ID# ____________________ Exam 2 PS306, Fall 1998

Psssst. Hey, wanna take a break this weekend? If you have not already done so, why not go see Arcadia? I'm going! It will give me a chance to see some of the non-experimental skills that some of my students have acquired. Should be a good time!

OK, so you can't escape a commercial even on an exam…how bad is that! ☺ As always, the Skidmore Honor Code is in effect. Keep your eyes focused on your own exam. Point values are as indicated. Keep in mind that I think of a point as a minute, so spend the appropriate amount of time on each question. Good luck!

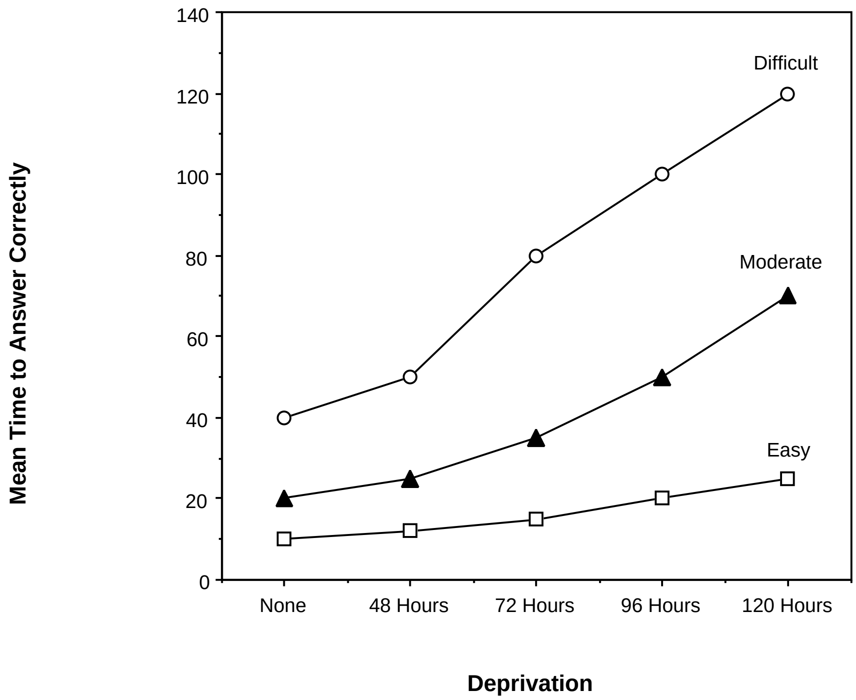
1a. Suppose that you are interested in conducting an experiment on the effects of sleep deprivation and task difficulty on performance. You operationally define performance as the time it takes a participant to complete the task correctly (better performance = faster time to complete the task). You decide that you want to use 5 levels of sleep deprivation. You also decide to use 3 levels of difficulty: easy tasks, moderate tasks, and difficult tasks. (You pre-test the tasks with a group of participants to determine the level of task difficulty.) You decide to use a completely between design (both factors independent groups), with an n = 25. Describe the study you would conduct in sufficient detail that I can tell that you know how to conduct such a study. Flesh out the details of the experiment: (1) the exact levels of sleep deprivation you would use; (2) how many people you would need to complete your study; (3) how you would run them through the study;

OK, for sleep deprivation it is critical that you have a control group that is not at all sleep deprived. So, my 5 levels would be Control (8 hours sleep the night before test), 48 Hours (no sleep for two days), 72 Hours (no sleep for three days), 96 Hours (no sleep for four days), and 120 Hours (no sleep for five days).

Because it's a 3x5 independent groups design with n = 25, I would need a total of 375 participants.

I would run the study in randomized replications. As such, I would create a random order of the 15 conditions (Control/Difficult; 96 Hours/Easy; 72 Hours/Easy; Control/Moderate; etc.) Then I would run the first 15 participants through the study with one in each of those conditions. After those participants have provided data, I would create a different random order of the 15 conditions and assign each of the next participants to one of each of those conditions.

(4) produce a figure or table to illustrate a set of results that you think you might obtain from this study and what the implications of the results would be in terms of the effects found in an ANOVA;



With these results, I would expect to find an interaction between Deprivation and Difficulty, as well as main effects for both Deprivation and Difficulty. I would focus my attention on the interaction, expecting that I would find that Easy problems are affected less by sleep deprivation (performance gets worse gradually) than Moderate and Difficult problems (performance decreases by a greater amount with increases in deprivation).

and (5) the d.f. that would be in your source table for the analysis of this experiment. [25 pts]

SOURCE	df
Sleep Deprivation	**4**
Task Difficulty	**2**
Depriv x Diff	**8**
Error	**360**
Total	**374**

1b. How would your design change if you made task difficulty a repeated factor? (Assume that you still want to have *at least* 25 scores in each condition.) Briefly describe how the number of participants would change, as well as the way you would run them through the various conditions. [10 pts]
With task difficulty as a repeated factor, I would need to counterbalance. With only 3 levels, I would use complete counterbalancing. Thus, I would need to run multiples of 6 participants. To achieve a minimum of 25 participants in each condition, I would need to run 30 participants for each of the 5 levels of the independent groups factor. Thus, I would need a total of 150 participants.

The 6 orders would be EMD, EDM, MED, MDE, DEM, DME. Thus, I would need to use each of the orders 5 times for each of the levels of sleep deprivation. I would still use randomized replications for the independent groups factor (sleep deprivation), so I would set up a random order of the 5 levels (96, Control, 48, 120, 72) for the first 5 participants and then also assign each of them to a particular order of the difficulty levels. Then I would have another random order of the 5 levels of sleep deprivation for the next 5 participants, each of whom would be assigned to an order of the difficulty levels that had not yet been used. I would keep collecting data along these lines until I had completed the study.

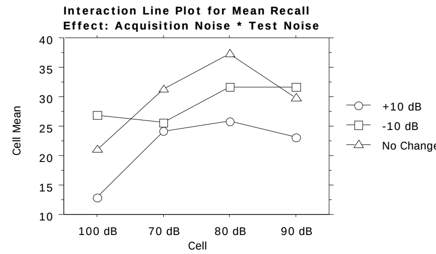
2. Several researchers have investigated the encoding specificity effect. The general finding is that people remember best when the testing situation is as similar as possible to the learning situation. (Thus, because the typical testing situation is a relatively quiet classroom, you'd best study/learn under conditions as similar to the testing situation as possible.) Dr. Ivana B. Loude was interested in further investigating this effect, to see the extent to which the learning and testing situations had to be similar. She decided to focus on the noise level of the room. Thus, she had 4 different noise levels (70 dB, 80 dB, 90 dB, and 100 dB) present while people watched a screen displaying pictures of 50 common objects (the acquisition phase of the experiment). After a brief distractor task, Dr. Loude tested some of the people from each noise level at one of 3 noise levels (10 dB less noise than acquisition, the same noise level, or 10 dB more noise than at acquistion). Thus, for instance, for the 70 dB acquisition group, one-third would be tested at 60 dB (-10 dB), one-third would be tested at 70 dB (same), and one-third would be tested at 80 dB (+10 dB). The dependent variable was the number of objects correctly recalled. Thus, this experiment is a 4x3 independent groups design. Below are the analyses of this experiment. The four acquisition noise levels are (1 = 70 dB, 2 = 80 dB, 3 = 90 dB, 4 = 100 dB). The three test noise levels are (1 = -10 dB from acquisition, 2 = Same as acquisition, 3 = +10 dB from acquisition). Interpret the results as completely as you can. [15 pts]

ANOVA Table for Mean Recall

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Acquisition Noise	3	2032.633	677.544	69.242	<.0001	207.726	1.000
Test Noise	2	1684.850	842.425	86.092	<.0001	172.184	1.000
Acquisition Noise * Test Noise	6	642.217	107.036	10.939	<.0001	65.632	1.000
Residual	108	1056.800	9.785				

Means Table for Mean Recall
Effect: Acquisition Noise * Test Noise

	Count	Mean	Std. Dev.	Std. Err.
100 dB, +10 dB	10	12.900	2.601	.823
100 dB, -10 dB	10	26.800	3.676	1.162
100 dB, No Change	10	21.000	4.643	1.468
70 dB, +10 dB	10	24.100	.738	.233
70 dB, -10 dB	10	25.700	2.111	.667
70 dB, No Change	10	31.300	2.163	.684
80 dB, +10 dB	10	25.800	2.486	.786
80 dB, -10 dB	10	31.600	5.358	1.694
80 dB, No Change	10	37.300	1.337	.423
90 dB, +10 dB	10	23.100	1.853	.586
90 dB, -10 dB	10	31.700	2.830	.895
90 dB, No Change	10	29.700	4.218	1.334



(First, note that the original output is from an older version of StatView. I've revised the output above so that it comes from a current version of StatView.)

Because there is a significant interaction, that's where I would focus my attention. With 12 means involved in the interaction, q = 4.74 and HSD = 4.7.

The interaction is fairly evident in the figure. As you can see, performance is usually best when the noise at acquisition and test are the same.

Using the HSD, I would claim that:
1) When the noise is 70 dB at acquisition, people perform significantly better if the noise stays the same at test, compared to increasing or decreasing the noise by 10 dB (either of those changes produces similar performance).
2) When the noise is 80 dB at acquisition, people perform significantly better if the noise stays that same at test, compared to increasing or decreasing the noise by 10 dB. For this level of noise, however, lowering the noise by 10 dB leads to better performance than increasing the noise by 10 dB.
3) When the noise is 90 dB at acquisition, people perform equally well if the noise stays the same or if the noise is decreased by 10 dB, and both of these levels of performance are better than increasing the noise by 10 dB at test.
4) When the noise is 100 dB at acquisition, people perform significantly better is the noise is decreased by 10 dB compared to keeping the noise the same or increasing it by 10 dB. Keeping the noise the same leads to better performance than increasing the noise by 10 dB.

In general, it appears that people perform better when the noise is not too loud. Because the 100 dB noise is fairly loud, increasing the noise to 110 dB (+10) or keeping it the same leads to worse performance than lowering the noise to 90 dB (-10). Thus, the encoding specificity effect has some limitations (given these completely fabricated data). That is,

having the noise level identical at acquisition and test does not always lead to better performance. Because the 90 dB noise is still fairly loud, increasing the noise to 100 dB (+10) leads to worse performance.

3. Dr. Julie Ard was interested in the effects of music on studying, also using an encoding specificity paradigm. 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 (1 = heavy metal, 2 = rock, 3 = classical, 4 = jazz, 5 = blues). People in these groups studied 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 (1 = same) and half of the people were tested with no music (2 = 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. [15 pts.]

The Source Table would look like this:

SOURCE	SS	df	MS	F	Pvalue
Music	**1872**	**4**	**468**	**21.1**	**.0001**
Test Situation	**823**	**1**	**823**	**37.1**	**.0001**
Music x Test	**150**	**4**	**37.5**	**1.7**	**.1579**
Error	**1999**	**90**	**22.2**		

Test Situation:		level 1	level 2	Totals:
Music	level 1	10 75.8	10 68.8	20 72.3
	level 2	10 82.6	10 75	20 78.8
	level 3	10 84.6	10 79.7	20 82.15
	level 4	10 86	10 84.7	20 85.35
	level 5	10 84	10 76.1	20 80.05
Totals:		50 82.6	50 76.86	100 79.73

In the absence of an interaction, I would focus my attention on the two main effects. Because the Test Situation has only two levels (Same, Different) no HSD would be necessary. Thus, I can say that people perform significantly better when the test setting is the same as the acquisition setting (M = 82.6) compared to situations in which the test situation differs from the acquisition setting (M = 76.86). However, you should note that there is a confound, in that there is no "No Music" acquisition group, so that the "Different" test is also always a No Music group as the experiment is currently designed.

For the main effect of music, I would compute HSD = 3.55 (q = 3.37). Thus, Jazz = Classical, but Jazz > Heavy Metal, Rock, & Blues. Classical = Blues and Rock, but Classical > Heavy Metal. Blues = Rock, but Blues > Heavy Metal.

4. People who are just starting to think about conducting experiments are often tempted to first collect some data on all the participants in a study and then match them on particular characteristics. In class, we discussed reasons why this approach might not be a great idea. Another common desire is to measure people on the dependent variable before putting people in any condition (pre-test) then measuring their performance on the dependent variable after the manipulation(s). Neither of these approaches is necessary with appropriate experimental design, right? If you use random assignment to conditions and have a sufficient number of participants you should be ok. First, tell me how random assignment to conditions works to solve the problems addressed by matching and pre-tests. Next, to show that you understand the nature of the problems, give me an example of a confounded experiment in which people are not randomly assigned to conditions and explain why the lack of random assignment would represent a confound. [10 pts]
With a sufficient number of participants and random assignment to conditions, your experiment should have conditions that are equivalent (on average). That is, each condition should have roughly equal levels of intelligence, motivation, etc. Attempting to match conditions on particular variables (like IQ or motivation) could lead to problems if you fail to match on a critical variable (e.g., suppose that SES is critical and you fail to match on that variable). The problem with pre-test/post-test designs is that you're letting the participants know what the test is like before you've even administered the treatment. That might lead to a lessening of the impact of the treatment.

Suppose that you allow participants to assign themselves to conditions. You are interested in conducting a study of the effects of two different styles of studying (massed vs. distributed) on learning. On your sign-up sheet, you let people know about the two conditions and let them choose which condition they want to be in. The problem, of course, is that the smarter people may select to be in the distributed studying condition and the dumber people may choose to be in the massed studying condition. Then, when you find a difference between the two conditions (i.e., distributed > massed studying) you wouldn't know if the difference was due to the manipulation or due to the fact that the people in the two conditions differed in intelligence (for example).