Final Exam PS 306, Spring 2003

1. You know about the Milgram study, but what about the Hofling, et al. (1966) study of obedience? [from Gross] [5 pts]

Identical boxes of capsules were placed in 22 wards of both public and private psychiatric hospitals. The capsules were, in fact, placebos (consisting of glucose). But the containers were labeled '5 mg capsules of Astrofen' (not a real drug); the labels also indicated that the normal dose is 5 mg with a maximum daily dose of 10 mg.

While the nurse was on duty, a 'doctor' ('Dr. Smith from the Psychiatric Department') instructed the nurse, by telephone, to give 20 mg of Astrofen to his patient, a Mr. Jones, as he was in a desperate hurry and the patient needed the capsules. He said that he would come in to see Mr. Jones in 10 minutes time and would sign the authorization document for the drug when he got there.

To comply with his request, the nurse would be breaking three basic procedural rules:

(i) the dose was above the maximum daily dose of 10 mg;

(ii) drugs should only be given after written authority has been obtained;

(iii) the nurse must be absolutely sure that 'Dr. Smith' is a genuine doctor.

A real doctor was posted nearby, unseen by the nurse, and observed what the nurse did following the telephone call—comply, refuse, or try to contact another doctor. Whatever the nurse's course of action, the observer-doctor then revealed to the nurse what was really going on.

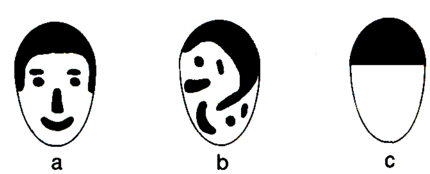
If you were a member of an IRB, would you have approved this study? Why or why not? [And although you would not know the results of the study as a member of an IRB, 21/22 nurses administered the 'drug.']

As always, for this type of question, you need to focus on the APA guidelines in your response. I'm less concerned with your actual decision than I am with your use of the guidelines in arriving at the decision.

2. How many participants would you need with a minimum n = 30? [8 pts]

A completely repeated measures (within) 3x5 design	**30**
A completely repeated measures (within) 2x9 design	**36**
A mixed 3x3 design, with the first factor repeated measures (within)	**90**
A mixed 2x9 design, with the second factor repeated measures (within)	**72**

3. Faces appear to be interesting stimuli to children (e.g., Fantz, 1961). To test that hypothesis, suppose that three different stimuli were presented to children of four different ages (1, 2, 3, and 4 months of age). The three different ovoids (seen below) were filled with face-like features (Face), filled with the same features in a scrambled fashion (Scrambled Face), or filled with an equivalent amount of black ink at the top of the ovoid (No Face). First of all, tell me why these particular stimuli were chosen. [2 pts]



The amount of black and white information was kept constant over the three "faces." However, the configuration of the black and white information is varied. In *a*, the configuration is face-like. In *b*, the configuration is as complex as the information in *a*, but it is not in a face-like configuration (controls for complexity). In *c*, the black and white information is not complex nor is it face-like.

The DV is the amount of time (in seconds) that the children spend looking at the stimuli in a 2-min test. Complete the source table below and interpret the results of this study as completely as you can. [15 pts]

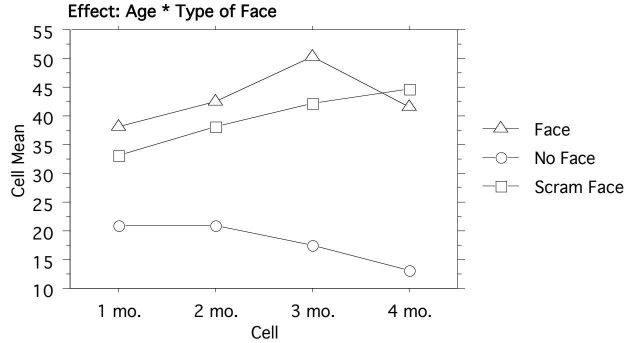
ANOVA Table for Looking Time

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Age	3	286.067	95.356	47.284	<.0001	141.851	1.000
Type of Face	2	7349.033	3674.517	1822.074	<.0001	3644.149	1.000
Age * Type of Face	6	724.833	120.806	59.904	<.0001	359.421	1.000
Residual	48	96.800	2.017				

Means Table for Looking Time
Effect: Age * Type of Face

	Count	Mean	Std. Dev.	Std. Err.
1 mo., Face	5	38.000	1.581	.707
1 mo., No Face	5	20.800	1.924	.860
1 mo., Scram Face	5	33.000	1.581	.707
2 mo., Face	5	42.400	2.074	.927
2 mo., No Face	5	20.800	1.483	.663
2 mo., Scram Face	5	38.200	1.304	.583
3 mo., Face	5	50.400	1.673	.748
3 mo., No Face	5	17.600	.894	.400
3 mo., Scram Face	5	42.200	.837	.374
4 mo., Face	5	41.600	1.140	.510
4 mo., No Face	5	13.000	1.000	.447
4 mo., Scram Face	5	44.800	.837	.374

Interaction Line Plot for Looking Time
Effect: Age * Type of Face



There is a significant interaction between Age and Type of Face, $F(6,48) = 59.9$, $MSE = 2.02$, $p < .001$. [There are also main effects for Age and Type of Face.] Because the interaction is significant, I would focus my attention on explaining the interaction. After looking at the graph above, I would have a sense of the differences that would lead to an interaction, so I would next compute Tukey's HSD:

$$HSD = 4.87\sqrt{\frac{2.02}{5}} = 3.1$$

Thus, at 1, 2, and 3 months of age, children preferred the facial stimuli (looked longer at them) than the scrambled faces, and they looked at longer at both facial and scrambled face thimuli than the non-face stimuli. However, at 4 months of age, children preferred (looked longer at) the scrambled faces than the facial stimuli, both of which were preferred to the non-face stimuli.

Alternatively, you might conclude that three-month-old children preferred the facial stimuli more than all other ages. Moreover, two- and four-moth-old children preferred the facial stimuli more than the one-month-old children (but did not differ from one another). However, three- and four-month-old children preferred the scrambled face stimuli more than the other two ages (but did not differ from one another). Moreover, the two-month-old children preferred the scrambled faces more than one-month-old children. Finally, the pattern was different from the non-face stimuli. For these non-face stimuli, one- and two-month-old children preferred the stimuli more than three- and four-month-old children (though they did not differ from one another). Moreover, three-month-old children preferred the non-face stimuli to the four-month-old children.

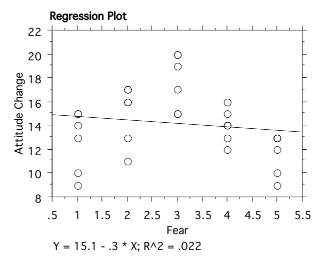
4. Rosenthal conducted a number of studies that were intended to illustrate the operation of experimenter expectancy effects. Other researchers have pointed out that various characteristics of experimenters are likely to affect the experimental outcome. [15 pts]
a. Describe the Rosenthal study that showed the problems that might emerge from the experimenter learning about how the participants were performing on a memory experiment *or* the Rosenthal study that showed the impact of telling the experimenter what to expect in terms of the ratings of pictures of people's faces.
b. What antidote(s) would you suggest to deal with such expectancy effects?
c. Which sort of effect (experimenter expectancy or experimenter characteristic) strikes you as the most problematic, and why.
d. How would you interpret the results of researchers like T. X. Barber, who do not find experimenter expectancy effects?

The answer to this question requires that you are able to carefully describe the results of the above studies and apply the ideas discussed in class relevant to this research.

5. Some researchers, such as McGuire (1968), have studied the relationship between the amount of fear invoked in a persuasive message and the extent of attitude change. Suppose that you observed a set of results such as those seen below. Interpret the results as completely as you can.
If a person had a Fear Score of 3, what would be your best estimate of that person's Attitude Change score? [5 pts]

Regression Summary
Attitude Change vs. Fear

Count	30
Num. Missing	0
R	.149
R Squared	.022
Adjusted R Squared	•
RMS Residual	2.924



ANOVA Table
Attitude Change vs. Fear

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	5.400	5.400	.632	.4335
Residual	28	239.400	8.550		
Total	29	244.800			

Regression Coefficients
Attitude Change vs. Fear

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	15.100	1.252	15.100	12.061	<.0001
Fear	-.300	.377	-.149	-.795	.4335

There is no significant linear relationship between attitude change and fear, *r*(28) = .149, *p* - .43. As such, you would not be able to use the regression equation to make a prediction of the Attitude Change Score of a person with a Fear score or 3. On the other hand, in looking at the scattergram, you should note that there is a fairly strong non-linear relationship between the two variables. Thus, you could compute two separate analyses looking at Fear from 1 to 3 and then from3 to 5. There appears to be a positive linear relationship between Fear and Attitude Change when Fear ranges from 1 to 3 (in which case, you might have an Attitude Change score of about 18). There also appears to be a negative linear relationship for Fear scores from 3 to 5.

6. In an attempt to determine the extent to which fear is an important tool in persuasive messages, Janis and Feshbach (1953) assigned high school students to one of four groups. The message was concerned with dental hygiene and degree of fear arousal was manipulated by the number and nature of consequences of improper care of teeth which were referred to (and shown in color slides); each message also contained factual messages about the causes of tooth decay and some advice about caring for teeth.

The *high fear* condition made 71 references to unpleasant effects, including toothache, painful treatment, and possible secondary diseases, including blindness and cancer; the *moderate fear* condition made 49 references and the *low fear* condition just 18. (Control participants heard a talk about the eye.)

After one week, the effectiveness of the persuasive communications was examined. Suppose that the DV was the extent to which the participants adopted better dental care behaviors (1 = adopted few, 10 = adopted many). Complete the source table below and interpret the results of this study. [10 pts]

ANOVA Table for Behav Adopted

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Fear Condition	3	508.938	169.646	191.719	<.0001	575.156	1.000
Residual	76	67.250	.885				

Means Table for Behav Adopted
Effect: Fear Condition

	Count	Mean	Std. Dev.	Std. Err.
Control	20	2.150	.813	.182
High	20	2.800	1.056	.236
Low	20	5.850	.933	.209
Moderate	20	8.450	.945	.211

To complete this ANOVA, you need to remember that the MS_{Error} for an independent groups ANOVA is the average of the variances of the groups. Because you have the standard deviations of the four groups, you first need to square the standard deviations to arrive at the variances for the groups (.66, 1.12, .87, and .89 for Control, High, Low, and Moderate, respectively). Thus, the average variance (and the MS_{Error}) would be .885. With that information, you would be able to complete the source table. There is a significant effect of Fear, $F(3,76) = 191.72$, $MSE = .885$, $p < .001$.

Given the significant effect, the next step would be to compute a post hoc test:

$$HSD = 3.72\sqrt{\frac{.885}{20}} = .78$$

Thus, you could conclude that Moderate fear leads to the adoption of more dental care behaviors than all the other fear conditions. Low fear leads to more dental care behaviors than High or Control, which do not differ.

7a. You are asked to determine if there is an effect of three different noise levels (80, 90, 100 decibels) in the working environments of people on an assembly line. (Assume that the typical noise level in the factory is 80 dB, so 90 dB is louder than normal and 100 dB is much louder than normal.) You decide to use the number of errors made in a 3 hour time period as your dependent variable. You also decide to use a repeated measures design, but you need to explain to the executives at this company why you've chosen to do so. Lay out the pros and cons of such a design now. Then describe *exactly* how you'd conduct the study (i.e. provide me with sufficient detail that I can determine that you would know how to conduct such a study). [15 pts]

This study will be a bit tricky to conduct as a repeated measures design, but you would want to do so to take advantage of the power and efficiency that come from conducting repeated measures designs. The disadvantage of conducting a repeated measures design is that the carryover effects may be substantial…to the point that the error variance left after counterbalancing may be quite large. Moreover, the demand characteristics (remember the Hawthorne Effect) may be substantial as well, which might argue for an independent groups design.

However, if you're going to go ahead with a repeated measures design, you've got to think about how you'd go about counterbalancing. You cannot simply ignore the carryover effects, so you'd need to use some form of counterbalancing. If you could have the workers come in to work on the assembly line in six different groups, then you would be able to use complete counterbalancing (80 -> 90 -> 100; 80 -> 100 -> 90; 90 -> 80 -> 100; 90 -> 100 -> 80; 100 -> 80 -> 90; 100 -> 90 -> 80). This approach is most likely impractical, because it would tie up the factory with a small staff for many days just to conduct the research.

If you could only divide the work force into three groups, you might use three different orders, even though doing so would be less than optimal (e.g., 80 -> 90 -> 100; 90 -> 100 -> 80; 100 -> 80 -> 90). But let's imagine that you could not divide the work force at all (because everyone is needed on the assembly line to do the work), you could have everyone present but change the noise levels for everyone using the complete counterbalancing order shown above. Thus, for one day, you might have one noise level (e.g., 80 dB) before lunch and a different noise level (e.g., 90 dB) after lunch. Then you'd keep on changing the noise levels over 12 days (2 different noise levels each day) and then measure the number of errors made during a 3-hour period (e.g., 9AM –12 PM and 1 PM – 4 PM). Then, you could average over the number of errors made under each of the three noise levels over the 12 workdays. I presume that assembly errors would be easy to determine (e.g., the object couldn't be assembled because of the error), but doing so may be fairly difficult in a real-life scenario.

7b. Suppose that the results of the study with three levels of the IV came out as seen below. Complete the source table and then interpret the results. [10 pts]

ANOVA Table for Noise Level

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	6.533	.726				
Category for Noise Level	2	24.800	12.400	22.622	<.0001	45.243	1.000
Category for Noise Level * Subject	18	9.867	.548				

Means Table for Noise Level
Effect: Category for Noise Level

	Count	Mean	Std. Dev.	Std. Err.
Normal.80dB	10	.600	.699	.221
Louder.90dB	10	1.400	.966	.306
Loudest.100dB	10	2.800	.632	.200

If you used the last approach to counterbalancing, you would be able to get away with n = 10 (even though that strikes me as a small number of people on an assembly line). However, if you took the first approach, you'd need to run in multiples of 6 people, which would likely result in a number far greater than 10 (and would have to be a multiple of 6).

There is a significant effect of noise level demonstrated, $F(2,18) = 22.62$, $MSE = .55$, $p < .001$. Thus, you'd need to conduct a post hoc test to determine which noise levels differ.

$$HSD = 3.61\sqrt{\frac{.5}{10}} = .81$$

Thus, the 100dB noise level leads to significantly more errors than the 80 dB and 90 dB noise levels, which don't differ from one another.