11.13 Prediction in MLR

- If we are predicting the response value Y^* for a new individual having x values $(x_1^*, x_2^*, \ldots, x_k^*)$, we would use:

- Similarly as in Sec. 11.7, it can be shown that

and

So a $100(1-\alpha)\%$ prediction interval for Y^* when $x_1 = x_1^*, \ldots, x_k = x_k^*$ is:

<u>Example 1 again</u>: Find a 90% PI for the graduation rate of a single college with median SAT score 1050 and per capita expenses \$15,000.

<u>Example 2</u>: Using the quadratic regression model, find a 90% PI for the 20-km race time of a skier with time to exhaustion of 9 minutes.

11.14 Reduced and Complete Models

- We have seen examples of testing whether a single $\beta_i = 0$ (whether the i-th independent variable is needed in the model) with a t-test.
- what about testing whether a set of independent variables is unnecessary?
- That is, if the complete model is

we could test a hypothesis like

- We do this by fitting the reduced model that would hold if H_0 were true:

and comparing the SSEs of the reduced and complete models.

Recall: We want SSE to be ________.

- Adding important variables to the model will improve the fit and ________ the SSE.
- If the set $(x_{g+1}, \dots, x_k)$ is important in explaining Y, then
- If the set $(x_{g+1}, \dots, x_k)$ is useless in explaining Y, then
- While $(SSE_R - SSE_c)$ cannot be negative, we see that ________ values of $(SSE_R - SSE_c)$ would lead us to reject $H_0: \beta_{g+1} = \cdots = \beta_k = 0$ and to conclude $(x_{g+1}, \dots, x_k)$ is important.

Cochran's Theorem (special case):

- If $H_0: \beta_{g+1} = \cdots = \beta_k = 0$ is true, then

- If we divide these r.v.'s by their d.f. and then take the ratio, we get:

which (under H_0) has ____________________

with ________________________________.

- Therefore, we reject $H_0: \beta_{g+1} = \cdots = \beta_k = 0$ if

__Example__: Test $H_0: \beta_1 = \beta_2 = 0$ in the quadratic regression model for the ski-racer data. Use $\alpha = .05$.

<u>Example 1 again:</u> Consider fitting a second-order regression model (conic surface) to the college data. Is this needed, or is the first-order model sufficient? Use $\alpha = .05$.

<u>Note:</u> Sec. 11.15 discusses some practical aspects of model fitting that will be useful for your project. See also the example R code on the course web page.