Exam 2 PS 306, Spring 2005

1. As we've discussed there are several different types of two-factor designs. In general, it makes sense to include as many repeated factors as possible. First, tell me why one would want to run a repeated measures design whenever possible. Second, tell me the kinds of independent variables that do not lend themselves to use as repeated factors. [5 pts]

1. RM design is more efficient and more powerful (smaller MS_{Error} is typical).
2. Can't use RM design if: factor produces permanent change, deception is involved, or if you are interested in a non-manipulated characteristic of a participant

2. For the following designs, tell me the number of participants needed and the total number of pieces of data that one would have in the study. [12 pts]

Design	Total Participants	Total Pieces of Data
A 5x7 completely independent groups (between) design, so that you achieve a minimum of 30 scores per cell.	**1050**	**1050**
A 5x7 completely repeated measures (within) design, so that you achieve a minimum of 30 scores per cell.	**70**	**2450**
A 5x7 mixed design, with the first factor repeated and the second factor independent groups, so that you achieve a minimum of 30 scores per cell.	**comp 840** **inc 210**	**comp 4200** **inc 1050**
A 5x7 mixed design, with the first factor independent groups and the second factor repeated measures, so that you achieve a minimum of 30 scores per cell.	**210**	**1470**
A 5x8 mixed design, with the first factor independent groups and the second factor repeated measures, so that you achieve a minimum of 30 scores per cell.	**160**	**1280**
A 4x6 mixed design, with the first factor repeated measures and the second factor independent groups, so that you achieve a minimum of 30 scores per cell.	**288**	**1152**

3. In our experiments, we're always interested in power. First, define power. Second, tell me what term is the complement of power. Third, tell me about the relationship between effect size and power (i.e., with low effect size...). [3 pts]

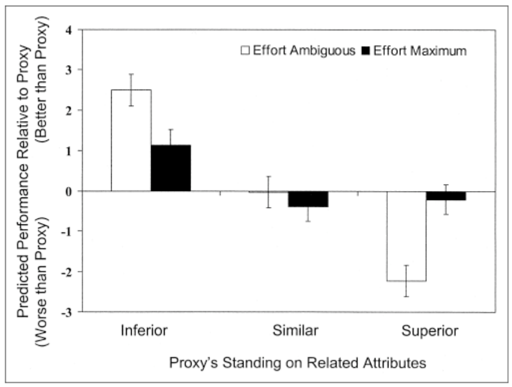
Power $(1-\beta)$ is the probability of correcting rejecting H_0.

Type II Error $(1-\beta)$ is the complement of power.

With smaller effect sizes, you would need more power to detect the effect.

4. People often compare themselves to other people, which has led social psychologists to study the phenomenon. In the proxy model, one estimates one's own likelihood of success by comparing oneself to a proxy (proxy = substitute). For example, suppose that you were trying to decide whether or not to go to graduate school. If you knew someone (a proxy) who was very much like you (in perceived intelligence, motivation, grades, etc.) and that person had failed in her attempt at graduate school, that knowledge might lead you to question your own likelihood of success in graduate school. The type of tasks studied varies quite a bit, but one area studied by Martin, Suls, and Wheeler (2002) involves grip strength (though they study other kinds of tasks as well). Participants begin by squeezing an exercise hand grip as many times as they could in a 30-sec period (Task1). Their performance was recorded and related to the participants. Then participants were shown a hand dynamometer, which measures hand grip force in kg/force exerted. They were asked to predict how much pressure they could exert on the dynamometer

(Task 2). To aid them in their predictions, participants were shown (fictitious) results from a participant "from last semester." Participants were randomly assigned to one of three proxy conditions: Inferior (the proxy hadn't squeezed as many times on Task 1 as the participant), Similar (the proxy had squeezed about as often on Task 1 as the participant), or Superior (the proxy had squeezed more often on Task 1 than the participant). In addition, based on written notes, the effort invoked by the proxy is either Ambiguous (participants were told that the experimenter did not know whether the proxy had exerted maximal effort or not) or Maximal (participants were told that the proxy had exerted maximal effort). The study was conducted as a 3x2 independent groups design. The dependent variable is a prediction difference score (PDS). Thus, if the participant expected to perform better than the proxy, the PDS would be positive. Solely from the figure below, predict what the authors found in their analyses of these data (treat any difference as significant) and interpret the results as best you can (as you would in a Discussion section). [15 pts]



Main Effect for Effort: Yes/No Why? **May be a main effect (Ambig = 0 and Max = .13)**

Main Effect for Proxy: Yes/No Why? **Likely a main effect (Inf = 1.75, Sim = -.2, Sup = -1.35)**

Interaction: Yes/No Why? **Likely an interaction (effect of effort not the same at all levels of proxy's standing)**

Interpretation

When the proxy doesn't do as well on the first task, participants think they will do better on the second task (which makes sense), but especially if the proxy's effort was ambiguous. Presumably, knowing that the proxy had exerted maximum effort is a signal that the participant will only be able to do a bit better on the second task, even if the participant exerts maximum effort also. However, if the participant anticipates exerting maximum effort, then he or she could reasonably expect to perform even better than a proxy who may well have exerted less-than-maximum effort.

When the proxy was similar on Task 1, participants thought they might do a bit worse than proxies who had exerted maximum effort on Task 1. In essence, given the similarity of the proxy, participants think that they will perform in a fashion roughly similar to the proxy. However, knowing that the proxy had exerted maximum effort leads participants to anticipate that they won't do quite as well on the second task.

If the proxy's performance was superior, participants thought they'd do a lot worse if the proxy exerted ambiguous effort, but less worse if the proxy had exerted maximum effort. Here the logic seems to reverse a bit. That is, the proxy did a lot better on the first task, but

if the proxy had to exert maximum effort to do so, then the participants probably think that they could "catch up" to the proxy by exerting a lot of effort. However, if the superior performance of the proxy may not have been due to maximum effort, then it's possible that even if the participants try really hard, they won't do as well.

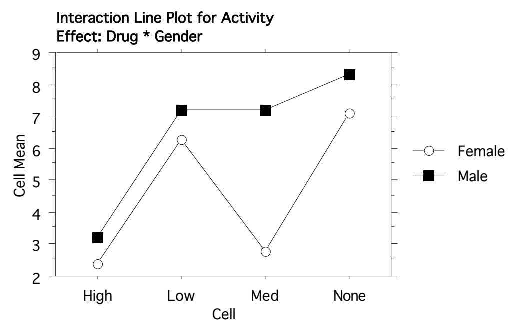
5. Dr. Mo Shun was interested in the impact of various dosages of a new drug (*Stay Put*) on the activity level of hyperactive children. She is fairly sure that, because of its chemical nature, *Stay Put* will be more effective for males than for females. To that end, she administers four dosage levels (None, Low, Medium, High) of *Stay Put* to an equal number of male and female children who exhibit similar levels of hyperactivity. The dependent variable is an activity measure, with higher numbers indicating greater activity. Analyze and interpret these data as completely as you can. [20 pts] {Johnson}

ANOVA Table for Activity

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Gender	1	66.613	66.613	21.045	<.0001	21.045	.998
Drug	3	278.538	92.846	29.333	<.0001	87.998	1.000
Gender * Drug	3	44.637	14.879	4.701	.0047	14.102	.890
Residual	72	227.900	3.165				

Means Table for Activity
Effect: Gender * Drug

	Count	Mean	Std. Dev.	Std. Err.
Female, High	10	2.400	1.265	.400
Female, Low	10	6.300	1.494	.473
Female, Med	10	2.800	.632	.200
Female, None	10	7.100	2.846	.900
Male, High	10	3.200	1.874	.593
Male, Low	10	7.200	1.814	.573
Male, Med	10	7.200	.789	.249
Male, None	10	8.300	2.359	.746



There is a significant interaction between Gender and Drug, *F*(3,72) = 4.7, *MSE* = 3.165, *p* < .005. There is also a main effect for Gender, *F*(1,72) = 21.045, *MSE* = 3.165, *p* < .001. There is also a main effect for Drug, *F*(3,72) = 29.333, *MSE* = 3.165, *p* < .001. As always, given the significant interaction, that's where I'd focus my attention. First, I'd produce a graph such as the one above. It appears that there is little difference between males and females for most drug levels. However, at a Medium dosage, it appears that Males score higher than Females.

To reinforce what my eyes are telling me, I would next compute Tukey's HSD:

$$HSD = 4.4\sqrt{\frac{3.2}{10}} = 2.49$$

With the HSD, I could now say that Males and Females don't differ for None, Low, or High levels of Drug. However, at Medium levels of Drug, Males scored significantly higher than Females. (I could also say that for Males, None, Low and Medium don't differ, but are greater than High. However, for Females, None and Low are equivalent and greater than Medium and High (which don't differ).

6. In Lab 2, we used difference scores as the DV so that the analysis would be a single-factor ANOVA. However, now that you're familiar with two-way ANOVA, we can think of the results from that lab in terms of actual response times in minutes (N.B. not milliseconds) as the only DV, so we'll ignore difficulty ratings. Let's suppose that we'd run that study as a completely between (independent groups) experiment. The two independent variables would be the same: TASK (actually solve anagrams vs. with solution present, estimate time for someone else to solve) and LENGTH (anagrams would be 4, 6, or 8 letters long). Note, however, that as a completely between design, a person would be exposed to only one of the six conditions. Complete the analysis below, interpret the results as completely as you can, and describe why these results may have arisen (as you would in a Discussion section). [25 pts]

ANOVA Table for Time

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Length	2	24.735	12.368	300.934	<.0001	601.868	1.000
Task	1	4.293	4.293	104.469	<.0001	104.469	1.000
Length * Task	2	.044	.022	.532	.5903	1.065	.130
Residual	54	2.219	.041				

Means Table for Time
Effect: Length

	Count	Mean	Std. Dev.	Std. Err.
4	20	.893	.315	.070
6	20	1.655	.369	.083
8	20	2.465	.331	.074

Means Table for Time
Effect: Task

	Count	Mean	Std. Dev.	Std. Err.
Estimate	30	1.403	.671	.122
Solve	30	1.938	.694	.127

There is no significant interaction, *F*(2,54) = .532, *MSE* = .041, *p* = .59. However, there is a main effect for Task, *F*(1,54) = 104.469, *MSE* = .041, *p* < .001. People take more minutes to solve the anagrams (*M* = 1.938) than they estimate it will take to solve the anagrams (*M* = 1.403). There is also a main effect of Length, *F*(2,54) = 300.9, *MSE* = .041, *p* < .001. To determine the source of this main effect, I would compute HSD:

$$HSD = 3.4\sqrt{\frac{.04}{20}} = .15$$

Thus, anagrams of length 8 took significantly longer (*M* = 2.465) than anagrams of 6 letters (*M* = 1.655) and both took longer than anagrams of 4 letters (*M* = .893).

Suppose that you had analyzed these same data as a one-way ANOVA on TASK. What would your resulting F-ratio have been?

Source	**df**	**SS**	**MS**	**F**
Task	1	4.29	4.29	9.5
Error	58	26.28	.45	
Total	59	30.57		