Multiple hypothesis testing worksheet 2

Ryan T. Godwin

F-test

1. (Review) Do you reject a hypothesis when the test statistic is large, or when it is small?

   When it is large! (For example, we reject when the t-test statistic is greater then 1.96 in absolute value).

2. What are the 3 formulas for the F-test that we have seen in class?

$$F = \frac{(R\boldsymbol{b} - \boldsymbol{q})' \left[R\left(X'X\right)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q})/J}{s^2}$$

$$F = \frac{\left(\boldsymbol{e}_*{}'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e}\right)/J}{s^2} = \frac{\left(\boldsymbol{e}_*{}'\boldsymbol{e}_* - \boldsymbol{e}'\boldsymbol{e}\right)/J}{\boldsymbol{e}'\boldsymbol{e}/(n-k)}$$

$$F = \frac{\left(R^2 - R_*^2\right)/J}{\left(1 - R^2\right)/(n-k)}$$

3. What is the intuition behind the version of the F-test that uses R-square?

   If the restricted model under the null hypothesis fits much worse than the unrestricted model under the alternative hypothesis, then $\left(R^2 - R_*^2\right)$ will be large, the F-test statistic will be large, and we'll reject.

Testing for differences

4. What is the starting point for differentiating (and testing for differences) between groups?

   Using a dummy variable that indicates which observation belongs to which group ($D = 0$ for one group, $D = 1$ for the other group).

5. Write down a population model that allows for the marginal effect of a variable to differ between two groups.

   For example:

$$\boldsymbol{y} = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_3 D X_1 + \beta_4 X_2 + \epsilon$$

   In this model, the marginal effect of X_1 on Y is $\beta_1 + D\beta_3$. It is β_1 for the $D = 0$ group and $\beta_1 + \beta_3$ for the $D = 1$ group.

6. Write down a population model that allows *everything* to differ between two groups.

   We could allow the dummy variable to interact fully with every variable in the model:

$$\boldsymbol{y} = X\boldsymbol{\beta}_1 + DX\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

7. Explain how to test for differences between groups.

   Since the dummy variable is allowing for a difference between groups for every variable that it interacts with, a joint hypothesis may be used to test the coefficients on any terms that the dummy interacts with.