

Least Squares Estimation with R: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

```
> trees[1:4,] # First 4 rows, all columns
  Girth Height Volume
1   8.3     70   10.3
2   8.6     65   10.3
3   8.8     63   10.2
4  10.5     72   16.4
> n = dim(trees)[1]; n
[1] 31
> attach(trees) # Makes variable names available
> int = numeric(n)+1 # Vector of ones, length n
> X = cbind(int, Girth, Height); Y = Volume
> X
      int Girth Height
[1,]   1   8.3     70
[2,]   1   8.6     65
[3,]   1   8.8     63
[4,]   1  10.5     72
[5,]   1  10.7     81
[6,]   1  10.8     83
[7,]   1  11.0     66
[8,]   1  11.0     75
[9,]   1  11.1     80
[10,]  1  11.2     75
[11,]  1  11.3     79
[12,]  1  11.4     76
[13,]  1  11.4     76
[14,]  1  11.7     69
[15,]  1  12.0     75
[16,]  1  12.9     74
[17,]  1  12.9     85
[18,]  1  13.3     86
[19,]  1  13.7     71
[20,]  1  13.8     64
[21,]  1  14.0     78
[22,]  1  14.2     80
[23,]  1  14.5     74
[24,]  1  16.0     72
[25,]  1  16.3     77
[26,]  1  17.3     81
```

```
[27,] 1 17.5 82
[28,] 1 17.9 80
[29,] 1 18.0 80
[30,] 1 18.0 80
[31,] 1 20.6 87
```

```
> XpX = t(X) %*% X; XpX
```

```
      int    Girth  Height
int    31.0   410.70 2356.0
Girth  410.7  5736.55 31524.7
Height 2356.0 31524.70 180274.0
```

```
> XpY = t(X) %*% Y; XpY
```

```
      [,1]
int    935.30
Girth 13887.86
Height 72962.60
```

```
> betahat = solve(XpX) %*% XpY
> betahat
```

```
      [,1]
int    -57.9876589
Girth   4.7081605
Height  0.3392512
```

```
> # Predict volume for a tree 12 inches in diameter, 80 feet tall
> betahat[1] + betahat[2]*12 + betahat[3]*80
```

```
[1] 25.65037
```

```
> # R does not actually calculate X'X-inverse. It solves the system of
> # linear equations numerically, like this:
> solve(XpX,XpY)
```

```
      [,1]
int    -57.9876589
Girth   4.7081605
Height  0.3392512
```

```
> # It's better (not just more convenient) to let R do the calculation
> treefit = lm(Volume ~ Girth+Height) # Produces a linked list
> treefit$coefficients
```

```
(Intercept)      Girth      Height
-57.9876589    4.7081605    0.3392512
```

```
> sum(Volume) # Sum of Y
[1] 935.3
> sum(treefit$fit) # Sum of Y-hat
[1] 935.3
> sum(treefit$residuals) # Sum of epsilon-hat
[1] 4.662937e-15
```

```
> # Try a model with no intercept
> treefit2 = lm(Volume ~ 0+Girth+Height)
```

```
> treefit2$coefficients
      Girth      Height
5.0440083 -0.4773192
> sum(treefit2$residuals)
[1] -11.71008
```

```
> # Prediction made easy
> newdata = data.frame(Girth=12,Height=80) # Creating a data frame
> newdata
  Girth Height
1    12     80
> predict(treefit2,newdata) # With an intercept, got 25.65037
```

```
      1
22.34256
```

```
> # WHICH PREDICTION DO YOU LIKE MORE?
```

This document is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License: http://creativecommons.org/licenses/by-sa/3.0/deed.en_US. Use any part of it as you like and share the result freely.