Chapter 15: Two-Factor ANOVA (Independent Measures)

To make the transition between one-way designs and two-way designs, let's start with a one-way design and then extend it to a two-way design. Suppose that you are interested in the effects of a particular study strategy on memory for verbal information. You decide to use two different study strategies: Repetition and Imagery. Your participants are shown a list of 30 words, one at a time. One half of your participants is told to repeat each word over and over as it appears. The other half of your participants is told to create an image of each word as it appears. [The independent variable, or factor, would then be the study strategy.] After presenting the list, you provide a distractor task (e.g., count backward from 571 by threes), then ask the participants to write down as many of the words as they can remember. [The dependent variable would then be the number of words recalled.]

For this factor, your null hypothesis would be: H_0: $\mu_{Repetition} = \mu_{Imagery}$

Compute a one-way ANOVA on these data:

	Repetition	**Imagery**	**Sum (T)**	**SS**
Males	14 13 10 11 17	19 24 25 20 22	175	258.47
Females	13 15 12 14 16	19 23 24 25 24	185	234.46
Sum (*T*)	135	225	360 (*G*)	$\Sigma X^2 = 6978$
SS	42.5	50.5		

Source	***SS***	***df***	***MS***	***F***
Strategy				
Within (Error)				
Total				

Now, suppose that you included an equal number of men and women in the experiment. In fact, the first 5 participants in each group were males and the second 5 participants were females. You could now reanalyze the data as a one-way ANOVA to look at the impact of gender. Thus, you would ignore the effects of strategy and analyze only for the impact of gender. Because the data are the same, what must be true about SS_{Total} and df_{Total}? For this factor (independent variable) you would again have two levels (Male and Female). Thus,

H_0: $\mu_{Male} = \mu_{Female}$

	Repetition	Imagery	**Sum (*T*)**	***SS***
Males	14 13 10 11 17	19 24 25 20 22	175	258.47
Females	13 15 12 14 16	19 23 24 25 24	185	234.46
Sum (T)	135	225	360 (*G*)	$\Sigma X^2 = 6978$
SS	42.5	50.5		

Source	***SS***	***df***	***MS***	***F***
Gender				
Within (Error)				
Total				

The major change from computing the two separate one-way ANOVAs to computing the two-way ANOVA is in the computation of the Within (Error) Term. Because we want the Error Term to be based on the variability among participants who are treated alike (so that the only sources of variability are individual differences and random variability), we need the *SS* for the smallest groups created by the experiment. In fact, you might want to think about this experiment as a one-way ANOVA on a single factor with four levels (Male/Repetition, Male/Imagery, Female/Repetition, and Female/Imagery). Gravetter and Wallnau refer to this variability as Between-Treatments. Thought about in this way, your summary data and source table might look like this:

	Male/Repetition	Male/Imagery	Female/Repetition	Female/Imagery	Sum
ΣX or *T*	65	110	70	115	360 (*G*)
SS	30	26	10	22	88

Source	***SS***	***df***	***MS***	***F***
Between	410	3	136.67	24.85
Within (Error)	88	16	5.5	
Total	498	19		

Okay, now we can think about computing a two-way ANOVA on the same data (as a 2x2 independent groups design). Instead of lumping our two factors together as a single factor (as I did above), we want to assess the independent effects of both factors, which we refer to as *main effects*. In addition, we will be able to assess the interactive effect of the two factors. For the two-way ANOVA, we will have three H_0's.

H_0: $\mu_{\text{Repetition}} = \mu_{\text{Imagery}}$

H_0: $\mu_{\text{Male}} = \mu_{\text{Female}}$

H_0: No Interaction

First of all, note that the way we will assess the two main effects is to compute a *MS* for the treatments and divide that *MS* by the MS_{Error}. The computation of the *MS* for the treatment is identical to the computation of $MS_{Treatment}$ for the one-way ANOVA. That is, you would compute the $MS_{Strategy}$ in exactly the same way that you did at the beginning of this handout. Then you would compute the MS_{Gender} in exactly the same way that you did earlier. You would compute MS_{Within} exactly as you did just above, using each of the conditions separately to estimate the population variance (σ^2), and then averaging over the four sample variances (s^2). That is, you are still pooling the separate condition variances in an effort to estimate the population variance (which is due to individual differences and random variability). Thus, the only "new" computation is for the interaction effect.

To best assess these effects, you should restructure the original data as in the table below:

	Repetition	Imagery	Marginal
Male	Sum = 65 *SS* = 30	Sum = 110 *SS* = 26	Sum (*T*) = 175
Female	Sum = 70 *SS* = 10	Sum = 115 *SS* = 22	Sum (*T*) = 185
Marginal	Sum (*T*) = 135	Sum (*T*) = 225	Sum (*G*) = 360

From this table, we can now compute the values for the source table for the two-way ANOVA.

$$SS_{Strategy} = \left(\frac{T^2_{Repetition}}{n_{Repetition}} + \frac{T^2_{Imagery}}{n_{Imagery}}\right) - \frac{G^2}{N} = \left(\frac{135^2}{10} + \frac{225^2}{10}\right) - \frac{360^2}{20} = 6885 - 6480 = 405$$

$$SS_{Gender} = \left(\frac{T^2_{Male}}{n_{Male}} + \frac{T^2_{Female}}{n_{Female}}\right) - \frac{G^2}{N} = \left(\frac{175^2}{10} + \frac{185^2}{10}\right) - \frac{360^2}{20} = 6485 - 6480 = 5$$

$$SS_{Error} = SS_{Male/Repetition} + SS_{Male/Imagery} + SS_{Female/Repetition} + SS_{Female/Imagery} = 30 + 26 + 10 + 22 = 88$$

$$SS_{Total} = \sum X^2 - \frac{G^2}{N} = 6978 - \frac{360^2}{20} = 498$$

Unfortunately, for the purposes of checking your math, there is no separate way to compute $SS_{Interaction}$. Instead, you simply add the *SS* for the two main effects and for error and then subtract that sum from SS_{Total}.

$$SS_{SxG} = SS_{Total} - \left(SS_{Strategy} + SS_{Gender} + SS_{Error}\right) = 498 - (405 + 5 + 88) = 0$$

The degrees of freedom are fairly easy to compute, because they follow closely what you've learned for the one-way ANOVA. That is:

df_{Total} = Total number of scores – 1 = 20 – 1 = 19

$df_{Strategy}$ = Total number of levels of strategy – 1 = 2 – 1 = 1

df_{Gender} = Total number of levels of gender – 1 = 2 – 1 = 1

$df_{SxG} = df_{Strategy} * df_{Gender} = 1 * 1 = 1$

df_{Error} = (Number of scores per condition – 1) * Number of conditions = 4 * 4 = 16

The computation of MS is straightforward as well:

$MS_{Strategy} = SS_{Strategy} / df_{Strategy}$

$MS_{Gender} = SS_{Gender} / df_{Gender}$

$MS_{SxG} = SS_{SxG} / df_{SxG}$

$MS_{Error} = SS_{Error} / df_{Error}$

With three null hypotheses, you'll be computing three F-ratios. In each case, the denominator will be MS_{Error}:

F	H_0	What's being compared
$F_{Strategy} = MS_{Strategy} / MS_{Error}$	$\mu_{\text{Repetition}} = \mu_{\text{Imagery}}$	$M_{\text{Repetition}}$ and M_{Imagery}
$F_{Gender} = MS_{Gender} / MS_{Error}$	$\mu_{\text{Male}} = \mu_{\text{Female}}$	M_{Male} and M_{Female}
$F_{SxG} = MS_{SxG} / MS_{Error}$	No interaction in population	Cell means

The source table would look like this:

Source	***SS***	***df***	***MS***	***F***
Strategy	405	1	405	73.64
Gender	5	1	5	.91
Strategy x Gender	0	1	0	0
Error	88	16	5.5	
Total	498	19		

Because this is an independent groups design, we would once again be interested in determining whether or not we had violated the homogeneity of variance assumption. That is, we need to compute F_{Max} and compare that value to $F_{Max\ Critical}$. When we have some concerns about heterogeneity of variance, we would evaluate our three F-ratios using $\alpha = .01$ instead of $\alpha = .05$.

In this example, the largest variance would be 7.5 and the smallest variance would be 2.5, so $F_{Max} = 3$. With four conditions and $(n - 1) = 4$, $F_{Max\ Critical}$ would be 20.6, so we wouldn't be concerned about heterogeneity of variance and we would use $\alpha = .05$ for each H_0.

For this particular analysis, the F-ratios for each of our null hypotheses would be evaluated with the same $F_{Crit}(1,16) = 4.49$. The particular F_{Crit} would be determined by the *df* associated with the effect (main effect or interaction) and the *df* associated with the error term. For each of our effects in this study the *df* would be 1, so the F_{Crit} is always the same.

What, then, would you decide about the two main effects and the interaction in this study?

Effect	**Decision**
Main effect for Strategy	
Main effect for Gender	
Interaction between Strategy and Gender	

Because there are only two levels to each main effect, no post hoc test is necessary. Of course, that will not always be the case, so you will often need to conduct post hoc analyses to allow you to interpret the main effects or the interaction.

In this particular case, of course, there is no significant interaction between Strategy and Gender ($F_{Obt} < F_{Crit}$). In fact, the $F_{SxG} = 0$. It's rare to have an interaction F of 0, but that tells you that there is not even the hint of an interaction. On some occasions, you may obtain a small (and non-significant) F for your interaction. But what does it mean to say that you have a significant interaction?

Here are a few ways of defining an interaction:

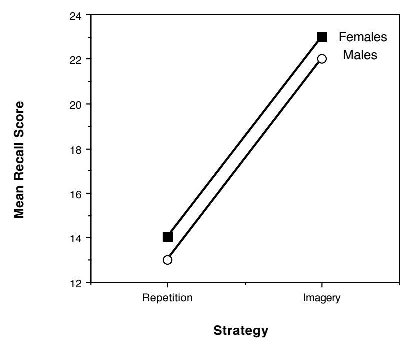
An interaction between two factors occurs whenever the mean differences between individual treatment conditions, or cells, are different from what would be predicted from the overall main effects of the factors.

When the effect of one factor depends on the different levels of a second factor, then there is an interaction between the two factors.

An interaction occurs when the effects of one factor are not the same at all levels of the other factor.

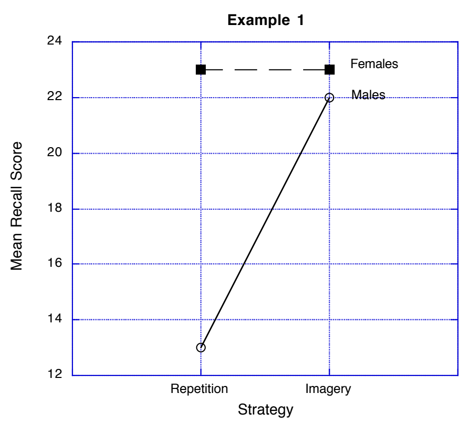
When the results of a two-factor study are presented in a graph, the existence of nonparallel lines (lines that cross over or converge) indicates an interaction between the two factors.

A graph of our data would look like this:

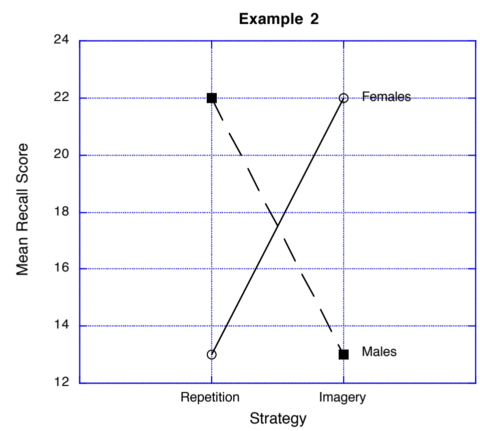


As illustrated in the figure above, the lines are perfectly parallel, which means that there is no interaction. (It is quite rare to have a situation like this one, where the lines are perfectly parallel. just as it's quite rare to have an interaction $F = 0$.) When the lines are not parallel, you may have an interaction (depending on the size of your F-ratio). For this particular set of results, the lack of an interaction means that males and females show a similar benefit for imagery over repetition. How would you interpret the results of the study? Keep in mind, of course, that you are not manipulating the gender of the participants.

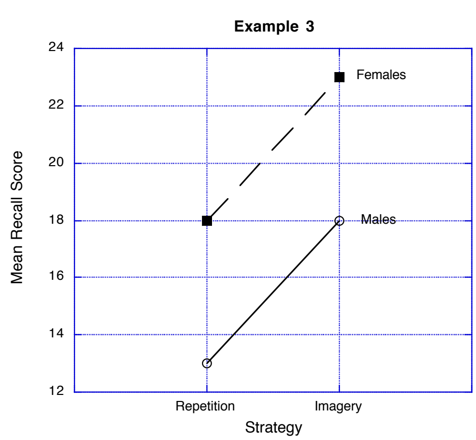
For the examples below, what would you predict about the presence of main effects and interactions in the source table?



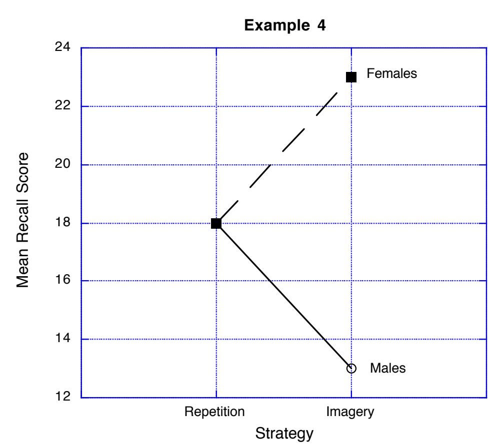
ME Strategy:
ME Gender:
Strat x Gen:



ME Strategy:
ME Gender:
Strat x Gen:



ME Strategy:
ME Gender:
Strat x Gen:



ME Strategy:
ME Gender:
Strat x Gen:

Effect Size

Just as you need to test three separate null hypotheses, you will also need to estimate three different effect sizes. Again, you will use η^2 as an index of effect size. In general the formula will be:

$$\eta^2 = \frac{SS_{Effect}}{SS_{Effect} + SS_{Within.Treatments}}$$

Thus, to assess the effect size for the main effect of Factor A:

$$\eta^2 = \frac{SS_A}{SS_A + SS_{Within.Treatments}} = \frac{SS_A}{SS_{Total} - SS_B - SS_{AxB}}$$

For the main effect of Factor B:

$$\eta^2 = \frac{SS_B}{SS_B + SS_{Within.Treatments}} = \frac{SS_B}{SS_{Total} - SS_A - SS_{AxB}}$$

For the interaction:

$$\eta^2 = \frac{SS_{AxB}}{SS_{AxB} + SS_{Within.Treatments}} = \frac{SS_{AxB}}{SS_{Total} - SS_A - SS_B}$$

In general, you are going to be most interested in estimating the effect size for the interaction.

For the example that we've been using, here are the estimates of the three effect sizes:

The effect size for the main effect of strategy would be:

$$\eta^2 = \frac{SS_{Strategy}}{SS_{Strategy} + SS_{Within.Treatments}} = \frac{SS_{Strategy}}{SS_{Total} - SS_{Gender} - SS_{SxG}} = \frac{405}{498 - 55 - 0} = .91$$

The effect size for the main effect of gender would be:

$$\eta^2 = \frac{SS_{Gender}}{SS_{Gender} + SS_{Within.Treatments}} = \frac{SS_{Gender}}{SS_{Total} - SS_{Strategy} - SS_{SxG}} = \frac{55}{498 - 405 - 0} = .59$$

The effect size for the interaction would be:

$$\eta^2 = \frac{SS_{SxG}}{SS_{SxG} + SS_{Within.Treatments}} = \frac{SS_{SxG}}{SS_{Total} - SS_{Strategy} - SS_{Gender}} = \frac{0}{498 - 405 - 55} = 0$$

Given the *F*-ratio of 0 for the interaction, it should be no surprise that the effect size is 0.

Here's another example of a 2x2 design. Suppose that you gave participants a test of self-esteem and divided your group into people with Low or High self-esteem (IV_1). Then you had each of your participants give a speech either Alone or in front of an Audience (IV_2). The dependent variable that you use is the number of errors made by the speaker. Analyze these data as completely as you can.

	Low Self Esteem		High Self Esteem		
	Alone	Audience	Alone	Audience	
	7 7 2 6 8 6	10 14 11 15 11 11	3 6 2 2 4 7	9 4 2 5 4 6	SUM
AB	36	72	24	36	168
SS	22	20	22	22	86
$\bar{X}$	6	12	4	6	
s^2	4.4	4	4.4	4.4	
ΣX^2	239	884	118	238	1478

	Alone	Audience	Marginal
High Self-Esteem	24	36	(*T*) 60
Low Self-Esteem	36	72	(*T*) 108
Marginal	(*T*) 60	(*T*) 108	(*G*) 168

Source	***SS***	***df***	***MS***	***F***
Self Esteem				
Audience				
SE x Aud				
Error				
Total				

Example 15.2 (G&W5) from Schacter

This example is derived from some work by Schacter (1968). The two “factors” were Weight (Normal vs. Obese) [which is actually a non-manipulated characteristic of the participant] and Fullness (half the people were given a full meal and half were left hungry). The participants are asked to taste and rate five different types of crackers. The DV is the number of crackers eaten.

The researchers were predicting an interaction. That is, they predicted that Obese participants would eat the same number of crackers regardless of fullness. On the other hand, they predicted that Normal participants would eat more crackers if hungry and fewer crackers if full.

Complete the analysis of these data and indicate if they are consistent with the predictions.

	Empty Stomach	Full Stomach	
Normal	$n = 20$ $\overline{X} = 22$ $T = 440$ $SS = 1540$	$n = 20$ $\overline{X} = 15$ $T = 300$ $SS = 1270$	$T = 740$
Obese	$n = 20$ $\overline{X} = 17$ $T = 340$ $SS = 1320$	$n = 20$ $\overline{X} = 18$ $T = 360$ $SS = 1266$	T = 700
	$T = 780$	$T = 660$	$G = 1440$

$\Sigma X^2 = 31836$ $N = 80$

Source	*SS*	*df*	*MS*	*F*
Weight (N vs. O)				
Fullness (E vs. F)				
Weight x Full				
Error				
Total				

A researcher was interested in the impact of a particular drug (Smart-O) on rats' performance in a maze. She decided to run an independent groups design, comparing Smart-O with a placebo. She also thought that the type of maze (simple vs. complex) might have an impact, so she introduced this second factor into the design — producing a 2x2 independent groups design. Her budget was pretty flush, so she decided to run 25 rats in each condition. She chose to use the number of errors the rats made (going down blind alleys) as her dependent variable. On completion of the study, she ran an analysis of the data, but absent-mindedly left her output where the rats could get to in and they nibbled away parts of the source table. As her research assistant, you are not the least bit perturbed, because you can generate the missing parts easily from the remaining numbers (right??). Do so now.

Source	***SS***	***df***	***MS***	***F***
Drug (D vs. P)	10			
Maze (S vs. C)	20			
Drug x Maze				
Error	192			
Total	262			

Dr. Smith was interested in the effects of different levels of a drug (Polypropahexadent) on performance of rats in a maze. The dependent variable used by Dr. Smith was the number of trials to learn the maze, so smaller numbers indicate increased performance. Dr. Smith was also interested in the extent to which the degree to which the rats were hungry would influence their performance. So Dr. Smith conducted a two-factor independent groups experiment in which both factors were manipulated. Complete the source table below and then answer the questions beneath the source table.

Source	***SS***	***df***	***MS***	***F***
Drug	6			1.0
Hunger	40			
Drug x Hunger		12		10.0
Error			2	
Total	966	359		

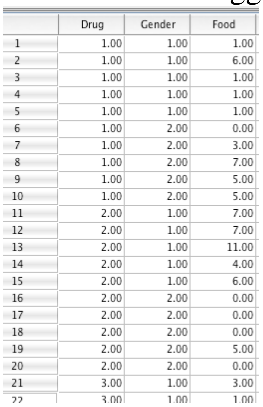
How many levels of the Drug factor were used?
How many levels of the Hunger factor were used?
Assuming an equal number of rats per condition, how many rats were in each condition?
Does it appear that Drug had an influence on performance in the maze? Why? (Careful!)

Dr. Mo Shun was interested in the impact of various dosages of a new drug (*Stay Put*) on the activity level of hyperactive children. She is fairly sure that, because of its chemical nature, *Stay Put* will be more effective for males than for females. To that end, she administers four dosage levels (None, Low, Medium, High) of *Stay Put* to an equal number of male and female children who exhibit similar levels of hyperactivity. The dependent variable is an activity measure, with higher numbers indicating greater activity. Analyze and interpret these data as completely as you can. {Johnson}

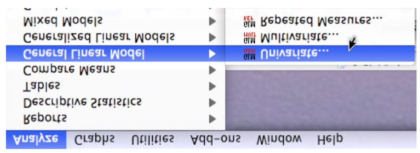
	Males				**Females**				
	None	**Low**	**Med**	**High**	**None**	**Low**	**Med**	**High**	
	10	8	4	3	12	9	3	5	
	11	7	3	4	8	6	6	2	
	8	10	5	5	10	7	5	3	
	7	9	7	2	9	5	2	1	
	12	8	6	7	7	6	3	2	
	4	5	5	1	5	4	4	4	
	8	4	3	3	4	5	2	4	
	6	7	2	1	5	6	3	2	
	8	6	4	4	3	7	3	1	
	9	8	4	2	8	8	5	1	Sum
ΣX (T)	83	72	43	32	71	63	36	25	425
ΣX^2	739	548	205	134	577	417	146	81	2847
SS	50.1	29.6	20.1	31.6	72.9	20.1	16.4	18.5	259.3

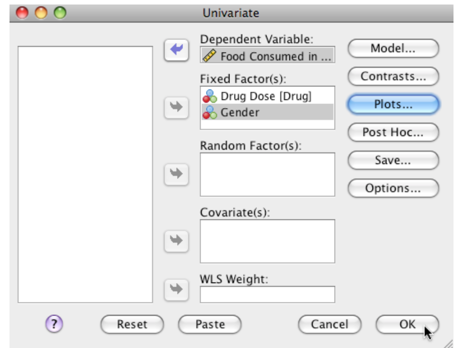
Two-Factor (2x3) ANOVA on SPSS

This example uses the data from G&W, p. 503, problem 25. Below left, note that you have to define two grouping variables (in this case Gender and Drug). Gender has 2 levels (1 = Male and 2 = Female) and Drug has 3 levels (1 = No Drug, 2 = Small Dose, and 3 = Large Dose). The final variable contains the scores for food consumed. Choosing *Univariate* from the *General Linear Model* under the *Analyze* menu produces the window on the right below. Note that I've dragged the DV (Food Consumed) into the appropriate box.

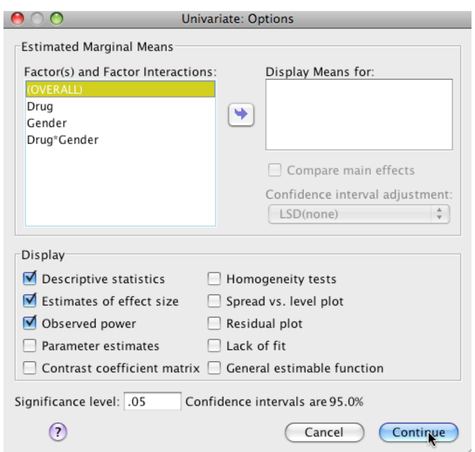


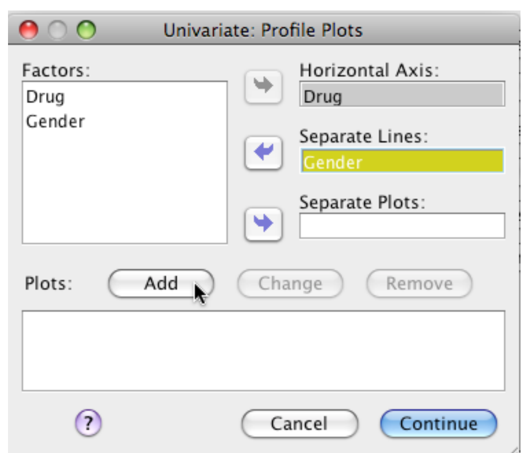
	Drug	Gender	Food
1	1.00	1.00	1.00
2	1.00	1.00	6.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
5	1.00	1.00	1.00
6	1.00	2.00	0.00
7	1.00	2.00	3.00
8	1.00	2.00	7.00
9	1.00	2.00	5.00
10	1.00	2.00	5.00
11	2.00	1.00	7.00
12	2.00	1.00	7.00
13	2.00	1.00	11.00
14	2.00	1.00	4.00
15	2.00	1.00	6.00
16	2.00	2.00	0.00
17	2.00	2.00	0.00
18	2.00	2.00	0.00
19	2.00	2.00	5.00
20	2.00	2.00	0.00
21	3.00	1.00	3.00
22	3.00	1.00	1.00





You'll also want to choose some options, so click on the *Options* button to reveal the window below left. Note that I've checked the boxes to produce descriptive statistics, estimates of effect size and power estimate. Clicking on the *Continue* button brings back the window above right. Now, click on the *Plots* button, which produces the window seen below right. Note that I've moved the Drug factor to the window that will cause it to be displayed on the horizontal axis and the Gender factor will appear as separate lines within the figure. To generate the plot, however, I first need to click on the *Add* button and then on the *Continue* button.



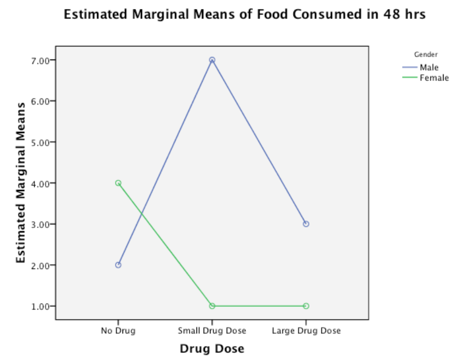


Once again, you'll return to the *Univariate* window, but now you're ready to click on the *OK* button. Doing so will produce the output seen next. As you can see in the source table, you'd have a significant interaction, as well as a main effect for Gender.

Descriptive Statistics

Dependent Variable:Food Consumed in 48 hrs

Drug Dose	Gender	Mean	Std. Deviation	N
No Drug	Male	2.0000	2.23607	5
	Female	4.0000	2.64575	5
	Total	3.0000	2.53859	10
Small Drug Dose	Male	7.0000	2.54951	5
	Female	1.0000	2.23607	5
	Total	4.0000	3.88730	10
Large Drug Dose	Male	3.0000	2.12132	5
	Female	1.0000	1.41421	5
	Total	2.0000	2.00000	10
Total	Male	4.0000	3.09377	15
	Female	2.0000	2.47848	15
	Total	3.0000	2.93610	30



Tests of Between-Subjects Effects

Dependent Variable:Food Consumed in 48 hrs

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	130.000[a]	5	26.000	5.200	.002	.520	26.000	.959
Intercept	270.000	1	270.000	54.000	.000	.692	54.000	1.000
Drug	20.000	2	10.000	2.000	.157	.143	4.000	.372
Gender	30.000	1	30.000	6.000	.022	.200	6.000	.652
Drug * Gender	80.000	2	40.000	8.000	.002	.400	16.000	.929
Error	120.000	24	5.000					
Total	520.000	30						
Corrected Total	250.000	29						

a. R Squared = .520 (Adjusted R Squared = .420)

b. Computed using alpha = .05

This source table is a bit more complex than it need be for your purposes. First of all, you can ignore the top two lines (Corrected Model and Intercept). You can also ignore the Total line. All the other lines are ones that you'll be used to from the source table in your textbook.

Unfortunately, SPSS uses colors to designate the lines within its plots, and they don't come out that well on a black-and-white printer. Furthermore, when the interaction is significant, you'll need to compute the post hoc tests yourself, because SPSS will only compute Tukey's HSD for the main effects.

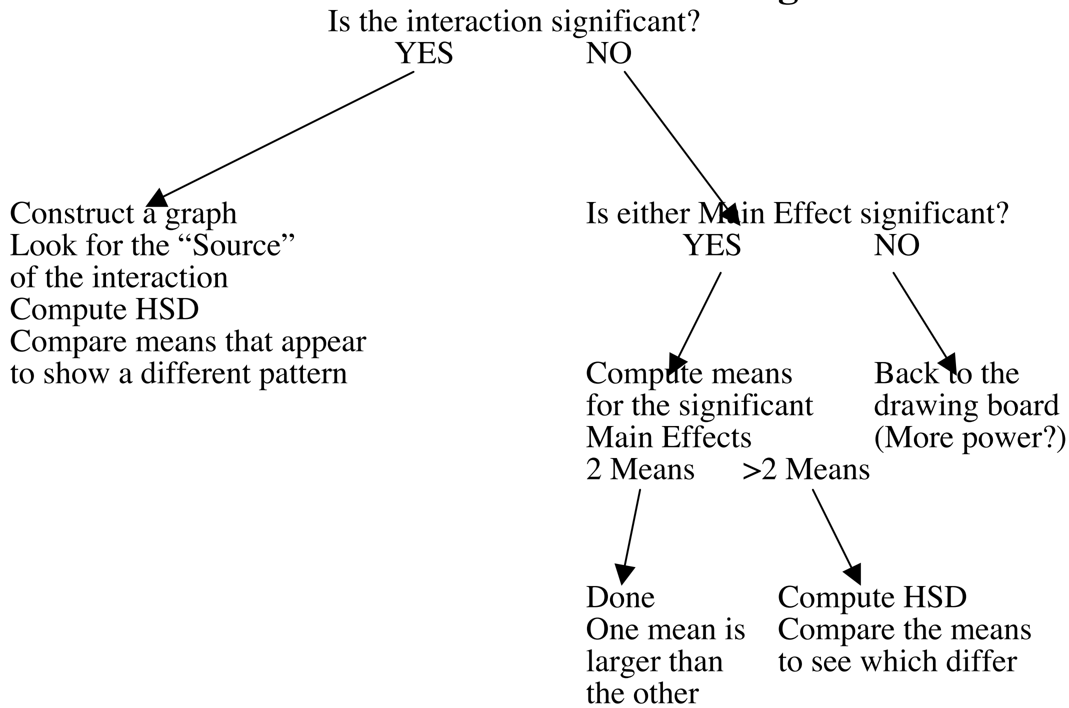
For this problem, with six means contributing to the interaction, your critical mean difference would be:

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 4.37\sqrt{\frac{5}{5}} = 4.37$$

Thus, any two means that differed by 4.37 or more would be considered significantly different. We could look at the simple effects for Drug, which would lead us to determine that the Males consumed significantly more food ($M = 7$) than Females ($M = 1$) when given a Small Dose of the drug. ***However***, Males and Females did not differ with No Drug or a Large Dose of the drug.

Alternatively, I could look at the simple effects of Gender, which would lead us to conclude that for Males, a Small Dose led to greater food consumed compared to No Drug or a Large Dose. ***However***, for Females, levels of drug had no impact on amount of food consumed.

Flow Chart for Two-Factor Designs



Example: 2x4 independent groups design with *n* = 20. Thus the *df* in the source table would be:

SOURCE	***df***
A	1
B	3
AxB	3
Error	152
Total	159

If the interaction were significant, you'd look up *q* with 8 treatment means and 152 *df* (q = 4.3). You'd compute $HSD = 4.3\sqrt{\frac{MS_{Error}}{20}}$. You'd use the resulting HSD to assess pairs of means in an effort to find a pattern where you could say, for instance, "A_1 and A_2 are equal at B_1, but A_1 is higher than A_2 at B_2, etc."

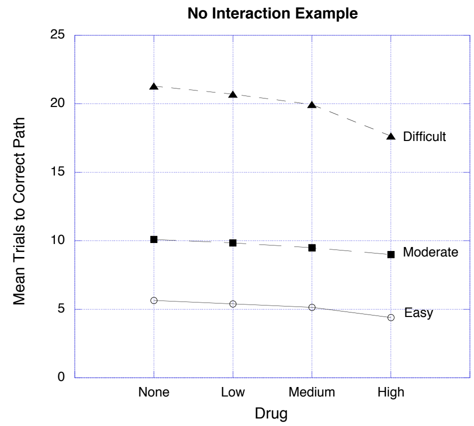
If the interaction is not significant, but the main effect for A is significant, you would need no post hoc test, because A only has two levels. If the main effect for B is significant, you would need a post hoc test, In this case, you'd look up *q* with 4 treatment means and 152 *df* (q = 3.65). You'd compute $HSD = 3.65\sqrt{\frac{MS_{Error}}{40}}$. You'd use the HSD to say something like, "B_1 is significantly higher than B_2 and B_3, which are equal, and both of which are greater than B_4."

Here's an example with no significant interaction:
A researcher was interested in studying the effects of different levels of a drug (None, Low Dose, Medium Dose, High Dose) and maze difficulty (Easy, Moderate, Difficult) on the time it took rats to learn a maze (trials to complete the maze with no errors). Complete the source table below and then analyze and interpret the outcome of this study as completely as you can.

Descriptive Statistics

Dependent Variable:Trials to Correct

Drug	Maze	Mean	Std. Deviation	N
High	Difficult	17.65000	3.543341	20
	Easy	4.40000	.940325	20
	Moderate	9.00000	.917663	20
	Total	10.35000	5.939882	60
Low	Difficult	20.70000	3.867544	20
	Easy	5.40000	1.231174	20
	Moderate	9.85000	1.225819	20
	Total	11.98333	6.912263	60
Medium	Difficult	19.95000	3.363504	20
	Easy	5.15000	1.089423	20
	Moderate	9.50000	1.000000	20
	Total	11.53333	6.601147	60
None	Difficult	21.30000	4.105196	20
	Easy	5.65000	1.308877	20
	Moderate	10.10000	1.333772	20
	Total	12.35000	7.116048	60
Total	Difficult	19.90000	3.915565	80
	Easy	5.15000	1.223194	80
	Moderate	9.61250	1.185287	80
	Total	11.55417	6.658212	240



Tests of Between-Subjects Effects

Dependent Variable:Trials to Correct

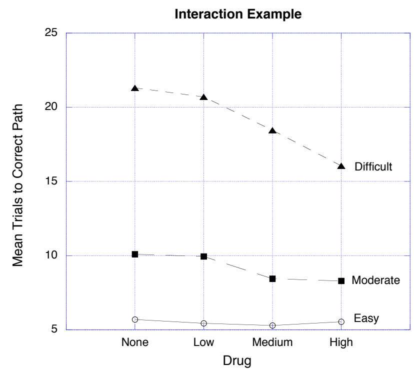
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Drug	136.1				.000	.098	24.703	.992
Maze	9154.9				.000	.879	1661.944	1.000
Drug * Maze	48.4				.192	.037	8.779	.563
Error	1255.9							
Corrected Total	10595.3							

Suppose that instead your data had come out as seen in the analyses below. Complete the analyses and interpret the results as completely as you can.

Descriptive Statistics

Dependent Variable:Trials to Correct

Drug	Maze	Mean	Std. Deviation	N
High	Difficult	16.05000	2.258901	20
	Easy	5.55000	1.234376	20
	Moderate	8.30000	.801315	20
	Total	9.96667	4.737040	60
Low	Difficult	20.70000	3.867544	20
	Easy	5.45000	1.276302	20
	Moderate	9.95000	1.316894	20
	Total	12.03333	6.893787	60
Medium	Difficult	18.45000	3.170173	20
	Easy	5.30000	1.341641	20
	Moderate	8.45000	.759155	20
	Total	10.73333	5.996798	60
None	Difficult	21.30000	4.105196	20
	Easy	5.70000	1.260743	20
	Moderate	10.10000	1.333772	20
	Total	12.36667	7.097306	60
Total	Difficult	19.12500	3.953447	80
	Easy	5.50000	1.262908	80
	Moderate	9.20000	1.353851	80
	Total	11.27500	6.287349	240

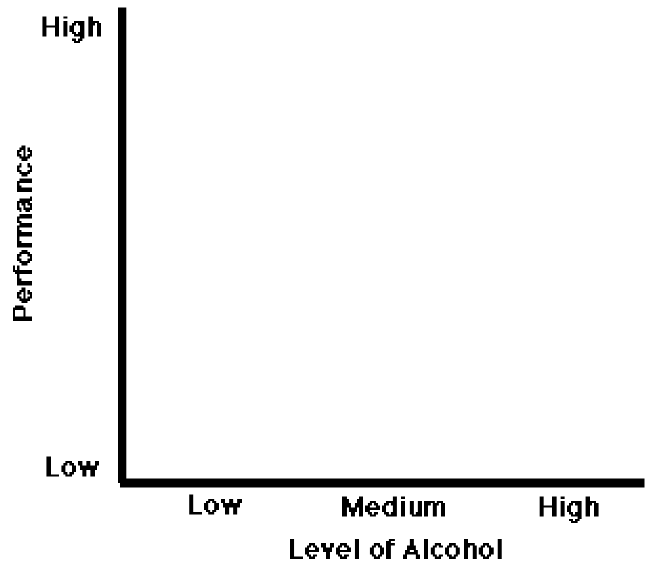


Tests of Between-Subjects Effects

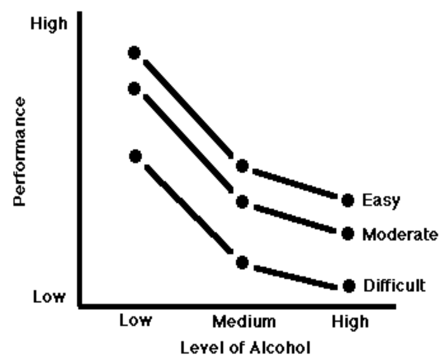
Dependent Variable:Trials to Correct

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Drug	226.3				.000	.170	46.634	1.000
Maze	7942.3				.000	.878	1636.552	1.000
Drug * Maze	172.7				.000	.135	35.593	.998
Error	1106.5							
Corrected Total	9447.8							

Dr. Rhoda Carr was interested in the impact of alcohol on driving ability. She was convinced that even fairly large amounts of alcohol would have only modest effects on performance in simple driving tasks, but that increased alcohol consumption would cause performance to drop drastically as the driving task became more difficult. To that end, she conducted an experiment in which participants were randomly assigned to one of 3 levels of alcohol (Low, Medium, High) and 3 levels of driving task difficulty (Easy, Moderate, Difficult). On the axes below, carefully and accurately draw a graph that would be completely consistent with Dr. Carr's hypotheses.



If the means from her experiment had turned out as seen below, what outcomes would you tell Dr. Carr to expect to find in any ANOVA she might compute? Why?



Dr. Carr collects her data and obtains the partially completed source table seen below. Complete the source table and tell Dr. Carr if her results might be consistent with her hypotheses and what she should do next.

Source	*SS*	*df*	*MS*	*F*
Alcohol Level	20			
Task Difficulty	10			
Alcohol x Diff	2			
Error	63			
Total		134		

Suppose that you are doing an experiment on memory for words under 4 different study strategies (Imagery, Repetition, Make-A-Story, No Instructions Control Group). In addition to strategy, you are also interested in motivation. For a third of the participants in each group, you offer $.25 for each word correctly recalled. For another third you offer $.50 for each correct word. For the final third of the participants, you offer $1.00 for each word correctly recalled. Suppose that you decide to run 10 participants in each condition of this experiment. Complete the following source table, tell me what you could reasonably conclude from the data, and what you would do next.

Source	***SS***	***df***	***MS***	***F***
Study Strategy	30			
Motivation	4			
Strategy x Motiv	96			
Error	216			
Total				

Suppose that you gave people different rewards, but were not interested in looking at that factor (i.e., you'd only included it for control purposes). Complete the source table below that you would have obtained from the one-way ANOVA on these same results.

Source	***SS***	***df***	***MS***	***F***
Study Strategy				
Error				
Total				

In the prior problem, an experiment was described in which people were randomly assigned to one of 12 conditions produced in a 3 x 4 independent groups design with motivation ($.25, $.50, $1.00 per word correctly recalled) and study strategy (Repetition, Imagery, Make-a-Story, and No instructions) as the two factors. The dependent variable is the number of words correctly recalled.

An appropriate control group is probably absent — people who participate in the study without any payment per word correctly recalled. Suppose that we want to change the experiment to include an additional level of motivation (i.e., $.00 per word correctly recalled — an intrinsic reward group). Suppose that in adding this group we will have to lower the number of people in each group to 5, so $n = 5$.

Below are the means for this experiment and a partially completed source table. Complete the source table, graph the means and interpret the results of this experiment as completely as possible.

	Repetition	Imagery	Make-a-Story	No Instructions	Marginals
No reward	5	10	10	6	
$.25/correct	3	6	6	3	
$.50/correct	4	8	8	3	
$1.00/correct	5	10	10	6	
Marginals					

Source	***SS***	***df***	***MS***	***F***
Strategy				
Motivation				8.0
Strategy x Motiv				1.8
Error			0.5	
Total	143			

Dr. Anna Mull has decided to conduct an experiment in which she tests learning in rats. She decides to run rats through 3 different mazes (Easy, Moderate, and Difficult). She decides to produce different motivation in her rats (Low, Medium, and High) by controlling their food intake. The rats in the Low motivation group are allowed to feed freely. Those in the Medium motivation group are fed reduced rations such that they are at 80% of their free-feeding weight. Those in the High motivation group are fed such that they are at 70% of their free-feeding weight.

Below are the summary data and a partially completed source table for Dr. Mull's research. Complete the source table and answer the following questions.

Source	***SS***	***df***	***MS***	***F***
Maze	20			
Motivation	10			
Maze x Motivation	100			
Error		81		
Total	292			

How many rats were in each condition?

What can you conclude from this study?

What would you do next?

What *F*-ratio would Dr. Mull have obtained had she analyzed these same data as a one-way ANOVA on the Motivation factor alone?

Dr. Robert Katt was interested in the role of taste aversions in the feeding behavior of cougars. He designed a study in which some cougars were exposed to tainted meat that made them ill, but did not kill them. Other cougars were exposed to the same meat, but it wasn't tainted, so they were not made ill. He was interested in whether the cougars developed taste aversions that led them to avoid killing/eating particular prey as a result of their taste aversions. To that end, he exposed a total of 80 hungry cougars to one of 5 kinds of meat: sheep, cow, dog, armadillo, and grizzly bear. For half of the cougars, the meat was poisoned, so that the cougars eating the meat were made ill. For the other half, the meat was not poisoned.

After a one-week delay, the cougars were again starved and then exposed to the types of meat they had eaten previously (but now none of it is poisoned). The dependent variable is the time delay (in minutes) before the cougar begins to eat the meat. If the cougar hasn't eaten the meat after 10 minutes, the cougar is given a score of 10. The source table and mean data are as seen below. Analyze these data and interpret as completely as possible. What would you tell Dr. Katt about the results?

Source	***SS***	***df***	***MS***	***F***
Poison/Not Poison	622			
Type of Meat	293			
Poison x Meat	236			
Error	41			
Total				

	Sheep	Cow	Dog	Armadillo	Grizzly Bear	Marginal
Poisoned at Time 1	9.375	9.25	9.125	9.375	9.875	9.4
Not Poisoned at Time 1	.875	.875	.875	7.75	8.75	3.825
Marginal	5.125	5.062	5	8.562	9.312	6.613

In a study of hyperactivity among elementary school boys, 63 students were randomly selected from a school population of ADHD, 7-year-old boys are randomly assigned to one of 9 groups (n = 7). (ADHD is Attention Deficits with Hyperactivity, and left untreated, it can prevent a child from attending to incoming learning stimuli and may also create major disruptions in the classroom.) The researcher wanted to study the classroom effects of both the drug Ritalin as well as a behavior modification program on the activity levels of the students. The drug administered at three levels: 5 mg Ritalin, 10 mg Ritalin, and 20 mg Ritalin. The behavior modification program consisted of giving each student 10 tokens to start the day and then taking away a token for each hyperactive infraction. The tokens that were saved could then be exchanged for some valued prize. The behavior modification program was varied across three levels: no program, program every other day, and program every day. After four weeks, all the children were evaluated for hyperactivity and were assigned scaled scores ranging from a possible low of 0 (no indication of hyperactivity) to a high of 40 (extreme hyperactivity). An SPSS output of the data is seen below. Analyze the data as completely as possible, providing a complete interpretation.

Descriptive Statistics

Dependent Variable:Hyperactivity Score

Treatment	Drug	Mean	Std. Deviation	N
BM Ev Oth Day	10 mg	19.0000	5.47723	7
	20 mg	17.1429	5.42920	7
	5 mg	22.4286	7.41299	7
	Total	19.5238	6.27391	21
BM Every Day	10 mg	12.2857	3.35233	7
	20 mg	8.7143	3.59232	7
	5 mg	14.4286	3.64496	7
	Total	11.8095	4.13061	21
No Beh Mod	10 mg	22.0000	3.95811	7
	20 mg	16.4286	4.27618	7
	5 mg	29.1429	5.81460	7
	Total	22.5238	6.98297	21
Total	10 mg	17.7619	5.86434	21
	20 mg	14.0952	5.78710	21
	5 mg	22.0000	8.28251	21
	Total	17.9524	7.39099	63

Tests of Between-Subjects Effects

Dependent Variable:Hyperactivity Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Treatment				26.3	.000	.493	52.583	1.000
Drug				13.5	.000	.333	26.934	.997
Treatment * Drug				1.3	.275	.089	5.277	.384
Error			24.4					
Corrected Total								

Several researchers have investigated the encoding specificity effect. The general finding is that people remember best when the testing situation is as similar as possible to the learning situation. (Thus, because the typical testing situation is a relatively quiet classroom, you'd best study/learn under conditions as similar to the testing situation as possible.) Dr. Julie Ard was interested in the effects of music on studying, as well as the encoding specificity effect. That is, she was interested in the extent to which the similarity of the study and test situations affected performance. To test her hypotheses, she used five acquisition conditions (studying while listening to: heavy metal, rock, classical, jazz, or blues). People in these groups studied written material while listening to a particular type of music. After a brief delay, half of the people in each condition were tested under identical music (same) and half of the people were tested with no music (different). The dependent variable was the percentage score on the test (100 = perfect performance). Complete the analysis and interpret the results below as completely as possible.

Descriptive Statistics

Dependent Variable:Score

Music	Test	Mean	Std. Deviation	N
Blues	Different	82.70000	1.251666	10
	Same	88.10000	2.131770	10
	Total	85.40000	3.250911	20
Classical	Different	88.30000	3.529243	10
	Same	94.80000	2.149935	10
	Total	91.55000	4.382681	20
Heavy Metal	Different	79.20000	3.011091	10
	Same	80.20000	2.394438	10
	Total	79.70000	2.696977	20
Jazz	Different	85.80000	3.047768	10
	Same	94.30000	1.888562	10
	Total	90.05000	5.010253	20
Rock	Different	84.50000	1.178511	10
	Same	86.10000	1.286684	10
	Total	85.30000	1.454575	20
Total	Different	84.10000	3.965412	50
	Same	88.70000	5.821319	50
	Total	86.40000	5.467997	100

Tests of Between-Subjects Effects

Dependent Variable:Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Music	1738.9				.000	.782	322.682	1.000
Test	529.0				.000	.522	98.165	1.000
Music * Test	207.1				.000	.299	38.431	1.000
Error	485.0							
Corrected Total	2960.0							

Researchers are interested in a phenomenon called *hindsight bias*. When you hear people talk about being a Monday Morning Quarterback, they're referring to this hindsight bias. Thus, knowing the results of professional football games on Sunday leads people to be more confident that they would have been able to predict the outcomes before the games (i.e., on Saturday). Psychologists have studied this phenomenon in a number of ways, but one way is through anagram solving (turning scrambled letters into words). That is, we can present some people with anagrams to solve and measure the time (in minutes) to solve the anagrams (the Worksight Condition). For other people, we can present the anagrams and the solutions simultaneously. Their task is to estimate the time they think it would take someone to solve the anagram (the Hindsight Condition). Thus, to the extent that hindsight bias is present, the estimated times in the Hindsight Condition will be less than the actual solution times in the Worksight Condition. In addition to this factor, we might also examine the extent to which anagram length has an impact on hindsight bias. Thus, the two independent variables would be: TASK (actually solve anagrams vs. with solution present, estimate time to solve) and LENGTH (anagrams would be 4, 6, or 8 letters long). Complete the analysis below, and interpret the results as completely as you can.

Descriptive Statistics

Dependent Variable:"Solution" Time

Task	Anagram ...	Mean	Std. Deviation	N
Hindsight	4-Letter	.6690	.08239	10
	6-Letter	1.3600	.10530	10
	8-Letter	2.0190	.11377	10
	Total	1.3493	.56908	30
Worksight	4-Letter	1.0920	.21049	10
	6-Letter	1.9480	.11545	10
	8-Letter	2.5950	.50322	10
	Total	1.8783	.69891	30
Total	4-Letter	.8805	.26700	20
	6-Letter	1.6540	.32024	20
	8-Letter	2.3070	.46194	20
	Total	1.6138	.68587	60

Dependent Variable:"Solution" Time

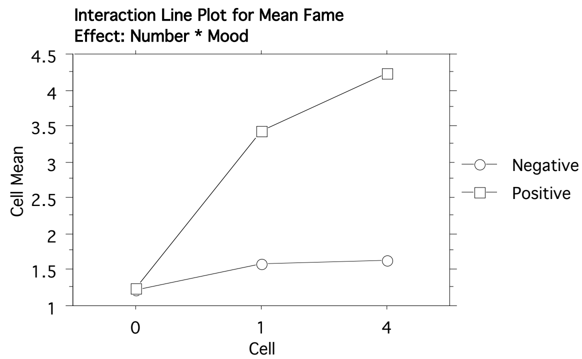
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
Task				73.711	.000	.577	1.000
Length				179.091	.000	.869	1.000
Task * Length				.743	.480	.027	.170
Error			.057				
Corrected Total	27.755						

2. Kitamura (2005) was interested in the impact of mood on cognitive processes. Kitamura thought that a positive mood leads to more automatic processing than a negative mood, which leads to more controlled processing. In one study, half of the participants were placed in a positive mood and half in a negative mood (using a mood induction technique). Then they were all given a list of non-famous companies either once or four times. Two days later they were asked to judge the fame of a list of companies, some of which were new (Number = 0) and some that had been seen previously (Number = 1 or 4). Let's pretend that the participants rated fame on a 7-point Likert-type scale (1 = "not famous" and 7 = " famous"). Suppose that the data had produced the results seen below. Complete the analysis and interpret the results as completely as you can. [15 pts]

Descriptive Statistics

Dependent Variable:Mean Fame

Mood	Number	Mean	Std. Deviation	N
Negative	0	1.20833	.178164	12
	1	1.58333	.327062	12
	4	1.63333	.486795	12
	Total	1.47500	.393791	36
Positive	0	1.23333	.182574	12
	1	3.42500	.748483	12
	4	4.23333	1.144420	12
	Total	2.96389	1.500124	36
Total	0	1.22083	.176879	24
	1	2.50417	1.097221	24
	4	2.93333	1.582147	24
	Total	2.21944	1.322038	72



Tests of Between-Subjects Effects

Dependent Variable:Mean Fame

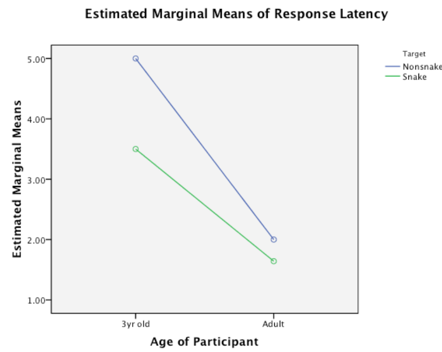
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Mood	39.9				.000	.614	105.055	1.000
Number	38.1				.000	.603	100.337	1.000
Mood * Number	21.0				.000	.456	55.320	1.000
Error	25.1							
Corrected Total	124.1							

In a study of early ability to detect a fear-relevant stimulus (a snake), LoBue and DeLoache (2008) presented 3-year-old children and adults (Age: 3-year old vs. adult) a series of 3x3 matrices of pictures. The subject's task was to point out a target by touching one of the nine pictures on a touch-screen (Target: either a *snake* among eight non-snake distractors or a non-snake animal, such as a *caterpillar*, among eight snake distractors). Thus, we can think of this study as a 2x2 independent groups design. Below is a partially completed source table that is consistent with their results (Experiment 3). Complete the source table and interpret the results as completely as you can.

Descriptive Statistics

Dependent Variable:Response Latency

Age of Parti...	Target	Mean	Std. Deviation	N
3yr old	Nonsnake	5.0000	.95346	12
	Snake	3.5000	.79772	12
	Total	4.2500	1.15156	24
Adult	Nonsnake	2.0000	.82572	12
	Snake	1.6000	.76396	12
	Total	1.8000	.80434	24
Total	Nonsnake	3.5000	1.76315	24
	Snake	2.5500	1.23500	24
	Total	3.0250	1.58053	48



Dependent Variable:Response Latency

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
Age	72.030				.000	.700	1.000
Target	10.830				.000	.259	.970
Age * Target	3.630				.028	.105	.604
Error	30.920						
Corrected Total	117.410						

Because old exams for this topic use StatView for analyses, here is an example of StatView output for a two-way ANOVA.

Dr. Mai Ayes was interested in studying the effects of task difficulty and sleep deprivation on performance, using a completely between (independent groups) design. The amounts of sleep deprivation that she decided to use are: 24, 36, 48, 60, and 72 hours. That is, participants were awake without sleep for one of those periods before being tested on either an easy, a moderate, or a difficult task. She measured performance on a 9-point scale (1 = lousy performance <-> 9 = excellent performance). Analyze these data as completely as you can.

ANOVA Table for Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Hours Deprived			16.4		<.0001	173.193	1.000
Task Difficulty			60.0		<.0001	315.860	1.000
Hours Deprived * Task Difficulty				.03	>.9999	.281	.058
Residual		22.8					

Means Table for Score
Effect: Hours Deprived * Task Difficulty

	Count	Mean	Std. Dev.	Std. Err.
24 Hours, Difficult	5	4.000	.707	.316
24 Hours, Easy	5	7.200	.447	.200
24 Hours, Moderate	5	6.200	.447	.200
36 Hours, Difficult	5	2.600	.548	.245
36 Hours, Easy	5	5.600	.548	.245
36 Hours, Moderate	5	4.600	.548	.245
48 Hours, Difficult	5	2.600	.548	.245
48 Hours, Easy	5	5.600	.548	.245
48 Hours, Moderate	5	4.600	.548	.245
60 Hours, Difficult	5	1.800	.837	.374
60 Hours, Easy	5	4.800	.837	.374
60 Hours, Moderate	5	3.800	.837	.374
72 Hours, Difficult	5	1.400	.548	.245
72 Hours, Easy	5	4.400	.548	.245
72 Hours, Moderate	5	3.400	.548	.245