Exam 1 PS 306, Fall 2003

1. Briefly define the following terms and explain why each term is important to experimental design. [10 pts]

floor effect

A floor effect occurs when the task (dv) is too difficult, so that scores are less variable and clustered at the low end of the response continuum. As a result, actual differences in the effect of treatment may be obscured.

Type I error

A Type I Error occurs when you reject H_0 and you should have retained H_0. Of course, you'd never know that you'd committed a Type I Error. You know that the probability of making a Type I Error is α, which is typically set at .05. It's important to recognize that whenever you reject H_0, there is a chance that you've done so as a result of a Type I Error.

random assignment to conditions

In an independent groups design, you would run a participant in each condition before running the second (or third…) participant in that condition. This procedure ensures that any effects of time (experimenter or equipment fatigue, major historical events) would be distributed fairly evenly over the conditions of the experiment.

counterbalancing

In a repeated measures experiment, you would run participants in different orders. Depending on the number of levels of your factor, you might use complete or incomplete counterbalancing. With orders ≤ 5 you should use complete counterbalancing, which (with k orders) would require k! orders. With orders ≥ 5 you should use incomplete counterbalancing, which would require k orders if k is even and 2k orders if k is odd.

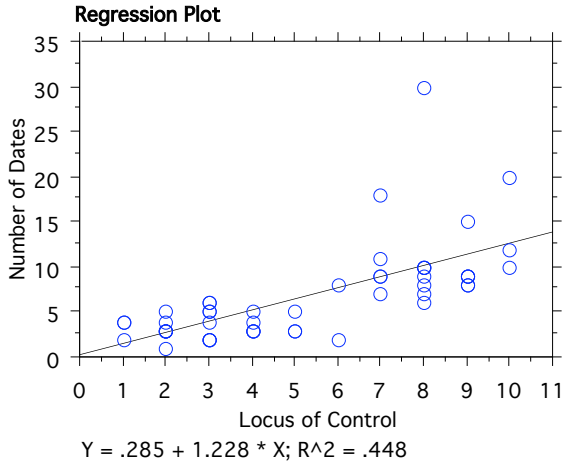
reliability

One has a reliable instrument when one achieves the same measure on multiple applications of the instrument to the same object. So, for instance, if one is measuring an individual's IQ, a reliable measure of IQ would be one that returned the same score every time the test was administered. Reliability is important for experimental design because an unreliable dependent variable, for instance, will lead to an inflated error term.

2. Dr. Sally Forth is interested in studying the relationship between Locus of Control (a measure developed by Dr. Julian Rotter) and the number of different people that a person has dated. She hypothesized that there would be a positive linear relationship between locus of control and the variety of a person's dating partners (higher locus of control leading to greater number of different people dated). Dr. Forth collected data from 50 college students on her scale of Locus of Control (0 = *Low* and 10 = *High*). Interpret her results (seen below) as completely as you can.
If a person had a Locus of Control score of 7, what would be your best estimate of the number of different people that person would have dated? Be very explicit in telling me why you would not be willing to accept the conclusion that one's Locus of Control affected the number of different people one would have dated. [10 pts]

Regression Summary
Number of Dates vs. Locus of Control

Count	50
Num. Missing	0
R	.669
R Squared	.448
Adjusted R Squared	.436
RMS Residual	3.957



ANOVA Table
Number of Dates vs. Locus of Control

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	609.318	609.318	38.918	<.0001
Residual	48	751.502	15.656		
Total	49	1360.820			

Regression Coefficients
Number of Dates vs. Locus of Control

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	.285	1.205	.285	.237	.8138
Locus of Control	1.228	.197	.669	6.238	<.0001

First of all, there is a positive linear relationship between Locus of Control and Number of Dates, such that people with higher Locus of Control tend to have had more dates than people with low Locus of Control, r(48) = .669, p < .001. The coefficient of determination (r^2) is .448, indicating that the two variables share a fair amount of variability. If a person had a Locus of Control score of 7, one would predict that the number of dates would be 8.88 (~9). Given the correlational design, one would not want to make a causal claim. It could very well be that one's Locus of Control is affected by the number of dates one has had (Causal Arrow problem). It may also be that some third variable, such as self esteem, may well produce the level of Locus of Control and the number of dates.

3. Dr. Nomar Gassé was interested in the impact of varying levels of depression on a person's ability to work effectively on a task, especially when tired. He selected people who were not clinically depressed, but who received high scores on the Beck Depression Inventory, as well as people who were diagnosed as clinically depressed and separated them into three groups (Low, Moderate, and Severe Depression). He then kept all participants awake for 48 hours. At the end of the 48-hour period, each participant was given a set of 10 problems to solve. The DV was the number of problems solved correctly in a 30-minute period. Complete the source table below and interpret the results of this study as completely as you can. [15 pts]

ANOVA Table for Prob Solved

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Depression	2	221.911	110.956	131.891	<.0001	263.781	1.000
Residual	42	35.333	.841				

Means Table for Prob Solved
Effect: Depression

	Count	Mean	Std. Dev.	Std. Err.
Low	15	6.600	.986	.254
Moderate	15	3.333	.976	.252
Severe	15	1.200	.775	.200

There is a significant effect of depression on number of problems solved, F(2,42) = 131.89, MSE = .841, p < .001. Note, however, that level of depression is a nonmanipulated characteristic of the participant, so one would not be able to make causal claims from this study. In order to determine which groups differ, one would need to compute a post hoc test.

$$HSD = 3.44\sqrt{\frac{.841}{15}} = .81$$

Thus, we could conclude that people with low levels of depression solved significantly more problems than people with moderate or severe levels of depression. People with moderate levels of depression solved significantly more problems than people with severe depression.

4. In Mook's article, he makes an argument about the value of experimental research that is not externally valid. For **three** of the following studies, indicate *why* the research is not externally valid and then tell me *why they are still useful* studies. [10 pts]

Need to provide solid answers to these questions using Mook notes.

Argyle study (glasses and intelligence)

Hecht study (dark adaptation)

Higgins & Marlatt study (anxiety and alcohol consumption)

Brown & Hanlon (parental role in children's acquisition of grammatical speech)

5. In order to study the power of reverse speech (back masking), Dr. Bob Reder had participants listen to music into which backward messages had been explicitly placed. Participants were asked to refrain from drinking any liquids for two hours prior to participating in the study. In the laboratory, each participant listened to four songs. In one song, the backward message was "Coke is heavenly." In the second song, the backward message was "Drink more Coke." In the third song, the backward message was "Don't drink Coke." And in the fourth song, the backward message was "Coke will kill you." As they listened to each song, participants had a large container of Coke in front of them and were told that they could drink as much Coke as they wanted. The DV was the number of ounces of Coke consumed during each of the four songs.

First, tell me very explicitly how many participants Dr. Reder should run in his study and how they should be exposed to the songs.

There would be 24 orders: CIH>DMC>DDC>CWK, CIH>DMC>CWK>DDC, CIH>DDC>DMC>CWK, CIH>DDC<CWK>DMK, etc. Thus, Dr. Reder would need to run in multiples of 24 participants.

Given the number of participants you propose, complete the following source table and tell me what Dr. Reder could conclude from his study. (You don't need to know F_{Crit}…right?) [10 pts]

Source	df	SS	MS	F
Subject	23	40.0	1.74	
Treatment	3	3.0	1.0	1.0
Error	69	69.0	1.0	

Even though the means for each condition are not provided to you here, what can you tell me about the means of the four conditions?

The means would be very similar to one another, given the low value of F.

6. Dr. Rick Call is interested in whether the type of mood reflected by a word (happy, neutral, or sad) affects how well the word is remembered. He constructs a list of 18 words: 6 happy words (e.g., joyful, bright), 6 neutral words (e.g., derive, convey), and 6 sad words (e.g., gloomy, lonely). He then presents the list repeatedly to 30 participants until they can recite the entire list correctly twice in a row. One week later, each participant attempts to recall the entire list. The number of items correctly recalled as a function of the type of word is analyzed. Complete the source table below and interpret the results of this study as completely as you can. [15 pts]

ANOVA Table for Type of Affect

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	29	27.656	.954				
Category for Type of Affect	2	166.756	83.378	80.718	<.0001	161.436	1.000
Category for Type of Affect * Subject	58	59.911	1.033				

Means Table for Type of Affect
Effect: Category for Type of Affect

	Count	Mean	Std. Dev.	Std. Err.
Happy	30	4.700	.952	.174
Neutral	30	1.633	1.066	.195
Sad	30	4.300	.988	.180

There was a significant effect of type of word (happy, neutral, sad) on recall, F(2,58) = 80.718, MSE = 1.033, p < .001. To determine which type of word is better remembered, one would need to compute a post hoc test.

$$HSD = 3.40\sqrt{\frac{1.033}{30}} = .63$$

Given these data, it appears that people recall happy and sad words better than neutral words, but the number of happy and sad words recalled is not different.