January 26, 1999

Useful Equations for Linear Regression

Simple linear regression: one predictor ($p = 1$):
Model: $E(y|x) = \alpha + \beta x$
$E(y)$ =expectation or long–term average of y $\quad$ | = conditional on
Alternate statement of model: $y = \alpha + \beta x + e$, $\quad e$ normal with mean zero for all x, $\quad var(e) = \sigma^2 = var(y|x)$

Assumptions:

1. Linearity
2. σ^2 is constant, independent of x
3. Observations (e's) are independent of each other
4. For proper statistical inference (CI, P–values), y (e) is normal conditional on x

Verifying some of the assumptions:

1. In a scattergram the spread of y about the fitted line should be constant as x increases
2. In a residual plot ($d = y - \hat{y}$ vs. x) there are no systematic patterns (no trend in central tendency, no change in spread of points with x)

Sample of size $n : (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

$$L_{xx} = \sum (x_i - \bar{x})^2 \qquad L_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{\beta} = b = \frac{L_{xy}}{L_{xx}} \qquad \hat{\alpha} = a = \bar{y} - b\bar{x}$$

$$\hat{y} = a + bx = \hat{E}(y|x) \qquad \text{estimate of } E(y|x) = \text{ estimate of y}$$

$$SST = \sum (y_i - \bar{y})^2 \qquad MST = \frac{SST}{n-1} = s_y^2$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2 \qquad MSR = \frac{SSR}{p}$$

$$SSE = \sum (y_i - \hat{y}_i)^2 \qquad MSE = \frac{SSE}{n-p-1} = s_{y \cdot x}^2$$

$$SST = SSR + SSE \qquad F = \frac{MSR}{MSE} = \frac{R^2/p}{(1-R^2)/(n-p-1)} \sim F_{p,n-p-1}$$

$$R^2 = \frac{SSR}{SST} \qquad \frac{SSR}{MSE} \dot{\sim} \chi_p^2$$

$$(p = 1) \quad \widehat{s.e.}(b) = \frac{s_{y \cdot x}}{\sqrt{L_{xx}}} \qquad t = \frac{b}{\widehat{s.e.}(b)} \sim t_{n-p-1}$$

$$1-\alpha \text{ two–sided CI for } \beta \qquad b \pm t_{n-p-1,1-\alpha/2}\widehat{s.e.}(b)$$

$$(p=1)\quad \widehat{s.e.}(\hat{y}) = s_{y\cdot x}\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{L_{xx}}}$$

$$1-\alpha \text{ two–sided CI for } y \qquad \hat{y} \pm t_{n-p-1,1-\alpha/2}\widehat{s.e.}(\hat{y})$$

$$(p=1)\quad \widehat{s.e.}(\hat{E}(y|x)) = s_{y\cdot x}\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{L_{xx}}}$$

$$1-\alpha \text{ two–sided CI for } E(y|x) \qquad \hat{y} \pm t_{n-p-1,1-\alpha/2}\widehat{s.e.}(\hat{E}(y|x))$$

Multiple linear regression: p predictors, $p > 1$:
Model: $E(y|x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + e$
Interpretation of β_j: effect on y of increasing x_j by one unit, holding all other x's constant

Assumptions: same as for $p = 1$ plus no interaction between the x's (x's act additively; effect of x_j does not depend on the other x's).

Verifying some of the assumptions:

1. When $p = 2$, x_1 is continuous, and x_2 is binary, the pattern of y vs. x_1, with points identified by x_2, is two straight, parallel lines
2. In a residual plot ($d = y - \hat{y}$ vs. $\hat{y}$) there are no systematic patterns (no trend in central tendency, no change in spread of points with $\hat{y}$). The same is true if one plots d vs. any of the x's.
3. Partial residual plots reveal the partial (adjusted) relationship between a chosen x_j and y, controlling for all other $x_i, i \neq j$, without assuming linearity for x_j. In these plots, the following quantities appear on the axes:

   y **axis:** residuals from predicting y from all predictors except x_j

   x **axis:** residuals from predicting x_j from all predictors except x_j (y is ignored)

When $p > 1$, least squares estimates are obtained using more complex formulas. But just as in the case with $p = 1$, all of the coefficient estimates are weighted combinations of the y's, $\sum w_i y_i$ [when $p = 1$, the w_i for estimating β are $\frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$].

Hypothesis tests with $p > 1$:

- Overall F test tests $H_0 : \beta_1 = \beta_2 = \ldots \beta_p = 0$ vs. the althernative hypothesis that at least one of the β's $\neq 0$.
- To test whether an individual $\beta_j = 0$ the simplest approach is to compute the t statistic, with $n - p - 1$ d.f.

- Subsets of the β's can be tested against zero if one knows the standard errors of all of the estimated coefficients and the correlations of each pair of estimates. The formulas are daunting.
- To test whether a subset of the β's are all zero, a good approach is to compare the model containing all of the predictors associated with the β's of interest with a sub–model containing only the predictors not being tested (i.e., the predictors being adjusted for). This tests whether the predictors of interest add response information to the predictors being adjusted for. If the goal is to test $H_0 : \beta_1 = \beta_2 = \ldots = \beta_q = 0$ regardless of the values of $\beta_{q+1}, \ldots, \beta_p$ (i.e., adjusting for $x_{q+1}, \ldots, x_p$), fit the full model with p predictors, computing SSE_{full} or R^2_{full}. Then fit the sub–model omitting $x_1, \ldots, x_q$ to obtain $SSE_{reduced}$ or $R^2_{reduced}$. Then compute the partial F statistic

$$F = \frac{(SSE_{reduced} - SSE_{full})/q}{SSE_{full}/(n-p-1)} = \frac{(R^2_{full} - R^2_{reduced})/q}{(1-R^2_{full})/(n-p-1)} \sim F_{q,n-p-1}$$

  Note that $SSE_{reduced} - SSE_{full} = SSR_{full} - SSR_{reduced}$.

Notes about distributions:

- If $t \sim t_b$, $\quad t \dot{\sim}$ normal for large b and $t^2 \dot{\sim} \chi^2_1$, so $[\frac{b}{\widehat{s.e.}(b)}]^2 \dot{\sim} \chi^2_1$
- If $F \sim F_{a,b}$, $\quad a \times F \dot{\sim} \chi^2_a$ for large b
- If $F \sim F_{1,b}$, $\quad \sqrt{F} \sim t_b$
- If $t \sim t_b$, $\quad t^2 \sim F_{1,b}$
- If $z \sim$ normal, $\quad z^2 \sim \chi^2_1$
- $y \sim D$ means y is distributed as the distribution D
- $y \dot{\sim} D$ means that y is approximately distributed as D for large n
- $\hat{\theta}$ means an estimate of θ