STOR 590:
ADVANCED LINEAR MODELS
Instructor: Richard L. Smith

Class Notes:

August 14, 2020

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL
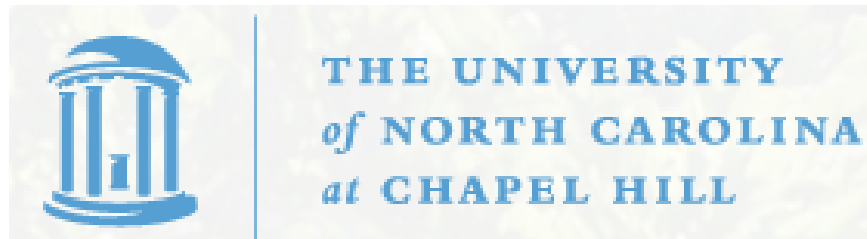
Summary Tables in R

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

1. *The residuals* — values $r_i = y_i - \widehat{y}_i$,

2. The standard errors, t-statistics and p-values of each of the parameter estimates.

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

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

is called the kth 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 $\widehat{\beta}_k$ at (two-sided) significance level α is

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

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 `lm1` and `lm2` where `lm1` is *nested* within `lm2` (in other words, every parameter that is in `lm2` is also in `lm1`, but not the other way round).

In the text, the two models are denoted by ω (`lm1`) and Ω (`lm2`). 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)`.