Exam 2 PS 306, Fall 2002

1. One problem that you might encounter in a study that could lead to a confound is mortality or attrition. Carefully describe a hypothetical experimental scenario in which such attrition would be a confound. Then describe a scenario in which your study might encounter attrition, but it would not lead to a confound. Clear logic and sufficient detail in your examples are both important. [10 pts]

Generally speaking, participant mortality or attrition is only a problem if it is extensive (proportionally lots of data lost) and falls unevenly on your conditions. For example, suppose that you were doing a study to look at the effectiveness of two weight-loss programs (A & B). You might gain access to a list of people who are interested in losing weight through some program and randomly assign them to one of three conditions (placebo control, A, & B). Let's assume that the "programs" would continue for a year, with weight assessments (the DV) taken at the end of the year. (Of course, you'd probably want to take weight measurements more often.) Suppose that you start out with 50 people in each of the three conditions, but at the end of the year when you went to take the weight measurement, only 30 people were left in the placebo condition, 25 people in the Program A condition, and 45 people in the Program B condition. The disproportionate number of people missing from the Program A condition should concern you. You might imagine that people in that group became discouraged, so they dropped out of the program. Or they might have been so successful that they discontinued the program. In either case, the 25 people remaining would not be an accurate representation of the effectiveness of Program A.

Suppose, however, that in the study you had less attrition or more similar attrition. Suppose that at the end of one year, 35 people were left in the placebo condition, 37 people were left in the Program A condition, and 36 people were left in the Program B condition. Unless you could think of some reason why the similar levels of attrition are due to different causes (leave because of success in Program A, but leave because of failure in Program B), I would think that the attrition is less of a problem in this scenario.

2a. Suppose that you are interested in determining the extent to which either or both of two drugs (Drug A and Drug B) might lead to improved performance on problem solving tasks for mentally retarded children. To gain a sense of the nature of the impact of the treatments, you decide to test each child on four different types of problems chosen to be appropriate for the mental abilities of such mentally retarded children (spatial problems, math problems, logic problems, and manual dexterity problems). Because you are concerned about the problems of carry-over effects, you decide to use a mixed design with the Drug factor between groups (independent groups) and the Problem factor within groups (repeated measures). Carefully describe the design of your study, the procedure you would use in conducting your study, the number of participants you would use in your study, and any relevant details that will make clear that you've thought carefully about the design of this study. [20 pts]

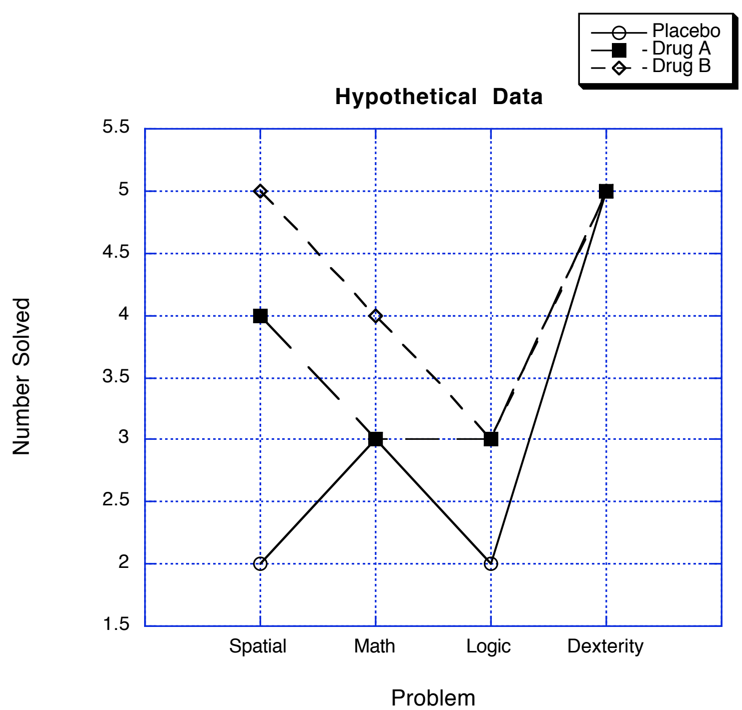
It is crucial that you provide sufficient detail in your depiction of the study you want to do. In this case, it is vital that you have at least three levels of your independent groups factor (Drug): Placebo, Drug A, and Drug B. Then, it is crucial that you use complete counterbalancing of the repeated measures factor (Problem). Complete counterbalancing would involve 24 different orders: Sp, Ma, Log, Dex; Sp, Ma, Dex, Log; Sp, Log, Ma, Dex; Sp, Log, Dex, Ma; Sp, Dex, Ma, Log; Sp, Dex, Log, Ma; etc. Thus, you'd need a minimum of 72 participants (24 x 3).

What might you use as the DV? Time to solve X problems? Number of problems correctly solved in X minutes? You'd need to be sensitive to the abilities of your participants, so that you didn't exceed their attention span, etc. So, you'd want to try to keep your total time under a certain number of minutes. And I think that you'd really want to keep the study constrained to one testing time rather than multiple days.

So, let's presume that we would have pre-tested 10 problems of each type (to ensure that they are appropriate) and we'll use number of problems solved in 5 minutes as our DV. We might also build in a brief rest/play period between problems. Thus, the whole procedure should take around a half hour.

The procedure might be something like: participant is given a pill to swallow (with the experimenter blind to the treatment), then play for a half hour to allow the drug to take effect, then Math problems (5 mins) , then rest (5 mins), then Logic problems, then rest, then Dexterity problems, then rest, then Spatial problems (as one possible order).

2b. In the "graph" below, indicate the means that might result from your study. Then tell me what your means would lead you to predict about the results of an ANOVA on the data. [10 pts]



These data certainly depict an interaction between the two factors. It appears that Dexterity problems aren't affected by the drugs, but the drugs both appear to help on Spatial problems, and (to a lesser extent) on Logic problems. Drug B seems to help on Math problems, but not Drug A. There also appears to be a main effect for Drug (Drug B > Drug A > Placebo) and a main effect for Problems (Dexterity > Spatial > Math > Logic). Of course, you would focus your attention on the interaction.

3. An experiment is done to test the effect of age and dress of a speaker on attitude of a college audience. Groups of college students listen to a speaker talk on why Slabotnik should be president. The speaker is either neatly dressed (tie and jacket) or grubbily dressed (ripped blue jeans and torn t-shirt). Additionally, the speaker is either in his 20's, 30's, or 40's. After the talk, the subjects rate their attitude about whether Slabotnik should be elected, on a scale ranging from 0 (should not be elected) to 7 (should be elected). Complete the source table below and interpret the study as completely as you can. [20 pts]

What are your three H_0's?

Age H_0: $\mu_{20} = \mu_{30} = \mu_{40}$
Dress H_0: $\mu_{neat} = \mu_{grubby}$
Interaction H_0: No interaction

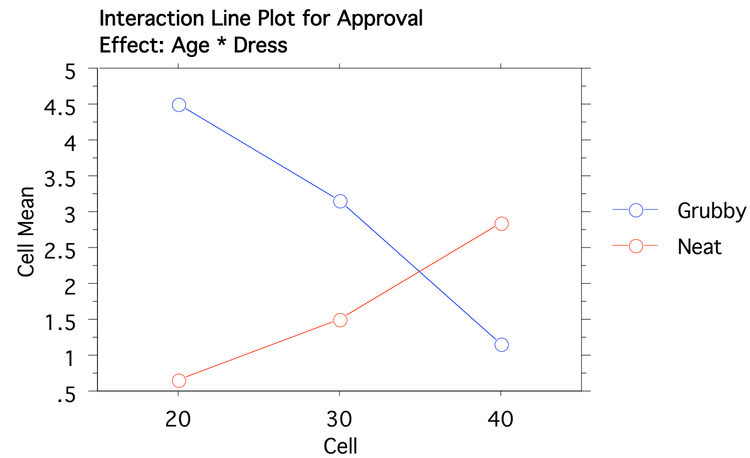
What value represents your best estimate of σ^2? **MS_{Error} = .628**

ANOVA Table for Approval

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Age	2	2.056	1.028	1.637	.2114	3.274	.307
Dress	1	14.694	14.694	23.407	<.0001	23.407	.999
Age * Dress	2	46.056	23.028	36.681	<.0001	73.363	1.000
Residual	30	18.833	.628				

Means Table for Approval
Effect: Age * Dress

	C...	M...	Std. ...	Std. ...
20, Gr...	6	4.5	1.049	.428
20, Neat	6	.667	.816	.333
30, Gr...	6	3.17	.753	.307
30, Neat	6	1.5	.548	.224
40, Gr...	6	1.17	.753	.307
40, Neat	6	2.83	.753	.307



The interaction is significant ($p < .0001$), so it is necessary to compute the HSD to interpret the results:

$$HSD = 4.3\sqrt{\frac{.628}{6}} = 1.39$$

When the speaker is 20 or 30 years old, the audience is more likely to approve of Slabotnik if the speaker is dressed in a grubby fashion rather than a neat fashion. However, when the speaker is 40 year old, the audience is more likely to approve of Slabotnik if the speaker is dressed in a neat fashion than in a grubby fashion. You should note that Age is not a manipulated factor, so you can't be sure that it's age *per se* that is producing the effect. It's also potentially problematic that there is only one speaker at each age group, which means that it might be the speaker producing the results, and not the age of the speaker.

4. Dr. B. N. Goode is an educational psychologist who is interested in the impact of different textbooks on learning. He wants to assess the benefits of three different statistics texts (by Drs. Aaron, Brown and Curtis). To conduct the study, he gets one of his colleagues to agree to use the Aaron text, another colleague to use the Brown text, and a third colleague to use the Curtis text as they teach their statistics courses. Because he's also interested in determining if one text might be more appropriate for a particular type of student, he divides each class into thirds, based on prior GPA (Low, Medium, High). Dr. Goode gives all three classes a standard test and uses the scores on the test as his dependent variable. Complete the source table below and interpret the results of this study as completely as possible. [20 pts]

ANOVA Table for Test Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Textbook	2	1312.111	656.056	168.614	<.0001	337.228	1.000
GPA	2	4700.861	2350.431	604.088	<.0001	1208.176	1.000
Textbook * GPA	4	31.889	7.972	2.049	.0982	8.196	.574
Residual	63	245.125	3.891				

Means Table for Test Score
Effect: Textbook * GPA

	Count	Mean	Std. Dev.	Std. Err.
Aaron, High	8	90.500	2.673	.945
Aaron, Low	8	70.625	1.061	.375
Aaron, Med	8	81.750	2.315	.818
Brown, High	8	87.500	1.604	.567
Brown, Low	8	66.500	2.204	.779
Brown, Med	8	75.625	1.188	.420
Curtis, High	8	96.000	1.309	.463
Curtis, Low	8	77.500	2.976	1.052
Curtis, Med	8	87.375	1.408	.498

First of all, you should note the confound between textbook and teacher. Thus, any differences that emerge may be due to textbook OR teacher.

That said, the interaction is not significant, so you would focus on the main effects, both of which are significant. The same HSD would work for both main effects:

$$HSD = 3.4\sqrt{\frac{3.891}{24}} = 1.37$$

The three textbook means are Aaron = 81, Brown = 76.5, Curtis = 87. Thus, the students scored significantly higher with Curtis than the other two texts and they scored higher with Aaron than Brown.

The GPA means were Low = 71.5, Medium = 81.6, High = 91.3. Thus, students with High GPA did significantly better than people with Low or Medium GPAs. Students with Medium GPAs did significantly better than people with Low GPAs. (GPA is a nonmanipulated characteristic of the participants.)