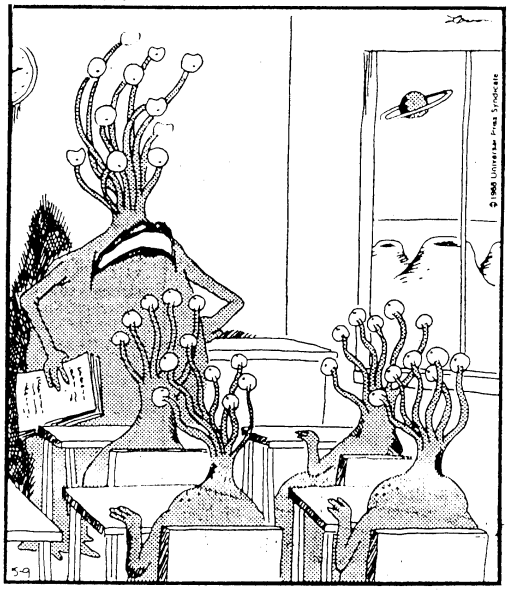
ID# _____ Exam 1 PS 306, Spring 2002



"And one final warning before we begin the exam — any stray eyeballs will be immediately thumped."

You should recognize the above cartoon as a fanciful way of reminding you that the Skidmore Honor Code is in effect. Work your way through the exam quickly and carefully, answering each question as completely as you can. Show as much of your work as possible, so that you can be sure of getting as much credit as possible. Don't hesitate to comment on particular designs, etc., even if it's not called for explicitly in the question (some questions are implicit). I think of a point as a minute, so you should expect to spend about 10 minutes on a 10-point question (for example). If you spend more time on a question than it is worth, you may not be able to complete the exam. There is a class following ours, so I will collect all the exams promptly at the end of the allotted time. Good Luck! Have a wonderful Spring Break!

1. Briefly provide an example of a nonmanipulated characteristic of a participant. Then, provide a clear explanation why you could not make a causal claim about such a variable if it were used in a study. (Be very explicit!) [5 pts]

A nonmanipulated characteristic of a participant is some aspect of the person that he or she brings to the study (age, personality, IQ, race, etc.). So, imagine a study with two groups (Low IQ and High IQ). Suppose that you use a DV of reaction time and you find a significant difference between the two groups. Would you be comfortable concluding that IQ *caused* the differences in RT? I would nope not! People with High IQ may be more confident, which makes them more prone to respond rapidly (so it's actually confidence that is producing the observed difference). Or people with Low IQ may be more cautious generally, which leads them to respond more slowly.

2. Hypothesis testing is essential to the research enterprise in psychology. Briefly define Type I Error, Type II Error, and power. Then, tell me why power is so important (or alternatively, why a Type II Error is so bad). Finally, tell me as many ways as you can to increase the power of a study. [5 pts]

Type I Error (α) = probability of incorrectly rejecting H_0
Type II Error (β) = probability of incorrectly retaining H_0
Power ($1 - \beta$) = probability of correctly rejecting H_0

Power is important (or one wants to avoid Type II Errors) because it's wasteful of one's time to look for differences (effects of a treatment) with little likelihood of actually finding such differences. You should do everything you can to ensure that your study has sufficient power to detect the sort of treatment effects that are typically investigated by psychologists. Such strategies include:
"large" sample size
"large" treatment effects
"small" error terms (by reducing individual differences or random error)
using a repeated measures design

3. Briefly explain why it is essential to develop an error term (MS) for the repeated measures design that includes only variability due to random factors? How is it computed? [5 pts]

The $MS_{Treatment}$ in a repeated measures design has no individual differences in it (because the same people contribute to each level of the factor). Thus, you need an error term that has no individual differences in it. As a result, your F-ratio will tend toward 1.0 when no treatment effect is present. You can think of computing MS_{Error} for a repeated measures design by removing the individual differences from the MS_{Error} you would have in an independent groups design. The error term in a repeated measures design indexes the interaction between participants and treatment. Thus, if each individual responds similarly to the levels of treatment, your error term will be small.

4. External validity is important in some circumstances, but Mook tries to argue that it's not particularly important in a lot of psychological research. First of all, define external validity. Then, using at least two of the studies cited in the article, explain circumstances in which Mook argues that external validity is not important (and why he thinks that way). Finally, using a single study that Mook discusses, indicate why a manipulation check would have been appropriate (or why a manipulation check would *not* have been needed). [10 pts]

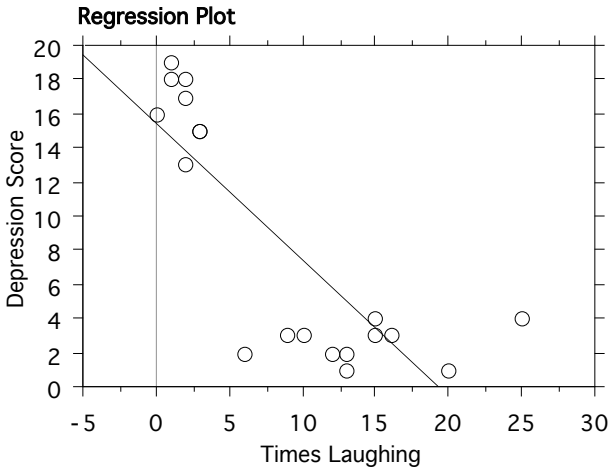
Clarity and detail would lead to a good response to the first part of this question.

A manipulation check would be useful when the IV is not directly observable. (Also, you probably need human participants to conduct a manipulation check, so the Harlow study wouldn't work.) Thus, you might consider the Higgins & Marlatt study, where participants were made more or less anxious by a manipulation of threat of shock. But were the people actually more or less anxious as a result of the manipulation? No real way to know unless you do a manipulation check. At the end of the study, you might ask each participant a series of questions, in which you embed a question asking the participant to rate how anxious he or she had been during the earlier phase of the experiment. You would hope that the high anxious group would give you a higher anxiety rating than the low anxious group.

5. Dr. Buster Gutt believes that laughter is a good antidote to depression. To test his hypothesis, he randomly samples 20 people and asks them to wear a counter device that they press every time they laugh during a given randomly selected day. At the end of the day of the study, he gives participants a device that tests their level of depression. Rather than use the Beck Depression Inventory, he chooses to use the less well known Degree of Unmanageable Depression Evaluation. Scores on the DUDE run from 0 (no depression) to 20 (very depressed). Dr. Gutt has analyzed his data as seen below.

Regression Summary
Depression Score vs. Times Laughing

Count	19
Num. Missing	1
R	.819
R Squared	.670
Adjusted R Squared	.651
RMS Residual	4.234



Y = 15.412 - .797 * X; R^2 = .67

ANOVA Table
Depression Score vs. Times Laughing

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	619.666	619.666	34.567	<.0001
Residual	17	304.755	17.927		
Total	18	924.421			

Regression Coefficients
Depression Score vs. Times Laughing

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	15.412	1.542	15.412	9.993	<.0001
Times Laughing	-.797	.135	-.819	-5.879	<.0001

Based on these analyses, interpret his results as completely as you can. (Be explicit!) If a person laughed 10 times a day, what would you predict that person's DUDE score to be? If a person laughed 20 times a day, what would you predict that person's DUDE score to be? [Careful…think!] Based on these results, Dr. Gutt begins advising his depressed patients to laugh at least 5 times each day. What might you want to say to the good doctor? [15 pts]

There is a significant negative linear relationship, ($F(1,17) = 34.567$, $p < .0001$). Thus, as the number of times laughing increases, the depression score decreases. If a person laughed 10 times a day, the regression equation would predict a depression score of 7.44. If a person laughed 20 times a day, the regression equation would predict a negative value (-.52) and the lowest score is 0, so your prediction should be 0. Because you cannot make causal claims from these correlational data, laughing may not be a means for lowering depression.

6. Dr. Jane Picasso is a clinical psychologist interested in autism. She is convinced that autistic children will respond better to treatment if they are in familiar settings. To that end, she randomly assigns 50 autistic children to treatment under 5 different conditions (Home, Familiar Office, Familiar Clinic, New Unfamiliar Clinic, and New Unfamiliar Office for every treatment). The children undergo treatment twice a week for one month under these conditions and are then tested on their progress using a 10-point rating scale (1=little progress and 10 = lots of progress). Below are incomplete analyses from her study. Complete the analyses and interpret the data as completely as you can. [15 pts]

ANOVA Table for Progress

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Setting	4	125.720	31.430	15.963	<.0001	63.853	1.000
Residual	45	88.600	1.969				

Means Table for Progress
Effect: Setting

	Count	Mean	Std. Dev.	Std. Err.
Fam Clinic	10	3.300	1.160	.367
Fam Office	10	4.100	1.370	.433
Home	10	6.200	2.394	.757
New Office	10	1.400	.516	.163
Unfam Clinic	10	2.800	.789	.249

H_0: $\mu_{FC} = \mu_{FO} = \mu_{H} = \mu_{NO} = \mu_{UC}$
H_1: Not H_0

With $F(4,45) = 15.963$, $MSE = 1.969$, $p < .0001$, I would reject H_0 and conclude that there were differences in progress due to the different locations. To determine which specific locations produced differences in progress, I would need to compute a post hoc test. With $q = 4$, I would get $HSD = 1.77$. Thus, I would conclude that Home leads to significantly more progress than all the other settings. The Familiar Office leads to significantly more progress than the New Office. The Familiar Clinic also leads to significantly more progress than the New Office. Thus, in general, familiar settings seem to lead to better performance, which supports Dr. Picasso's hypothesis.

7. Dr. Ty Pest is a human factors psychologist who is interested in testing the impact of different keyboard layouts on typing speed. To that end, he chooses four different conditions: a normal keyboard with normal (QWERTY) keyboard layout, a normal keyboard with the Dvorak layout of keys (supposedly a more efficient layout), an ergonomic (split) keyboard with normal keyboard layout, and an ergonomic (split) keyboard with the Dvorak layout of keys. He chooses as his participants secretaries who have at least 5 years of typing experience. The DV is the number of words per minute of typing of complex textual material in 15 minutes.

Dr. Pest should use counterbalancing in his experiment, of course. What kind would he use and how many different orders would that generate?

Using complete counterbalancing, Ty would need 24 orders.

Below are the incomplete analyses of his experiment. Complete the analyses and interpret the results as completely as you can. [15 pts]

ANOVA Table for Keyboard

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	3415.625	379.514				
Category for Keyboard	3	11436.875	3812.292	45.608	<.0001	136.824	1.000
Category for Keyboard * Subject	27	2256.875	83.588				

Means Table for Keyboard
Effect: Category for Keyboard

	Count	Mean	Std. Dev.	Std. Err.
Normal-QWERTY	10	74.000	11.738	3.712
Normal-Dvorak	10	43.000	12.737	4.028
Ergo-QWERTY	10	70.500	13.006	4.113
Ergo-Dvorak	10	35.000	12.693	4.014

Based on your earlier response, you should note immediately that the study could not have been appropriately counterbalanced with only 10 participants!

H_0: $\mu_{NQ} = \mu_{ND} = \mu_{EQ} = \mu_{ED}$
H_1: Not H_0

With F(3,27) = 45.608, MSE = 83.588, p < .0001, you would reject H_0 and conclude that the keyboard formats produced a difference in typing speed. To determine which specific conditions differed, you would need to compute Tukey's post hoc test. With q = 3.87, HSD = 11.19. Thus, N-Q leads to significantly faster typing speeds than N-D and E-D. The E-Q keyboard also leads to significantly faster typing speeds than N-D and E-D. No other differences were significant. It appears that these experienced typists are fastest when using a familiar keyboard layout. The Dvorak keyboard has a more efficient layout, so one can only presume that with sufficient practice, a typist would become even faster using a Dvorak keyboard.