Activity 7: AxB (Two-Way) ANOVA

Scenario: To test the effectiveness of a new cholesterol drug, you gather 40 subjects and randomly assign half of them to take the drug and the other half to take a placebo. After 6 months, you measure the cholesterol level of each group. Here's the data you obtain:

Placebo	Drug	Total
n = 20	n = 20	N = 40
mean = 98.0	mean = 88.0	mean = 93.0
std. dev = 8.8	std. dev = 8.4	std. dev = 9.887

You calculate the following independent samples t-test:

$$t_{n_1+n_2-2} = \frac{\text{(Observed) - (Hypothesized)}}{\text{(Standard Error)}} = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - 0}{\sqrt{s^2_{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(98-88)}{\sqrt{74}\sqrt{\frac{1}{20}+\frac{1}{20}}} = 3.676$$

where:

$$s^2_{pooled} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(20-1)8.8^2 + (20-1)8.4^2}{20+20-2} = 74$$

and a critical value for t is found to be: $t^{\alpha=0.025}_{df=38} = 2.024$

1) Notice that we could have also chosen to conduct an ANOVA on this data. Based on the results from the t-test, fill-in the blanks in the following ANOVA summary table. What conclusions can we make from either the t-test or the ANOVA?

Source	SS	df	MS	MSR
Drug	1000	________	________	________
Error	2812	________	________	F^1_{38} = ______
Total	3812	________	________	

2) You publish the results of this ground-breaking study in a prestigious journal and receive the following feedback:

I question the results of your study stating that this new drug lowers cholesterol. It is widely known that individuals who are overweight have significantly higher cholesterol levels than individuals who are not overweight. I believe your "placebo" group had more obese individuals than your "drug" group. The "placebo" group's cholesterol level was higher simply because the group contained many obese individuals.

Uh oh - you didn't weigh the subjects in this study. Is this a legitimate concern?

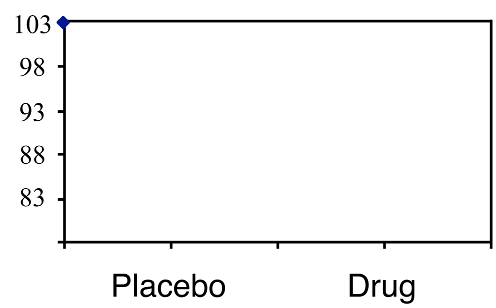
3) We're going to learn how to conduct an ANOVA when we have two independent variables of interest (drug and weight, in this example) and one dependent variable (cholesterol). This is called an AxB ANOVA (also called a two-way ANOVA or a factorial design). To see how this is going to work, it may help to visualize the data in a 2-dimensional table:

	Placebo (D_1)	**Drug (D_2)**	**Total**
Not obese (W_1)	n_{11} $\bar{X}_{11}$	n_{12} $\bar{X}_{12}$	$n_{1\bullet}$ $\bar{X}_{1\bullet}$
Obese (W_2)	n_{21} $\bar{X}_{21}$	n_{22} $\bar{X}_{22}$	$n_{2\bullet}$ $\bar{X}_{2\bullet}$
Total	$n_{\bullet 1}$ $\bar{X}_{\bullet 1}$	$n_{\bullet 2}$ $\bar{X}_{\bullet 2}$	N M

Before we begin worrying about the concepts and calculations behind an AxB ANOVA, let's consider some possible outcomes from our study. For each of the following scenarios, graph the mean cholesterol levels for the non-obese and obese groups. Also, describe the outcome(s) of the study in each scenario:

<u>Scenario A</u>

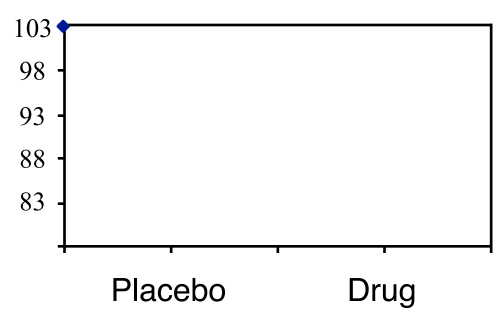
	Placebo (D_1)	**Drug (D_2)**	**Total**
Not obese (W_1)	$n_{11} = 10$ $\bar{X}_{11} = 93$	$n_{12} = 10$ $\bar{X}_{12} = 93$	$n_{1\bullet} = 20$ $\bar{X}_{1\bullet} = 93$
Obese (W_2)	$n_{21} = 10$ $\bar{X}_{21} = 93$	$n_{22} = 10$ $\bar{X}_{22} = 93$	$n_{2\bullet} = 20$ $\bar{X}_{2\bullet} = 93$
Total	$n_{\bullet 1} = 20$ $\bar{X}_{\bullet 1} = 93$	$n_{\bullet 2} = 20$ $\bar{X}_{\bullet 2} = 93$	$N = 40$ $M = 93$



Does the drug impact cholesterol? ____

Does obesity impact cholesterol? ____

<u>Scenario B</u>



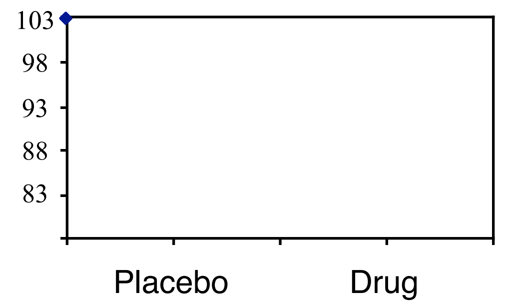
	Placebo (D_1)	**Drug (D_2)**	**Total**
Not obese (W_1)	Mean = 93	Mean = 83	Mean = 88
Obese (W_2)	Mean = 103	Mean = 93	Mean = 98
Total	Mean = 98	Mean = 88	Mean = 93

Does the drug impact cholesterol? ____

Does obesity impact cholesterol? ____

Scenario C

	Placebo (D_1)	Drug (D_2)	Total
Not obese (W_1)	Mean = 88	Mean = 83	Mean = 85.5
Obese (W_2)	Mean = 98	Mean = 88	Mean = 93
Total	Mean = 93	Mean = 85.5	Mean=89.25

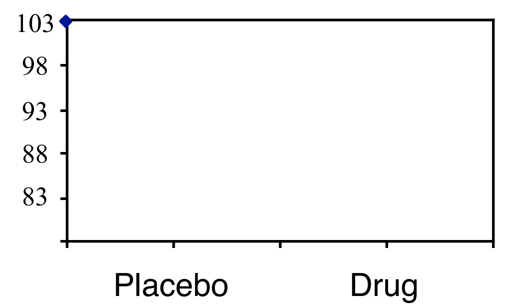


Does the drug impact cholesterol? ____

Does obesity impact cholesterol? ____

Scenario D

	Placebo (D_1)	Drug (D_2)	Total
Not obese (W_1)	Mean = 88	Mean = 98	Mean = 93
Obese (W_2)	Mean = 98	Mean = 88	Mean = 93
Total	Mean = 93	Mean = 93	Mean = 93



Does the drug impact cholesterol? ____

Does obesity impact cholesterol? ____

In an AxB ANOVA, we will analyze the differences (distances) among the means of the row variables and the differences among the means of the column variables separately. These two sets of differences are called *main effects*. Calculate the main effects (and the sum of those main effects) for each scenario:

Scenario A: Drug main effect = Obesity main effect = Sum =

Scenario B: Drug main effect = Obesity main effect = Sum =

Scenario C: Drug main effect = Obesity main effect = Sum =

Scenario D: Drug main effect = Obesity main effect = Sum =

4) In addition to main effects, an AxB ANOVA allows us to investigate the *interaction* between our two independent variables. Interaction is present when the two independent variables in combination produce effects that are different from the sum of their main effects.

Take Scenario B, for example. We found the drug main effect was equal to 5 points (that's the average distance from the drug and placebo means to the overall mean). We also found the obesity main effect to equal 5 points. The sum of the effects is, thus, 10. Now let's look at the combination of the drug and obesity effects in the study. Suppose we have an average person (with a cholesterol level of 93, the grand mean). If that person becomes non-obese **and** takes the drug, we expect that person to have a cholesterol of 83. So, in combination, the drug and obesity effects equal 10 (which is what we might expect). This is an example of a scenario with **no interaction** -- we can tell because the graph of means shows parallel lines.

Now look again at Scenario D. In this scenario, we found no drug or obesity main effects. However, when you take an average person (with a cholesterol of 93) and that person becomes non-obese **and** takes the drug, we expect that person to have a cholesterol of 98. So even though the drug and obesity effects, taken separately, should not change that person's cholesterol, the combination of those effects does have an impact. This is an example of interaction -- we can tell because the graph of means shows nonparallel lines.

A more subtle form of interaction is found in Scenario C. Here we find main effects of 3.75 (for Drug) and 3.75 (for Obesity). Summed, we would expect a total effect of 7.5 points. Our average person with a cholesterol of 89.25 drops down to 83 (only 6.25 points) with non-obesity and the drug. Once again, the drug and obesity effects in combination differ from their simple sum, and we can see this with the non-parallel lines in our graph.

What does the presence of interaction mean? We'll investigate this more in a bit. For now, we'll just mention:

Interaction means the effect of one independent variable (on the dependent var.) depends on the other independent var.

Procedurally, an AxB ANOVA is similar to the one-way ANOVA we've been studying. In a one-way ANOVA, we partitioned the total variation into two components: a between-groups component and a within-groups (error) component. We then took the ratio of those two estimates of variance and compared it to an F-distribution.

In an AxB ANOVA, we're going to partition the total variation into between-groups, within-groups, and interaction components. Before we do that, let's investigate the formal model for an AxB ANOVA:

Pretend you're an individual in this cholesterol study. What factors could influence your cholesterol score? Which of these factors are of interest to the researcher?

An individual's cholesterol = __________________ + __________________ + __________________ + __________________

We'll write the formal model as: $x_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}$ where

$$\alpha_j = \mu_{Aj} - \mu$$
$$\beta_k = \mu_{Bk} - \mu$$
$$\alpha\beta_{jk} = \mu_{AjBk} - \mu_{Aj} - \mu_{Bk} + \mu$$

5) The table below displays the actual data from this cholesterol study. As you can see, the first individual in the study had a cholesterol level of 85.08. According to our formal model, why did this individual have this cholesterol level? What are the values of α, β, αβ, and ε?

	Placebo (D_1)		Drug (D_2)		Total
Not obese (W_1)	**85.08**	92.68	80.11	73.14	Overall Non-Obese
	90.60	89.86	95.13	75.59	20 subjects
	76.85	77.63	**79.20**	67.83	Avg. Cholesterol = 85.1
	90.24	107.21	70.05	91.61	St. Deviation = 9.7
	91.38	90.24	89.25	87.96	
	Mean = 89.2 StDev = 8.5		Mean = 81.0 StDev = 9.5		
Obese (W_2)	103.13	95.17	92.97	101.78	Overall Obese
	116.50	100.63	81.51	90.67	20 subjects
	102.87	115.14	86.53	90.39	Avg. Cholesterol = 97.8
	101.10	100.16	83.55	102.52	St. Deviation = 10.0
	109.19	94.27	104.24	82.86	
	Mean = 103.8 StDev = 7.6		Mean = 91.7 StDev = 8.5		
Total	20 subjects Avg. Cholesterol = 96.5 St. Deviation = 10.8		20 subjects Avg. Cholesterol = 86.3 St. Deviation = 10.4		40 subjects Avg. Cholesterol = 91.4 St. Deviation = 11.7

Individual Score X_{ijk}	Overall Mean μ	Drug effect $\alpha_j = \mu_{Aj} - \mu$	Obesity effect $\beta_k = \mu_{Bk} - \mu$	Interaction $\alpha\beta_{jk} = \mu_{AjBk} - \mu_{Aj} - \mu_{Bk} + \mu$	Error ε
85.08					
79.20					
109.19					

So, all Greek letters aside, our formal model is simply stating our expectations for the factors that may attribute to variation in cholesterol levels.

6) Assuming the model makes sense, let's begin partitioning the variation in cholesterol levels. Explain each step of the derivation, paying particular attention to any assumptions we're making along the way:

We start with this tautology. What makes it true?

$$X_{ijk} = M + \left(\overline{X}_{Aj} - M\right) + \left(\overline{X}_{Bk} - M\right) + \left(\overline{X}_{AjBk} - \overline{X}_{Aj} - \overline{X}_{Bk} + M\right) + \left(X_{ijk} - \overline{X}_{AjBk}\right)$$

$$\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(X_{ijk} - M\right) = \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(\overline{X}_{Aj} - M\right) + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(\overline{X}_{Bk} - M\right) + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(\overline{X}_{AjBk} - \overline{X}_{Aj} - \overline{X}_{Bk} + M\right) + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(X_{ijk} - \overline{X}_{AjBk}\right)$$

$$\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(X_{ijk} - M\right)^2 = \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(\overline{X}_{Aj} - M\right)^2 + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(\overline{X}_{Bk} - M\right)^2 + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(\overline{X}_{AjBk} - \overline{X}_{Aj} - \overline{X}_{Bk} + M\right)^2 + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n_{ajbk}}\left(X_{ijk} - \overline{X}_{AjBk}\right)^2$$

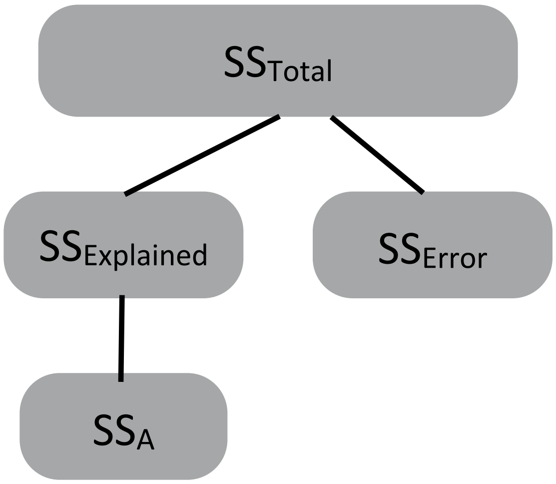
$$SS_{\text{Total}} = SS_{\text{A}} + SS_{\text{B}} + SS_{\text{AxB}} + SS_{\text{Error}}$$

Explain what these sums of squares represent.

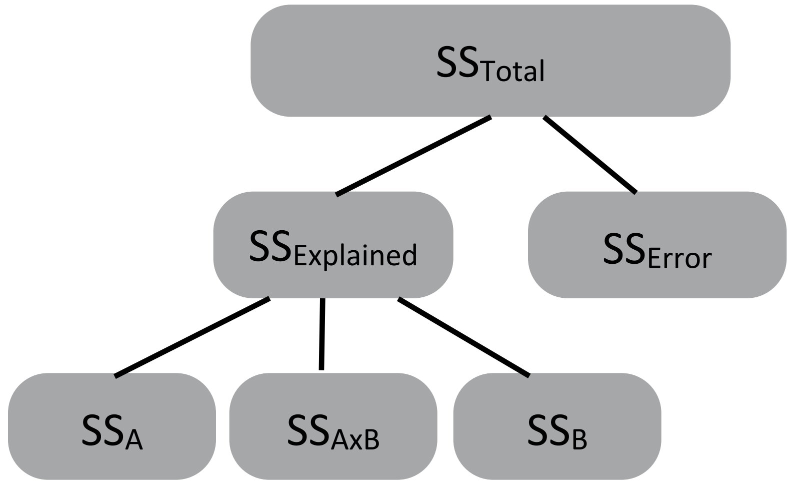
7) We will convert these SS values into MS values by dividing each component by its degrees of freedom. We'll then take ratios of these MS values to compare them. What ratio of mean squares will we calculate to determine if the drug had a significant effect on cholesterol? What about to determine the significance of the obesity effect? How will we test for interaction? Which of these MS ratios will we want to calculate first? Why?

8) In a visual form, here's how a one-way and an AxB ANOVA partition variance:

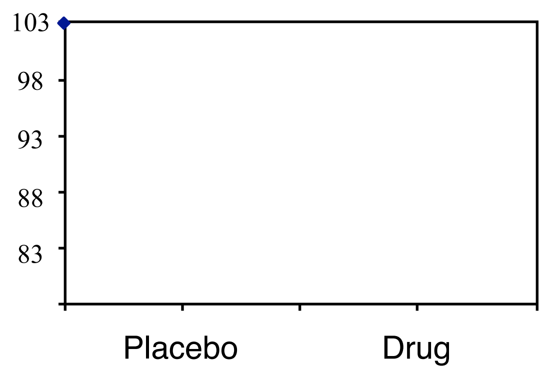
one-way ANOVA



AxB ANOVA



9) Before we get into the formulas to calculate SS, df, and MS values, let's take a quick moment to graph the means in our study. This graph will give us an idea of which effects we should expect to be significant. Based on the graph, do you expect to find a significant drug effect? obesity effect? interaction effect? Explain.



Source	SS	df	MS	MSR
A (Drug)	$\sum n_{Aj}\left(\bar{X}_{Aj} - M\right)^2$	a – 1	SS / df	MS_A / MS_E
B (Obesity)	$\sum n_{Bk}\left(\bar{X}_{Bk} - M\right)^2$	b – 1	SS / df	MS_B / MS_E
AxB (Interaction)	$\sum\sum n_{AjBk}\left(\bar{X}_{AjBk} - \bar{X}_{Aj} - \bar{X}_{Bk} + M\right)^2$	(a – 1)(b – 1)	SS / df	MS_{AxB} / MS_E
Error	$\sum\sum\sum\left(X_{ijk} - \bar{X}_{AjBk}\right)^2 = \sum\left(n_{AjBk} - 1\right)s^2_{AjBk}$	N – ab	SS / df	
Total	$SS_{\text{A}} + SS_{\text{B}} + SS_{\text{AxB}} + SS_{\text{Error}} = \sum\left(X_{ijk} - M\right)^2$	N – 1	(total variance of all N observations)	

10) I had Stata compute everything for me in the following table. Verify the calculations and interpret the output. How would we determine which MSR values are significant? What conclusions can we make from this study? Calculate an effect size.

Source	SS	df	MS	MSR	p-value
A (Drug)	1030.784	1	1030.784	14.063	0.001
B (Obesity)	1606.936	1	1606.936	21.923	<.001
AxB (Interaction)	38.426	1	38.426	0.524	0.474
Error	2638.774	36	73.299		
Total	5314.919	39			

At this point, you should go to Activity #9 and complete exercises #1 (cows) and #2 (haircuts).