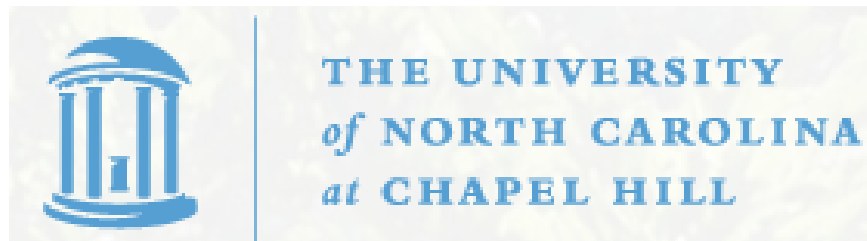


STOR 556: ADV METH DATA ANAL

Instructor: Richard L. Smith

**Class Notes #2:
January 15, 2019**



BASICS OF LINEAR REGRESSION

$$y_i = x_{i0}\beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \epsilon_i, \quad i = 1, \dots, n$$

where y_i is i th value of the observation of interest, x_{i0}, \dots, x_{ip} are the associated covariates, and $\epsilon_1, \dots, \epsilon_n$ are random errors. Here β_0, \dots, β_p are the unknown parameters, or regression coefficients. Usually we assume $x_{i0} = 1$ and in that case we call β_0 the intercept. Matrix form:

$$y = X\beta + \epsilon.$$

Principle of least squares: Find β_0, \dots, β_p to minimize

$$L = \sum_i \left(y_i - \sum_j x_{ij}\beta_j \right)^2.$$

Solve by calculus.

$$\begin{aligned} \frac{\partial L}{\partial \beta_0} &= -2 \sum_i \left(y_i - \sum_j x_{ij} \beta_j \right) x_{i0}, \\ \frac{\partial L}{\partial \beta_1} &= -2 \sum_i \left(y_i - \sum_j x_{ij} \beta_j \right) x_{i1}, \\ &\dots \\ \frac{\partial L}{\partial \beta_p} &= -2 \sum_i \left(y_i - \sum_j x_{ij} \beta_j \right) x_{ip}. \end{aligned}$$

We find the minimizing $\hat{\beta}_0, \dots, \hat{\beta}_p$ by setting all the partial derivatives to 0, hence

$$\sum_i \left(y_i - \sum_j x_{ij} \hat{\beta}_j \right) x_{ik} = 0, \quad k = 0, \dots, p.$$

Matrix notation:

$$X^T y - X^T X \hat{\beta} = 0.$$

The Normal Equations

Predicted values, R^2 and R_a^2

$$\hat{y}_i = \sum_k x_{ik} \hat{\beta}_k$$

or in matrix notation

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y.$$

We define (in case $x_{i0} \equiv 1$)

$$RSS = \sum_i (\hat{y}_i - y_i)^2,$$

$$TSS = \sum_i (y_i - \bar{y})^2,$$

$$R^2 = 1 - \frac{RSS}{TSS}.$$

An alternative is the *adjusted* R^2 given by

$$R_a^2 = 1 - \frac{RSS/(n-p)}{TSS/(n-1)}.$$

Summary Tables in R

The `summary` command in R produces a table of values that includes information about

1. *The residuals* — values $r_i = y_i - \hat{y}_i$,
2. The standard errors, t-statistics and p-values of each of the parameter estimates.

For a parameter estimate $\hat{\beta}_k$, R will give us a standard error s_k , then

$$t_k = \frac{\hat{\beta}_k}{s_k}$$

is called the k th t statistic, so called because it has a t_{n-p} distribution under the null hypothesis that $\beta_k = 0$.

Confidence Interval for a Single Parameter

The confidence interval for $\hat{\beta}_k$ at (two-sided) significance level α is

$$\hat{\beta}_k \pm t_{n-p}^{\alpha/2} s_k$$

where $t_{n-p}^{\alpha/2}$ is the value exceeded with probability $\alpha/2$ by the t_{n-p} distribution (in R: `qt(1-alpha/2,n-p)`).

Confidence Interval for a Single Parameter

The confidence interval for $\hat{\beta}_k$ at (two-sided) significance level α is

$$\hat{\beta}_k \pm t_{n-p}^{\alpha/2} s_k$$

where $t_{n-p}^{\alpha/2}$ is the value exceeded with probability $\alpha/2$ by the t_{n-p} distribution (in R: `qt(1-alpha/2,n-p)`).

F Tests

Useful for testing whether a whole group of parameters is 0.

Suppose we have two models l_{m1} and l_{m2} where l_{m1} is *nested* within l_{m2} (in other words, every parameter that is in l_{m2} is also in l_{m1} , but not the other way round).

In the text, the two models are denoted by ω (l_{m1}) and Ω (l_{m2}). Suppose they have respectively q and p parameters, with $q < p$.

If model ω is true, then we have

$$F = \frac{(RSS_{\omega} - RSS_{\Omega}) / (p - q)}{RSS_{\Omega} / (n - p)} \sim F_{p-q, n-p}.$$

In R: `anova(lm1, lm2)`.