Final Exam PS 306, Spring 2001

1. Historically, psychology may be characterized as the study of the behavior of white males. A while ago, there was a book titled *Even the Rat was White*, illustrating the bias present in psychological research. Of course, such bias is a two-edged sword. What is *good* about collecting data from white males only? What is *bad* about collecting data from white males only? [5 pts]

Studies are more powerful if they have smaller error terms. By studying only white males, one would be doing more powerful research. OTOH, studying only white males would limit the generalizability of the research.

2a. Two researchers were interested in studying the effects of reward magnitude on performance. Both researchers draw from the same pool of participants, use the same total number of participants (24), the same type of reward and reward magnitudes ($1, $5, $20), the same apparatus, the same task, and the same performance measure (DV). One researcher uses an independent groups design and, on the basis of the results, cannot reject the null hypothesis (that reward has no effect on performance). The other researcher uses a repeated measures design and finds a statistically significant effect of reward magnitude — larger rewards lead to better performance. Assume that neither study has a major flaw (e.g., repeated measures design is properly counterbalanced). There are three fundamental reasons why the two researchers might have reached different conclusions. One reason concerns the sensitivity of the test of the null hypothesis. Another reason concerns the nature of the participant's experience in the two studies. A final reason has to do with the tentativeness of hypothesis testing, regardless of whether or not a researcher rejects the null hypothesis. Provide me with a clear explanation of the reasons that the two researchers may have reached such different conclusions. Would you trust the results of one study more than the other? Why? [15 pts]

a. You would expect the two studies to differ in their results because of the fact that the repeated measures design is more powerful. Thus, you might reject H_0 with a repeated measures design while retaining H_0 with an independent groups design.

b. Because of carry-over effects in the repeated measures design, you might get tainted results. Imagine, for instance, that you've been paid $20 and now the experimenter wants you to do the same task, but will only pay $1. How hard are you likely to work? Thus, the $20->$1->$5 order may produce quite different results from the $1->$5->$20 order. The carry-over effects, even given appropriate counterbalancing, may limit the interpretability of the data.

c. All of our conclusions are tentative because of the possibility that we've made an error. That is, one researcher may appropriately reject H_0 and the other researcher may inappropriately retain H_0 (i.e., a Type II error).

I would accept any conclusion, as long as the rationale was compelling. For myself, I'd probably prefer the independent groups design in this case because of concerns about the strong carry-over effects. I would, however, probably use a larger sample size to increase power.

2b. Finally, complete the source tables for the two experiments, as seen below. {Remember, the RM design is more efficient, so participants generate more than one piece of data.} [10 pts]

Independent Groups Design:

Source	SS	df	MS	F	
Between Treatments	24	**2**	**12**	**3.40**	$F_{Crit} = 3.47$
Within Treatments (Error)	**74**	**21**	**3.52**		
Total	98				

Repeated Measures Design:

Source	SS	df	MS	F	
Between Treatments	14	**2**	**7**	**3.5**	$F_{Crit} = 3.20$
Between Subjects	88	**23**	**3.82**		
Error (Residual)	**92**	**46**	**2**		
Total	194				

3. Suppose that you were interested in conducting a 4x6 mixed design, with the first variable (with 4 levels) an independent groups factor and the second variable a repeated measures variable. Suppose, further, that you needed to obtain a minimum of 35 pieces of data per condition (due to power considerations). What's the minimum number of participants that you would need for your study? [5 pts]

	b1	b2	b3	b4	b5	b6
a1						
a2						
a3						
a4						

First, for a1 I would need to counterbalance 6 levels (conditions). That would lead me to use incomplete counterbalancing, and with an even number of levels (6) I would need 6 orders. However, to get above 35, I would need to use each order 6 times and would run 36 people. That would complete a1. I would need to do the same for a2, a3, and a4, which would lead to a total of 144 participants.

4. We discussed the study that Doob & Wood (1972) conducted, reported in "Catharsis and aggression: Effects of annoyance and retaliation on aggressive behavior." As they state, "The catharsis hypothesis of aggression usually refers to a decrease in aggression after the expression of aggression. The assumption as stated by Freud is that there is a certain amount of aggression that has to be expressed, and that once this has happened, there is less left to be expressed later on."

As you may recall, half the participants were first annoyed by the confederate (who made all sorts of nasty personal comments), and the other half were not annoyed by the confederate. Then one-third of each group: (1) shocked the confederate, who had been moved into an adjoining room, for errors in a learning task, (2) watched the experimenter shock the confederate, who had been moved into an adjoining room, for errors in a learning task, or (3) saw nothing, as the experimenter took the confederate into an adjoining room to administer the "learning test." In the final phase of the experiment, participants judged the "creativity" of associations made by the confederate to words presented by the subject. If they thought that a response was uncreative they gave the confederate a shock. They were told to increase the duration of the shock for increasingly uncreative responses. Confederates (who did not really receive any shocks) were cautioned by the experimenter not to comment if a shock was delivered, and then the experimenter left the room.

Doob & Wood were interested in testing the hypothesis that participants would experience a cathartic effect from observing the boorish confederate being shocked, or shocking her themselves. This would lead them to be less likely to shock the confederate in the final (judging creativity) phase of the experiment.

Participants were given an informed consent form that told them only that they were in a learning experiment and that they might be given electric shocks. They were told that they could withdraw from the experiment, etc. Finally, the participants were debriefed about the fact that the other “participant” was actually a confederate of the experimenter and that they had not really shocked the confederate. If you were on a participant review board presented with a proposal for this experiment, what would be your reaction, and why? Be sure to draw specifically on the APA guidelines. [5 pts]

For ethical questions such as this one, I’m less interested in your decision (approve or reject) than I am in your reasoning using the APA guidelines as articulated in class and as seen in your textbook.

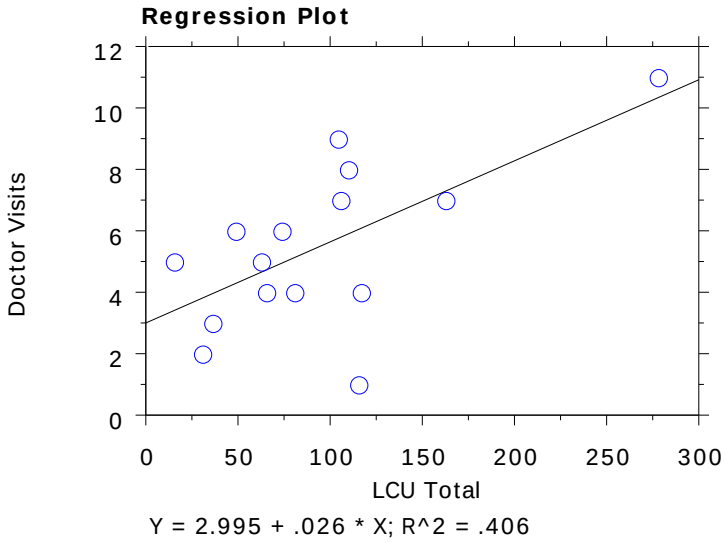
5. Briefly, but clearly, describe the design and major results of the two studies by Ross, Lepper & Hubbard (1975) on perseverance in self-perception and the effectiveness of debriefing. What are the implications for ethical treatment of human subjects? [10 pts]

A clear summary of the studies in the Ross, et al. paper and their implications for debriefing was necessary for a good answer to this question.

6. Studies have suggested that the stress of major life changes is related to subsequent physical illness. Holmes and Rahe (1967) devised the Social Readjustment Rating Scale (SRRS) to measure the amount of stressful change in one's life. Each event is assigned a point value, which measures its severity. For example, at the top of the list, death of a spouse is assigned 100 life change units (LCU). Divorce is 73 LCUs, retirement is 45, change of career is 36, the beginning or end of school is 26, and so on. The more life change units one has accumulated in the past year, the more likely he or she is to have an illness. The following StatView analyses show the results from a hypothetical set of data. Interpret these results as completely as you can. For these data, if a person had accumulated 100 LCUs, how many doctor visits would you predict? Provide three possible interpretations for the observed relationship. [15 pts]

Regression Summary
Doctor Visits vs. LCU Total

Count	15
Num. Missing	0
R	.637
R Squared	.406
Adjusted R Squared	.360
RMS Residual	2.135



ANOVA Table
Doctor Visits vs. LCU Total

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	40.469	40.469	8.877	.0107
Residual	13	59.264	4.559		
Total	14	99.733			

Regression Coefficients
Doctor Visits vs. LCU Total

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	2.995	.996	2.995	3.006	.0101
LCU Total	.026	.009	.637	2.979	.0107

First of all, you should note that there is a significant linear relationship (r = .637) between doctor visits and LCU's (F(1,13) = 8.877, p = .0107). With r2 = .406, these two variables share roughly 40% of their variability. That is, ~40% of the variability in Doctor Visits is associated with LCU's and 60% is associated with other variables. Given the significant linear relationship, 100 LCU's would lead you to predict 5.6 doctor visits. It could be that high LCU's cause illness, which leads to doctor visits. It could be that going to the doctor leads to greater LCU's. Or, it could be that people who are really sick (or unlucky?) tend to go to the doctors a lot AND to have high LCU's.

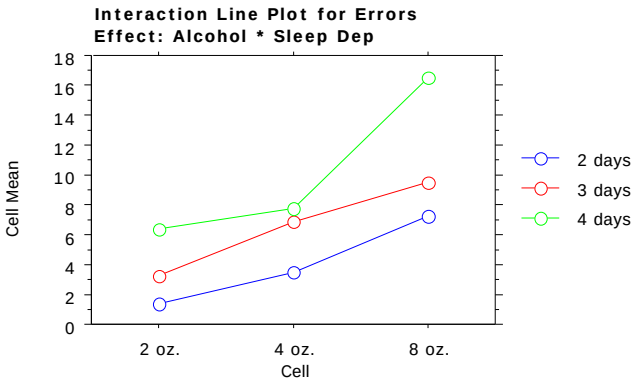
7. Dr. Noah Weisser was interested in investigating the effects of sleep deprivation and alcohol on driving ability. To that end, he conducted a 3x3 independent groups study. Participants were given identical 32 oz. drinks that could contain 2, 4, or 8 ozs. of vodka (with the remainder of the glass filled with orange juice). Within each level of alcohol, one-third of the participants were deprived of sleep for 2 days, one-third were deprived of sleep for 3 days, and the final third were deprived of sleep for 4 days. Each of the participants operated a driving simulator for 30 minutes, during which the number of driving errors (objects hit, crossing into the oncoming lane, etc.) were recorded. Complete the analysis shown below and interpret the results as completely as you can. [20 pts]

ANOVA Table for Errors

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Alcohol	2	688.528	344.264	167.966	<.0001	335.932	1.000
Sleep Dep	2	461.778	230.889	112.651	<.0001	225.301	1.000
Alcohol * Sleep Dep	4	93.222	23.306	11.371	<.0001	45.483	1.000
Residual	63	129.125	2.050				

Means Table for Errors
Effect: Alcohol * Sleep Dep

	Count	Mean	Std. Dev.	Std. Err.
2 oz., 2 days	8	1.375	1.061	.375
2 oz., 3 days	8	3.250	1.035	.366
2 oz., 4 days	8	6.375	1.188	.420
4 oz., 2 days	8	3.500	.926	.327
4 oz., 3 days	8	6.875	.991	.350
4 oz., 4 days	8	7.750	.707	.250
8 oz., 2 days	8	7.250	1.282	.453
8 oz., 3 days	8	9.500	1.195	.423
8 oz., 4 days	8	16.500	3.071	1.086



Because the interaction is significant (p < .0001), I would focus my attention on explaining that outcome. The figure illustrates the nature of the interaction. I would next compute post hoc analyses to determine which of the means differed. With 9 means and dfError = 63, I would get q = 4.55 and HSD = 2.3. Thus, with 2 oz. of alcohol, more errors were made be people with 4 days of sleep deprivation compared to people with either 3 or 2 days of sleep deprivation. People with 2 oz. of alcohol and 2 or 3 days of sleep deprivation did not differ. The same basic pattern is also observed with 8 oz. of alcohol. However, with 4 oz. of alcohol 3 and 4 days of sleep deprivation both lead to more errors than 2 days of sleep deprivation, and 3 and 4 days don't differ. You should note the obvious lack of appropriate control groups (0 oz. alcohol and 0 sleep deprivation).

8. Briefly define experimenter expectancy effects and demand characteristics. Then, tell me why experimenter expectancy effects might be considered demand characteristics. Use Rosenthal's study on the "early data returns" effect to illustrate your point. [15 pts]

Answering this question requires that you've mastered the information on experimenter expectancy effects and demand characteristics. If you have, then it should make sense to you that the experimenter is another aspect of the participant's experience in the experiment and may well provide the active participant with clues as to the purpose of the study.