Exam 2 PS 306, Spring 2010

1. Well, of course you expect to tell me about the impact of various designs on the number of participants needed. For this problem, assume that we want to have a minimum of 30 pieces of data in each cell/condition. [10 pts]

Design	# of participants	# of pieces of data
A 5x6 completely between (independent groups) design	**900**	**900**
A 5x6 completely within (repeated measures) design	**30**	**900**
A 5x6 mixed design, with the first factor between (independent groups) and the second factor within (repeated measures)	**150**	**900**
A 5x6 mixed design, with the first factor within (repeated measures) and the second factor between (independent groups)	**Inc: 180** **Comp: 720**	**Inc: 900** **Comp: 3600**
A 4x7 mixed design, with the first factor within (repeated measures) and the second factor between (independent groups)	**336**	**1344**

2. Suppose that you are interested in the effects of a particular drug on the concentration level of young children with Attention Deficit Disorder. You conduct a simple experiment in which an experimental group gets the drug and a control group gets a placebo. When you analyze your data, you find that your effect is not significant, with $p < .09$. Initially, you are disappointed, but then you remember that you had an equal number of male and female students in each group. How might that information help you at this stage of your experiment? In other words, tell me what's lacking in your initial analysis of the study and how you might fix that problem by taking advantage of the fact that you have data from an equal number of males and females in each of your two groups. When would this approach not be helpful? Be as explicit as possible. (5 pts)

The error term may be inflated by differences between males and females. By introducing gender as a factor, the new error term (MS_{Error}) would be smaller—*if* males and females differed in their response. And, of course, a smaller denominator will yield a larger *F*-ratio.

3. For Lab 3, you collected data for a 2x2 independent groups design in which the two factors were Perceived Time (Distant Past vs. Recent Past) and Type of Event (Excelled vs. Suffered). There were a number of different dependent variables, including a *Purpose in Life* subscale. Complete the source table below, then interpret the results as completely as you can. Finally, briefly discuss/interpret the results you obtained (as you might in your discussion section). (15 pts)

Descriptive Statistics

Dependent Variable:Purpose in Life

Time	Event	Mean	Std. Deviation	N
Recent	Suffered	71.6	12.6	20
	Excelled	67.7	14.9	20
	Total	69.9	13.6	40
Distant	Suffered	68.7	13.8	20
	Excelled	76.3	9.3	20
	Total	72.9	12.0	40
Total	Suffered	70.3	13.1	40
	Excelled	72.6	12.6	40
	Total	71.4	12.8	80

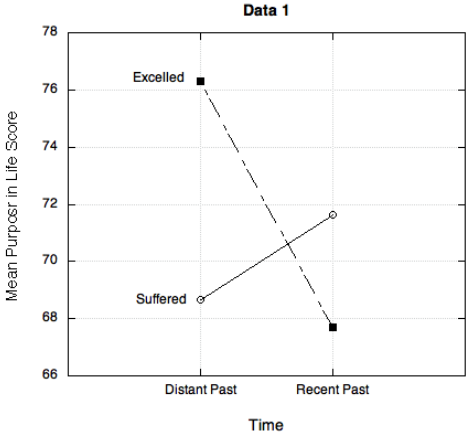
Levene's Test of Equality of Error Variances[a]

Dependent Variable:Purpose in Life

F	df1	df2	Sig.
1.479	3	75	.227

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Time + Event + Time * Event



Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
Time	68.39	1	68.39	.97	.327	.013	.164
Event	29.61	1	29.61	.42	.517	.006	.098
Time * Event	289.05	1	289.05	4.10	.047	.052	.515
Error	5358.00	76	70.5				
Corrected Total	5745.05	79					

The Levene test indicates that there's no concern about heterogeneity of variance, so use α = .05.

$$HSD = 3.72\sqrt{\frac{70.5}{20}} = 6.98$$

There is no main effect of Time, $F(1,76) = .97$, $MSE = 70.5$, $p = .327$, $\eta^2 = .013$. There is no main effect of Event, $F(1,76) = .42$, $p = .517$, $\eta^2 = .006$. However, there is a significant interaction between Time and Event, $F(1,76) = 4.10$, $p = .047$, $\eta^2 = .052$. Post hoc tests using Tukey's HSD indicate that for the Distant Past, Purpose in Life scores are higher for Excelled ($M = 76.3$) than for Suffered ($M = 68.7$). However, for the Recent Past, Purpose in Life scores were similar for Excelled ($M = 67.7$) and Suffered ($M = 71.6$). OR Post hoc tests using Tukey's HSD indicate that for Excelled, Purpose in Life scores are higher for the Distant Past ($M = 76.3$) than for the Recent Past ($M = 67.7$). However, for Suffered, Purpose in Life scores were similar for the Distant Past ($M = 68.7$) and the Recent Past ($M = 71.6$).

4. A common desire—especially in novice researchers—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). (Recognize anyone, Brendan? ☺) This approach isn't 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 pre-tests. Next, to show that you understand the nature of the problem, 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. Finally, tell me one downside of doing a pre-test. [10 pts]

With random assignment to conditions, along any dimension you might choose, the average score for one group should be about the same as the average score for any other group. Of course, that will be most true when your samples are sufficiently large so as to allow the averaging process to work for you.

Here's one example of the lack of random assignment. Suppose that you set up two times to participate in your experiment (1:00 and 5:00). If you run Condition 1 at 1:00 and Condition 2 at 5:00, you run the risk that the people who can participate at 1:00 differ from those who can participate at 5:00. For example, those at 1:00 may be coming right after

lunch, which may make them drowsy. ☺ However, they may also be science majors who have labs at 5:00, so they have to come to the 1:00 experiment. Note that with the problems I've articulated here, even random assignment (the experimenter determines which time the subject comes) wouldn't work, because some subjects would simply be unable to come to one time or the other.

The problems with a pre-test are many. First of all, if you're using a different test for the pre-test, then you have to construct at least two different tests. Secondly, you are indirectly informing your subjects about the nature of the study. And finally, assuming the same "test" at both pre-test and post-test, there is the possibility that the subjects will try to ensure that their responses are uniform, trying to give the same response on the post-test as they had done on the pre-test.

5. Helmrich, Aronson, and LeFan (1970) examined the effects of seeing a person commit a social blunder. The participants were divided into three conditions on the basis of their self-esteem scores (Low, Medium, High). Some participants in each self-esteem group saw a competent person accidentally spill a cup of coffee on the floor (Spill). The other participants saw the competent person in the same situation but not spilling the coffee (No Spill). All participants were asked to indicate how much they liked the person on a 20-point scale (20 = like a lot). Below is a partially completed source table consistent with results from this study. Complete the source table and then interpret the results as completely as you can. (15 pts)

Descriptive Statistics

Dependent Variable:Liking

SelfEsteem	CoffeeSpill	Mean	Std. Deviation	N
Low	No Spill	12.2000	2.04396	10
	Spill	5.0000	1.24722	10
	Total	8.6000	4.04449	20
Medium	No Spill	8.0000	1.76383	10
	Spill	13.8000	1.31656	10
	Total	10.9000	3.33877	20
High	No Spill	13.3000	1.49443	10
	Spill	10.3000	1.56702	10
	Total	11.8000	2.14231	20
Total	No Spill	11.1667	2.88974	30
	Spill	9.7000	3.91417	30
	Total	10.4333	3.49026	60

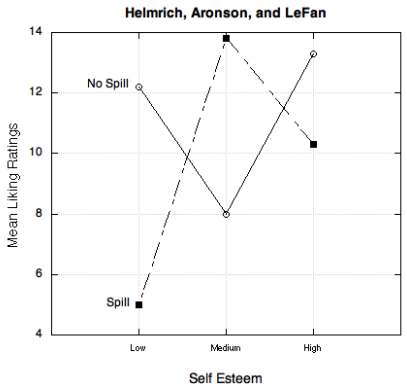
Levene's Test of Equality of Error Variances[a]

Dependent Variable:Liking

F	df1	df2	Sig.
.651	5	54	.662

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + SelfEsteem + CoffeeSpill + SelfEsteem * CoffeeSpill



Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
SelfEsteem	108.7	2	54.4	21.4	.000	.442	1.000
CoffeeSpill	32.3	1	32.3	12.7	.001	.190	.938
SelfEstm * CoffSpill	439.4	2	219.7	86.5	.000	.762	1.000
Error	137.4	54	2.5				
Corrected Total	717.8	59					

The Levene test indicates that there's little reason to be concerned about heterogeneity of variance. Thus, analyses were conducted with $\alpha = .05$.

$$HSD = 4.18\sqrt{\frac{2.5}{10}} = 2.11$$

There is a significant main effect of Self Esteem, $F(2,54) = 54.36$, $MSE = 2.5$, $p < .001$, $\eta^2 = .442$. There is also a significant main effect of Coffee Spill, $F(1, 54) = 12.7$, $p = .001$, $\eta^2 = .190$. The interaction between Self Esteem and Coffee Spill is also significant, $F(2.54) = 219.7$, $p < .001$, $\eta^2 = .762$. Post hoc analyses using Tukey's HSD indicate that the interaction is due to the fact that for people with high or low self esteem, the Liking rating was significantly higher for the person who didn't spill the coffee ($M = 13.3$ for High and $M = 12.2$ for Low) than for the person who did spill the coffee ($M = 10.3$ for High and $M = 5.0$ for Low). However, for people with medium self esteem, people who didn't spill the coffee had significantly lower Liking ratings ($M = 8.0$) than did people who did spill the coffee ($M = 13.8$). OR Post hoc analyses using Tukey's HSD indicate that the interaction is due to the fact that for people who spilled the coffee, people with medium self esteem liked the person more ($M = 13.8$) than did people with high self esteem ($M = 10.3$) or people with low self esteem ($M = 5.0$). However, for people who didn't spill the coffee, people with high or low self esteem liked the person more ($M = 13.3$ for High and $M = 12.2$ for Low) than did people with medium self esteem (M = 8.0).

6. You should know a bit about the bystander apathy effect (e.g., some discussion in the textbook and some discussion in class—Kitty Genovese). Remember that Lance Shotland studied the extent to which people were more or less likely to go to the aid of a woman being accosted by a man if it did ("I don't know why I married you!") or did not ("I don't know you!") appear that the two people were in a relationship. Suppose that you were interested in designing a study in which you looked at this variable (Type of Relationship) and also at the number of bystanders present. You are limited to 3 levels of the Number of Bystanders variable, but you are free to choose the specific levels. (20 pts)

a. Briefly describe the design that you would use for this study (completely between/independent groups, completely within/repeated measures, or mixed) and tell me why you would choose to do so.

Because the study involved deception, it would be virtually impossible to conduct the study as a repeated measures design. (And you certainly wouldn't want participants who couldn't catch on to the fact that you'd fooled them on earlier occasions.) Thus, you'd use an independent groups design.

b. You want to have a minimum of 25 scores per cell for this study. Describe the study you would conduct in sufficient detail that I can see that you have given thought to how you would actually go about constructing the experiment and collecting the data.

In my 2x3 independent groups design, I would have two levels of the Relationship variable (Know One Another vs. Strangers) and three levels of the number of bystanders (0, 1, and 5). Thus, I'd need to run a total of 150 people. I'd probably set it up so that I sent the subject (or subject plus 1 or 4 confederates) from a waiting room to an experimental room, but along the way would be a staged altercation between a man and a woman, where the woman would either indicate that she knew the attacker ("I don't know why I married you!") or didn't know the attacker ("I don't know you!"). I would time the subject's response to help the woman (waiting 3 minutes…so if no intervention by 3 minutes, that would be the maximum response time…essentially no response). Thus, the DV would be time in seconds to respond, from a minimum of 0 to a maximum of 180 seconds. I would operationally define intervention as any effort (even verbal) to get the man to stop abusing the woman.