Regression Let's say a given system has p inputs and one output. Looking at the the historical inputs $\{x_1, \ldots, x_n\}$ (each might be a vector) and the corresponding outputs $\{y_1, \ldots, y_n\}$, we would like to make a guess what y_i will be for an a new x_i.

Simple Linear Regression In simple linear regression, there is only one input and our guess of y_i will be given by the following:

$$y_i = \beta_0\beta_1 x_i + \epsilon_i$$

Where $E(\epsilon_i) = 0$ and the variance of ϵ is σ_2.
The mean of the observed inputs is

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

The mean of the observed outputs is

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Least Squared Estimate (SLR)

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Using β_1, you can estimate β_0:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

The prediction for x_i is

$$\hat{y}_i = \bar{\beta}_0 + \hat{\beta}_1 x_i$$

The prediction error (or residual) is

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

ANOVA

$$SST = S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = SSR + SSE$$

(SST has $n-1$ degrees of freedom) Note that

$$SSR = \hat{\beta}_1 S_{xy} = SST - SSE$$

(SSR has 1 degree of freedom) and

$$SSE = \sum_{i=1}^n \hat{\epsilon}_i{}^2 = SST + SSR$$

(SSE has $n-2$ degrees of freedom).

$$MSR = \frac{SSR}{df(SSR)}$$

$$\hat{\sigma}^2 = MSE = \frac{SSE}{df(SSE)}$$

$$Fstatistic = \frac{MSR}{MSE}$$

which follows Snedecor's F-distribution with $df_1 = df(SSR)$ and $df_2 = df(SSE)$. The p-value is the tail probability of the observed F-statistic. Anything smaller than 0.05 is pretty good.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \hat{\beta}_1{}^2 \frac{S_{xx}}{S_{yy}} = \hat{\beta}_1 \frac{S_{xy}}{S_{yy}}$$

Quality of Parameters The standard error of our estimate of $\hat{\beta}_1$ is

$$s.e.(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

The T-statistic for $\hat{\beta}_1$:

$$\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)}$$

which follows Student's distribution with $df = n-2$. The p-value is the tail probability of the observed t-statistic. Once again, anything smaller than 0.05 is pretty good.

Confidence Interval of Expectation The prediction of the mean response at $x = x_0$ is given by

$$E(Y) = \hat{y_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

The standard error of the prediction of $E(Y)$ at x_0 is given by

$$s.e.(prediction) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{x_0 - \bar{x}^2}{S_{xx}} \right)}$$

Thus, $100(1-\alpha)$ confidence interval of $E(Y)$ at $x = x_0$ is

$$PointPredition \pm \left(t_{\alpha/2, df=n-2}\right)(s.e.(prediction))$$

Confidence Interval of New Observation The prediction is the same. But the standard error is bigger:

$$s.e.(prediction) = \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{x_0 - \bar{x}^2}{S_{xx}} \right)}$$

The confidence interval is calculated the same as above using the t-distribution.

Adjusted R^2

$$R^2_{adj} = 1 - \left(1 - R^2\right) \frac{n-1}{n-k-1}$$

where k is the number of parameters.

Variance Inflation Factor For a variable X_j that is suspected of being correlated with other variables, we remove it if its VIF is greater than 5.

$$VIF(X_j) = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R^2 of the regression run without X_j.

MLR in matrices Let X be the matrix where the inputs for each sample are a row and the first item in the row is 1. Let Y be the column vector of outputs. Let β be the column vector of coefficients. Let Σ be a column vector of residuals.

$$Y = X\beta + \Sigma$$