Exam 2 PS 306, Spring 2003

1. Yael is always quick to spot confounds in which a repeated measures design is not properly counterbalanced. Using a specific example that you create, illustrate *why* it is a confound to conduct a repeated measures experiment in which you use the same order for all participants. Be as explicit as you can. Then tell me what impact counterbalancing will have on the variability of scores within a condition if there are order or carryover effects. [10 pts]

Imagine a drug study in which you have two conditions: Placebo and Drug. If you did not counterbalance, but ran all participants through the study with Placebo first and Drug second, you would not know if any differences were due to the Drug or to some possible order effects. The DV is number of problems solved in an hour. Your data might look like this:

	Placebo	Drug
P_1	4	6
P_2	8	10
P_3	2	4
P_4	7	9
P_5	8	10
P_6	4	6

Drug produces performance that is 2 problems better than Placebo. Or does it? Without counterbalancing, it's possible that there is no effect of Drug at all, and that all that you've observed is a practice effect.

Now suppose that you do have a practice effect of +2. And let's assume that the Drug has no effect at all. What impact would counterbalancing have on your F-ratio? Well, let's assume that the first three participants get the order P->D and the last three participants get the order D->P. Your data would look like this:

	Placebo	Drug
P_1	4	6
P_2	8	10
P_3	2	4
P_4	9	7
P_5	10	8
P_6	6	4

It may not be immediately obvious to you, but the variability within both Placebo and Drug will now be higher. As a result, your MS_{Error} will be larger…and your F-ratio will be smaller.

2. For the following designs, tell me exactly how many participants would be needed. [15 pts]

a. In a 3x4 completely between (independent groups) design with n = 20. **240**

b. In a 3x7 completely within (repeated measures) design with a minimum n of 25. **42**

c. In a 3x6 completely within (repeated measures) design with a minimum n of 30. **36**

d. In a 4x4 mixed design, with the first factor between (independent groups) and the second factor within (repeated measures). Assume a minimum n of 25. **192**
e. In a 5x7 mixed design, with the first factor between (independent groups) and the second factor within (repeated measures). Assume a minimum n of 20. **140**

3. Suppose that you are interested in evaluating the effectiveness of different teaching techniques (e.g., Method A, Method B). What are two possible ways to avoid confounding teacher and technique? Which way seems most effective to you and why? [5 pts]

a. You could have the same teacher use both methods. This approach may be problematic because the teacher may have a preference for one of the two methods.
b. You could use multiple teachers who are trained in either of the two methods. That is, suppose that you have 10 teachers who are trained in Method A and 10 teachers who are trained in Method B. With that many teachers in each Method, it seems unlikely that either group would be comprised of people who are better teachers irrespective of method. So, I'd use this approach.

4. Dr. Alucard has a total of 60 bats at his disposal, 30 male and 30 female. He is interested in testing the effectiveness of a drug (Hearmore) that claims to permanently improve auditory abilities. To test the effectiveness of the drug, Dr. Alucard has decided to use a maze that must be traversed by the bats in total darkness, using only their sonar abilities. After training the bats to fly from one end of the maze to the other to obtain food, Dr. Alucard wants to use four different diameter wires (.25, .5, 1.0, and 1.5 mm) to construct the mazes through which the bats must fly. Each maze is constructed by stringing the wires of a particular diameter from floor to ceiling. (Think of the kitten video at the beginning of the semester.) Dr. Alucard intends for each bat to fly through each of the mazes as a means of assessing auditory discrimination abilities.

Dr. Alucard decides that the best dependent variable to use is the number of times that a bat touches one of the wires in the maze. Thus, more touches indicate worse auditory discrimination. As Dr. Alucard's research assistant, you are expected to complete the design of this study, based on the rough ideas that Dr. Alucard has established. Do so now, providing *explicit detail* about exactly how you would administer this experiment (nature of the design, procedure, etc.). [20 pts]

First of all, I'd want to use a completely repeated measures design here, but the effects or Hearmore are thought to be permanent. Therefore, I'd avoid using the drug as a level of a repeated factor. I would be comfortable using the type of maze (wire diameter) as a repeated factor, so I'd propose a mixed design with two levels of Drug (Placebo vs. Hearmore) and four levels of wire diameter (.25, .5, 1.0, and 1.5 mm), for a 2x4 mixed design.

With 4 levels of the repeated factor, I'd use complete counterbalancing, which would produce 24 orders (4!): .25 -> .5 -> 1.0 -> 1.5; .25 -> .5 -> 1.5 -> 1.0; .25 -> 1.0 -> .5 -> 1.5; etc. Thus, I'd need a total of 48 bats (24 with Placebo and 24 with Hearmore). To ensure that sex has no impact, I'd have the bats refrain from sex for…no, what I mean is that I'd have half of the bats in each condition male (12) and half female (12). Thus, 12 bats (6 male and 6 female) would get the day off.

So, I'd select a bat (say a male bat) and administer the Placebo treatment. I'd wait a half hour for the "drug" to take effect and then place him in Order 1 (.25, .5, 1.0, 1.5). That means that he'd first go through the maze with .25 mm diameter wires. Then, he'd go through the maze with .5 mm diameter wires. Then, he'd go through the maze with 1.0 mm diameter wires. Then, he'd go through the maze with 1.5 mm diameter wires. As the bat went through each maze, I would tally the number of times the bat touched one of the wires (measured electronically through pressure sensors, because it's totally dark in there). I'm presuming that the food rewards are so small that the bat would willingly keep working for food for all four mazes.

Then, I'd select another bat (say a female bat) and administer the Hearmore treatment. Then I'd wait for a half hour and place her in Order 24 and collect data from all four mazes. Then I'd run a female bat through the Hearmore treatment with Order 2. Then I'd run a male bat through the Placebo treatment with Order 23. And so forth… [Note that I'd be sure that an equal number of male and female bats were placed in the Placebo and Hearmore conditions.] When all the data were collected, I'd analyze the data with a 2x4 mixed model ANOVA.

5. As you may recall, Schacter has studied the number of speech fillers ("um," "ah," etc.) found in the speech of faculty from different areas under different circumstances. Suppose that he had looked at faculty from 4 different departments (English, History, Chemistry, and Biology). Suppose, also, that he had looked at the number of fillers found in 3 different social situations (Classroom Lecture, Conference Lecture (to peers), and Party situation), sampled over a 30-minute period in each situation. He might have found data that would lead to a source table such as that seen below. Complete the source table and interpret the results from this study as completely as you can. [10 pts]

ANOVA Table for Speech Fillers

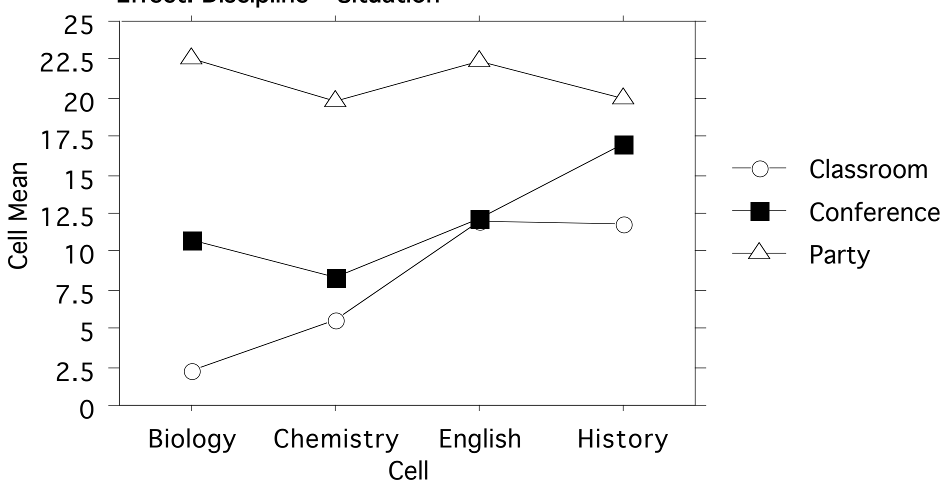
	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Situation	2	1848.933	924.467	78.289	<.0001	156.579	1.000
Discipline	3	288.400	96.133	8.141	.0002	24.423	.992
Situation * Discipline	6	291.600	48.600	4.116	.0021	24.694	.964
Residual	48	566.800	11.808				

Means Table for Speech Fillers
Effect: Situation * Discipline

	Count	Mean	Std. Dev.	Std. Err.
Classroom, Biology	5	2.200	1.304	.583
Classroom, Chemistry	5	5.600	2.074	.927
Classroom, English	5	12.000	5.148	2.302
Classroom, History	5	11.800	6.380	2.853
Conference, Biology	5	10.800	2.168	.970
Conference, Chemistry	5	8.400	1.140	.510
Conference, English	5	12.200	4.147	1.855
Conference, History	5	17.000	3.464	1.549
Party, Biology	5	22.600	2.074	.927
Party, Chemistry	5	19.800	4.207	1.881
Party, English	5	22.400	2.702	1.208
Party, History	5	20.000	2.000	.894

Given the significant interaction ($F(6,48) = 4.116$, $MSE = 11.81$, $p = .002$), that's where I'd focus my attention.

Interaction Line Plot for Speech Fillers
Effect: Discipline * Situation



$HSD = 4.85\sqrt{\frac{11.81}{5}} = 7.45$**, so two means would need to differ by 7.45 in order to be considered significant. So, we could conclude that in the Party situation, the 4 disciplines used an equal number of speech fillers. However, in the Classroom situation, both English and History faculty used significantly more fillers than Biology faculty (with no other**

differences significant). Moreover, in the Conference situation, History faculty produce more fillers than Chemistry faculty (with no other differences significant).

6. Dr. Alphonse Dente studies taste perception. In a recent study, he was interested in studying the impact of amount of salt added to a tomato sauce on ratings of the quality of the gustatory experience. He used three levels of salt (1 tablespoon per quart, 2 tablespoons per quart, and 3 tablespoons per quart). Other than the level of salt, the composition of the tomato sauce was identical. An equal amount of one sauce with one of the three salt levels was poured over spaghetti and served to each participant. Because Dr. Dente thought that the accompanying beverage might have an impact on the ratings of the food quality, one third of the participants for each level of salt consumed a beer along with their spaghetti, one third of the participants consumed a glass of wine, and one third of the participants consumed a glass of water. The dependent variable was a rating by each participant of the overall quality of the spaghetti using a 9-pt rating scale (1 = not so good and 9 = great). Complete the source table below and interpret the results of this study as completely as you can. [10 pts]

ANOVA Table for Rating

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Salt	2	89.911	44.956	69.759	<.0001	139.517	1.000
Beverage	2	69.378	34.689	53.828	<.0001	107.655	1.000
Salt * Beverage	4	1.956	.489	.759	.5591	3.034	.216
Residual	36	23.200	.644				

Means Table for Rating
Effect: Salt * Beverage

	Count	Mean	Std. Dev.	Std. Err.
1, Beer	5	8.000	.707	.316
1, Water	5	6.000	.707	.316
1, Wine	5	8.200	.837	.374
2, Beer	5	7.400	.548	.245
2, Water	5	4.200	.837	.374
2, Wine	5	7.200	.837	.374
3, Beer	5	5.000	1.000	.447
3, Water	5	2.200	.837	.374
3, Wine	5	4.800	.837	.374

Given the non-significant interaction, you should focus your attention on the two main effects. Because each factor has 3 levels, we need to compute Tukey's HSD:

$$HSD = 3.45\sqrt{\frac{.64}{15}} = .71$$

Next, you'd need to compute the means for the three levels of each factor:

Salt			**Beverage**		
1	**2**	**3**	**Beer**	**Wine**	**Water**
7.4	**6.3**	**4**	**6.8**	**6.7**	**4.1**

Then, you would be able to conclude that the quality of the spaghetti was rated as significantly higher when people tasted sauce with 1 tablespoon of salt compared to 2 or 3 tablespoons of salt. Moreover, 2 tablespoons of salt in the tomato sauce was rated as significantly better than 3 tablespoons of salt. Note, however, the absence of a crucial control group with no salt in the tomato sauce! For the beverages accompanying the meal, beer and wine produced no difference in the rating of the spaghetti, however both led to a significantly higher rating for the spaghetti compared to water as the accompanying beverage.