Regression Analysis

Multiple Regression Analysis

Linear Models

- Deterministic model

$$y_i = \beta_0 + \beta_1 X_i$$

$$y_i = a + bX_i$$

- Stochastic (Random) model

$$y_i = \beta_0 + \beta_1 X_i + e_i$$

Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 X_i + e_i \qquad i = 1, \dots, n$$

Multiple Linear Regression Model

$$y_i = \beta_0 + \underbrace{\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}}_{} + e_i$$

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ji} + e_i$$

$j = 1, \dots, p$ and $i = 1, \dots, n$

Multiple Linear Regression Model

No	Y	X1	X2	X3	X4
1	32	70	78	66	67
2	29	62	74	64	59
3	32	65	77	65	65
4	39	66	80	75	66
5	27	61	74	62	58
6	39	64	84	75	60
7	35	64	83	77	65
8	35	65	81	66	66
9	33	64	81	67	63
10	26	66	73	64	61
11	32	66	82	66	65
12	34	65	82	60	66
13	33	68	80	66	68
14	33	63	71	63	63
15	32	60	79	65	62
16	36	65	80	65	67
17	31	63	71	61	62
18	40	68	85	65	68
19	37	67	76	67	67

No	Y	X1	X2	X3	X4
20	33	67	74	64	68
21	42	68	85	69	67
22	43	67	84	65	65
23	35	66	79	67	67
24	45	70	90	70	71
25	48	73	92	75	74
26	50	76	92	75	77
27	46	73	86	70	71
28	54	71	94	71	72
29	45	69	87	70	71
30	43	74	88	72	72
31	40	72	64	75	71
32	36	66	86	70	69
33	42	74	95	65	75
34	40	71	84	70	72
35	39	75	83	70	74
36	38	69	85	70	68
37	45	75	87	70	78
38	48	73	85	78	70

Multiple Linear Regression Model

No	Y	X1	X2	X3	X4
1	32	70	78	66	67
2	29	62	74	64	59
3	32	65	77	65	65
4	39	66	80	75	66

No	Y	X1	X2	X3	X4
20	33	67	74	64	68
21	42	68	85	69	67
22	43	67	84	65	65
23	35	66	79	67	67

- Y: Dependent variable
- Independent X variables
  - X1 independent variable
  - X2 independent variable
  - X3 independent variable
  - X4 independent variable

Multiple Linear Regression Model

No	Y	X1	X2	X3	X4
1	32	70	78	66	67
2	29	62	74	64	59
3	32	65	77	65	65
4	39	66	80	75	66
20	33	67	74	64	68
21	42	68	85	69	67
22	43	67	84	65	65
23	35	66	79	67	67

- Y: dependent variable
- X1, X2, X3 and X4 independent variables
- Four indepenent simple linear regression models

Multiple Linear Regression Model

No	Y	X1	X2	X3	X4
1	32	70	78	66	67
2	29	62	74	64	59
20	33	67	74	64	68
21	42	68	85	69	67

- Four indepenent simple linear regression models

Simple linear regression model with X1 variable

$$Y_i = \beta_0 + \beta_1 X1_i + e_i \qquad i = 1, \ldots., n$$

Simple linear regression model with X2 variable

$$Y_i = \beta_0 + \beta_2 X2_i + e_i \qquad i = 1, \ldots., n$$

Simple linear regression model with X3 variable

$$Y_i = \beta_0 + \beta_3 X3_i + e_i \qquad i = 1, \ldots., n$$

Simple linear regression model with X4 variable

$$Y_i = \beta_0 + \beta_4 X4_i + e_i \qquad i = 1, \ldots., n$$

No	Y	X1
1	32	70
2	29	62
3	32	65
4	39	66
5	27	61
6	39	64
7	35	64
8	35	65
9	33	64
10	26	66
11	32	66
12	34	65
13	33	68
14	33	63
15	32	60
16	36	65
17	31	63
18	40	68
19	37	67
20	33	67
21	42	68
22	43	67
23	35	66
24	45	70
25	48	73
26	50	76
27	46	73
28	54	71
29	45	69
30	43	74
31	40	72
32	36	66
33	42	74
34	40	71
35	39	75
36	38	69
37	45	75
38	48	73

Multiple Linear Regression Model

Regression model with X1 variable

$$Y_i = \beta_0 + \beta_1 X1_i + e_i \qquad i = 1, \dots, n$$

$$\hat{\beta}_1 = \frac{\sum X1_i Y_i - \frac{(\sum X1_i)(\sum Y_i)}{n}}{\sum X1_i^2 - \frac{(\sum X1_i)^2}{n}} = \frac{774.2368}{658.7632} = 1.175$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \overline{X1} = 38.080 - 1.175 * 67.921 = -41.727$$

No	Y	X1
1	32	70
2	29	62
3	32	65
4	39	66
5	27	61
6	39	64
7	35	64
8	35	65
9	33	64
10	26	66
11	32	66
12	34	65
13	33	68
14	33	63
15	32	60
16	36	65
17	31	63
18	40	68
19	37	67
20	33	67
21	42	68
22	43	67
23	35	66
24	45	70
25	48	73
26	50	76
27	46	73
28	54	71
29	45	69
30	43	74
31	40	72
32	36	66
33	42	74
34	40	71
35	39	75
36	38	69
37	45	75
38	48	73

Multiple Linear Regression Model

Regression equation with X1 variable

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X1_i$$

$$\hat{Y}_i = -41.727 + 1.175 X1_i$$

Estimation of Residuals

$$\hat{e}_i = Y_i - \hat{Y}_i$$

Variation explained by X1 variable

$$\textbf{T}otal\ \textbf{S}um\ of\ \textbf{S}quares\ (\Sigma d_Y^2) = 1598.763$$

$$\textbf{R}egression\ \textbf{M}odel\ \textbf{S}um\ of\ \textbf{S}quares\ (\Sigma d_{\hat{Y}}^2) = 909.952$$

$$\textbf{R}esidual\ \textbf{S}um\ of\ \textbf{S}quares\ (\Sigma d_{\hat{e}}^2) = 688.811$$

$$R\ \left(Correlation\ coefficient\ r_{Y,X1}\right) = 0.754$$

$$R^2\ (Coefficient\ of\ determination) = 0.569$$

No	Y	X2
1	32	78
2	29	74
3	32	77
4	39	80
5	27	74
6	39	84
7	35	83
8	35	81
9	33	81
10	26	73
11	32	82
12	34	82
13	33	80
14	33	71
15	32	79
16	36	80
17	31	71
18	40	85
19	37	76
20	33	74
21	42	85
22	43	84
23	35	79
24	45	90
25	48	92
26	50	92
27	46	86
28	54	94
29	45	87
30	43	88
31	40	64
32	36	86
33	42	95
34	40	84
35	39	83
36	38	85
37	45	87
38	48	85

Multiple Linear Regression Model

Regression model with X2 variable

$$Y_i = \beta_0 + \beta_2 X2_i + e_i \qquad i = 1, \ldots, n$$

$$\hat{\beta}_2 = \frac{\sum X2_i Y_i - \frac{(\sum X2_i)(\sum Y_i)}{n}}{\sum X2_i^2 - \frac{(\sum X2_i)^2}{n}} = \frac{1217.395}{1692.342} = 0.719$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_2 \overline{X2} = 38.080 - 0.719 * 81.868 = -20.814$$

No	Y	X2
1	32	78
2	29	74
3	32	77
4	39	80
5	27	74
6	39	84
7	35	83
8	35	81
9	33	81
10	26	73
11	32	82
12	34	82
13	33	80
14	33	71
15	32	79
16	36	80
17	31	71
18	40	85
19	37	76
20	33	74
21	42	85
22	43	84
23	35	79
24	45	90
25	48	92
26	50	92
27	46	86
28	54	94
29	45	87
30	43	88
31	40	64
32	36	86
33	42	95
34	40	84
35	39	83
36	38	85
37	45	87
38	48	85

Multiple Linear Regression Model

Regression equation with X2 variable

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_2 X2_i$$

$$\hat{Y}_i = -20.814 + 0.719 X2_i$$

Estimation of Residuals

$$\hat{e}_i = Y_i - \hat{Y}_i$$

Variation explained by X2 variable

$$\textbf{T}otal\ \textbf{S}um\ of\ \textbf{S}quares\ (\Sigma d_Y^2) = 1598.763$$

$$\textbf{R}egression\ \textbf{M}odel\ \textbf{S}um\ of\ \textbf{S}quares\ (\Sigma d_{\hat{Y}}^2) = 875.739$$

$$\textbf{R}esidual\ \textbf{S}um\ of\ \textbf{S}quares\ (\Sigma d_{\hat{e}}^2) = 723.024$$

$$R\ (Correlation\ coefficient\ r_{Y,X2}) = 0.740$$

$$R^2\ (Coefficient\ of\ determination) = 0.548$$

No	Y	X3
1	32	66
2	29	64
3	32	65
4	39	75
5	27	62
6	39	75
7	35	77
8	35	66
9	33	67
10	26	64
11	32	66
12	34	60
13	33	66
14	33	63
15	32	65
16	36	65
17	31	61
18	40	65
19	37	67
20	33	64
21	42	69
22	43	65
23	35	67
24	45	70
25	48	75
26	50	75
27	46	70
28	54	71
29	45	70
30	43	72
31	40	75
32	36	70
33	42	65
34	40	70
35	39	70
36	38	70
37	45	70
38	48	78

Multiple Linear Regression Model

Regression model with X3 variable

$$Y_i = \beta_0 + \beta_3 X3_i + e_i \qquad i = 1, \dots, n$$

$$\hat{\beta}_3 = \frac{\sum X3_i Y_i - \frac{(\sum X3_i)(\sum Y_i)}{n}}{\sum X3_i^2 - \frac{(\sum X3_i)^2}{n}} = \frac{720.132}{775.816} = 0.928$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \overline{X3} = 38.080 - 0.928 * 68.290 = -25.309$$

No	Y	X3
1	32	66
2	29	64
3	32	65
4	39	75
5	27	62
6	39	75
7	35	77
8	35	66
9	33	67
10	26	64
11	32	66
12	34	60
13	33	66
14	33	63
15	32	65
16	36	65
17	31	61
18	40	65
19	37	67
20	33	64
21	42	69
22	43	65
23	35	67
24	45	70
25	48	75
26	50	75
27	46	70
28	54	71
29	45	70
30	43	72
31	40	75
32	36	70
33	42	65
34	40	70
35	39	70
36	38	70
37	45	70
38	48	78

Multiple Linear Regression Model

Regression equation with X3 variable

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_3 X3_i$$

$$\hat{Y}_i = -25.309 + 0.928 X3_i$$

Estimation of Residuals

$$\hat{e}_i = Y_i - \hat{Y}_i$$

Variation explained by X3 variable

$$\boldsymbol{T}otal\ \boldsymbol{S}um\ of\ \boldsymbol{S}quares\ (\Sigma d_Y^2) = 1598.763$$

$$\boldsymbol{R}egression\ \boldsymbol{M}odel\ \boldsymbol{S}um\ of\ \boldsymbol{S}quares\ (\Sigma d_{\hat{Y}}^2) = 668.444$$

$$\boldsymbol{R}esidual\ \boldsymbol{S}um\ of\ \boldsymbol{S}quares\ (\Sigma d_{\hat{e}}^2) = 930.319$$

$$R\ (Correlation\ coefficient\ r_{Y,X3}) = 0.647$$

$$R^2\ (Coefficient\ of\ determination) = 0.418$$

No	Y	X4
1	32	67
2	29	59
3	32	65
4	39	66
5	27	58
6	39	60
7	35	65
8	35	66
9	33	63
10	26	61
11	32	65
12	34	66
13	33	68
14	33	63
15	32	62
16	36	67
17	31	62
18	40	68
19	37	67
20	33	68
21	42	67
22	43	65
23	35	67
24	45	71
25	48	74
26	50	77
27	46	71
28	54	72
29	45	71
30	43	72
31	40	71
32	36	69
33	42	75
34	40	72
35	39	74
36	38	68
37	45	78
38	48	70

Multiple Linear Regression Model

Regression model with X4 variable

$$Y_i = \beta_0 + \beta_4 X4_i + e_i \qquad i = 1, \ldots, n$$

$$\hat{\beta}_4 = \frac{\sum X4_i Y_i - \frac{(\sum X4_i)(\sum Y_i)}{n}}{\sum X4_i^2 - \frac{(\sum X4_i)^2}{n}} = \frac{900.105}{854.842} = 1.053$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_4 \overline{X4} = 38.080 - 1.053 * 67.632 = -33.134$$

Multiple Linear Regression Model

No	Y	X4
1	32	67
2	29	59
3	32	65
4	39	66
5	27	58
6	39	60
7	35	65
8	35	66
9	33	63
10	26	61
11	32	65
12	34	66
13	33	68
14	33	63
15	32	62
16	36	67
17	31	62
18	40	68
19	37	67
20	33	68
21	42	67
22	43	65
23	35	67
24	45	71
25	48	74
26	50	77
27	46	71
28	54	72
29	45	71
30	43	72
31	40	71
32	36	69
33	42	75
34	40	72
35	39	74
36	38	68
37	45	78
38	48	70

Regression equation with X4 variable

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_4 X4_i$$

$$\hat{Y}_i = -33.134 + 1.053 X4_i$$

Estimation of Residuals

$$\hat{e}_i = Y_i - \hat{Y}_i$$

Variation explained by X4 variable

$$\boldsymbol{T}otal\ \boldsymbol{S}um\ of\ \boldsymbol{S}quares\ (\Sigma d_Y^2) = 1598.763$$

$$\boldsymbol{R}egression\ \boldsymbol{M}odel\ \boldsymbol{S}um\ of\ \boldsymbol{S}quares\ (\Sigma d_{\hat{Y}}^2) = 947.765$$

$$\boldsymbol{R}esidual\ \boldsymbol{S}um\ of\ \boldsymbol{S}quares\ (\Sigma d_{\hat{e}}^2) = 650.998$$

$$R\ (Correlation\ coefficient\ r_{Y,X4}) = 0.770$$

$$R^2\ (Coefficient\ of\ determination) = 0.593$$

Multiple Linear Regression Model

$\boldsymbol{Regression\ Equation}$	$\boldsymbol{TSS}$	$\boldsymbol{RMSS}$	$\boldsymbol{RSS}$	$\boldsymbol{R^2}$
$\hat{Y}_i = -41.727 + 1.175X1_i$	1598.763	909.952	688.811	0.569
$\hat{Y}_i = -20.814 + 0.719X2_i$	1598.763	875.739	723.024	0.548
$\hat{Y}_i = -25.309 + 0.928X3_i$	1598.763	668.444	930.319	0.418
$\hat{Y}_i = -33.134 + 1.053X4_i$	1598.763	947.765	650.998	0.593

Multiple Linear Regression Model

$$Y_i = \beta_0 + \underbrace{\beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + \beta_4 X4_i}_{} + e_i$$

$$Y_i = \beta_0 + \sum_{j=1}^{p=4} \beta_j X_{ji} + e_i$$

$j = 1, \dots, p = 4$ and $i = 1, \dots, n$

Multiple Linear Regression Model
Matrix Notation

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} \\ 1 & x_{12} & x_{22} & x_{32} & x_{42} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & x_{4n} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{pmatrix}$$

$$\begin{pmatrix} 32 \\ 29 \\ \vdots \\ \vdots \\ 48 \end{pmatrix} = \begin{pmatrix} 1 & 70 & 78 & 66 & 67 \\ 1 & 62 & 74 & 64 & 59 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 73 & 85 & 78 & 70 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}_{n\times 5}\boldsymbol{\beta}_{5\times 1}$$

$$\boldsymbol{X}^{T}_{5\times n}\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}^{T}_{5\times n}\boldsymbol{X}_{n\times 5}\boldsymbol{\beta}_{5\times 1}$$

$$\boldsymbol{X}^{T}_{5\times n}\boldsymbol{X}_{n\times 5}\boldsymbol{\beta}_{5\times 1} = \boldsymbol{X}^{T}_{5\times n}\boldsymbol{Y}_{n\times 1}$$

Multiple Linear Regression Model
Matrix Notation

$$\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\boldsymbol{\beta}_{5\times 1} = \boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\boldsymbol{X}_{n\times 5} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} \\ 1 & x_{12} & x_{22} & x_{32} & x_{42} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & x_{4n} \end{pmatrix}_{n\times 5} = \begin{pmatrix} 1 & 70 & 78 & 66 & 67 \\ 1 & 62 & 74 & 64 & 59 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 73 & 85 & 78 & 70 \end{pmatrix}_{n\times 5}$$

$$\boldsymbol{X}_{5\times n}^{T} = \begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ x_{11} & x_{12} & \cdots & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2n} \\ x_{31} & x_{32} & \cdots & \cdots & x_{3n} \\ x_{41} & x_{42} & \cdots & \cdots & x_{4n} \end{pmatrix}_{5\times n} = \begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 70 & 62 & \cdots & \cdots & 73 \\ 78 & 74 & \cdots & \cdots & 85 \\ 66 & 64 & \cdots & \cdots & 78 \\ 67 & 59 & \cdots & \cdots & 70 \end{pmatrix}_{5\times n}$$

Multiple Linear Regression Model Matrix Notation

$$\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\boldsymbol{\beta}_{5\times 1} = \boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5} = \begin{pmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \Sigma X_4 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \Sigma X_1 X_3 & \Sigma X_1 X_4 \\ \Sigma X_2 & \Sigma X_2 X_1 & \Sigma X_2^2 & \Sigma X_2 X_3 & \Sigma X_2 X_4 \\ \Sigma X_3 & \Sigma X_3 X_1 & \Sigma X_3 X_2 & \Sigma X_3^2 & \Sigma X_3 X_4 \\ \Sigma X_4 & \Sigma X_4 X_1 & \Sigma X_4 X_2 & \Sigma X_4 X_3 & \Sigma X_4^2 \end{pmatrix}$$

$$\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5} = \begin{pmatrix} 38 & 2581 & 3111 & 2595 & 2570 \\ 2581 & 175963 & 211895 & 176630 & 175247 \\ 31111 & 211895 & 256385 & 212910 & 211151 \\ 2595 & 176630 & 212910 & 177987 & 175888 \\ 2570 & 175247 & 211151 & 175888 & 174668 \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

$$\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\boldsymbol{\beta}_{5\times 1} = \boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1} = \begin{pmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \Sigma X_3 Y \\ \Sigma X_4 Y \end{pmatrix} = \begin{pmatrix} 1447 \\ 99056 \\ 119681 \\ 99535 \\ 98763 \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

$$\left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times\mathbf{5}}\right)^{-1}\left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times\mathbf{5}}\right)\boldsymbol{\beta}_{5\times 1} = \left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\right)^{-1}\boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\boldsymbol{I}_{5\times\mathbf{5}} = \left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\right)^{-1}\left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

Least Square Estimates

$$\widehat{\boldsymbol{\beta}}_{5\times 1} = \left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times \mathbf{5}}\right)^{-1}\boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\widehat{\boldsymbol{\beta}}_{5\times 1} = \begin{pmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \Sigma X_4 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \Sigma X_1 X_3 & \Sigma X_1 X_4 \\ \Sigma X_2 & \Sigma X_2 X_1 & \Sigma X_2^2 & \Sigma X_2 X_3 & \Sigma X_2 X_4 \\ \Sigma X_3 & \Sigma X_3 X_1 & \Sigma X_3 X_2 & \Sigma X_3^2 & \Sigma X_3 X_4 \\ \Sigma X_4 & \Sigma X_4 X_1 & \Sigma X_4 X_2 & \Sigma X_4 X_3 & \Sigma X_4^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \Sigma X_3 Y \\ \Sigma X_4 Y \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

$$\widehat{\boldsymbol{\beta}}_{5\times 1} = \left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times \mathbf{5}}\right)^{-1}\boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times \mathbf{5}}\right)^{-1} = \begin{pmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \Sigma X_4 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \Sigma X_1 X_3 & \Sigma X_1 X_4 \\ \Sigma X_2 & \Sigma X_2 X_1 & \Sigma X_2^2 & \Sigma X_2 X_3 & \Sigma X_2 X_4 \\ \Sigma X_3 & \Sigma X_3 X_1 & \Sigma X_3 X_2 & \Sigma X_3^2 & \Sigma X_3 X_4 \\ \Sigma X_4 & \Sigma X_4 X_1 & \Sigma X_4 X_2 & \Sigma X_4 X_3 & \Sigma X_4^2 \end{pmatrix}^{-1}$$

$$\left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times \mathbf{5}}\right)^{-1} = \begin{pmatrix} 38 & 2581 & 3111 & 2595 & 2570 \\ 2581 & 175963 & 211895 & 176630 & 175247 \\ 31111 & 211895 & 256385 & 212910 & 211151 \\ 2595 & 176630 & 212910 & 177987 & 175888 \\ 2570 & 175247 & 211151 & 175888 & 174668 \end{pmatrix}^{-1}$$

Multiple Linear Regression Model
Matrix Notation

$$\widehat{\boldsymbol{\beta}}_{5\times1} = \left(\boldsymbol{X}^T_{5\times n}\boldsymbol{X}_{n\times\mathbf{5}}\right)^{-1}\boldsymbol{X}^T_{5\times n}\boldsymbol{Y}_{n\times1}$$

$$\left(\boldsymbol{X}^T_{5\times n}\boldsymbol{X}_{n\times\mathbf{5}}\right)^{-1} = \begin{pmatrix} 38 & 2581 & 3111 & 2595 & 2570 \\ 2581 & 175963 & 211895 & 176630 & 175247 \\ 31111 & 211895 & 256385 & 212910 & 211151 \\ 2595 & 176630 & 212910 & 177987 & 175888 \\ 2570 & 175247 & 211151 & 175888 & 174668 \end{pmatrix}^{-1}$$

$$\left(\boldsymbol{X}^T_{5\times n}\boldsymbol{X}_{n\times\mathbf{5}}\right)^{-1} = \begin{pmatrix} 8.95731883 & -0.1020743696 & -0.0171773919 & -0.0482915112 & 0.0400121284 \\ -0.1020743696 & 0.0105975989 & 0.0002669055 & -0.0011957086 & -0.0082494381 \\ -0.0171773919 & 0.0002669055 & 0.0009968662 & -0.0002350703 & -0.0009834185 \\ -0.0482915112 & -0.0011957086 & -0.0002350703 & 0.0018349577 & 0.0003466108 \\ 0.0400121284 & -0.0082494381 & -0.0009834185 & 0.0003466108 & 0.0085335792 \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

$$\hat{\boldsymbol{\beta}}_{5\times 1} = \left(\boldsymbol{X}_{5\times n}^{T}\boldsymbol{X}_{n\times 5}\right)^{-1}\boldsymbol{X}_{5\times n}^{T}\boldsymbol{Y}_{n\times 1}$$

$$\hat{\boldsymbol{\beta}}_{5\times 1} = \begin{pmatrix} 8.95731883 & -0.1020743696 & -0.0171773919 & -0.0482915112 & 0.0400121284 \\ -0.1020743696 & 0.0105975989 & 0.0002669055 & -0.0011957086 & -0.0082494381 \\ -0.0171773919 & 0.0002669055 & 0.0009968662 & -0.0002350703 & -0.0009834185 \\ -0.0482915112 & -0.0011957086 & -0.0002350703 & 0.0018349577 & 0.0003466108 \\ 0.0400121284 & -0.0082494381 & -0.0009834185 & 0.0003466108 & 0.0085335792 \end{pmatrix} \begin{pmatrix} 1447 \\ 99056 \\ 119681 \\ 99535 \\ 98763 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}}_{5\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} -60.6235715 \\ 0.2435507 \\ 0.3657660 \\ 0.4214622 \\ 0.3464976 \end{pmatrix}$$

Multiple Linear Regression Model
Matrix Notation

$$\widehat{\boldsymbol{\beta}}_{5\times 1} = \left(\boldsymbol{X}^T_{5\times n}\boldsymbol{X}_{n\times 5}\right)^{-1}\boldsymbol{X}^T_{5\times n}\boldsymbol{Y}_{n\times 1}$$

$$\widehat{\boldsymbol{\beta}}_{5\times 1} = \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{pmatrix} = \begin{pmatrix}-60.6235715\\ 0.2435507\\ 0.3657660\\ 0.4214622\\ 0.3464976\end{pmatrix}$$

Regression Equation

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X1_i + \hat{\beta}_2 X2_i + \hat{\beta}_3 X3_i + \hat{\beta}_4 X4_i$$

$$\hat{Y}_i = -60.62 + 0.24X1_i + 0.37X2_i + 0.42X3_i + 0.35X4_i$$

Multiple Linear Regression Model
Matrix Notation

Y	$\hat{Y}$	$\hat{e}$
32	35,98657	-3,98657
29	28,96020	0,03980
32	33,28860	-1,28860
39	39,19056	-0,19056
27	27,52722	-0,52722
39	38,08754	0,91246
35	40,29719	-5,29719
35	35,51962	-0,51962
33	34,65804	-1,65804
26	30,26163	-4,26163
32	35,78244	-3,78244
34	33,35661	0,64339
33	36,57750	-3,57750
33	29,07098	3,92902
32	31,76288	0,23712
36	35,07889	0,92111
31	27,88156	3,11844
40	37,98487	2,01513
37	34,94585	2,05415
33	33,29643	-0,29643
42	39,32422	2,67578
43	36,33606	6,66394
35	35,79960	-0,79960
45	43,44760	1,55240
48	48,05659	-0,05659
50	49,82674	0,17326
46	42,71519	3,28481
54	45,92218	8,07782
45	42,10676	2,89324
43	44,87970	-1,87970
40	36,53210	3,46790
36	40,31734	-4,31734
42	45,52932	-3,52932
40	41,84306	-1,84306
39	43,14449	-4,14449
38	40,33573	-2,33573
45	45,99354	-0,99354
48	45,37463	2,62537

Regression Equation

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X1_i + \hat{\beta}_2 X2_i + \hat{\beta}_3 X3_i + \hat{\beta}_4 X4_i$$

$$\hat{Y}_i = -60.62 + 0.24X1_i + 0.37X2_i + 0.42X3_i + 0.35X4_i$$

Estimation of Residuals

$$\hat{e}_i = Y_i - \hat{Y}_i$$

Multiple Linear Regression Model
Matrix Notation

Y	$\hat{Y}$	$\hat{e}$
32	35,98657	-3,98657
29	28,96020	0,03980
32	33,28860	-1,28860
39	39,19056	-0,19056
27	27,52722	-0,52722
39	38,08754	0,91246
35	40,29719	-5,29719
35	35,51962	-0,51962
33	34,65804	-1,65804
26	30,26163	-4,26163
32	35,78244	-3,78244
34	33,35661	0,64339
33	36,57750	-3,57750
33	29,07098	3,92902
32	31,76288	0,23712
36	35,07889	0,92111
31	27,88156	3,11844
40	37,98487	2,01513
37	34,94585	2,05415
33	33,29643	-0,29643
42	39,32422	2,67578
43	36,33606	6,66394
35	35,79960	-0,79960
45	43,44760	1,55240
48	48,05659	-0,05659
50	49,82674	0,17326
46	42,71519	3,28481
54	45,92218	8,07782
45	42,10676	2,89324
43	44,87970	-1,87970
40	36,53210	3,46790
36	40,31734	-4,31734
42	45,52932	-3,52932
40	41,84306	-1,84306
39	43,14449	-4,14449
38	40,33573	-2,33573
45	45,99354	-0,99354
48	45,37463	2,62537

Analysis of Variance for Dependent Variable

$$Dependent\ Variable = Regression\ Model + Residual$$

$$Y_i = \hat{Y}_i + \hat{e}_i$$

$$TSS = RMSS + RSS$$

$$\frac{TSS}{TSS} = \frac{RMSS}{TSS} + \frac{RSS}{TSS}$$

$$1 = R^2 + E^2$$

Multiple Linear Regression Model Matrix Notation

Y	$\hat{Y}$	$\hat{e}$
32	35,98657	-3,98657
29	28,96020	0,03980
32	33,28860	-1,28860
39	39,19056	-0,19056
27	27,52722	-0,52722
39	38,08754	0,91246
35	40,29719	-5,29719
35	35,51962	-0,51962
33	34,65804	-1,65804
26	30,26163	-4,26163
32	35,78244	-3,78244
34	33,35661	0,64339
33	36,57750	-3,57750
33	29,07098	3,92902
32	31,76288	0,23712
36	35,07889	0,92111
31	27,88156	3,11844
40	37,98487	2,01513
37	34,94585	2,05415
33	33,29643	-0,29643
42	39,32422	2,67578
43	36,33606	6,66394
35	35,79960	-0,79960
45	43,44760	1,55240
48	48,05659	-0,05659
50	49,82674	0,17326
46	42,71519	3,28481
54	45,92218	8,07782
45	42,10676	2,89324
43	44,87970	-1,87970
40	36,53210	3,46790
36	40,31734	-4,31734
42	45,52932	-3,52932
40	41,84306	-1,84306
39	43,14449	-4,14449
38	40,33573	-2,33573
45	45,99354	-0,99354
48	45,37463	2,62537

Analysis of Variance for Dependent Variable

$$Dependent\ Variable = Regression\ Model + Residual$$

$$Y_i = \widehat{Y_i} + \widehat{e_i}$$

$$TSS = 1598.763 \qquad RMSS = 1249.24 \qquad RSS = 349.523$$

$$\frac{TSS = 1598.763}{TSS = 1598.763} = \frac{RMSS = 1249.24}{TSS = 1598.763} + \frac{RSS = 349.523}{TSS = 1598.763}$$

$$1 = R^2 = 0.78 + E^2 = 0.22$$

Multiple Linear Regression Models

$\boldsymbol{Regression\ Equation}$	$\boldsymbol{TSS}$	$\boldsymbol{RMSS}$	$\boldsymbol{RSS}$	$\boldsymbol{R^2}$
$\hat{Y}_i = -41.727 + 1.175X1_i$	1598.763	909.952	688.811	0.569
$\hat{Y}_i = -20.814 + 0.719X2_i$	1598.763	875.739	723.024	0.548
$\hat{Y}_i = -25.309 + 0.928X3_i$	1598.763	668.444	930.319	0.418
$\hat{Y}_i = -33.134 + 1.053X4_i$	1598.763	947.765	650.998	0.593

$\boldsymbol{Regression\ Equation}$

$$\hat{Y}_i = -60.62 + 0.24X1_i + 0.37X2_i + 0.42X3_i + 0.35X4_i$$

$\boldsymbol{TSS}$	$\boldsymbol{RMSS}$	$\boldsymbol{RSS}$	$\boldsymbol{R^2}$
1598.763	1249.24	349.523	0.781

Multiple Linear Regression Models

Models with one independent variable

$$Y_i = \beta_0 + \beta_1 X1_i + e_i$$
$$Y_i = \beta_0 + \beta_2 X2_i + e_i$$
$$Y_i = \beta_0 + \beta_3 X3_i + e_i$$
$$Y_i = \beta_0 + \beta_4 X4_i + e_i$$

Models with two independent variables

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + e_i$$
$$Y_i = \beta_0 + \beta_1 X1_i + \beta_3 X3_i + e_i$$
$$Y_i = \beta_0 + \beta_1 X1_i + \beta_4 X4_i + e_i$$

$$Y_i = \beta_0 + \beta_2 X2_i + \beta_3 X3_i + e_i$$
$$Y_i = \beta_0 + \beta_2 X2_i + \beta_4 X4_i + e_i$$

$$Y_i = \beta_0 + \beta_3 X3_i + \beta_4 X4_i + e_i$$

Models with three independent variables

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + e_i$$
$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_4 X4_i + e_i$$

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_3 X3_i + \beta_4 X4_i + e_i$$

$$Y_i = \beta_0 + \beta_2 X2_i + \beta_3 X3_i + \beta_4 X4_i + e_i$$

Models with four independent variables

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + \beta_4 X4_i + e_i$$