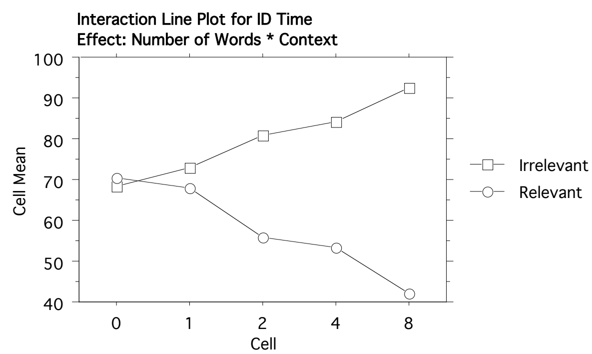
1. In a classic study, Tulving and Gold (1963) studied word identification under conditions of amounts of relevant and irrelevant context. Let's conceive of their study as a 2x5 independent groups design, with Context (Relevant vs. Irrelevant) and Number of Words (0, 1, 2, 4, and 8) as the two factors. For the very briefly presented target word *performer*, for instance, some participants would see no preceding words (0) or the first 1, 2, 4, or 8 words from one of two sentences. The Relevant sentence was "The actress received praise for being an outstanding..." The Irrelevant sentence was "The dog retrieved the burrito from the neighbor's..." Obviously, the issue was the extent to which the preceding context would help the participant to identify the target word that was presented very briefly. The dependent variable was the time it took the participant to identify the target word. Complete the source table below and analyze the data as completely as you can, providing an interpretation for the obtained results. [15 pts]

ANOVA Table for ID Time

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Number of Words	4	132.467	33.117	3.375	.0120	13.500	.841
Context	1	14344.533	14344.533	1461.920	<.0001	1461.920	1.000
Number of Words * Context	4	10560.467	2640.117	269.067	<.0001	1076.267	1.000
Residual	110	1079.333	9.812				

Means Table for ID Time
Effect: Number of Words * Context

	Count	Mean	Std. Dev.	Std. Err.
0, Irrelevant	12	68.333	3.200	.924
0, Relevant	12	70.500	3.989	1.151
1, Irrelevant	12	73.083	2.314	.668
1, Relevant	12	67.917	2.778	.802
2, Irrelevant	12	80.917	3.147	.908
2, Relevant	12	55.917	3.679	1.062
4, Irrelevant	12	84.333	2.964	.856
4, Relevant	12	53.333	3.025	.873
8, Irrelevant	12	92.500	3.425	.989
8, Relevant	12	42.167	2.406	.694



First, I would test for the homogeneity of the variances in the groups. Using Hartley's $F_{Max\ Crit}$ = 8.0, with F_{Max} = 2.97 I would presume that there is no concern about heterogeneity of variance and use α = .05 for the ANOVA.

$$F_{Max} = \frac{15.91}{5.35} = 2.97$$

There is a significant main effect for number of words, $F(4,110) = 3.375$, $MSE = 9.812$, $p = .012$. There is also a significant main effect for context, $F(1,110) = 1461.92$, $MSE = 9.812$, $p < .001$. There is also a significant interaction, $F(4,110) = 269.067$, $MSE = 9.812$, $p < .001$. To interpret the interaction, I would first compute Tukey's HSD:

$$HSD = 4.58\sqrt{\frac{9.812}{12}} = 4.14$$

Thus, when there are no words present (so that there is not really any context), people don't differ in their speed of identifying the target word. However, with one or more words

of context present, people take significantly longer to identify the target word when the word(s) provided an irrelevant context compared to those shown the word(s) in a relevant context. As seen in the figure, it appears that the more words of context provided, the greater the impact, so that the more irrelevant context you have, the more difficult it is to identify the target word and the more relevant context you have, the easier it is to identify the target word.

2. 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). A partially completed StatView output of the data is seen below. Complete the source table and then analyze the data as completely as possible, providing a complete interpretation. Behavior Mod (None, Every other day, Every day). Ritalin (5 mg, 10 mg, 20 mg). [10 pts]

ANOVA Table for Hyperactivity Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Behavior Mod	2	1283.143	641.571	26.292	<.0001	52.583	1.000
Ritalin	2	657.238	328.619	13.467	<.0001	26.934	.999
Behavior Mod * Ritalin	4	128.762	32.190	1.319	.2746	5.277	.376
Residual	54	1317.714	24.402				

Means Table for Hyperactivity Score
Effect: Behavior Mod

	Count	Mean	Std. Dev.	Std. Err.
BM Ev Oth Day	21	19.524	6.274	1.369
BM Every Day	21	11.810	4.131	.901
No Beh Mod	21	22.524	6.983	1.524

Means Table for Hyperactivity Score
Effect: Ritalin

	Count	Mean	Std. Dev.	Std. Err.
10 mg	21	17.762	5.864	1.280
20 mg	21	14.095	5.787	1.263
5 mg	21	22.000	8.283	1.807

First, I would test for the homogeneity of the variances in the groups. Using Hartley's $F_{Max\ Crit} = 17.5$, with $F_{Max} = 4.89$ I would presume that there is no concern about heterogeneity of variance and use $\alpha = .05$ for the ANOVA.

$$F_{Max} = \frac{54.95}{11.24} = 4.89$$

There is a significant main effect for behavior modification, $F(2,54) = 26.292$, $MSE = 24.402$, $p < .001$. There is also a significant main effect for Ritalin, $F(2,54) = 13.467$, $MSE = 24.402$, $p < .001$. There was no significant interaction, $F(4,54) = 1.319$, $MSE = 24.402$, $p = .275$. Thus, I would focus on the main effects, which would use the same Tukey's HSD:

$$HSD = 3.42\sqrt{\frac{24.402}{21}} = 3.69$$

The means for Behavior Modification would be 19.5, 11.8, and 22.5 for Every Other Day, Every Day, and No Beh Mod, respectively. The means for Ritalin would be 17.8, 14.1, and 22 for 10mg, 20 mg, and Placebo, respectively.

For Behavior Mod: Administering the Behavior Mod every day leads to less hyperactivity than administering Beh Mod every other day or not using Beh Mod (and those two levels produce statistically similar results).

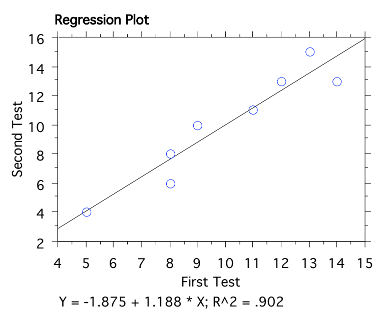
For Ritalin: Dosage 20 leads to significantly less hyperactivity than Dosage 10 or Placebo. Dosage 10 leads to significantly less hyperactivity than Placebo.

3. A researcher has developed a new test of self-esteem. To evaluate the reliability of the test, the researcher takes a sample of $n = 8$ participants. Each individual takes the test on a Monday evening, then returns two weeks later to take the test again. The two scores for each individual are reported below. Are the participant's scores on the two tests related to one another? (This measure is often referred to as test-retest reliability.) [15 pts]

	First Test	Second Test
	13	15
	5	4
	12	13
	11	11
	9	10
	14	13
	8	8
	8	6
Sum	80	80
SS	64	100

Regression Summary
Second Test vs. First Test

Count	8
Num. Missing	0
R	.950
R Squared	.902
Adjusted R Squared	.886
RMS Residual	1.275



ANOVA Table
Second Test vs. First Test

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	90.250	90.250	55.538	.0003
Residual	6	9.750	1.625		
Total	7	100.000			

Regression Coefficients
Second Test vs. First Test

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	-1.875	1.656	-1.875	-1.132	.3007
First Test	1.188	.159	.950	7.452	.0003

There is a significant correlation between the two variables, $r(6) = .95$, $p < .001$.

4a. Before making a decision about his advertising campaign, a publisher ran an experiment to discover whether readers' responses to certain ads differed. He wanted to test responses to three kinds of ads: ads with a color picture, ads with a black-and-white picture, and ads with no picture. Each ad was inserted with other material intended to draw attention away from the material being evaluated. Participants rated the critical ad on an 11-point scale (1 = little preference for the ad, 11 = great preference for the ad). Results for the 24 participants are given below. Analyze the results as completely as you can and then interpret the results. [15 pts]

	Color Picture	Black-and-White Picture	No Picture
	3	4	10
	3	7	7
	7	5	8
	6	3	5
	8	9	9
	1	8	7
	5	7	6
	3	5	8
Sum	36	48	60
SS	40	30	18

ANOVA Table for Pref Rating

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Type of Ad	2	36.000	18.000	4.295	.0273	8.591	.683
Residual	21	88.000	4.190				

Means Table for Pref Rating
Effect: Type of Ad

	Count	Mean	Std. Dev.	Std. Err.
B&W Pic	8	6.000	2.070	.732
Color Pic	8	4.500	2.390	.845
No Pic	8	7.500	1.604	.567

There is an effect of type of picture, $F(2,21) = 4.295$, $MSE = 4.19$, $p = .027$. To determine the source of the effect, I would compute HSD:

$$HSD = 3.57\sqrt{\frac{4.19}{8}} = 2.58$$

Thus, preference for the ad was significantly higher with no picture ($M = 7.5$) compared to the ad with the color picture ($M = 4.5$). [Non-intuitive results, but, heck, the data are all made up anyway!]

4b. [No computation is necessary to answer this part of the question.] Suppose that you were told that the first four participants in each group (of 4a) were males and the remaining participants in each group were females. This information would enable you to analyze the data in a new way, should you introduce this factor. How would your interpretation of the results be most likely to differ? Under which conditions would the F-ratio for Type of Ad be larger as a result of the new analysis? Under which conditions would the F-ratio be smaller as a result of the new analysis? [Examples, such as a possible source table, might help here.] [5 pts]

You'd be likely to find a larger *F* for the Type of Ad, presuming that there were actually differences between the males and females.

Source	*SS*	*df*	*MS*	*F*
Type of Ad	36	2	18	18
Gender	20	1	20	20
Type of Ad x Gender	50	2	25	25
Error	18	18	1	

If the males and females were very similar, then your *F* for Type of Ad would likely decrease, as seen below:

Source	*SS*	*df*	*MS*	*F*
Type of Ad	36	2	18	3.77
Gender	1	1	1	.21
Type of Ad x Gender	1	2	.5	.10
Error	86	18	4.78	

4c. [No computation is necessary to answer this part of the question.] Suppose that the *same 24 pieces of data* had been obtained from only 8 participants (in 4a) in a repeated measures analysis. How would your interpretation of the results be most likely to differ? Under which conditions would the F-ratio for Type of Ad be larger as a result of the new analysis? Under which conditions would the F-ratio be smaller as a result of the new analysis? [Examples, such as a possible source table, might help here.] [5 pts]

Because the repeated measures ANOVA is more powerful, you would expect that your *F*-ratio would be larger. The larger *F* would be due to the removal of the individual differences, as seen below:

Source	*SS*	*df*	*MS*	*F*
Type of Ad	36	2	18	18
Within	88	21		
Subject	70	7		
Error	18	18	1	

However, if there were few individual differences, the *F* would actually be smaller, as seen below:

Source	*SS*	*df*	*MS*	*F*
Type of Ad	36	2	18	3.72
Within	88	21		
Subject	1	7		
Error	87	18	4.83	

5. Although psychologists do not completely understand the phenomenon of dreaming, it does appear that people need to dream (or, at the very least, need REM sleep). One experiment demonstrating this fact shows that people who are deprived of REM sleep one night tend to have more REM sleep (dreams?) the following night, as if they were trying to make up for the lost REM sleep. In a typical version of this experiment, the psychologist first records the number of periods of REM sleep during a normal night's sleep. The next night, each participant is prevented from REM sleep by being awakened as soon as he or she begins to exhibit REM sleep. During the third night, the psychologist once again records the number of periods of REM sleep. Hypothetical data from this experiment are seen below. Analyze and interpret these results as completely as you can. [15 pts]

Participant	First Night	Night After Deprivation
1	4	7
2	5	5
3	4	8
4	6	7
5	4	9
6	5	7
7	4	7
8	4	6
Sum	36	56
SS	4	10

Note, at the outset, that this repeated measures design did not use proper counterbalancing.

ANOVA Table for Night

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	7	5.000	.714				
Category for Night	1	25.000	25.000	19.444	.0031	19.444	.970
Category for Night * Subject	7	9.000	1.286				

Means Table for Night
Effect: Category for Night

	Count	Mean	Std. Dev.	Std. Err.
First Night	8	4.500	.756	.267
Night After Dep	8	7.000	1.195	.423

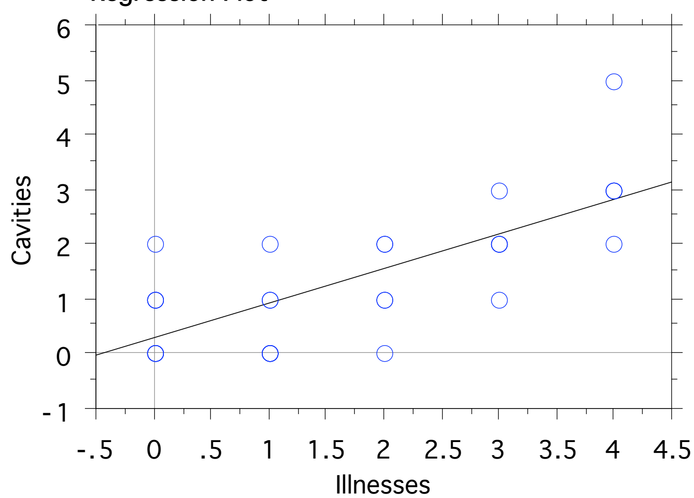
There is a significant effect present (which may be due to deprivation or it may be due to the greater familiarity with the research setting), $F(1,7) = 19.44$, $MSE = 1.286$, $p = .003$. Thus, the amount of REM sleep is greater on the third night ($M = 7.0$) than on the first night ($M = 4.5$).

6. Dr. Ginger Vitas is a health psychologist who is interested in the relationship between dental heath (operationally defined as number of cavities found in an annual checkup) and general health (operationally defined as the number of illnesses experienced in the preceding year). Analyze the output seen below as completely as you can. If a person had 3 illnesses in a year, how many cavities would you predict? If a person had 6 illnesses, how many cavities would you predict? What proportion of variability do these two variables share? [5 pts]

Regression Summary
Cavities vs. Illnesses

Count	30
Num. Missing	0
R	.738
R Squared	.545
Adjusted R Squared	.529
RMS Residual	.855

Regression Plot



Y = .276 + .629 * X; R^2 = .545

ANOVA Table
Cavities vs. Illnesses

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	24.501	24.501	33.521	<.0001
Residual	28	20.466	.731		
Total	29	44.967			

Regression Coefficients
Cavities vs. Illnesses

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	.276	.245	.276	1.128	.2691
Illnesses	.629	.109	.738	5.790	<.0001

7a. Dr. Kip Werkin is an industrial/organizational psychologist who is interested in the impact of environmental factors (such as noise) on productivity. He has a group of ten workers experience each of a set of background noise levels (70 dB, 80 dB, 90 dB, and 100 dB SPL) as they work on a project that involves creating delicate instruments. (SPL = Sound Pressure Level) The dependent variable is the number of errors made in the construction of the pieces. Complete the source table and tell Dr. Werkin what he should conclude from this study. [10 pts]

ANOVA Table for SPL

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	4.000	.444				
Category for SPL	3	13.900	4.633	22.339	<.0001	67.018	1.000
Category for SPL * Subject	27	5.600	.207				

Means Table for SPL
Effect: Category for SPL

	Count	Mean	Std. Dev.	Std. Err.
SPL 70 dB	10	.200	.422	.133
SPL 80 dB	10	.200	.422	.133
SPL 90 dB	10	1.000	.667	.211
SPL 100 dB	10	1.600	.516	.163

First, note that the study could not have been appropriately counterbalanced, because with four levels of noise, there should be 24 orders used.

There is a significant effect of SPL, $F(3,27) = 22.339$, $MSE = .207$, $p < .001$. To determine which levels differ, you would compute HSD:

$$HSD = 3.87\sqrt{\frac{.207}{10}} = .56$$

Thus, the 100 dB condition led to significantly more errors ($M = 1.6$) than the 90 dB ($M = 1.0$), 80 dB ($M = .2$) and the 70 dB ($M = .2$) conditions. The 90 dB condition led to more errors than the 80 dB and the 70 dB conditions (neither of which differed).

7b. If the ***same*** data were analyzed with an independent groups ANOVA, complete the source table below. [5 pts]

Source	*df*	*SS*	*MS*	*F*
Treatment	3	13.9	4.633	17.37
Error	36	9.6	.27	
Total	39	23.5		

Note that the *F* decreases, as you'd expect.