Exam 2, PS 306, Spring 2002

1. We'd briefly discussed the Dutton & Aron bridge study in class. In that study, elevated physiological arousal of the male participants was presumed to be the root cause of a greater attraction to the young woman waiting at the end of the high rope bridge as compared to the perceived attraction from the non-elevated physiological arousal that followed crossing a stable concrete bridge. But maybe the perceived attraction had to do with elements other than physiological arousal. Dutton & Aron also used only male participants. As an extension of their study, suppose that male and female participants are given one of three levels of epinephrine [None (saline), Small, Large] to produce varying levels of physiological arousal. After 5 minutes (to allow the drug to have its effect), participants were shown a picture of the face a person who had previously been judged by other people to be moderately attractive (4 on a 7-point scale of attractiveness). For the male participants, the picture was of a female face. For the female participants, the picture was of a male face. Each participant rated the attractiveness of the target face on a 7-point scale. Complete the analysis of this 2x3 independent groups design and interpret the results as completely as you can. [10 pts]

ANOVA Table for Attraction Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Gender	1	6.021	6.021	5.282	.0266	5.282	.608
Epinephrine	2	19.625	9.812	8.608	.0007	17.217	.967
Gender * Epinephrine	2	3.792	1.896	1.663	.2018	3.326	.319
Residual	42	47.875	1.140				

Means Table for Attraction Score
Effect: Gender * Epinephrine

	Count	Mean	Std. Dev.	Std. Err.
Female, Large	8	4.875	1.126	.398
Female, None	8	4.000	.926	.327
Female, Small	8	4.500	1.195	.423
Male, Large	8	6.250	.707	.250
Male, None	8	4.000	1.309	.463
Male, Small	8	5.250	1.035	.366

Because the interaction is not significant, I would focus my attention on the two main effects, which are significant. Gender requires no additional analysis, beyond the computation of the means for Males (5.2) and Females (4.5). Thus, I could conclude that males rated the pictures as significantly more attractive than did females. For Epinephrine, I would need to compute Tukey's HSD = .92 (q = 3.44). Thus, I would conclude that Large doses of epinephrine lead to higher ratings (*M* = 5.56) than Small doses (*M* = 4.88) or None (*M* = 4.00), neither of which differed.

2. Power…it's all about power! [15 pts]
First of all, clearly define power.

Power is the probability of correctly rejecting H_0.

Jacob Cohen and others have advocated working to achieve a level of power of roughly .80. What does such a level of power say about the level of Type II Error that these folks are willing to tolerate? What does *that* say about people's tolerance for Type II Errors compared to Type I Errors?

Because power and Type II errors are complimentary, power of .80 means that the probability of a Type II error would be .20. That then means that people are willing to tolerate 4x's as much Type II as Type I error.

Finally, tell me very explicitly how you would design a study to maximize its power (i.e., talk about the aspects of the design on which you would focus to achieve the greatest power).

Examples would help, but the basic notion is to increase n (or have the largest n possible), increase treatment effect (or have the largest treatment effect you could imagine as reasonable), and decrease the error (or design the study to have the least individual differences and random variability that you could achieve). Details about how you would accomplish those goals would be useful.

3. Repeated measures designs are more powerful than independent groups designs. They are also more efficient. First, determine the number of participants needed to achieve a minimum of 25 scores per cell in the following 3x5 designs. Then use that information to illustrate (briefly) the efficiency of a repeated measures design. [15 pts]

	B1	B2	B3	B4	B5
A1					
A2					
A3					

Completely Between How many total participants would you need? **375** How many total pieces of data would occur in the entire experiment? **375**
Completely Within How many total participants would you need? **30 (incomplete counterbalancing)** How many total pieces of data would occur in the entire experiment? **450**
Mixed (A independent groups and B repeated measures) How many total participants would you need? **90 (incomplete), 360 (complete)** How many total pieces of data would occur in the entire experiment? **450 OR 1800**
Mixed (A repeated measures and B independent groups) How many total participants would you need? **150 (complete)** How many total pieces of data would occur in the entire experiment? **450**
What does all the above information say about efficiency? **Note that for a completely between design, you would need 375 people and only end up with 375 pieces of data. However, if you could use a completely within design, you would need only 30 people and they would generate 450 pieces of data for you. The two mixed designs fall in between these two extremes.**

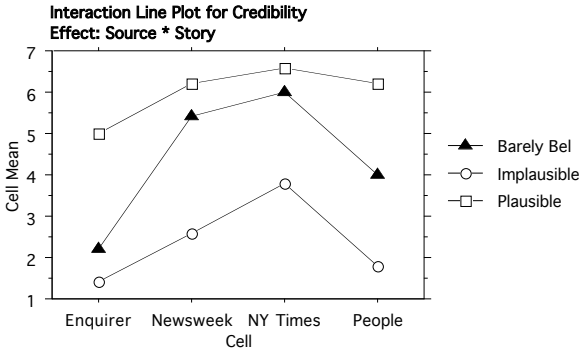
4. Dr. Reeder was interested in the extent to which people believe reports that they read in various media. To that end, he constructs three stories: Plausible, Barely Believable, and Implausible. For example, the Plausible story would describe the details of an actual event (such as an armed robbery). The Barely Believable story would describe a situation that might be true, but seems unlikely (such as the number of women with whom Wilt Chamberlain claimed to have had sexual relations). Finally, the Implausible story would describe a situation that seems extraordinarily unlikely (such as the sighting of a thin Elvis Presley). Each of the three stories was set up (on proper paper, with proper font, etc.) so that they appeared to have come from one of four sources: The New York Times, The National Enquirer, Newsweek, and People Magazine. Each participant read one of the 12 possible stories and rated its credibility on a scale from 1 (Absolutely convinced that the story was not true) to 7 (Absolutely convinced that the story was true). Thus, this is a 3x4 independent groups design. Complete the analysis below and tell Dr. Reeder what he should conclude from the analyses of these data. [20 pts]

ANOVA Table for Credibility

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Source	3	55.333	18.444	25.441	<.0001	76.322	1.000
Story	2	130.133	65.067	89.747	<.0001	179.494	1.000
Source * Story	6	11.467	1.911	2.636	.0273	15.816	.812
Residual	48	34.800	.725				

Means Table for Credibility
Effect: Source * Story

	Count	Mean	Std. Dev.	Std. Err.
Enquirer, Barely Bel	5	2.200	1.095	.490
Enquirer, Implausible	5	1.400	.548	.245
Enquirer, Plausible	5	5.000	.707	.316
Newsweek, Barely Bel	5	5.400	.894	.400
Newsweek, Implausible	5	2.600	1.140	.510
Newsweek, Plausible	5	6.200	.837	.374
NY Times, Barely Bel	5	6.000	.707	.316
NY Times, Implausible	5	3.800	.837	.374
NY Times, Plausible	5	6.600	.548	.245
People, Barely Bel	5	4.000	1.000	.447
People, Implausible	5	1.800	.837	.374
People, Plausible	5	6.200	.837	.374



Focusing on the interaction, I would compute HSD = 1.86 (q = 4.88). Using HSD in conjunction with the figure, I would interpret the interaction as follows: For the *Enquirer*, people rated the plausible stories as significantly higher than the barely plausible and the implausible stories (neither of which differed). However, for *People*, people rated the plausible stories significantly more credible than the barely plausible stories and the implausible stories, and the barely plausible stories were rated higher than the implausible stories. Different still were the results for *Newsweek* and the *NY Times*, where plausible stories and barely plausible stories were rated as equally credible and significantly more credible than the implausible stories.

5. Adapted from Huck & Sandler's *Rival Hypotheses*:

a. Current theories of hunger place the main responsibility on…the hypothalamus. The hypothalamus monitors the chemical content of the blood and triggers eating responses at appropriate times. It has even been shown that certain surgical lesions in the hypothalamus of rats can lead to nonstop eating — the rats literally eat themselves to death. Earlier theories of hunger, however, were not so sophisticated. They were based mainly on the assumption that the stomach had a causal influence on the brain — that is, since hunger pangs were often reported by those who were hungry, it seemed logical to look at the influence of stomach contractions on hunger.

One of the earliest studies in this area was done by two researchers (Cannon and Washburn, 1912) who had human participants swallow a small balloon that was then inflated. The air pressure in the "gastric balloon" was affected by stomach contractions that were transmitted by a recording device. The participants were also asked to indicate each time a hunger pang was felt. The researchers took the strong positive correlation between stomach contractions and hunger pangs to mean that the contractions caused the pangs. What would you say? [10 pts]

This is a correlational study, which means that you've got the usual problems associated with making causal claims from such designs. In this case, the stomach contractions might be causing the hunger pangs (the preferred interpretation), the hunger pangs might be causing the stomach contractions, or there might be a third variable producing the correlation. For instance, a change in the state of the hypothalamus may be producing both hunger pangs and stomach contractions.

b. One of our least favorite events is to have the phone ring at an inopportune moment, especially when we find that the caller has dialed a wrong number. Occasionally such callers infuriate us even more by hanging up with little indication of remorse on their part — undoubtedly auguring the ultimate decline of modern society. Nonetheless, we wonder what our reaction would have been as participants in the following study.

In an attempt to study the helping behavior of blacks and whites towards members of their own and the other race, two psychologists called people in Brooklyn and said that the caller's car had broken down on the highway and the caller ("George Williams") was attempting to reach his garage mechanic with his last quarter. The recipient of the call was then asked to call the mechanic and report the caller's location. Helping was defined as the participant's actually calling the mechanic's number (really the phone number of a member of the research team). The caller was either obviously white (Brooklyn accent) or obviously black (Southern black accent). Five hundred forty participants were categorized as black on the basis of their having common black last names, as well as on the basis of the neighborhood in which they lived. Five hundred sixty-nine whites were classified on the basis of their geographic location. Voice characteristics were used to confirm the other selection criteria, and fewer than 1 percent were eliminated on this basis.

Statistical analyses showed a significant effect of race of the participant and the race of the caller in helping, with blacks helping both blacks and whites equally, and whites helping whites more often than helping blacks. The general conclusion of the researchers was that "race of the victim (inferred from his dialect) has a small but detectable influence on helping behavior" (Gaertuer & Bickman, 1972).

Given that the researchers recognized socioeconomic differences between the two samples as possibly confounding their results, are there other plausible rival hypotheses to be considered? [10 pts]

I'd be willing to accept that the people called were correctly identified (though some people may take issue with that possibility). Nonetheless, there are a number of potential confounds, including:
1. One person with a Brooklyn accent and one person with a Southern black accent were calling. As we discussed in terms of therapists, it might be the case that one person (i.e., Brooklyn accent) was more persuasive (independent of race).
2. The results may have less to do with race than with familiarity. Doing the research in Brooklyn and using a Brooklyn accent for one of the people calling means that all of the people called would be more likely to be familiar with a Brooklyn accent.