

Applied Regression With R

Bruce Campbell

July 17, 2017

Tue Aug 22 11:47:13 2017

Chapter 2 examples

```
if (!require(faraway)) {  
  install.packages("faraway")  
  library(faraway)  
}
```

```
## Loading required package: faraway
```

```
if (!require(HistData)) {  
  install.packages("HistData")  
  library(HistData)  
}
```

```
## Loading required package: HistData
```

```
library(faraway)  
data(gala, package = "faraway")  
head(gala[, -2])
```

```
##           Species Area Elevation Nearest Scrutz Adjacent  
## Baltra          58 25.09      346      0.6   0.6      1.84  
## Bartolome       31  1.24      109      0.6  26.3     572.33  
## Caldwell        3  0.21      114      2.8  58.7       0.78  
## Champion       25  0.10       46      1.9  47.4       0.18  
## Coamano         2  0.05       77      1.9   1.9     903.82  
## Daphne.Major    18  0.34      119      8.0   8.0       1.84
```

```
lmod <- lm(Species ~ Area + Elevation + Nearest + Scrutz + Adjacent, data = gala)  
summary(lmod)
```

```
##  
## Call:  
## lm(formula = Species ~ Area + Elevation + Nearest + Scrutz + Adjacent,  
##     data = gala)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -111.679  -34.898   -7.862   33.460  182.584   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  7.068221  19.154198   0.369 0.715351      
## Area        -0.023938   0.022422  -1.068 0.296318      
## Elevation    0.319465   0.053663   5.953 3.82e-06 ***
```

```
## Nearest      0.009144    1.054136    0.009 0.993151
## Scruz       -0.240524    0.215402   -1.117 0.275208
## Adjacent    -0.074805    0.017700   -4.226 0.000297 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.98 on 24 degrees of freedom
## Multiple R-squared:  0.7658, Adjusted R-squared:  0.7171
## F-statistic: 15.7 on 5 and 24 DF,  p-value: 6.838e-07
```

```
require(faraway)
summary(lmod)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.068221  19.154198  0.3690 0.7153508
## Area        -0.023938   0.022422 -1.0676 0.2963180
## Elevation    0.319465   0.053663  5.9532 3.823e-06
## Nearest      0.009144    1.054136  0.0087 0.9931506
## Scruz       -0.240524    0.215402 -1.1166 0.2752082
## Adjacent    -0.074805    0.017700 -4.2262 0.0002971
##
```

```
## n = 30, p = 6, Residual SE = 60.97519, R-Squared = 0.77
```

```
x <- model.matrix(~Area + Elevation + Nearest + Scruz + Adjacent, gala)
y <- gala$Species
xtxi <- solve(t(x) %*% x)
xtxi %*% t(x) %*% y
```

```
##              [,1]
## (Intercept)  7.068220709
## Area        -0.023938338
## Elevation    0.319464761
## Nearest      0.009143961
## Scruz       -0.240524230
## Adjacent    -0.074804832
```

```
solve(crossprod(x, x), crossprod(x, y))
```

```
##              [,1]
## (Intercept)  7.068220709
## Area        -0.023938338
## Elevation    0.319464761
## Nearest      0.009143961
## Scruz       -0.240524230
## Adjacent    -0.074804832
```

```
names(lmod)
```

```
## [1] "coefficients" "residuals"      "effects"      "rank"
## [5] "fitted.values" "assign"          "qr"           "df.residual"
## [9] "xlevels"       "call"           "terms"        "model"
```

```
lmodsum <- summary(lmod)
```

```
names(lmodsum)
```

```
## [1] "call"          "terms"          "residuals"      "coefficients"
## [5] "aliases"       "sigma"          "df"             "r.squared"
## [9] "adj.r.squared" "fstatistic"     "cov.unscaled"
```

```

sqrt(deviance(lmod)/df.residual(lmod))

## [1] 60.97519
lmodsum$sigma

## [1] 60.97519
xtxi <- lmodsum$cov.unscaled
sqrt(diag(xtxi)) * 60.975

## (Intercept)      Area      Elevation      Nearest      Scruz      Adjacent
## 19.15413865  0.02242228  0.05366264  1.05413269  0.21540158  0.01770013
lmodsum$coef[, 2]

## (Intercept)      Area      Elevation      Nearest      Scruz      Adjacent
## 19.15419782  0.02242235  0.05366280  1.05413595  0.21540225  0.01770019
qrx <- qr(x)
dim(qr.Q(qrx))

## [1] 30  6
(f <- t(qr.Q(qrx)) %*% y)

##           [,1]
## [1,] -466.842193
## [2,]  381.405574
## [3,]  256.250473
## [4,]   5.407646
## [5,] -119.498340
## [6,]  257.694369
backsolve(qr.R(qrx), f)

##           [,1]
## [1,]  7.068220709
## [2,] -0.023938338
## [3,]  0.319464761
## [4,]  0.009143961
## [5,] -0.240524230
## [6,] -0.074804832
gala$Adiff <- gala$Area - gala$Adjacent
lmod <- lm(Species ~ Area + Elevation + Nearest + Scruz + Adjacent + Adiff,
  gala)
sumary(lmod)

##
## Coefficients: (1 not defined because of singularities)
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.068221  19.154198  0.3690 0.7153508
## Area        -0.023938   0.022422 -1.0676 0.2963180
## Elevation     0.319465   0.053663  5.9532 3.823e-06
## Nearest       0.009144   1.054136  0.0087 0.9931506
## Scruz        -0.240524   0.215402 -1.1166 0.2752082
## Adjacent     -0.074805   0.017700 -4.2262 0.0002971
##

```

```
## n = 30, p = 6, Residual SE = 60.97519, R-Squared = 0.77
```

```
set.seed(123)
Adiffe <- gala$Adiff + 0.001 * (runif(30) - 0.5)
lmod <- lm(Species ~ Area + Elevation + Nearest + Scrutz + Adjacent + Adiffe,
          gala)
summary(lmod)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.2964e+00  1.9434e+01  0.1696   0.8668
## Area        -4.5123e+04  4.2583e+04 -1.0596   0.3003
## Elevation    3.1302e-01  5.3870e-02  5.8107 6.398e-06
## Nearest      3.8273e-01  1.1090e+00  0.3451   0.7331
## Scrutz       -2.6199e-01  2.1581e-01 -1.2140   0.2371
## Adjacent     4.5123e+04  4.2583e+04  1.0596   0.3003
## Adiffe       4.5123e+04  4.2583e+04  1.0596   0.3003
##
```

```
## n = 30, p = 7, Residual SE = 60.81975, R-Squared = 0.78
```

```
data(odor, package = "faraway")
odor
```

```
##   odor temp gas pack
## 1   66  -1  -1   0
## 2   39   1  -1   0
## 3   43  -1   1   0
## 4   49   1   1   0
## 5   58  -1   0  -1
## 6   17   1   0  -1
## 7   -5  -1   0   1
## 8  -40   1   0   1
## 9   65   0  -1  -1
## 10   7   0   1  -1
## 11  43   0  -1   1
## 12 -22   0   1   1
## 13 -31   0   0   0
## 14 -35   0   0   0
## 15 -26   0   0   0
```

```
cov(odor[, -1])
```

```
##           temp           gas           pack
## temp 0.5714286 0.0000000 0.0000000
## gas  0.0000000 0.5714286 0.0000000
## pack 0.0000000 0.0000000 0.5714286
```

```
lmod <- lm(odor ~ temp + gas + pack, data = odor)
summary(lmod, cor = T)
```

```
##
## Call:
## lm(formula = odor ~ temp + gas + pack, data = odor)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50.200 -17.138   1.175  20.300  62.925
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.200      9.298   1.635   0.130
## temp        -12.125     12.732  -0.952   0.361
## gas         -17.000     12.732  -1.335   0.209
## pack        -21.375     12.732  -1.679   0.121
##
## Residual standard error: 36.01 on 11 degrees of freedom
## Multiple R-squared:  0.3337, Adjusted R-squared:  0.1519
## F-statistic: 1.836 on 3 and 11 DF,  p-value: 0.1989
##
## Correlation of Coefficients:
##      (Intercept) temp gas
## temp  0.00
## gas   0.00      0.00
## pack  0.00      0.00 0.00

lmod <- lm(odor ~ gas + pack, odor)
summary(lmod)

##
## Call:
## lm(formula = odor ~ gas + pack, data = odor)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50.200 -26.700   1.175  26.800  50.800
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.200      9.262   1.641   0.127
## gas         -17.000     12.683  -1.340   0.205
## pack        -21.375     12.683  -1.685   0.118
##
## Residual standard error: 35.87 on 12 degrees of freedom
## Multiple R-squared:  0.2787, Adjusted R-squared:  0.1585
## F-statistic: 2.319 on 2 and 12 DF,  p-value: 0.1408

x <- 1:20
y <- x + rnorm(20)
```