OpenIntro online supplement

This material is an online resource of *OpenIntro Statistics*, a textbook available for free in PDF at openintro.org and in paperback for about $10 at amazon.com. This document is licensed to you under a Creative Commons license, and you are welcome to share it with others. For additional details on the license this document is under, see www.openintro.org/rights.php.

ANOVA Calculations

In today's digital world, ANOVA calculations are automated by statistical software. So while you are unlikely to carry out ANOVA by hand, familiarity with the underlying mathematics can be useful as you advance onto more complex statistical methods or when you encounter unique data contexts.

To see how the calculations for ANOVA are done, we can follow the organization of the ANOVA table. Table 1 collects all of the necessary formulas.

	Df	Sum Sq	Mean Sq	F value
Group	$df_G = k - 1$	$SSG = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$	$MSG = \frac{1}{df_G} SSG$	$F = \frac{MSG}{MSE}$
Residuals	$df_E = n - k$	$SSE = \sum_{i=1}^{k} (n_i - 1)s_i^2$	$MSE = \frac{1}{df_E} SSE$	
Total	$df_T = n - 1$	$SST = \sum_{i=1}^{n} (x_i - \bar{x})^2$		

Table 1: Formulas for ANOVA calculations.

With multiple samples, means, variances, etc., there are a lot of symbols involved here, so let's do our best to keep them straight. Reading across the rows of the ANOVA table, we have the symbols:

df degrees of freedom

SS sum of squares

MS mean square

F ratio of variability in the sample means relative to variability within the groups

k number of groups

n total number of observations

n_i number of observations in group i

$\bar{x}_i$ the mean of the n_i observations in group i

$\bar{x}$ the grand mean of all n observations

s_i^2 the variance of the n_i observations in group i

Due to the underlying mathematics, we can also calculate the total sum of squares SST like so:

$$SSG + SSE = SST$$

Before computers became commonplace, statisticians used that fact to double check these complicated calculations.

● **Example 1** College departments commonly run multiple lectures of the same introductory course each semester because of high demand. Consider a statistics department that runs three lectures of an introductory statistics course. We might like

Class A	Class B	Class C
90	87	95
87	85	93
88	80	90
78		88
		85

Table 2: Individual course grades from classes A, B, and C.

to determine whether there are statistically significant differences in first exam scores in these three classes (A, B, and C). Describe appropriate hypotheses to determine whether there are any differences between the three classes.

The hypotheses may be written in the following form:

H_0: The average score is identical in all lectures. Any observed difference is due to chance. Notationally, we write $\mu_A = \mu_B = \mu_C$.

H_A: The average score varies by class. We would reject the null hypothesis in favor of the alternative hypothesis if there were larger differences among the class averages than what we might expect from chance alone.

We'll use the small, hypothetical data set in Table 2 to show how ANOVA calculations work. Let's carefully work through the calculations in Table 1 one at time. We'll work in this order:

- Degrees of freedom for Groups df_G

- Sum of Squares between Groups SSG

- Mean Square between Groups MSG

- Degrees of freedom for Error df_E

- Sum of Squared Error SSE

- Mean Square Error MSE

- Totals df_T and SST

- F Statistic and p-value

Remember to refer back to the list of symbols above as needed to keep these calculations straight. All results below are rounded to the nearest hundredth.[1]

Degrees of freedom for Groups df_G

We calculate degrees of freedom for k groups like so: $df_G = k - 1$. So in our example, we have $k = 3$ groups, and $df_G = 3 - 1 = 2$.

---

[1]To avoid rounding error, a spreadsheet was used for calculations. If you do these calculations "by hand," be sure to hold onto as much accuracy as possible as you work.

Sum of Squares between Groups SSG

Statisticians measure variation using squared differences. The sum of squares for the groups SSG measures the differences between the group means $\bar{x}_i$ and the grand mean $\bar{x}$ like so:

$$SSG = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$$

With exam scores from classes A, B, and C from Table 2, first we calculate the three classes' group means:

$$\bar{x}_A = 85.75 \qquad \bar{x}_B = 84 \qquad \bar{x}_C = 90.2$$

We also calculate the grand mean of all $n = 12$ observations:

$$\bar{x} = 87.17$$

Then we calculate SSG as follows:

$$
\begin{aligned}
SSG &= \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2 \\
&= n_A(\bar{x}_A - \bar{x})^2 + n_B(\bar{x}_B - \bar{x})^2 + n_C(\bar{x}_C - \bar{x})^2 \\
&= 4(85.75 - 87.17)^2 + 3(84.00 - 87.17)^2 + 5(90.20 - 87.17)^2 \\
&= 84.12
\end{aligned}
$$

Mean Square between Groups MSG

The mean square between groups MSG is calculated using both the groups' degrees of freedom df_G and the sum of squares between groups SSG like so:

$$MSG = \frac{1}{df_G} SSG$$

That means that to find the mean square between groups MSG, we simply divide the sum of squares for the groups SSG by the groups' degrees of freedom df_G. In our example we have:

$$MSG = \frac{1}{df_G} SSG = \frac{1}{2} \times 84.12 = 42.06$$

We have completed all of the calculations for the groups, and the MSG provides our measurement of the variation among the sample means. Our goal in ANOVA is to compare that variation between the groups to the variation within the groups. Variation within the groups can be seen as variation among the residuals, i.e. the deviations from the group means. Let's complete the calculations for that error.

Degrees of freedom for Error df_E

Degrees of freedom for residuals are calculated by finding the sum of the degrees of freedom for each of the k samples like so:

$$\begin{aligned}
df_E &= df_1 + df_2 + \ldots + df_k \\
&= (n_1 - 1) + (n_2 - 1) + \ldots + (n_k - 1) \\
&= n - k
\end{aligned}$$

For our example, we have

$$df_E = n - k = 12 - 3 = 9.$$

Sum of Squared Errors SSE

The sum of squared errors SSE is the sum of the squares of the differences between each sample's observations with each respective sample mean. The calculation is simplified by first finding the variance s_i^2 of each sample:

$$s_A^2 = 28.25 \qquad s_B^2 = 13.00 \qquad s_C^2 = 15.70$$

Then we calculate the sum of squared errors like so:

$$SSE = \sum_{i=1}^{k} (n_i - 1)s_i^2$$

For our example, we have:

$$\begin{aligned}
SSE &= \sum_{i=1}^{k} (n_i - 1)s_i^2 \\
&= (n_A - 1)s_A^2 + (n_B - 1)s_B^2 + (n_C - 1)s_C^2 \\
&= (4 - 1)(28.25)^2 + (3 - 1)(13.00)^2 + (5 - 1)(15.70)^2 \\
&= 173.55
\end{aligned}$$

Do you see how this calculation produces a sum of squares? (Consider the effect of multiplying the formula for sample variance by one less than the sample size.)

Mean Square Error MSE

With df_E and MSE already calculated, computing MSE is simple. Like the mean square for the groups (MSG), the mean square error (MSE) is the ratio of the corresponding sum of squares with the degrees of freedom:

$$MSE = \frac{1}{df_E} SSE$$

For our example, we have

$$MSE = \frac{1}{df_E} SSE = \frac{1}{9} \times 173.55 = 19.28$$

We've now completed most of the calculations in the ANOVA Table. Let's turn to the Total row.

Totals df_T and SST

We complete the total degrees of freedom (df_T) and total sum of squares by finding the sums of the corresponding columns. So our total degrees of freedom are

$$df_T = 2 + 9 = 11$$

and the sum of squares total is

$$SST = SSG + SSE = 84.12 + 173.55 = 257.67.$$

The sum of squares total (SST) can also be calculated like so:

$$SST = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Notice that this is the difference of all n observations with the grand mean $\bar{x}$. Calculating SST both as the sum $SSG + SSE$ and as the more complicated calculation above allows us to double check that our ANOVA calculations have been completed correctly. (This was very important back when these calculations were completed by hand.)

We now have a nearly-complete ANOVA table (Table 3 below). All that's left to do is calculate the F statistic and the p-value.

	Df	Sum Sq	Mean Sq	F value
Classes	2	84.12	42.06	
Residuals	9	173.55	19.28	
Total	11	257.67		

Table 3: ANOVA table for course grades from classes A, B, and C.

F Statistic and p-value

We now have comparable measures of variability both between the groups (MSG) and within the groups (MSE). And if variation among group means is simply random, the ratio $F = MSG/MSE$ is typically close to 1.

In our example, we have:

$$F = \frac{MSG}{MSE} = \frac{42.06}{19.28} = 2.18$$

	df	SS	MS	F	P
Between Groups	2	84.1167	42.0583	2.1811	0.1689
Within Groups	9	173.5500	19.2833		
Total	11	257.6667			

Table 5: ANOVA table from GeoGebra.

Knowing degrees of freedom (both between and within groups) and the F statistic, we can use any technology with the F distribution to compute a p-value that communicates the likelihood of observing an F statistic this large if the differences between the observed group means are purely due to chance. There are many technology options—including traditional books of statistical tables, graphing calculators, spreadsheet software, and statistical software.

For our example, we used the F Distribution from online math app GeoGebra's Probability Calculator. With input df1 $= 2$, df2 $= 9$, a right-tailed calculation yields $P(2.18 \leq X) = 0.169$. See Figure 4 for a visual idea of this p-value.
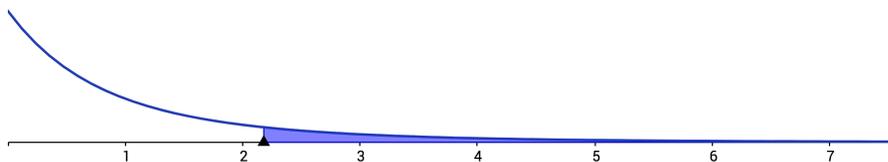

Figure 4: F distribution with degrees of freedom 2 and 9 (graphic from GeoGebra) showing $P(2.18 \leq X) = 0.169$.

The p-value is the final entry in the ANOVA table. Table 5 shows how GeoGebra presents its ANOVA output for our class exam scores example.

You'll notice that, depending on technology choice, the ANOVA table's rows may be labeled differently. The first row may be labeled Groups, Treatments, Between Groups, or Factors. Or it may be labeled to identify which groups' means are being compared. In our example, that would be the label Classes. The second row may be labeled Residuals, Error, Within Groups, or Effect. While the labels may vary, the layout of the ANOVA table is fairly standard.

Whether we compute these values one at a time or use technology to quickly find the p-value, now we can analyze our hypotheses. In our example, if all three classes in fact have the same average 1st exam score, then the likelihood of observing the differences in average scores between the classes is about 17%. That's not unlikely, so the observed differences are not statistically significant.

Exercises

1 Diet plans. Three different diet plans are to be tested for mean weight loss. The entries in the table below are the weight losses for the different plans. Test the hypothesis that all three diet plans have the same mean weight loss by performing ANOVA. [2]

Plan 1	Plan 2	Plan 3
3	3.5	8
4.5	7	4
4	4.5	3
3		

2 Men's weights. Groups of men from three different areas of the country are to be tested for mean weight. The entries in the table are the weights for the different groups. Test the hypothesis that all three groups of men have the same mean weight by performing ANOVA. [3]

Group 1	Group 2	Group 3
216	202	170
198	213	165
240	284	182
187	228	197
176	210	201

3 Class delivery types. Are the means for the final exams the same for all statistics class delivery types? The table below shows the scores on final exams from several randomly selected classes that used the different delivery types. Perform a hypothesis test using ANOVA. [4]

Online	Hybrid	Face-to-face
72	83	80
84	73	78
77	84	84
80	81	81
81		86
		79
		82

Solutions

Inference for numerical data

1   Diet plans. The hypotheses: H_0: All three diet plans result in the same mean weight loss, that is, $\mu_1 = \mu_2 = \mu_3$. H_A: Mean weight loss differs among at least two of the plans.

We can use technology to produce the ANOVA Table 6. Or we can perform ANOVA calculations as follows:

For 3 diet plans, $k = 3$. So $df_G = k - 1 = 3 - 1 = 2$.

Mean weight loss $\bar{x}_i$ for each plan:

$$\bar{x}_1 = \frac{3 + 4.5 + 4 + 3}{4} = 3.625$$

$$\bar{x}_2 = \frac{3.5 + 7 + 4.5}{3} = 5$$

$$\bar{x}_3 = \frac{8 + 4 + 3}{3} = 5.$$

The grand mean $\bar{x}$:

$$\bar{x} = \frac{3 + 4.5 + 4 + 3 + 3.5 + 7 + 4.5 + 8 + 4 + 3}{10} = 4.45$$

So for the sum of squares between the diet plans SSG we have

$$SSG = 4(3.625 - 4.45)^2 + 3(5 - 4.45)^2 + 3(5 - 4.45)^2 = 4.5375.$$

So the mean square between the diet plans MSG is:

$$MSG = \frac{1}{df_G} SSG = \frac{4.5375}{2} = 2.2688$$

With 10 total observations among the 3 diet plans, we have $df_E = n - k = 10 - 3 = 7$. The variance s_i^2 for each plan:

$$s_1^2 = 0.5625 \qquad s_2^2 = 3.25 \qquad s_3^2 = 7$$

So the sum of squared errors SSE is:

$$SSE = (4 - 1)(0.5625) + (3 - 1)(3.25) + (3 - 1)(7) = 22.1875$$

And the mean squared error is:

$$MSG = \frac{1}{df_E} SSE = \frac{22.1875}{7} = 3.1696$$

So we have F statistic:

$$F = \frac{MSG}{MSE} = \frac{2.2688}{3.1696} = 0.7158$$

Whenever the F statistic is less than 1, we should double check the ANOVA assumptions. Which ANOVA assumption(s) is (are) violated here?

For the p-value, we use technology that has the F-distribution $X = F_{2,7}$ to find $P(0.7158 \le X) = 0.5214$. That means that the probability is about 0.52 that we'd observe the variation of

	Df	Sum sq	Mean sq	F value	Pr(>F)
Diet plans	2	4.5375	2.2687	0.5214	0.5214
Residuals	7	22.1875	3.1696		
Total	9	26.725			

Table 6: ANOVA table for mean weight loss.

	Df	Sum sq	Mean sq	F value	Pr(>F)
Regions	2	4939.2	2469.6	3.7416	0.0546
Residuals	12	7920.4	660.0333		
Total	14	12859.6			

Table 7: ANOVA table for mean men's weight.

	Df	Sum sq	Mean sq	F value	Pr(>F)
Regions	2	20.1732	10.0866	0.6388	0.5437
Residuals	13	205.2643	15.7896		
Total	15	225.4375			

Table 8: ANOVA table for mean final exam scores.

mean weight loss between our samples *if* these three diet plans actually produce the same mean weight loss. Such a high p-value throws no doubt on the null hypothesis. We conclude that these data provide no evidence that any of these diet plans produce different mean weight loss.

2 Men's weights.

The hypotheses: H_0: In three different areas of the country, the mean men's weight is the same, that is, $\mu_1 = \mu_2 = \mu_3$. H_A: Mean men's weight differs among at least two regions.

Technology (or careful calculations) produces the ANOVA Table 7.

That means that the probability is about 0.055 that we'd observe the variation of mean weight between our samples *if* the mean weight of men in all three regions was the same. This is a fairly low p-value, but it's not significant at the 0.05 level.

3 Class delivery types.
The hypotheses: H_0: Mean final exam scores are the same for all three class delivery types, that is, $\mu_1 = \mu_2 = \mu_3$. H_A: Mean final exam scores differ for at least two class delivery types.

Technology (or careful calculations) produce the ANOVA Table 8.

Again, the F-statistic is below 1. Which ANOVA assumption(s) was (were) not met?

The p-value 0.5437 says that these data provide no evidence of any difference in mean final exam scores among these three class delivery methods.

References

Mosteller et al, *Beginning Statistics with Data Analysis*

OpenStax, *Introductory Statistics*. OpenStax CNX. Jun 13, 2017 http://cnx.org/contents/30189442-6998-4686-ac05-ed152b91b9de@18.55.