Comparing Nested Models

Two models are *nested* if one model contains all the terms of the other, and at least one additional term.

The larger model is the *complete* (or *full*) model, and the smaller is the *reduced* (or *restricted*) model.

Example: with two independent variables x_1 and x_2, possible *terms* are x_1, $x_1 x_2$, x_1^2, and so on.

Consider three models:

- First order:
$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2;$$
- Interaction:
$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2;$$
- Full second order:
$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2.$$

The first order model is nested within both the Interaction model and the Full second order model.

The Interaction model is nested within the Full second order model.

We usually want to use the simplest (most *parsimonious*) model that adequately fits the observed data.

One way to decide is by testing

- H_0: reduced model is adequate;
- H_a: full model is better.

When the full model has exactly *one* more term than the reduced model, we can use a t-test.

E.g., testing the Interaction model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2;$$

against the First order model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

- H_0: "reduced model is adequate" is the same as $H_0 : \beta_3 = 0$.
- So the usual t-statistic is the relevant test statistic.

When the full model has *more than one* additional term, we use an F-test, which generalizes the t-test.

Basic idea: fit both models, and test whether the full model fits significantly better than the reduced model:

$$F = \frac{\left(\dfrac{\text{Drop in SSE}}{\text{Number of extra terms}}\right)}{s^2 \text{ for full model}}$$

where SSE is the sum of squared residuals.

When H_0 is true, F follows the F-distribution, which we use to find the P-value.

E.g., testing the Full second order model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2;$$

against the Interaction model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

Here H_0 is $\beta_4 = \beta_5 = 0$, and H_a is the opposite.

In R, the `lm()` method is not convenient for carrying out this test; `aov()` is better.

```
summary(aov(Cost ~ Weight + Distance + I(Weight * Distance) +
            I(Weight^2) + I(Distance^2), express))
                     Df Sum Sq Mean Sq  F value   Pr(>F)
Weight                1 270.55  270.55 1380.001 2.17e-15 ***
Distance              1 143.63  143.63  732.616 1.72e-13 ***
I(Weight * Distance)  1  31.27   31.27  159.487 4.84e-09 ***
I(Weight^2)           1   3.80    3.80   19.383 0.000602 ***
I(Distance^2)         1   0.09    0.09    0.451 0.512657
Residuals            14   2.74    0.20
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
summary(aov(Cost ~ Weight + Distance + I(Weight * Distance),
            express))
                     Df Sum Sq Mean Sq F value   Pr(>F)
Weight                1 270.55  270.55  652.59 2.14e-14 ***
Distance              1 143.63  143.63  346.45 2.89e-12 ***
I(Weight * Distance)  1  31.27   31.27   75.42 1.88e-07 ***
Residuals            16   6.63    0.41
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

These are *sequential* sums of squares, adding each term to the model in order.

See `Residuals` line in each set of results:

$$\begin{aligned} \text{SSE(Full second order)} &= 2.74, \\ \text{SSE(Interaction)} &= 6.63, \end{aligned}$$

so

$$F = \frac{(6.63 - 2.74)/2}{0.20} = 9.75, P < .01$$

You can, in fact, calculate F from the output for the full model:

- Note that, because the terms are added sequentially, the sums of squares for the common terms (`Weight`, `Distance`, and `Weight * Distance`) are the same in both models.
- In the reduced model, the extra terms (`Weight^2` and `Distance^2`) have gone away.
- Their *combined* sum of squares, $3.80 + 0.09 = 3.89$, is exactly the increase in SSE, 6.63 - 2.74 $= 3.89$.

So we can also calculate

$$F = \frac{(\text{Sum Sq for } \texttt{Weight\^{} 2} + \text{Sum Sq for } \texttt{Distance\^{} 2})/2}{\text{Mean Square for Residuals}}$$

using only the output for the full model.

Note that F was calculated imprecisely, because of rounding.

We can get more digits using `print(summary(...), digits = 8)` for example, or calculate F to full precision:

```
s = summary(aov(Cost ~ Weight + Distance + I(Weight * Distance) +
                I(Weight^2) + I(Distance^2), express))[[1]]
sum(s[c("I(Weight^2)", "I(Distance^2)"), "Sum Sq"]) / 2 /
    s["Residuals", "Mean Sq"]
```

We could use the same F-test when there is only one additional term in the full model, based on just one line in the ANOVA table, provided it is the last term in the formula.

It appears very different from the t-test described earlier.

Some matrix algebra shows that it is, in fact, exactly the same test:

- The F-statistic is exactly the square of the t-statistic.
- The F critical values are exactly the squares of the (two-sided) t critical values.
- So the P-value is exactly the same.

Complete Example: Road Construction Cost

Data from the Florida Attorney General's office

y = successful bid;

x_1 = DOT engineer's estimate of cost

x_2 = indicator of fixed bidding:

$$x_2 = \begin{cases} 1 & \text{if fixed} \\ 0 & \text{if competitive} \end{cases}$$

Get the data and plot them:

```
flag = read.table("Text/Exercises&Examples/FLAG.txt",
                  header = TRUE)
pairs(flag[,-1])
```

Section 4.14 suggests beginning with the full second order model, and simplifying it as far as possible (but no further!).

We'll take the opposite approach: begin with the first order model, and complicate it as far as necessary.

Because x_2 is a dummy variable, the first order model is a pair of parallel straight lines:

```
summary(lm(COST ~ DOTEST + STATUS, flag))
```

First order model

```
Call:
lm(formula = COST ~ DOTEST + STATUS, data = flag)

Residuals:
     Min       1Q   Median       3Q      Max
-2199.94   -73.83     7.76    53.68  1722.42

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -20.537724  26.817718  -0.766 0.444558
DOTEST        0.930781   0.009744  95.519  < 2e-16 ***
STATUS      166.357224  49.287822   3.375 0.000864 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 306.3 on 232 degrees of freedom
Multiple R-squared: 0.9755, Adjusted R-squared: 0.9752
F-statistic:  4610 on 2 and 232 DF,  p-value: < 2.2e-16
```

Both variables look important.

The DOTEST coefficient is close to 1, so the winning bids roughly track the estimated cost.

The positive STATUS coefficient means the line for STATUS $= 1$ is higher than the line for STATUS $= 0$.

Are the slopes different? Try the interaction model:

```
summary(lm(COST ~ DOTEST * STATUS, flag))
```

Interaction model

```
Call:
lm(formula = COST ~ DOTEST * STATUS, data = flag)

Residuals:
     Min       1Q   Median       3Q      Max
-2143.12   -43.21     1.39    40.17  1765.99

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   -6.428025  26.208287  -0.245    0.806
DOTEST         0.921338   0.009723  94.755  < 2e-16 ***
STATUS        28.673189  58.661711   0.489    0.625
DOTEST:STATUS  0.163282   0.040431   4.039 7.32e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 296.7 on 231 degrees of freedom
Multiple R-squared: 0.9771, Adjusted R-squared: 0.9768
F-statistic:  3281 on 3 and 231 DF,  p-value: < 2.2e-16
```

The interaction term is highly significant: reject the first order model in favor of the interaction model.

The slopes are 0.921338 for $\texttt{STATUS} = 0$, and $0.921338 + 0.163282 = 1.08462$ for $\texttt{STATUS} = 1$.

So the competitive auctions are won with bids that fall slightly below the estimated cost, while the fixed winning bids fall slightly above the estimated cost.

To validate the interaction model, we compare it with (finally!) the full second order model.

Note

When some variables are qualitative, the "full second order model" consists of the full second order (i.e., quadratic) model in the *quantitative* variables, plus the *interactions* of those terms with the *qualititative* variables:

```
summary(lm(COST ~ DOTEST + STATUS + I(DOTEST * STATUS) +
           I(DOTEST^2) + I(DOTEST^2 * STATUS), flag))
```

Full second order model

```
Call:
lm(formula = COST ~ DOTEST + STATUS + I(DOTEST * STATUS) +
             I(DOTEST^2) + I(DOTEST^2 * STATUS), data = flag)

Residuals:
     Min       1Q   Median       3Q      Max
-2143.50   -35.38     1.27    46.58  1771.19

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -2.972e+00  3.089e+01  -0.096  0.92344
DOTEST                9.155e-01  2.917e-02  31.385  < 2e-16 ***
STATUS               -3.673e+01  7.477e+01  -0.491  0.62375
I(DOTEST * STATUS)    3.242e-01  1.192e-01   2.721  0.00702 **
I(DOTEST^2)           7.191e-07  3.404e-06   0.211  0.83288
I(DOTEST^2 * STATUS) -3.576e-05  2.478e-05  -1.443  0.15041
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 296.6 on 229 degrees of freedom
Multiple R-squared: 0.9773, Adjusted R-squared: 0.9768
F-statistic:  1970 on 5 and 229 DF,  p-value: < 2.2e-16
```

Full second order model, ANOVA

```
summary(aov(COST ~ DOTEST + STATUS + I(DOTEST * STATUS) +
                   I(DOTEST^2) + I(DOTEST^2 * STATUS), flag))
                      Df    Sum Sq   Mean Sq  F value   Pr(>F)
DOTEST                 1 864038187 864038187 9818.947  < 2e-16 ***
STATUS                 1   1069006   1069006   12.148 0.000589 ***
I(DOTEST * STATUS)     1   1435733   1435733   16.316 7.32e-05 ***
I(DOTEST^2)            1        15        15    0.000 0.989487
I(DOTEST^2 * STATUS)   1    183210    183210    2.082 0.150411
Residuals            229  20151321     87997
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The significance of the two terms that were added is tested using an F statistic with 2 degrees of freedom in the numerator; the value is only slightly greater than 1, and is completely consistent with the null hypothesis that the interaction model is adequate.