      0.08
## 133  -0.03   0.05     0.04    -0.08     0.00    -0.05     0.00     -0.06
## 137  -0.14   0.11     0.26    -0.42     0.04    -0.33     0.02     -0.24
## 139   0.00   0.00     0.00     0.00     0.00     0.00     0.00      0.00
## 141  -0.07   0.06    -0.03     0.04     0.07     0.00     0.13      0.05
## 146   0.05  -0.09    -0.05     0.11     0.00     0.08    -0.01      0.10
## 180   0.00   0.01    -0.01     0.01     0.00    -0.05     0.00      0.00
## 229  -0.04  -0.10     0.17    -0.29     0.08    -0.17     0.02      0.02
## 232  -0.04  -0.09     0.17    -0.26     0.08    -0.16     0.01      0.02
## 240   0.01  -0.05    -0.01    -0.02     0.04     0.43     0.00      0.07
## 241   0.12  -0.12    -0.05     0.00    -0.15    -0.01     0.03     -0.01
## 274   0.00  -0.01    -0.01     0.01     0.01     0.06     0.00      0.02
## 276  -0.01   0.00    -0.01     0.04     0.00     0.00     0.17      0.01
## 280  -0.02   0.02     0.01     0.03     0.02     0.00     0.10      0.01
## 286   0.02  -0.01    -0.01    -0.03    -0.01     0.00    -0.13     -0.01
## 311  -0.02   0.01     0.03     0.03     0.02     0.00     0.11      0.01
## 349  -0.04   0.11     0.20    -0.23    -0.15    -0.04    -0.18     -0.28
## 368   0.00   0.11     0.17    -0.12    -0.25     0.01     0.01     -0.33
## 400   0.02   0.00     0.01     0.09    -0.08     0.07    -0.01     -0.05
## 413   0.13  -0.11    -0.20     0.20    -0.04     0.11     0.07      0.16
## 418  -0.11   0.11    -0.06     0.01     0.13     0.01     0.00      0.04
## 421   0.01  -0.01    -0.02    -0.02     0.01    -0.01    -0.01      0.02
## 422  -0.25   0.26    -0.15     0.02     0.31     0.03    -0.03      0.10
## 424   0.31  -0.31     0.17    -0.03    -0.38    -0.03    -0.01     -0.11
## 425  -0.17   0.17    -0.09     0.02     0.21     0.02     0.00      0.06
## 426  -0.17   0.10     0.14    -0.01     0.14    -0.03     0.00     -0.05
## 427  -0.08   0.08    -0.05     0.01     0.10     0.01     0.00      0.03
## 428  -0.13   0.13    -0.08     0.01     0.16     0.01    -0.01      0.05
## 434  -0.21   0.17     0.26     0.14     0.01     0.06     0.15     -0.24
```

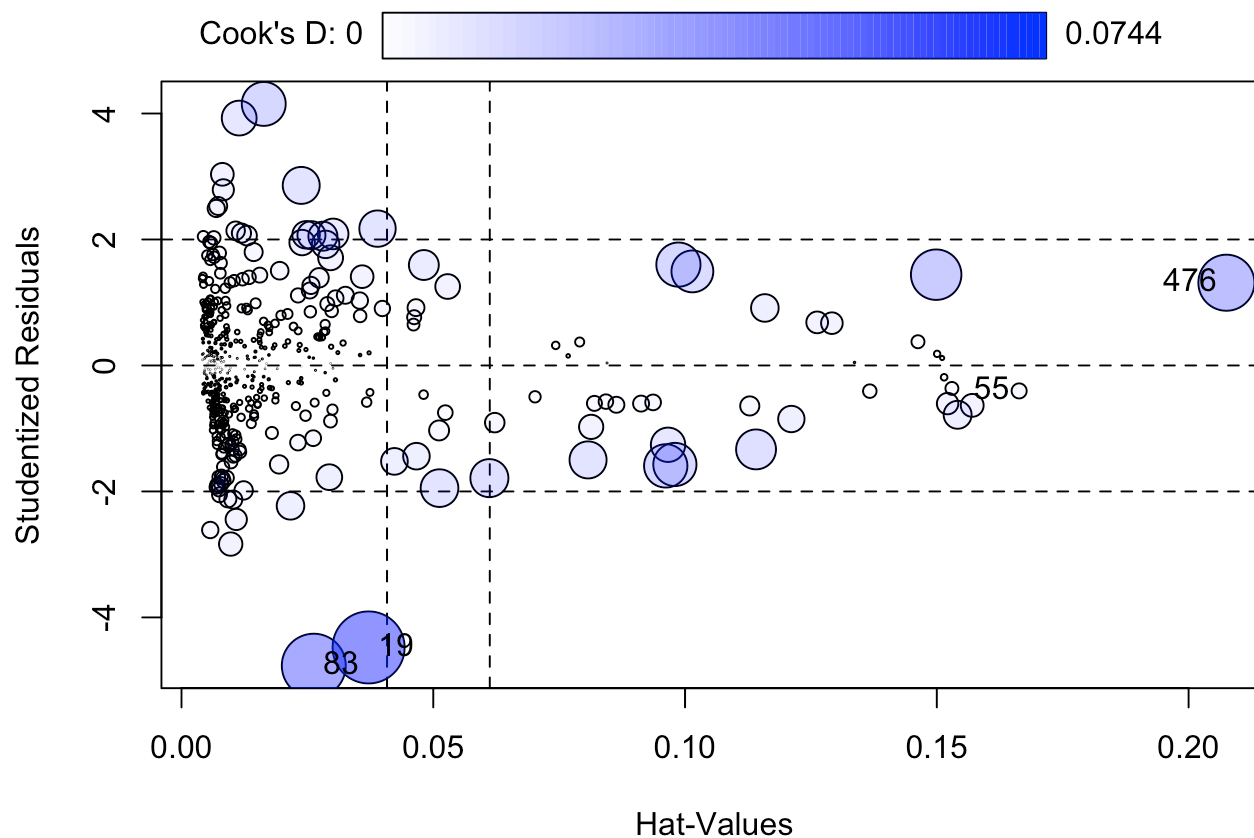```
441 -0.16   0.15   0.05   0.09   0.18   0.01   0.30   -0.07
455 -0.06   0.06   -0.01   0.01   0.07   0.01   0.00   0.00
465 -0.07   0.09   0.17   0.08   -0.11   0.06   -0.01   -0.21
471 0.10 -0.09   0.01   -0.03   -0.11   -0.01   -0.07   0.00
472 0.01 -0.01   -0.01   0.00   0.01   0.00   0.00   0.01
476 0.57 -0.44   -0.44   0.08   -0.39   0.10   0.00   0.27
480 -0.12   0.10   0.11   -0.01   0.07   -0.02   0.00   -0.08
487 0.01 -0.02   -0.03   -0.01   0.02   -0.01   0.00   0.04
dfb.PTL:(dfb.tP:(dffit   cov.r   cook.d hat
19   0.10   -0.47   -0.88_* 0.70_* 0.07   0.04
22   -0.06   0.16   0.33   1.13_* 0.01   0.12_*
26   -0.03   0.06   -0.10   1.07_* 0.00   0.05
29   0.00   0.02   0.16   1.19_* 0.00   0.15_*
41   -0.03   0.10   -0.23   1.14_* 0.01   0.11_*
42   -0.19   -0.05   0.42   0.75_* 0.02   0.01
43   0.01   -0.03   -0.19   1.11_* 0.00   0.09_*
45   0.00   0.01   -0.17   1.11_* 0.00   0.08_*
47   0.02   -0.05   -0.19   1.12_* 0.00   0.09_*
49   -0.05   0.14   0.26   1.16_* 0.01   0.13_*
50   0.08   0.04   -0.28   0.87_* 0.01   0.01
52   -0.06   -0.04   -0.26   0.91_* 0.01   0.01
55   0.01   -0.06   -0.18   1.22_* 0.00   0.17_*
71   -0.03   0.07   -0.27   1.20_* 0.01   0.16_*
74   0.02   -0.08   -0.33   1.19_* 0.01   0.15_*
83   0.02   -0.22   -0.78_* 0.66_* 0.06   0.03
85   -0.04   0.16   -0.41   1.09_* 0.02   0.10_*
128 -0.24   -0.09   0.45_* 0.88_* 0.02   0.02
133 -0.01   0.01   0.11   1.11_* 0.00   0.08_*
137 0.06   -0.23   0.50_* 1.08_* 0.03   0.10_*
139 0.00   -0.01   0.01   1.12_* 0.00   0.08_*
141 -0.08   -0.02   0.26   0.88_* 0.01   0.01
146 0.02   -0.05   -0.18   1.10_* 0.00   0.08_*
180 -0.01   0.02   -0.08   1.20_* 0.00   0.15_*
229 0.10   -0.42   -0.52_* 1.08_* 0.03   0.10_*
232 0.09   -0.40   -0.48_* 1.11_* 0.02   0.11_*
240 0.00   -0.11   0.61_* 1.15_* 0.04   0.15_*
241 0.16   0.12   0.44_* 0.96   0.02   0.04
274 -0.01   0.00   0.08   1.20_* 0.00   0.15_*
276 0.00   0.03   0.22   0.90_* 0.00   0.01
280 -0.02   -0.03   0.21   0.90_* 0.00   0.01
286 0.01   0.01   -0.20   0.89_* 0.00   0.01
311 -0.02   -0.05   0.27   0.85_* 0.01   0.01
349 0.10   0.11   -0.46_* 1.02   0.02   0.06
368 0.14   0.30   -0.52_* 1.07_* 0.03   0.10_*
400 0.03   0.11   -0.16   1.18_* 0.00   0.14_*
413 0.00   0.13   0.26   1.16_* 0.01   0.13_*
418 -0.15   0.04   -0.19   1.12_* 0.00   0.09_*
421 0.00   -0.01   0.04   1.11_* 0.00   0.08_*
422 -0.34   0.09   -0.45_* 0.99   0.02   0.05
424 0.41   -0.11   0.53_* 1.07_* 0.03   0.10_*
425 -0.22   0.06   -0.29   1.09_* 0.01   0.08_*
426 -0.08   -0.17   -0.26   1.20_* 0.01   0.15_*
```

```
## 427 -0.10      0.03    -0.14      1.09_*  0.00    0.07_*
## 428 -0.18      0.05    -0.23      1.07_*  0.01    0.06_*
## 434 -0.01      0.03    -0.44_*    1.06    0.02    0.08_*
## 441 -0.19      0.07     0.54_*    0.73_*  0.03    0.02
## 455 -0.07      0.02     0.09      1.10_*  0.00    0.07_*
## 465  0.06      0.12    -0.32      1.14_*  0.01    0.12_*
## 471  0.12     -0.03    -0.18      1.06_*  0.00    0.05
## 472  0.00     -0.01     0.02      1.18_*  0.00    0.13_*
## 476  0.28      0.32     0.67_*    1.24_*  0.05    0.21_*
## 480 -0.05     -0.05    -0.15      1.20_*  0.00    0.15_*
## 487 -0.01     -0.02     0.05      1.20_*  0.00    0.15_*
```

```
# influential Plot
influencePlot(fit_full_interaction_6, scale=5, xlab="Hat-Values",
              ylab="Studentized Residuals",
              fill.col=carPalette()[2], fill.alpha=0.5, id=TRUE)
```



```
##         StudRes        Hat         CookD
## 19   -4.4764820  0.03715739  0.074382439
## 55   -0.4039633  0.16635168  0.003262015
## 83   -4.7664314  0.02627265  0.058645401
## 476   1.3135268  0.20750861  0.045109048
```

Full Interaction Model Validation

```r
# residual vs fitted Model
residual_plot <-  ggplot(fit_full_interaction_6, aes(x = .fitted, y = .resid)) +
                  geom_point() +
                  geom_hline(yintercept = 0, col = "red") +
                  labs(title="Fitted",
                    x = "Fitted", y = "Residual") +
                  # move the title text to the middle
                  theme(plot.title=element_text(hjust=0.5))  +
                  theme(text = element_text(size = 10))  +
                  theme(axis.title = element_text(size = 10))


# Page Total Likes vs residuals
PTLike_Full_plot <- ggplot(mydata, aes(x = PTLike, y =rstandard(fit_full_interaction_
6))) +
                geom_point() +
                geom_hline(yintercept = 0, col = "red") +
                labs(title="Page Total Likes",
                x = "Page Total Likes", y = "Residual") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10))


#create Q-Q plot
qq_plot <- ggplot(fit_full_interaction_6, aes(sample=rstandard(fit_full_interaction_6)))
+
                stat_qq(size=1.5, color='blue') +
                stat_qq_line(col = "red") +
                labs(title="Q-Q Plot",
                  x = "Theoretical Quantiles", y = "Sample Quantiles") +
                # move the title text to the middle
                theme(plot.title=element_text(hjust=0.5))  +
                theme(text = element_text(size = 10))  +
                theme(axis.title = element_text(size = 10))


# combine all plots
fit_inter_Paid_plot <- ggarrange(residual_plot, PTLike_Paid_plot, qq_plot,
                labels = c("Fig A", "Fig B", "Fig C"),
                font.label = list(size = 9, color = "blue"))
# plot all
fit_inter_Paid_plot
```
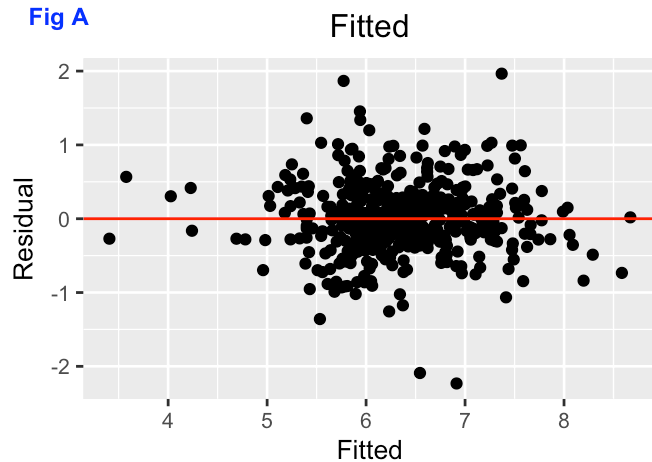
Fig A

Fitted

Fig B

Page Total Likes

Fig C

Q-Q Plot
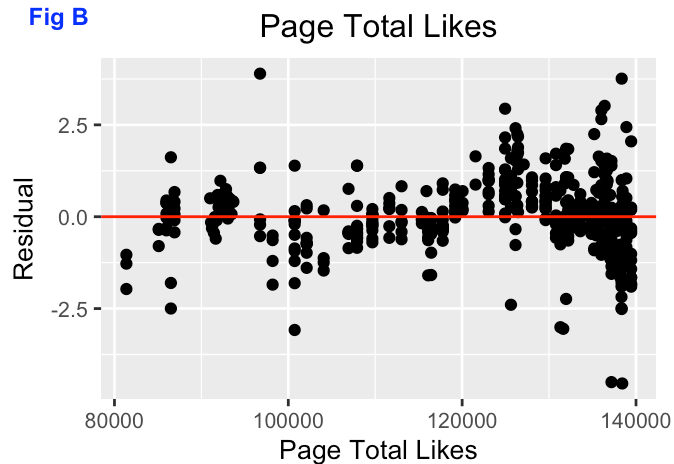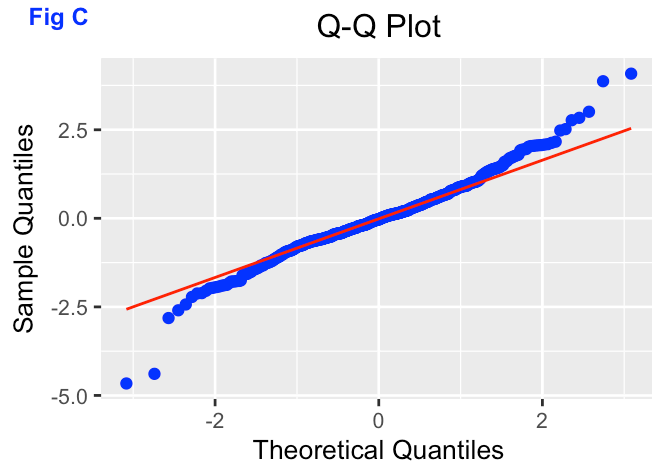

Model Validation

```
# setting the seed to get the same result when knit
set.seed(2500)

# split samples (75% for training and 25% for testing)
select.mydata <- sample(1:nrow(mydata), 0.75*nrow(mydata))

# selecting 75% of the data for training purpose
train.mydata <- mydata[select.mydata,]

# selecting 25% (remaining) of the data for testing
test.mydata <- mydata[-select.mydata,]


# Model: 1 : fit_full_3
fit_m1_trn <- lm(formula = log(LPConsumer) ~ PTLike + log(TotalInterac) + typeP +
                    typeS + typeV + category1, data = mydata)

# summary of fit_m1_trn
summary(fit_m1_trn)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + log(TotalInterac) + typeP +
##      typeS + typeV + category1, data = mydata)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.32981 -0.26168  0.00091  0.27016  2.00912
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.620e+00  2.330e-01  24.124  < 2e-16 ***
## PTLike            -2.077e-05  1.490e-06 -13.935  < 2e-16 ***
## log(TotalInterac)  4.149e-01  2.291e-02  18.115  < 2e-16 ***
## typeP              1.042e+00  1.172e-01   8.894  < 2e-16 ***
## typeS              2.247e+00  1.443e-01  15.568  < 2e-16 ***
## typeV              1.605e+00  2.299e-01   6.979 9.85e-12 ***
## category1          4.202e-01  5.307e-02   7.918 1.67e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5196 on 483 degrees of freedom
## Multiple R-squared:  0.6146, Adjusted R-squared:  0.6098
## F-statistic: 128.4 on 6 and 483 DF,  p-value: < 2.2e-16
```

```r
# create fitted values using test.mydata
y_pred <- predict.glm(fit_m1_trn,test.mydata)
y_obs  <- log(test.mydata[,"LPConsumer"])

# validation statistics
# RMSE of prediction error
rmse_m1 <-sqrt((y_obs-y_pred)%*%(y_obs-y_pred)/nrow(test.mydata))

# compute MAE
mae_m1 <- mean(abs(y_obs-y_pred))

# compute MAPE
mape_m1 <- mean(abs((y_obs-y_pred)/y_obs))*100

# compute cross-validated R^2_pred
r2_pred  <-  cor(cbind(y_obs,y_pred))**2
r2_train <-  summary(fit_m1_trn)$r.squared
diffr2_m1 <- abs(r2_train-r2_pred)

# print difference of cross-validate R2 and R2
# diffr2_m1[1,2]


# Model: 2 : fit_inter_Paid_3
fit_int_m1_trn <- lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS +
                        log(TotalInterac) + typeV + category1, data = mydata)


# summary of fit_int_m1_trn
summary(fit_int_m1_trn)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInterac) +
##     typeV + category1, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32981 -0.26168  0.00091  0.27016  2.00912
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.620e+00  2.330e-01  24.124  < 2e-16 ***
## PTLike            -2.077e-05  1.490e-06 -13.935  < 2e-16 ***
## typeP              1.042e+00  1.172e-01   8.894  < 2e-16 ***
## typeS              2.247e+00  1.443e-01  15.568  < 2e-16 ***
## log(TotalInterac)  4.149e-01  2.291e-02  18.115  < 2e-16 ***
## typeV              1.605e+00  2.299e-01   6.979 9.85e-12 ***
## category1          4.202e-01  5.307e-02   7.918 1.67e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5196 on 483 degrees of freedom
## Multiple R-squared:  0.6146, Adjusted R-squared:  0.6098
## F-statistic: 128.4 on 6 and 483 DF,  p-value: < 2.2e-16
```

```r
# create fitted values using test.mydata
y_pred2 <- predict.glm(fit_int_m1_trn,test.mydata)
y_obs2  <- log(test.mydata[,"LPConsumer"])

# validation statistics
# RMSE of prediction error
rmse_m1_2 <-sqrt((y_obs2-y_pred2)%*%(y_obs2-y_pred2)/nrow(test.mydata))

# compute MAE
mae_m1_2 <- mean(abs(y_obs2-y_pred2))

# compute MAPE
mape_m1_2 <- mean(abs((y_obs2-y_pred2)/y_obs2))*100

# compute cross-validated R^2_pred
r2_pred2    <-  cor(cbind(y_obs2,y_pred2))**2
r2_train2   <-  summary(fit_int_m1_trn)$r.squared
diffr2_m1_2 <- abs(r2_train2-r2_pred2)

# print difference of cross-validate R2 and R2
# diffr2_m1_2[1,2]


# Model 3 : fit_full_interaction_6
fit_int_m2_trn <- lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInter
ac) +
                typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
                typeP:log(TotalInterac), data = mydata)

# summary of fit_full_interaction_4
summary(fit_int_m2_trn)
```

```
##
## Call:
## lm(formula = log(LPConsumer) ~ PTLike + typeP + typeS + log(TotalInterac) +
##      typeV + category1 + PTLike:typeP + PTLike:log(TotalInterac) +
##      typeP:log(TotalInterac), data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23186 -0.27005 -0.01573  0.26218  1.96499
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               9.930e-01  8.336e-01   1.191  0.23414
## PTLike                    3.807e-06  6.754e-06   0.564  0.57328
## typeP                     4.392e+00  5.991e-01   7.330 9.84e-13 ***
## typeS                     1.561e+00  1.632e-01   9.568  < 2e-16 ***
## log(TotalInterac)         1.218e+00  1.658e-01   7.349 8.65e-13 ***
## typeV                     7.625e-01  2.422e-01   3.148  0.00174 **
## category1                 3.893e-01  4.978e-02   7.821 3.35e-14 ***
## PTLike:typeP             -1.194e-05  4.683e-06  -2.551  0.01106 *
## PTLike:log(TotalInterac) -3.006e-06  1.255e-06  -2.395  0.01703 *
## typeP:log(TotalInterac)  -4.884e-01  6.780e-02  -7.203 2.29e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4851 on 480 degrees of freedom
## Multiple R-squared:  0.666,  Adjusted R-squared:  0.6598
## F-statistic: 106.4 on 9 and 480 DF,  p-value: < 2.2e-16
```

```r
# create fitted values using test.mydata
y_pred3 <- predict.glm(fit_int_m2_trn,test.mydata)
y_obs3  <- log(test.mydata[,"LPConsumer"])

# validation statistics
# RMSE of prediction error
rmse_m3 <- sqrt((y_obs3-y_pred3)%*%(y_obs3-y_pred3)/nrow(test.mydata))

# compute MAE
mae_m3 <- mean(abs(y_obs3-y_pred3))

# compute MAPE
mape_m3 <- mean(abs((y_obs3-y_pred3)/y_obs3))*100

# compute cross-validated R^2_pred
r2_pred3  <- cor(cbind(y_obs3,y_pred3))**2
r2_train3 <- summary(fit_int_m2_trn)$r.squared
diffr2_m3 <- abs(r2_train3-r2_pred3)

# print difference of cross-validate R2 and R2
# diffr2_m3[1,2]

# create dataframe
Model   <-  c("fit_full_3", "fit_inter_Paid_3", "fit_full_interaction_6")
RMSE    <-  c(rmse_m1, rmse_m1_2, rmse_m3)
MAE     <-  c( mae_m1, mae_m1_2, mae_m3)
MAPE    <-  c(mape_m1, mape_m1_2, mape_m3)
Diff_R2 <-  c(diffr2_m1[1,2], diffr2_m1_2[1,2], diffr2_m3[1,2])

df <- data.frame(Model, RMSE, MAE, MAPE, Diff_R2)

# print Model Info
df
```

```
##                     Model      RMSE       MAE     MAPE      Diff_R2
## 1             fit_full_3 0.4944894 0.3766650 6.094994 0.008043312
## 2       fit_inter_Paid_3 0.4944894 0.3766650 6.094994 0.008043312
## 3 fit_full_interaction_6 0.4663610 0.3629355 5.868759 0.004332479
```