

R Programming Fundamentals for Business Students— Multiple Regressions

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

ACTION

- FIRST, do R-Programming Tutorials 5 & 6, where you create the `foodsurvey.csv` table in Excel, import this table into R inside the variable `m`, and experiment with matrix indexing (not shown).
- Type `m` just to be sure you have the right data.

Now suppose you have a hunch that Blakez' scores depend on what users think of WiseGuys and SubYay, i.e.,

$$\text{Blakez} = b_2 * \text{WiseGuys} + b_3 * \text{SubYay} + Y_{\text{intercept}}$$

REACTION

```
R Console
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> m=read.csv("foodsurvey.csv",header=TRUE,row.names=1)
> m
      Blakez WiseGuys SubYay
Fred      4         5      1
Wilma    1         3      5
Pebbles  4         4      2
Barney   5         4      1
Betty    1         3      5
BamBam   5         4      2
> |
```

FINDING THE BEST FIT LINE

- Type `bestline=lm(m[,1]~m[,2]+m[,3])`
- Type `bestline`

Explanation: `m[,1]~m[,2]+m[,3]` is the formula stating column1 is a linear function of columns 2 and 3. We assign the results to the variable `bestline` (you could have called it something else).

The actual formula is:

$$\text{Blakez} = -1.222 * \text{WiseGuys} + -1.389 * \text{SubYay} + 11.722$$

```
R Console
> m
      Blakez WiseGuys SubYay
Fred      4         5      1
Wilma    1         3      5
Pebbles  4         4      2
Barney   5         4      1
Betty    1         3      5
BamBam   5         4      2
> bestline=lm(m[,1]~m[,2]+m[,3])
> bestline

Call:
lm(formula = m[, 1] ~ m[, 2] + m[, 3])

Coefficients:
(Intercept)      m[, 2]      m[, 3]
    11.722      -1.222     -1.389
> |
```

CHECKING THE R-SQUARED

- Type `summary(bestline)`

Explanation. Like Excel's linear regression tool, R's `summary` provides extra information about the goodness-of-fit for the line. In this case the R-squared is .9327 (remember: 1 is perfect, 0 is bad), so this line is a good fit.

```
R Console
(Intercept)      m[, 2]      m[, 3]
    11.722      -1.222     -1.389

> summary(bestline)

Call:
lm(formula = m[, 1] ~ m[, 2] + m[, 3])

Residuals:
    1      2      3      4      5      6
-0.22222 -0.11111 -0.05556 -0.44444 -0.11111  0.94444

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.7222   4.1876   2.799  0.0679 .
m[, 2]       -1.2222   0.8654  -1.412  0.2527
m[, 3]       -1.3889   0.3499  -3.969  0.0286 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6236 on 3 degrees of freedom
Multiple R-squared:  0.9327,    Adjusted R-squared:  0.8878
F-statistic: 20.79 on 2 and 3 DF,  p-value: 0.01746

> |
```