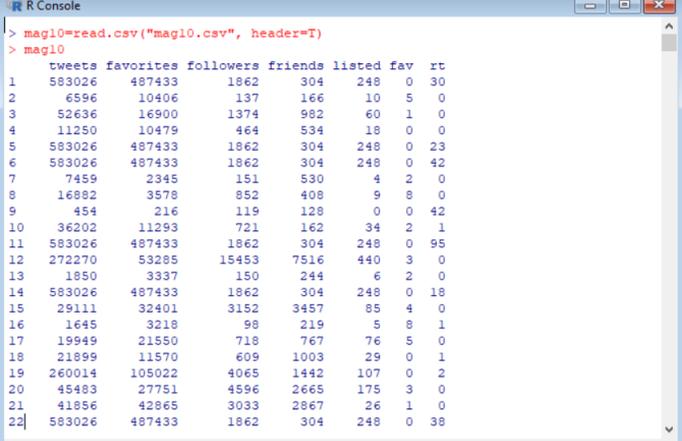


R-Programming Fundamentals for Business Students — Multiple Linear Regressions

Nick V. Flor, University of New Mexico (nickflor@unm.edu)

Assumptions. This tutorial assumes (1) that you had an Excel worksheet with the following columns: *tweets*, *favorites*, *followers*, *friends*, *listed*, *fav*, *rt*; (2) that you saved that file as *mag10.csv* and, finally, (3) that you ran R and did a *File > Change dir...* to the folder containing *mag10.csv*

ACTION	REACTION
IMPORTING CSV DATASET INTO R	
<ul style="list-style-type: none"> Type <code>mag10=read.csv("mag10.csv", header=T)</code> then press Enter <p><u>Explanation:</u> This reads the csv file into the variable <code>mag10</code>. Remember, this particular dataset is assumed to have following columns: <i>tweets</i>, <i>favorites</i>, <i>followers</i>, <i>friends</i>, <i>listed</i>, <i>fav</i>, <i>rt</i>.</p> <p>We will NOT use <i>fav</i> and <i>rt</i> in our model.</p>	 <pre>R Console > mag10=read.csv("mag10.csv", header=T) > </pre>
<ul style="list-style-type: none"> Type <code>mag10</code> (or whatever you named your variable) then press Enter to see your data <p><u>Explanation:</u> Typing any variable allows you to see its content.</p> <p><u>Note:</u> I scrolled up some so you could see the column headings</p>	 <pre>R Console > mag10=read.csv("mag10.csv", header=T) > mag10 tweets favorites followers friends listed fav rt 1 583026 487433 1862 304 248 0 30 2 6596 10406 137 166 10 5 0 3 52636 16900 1374 982 60 1 0 4 11250 10479 464 534 18 0 0 5 583026 487433 1862 304 248 0 23 6 583026 487433 1862 304 248 0 42 7 7459 2345 151 530 4 2 0 8 16882 3578 852 408 9 8 0 9 454 216 119 128 0 0 42 10 36202 11293 721 162 34 2 1 11 583026 487433 1862 304 248 0 95 12 272270 53285 15453 7516 440 3 0 13 1850 3337 150 244 6 2 0 14 583026 487433 1862 304 248 0 18 15 29111 32401 3152 3457 85 4 0 16 1645 3218 98 219 5 8 1 17 19949 21550 718 767 76 5 0 18 21899 11570 609 1003 29 0 1 19 260014 105022 4065 1442 107 0 2 20 45483 27751 4596 2665 175 3 0 21 41856 42865 3033 2867 26 1 0 22 583026 487433 1862 304 248 0 38</pre>

(continued on next page)

DO THE REGRESSION (MODEL THE DATA SET)

- Type:
`model=lm(followers~tweets+favorites+friends+listed,mag10)`

Explanation:

Suppose you were interested in how to gain more followers on Twitter, and you suspected it had to do with the number of times you tweeted & favorited, as well as the number of times you were friended and listed by others.

This creates a linear model, where followers is a linear combination of the tweets, tweets, favorites, friends, and the number of times listed.

```
R Console
1877 260014 105022 4065 1442 107 2 0
1878 193298 12629 1345 2027 37 0 43
1879 4628 12852 504 1182 18 0 0
1880 41917 14080 1117 1771 51 0 43
1881 14737 2140 369 257 7 2 0
1882 76709 42361 948 2369 52 0 0
1883 1850 3337 150 244 6 2 0
1884 389 816 47 269 1 4 1
1885 161036 7039 4693 674 86 0 0
1886 4940 12000 278 1402 7 0 0
1887 51758 54454 8420 7034 423 0 3
1888 11250 10479 464 534 18 0 0
1889 18139 30872 395 290 21 1 0
1890 4306 1098 54 452 2 0 0
1891 12815 3436 240 391 18 0 0
1892 260353 34675 5167 5601 188 1 0
1893 8637 3986 113 69 2 2 0
1894 7775 8650 3989 348 198 9 1
1895 23903 33319 205 666 17 0 43
1896 6304 4566 177 469 4 2 0
1897 16882 3578 852 408 9 13 3
1898 2511 7772 343 410 2 0 43
1899 147277 824 1243 1906 117 0 0
> model=lm(followers~tweets+favorites+friends+listed,mag10)
> |
<
```

ANALYSE THE REGRESSION MODEL

- Type: `model1`

Explanation: This prints out the coefficients and the intercept.

Equation:

Followers = $0.06702 \cdot \text{tweets} - 0.07931 \cdot \text{favorites} - 1.61977 \cdot \text{friends} + 58.63343 \cdot \text{listed}$

Implication: The best strategy for gaining followers is getting listed by others.

```
R Console
1887 51758 54454 8420 7034 423 0 3
1888 11250 10479 464 534 18 0 0
1889 18139 30872 395 290 21 1 0
1890 4306 1098 54 452 2 0 0
1891 12815 3436 240 391 18 0 0
1892 260353 34675 5167 5601 188 1 0
1893 8637 3986 113 69 2 2 0
1894 7775 8650 3989 348 198 9 1
1895 23903 33319 205 666 17 0 43
1896 6304 4566 177 469 4 2 0
1897 16882 3578 852 408 9 13 3
1898 2511 7772 343 410 2 0 43
1899 147277 824 1243 1906 117 0 0
> model=lm(followers~tweets+favorites+friends+listed,mag10)
> model
Call:
lm(formula = followers ~ tweets + favorites + friends + listed,
    data = mag10)
Coefficients:
(Intercept)      tweets      favorites      friends      listed
 539.70555      0.06702     -0.07931     -1.61977     58.63343
> |
<
```

- Type: `summary(model1)`

Explanation: This gives additional information about the regression, the most important of which is the r-squared, which in our example is 0.4933. That value means the model only accounts for 49.33% of the variance in the data, i.e., 50.67% is explained by other variables.

Nevertheless, this model gives useful heuristics to try in order to get more followers. You can also do what-if analyses.

```
R Console
> summary(model)
Call:
lm(formula = followers ~ tweets + favorites + friends + listed,
    data = mag10)
Residuals:
    Min       1Q   Median       3Q      Max
-68110  -1240   -373     790   64891
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  539.70555    282.036408     1.914  0.0558 .
tweets       0.067022    0.003018    22.210 <2e-16 ***
favorites    -0.079305    0.005648   -14.041 <2e-16 ***
friends     -1.619766    0.127553   -12.699 <2e-16 ***
listed       58.633432    2.449125    23.941 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

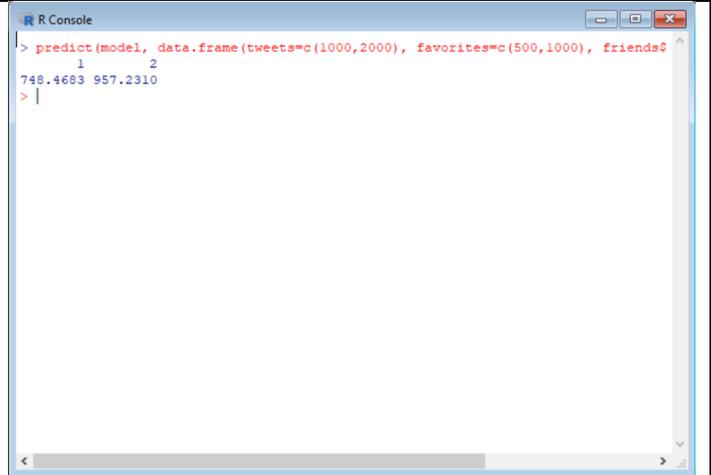
Residual standard error: 9914 on 1894 degrees of freedom
Multiple R-squared:  0.4933,    Adjusted R-squared:  0.4922
F-statistic: 460.9 on 4 and 1894 DF,  p-value: < 2.2e-16
> |
<
```

PREDICTING NEW VALUES (WHAT-IF ANALYSES)

Supposed I want to know how many followers I would have if I tweeted a total of 1000 vs 2000 times, favorited 500 vs 1000 other tweets, was friended 250 times vs 500, and was put on 10 lists vs 20 lists. Use the predict function:

- `predict(model, data.frame(tweets=c(1000,2000), favorites=c(500,1000), friends=c(250,500), listed=c(10,20)))`

Explanation: The prediction function takes a model as input and a data.frame with new values. In this case it predicts I'd have 748 vs 957 followers.



```
R Console
> predict(model, data.frame(tweets=c(1000,2000), favorites=c(500,1000), friends=c(250,500), listed=c(10,20)))
      1      2
748.4683 957.2310
> |
```

Very easy, but remember to check the r-squared to determine the quality of your model.