

# ANOVA

Thomas Elliott

January 14, 2014

ANOVA stands for analysis of variance and is one of the basic statistical tests we can use to find relationships between two or more variables. ANOVA compares the group means of a dependent variable where the groups are defined by one or more independent variables. Using ANOVA, we can find out if the group means are statistically significant, but also how much variance in the dependent variable is explained by our independent variables.

## 1 Oneway ANOVA

Oneway ANOVA involves testing the relationship between two and only two variables. Let's consider the example in which income in thousands is our dependent variable and gender, coded male and female, is our independent variable. To do ANOVA, we need to calculate the mean income for each group in our independent variable.

Table 1: Group Means

Group	Mean	N
Male	49.9	95
Female	30.9	105
Total	39.9	200

### 1.1 Sum of Squares

ANOVA involves calculating three different sums of squares. Each also has an associated degrees of freedom ( $df$ ), with  $N$  = number of cases and  $K$  = number of groups. First, the total sum of squares (TSS) is the sum of the square deviations of each case from the grand or total mean:

$$TSS = \sum (Y - \bar{Y}_{grand})^2$$
$$df = N - 1$$

Second, the within sum of squares (WSS), which is also called the residual sum of squares or the error sum of squares, is the sum of the square deviations of each case from their group mean:

$$WSS = \sum (Y - \bar{Y}_{group})^2$$

$$df = N - K$$

Third, the between sum of squares (BSS), which is also called the explained sum of squares, the regression sum of squares, or the model sum of squares, is the sum of the square deviations of the group means from the grand mean for each case.

$$BSS = \sum (\bar{Y}_{group} - \bar{Y}_{grand})^2$$

$$df = K - 1$$

Table 2 shows how to calculate TSS by hand with a sample of the data presented in Table 1. You calculate the deviation of each income from the grand mean by subtracting the grand mean from the individual scores. Then you square the deviation, and then sum the squares. Tables 3 and 4 show how to compute WSS and BSS respectively.

Table 2: Calculating TSS

Case	Gender	Income	Grand Mean	Deviation	Sqr Dev
1	Male	59.3	39.9	19.4	376.36
2	Male	46.5	39.9	6.6	43.56
3	Male	49.7	39.9	9.8	96.04
4	Male	45.7	39.9	5.8	33.64
5	Male	46.9	39.9	7	49
6	Female	32.2	39.9	-7.7	59.29
7	Female	36.5	39.9	-3.4	11.56
8	Female	28.9	39.9	-11	121
9	Female	49.5	39.9	9.6	92.16
10	Female	33.1	39.9	-6.8	46.24
TSS:					928.85

Table 3: Calculating WSS

Case	Gender	Income	Group Mean	Deviation	Sqr Dev
1	Male	59.3	49.9	9.4	88.36
2	Male	46.5	49.9	-3.6	11.56
3	Male	49.7	49.9	-0.2	0.04
4	Male	45.7	49.9	-4.2	17.64
5	Male	46.9	49.9	-3	9
6	Female	32.2	30.9	1.3	1.69
7	Female	36.5	30.9	5.6	31.36
8	Female	28.9	30.9	-2	4
9	Female	49.5	30.9	18.6	345.96
10	Female	33.1	30.9	2.2	4.84
					514.45

Since these tables only use 10 of the 200 cases, these are not the final sums of squares for our ANOVA. The sum of squares for the full sample is in Table 5.

Table 4: Calculating BSS

Case	Gender	Group Mean	Grand Mean	Deviation	Sqr Dev
1	Male	49.9	39.9	10	100
2	Male	49.9	39.9	10	100
3	Male	49.9	39.9	10	100
4	Male	49.9	39.9	10	100
5	Male	49.9	39.9	10	100
6	Female	30.9	39.9	-9	81
7	Female	30.9	39.9	-9	81
8	Female	30.9	39.9	-9	81
9	Female	30.9	39.9	-9	81
10	Female	30.9	39.9	-9	81
					905

Table 5: Sum of Squares

	Sum of Squares	Degrees of Freedom
Between	17929.5	1
Within	20448.3	198
Total	38377.8	199

## 1.2 F-Test

We use these sum of squares to calculate F, our statistic for whether there is a significant difference in group means. F is a ratio of the between mean squares to the within mean squares. Mean squares are themselves a ratio, with the sum of squares divided by the degrees of freedom.

$$F = \frac{\frac{BSS}{K-1}}{\frac{WSS}{N-K}} = \frac{BMS}{WMS}$$

$$F = \frac{\frac{17929.5}{1}}{\frac{20448.3}{198}} = 173.61$$

This gives us the F-statistic, but now we need figure out if this is large enough to be significant. We do that by looking up critical values of F for a given alpha. There are different tables for different levels of alpha, usually 0.10, 0.05, and 0.01. The table in Figure 1 contains the critical values of F for  $\alpha = 0.05$ . Along the columns are values of degrees of freedom for the numerator, or the between sum of squares. Along the rows are values of degrees of freedom for the denominator, or the within sum of squares. Our BSS df is 1, and our WSS df is 198, so we look down the column for 1 df and look for the row for 198 df. In this table, the largest df the rows go to is 120, and then it has a row for infinity, for degrees of freedom that are very large. We can use this row. So our critical value is 3.84. If our F-statistic is larger than the critical value, then we have a significant difference in group means (we are able to reject our null hypothesis that the group means are equal).

Figure 1: Critical Values of F for  $\alpha = 0.05$ 

$v_1$ = degrees of freedom for numerator																					
$v_2$	$v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$	
$v_2$ = degrees of freedom for denominator	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.47	19.48	19.49	19.49	19.50	
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71		
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69		
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67		
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65		
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64		
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62		
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51		
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39		
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25		
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00		

### 1.3 Oneway ANOVA in Stata

To perform oneway ANOVAs in Stata, use the `anova` command followed by your dependent variable then your independent variable:

```
. anova income gender
```

```
Number of obs =      200      R-squared      =  0.4672
Root MSE      = 10.1624      Adj R-squared =  0.4645
```

Source	Partial SS	df	MS	F	Prob > F
-----+-----					
Model	17929.5346	1	17929.5346	173.61	0.0000
gender	17929.5346	1	17929.5346	173.61	0.0000
Residual	20448.2833	198	103.274158		
-----+-----					
Total	38377.8179	199	192.853356		

As you can see, Stata gives us the model (between), residual (within), and total sums of squares, the degrees of freedom, the mean sums of squares, and the F statistic.

## 2 ANOVA with two or more independent variables

We are not restricted to only two variables with ANOVA. We can investigate whether there is a relationship among three or more variables. In other words, we can see if our dependent variable varies according to values of two or more categorical independent variables. When we do this, we are investigating whether variable A has an effect on our dependent variable in the presence of, or controlling for, variable B. Sometimes this is referred to as factorial ANOVA.

Extending our example from before, let's say we add an education variable. It has three categories for arbitrarily low, medium, and high levels of education. The income means and frequencies of each group defined by gender and education are in Table 6.

Table 6: Means and Frequencies of Income

Gender	Low	Medium	High	Total
Female	21.24	36.23	53.14	30.90
	43	57	5	105
Male	25.53	41.18	57.33	49.86
	1	42	52	95
Total	21.34	38.33	56.96	39.90
	44	99	57	200

It looks like there is a relationship between education and income, as well as between sex and income. Often, we'll want to test not only if a relationship exists between the two independent variables and the dependent variable, but if the effect on one variable depends on the value of the other variable. In our example, we could imagine that the effect of education on income might be different depending on whether the case is a man or a woman. In regression analysis, we would say the slopes of income on education differ for men and women. When this happens, we call it an interaction effect. Two independent variables are interacting with one another to produce the dependent variable. We often include interaction terms in factorial ANOVA.

To perform a factorial ANOVA in Stata, we use the `anova` command:

```
. anova income gender educ gender#educ
```

```
Number of obs =      200      R-squared      =  0.8518
Root MSE      =  5.41373      Adj R-squared =  0.8480
```

Source	Partial SS	df	MS	F	Prob > F
Model	32691.9761	5	6538.39521	223.09	0.0000
gender	140.544649	1	140.544649	4.80	0.0297
educ	5411.68372	2	2705.84186	92.32	0.0000
gender#educ	2.50078462	2	1.25039231	0.04	0.9582
Residual	5685.84179	194	29.3084628		
Total	38377.8179	199	192.853356		

Like in `oneway` we type the dependent variable immediately after the command, then a list of our independent variables. Notice what we had to do for our interaction term. In Stata, we specify an interaction term with the pound sign (`#`). We could have also typed the command:

```
anova income gender##educ
```

This would have produced the same results. Typing two pound signs in a row tells Stata include all possible combinations of variables, making it easier to specify three-way interactions (interactions between three, or more, variables). For more information about how Stata works with factorial variables, you can type `help fvvarlist`.

If we look back at our results, we'll find several F-tests. Let's go from bottom to top. The first F-test is for the interaction term. Here the interaction term is testing whether the effect of education on income depends on the value of gender (or whether the effect of gender on income depends on the value of education, both are correct). If we think about the two by two table of means (Table 6), the difference in mean income between men and women is different for each value of education. According to our F-test, this is not the case. We have an incredibly small F statistic. So the effect of education on income is the same for men and women (or the effect of gender on income is the same for all values of education).

The next F-test is for the education term alone. This tests whether the mean income differs among values of education, controlling for gender. We can see from the F-test that this is the case - our F statistic is very large, indicating that the different groups defined by education have different mean incomes. Similarly, the F-test for gender tests whether men and women have different incomes, controlling for education. This F-test is a little less clear. Our F statistic may be significant depending on the value of  $\alpha$  we choose. Finally, we have the F-test for the overall model. This tests whether the three terms taken overall have significant differences in their groups means. Or we can think about it as using gender, education, and an interaction term is better at predicting income than the grand mean of income by itself.

## 2.1 Incremental F-test

What's nice about Stata is that it does a lot of work for us. For example, each of the F-tests for individual terms in the output above are incremental F-tests. Incremental F-tests test the significance of individual terms, or groups of terms, in a model by comparing that model (called the full model) to a model missing the term or terms you want to test the significance of (this smaller model is called the nested model). Normally, it would be pretty tedious for us to calculate those by hand. To see how it works, though, let's calculate one by hand. We have a full, saturated interaction model above. Let's calculate the incremental F-test for the interaction term. We'll need the additive model to do this. Remember, the additive model is model with no interactions.

```
. anova income gender educ
```

```

      Number of obs =      200      R-squared      =  0.8518
      Root MSE      =  5.38722      Adj R-squared =  0.8495

```

Source	Partial SS	df	MS	F	Prob > F
Model	32689.4753	3	10896.4918	375.45	0.0000
gender	688.562334	1	688.562334	23.73	0.0000
educ	14759.9407	2	7379.97035	254.29	0.0000
Residual	5688.34257	196	29.022156		
Total	38377.8179	199	192.853356		

So we have our interaction model with Model SS = 32691.98 and df = 5, and we have our additive model with Model SS = 32689.48 and df = 3. The equation for the incremental F-test is:

$$F = \frac{\frac{\text{Reg SS, full} - \text{Reg SS, nested}}{df, \text{full} - df, \text{nested}}}{\frac{\text{Res SS, full}}{df, \text{full}}}$$

Where the full model is the model with the interaction terms and the nested model is the additive model. All terms in the nested model should appear in the full model. So if we plug in our values:

$$\begin{aligned}
 F &= \frac{\frac{32691.98 - 32689.48}{5 - 3}}{\frac{5685.84}{194}} \\
 &= \frac{2.5}{29.31} \\
 &= 0.0426
 \end{aligned}$$

Which is the same F statistic Stata gave us for the interaction term. When we do an incremental F-test, we are testing the significance of the term(s) we left out of the nested model. Here, we left out the interaction term in the nested model, so that is the F-test we calculated.