5.05 Analysis of Variance: Factorial ANOVA assumptions and tests

In this video we'll discuss the overall F-test in factorial ANOVA. We'll look at the assumptions, the null and alternative hypothesis, we'll see how to calculate the test statistics for the main and interaction effects, determine the p-values and how to interpret the results. We'll also see how to perform post-hoc comparisons.

Suppose we want to investigate the main effect of raw meat versus canned cat food, the main effect of portion control versus free-feeding and their interaction on cat health, rated on a scale from zero to ten. Cats are randomly assigned to one of four conditions that vary in terms of diet and feeding pattern.

Assumptions

First we need to check the assumptions. These are the same as in one-way ANOVA. The observations should be independent; in our case random assignment takes care of this assumption. If the observations are dependent by design, involving repeated measurements or paired observations we should use repeated measures ANOVA.
Second, the response variable should be normally distributed in each group. The histograms in our example look fairly normal. ANOVA is robust against moderate violation of normality and our samples are large enough with at least ten observations in each group, so we don't have to consider using a non-parametric test.
Thirdly, in ANOVA the variances need to be homogeneous, or the same. In our example, with unequal sample sizes, the largest standard deviation is no more than twice the size of the smallest standard deviation, so we'll proceed.

Statistical hypotheses

We specify statistical hypotheses for each of the main and interaction effects. For each main effect the null hypothesis states that the relevant marginal population means are equal. The alternative hypothesis states that *at least* one population mean differs from the rest. The interaction effect specifies that the difference between means on one factor are the same for each level of the other factor, so the differences in population mean health for raw meat and canned food are expected to be the same for cats fed in controlled portions compared to free-fed cats: $\mu_{rp} - \mu_{cp} = \mu_{rf} - \mu_{cf}$.
If this is the case the effect of diet is not influenced by feeding pattern. The alternative hypothesis is that the differences are unequal for the portion control and free-fed groups.

Test statistic

For each effect the test statistic F equals the between-group variance, divided by the within-group variance:
Diet: $F = \frac{MS_D}{MS_{within}}$, Feeding Pattern: $F = \frac{MS_F}{MS_{within}}$, Interaction: $F = \frac{MS_{DF}}{MS_{within}}$.
The within-group variance is the same for all effects, it's calculated by dividing the within - or error - sum of squares by the total number of observations minus the total number of groups: MS_{within}=$\frac{SS_{within}}{n-d\cdot f}$; the between-group variance is calculated using the sums of squares relevant to the particular effect and dividing by the appropriate degrees of freedom. For main effects this is the number of levels for a particular factors minus one. For interaction effects we take the number of levels for a factor minus one, we do this for all factors and multiply the results.
Diet: $F = \frac{MS_D}{MS_{within}}$=$\frac{\frac{SS_D}{d-1}}{\frac{SS_{within}}{n-d\cdot f}}$, Feeding Pattern: $F = \frac{MS_F}{MS_{within}}$=$\frac{\frac{SS_F}{f-1}}{\frac{SS_{within}}{n-d\cdot f}}$, Interaction:
$F = \frac{MS_{DF}}{MS_{within}}$=$\frac{\frac{SS_{DF}}{(d-1)(f-1)}}{\frac{SS_{within}}{n-d\cdot f}}$.

Each F-test statistic is associated with two degrees of freedom, the *error*, *denominator* or *within* degrees of freedom are the same for all effects. The *numerator* or *between* degrees of freedom depend on the effect. They are the values we used to divide the sums of squares by when we calculated the appropriate between-group variances:

Diet:	df_1= d - 1	df_2= $n - d \cdot f$
Feeding Pattern:	df_1= f - 1	df_2= $n - d \cdot f$
Interaction:	df_1= (d - 1)(f - 1)	df_2= $n - d \cdot f$.

Test statistic distribution and p-value

The test statistics follow an F distribution, with two degrees of freedom. We always start by inspecting the interaction effect. In our example - for the interaction between diet and feeding pattern - we find an F of 4.475 with 1 and 84 degrees of freedom. We look in the right tail of the distribution. With the significance level set at 0.05, using a table we find the critical F-value with 1 and 60 (rounding down from 84) degrees of freedom is 4.0012.
The observed F-value exceeds this value, so we know the p-value is smaller than 0.05. Software provides an exact p-value of 0.0374. We can reject the null hypothesis and conclude that there is an interaction between diet and feeding pattern.

If we perform pairwise tests or confidence intervals for the simple effects - using Tukey's correction for overall-alpha - we see that a raw meat diet shows a significantly higher mean health score than canned food, for both portion controlled and free-fed cats. This effect of diet is *stronger* when cats are free-fed. The effect of raw meat is the same for portion controlled and free-fed cats, however the health rating of cats free-fed on canned food is significantly lower than the health rating of cats fed controlled portions of canned food.

Since the effect of diet appears for both levels of feeding pattern the main effect can be meaningfully interpreted. We've already compared the groups in a pairwise fashion, so interpretation of the main effects isn't really necessary but let's look at them anyway.
For the factor diet we find an F of 36.986 with 1 and 84 degrees of freedom, a very high value - which is due to the fact that I generated the data myself. For feeding pattern we find an F of 3.507 with 1 and 84 degrees of freedom. Using a table we find the critical F-value with 1 and 60 degrees of freedom is also 4.0012 for diet and feeding pattern, since they both happen to have the same numerator degrees of freedom as the interaction term.

The observed F-value exceeds the critical value for the main effect of diet, but not for the main effect of feeding pattern. Software provides an exact p-value of 0.00000003 for diet and 0.0646 for feeding pattern. p-values as tiny as the one for diet are usually reported as smaller than 0.001.

We can reject the null hypothesis for the main effect diet, but not for feeding pattern. When we ignore feeding pattern, the difference between raw meat and canned food is large enough to be significant, with cats fed on raw meat receiving a higher mean health score. When we ignore diet, the difference between portion controlled and free-feeding is not quite large enough to be significant.