

16.5: The F test as a model comparison

At this point, I want to talk in a little more detail about what the F-tests in an ANOVA are actually doing. In the context of ANOVA, I've been referring to the F-test as a way of testing whether a particular term in the model (e.g., main effect of Factor A) is significant. This interpretation is perfectly valid, but it's not necessarily the most useful way to think about the test. In fact, it's actually a fairly limiting way of thinking about what the F-test does. Consider the clinical trial data we've been working with in this chapter. Suppose I want to see if there are *any* effects of any kind that involve `therapy`. I'm not fussy: I don't care if it's a main effect or an interaction effect.²³⁸ One thing I could do is look at the output for `model.3` earlier: in this model we did see a main effect of therapy ($p=.013$) but we did not see an interaction effect ($p=.125$). That's kind of telling us what we want to know, but it's not quite the same thing. What we really want is a single test that *jointly* checks the main effect of therapy and the interaction effect.

Given the way that I've been describing the ANOVA F-test up to this point, you'd be tempted to think that this isn't possible. On the other hand, if you recall the chapter on regression (in Section 15.10), we were able to use F-tests to make comparisons between a wide variety of regression models. Perhaps something of that sort is possible with ANOVA? And of course, the answer here is yes. The thing that you really need to understand is that the F-test, as it is used in both ANOVA and regression, is really a comparison of *two* statistical models. One of these models is the full model (alternative hypothesis), and the other model is a simpler model that is missing one or more of the terms that the full model includes (null hypothesis). The null model cannot contain any terms that are not in the full model. In the example I gave above, the full model is `model.3`, and it contains a main effect for therapy, a main effect for drug, and the drug by therapy interaction term. The null model would be `model.1` since it contains only the main effect of drug.

test comparing two models

Let's frame this in a slightly more abstract way. We'll say that our full model can be written as an R formula that contains several different terms, say $Y \sim A + B + C + D$. Our null model only contains some subset of these terms, say $Y \sim A + B$. Some of these terms might be main effect terms, others might be interaction terms. It really doesn't matter. The only thing that matters here is that we want to treat some of these terms as the "starting point" (i.e. the terms in the null model, A and B), and we want to see if including the other terms (i.e., C and D) leads to a significant improvement in model performance, over and above what could be achieved by a model that includes only A and B . In essence, we have null and alternative hypotheses that look like this:

Hypothesis	Correct model?	R formula for correct model
Null	M0	$Y \sim A + B$
Alternative	M1	$Y \sim A + B + C + D$

Is there a way of making this comparison directly?

To answer this, let's go back to fundamentals. As we saw in Chapter 14, the F-test is constructed from two kinds of quantity: sums of squares (SS) and degrees of freedom (df). These two things define a mean square value ($MS = SS/df$), and we obtain our F statistic by contrasting the MS value associated with "the thing we're interested in" (the model) with the MS value associated with "everything else" (the residuals). What we want to do is figure out how to talk about the SS value that is associated with the *difference* between two models. It's actually not all that hard to do.

Let's start with the fundamental rule that we used throughout the chapter on regression:

That is, the total sums of squares (i.e., the overall variability of the outcome variable) can be decomposed into two parts: the variability associated with the model SS_M , and the residual or leftover variability, SS_R . However, it's kind of useful to rearrange this equation slightly, and say that the SS value associated with a model is defined like this...

$$SS_M = SS_T - SS_R$$

Now, in our scenario, we have two models: the null model (M0) and the full model (M1):

$$SS_{M0} = SS_T - SS_{R0}$$

$$SS_{M1} = SS_T - SS_{R1}$$

Next, let's think about what it is we *actually* care about here. What we're interested in is the *difference* between the full model and the null model. So, if we want to preserve the idea that what we're doing is an “analysis of the variance” (ANOVA) in the outcome variable, what we should do is define the SS associated with the difference to be equal to the difference in the SS:

$$\begin{aligned} SS_{\Delta} &= SS_{M1} - SS_{M0} \\ &= (SS_T - SS_{R1}) - (SS_T - SS_{R0}) \\ &= SS_{R0} - SS_{R1} \end{aligned}$$

Now that we have our degrees of freedom, we can calculate mean squares and F values in the usual way. Specifically, we're interested in the mean square for the difference between models, and the mean square for the residuals associated with the *full* model (M1), which are given by

$$MS_{\Delta} = \frac{SS_{\Delta}}{df_{\Delta}}$$

$$MS_{R1} = \frac{SS_{R1}}{df_{R1}}$$

Finally, taking the ratio of these two gives us our F statistic:

$$F = \frac{MS_{\Delta}}{MS_{R1}}$$

Running the test in R

At this point, it may help to go back to our concrete example. The null model here is `model.1`, which stipulates that there is a main effect of drug, but no other effects exist. We expressed this via the model formula `mood.gain ~ drug`. The alternative model here is `model.3`, which stipulates that there is a main effect of drug, a main effect of therapy, *and* an interaction. If we express this in the “long” format, this model corresponds to the formula `mood.gain ~ drug + therapy + drug:therapy`, though we often express this using the `*` shorthand. The key thing here is that if we compare `model.1` to `model.3`, we're lumping the main effect of therapy and the interaction term together. Running this test in R is straightforward: we just input both models to the `anova()` function, and it will run the exact F-test that I outlined above.

```
anova( model.1, model.3 )
```

```
## Analysis of Variance Table
##
## Model 1: mood.gain ~ drug
## Model 2: mood.gain ~ drug * therapy
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      15 1.39167
## 2      12 0.65333   3   0.73833 4.5204 0.02424 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's see if we can reproduce this F-test ourselves. Firstly, if you go back and look at the ANOVA tables that we printed out for `model.1` and `model.3` you can reassure yourself that the RSS values printed in this table really do correspond to the residual sum of squares associated with these two models. So let's type them in as variables:

```
ss.res.null <- 1.392
ss.res.full <- 0.653
```

Now, following the procedure that I described above, we will say that the “between model” sum of squares, is the difference between these two residual sum of squares values. So, if we do the subtraction, we discover that the sum of squares associated with those terms that appear in the full model but not the null model is:

```
ss.diff <- ss.res.null - ss.res.full
ss.diff
```

```
## [1] 0.739
```

Right. Next, as always we need to convert these SS values into MS (mean square) values, which we do by dividing by the degrees of freedom. The degrees of freedom associated with the full-model residuals hasn't changed from our original ANOVA for `model.3` : it's the total sample size N , minus the total number of groups G that are relevant to the model. We have 18 people in the trial and 6 possible groups (i.e., 2 therapies \times 3 drugs), so the degrees of freedom here is 12. The degrees of freedom for the null model are calculated similarly. The only difference here is that there are only 3 relevant groups (i.e., 3 drugs), so the degrees of freedom here is 15. And, because the degrees of freedom associated with the difference is equal to the difference in the two degrees of freedom, we arrive at the conclusion that we have $15 - 12 = 3$ degrees of freedom. Now that we know the degrees of freedom, we can calculate our MS values:

```
ms.res <- ss.res.full / 12
ms.diff <- ss.diff / 3
```

Okay, now that we have our two MS values, we can divide one by the other, and obtain an F-statistic ...

```
F.stat <- ms.diff / ms.res
F.stat
```

```
## [1] 4.526799
```

... and, just as we had hoped, this turns out to be identical to the F-statistic that the `anova()` function produced earlier.

This page titled [16.5: The F test as a model comparison](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.