25. Two-way ANOVA

The Analysis of Variance seeks to identify sources of variability in data with when the data is partitioned into differentiated groups. In the prior section, we considered two sources of variation that might contribute to the total variability in the sample:

- Variability between groups; and
- Variability within groups.

With inter-rater reliability we considered an additional source of variability by testing to see if the items differentiated between subjects.

In two-way ANOVA, we will have one measurement variable and a nominal variable,with the values of the nominal variable used to divide the sample into groups. The division might, for example, involve the experimental protocol as in the example on alcohol consumption and blood

---

pressure, or it might be the location of the subjects, as in the care facility example. In any case, it is some attribute that is important to the quantitative measurements. The nominal variable is sometimes called a factor.

Two-way ANOVA involves a measurement variable and two factors or nominal variables.

The test for inter-rater reliability we did in the prior section is an example of a special case of two-way ANOVA. In that test, the raters are an obvious category. However, we didn't have a second category to divide up the rows in the sample. This kind of two-way ANOVA is sometimes called Two-factor ANOVA without replication since we don't replicate our sample with another group. For example, if we divided our subjects (the ratees) into male and female groups, we'd have Two-factor ANOVA with replication, which is the topic of this section.

As an example, let's return to the scale that screened for substance abuse.

Now the location where the screen is administered is one factor. The second factor we've added is the gender of the subject. If the researcher gathered data on five females and five males from each setting, the data might look something like the table at right.

	ER	Physician Clinic	DHS Clinic
Males	1	3	1
	2	9	2
	1	7	1
	8	2	2
	1	3	1
Females	2	2	1
	7	3	2
	5	1	2
	3	3	1
	3	2	1

There are several sources of variation in this example. For example, we could disregard location and just test to see if there is difference between genders, or we could disregard gender and just to see if there is difference between locations. Each of these would be a one-way ANOVA. Summary statistics for the columns and rows would look like

---

	Total	ER	Physician	DHS
Count		10	10	10
Sum		33	35	14
Average		3.3	3.5	1.4
Variance		6.455556	6.277778	0.266667

	Males	Females
Count	15	15
Sum	44	38
Average	2.93333	2.53333
Variance	7.495238	2.695238

But the two one-way ANOVAS lose the ways that the factors of gender and location work together to influence the outcomes.

One-way ANOVA calculates the difference between the individual observations in, say, column one and the overall mean:

$$\sum (x_{1,j} - \bar{\bar{x}})^2$$

One-way ANOVA assumes that the total variability comes from two sources: differences within the factors and the differences between the factors. It tests the null hypothesis that all the groups have the same mean against the alternative that they do not.

In two-way ANOVA, we can calculate the row effects of the vertical category by looking at the average data for males and females

males	females
2.933	2.533

Each of these averages are based on a sample of size fifteen and the overall average of all thirty subjects is 2.733. Just as with one-way ANOVA, then, the variability comes from

$$SS_{Rows} = 15 \times ((2.933 - 2.733)^2 + (2.533 - 2.733)^2) = 1.2$$

multiplying by 15 since there are fifteen subjects in each of the male and female samples.

In exactly the same way, we could find the variability due to the locations–the columns–using the overall average of 2.733 and the averages for each column

ER	Physician	DHS
3.3	3.5	1.4

and obtain

$$SS_{Cols} = 10 \times ((3.3 - 2.733)^2 + (3.5 - 2.733)^2 + (1.4 - 2.733)^2)$$
$$= 26.997$$

this time multiplying by ten, the number of subjects in each column.

The above accounts for the variability between locations (columns) and genders (rows), but does not account for the variability within our groups. The data is divided into six groups: three locations for each of two genders.

	ER	Physician Clinic	DHS Clinic
Males	1	3	1
	2	9	2
	1	7	1
	8	2	2
	1	3	1
Females	2	2	1
	7	3	2
	5	1	2
	3	3	1
	3	2	1

That means that we can find the average for each group, and then find the sum of squares within the groups. The calculations are in the table below.

	ER	Physician Clinic	DHS Clinic
Males	37.2	36.8	1.2
Females	16	2.8	1.2

so the total variability within the groups is the sum of the above entries, or 95.2.

We can account for the total variability in exactly the same way as with one-way ANOVA:

$$\sum (x_i - \bar{x})^2 = 143.8663.$$

Thus far we've accounted for

	SS
Rows	1.2
Cols	28.667
Within	95.2

The total we've accounted for so far is 123.2667, leaving

$$20.62 = 143.8663 - 123.2667$$

still not covered. This last variability comes from the interaction between the two factors.

While it's useful to see how the calculations work, Excel has built-in functions to do all of this. It's a little less straightforward to use than the built-in tools in the AnalyzeThis spreadsheet, but not terribly hard.

Before turning to Excel let's summarize our the general setup.
- We have a numeric data and attribute data on our subjects.
- We have two attribute variables which partition the data into groups of rows and columns.
- The groups are all the same size: the number of rows per column is the same for each column, and the number of columns is the same for each row.
- we test hypotheses on the row and column effects:

$$H_0 : \text{ means in column groups are all the same}$$

$$H_0 : \text{ means in row groups are all the same}$$

The alternative in each case is that the means are not the same.

- We also test for the interaction between the factors:

$$H_0 : \text{ there is no interaction}$$

The alternative is that that there is interaction.
The assumptions are:
- All samples are drawn from normally distributed populations;
- all populations have a common variance;
- all samples are drawn independently from each other;
- within each sample, the observations are sampled randomly and independently of each other.

25.1. Example.

The first example returns to the data involving the substance abuse scale given in different settings. We will perform a two a two-way ANOVA. This is with replication since we have categories for each of the factors.

	ER	Physician Clinic	DHS Clinic
Males	1	3	1
	2	9	2
	1	7	1
	8	2	2
Females	1	3	1
	2	2	2
	7	1	2
	5	3	1
	3	2	1

Solution. Once the data analysis tool is installed, it's pretty easy to use. You can find it under the DATA tab in excel. Note that to use it, you must have the same number of subjects in each of the column groups, i.e., a balanced sample with respect to gender, divided equally between the locations.

We'll break the tool's analysis into two parts for convenience.

The first section reports summary statistics for factors. From this we can see that the overall average for the DHS clinic (1.4) is lower than the other two settings (3.3 an 3.5). We can also see differences in scores by gender at the DR and Physician offices, but not at the DHS clinic.

Anova: Two-Factor With Replication

SUMMARY	ER	Physician	DHS Clinic	Total
Males				
Count	5	5	5	15
Sum	13	24	7	44
Average	2.6	4.8	1.4	2.933333
Variance	9.3	9.2	0.3	7.495238
Females				
Count	5	5	5	15
Sum	20	11	7	38
Average	4	2.2	1.4	2.533333
Variance	4	0.7	0.3	2.695238
Total				
Count	10	10	10	
Sum	33	35	14	
Average	3.3	3.5	1.4	
Variance	6.455556	6.277778	0.266667	

This gives the inferential statistics for the row factors (gender), column factors (location) and the interaction between the two factors.

Since all the *p*-values exceed 5%, we reject the relevant hypotheses in each case: there are no statistically significant differences by gender between the locations. Similarly, there are no statistically significant differences by location between the genders. Finally, there is no interaction between location and gender.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	1.2	1	1.2	0.302524	0.587389	4.259677
Columns	26.86667	2	13.43333	3.380555	0.050608	3.402826
Interaction	20.6	2	10.3	2.596639	0.09531	3.402826
Within	95.2	24	3.966667			
Total	143.8667	29				

25.2. Example.

Suppose an HR director has four different training curricula covering the same subject. Each curricula is designed so that it can be given either in a group classroom or as a self-paced course, with the self-paced course available in booklet form or online. Thus there are $4 \times 3 = 12$ different combinations of curricula and delivery. The Director wants to know if there are differences between the curricula, between the delivery methods, and if there is any interaction between the delivery method and the curriculum.

Setting		A	B	C	D
			Training Curriculum		
Self-paced, online		123	128	166	151
		156	150	178	125
		112	174	187	117
		100	116	153	155
		168	109	195	158
Self-paced, written		135	175	140	167
		130	132	145	183
		176	120	159	142
		120	187	131	167
		155	184	126	168
Group Class		156	186	185	175
		180	138	206	173
		147	178	188	154
		146	176	165	191
		193	190	188	169

Solution. Running the data analysis tool gives the summary statistics at right. Notice the averages for the curricula range from a low of 146 to a high of 167, and the averages for the delivery modes range from a low of 146 to a high of 174.

The ANOVA section of the tool's output lets us decide if these differences are statistically significant.

Anova: Two-Factor With Replication

SUMMARY	A	B	C	D	Total
Self-paced, online					
Count	5	5	5	5	20
Sum	659	677	879	706	2921
Average	131.8	135.4	175.8	141.2	146.05
Variance	844.2	707.8	278.7	354.2	782.3658
Self-paced, written					
Count	5	5	5	5	20
Sum	716	798	701	827	3042
Average	143.2	159.6	140.2	165.4	152.1
Variance	498.7	978.3	165.7	217.3	511.0421
Group Class					
Count	5	5	5	5	20
Sum	822	868	932	862	3484
Average	164.4	173.6	186.4	172.4	174.2
Variance	443.3	428.8	212.3	175.8	330.6947
Total					
Count	15	15	15	15	
Sum	2197	2343	2512	2395	
Average	146.4667	156.2	167.4667	159.6667	
Variance	705.8381	871.0286	605.981	404.9524	

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	8782.9	2	4391.45	9.933347	0.000245	3.190727
Columns	3411.65	3	1137.217	2.572355	0.064944	2.798061
Interaction	6225.9	6	1037.65	2.347138	0.045555	2.294601
Within	21220.4	48	442.0917			
Total	39640.85	59				

The analysis shows that the differences in the curricula and in the delivery are statistically significant, and that there is statistically significant interaction between the delivery and the curricula.

■