Design Matrix Approximation

There is another matrix that corresponds to this, called design matrix.

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^k \\ 1 & x_2 & x_2^2 & x_2^k \\ 1 & x_3 & x_3^2 & x_3^k \\ . & & & \\ . & & & \\ . & & & \\ 1 & x_n & x_n^2 & x_n^k \end{bmatrix}$$

It is easy to show that $A^T A$ is just the coefficient matrix.

This means that we can write normal equations in matrix form as

$$A^T A\beta = A^T y$$

In that case we can compute the least-squares estimator explicitly by inverting the k*k matrix $A^T A$

$$\hat{\beta}_{LeastSquares} = (A^T A)^{-1} A^T Y$$

Example (Solved Before)

i	X_i	Y_i	X_i^2	$X_i Y_i$
1	20.5	765	420.25	15682.5
2	32.7	826	1069.29	27010.2
3	51.0	873	2601.00	44523.0
4	73.2	942	5358.24	68954.4
5	95.7	1032	9158.49	98762.4
	273.1	4438	18607.3	254933

Normal equations

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i + \hat{\beta}_0 n = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \hat{\beta}_0 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

$$273.1\hat{\beta}_1 + 5\hat{\beta}_0 = 4438$$

$$18607.3\hat{\beta}_1 + 273.1\hat{\beta}_0 = 254933$$

From these we find $\hat{\beta}_1$ = 3.395, $\hat{\beta}_0$ = 702.2, and hence we write estimated linear equation as

$$\hat{y} = 702.2 + 3.395x$$

(Estimated Regression Line)

Solution Using Design Matrix

```
MTB > print c1-c3
```

Data Display

```
Row  C1    C2    C3
  1   1  20,5   765
  2   1  32,7   826
  3   1  51,0   873
  4   1  73,2   942
  5   1  95,7  1032
```

```
MTB >
MTB > copy c1 c2 m1
```

```
MTB > print m1
```

Data Display

```
 Matrix M1
```

Design Matrix

```
1  20,5
1  32,7
1  51,0
1  73,2
1  95,7
```

```
MTB > tran m1 m2
```

```
MTB > print m2
```

Transpose of the Design Matrix

Data Display

```
 Matrix M2
```

```
 1,0   1,0   1   1,0   1,0
20,5  32,7  51  73,2  95,7
```

```
MTB > mult m2 m1 m3

MTB > print m3
```

Data Display

```
 Matrix M3

  5,0    273,1
273,1  18607,3
```

```
MTB > copy c3 m4

MTB > print m4
```

Data Display

```
 Matrix M4

 765
 826
 873
 942
1032
```

```
MTB > mult m2 m4 m5

MTB > print m5
```

Data Display

```
 Matrix M5

  4438
254933
```

```
MTB > inve m3 m6

MTB > print m6
```

Data Display

```
 Matrix M6

 1,00837  -0,0148000
-0,01480   0,0002710


MTB > mult m6 m5 m7

MTB > print m7
```

Data Display

```
 Matrix M7

702,172
  3,395
```

$$\hat{y} = 702.2 + 3.395x$$

(Estimated Regression Line)

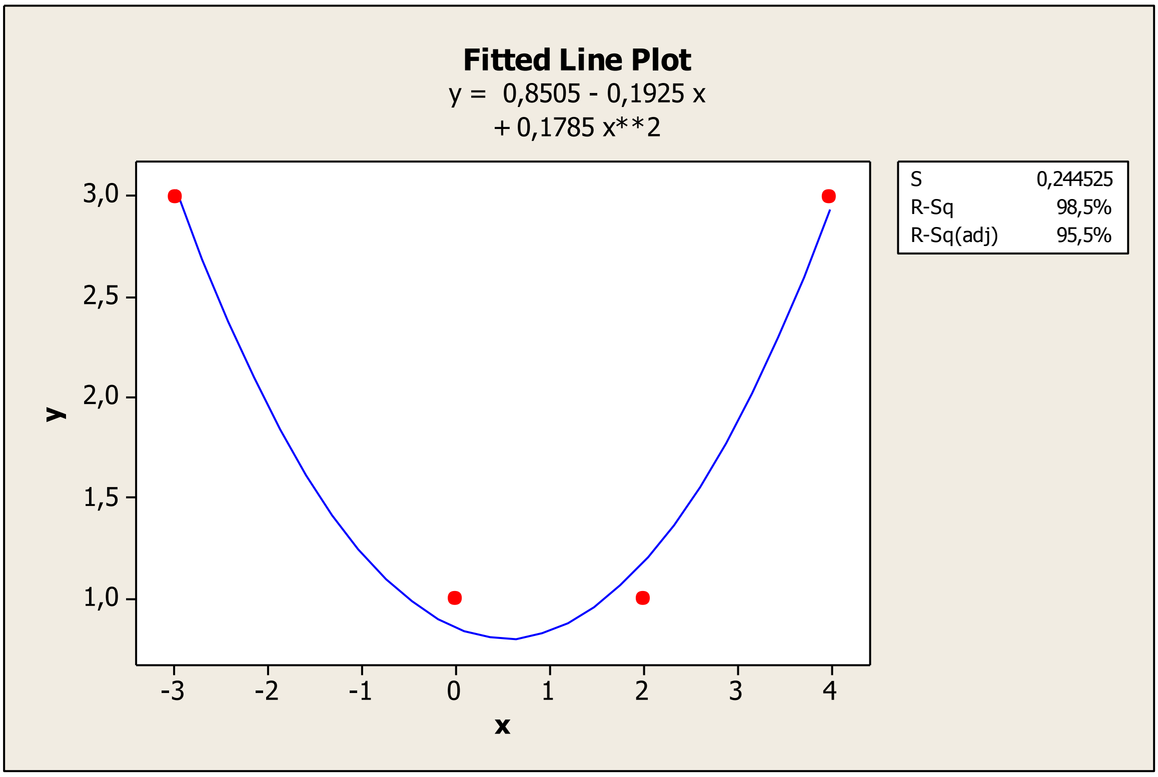
2nd degree-Polynomial Regression Model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 e$$

$$Sum = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)^2$$

Example: Find the least squares parabola for the following data

x	y
-3	3
0	1
2	1
4	3



Design Matrix

Matrix M1

```
1  -3   9
1   0   0
1   2   4
1   4  16
```

Transpose of the Design Matrix

```
MTB > trans m1 m2
MTB > print m2
```

Data Display

```
 Matrix M2

 1  1  1   1
-3  0  2   4
 9  0  4  16
```

The matrix of $A^T A$

```
MTB > mult m2 m1 m3
MTB > print m3
```

Data Display

```
 Matrix M3

 4   3   29
 3  29   45
29  45  353
```

The matrix of $(A^T A)^{-1}$

```
MTB > inverse m3 m4
MTB > print m4
```

Data Display

```
 Matrix M4

 0,626297   0,0187614  -0,0538438
 0,018761   0,0435479  -0,0070927
-0,053844  -0,0070927   0,0081605
```

The matrix of $A^T Y$

```
MTB > print m5
```

Data Display

```
 Matrix M5
3
1
1
3
MTB > mult m2 m5 m6
MTB > print m6
```

Data Display

```
 Matrix M6
 8
 5
79
```

The matrix of $\hat{\beta}_{LeastSquares} = (A^T A)^{-1} A^T Y$

```
MTB > mult m4 m6 m7
MTB > print m7
```

Data Display

```
 Matrix M7

 0,850519
-0,192495
 0,178462
```

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$$

$$\hat{y} = 0.851 - 0.192x + 0.178x^2$$

Exercise:

Convert Operations to MATLAB

What Degree of Polynomial should be used?

In general case, we may wonder what degree of polynomial should be used. As we use high-degree polynomials, we of course will reduce the deviations of the points from the curve until, when the degree of the polynomial, k, equals n-1, there is an exact match (assuming no duplicate data at the same x value) and *we have an interpolating polynomial*.

The answer to this problem is found in statistics. One increases the degree of approximating polynomial as long as there is a statistically significant decrease in the variance, σ^2, which is computed by

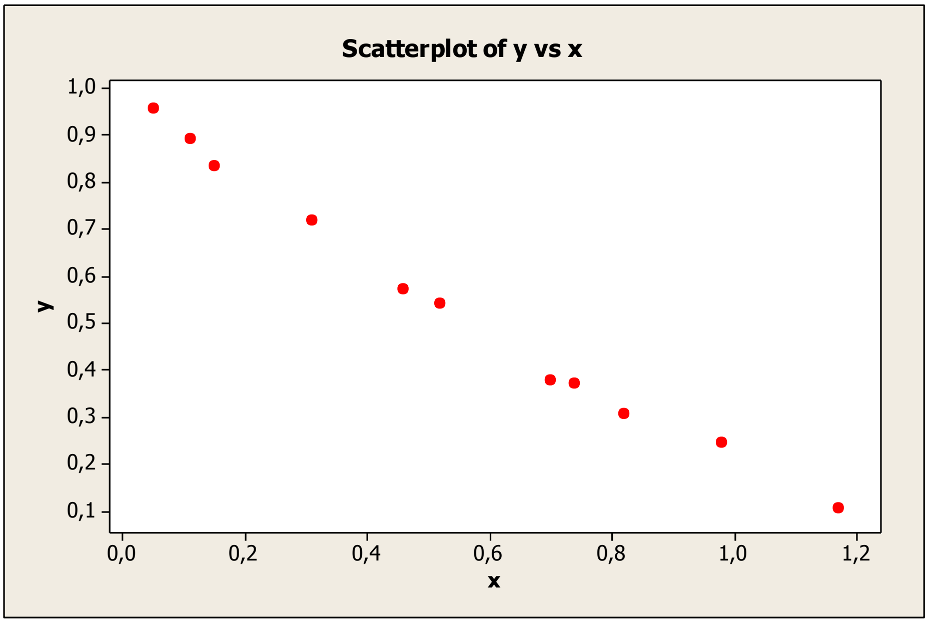
$$\sigma^2 = \frac{\sum e_i^2}{n-k-1}.$$

Example:

Data to illustrate curve fitting

y	x
0,956	0,050
0,890	0,110
0,832	0,150
0,717	0,310
0,571	0,460
0,539	0,520
0,378	0,700
0,370	0,740
0,306	0,820
0,242	0,980
0,104	1,171

N=11


Scatterplot of y vs x
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
y
0,0
0,2
0,4
0,6
0,8
1,0
1,2
x

First Degree

```
MTB > regress c1 1 c2
```

Regression Analysis: y versus x

```
The regression equation is
y = 0,952 - 0,760 x

Analysis of Variance

Source            DF        SS        MS        F
P
Regression         1   0,79259   0,79259   774,58
0,000
Residual Error     9   0,00921   0,00102
Total             10   0,80180
```

Second Degree

```
MTB > regress c1 2 c2 c3
```

Regression Analysis: y versus x; x2

```
The regression equation is
```

$y = 0,998 - 1,02\ x + 0,225\ x^2$

```
Analysis of Variance

Source          DF        SS        MS        F
P
Regression       2   0,79994   0,39997  1722,77
0,000
Residual Error   8   0,00186   0,00023
Total           10   0,80180
```

Third Degree

MTB > regress c1 3 c2 c3 c4

Regression Analysis: y versus x; x2; x3

The regression equation is
$y = 1,00 - 1,08\ x + 0,349\ x^2 - 0,067\ x^3$

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0,79999	0,26666	1034,01	0,000
Residual Error	7	0,00181	0,00026		
Total	10	0,80180			

Fourth Degree

```
MTB > regress c1 4 c2 c3 c4 c5
```

Regression Analysis: y versus x; x2; x3; x4

```
The regression equation is
```

$y = 0{,}988 - 0{,}838\ x - 0{,}52\ x^2 + 1{,}04\ x^3 - 0{,}452\ x^4$

```
Analysis of Variance

Source              DF        SS         MS         F
P
Regression           4   0,80016    0,20004    731,66
0,000
Residual Error       6   0,00164    0,00027
Total               10   0,80180
```

$$\sigma^2 = \frac{\sum e_i^2}{n-k-1}.$$

N=11

In this example smallest value of σ^2 is at degree-2 as we expect.

Degree	n-k-1	$\sum e_i^2$	σ^2
1	11-1-1=9	0.00918	0.00102
2	11-2-1=8	0.00187	0.00023
3	11-3-1=7	0.00181	0.00026
4	11-4-1=6	0.00165	0.00027

Other Related Subjects:

- **Orthogonal Polynomials**
- **Least-squares estimation using singular values**