F-tests and Nested Models

Nested Models: A core concept in statistics is comparing **nested models**. Consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon. \tag{1}$$

The following **reduced models** are special cases (nested within) the **full** or **complete** model (1):

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \epsilon \\ Y &= \beta_0 + \beta_2 x_2 + \epsilon \\ Y &= \beta_0 + \epsilon \end{aligned}$$

By setting certain coefficients in (1) to zero, we get the various reduced models.

Assessing Predictors and Testing Coefficients via Nested Models: By comparing a given reduced model with the complete model we can assess the usefulness of one or more predictors and formally test whether or not the corresponding coefficients are nonzero. There are various methods and statistics for comparing nested statistical models. The most popular method for comparing **nested regression models** uses the **Sum of Squares Error, SSE:**

$$\text{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

For example, suppose we want to compare the following two models:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \tag{2}$$
$$Y = \beta_0 + \beta_1 x_1 + \epsilon \tag{3}$$

The bottom, reduced model (3) is nested in the top, complete model (2), that is, (3) is the special case of (2) when $\beta_2 = 0$. Let SSE(reduced) denote the SSE for model (3) and SSE(full) denote the SSE for model (2). If

$$\Delta\text{SSE} = \text{SSE(reduced)} - \text{SSE(full)}$$

is large then we have evidence that

1. the full model (2) is superior to the reduced model (3),
2. x_2 is a useful predictor, and
3. β_2 is nonzero.

F-test for Formally Comparing Models/Testing Coefficients: We formally compare nested models and test that the corresponding coefficients are nonzero using the F-statistic or F-ratio, a scaled version of ΔSSE; see next page.

F-test for Formally Comparing Models and Testing Coefficients

Assumptions: Full/complete model satisfies all 5 regression assumptions. **Note: For F-test purposes we don't care if the reduced model meets the regression assumptions.**

F-Statistic: The F-test statistic or F-ratio is simply a scaled version of ΔSSE:

$$\begin{aligned} F &= \frac{[\text{SSE(R)} - \text{SSE(F)}]/\Delta p}{\hat{\sigma}^2_{\text{F}}} \\ &= \frac{\Delta\text{SSE}/\Delta p}{\text{MSE}_{\text{F}}} \end{aligned}$$

where

1. SSE(R) is the reduced model SSE
2. SSE(F) is the full model SSE
3. Δp is the number of coefficients being tested
4. $\hat{\sigma}^2_{\text{F}} = \text{MSE}_{\text{F}}$ is the full-model estimate of the random error variance σ^2.

Note that the numerator of F is essentially the average reduction in SSE per predictor eliminated from the full model. Since the numerator is in units of Y squared and the denominator $\hat{\sigma}^2_{\text{F}}$ is also in units of Y squared, F is dimensionless and hence invariant to changes in units.

Hypotheses: The F-test hypotheses are

H$_o$: All coefficients under consideration are zero

H$_a$: At least one of the coefficients in nonzero

Null Distribution of F**:** Assuming the full model satisfies the 5 regression assumptions and H$_o$ is true, the distribution of the F-statistic F is F_{ν_1,ν_2} where

ν_1**:** numerator degrees of freedom, Δp

ν_2**:** denominator degrees of freedom, $n - p$, where p is the number of coefficients in the full model.

P-value Computation: Since a large ΔSSE yields a large value for F, all things being equal, large values of F (F >> 1) provide strong evidence against H$_o$ in favor of H$_a$. Therefore the F-test p-value is $P(F_{\Delta p,n-p} \geq \text{F})$ where F is the observed F-ratio.

Types of F tests: There are various types of F tests and corresponding hypotheses. The two most important types of F tests are nested model F tests:

Overall F Test/F Test for Regression Relation: Suppose we are interested in the regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} X_{p-1} + \epsilon$$

The fundamental question with respect to this model is whether or not any of the $p-1$ predictor variables $x_1, x_2, \ldots, x_{p-1}$ are useful predictors of Y, i.e., whether or not any of the $p-1$ regression coefficients $\beta_1, \beta_2, \ldots, \beta_{p-1}$ are nonzero. In order to answer this question we test H$_o$: $\beta_1 = \beta_2 = \cdots = \beta_{p-1}$ vs. H$_a$: *at least one of* $\beta_1, \beta_2, \ldots, \beta_{p-1}$ *is nonzero.* by comparing the full model above with the constant-only reduced model

$$Y = \beta_0 + \epsilon$$

The corresponding F test has various names: "Model utility Test," "Overall F Test," and "F Test for Regression Relation."

Partial F Test: The "Partial F Test" is the term used for nested model F tests in which the reduced model is something other than the constant-only model. For example, we may wish to compare the full model above with the reduced model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q X_q + \epsilon$$

Exercise 1: Simple Linear Model Overall F test. Suppose we fit the simple linear model

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{4}$$

to the following set of three (x, y) pairs: $\{(-1, 1), (0, 5), (1, 3)\}$. Answer/do the following:

i. Enter the data into Minitab and verify that the least squares estimates of β_0 and β_1 are $b_0 = 3$ and $b_1 = 1$.

ii. In order to test H$_o$: $\beta_1 = 0$ vs. H$_a$: $\beta_1 \neq 0$ we need to compare the full model (4) with the reduced model

$$Y = \beta_0 + \epsilon \tag{5}$$

In order to do this test certain assumptions must be met. What are they?

iii. Using your least squares estimates from i manually compute SSE(Full), SSE for the full model (4) above.

iv. Using the fact that the least squares estimate of β_0 in the reduced model (5) is $\bar{y} = b_0 = 3$, manually compute SSE(Reduced), SSE for the reduced model (5).

v. Manually compute the F-statistic for testing H$_o$: $\beta_1 = 0$ vs. H$_a$: $\beta_1 \neq 0$.

vi. Determine the numerator and denominator degrees of freedom for your F-statistic.

vii. See next page..

vii. Compute your p-value as follows using Minitab's `Graph -> Probability Distribution Plot -> View Probability`.

1. Select `F` for `Distribution` and enter the numerator and denominator degrees of freedom.
2. Next click on the `Shaded Area` tab, select the `x value` radio button, enter your the value of your F-ratio, then click OK.

viii. Compare your results above with the `Analysis of Variance` output at the bottom of your regression output from fitting the full model. Do you notice any similarities?

Regression Analysis of Variance

Simple Linear Model ANOVA: As you discovered in the previous excercise, the overall F test corresponds to an **Analysis of Variance (ANOVA)** of the regression model. Recall that ANOVA refers to an analysis in which we partition the sums of squares and degrees of freedom of the response variable Y. The ANOVA corresponding to the overall F test is based on the following identity,

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\ \text{SSTO} &= \text{SSE} + \text{SSR}.\end{aligned}$$

To relate this ANOVA identity to the overall F test in exercise 1, recall that this F test compares the full model

$$Y = \beta_0 + \beta_1 x_1 + \epsilon$$

with the reduced, constant-only model

$$Y = \beta_0 + \epsilon$$

We connect the overall F test with our ANOVA identity by noting two things:

1. $\text{SSE(full)} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \text{SSE}$.
2. $\text{SSE(reduced)} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \text{SSTo}$. To see this recall that the least squares estimate of β_0 in the constant-only model is $b_0 = \bar{Y}$. Thus for the reduced, constant-only model $\hat{Y}_i = \bar{Y}$ and therefore

$$\begin{aligned}\text{SSE(reduced)} &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \\ &= \text{SSTO}\end{aligned}$$

Thus for ΔSSE we have

$$\begin{aligned}\Delta\text{SSE} &= \text{SSE(reduced)} - \text{SSE(full)} \\ &= \text{SSTO} - \text{SSE} \\ &= \text{SSR}\end{aligned}$$

where the last line follows from the ANOVA identity above.

Our overall F test F-ratio is thus

$$\begin{aligned} \mathrm{F} &= \frac{\Delta\mathrm{SSE}/\Delta p}{\hat{\sigma^2}_{\mathrm{F}}} \\ &= \frac{\mathrm{SSR}/1}{\mathrm{SSE}/(n-2)} \\ &= \frac{MSR}{MSE} \end{aligned}$$

Arranging the calculations for this F-ratio in tabular format, we get the ANOVA table for the simple linear regression model (Table 2.2, page 67):

Source of variation	**SS**	**df**	**MS**	**F**	**P**
Regression	SSR	1	MSR	F = MSR/MSE	p-value
Error	SSE	n-2	MSE		
Total	SSTO	n-1			

Compare this with your calculations for exercise 1 and the corresponding Minitab regression analysis of variance table below:

```
Source          DF     SS     MS     F      P
Regression       1  2.000  2.000  0.33  0.667
Residual Error   1  6.000  6.000
Total            2  8.000
```

Multiple Regression Model ANOVA: In the multiple regression model case where we have $p-1$ predictors the overall F test compares the full model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1} + \epsilon$$

with the reduced model

$$Y = \beta_0 + \epsilon$$

The ANOVA identity is the same as before as is the ANOVA table except that the degrees of freedom differ since there are $p-1$ predictors:

Source of variation	**SS**	**df**	**MS**	**F**	**P**
Regression	SSR	p-1	MSR	F = MSR/MSE	p-value
Error	SSE	n-p	MSE		
Total	SSTO	n-1			

Compare this with Table 6.1, page 225. Note that Table 6.1 provides matrix formulas for SSR, SSE, and SSTO. We will discuss these later.

Exercise 2: Partial F-test of a Polynomial Model. Open the data set for problem 6 of hw 2 on the course website

`www.rose-hulman.edu/class/ma/inlow/Math485`

An important extension of the simple linear model is the polynomial model in which higher order terms, e.g., quadratics and cubics, are added. Since, as we determined earlier, the relationship between `citympg` and `wt` is nonlinear we might consider fitting the second order model

$$\texttt{citympg} = \beta_0 + \beta_1\texttt{wt} + \beta_2\texttt{wt}^2 + \epsilon.$$

We can fit this model using Minitab by doing the following:

1. Create a column called `wt2` and, using Minitab's calculator menu, fill it with the squares of the `wt` values.
2. Use `regression -> regression` to fit the model by specifying both `wt` and `wt2` as predictors.

2.1: Determine the SSE for the second order (full) model from the output (HINT: it's provided in the ANOVA table) and compare it with the SSE for the simple linear (reduced) model using in order to test the significance of the quadratic term using a partial F-test. What are the corresponding hypotheses and what do you conclude at $\alpha = 0.05$?

2.2: What is the F-statistic for the model utility test for the second order model and what do you conclude at $\alpha = 0.05$?

2.3 What assumptions must be met for these F tests to be valid? Check them as completely as possible. What do you conclude?