

# Social Media Multi regression

Yuefei Chen

2024-04-14

## The second part: Social media data

### Model development

Running the following code, we build a multiple regression model based on rent house data. Its independent variables “Instagram\_value”, “Linkedin\_value”, “Snapchat\_value”, “Twitter\_value”, “Whatsapp\_Wechat\_value”, “Youtube\_value”, “OTT\_Netflix\_Hulu\_Prime\_video\_value”, “Reddit\_value”, “job\_interview\_calls”, “networking\_done\_with\_coffee\_chats”, “learning\_done\_in\_terms\_of\_items\_created”. The dependent variable is “felt\_the\_entire\_week”.

```
library(readr)
APP_data <- read_csv("Dataset/Social Media_cleaned.csv")
```

```
## New names:
## Rows: 23 Columns: 33
## -- Column specification
## ----- Delimiter: "," chr
## (15): ID, Instagram, Linkedin, Snapchat, Twitter, Whatsapp_Wechat, Yout... dbl
## (12): Instagram_value, Linkedin_value, Snapchat_value, Twitter_value, W... time
## (6): Hours_spent...3, Hours_spent...6, Hours_spent...9, Hours spent, H...
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * 'Hours_spent' -> 'Hours_spent...3'
## * 'Hours_spent' -> 'Hours_spent...6'
## * 'Hours_spent' -> 'Hours_spent...9'
## * 'Hours_spent' -> 'Hours_spent...15'
## * 'Hours_spent' -> 'Hours_spent...18'
## * 'Hours_spent' -> 'Hours_spent...21'
## * 'Hours_spent' -> 'Hours_spent...24'
```

```
APP_data <- APP_data[c(1:22), c(1:2, 4:5, 7:8, 10:11, 13:14, 16:17, 19:20, 22:23, 25:33)]
str(APP_data)
```

```
## tibble [22 x 25] (S3: tbl_df/tbl/data.frame)
## $ ID : chr [1:22] "masinl" "peace" "Patty" "Bunny" ...
## $ Instagram : chr [1:22] "Yes" "Yes" "Yes" "Yes" ...
## $ Instagram_value : num [1:22] 3.5 7.73 3.77 5.38 0 2.33 5.37 7 8.65 0.17
## $ Linkedin : chr [1:22] "Yes" "Yes" "Yes" "Yes" ...
## $ Linkedin_value : num [1:22] 4 5.2 7 5.32 0.58 7 4 4 10 0 ...
## $ Snapchat : chr [1:22] "Yes" "Yes" "Yes" "Yes" ...
```

```
## $ Snapchat_value : num [1:22] 1 3.68 0.53 1.3 0 0.47 0 3 3.83 0 ...
## $ Twitter : chr [1:22] "Yes" "No" "No" "No" ...
## $ Twitter_value : num [1:22] 5 0 0 0 0.67 0 0 0 0 0 ...
## $ Whatsapp_Wechat : chr [1:22] "Yes" "Yes" "Yes" "Yes" ...
## $ Whatsapp_Wechat_value : num [1:22] 1 4.18 9.83 5.3 3 12 6 10 6.15 1 ...
## $ Youtube : chr [1:22] "Yes" "Yes" "Yes" "Yes" ...
## $ Youtube_value : num [1:22] 2.5 4.25 1.85 2 3.5 7 3 2 4 3 ...
## $ OTT_Netflix_Hulu_Prime_video : chr [1:22] "Yes" "No" "Yes" "Yes" ...
## $ OTT_Netflix_Hulu_Prime_video_value : num [1:22] 14.5 0 2 2 2 3 0 3 3 0 ...
## $ Reddit : chr [1:22] "Yes" "No" "No" "No" ...
## $ Reddit_value : num [1:22] 2.5 0 0 0 1 0 0 0 0 0 ...
## $ Application_type_Social_media_OTT_Learning: chr [1:22] "OTT" "Social Media" "Social Media" "Social Media" ...
## $ job_interview_calls : num [1:22] 0 0 0 2 0 0 0 0 1 0 ...
## $ networking_done_with_coffee_chats : num [1:22] 0 1 0 0 2 0 2 0 0 0 ...
## $ learning_done_in_terms_of_items_created : num [1:22] 3 3 4 4 4 4 3 2 6 2 ...
## $ Mood_Productivity : chr [1:22] "Yes" "Yes" "Yes" "Yes" ...
## $ Tired_waking_up_in_morning : chr [1:22] "No" "No" "No" "No" ...
## $ Trouble_falling_asleep : chr [1:22] "No" "Yes" "No" "No" ...
## $ felt_the_entire_week : num [1:22] 3 3 4 4 3 5 4 4 3 2 ...
```

```
fit <- lm(felt_the_entire_week ~ Instagram_value + Linkedin_value + Snapchat_value + Twitter_value + Whatsapp_Wechat_value + Youtube_value + OTT_Netflix_Hulu_Prime_video_value + Reddit_value + job_interview_calls + networking_done_with_coffee_chats + learning_done_in_terms_of_items_created, data = APP_data)
fit
```

```
##
## Call:
## lm(formula = felt_the_entire_week ~ Instagram_value + Linkedin_value +
##     Snapchat_value + Twitter_value + Whatsapp_Wechat_value +
##     Youtube_value + OTT_Netflix_Hulu_Prime_video_value + Reddit_value +
##     job_interview_calls + networking_done_with_coffee_chats +
##     learning_done_in_terms_of_items_created, data = APP_data)
##
## Coefficients:
##                (Intercept)
##                   3.38572
##             Instagram_value
##              -0.09842
##             Linkedin_value
##               0.19780
##             Snapchat_value
##              -0.10269
##             Twitter_value
##               0.24008
##             Whatsapp_Wechat_value
##               0.10295
##             Youtube_value
##               0.02357
## OTT_Netflix_Hulu_Prime_video_value
##              -0.05841
##             Reddit_value
##              -0.11598
##             job_interview_calls
##               0.72837
##             networking_done_with_coffee_chats
##               0.04368
```

```
## learning_done_in_terms_of_items_created
## -0.28138
```

## Model Acceptance

In the summary of the model, we focus on R squared value, coefficients, and P-value of each coefficient. The R-squared value is 0.4434 and Adjust R-squared value is -0.1688. It shows there is a low proportion of variance in the dependent variable can be explained by the independent variables. Therefore, we use stepAIC to find an optimal model.

```
summary(fit)
```

```
##
## Call:
## lm(formula = felt_the_entire_week ~ Instagram_value + LinkedIn_value +
##   Snapchat_value + Twitter_value + Whatsapp_Wechat_value +
##   Youtube_value + OTT_Netflix_Hulu_Prime_video_value + Reddit_value +
##   job_interview_calls + networking_done_with_coffee_chats +
##   learning_done_in_terms_of_items_created, data = APP_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97988 -0.38761 -0.06981  0.39474  1.51971
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.38572    0.75072   4.510  0.00113 **
## Instagram_value  -0.09842    0.10388  -0.947  0.36574
## LinkedIn_value    0.19780    0.12405   1.595  0.14190
## Snapchat_value   -0.10269    0.20798  -0.494  0.63215
## Twitter_value     0.24008    0.35856   0.670  0.51829
## Whatsapp_Wechat_value 0.10295    0.08365   1.231  0.24660
## Youtube_value     0.02357    0.14998   0.157  0.87827
## OTT_Netflix_Hulu_Prime_video_value -0.05841    0.12293  -0.475  0.64488
## Reddit_value     -0.11598    0.14469  -0.802  0.44141
## job_interview_calls  0.72837    0.53742   1.355  0.20514
## networking_done_with_coffee_chats  0.04368    0.29401   0.149  0.88486
## learning_done_in_terms_of_items_created -0.28138    0.27268  -1.032  0.32642
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7985 on 10 degrees of freedom
## Multiple R-squared:  0.4434, Adjusted R-squared:  -0.1688
## F-statistic: 0.7242 on 11 and 10 DF,  p-value: 0.6985
```

```
coefficients(fit)
```

```
##              (Intercept)              Instagram_value
##              3.38572032              -0.09842050
##              LinkedIn_value              Snapchat_value
##              0.19779691              -0.10268982
##              Twitter_value              Whatsapp_Wechat_value
```

```
##              0.24008117              0.10295052
##      Youtube_value      OTT_Netflix_Hulu_Prime_video_value
##              0.02356621              -0.05840958
##      Reddit_value              job_interview_calls
##      -0.11598337              0.72837196
##      networking_done_with_coffee_chats learning_done_in_terms_of_items_created
##              0.04367604              -0.28137960
```

```
library(MASS)
step <- stepAIC(fit, direction="both")
```

```
## Start:  AIC=-3.25
## felt_the_entire_week ~ Instagram_value + Linkedin_value + Snapchat_value +
##      Twitter_value + Whatsapp_Wechat_value + Youtube_value + OTT_Netflix_Hulu_Prime_video_value +
##      Reddit_value + job_interview_calls + networking_done_with_coffee_chats +
##      learning_done_in_terms_of_items_created
##
##              Df Sum of Sq    RSS    AIC
## - networking_done_with_coffee_chats      1  0.01407  6.3896 -5.2001
## - Youtube_value                          1  0.01574  6.3913 -5.1944
## - OTT_Netflix_Hulu_Prime_video_value      1  0.14394  6.5195 -4.7575
## - Snapchat_value                         1  0.15543  6.5310 -4.7187
## - Twitter_value                         1  0.28583  6.6614 -4.2838
## - Reddit_value                          1  0.40968  6.7852 -3.8785
## - Instagram_value                       1  0.57231  6.9478 -3.3574
## <none>                                   6.3755 -3.2486
## - learning_done_in_terms_of_items_created 1  0.67890  7.0544 -3.0225
## - Whatsapp_Wechat_value                  1  0.96566  7.3412 -2.1459
## - job_interview_calls                    1  1.17108  7.5466 -1.5387
## - Linkedin_value                        1  1.62100  7.9965 -0.2647
##
## Step:  AIC=-5.2
## felt_the_entire_week ~ Instagram_value + Linkedin_value + Snapchat_value +
##      Twitter_value + Whatsapp_Wechat_value + Youtube_value + OTT_Netflix_Hulu_Prime_video_value +
##      Reddit_value + job_interview_calls + learning_done_in_terms_of_items_created
##
##              Df Sum of Sq    RSS    AIC
## - Youtube_value                          1  0.02271  6.4123 -7.1221
## - Snapchat_value                         1  0.14326  6.5329 -6.7123
## - OTT_Netflix_Hulu_Prime_video_value      1  0.19634  6.5859 -6.5343
## - Twitter_value                         1  0.31760  6.7072 -6.1329
## - Reddit_value                          1  0.40306  6.7927 -5.8544
## - Instagram_value                       1  0.56822  6.9578 -5.3259
## <none>                                   6.3896 -5.2001
## - learning_done_in_terms_of_items_created 1  0.68022  7.0698 -4.9746
## - Whatsapp_Wechat_value                  1  0.95246  7.3421 -4.1433
## - job_interview_calls                    1  1.18669  7.5763 -3.4524
## + networking_done_with_coffee_chats      1  0.01407  6.3755 -3.2486
## - Linkedin_value                        1  1.64753  8.0371 -2.1533
##
## Step:  AIC=-7.12
## felt_the_entire_week ~ Instagram_value + Linkedin_value + Snapchat_value +
##      Twitter_value + Whatsapp_Wechat_value + OTT_Netflix_Hulu_Prime_video_value +
##      Reddit_value + job_interview_calls + learning_done_in_terms_of_items_created
```

```

##
##
##      Df Sum of Sq   RSS   AIC
## - Snapchat_value      1  0.17215 6.5845 -8.5392
## - OTT_Netflix_Hulu_Prime_video_value      1  0.17363 6.5859 -8.5343
## - Twitter_value      1  0.30075 6.7131 -8.1137
## - Reddit_value      1  0.38044 6.7928 -7.8541
## - Instagram_value      1  0.56298 6.9753 -7.2707
## <none>                      6.4123 -7.1221
## - learning_done_in_terms_of_items_created      1  0.75024 7.1626 -6.6878
## - Whatsapp_Wechat_value      1  0.93212 7.3444 -6.1362
## + Youtube_value      1  0.02271 6.3896 -5.2001
## + networking_done_with_coffee_chats      1  0.02104 6.3913 -5.1944
## - job_interview_calls      1  1.32302 7.7353 -4.9953
## - LinkedIn_value      1  1.65446 8.0668 -4.0723
##
## Step:  AIC=-8.54
## felt_the_entire_week ~ Instagram_value + LinkedIn_value + Twitter_value +
##      Whatsapp_Wechat_value + OTT_Netflix_Hulu_Prime_video_value +
##      Reddit_value + job_interview_calls + learning_done_in_terms_of_items_created
##
##      Df Sum of Sq   RSS   AIC
## - Reddit_value      1  0.29735 6.8818 -9.5675
## <none>                      6.5845 -8.5392
## - learning_done_in_terms_of_items_created      1  0.66755 7.2520 -8.4148
## - OTT_Netflix_Hulu_Prime_video_value      1  0.70119 7.2857 -8.3130
## - Twitter_value      1  0.79954 7.3840 -8.0180
## + Snapchat_value      1  0.17215 6.4123 -7.1221
## + Youtube_value      1  0.05160 6.5329 -6.7123
## + networking_done_with_coffee_chats      1  0.00611 6.5784 -6.5597
## - LinkedIn_value      1  1.48267 8.0671 -6.0714
## - Instagram_value      1  1.71262 8.2971 -5.4530
## - Whatsapp_Wechat_value      1  1.79682 8.3813 -5.2309
## - job_interview_calls      1  1.80555 8.3900 -5.2080
##
## Step:  AIC=-9.57
## felt_the_entire_week ~ Instagram_value + LinkedIn_value + Twitter_value +
##      Whatsapp_Wechat_value + OTT_Netflix_Hulu_Prime_video_value +
##      job_interview_calls + learning_done_in_terms_of_items_created
##
##      Df Sum of Sq   RSS   AIC
## - learning_done_in_terms_of_items_created      1  0.42965 7.3115 -10.2352
## <none>                      6.8818 -9.5675
## - Twitter_value      1  0.76971 7.6515 -9.2350
## - OTT_Netflix_Hulu_Prime_video_value      1  0.80515 7.6870 -9.1333
## + Reddit_value      1  0.29735 6.5845 -8.5392
## - LinkedIn_value      1  1.22294 8.1048 -7.9690
## + Snapchat_value      1  0.08905 6.7928 -7.8541
## + networking_done_with_coffee_chats      1  0.01095 6.8709 -7.6026
## + Youtube_value      1  0.00626 6.8756 -7.5875
## - Instagram_value      1  1.44983 8.3317 -7.3616
## - job_interview_calls      1  1.78437 8.6662 -6.4955
## - Whatsapp_Wechat_value      1  1.91463 8.7965 -6.1673
##
## Step:  AIC=-10.24

```

```

## felt_the_entire_week ~ Instagram_value + Linkedin_value + Twitter_value +
##   Whatsapp_Wechat_value + OTT_Netflix_Hulu_Prime_video_value +
##   job_interview_calls
##
##           Df Sum of Sq   RSS   AIC
## - Twitter_value          1  0.41685 7.7283 -11.0153
## - OTT_Netflix_Hulu_Prime_video_value      1  0.44635 7.7578 -10.9315
## <none>                                7.3115 -10.2352
## - Linkedin_value          1  0.80531 8.1168  -9.9364
## + learning_done_in_terms_of_items_created  1  0.42965 6.8818  -9.5675
## - Instagram_value         1  1.05710 8.3686  -9.2643
## + networking_done_with_coffee_chats        1  0.08350 7.2280  -8.4879
## - job_interview_calls       1  1.37013 8.6816  -8.4564
## + Snapchat_value           1  0.06601 7.2455  -8.4347
## + Reddit_value             1  0.05944 7.2520  -8.4148
## + Youtube_value            1  0.05443 7.2570  -8.3996
## - Whatsapp_Wechat_value     1  1.55121 8.8627  -8.0023
##
## Step: AIC=-11.02
## felt_the_entire_week ~ Instagram_value + Linkedin_value + Whatsapp_Wechat_value +
##   OTT_Netflix_Hulu_Prime_video_value + job_interview_calls
##
##           Df Sum of Sq   RSS   AIC
## - OTT_Netflix_Hulu_Prime_video_value      1  0.08759 7.8159 -12.7674
## - Linkedin_value          1  0.71872 8.4470 -11.0590
## <none>                                7.7283 -11.0153
## - job_interview_calls       1  1.02582 8.7541 -10.2734
## + Twitter_value          1  0.41685 7.3115 -10.2352
## - Instagram_value         1  1.10750 8.8358 -10.0691
## + Snapchat_value          1  0.34127 7.3870 -10.0089
## - Whatsapp_Wechat_value     1  1.16172 8.8900  -9.9345
## + Reddit_value             1  0.12358 7.6047  -9.3700
## + Youtube_value            1  0.09128 7.6370  -9.2767
## + learning_done_in_terms_of_items_created  1  0.07678 7.6515  -9.2350
## + networking_done_with_coffee_chats        1  0.02636 7.7020  -9.0905
##
## Step: AIC=-12.77
## felt_the_entire_week ~ Instagram_value + Linkedin_value + Whatsapp_Wechat_value +
##   job_interview_calls
##
##           Df Sum of Sq   RSS   AIC
## - Linkedin_value          1  0.64719 8.4631 -13.017
## <none>                                7.8159 -12.767
## + Snapchat_value          1  0.42553 7.3904 -11.999
## - job_interview_calls       1  1.19162 9.0075 -11.646
## - Whatsapp_Wechat_value     1  1.27121 9.0871 -11.452
## + Reddit_value             1  0.17296 7.6429 -11.260
## - Instagram_value         1  1.36199 9.1779 -11.233
## + Youtube_value            1  0.09129 7.7246 -11.026
## + OTT_Netflix_Hulu_Prime_video_value      1  0.08759 7.7283 -11.015
## + Twitter_value          1  0.05809 7.7578 -10.931
## + learning_done_in_terms_of_items_created  1  0.04679 7.7691 -10.899
## + networking_done_with_coffee_chats        1  0.00387 7.8120 -10.778
##

```

```
## Step: AIC=-13.02
## felt_the_entire_week ~ Instagram_value + Whatsapp_Wechat_value +
##     job_interview_calls
##
##              Df Sum of Sq    RSS    AIC
## <none>                        8.4631 -13.017
## + LinkedIn_value             1  0.64719  7.8159 -12.767
## - Instagram_value            1  1.13255  9.5956 -12.254
## - Whatsapp_Wechat_value      1  1.54766 10.0107 -11.322
## + Snapchat_value             1  0.11643  8.3467 -11.322
## + Reddit_value               1  0.11323  8.3499 -11.314
## + Twitter_value              1  0.09719  8.3659 -11.271
## + learning_done_in_terms_of_items_created 1  0.08612  8.3770 -11.242
## + OTT_Netflix_Hulu_Prime_video_value      1  0.01606  8.4470 -11.059
## + Youtube_value               1  0.00496  8.4581 -11.030
## + networking_done_with_coffee_chats       1  0.00080  8.4623 -11.019
## - job_interview_calls          1  2.15468 10.6178 -10.027

fit2 <- lm(felt_the_entire_week ~ Instagram_value + Whatsapp_Wechat_value +
  job_interview_calls, data = APP_data)
```

## Residual Analysis

QQ plot is used in these residual analysis. We can conclude that most of residual points are located in a straight line. It satisfies normal distribution.

```
confint(fit2,level=0.95)
```

```
##              2.5 %    97.5 %
## (Intercept)      2.54631223 3.8923436
## Instagram_value -0.19785737 0.0297300
## Whatsapp_Wechat_value -0.01302441 0.1779132
## job_interview_calls  0.01182008 1.2592082
```

```
fitted(fit2)
```

```
##      1      2      3      4      5      6      7      8
## 3.007549 2.914133 3.712836 4.475049 3.466661 4.012792 3.262572 3.455326
##      9     10     11     12     13     14     15     16
## 3.634724 3.287481 2.916761 3.241967 3.736560 2.996906 3.364320 3.678618
##     17     18     19     20     21     22
## 3.496136 3.220683 3.478091 3.222213 3.882471 3.536151
```

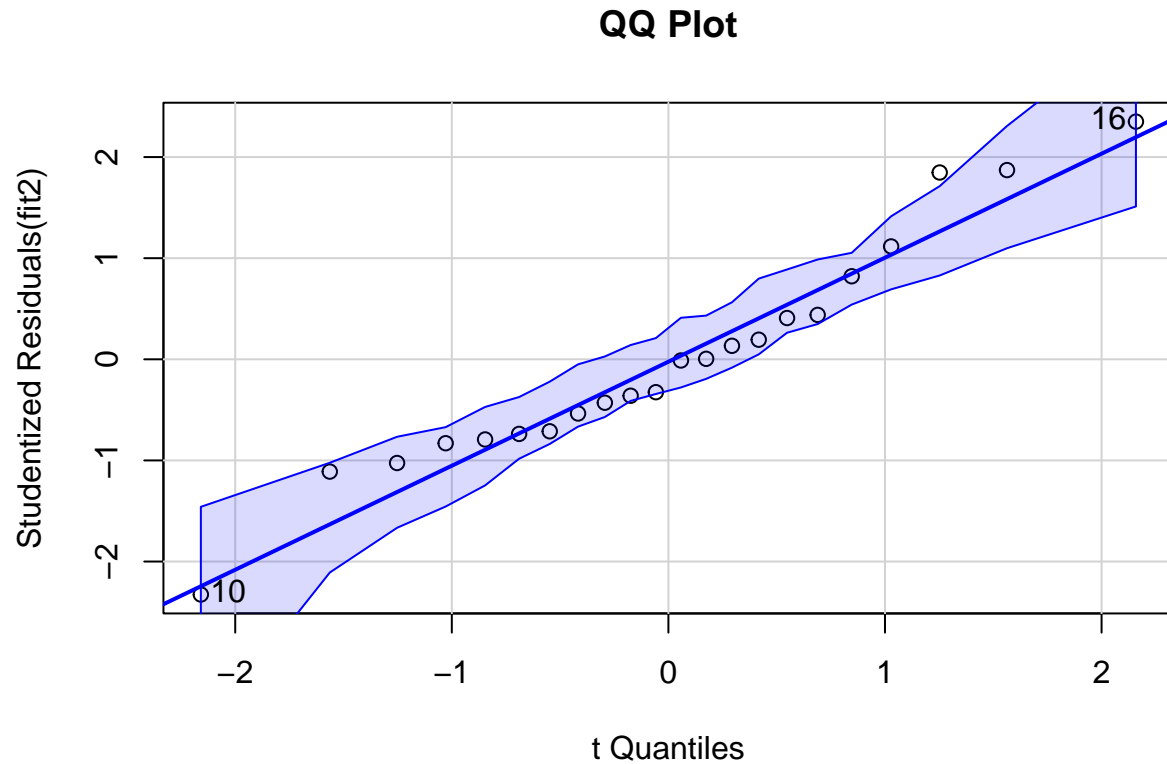
```
residuals(fit2)
```

```
##      1      2      3      4      5      6
## -0.007549419 0.085866852 0.287163972 -0.475048787 -0.466661048 0.987207984
##      7      8      9     10     11     12
## 0.737427807 0.544674120 -0.634724119 -1.287481479 1.083239357 -0.241966718
##     13     14     15     16     17     18
## 0.263439858 0.003094405 -0.364320230 1.321381834 -0.496136217 -0.220683282
##     19     20     21     22
## -0.478090655 -0.222212542 0.117529247 -0.536150941
```

```
library(car)
```

```
## Loading required package: carData
```

```
qqPlot(fit2, main="QQ Plot")
```



```
## [1] 10 16
```

### Prediction

We set a data point with Instagram\_value = 5, Whatsapp\_Wechat\_value = 5 and, job\_interview\_calls = 0, then the feeling score of the entire week we predict is approximate to 3

```
predict.lm(fit2, data.frame(Instagram_value = 5, Whatsapp_Wechat_value = 5, job_interview_calls = 0))
```

```
##          1  
## 3.211231
```

### Model Accuracy

The accuracy is based on summary of the model and we also calculate the MSE and RMSE for the model. The MSE is 0.3846858 and RMSE is 0.6202304.



```
summary(fit2)
```

```
##
## Call:
## lm(formula = felt_the_entire_week ~ Instagram_value + Whatsapp_Wechat_value +
##     job_interview_calls, data = APP_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2875 -0.4729 -0.1141  0.2812  1.3214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.21933    0.32034   10.050 8.28e-09 ***
## Instagram_value  -0.08406    0.05416   -1.552  0.1381
## Whatsapp_Wechat_value 0.08244    0.04544    1.814  0.0863 .
## job_interview_calls  0.63551    0.29687    2.141  0.0462 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6857 on 18 degrees of freedom
## Multiple R-squared:  0.2612, Adjusted R-squared:  0.138
## F-statistic: 2.121 on 3 and 18 DF,  p-value: 0.1332
```

```
predictions <- predict(fit2, APP_data)
mse <- mean((APP_data$felt_the_entire_week - predictions)^2)
rmse <- sqrt(mse)
cat("MSE: ", mse, "\n")
```

```
## MSE:  0.3846858
```

```
cat("RMSE: ", rmse, "\n")
```

```
## RMSE:  0.6202304
```