1. Dr. June Bugg was interested in examining the extent to which a person's attention is captured by a problem. She sets up her experiment on the computer so that the participant is presented with a series of problems, one at a time. Participants are told to solve as many problems as they can within a 10-minute period. At the same time, participants are told that when a small dot appears in the bottom left corner of the computer screen, they should hit the space bar. The experiment lasts about an hour, with 5 different types of problems that vary in difficulty (Very Easy, Easy, Moderate, Difficult, and Very Difficult). Dr. Bugg wants to "warm up" the participants for the Very Difficult problems, so she runs every participant through the Very Easy problems first (for 10 minutes), then through the Easy problems (for the next 10 minutes), then through the Moderate problems (10 minutes), then through the Difficult problems (10 minutes), and finally through the Very Difficult problems (10 minutes). [Obviously, in a 10-minute interval people will solve more easy problems than difficult problems, but that's not the dependent variable.] At 10 random times within each 10-minute period, a dot appears in the bottom left corner of the screen. The dependent variable in this experiment is the number of times that the participant detects the dot at the bottom of the screen. Dr. Bugg reasons that as tasks become more difficult, the participant's attention will be more absorbed by the problem, leading the participant to miss the appearance of the dot. Below is a partially completed source table for this experiment. Complete the table and interpret the results as completely as you can. [20 pts]

ANOVA Table for Task Difficulty

	DF	Sum of S...	Mean S...	F-Value	P-Value	Lambda	Power
Subject	11	20.733	1.885				
Category for Task Difficulty	4	145.500	36.375	88.425	<.0001	353.702	1.000
Category for Task Difficulty * Subject	44	18.100	.411				

Means Table for Task Difficulty
Effect: Category for Task Difficulty

	Count	Mean	Std. Dev.	Std. Err.
Very Easy	12	9.333	.778	.225
Easy	12	9.083	.669	.193
Moderate	12	8.250	.622	.179
Difficult	12	7.500	.522	.151
Very Difficult	12	5.000	1.348	.389

First of all, you should note that this repeated measures design has 5 levels. As such, if you were to use incomplete counterbalancing you would need to run participants in multiples of 10. Thus, with 12 participants the study could not have been appropriately counterbalanced.

Because you would reject H_0 ($p < .05$), you next need to compute a post-hoc analysis to determine which groups are from populations with different means. In this case, with q = 4.02, HSD = .74. Thus, performance on Very Easy and Easy problems were equivalent and both were better than Moderate problems. Moderate problems were better than Difficult problems (barely), and Difficult problems were better than Very Difficult problems.

2. Professor Ty Knott was interested in the relationship between the longevity of divorced women's marriages and the longevity of their divorced parent's marriages. To that end, he collected a sample of 25 divorced women whose parents had also been divorced. The StatView analysis of the data is seen below. Interpret these data as completely as you can. If a woman's parents had been married for 10 years prior to their divorce, how long would you predict that the woman's marriage would last before a divorce? If a woman's parents had never been divorced, what prediction could you make about the length of time before the woman's marriage might end in divorce? [10 pts]

Regression Summary
Yrs. Person Married vs. Yrs. Parents Married

Count	25
Num. Missing	0
R	.852
R Squared	.726
Adjusted R Squared	.714
RMS Residual	1.510

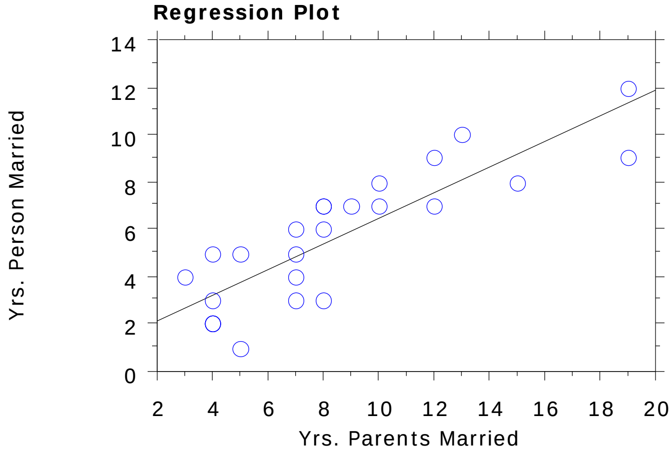
ANOVA Table
Yrs. Person Married vs. Yrs. Parents Married

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	139.013	139.013	60.985	<.0001
Residual	23	52.427	2.279		
Total	24	191.440			

Regression Coefficients
Yrs. Person Married vs. Yrs. Parents Married

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	1.108	.659	1.108	1.682	.1061
Yrs. Parents Married	.539	.069	.852	7.809	<.0001

Regression Plot



Y = 1.108 + .539 * X; R^2 = .726

There is a significant positive linear relationship (p < .05) between the number of years a woman is married and the number of years that woman's divorced parents were married. I her parents had been married for 10 years, the regression equation would lead to a prediction that she would be married for 6.5 years. Because the study did not look at women whose parents were not divorced, you could not really predict the duration of the marriage.

3. Hess has done work that suggests that (eye) pupil size is affected by emotional arousal. In order to ascertain if the type of arousal makes a difference, Dr. Bo Ring decides to measure pupil size under three different arousal conditions (Neutral, Pleasant, Aversive). Participants look at pictures that differ according to the condition (neutral = plain brick building, pleasant = young man and woman sharing an ice cream cone, aversive = graphic photograph of an automobile accident). The pupil size after viewing each photograph is measured in millimeters. The data are seen below:

	Neutral	Pleasant	Aversive	
	4	8	3	
	3	7	7	
	2	7	3	
	3	8	4	
	3	6	7	
	5	7	4	
	4	8	4	
	3	7	6	
	5	6	5	
	2	7	6	
				SUM
Sum	34	71	49	154.
SS	10.4	4.9	20.9	36.2
ΣX^2	126.	509.	261.	896.

a. Due to miscommunication between Dr. Ring and her research assistant Igor, she has no idea if the study was run as an independent groups design or as a repeated measures design. Igor has left the lab in a cloud of controversy amidst a whole range of allegations (involving improper use of lab alcohol, monkeys, and videotapes, but that's another story entirely). With no immediate hope of clarifying the way in which the data were collected, Dr. Ring decides that the most reasonable strategy would be to analyze the data as an independent groups design. Her logic should make sense to you, so explain to me why she would be smart to analyze the data as though they were collected in an independent groups design. [5 pts]

The logic here is that the repeated measures analysis is more powerful than an equivalent independent groups analysis. As such, if you didn't know the true nature of the design, you'd be better served by computing an independent groups ANOVA. If your results are significant using that approach, they should also be significant using a repeated measures ANOVA.

Of course, you could never publish results of such a study. Can you imagine: "I forgot how my data were collected, so I'm going to assume that the study was an independent groups design." The editors reviewing your paper would have a good laugh at your expense! ☺

b. OK, Dr. Ring asks you to please analyze the data as though they were collected in an independent groups design. Please do so now, analyzing and interpreting the data as completely as you can. [25 pts]

F_{Max} = 4.3, so with $F_{MaxCrit}$ = 5.34, you would retain H_0: $\sigma^2_N = \sigma^2_P = \sigma^2_A$. There is no evidence that you need to be concerned about heterogeneity of variance.

For the ANOVA, H_0: $\mu_N = \mu_P = \mu_A$; H_1: Not H_0

Source	SS	df	MS	F
Treatment	69.3	2	34.7	25.86
Error	36.2	27	1.34	
Total	105.5	29		

With F_{Crit} (2,27) = 3.35, we would reject H_0, because $F_{Obtained} \geq F_{Crit}$.

In order to determine which means differed, the next step would be to compute post-hoc analyses. In this case, q = 3.5 and HSD = 1.28.

	Neutral	Pleasant	Aversive
Neutral			
Pleasant	**3.7***		
Aversive	**1.5***	**2.2***	

All of the differences exceed the critical mean difference (1.28), so we could conclude that Pleasant stimuli lead to significantly larger pupil sizes than Aversive stimuli and that both of those types of stimuli lead to larger pupil sizes than Neutral stimuli.

c. As you noodle around the lab, having completed the analyses that Dr. Ring wanted, you stumble across the original data from the study and realize that Igor had actually run the study as a repeated measures design. Now that you know that the data were collected as a repeated measures design, you re-compute the analyses appropriately, using StatView. [Notice that the means, etc., aren't displayed below, because they would not change from the earlier analysis, right? Note, also, that other parts of the source table don't change as you move to a repeated measures analysis.] Once again, analyze these data as completely as you can. Given what you know about the relationship between independent groups analyses and repeated measures analyses, the results might surprise you somewhat. However, you can readily explain the anomaly, right? [10 pts]

ANOVA Table for Arousal

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	5.467	.607				
Category for Arousal	2	69.267	34.633	20.284	<.0001	40.568	1.000
Category for Arousal * Subject	18	30.733	1.707				

First of all, you should note that the "Treatment" line (in this case "Category for Arousal") is the same as you found in the Independent Groups ANOVA. The only difference is in the Error term (in this case "Category for Arousal * Subject).

You should also note that for a repeated measures design, appropriate counterbalancing (complete) would lead to 6 orders, which means that you would need to run multiples of 6 participants. Ten wouldn't be appropriate.

You would still reject H_0 with this analysis ($p < .05$). Then you would need to compute a post hoc analysis. With q = 3.61, HSD = 1.49. Thus, your interpretation would be identical to that resulting from the independent groups ANOVA: Pleasant > Aversive > Neutral.

Why is your $F_{Obtained}$ smaller for the repeated measures ANOVA? Because the individual differences (as indexed by the $MS_{Subject}$) are relatively modest. Compared to the independent groups ANOVA, you've lost df in your error term (from 27 down to 18). Unless you also reduce the SS_{Error} by a proportionately greater amount, you will end up with a smaller F-ratio (as you did in this case). As an exercise, can you determine the size of the SS_{Error} that you would need to achieve to produce the same F-ratio in the repeated measures ANOVA as you had obtained in the independent groups ANOVA?

4. 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. [20 pts.]

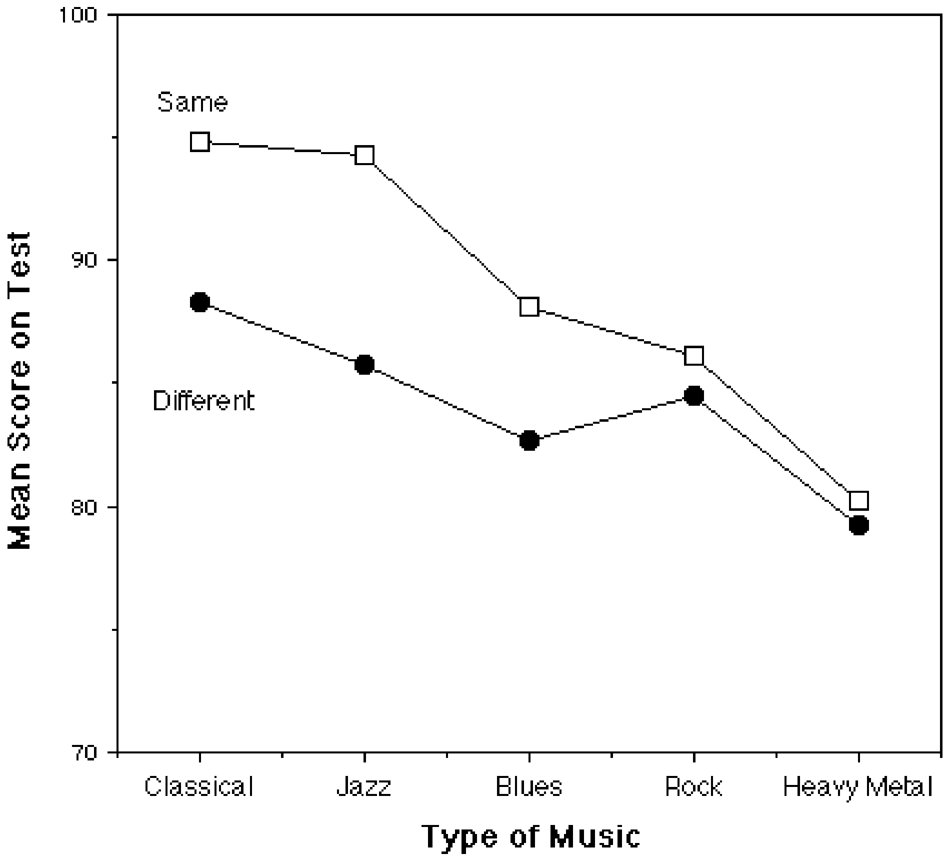
ANOVA Table for Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Music	4	1738.900	434.725	80.671	<.0001	322.682	1.000
Test	1	529.000	529.000	98.165	<.0001	98.165	1.000
Music * Test	4	207.100	51.775	9.608	<.0001	38.431	1.000
Residual	90	485.000	5.389				

Means Table for Score
Effect: Music * Test

	Count	Mean	Std. Dev.	Std. Err.
Blues, Different	10	82.700	1.252	.396
Blues, Same	10	88.100	2.132	.674
Classical, Different	10	88.300	3.529	1.116
Classical, Same	10	94.800	2.150	.680
Heavy Metal, Different	10	79.200	3.011	.952
Heavy Metal, Same	10	80.200	2.394	.757
Jazz, Different	10	85.800	3.048	.964
Jazz, Same	10	94.300	1.889	.597
Rock, Different	10	84.500	1.179	.373
Rock, Same	10	86.100	1.287	.407

Given the significant interaction, that's where I would focus my attention. My first step would be to create a graph to see what happened in the study. I've produced just such a graph below:



To my eyes, here's what happened. Performance was better when the test conditions matched the learning conditions (Same) compared to situations when they differed (Different) for Classical, Jazz, and Blues, but not for Rock and Heavy Metal. To gain some statistical confidence in what my eyes are telling me, I would compute Tukey's HSD. In this case, with 10 treatment conditions and 90 df_{Error}, I would get q = 4.6 and HSD = 3.38. Thus, if two means differ by 3.38 or more, they would be significant. Now I can see that my analyses support what my eyes had told me. That is, Same produced better performance than Different for Classical, Jazz, and Blues. However, for Rock and Heavy Metal keeping the conditions the Same from Learning to Test did not lead to better performance than changing the conditions at test.

You may want to comment on the fact that the "Different" condition was no music at all. Thus, for all conditions you have Music/Music for the Same condition and Music/No Music for the Different condition. Thus, it's hard to know if the effects observed in the Different condition are due to a difference between Learning and Test or are due to having No Music at test. However, if you had music at test for the Different condition, it's hard to know what kind of different music you would use. That could be worked out, however.

5. 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. [20 pts]

ANOVA Table for Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Hours Deprived	4	65.813	16.453	43.298	<.0001	173.193	1.000
Task Difficulty	2	120.027	60.013	157.930	<.0001	315.860	1.000
Hours Deprived * Task Difficulty	8	.107	.013	.035	>.9999	.281	.058
Residual	60	22.800	.380				

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

For this 3x5 independent group ANOVA, the interaction is not significant. Because both main effects are significant, that's where I would focus my attention.

First, for the main effect for Hours of Deprivation, the means are: 5.8, 4.3, 4.3, 3.5, and 3.1 for 24, 36, 48, 60, and 72 hours of deprivation, respectively. To determine which means are significantly different, I would compute HSD = .63 (q = 3.98). Thus, performance after 24 hours of deprivation was rated as significantly better than performance after 36 and 48 hours of deprivation (with performance equal in those two groups). Performance in the 24, 36, and 48 hour deprivation groups was rated as better than performance in the 60 and 72 hours of deprivation groups (with performance equal in those two groups).

For the main effect for Task Difficulty, the means are: 2.48, 4.52, and 5.52 for Easy, Moderate, and Difficult, respectively. To determine which means are significantly different, I would compute HSD = .42 (q = 3.4). Thus, Easy problems lead to significantly better ratings than Moderate or Difficult problems and Moderate problems lead to significantly better ratings of performance than Difficult problems.

6. [10 pts]

a. What is the general definition of a standard score?

$$\frac{\textbf{Score - Mean of Distribution}}{\textbf{Standard Deviation of Distribution}}$$

b. How does a *t* differ from a *z*?

t uses an estimate of σ, while z requires that you know σ

c. What is the purpose (use for) a z-score?

Compare scores on a common scale of standard deviation units.

d. What role does a z-score play in a correlation coefficient?

The correlation coefficient is actually the average of the product of two z scores:

$$\frac{z_x * z_y}{n}$$

e. What is the relationship between a *t* and an *F*?

$t^2 = F$

f. Increasing power would do what to a *t* score? Why?

Increase t. Power is the probability of correctly rejecting H_0, so an increase in power will lead to a larger value of the statistic.

g. What is the impact on *r* of adding a constant to each score in the analysis?

Nothing…because r involves z scores, so the addition of a constant would have no impact on the z scores.

7. On a standardized spatial skills task, normative data reveal that people typically get $\mu = 15$ correct solutions. A psychologist tests n = 7 individuals who have brain injuries in the right cerebral hemisphere. For the following data, determine whether or not right-hemisphere damage results in significantly reduced performance on the spatial skills task. [10 pts]

Participant 1	12
Participant 2	16
Participant 3	09
Participant 4	08
Participant 5	10
Participant 6	17
Participant 7	10

H_0: $\mu = 15$
H_1: $\mu \neq 15$

ss = 73.4
$s^2 = 12.23$
s = 3.5
sX = 1.32

$$t = \frac{11.7 - 15}{1.32} = -2.5$$

With $t_{Crit}(6) = 2.447$, you would reject H_0 because $t_{Obtained} \geq t_{Critical}$. Thus, you would conclude that sample is more likely to have been drawn from a population with $\mu < 15$.

8. A psychologist suspects that LSD affects the speech center in the brain. Specifically, he believes that repeated use of LSD reduces a person's ability to retrieve verbal information. To see if he can obtain any evidence regarding the relationship between LSD and retrieval of verbal information, the psychologist advertises for people who have taken LSD at least once. Nine people of comparable IQ and education are selected from the applicants. All participants are given a 50-item test. Each item consists of a definition of a low-frequency English word; the participant's task is to produce the target word. Sample items might be:

Definition	Target Word
To make things thinner or weaker by the addition of water	dilute
Patronage bestowed in consideration of family relationship and not merit	nepotism

Here are the relevant data. Analyze these data as completely as you can. [20 pts]

	Number of reported LSD trips	Number of failures to produce the target word (errors)
Participant 1	2	2
Participant 2	3	2
Participant 3	1	2
Participant 4	2	0
Participant 5	10	13
Participant 6	1	2
Participant 7	3	1
Participant 8	3	4
Participant 9	2	3
Sum	27	29
SS	60	117.56

H0: $\rho = 0$
H1: $\rho \neq 0$

$$\boldsymbol{r} = \frac{165 - \frac{(27)(29)}{9}}{\sqrt{(60)(117.56)}} = \frac{78}{83.99} = .93$$

Because $r_{Crit}(7) = .666$, you would reject H_0. Thus, you would conclude that there is a significant positive linear relationship between the number of reported LSD trips and the number of failures to produce the target word. Given the significant linear relationship, it would make sense to determine the regression equation. $r^2 = .86$. If you created a scattergram, you would note that Participant 5 is something of an outlier.

$b = \frac{78}{60} = 1.3$

a = 3.2 - (1.3 * 3) = -.7
Regression equation: y = 1.3x - .7