Least Squares

Jeremy Teitelbaum

Example Application: Least Squares Regression

	TV	Radio	Newspaper	Sales
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9
6	8.7	48.9	75.0	7.2
7	57.5	32.8	23.5	11.8
8	120.2	19.6	11.6	13.2
9	8.6	2.1	1.0	4.8
10	199.8	2.6	21.2	10.6

10 rows of a 200 row table. Predict sales using a linear model:

$$sales = A(TV) + B(Radio) + C(Newspaper) + U$$

Least squares

Create a 200×4 matrix D by removing the sales column and adding a column of ones. Let Y be the column of sales data.

If sales *were* a perfect linear function, then we would have

$$Y = D \begin{bmatrix} U \\ A \\ B \\ D \end{bmatrix}$$

But this isn't true. So instead we try to minimize $\|Y - DM\|$ where

$$M = \begin{bmatrix} U \\ A \\ B \\ D \end{bmatrix}$$

is a variable.

Geometry

The four columns of the data matrix span a four dimensional space W in $\mathbf{R}^{200}$. We try to find a point in this span that is as close to Y as possible.

We know this is given by orthogonal projection. That is, we want to find $\hat{Y} = DM$ so that $\hat{Y} - Y$ is perpendicular to the vectors spanning W. This boils down to the requirement that

$$D^T(\hat{Y} - Y) = 0$$

Least Squares solution

This gives the equation

$$D^T DM - D^T Y = 0$$

or

$$M = (D^T D)^{-1} D^T Y$$

The predicted values are

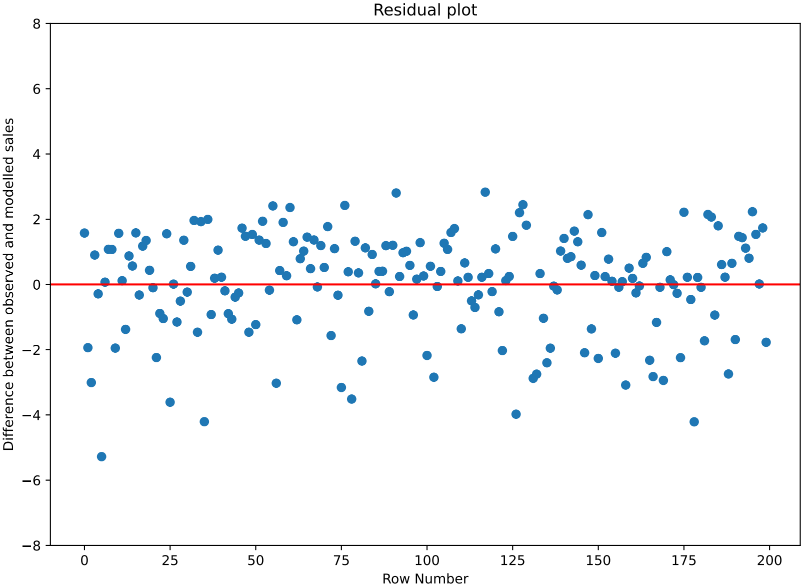
$$\hat{Y} = DM = D(D^T D)^{-1} D^T Y$$

Computer Calculation

This shows the true and computed values for sales.

```
U=2.9389 A=0.0458 B=0.1885 C = -0.0010
```

Residuals



Residual plot
Difference between observed and modelled sales
Row Number
8
6
4
2
0
−2
−4
−6
−8
0
25
50
75
100
125
150
175
200