2.6: Multiple Inference and Multicollinearity

Dr. Bean - Stat 5100

Think of an real-world example of where a "subset" F test might be useful.

My own research: I want to know if temperature related variables (mean annual temperature, temperature difference between the warmest and coldest month of the year, daily temperature difference) are related to annual snow *accumulation* after accounting for the effects of elevation and mean annual precipitation. The key here is that I want to know if *any* of the temperature variables have an influence in the prediction but I am not as much interested in which specific temperature variable is causing the influence.

Example: Bodyfat Dataset (Handout 2.6.1) Y = body, X_1 = triceps, X_2 = thigh, X_3 = midarm

$$Y \;=\; \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

How would you describe the hypothesis $H_0 : \beta_2 = \beta_3 = 0$ in an English sentence?

We want to know if thigh and midarm measurements share a significant relationship with bodyfat *after* accounting for the variance in bodyfat already explained by tricep measurements.

True or False (and explain): Because the Type I SS associated with X_1 is greatest, it means that X_1 is the most significant coefficient in the model.

FALSE The first of the Type I SS will often be the largest because no other predictors have yet been accounted for. This is why order matters in the Type I SS calculation.

What other advantages (besides help with multicollinearity) might standardizing our variables provide us?

Perhaps most notably: the slopes of each b_k coefficient are now directly comparable to each other.

True or False: Eliminating multicollinearity should improve the predictive power of my linear model.

FALSE Multicollinearity has nothing to do with a model's prediction capability. It only affects our ability to make inference on the coefficients.

True or False: The p-value for the model F-test is unreliable when the model contains multicollinearity.

FALSE The model F-test is unaffected by multicollinearity. It is only inference on particular model coefficients that is ruined.