Final Exam PS 217, Spring 2004

1. What is the relationship between power and effect size? That is, when you are considering a research design in which there is a large effect size, what are the implications for power? When you are considering a research design in which there is a small effect size, what are the implications for power? How would you typically add power to a study? [10 pts]

With large effect size, you would not need a lot of power to detect the effect. With small effect size, you would need a lot of power to detect the effect.

To increase power, you would typically increase sample size (n). In an experiment, you could also consider ways in which you could increase the treatment effect. You could also consider ways in which you could decrease error due to individual differences (possibly using a more homogeneous sample) or random variability (possibly by improving instructions, eliminating disturbances to participants, etc.).

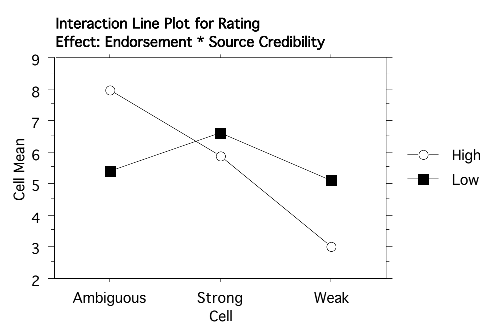
2. How seriously do people take product reviews? Chaiken and Maheswaran (1992) conducted an interesting experiment in which they varied the credibility of the review source and the general message of the review. They asked college students to read a review of a new telephone answering machine. The researchers told half the participants that the review came from a flyer printed by the discount store Kmart (low credibility) or from the magazine Consumer Reports (high credibility). Each participant then read one of three types of review, an unambiguous strong review, an ambiguous review (the answering machine was better than some machines but not as good as others), or an unambiguous weak review. The researchers then asked the participants to rate on a 10-point scale their willingness to buy the answering machine for $50 (10 = very willing to buy). The results of their study are replicated below. Complete the analysis and interpret the results of this study as completely as you can (as in a Discussion section). [10 pts] {Pittenger}

ANOVA Table for Rating

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Source Credibility	1	.067	.067	.056	.8143	.056	.056
Endorsement	2	80.433	40.217	33.618	<.0001	67.235	1.000
Source Credibility * Endorsement	2	58.233	29.117	24.339	<.0001	48.678	1.000
Residual	54	64.600	1.196				

Means Table for Rating
Effect: Source Credibility * Endorsement

	Count	Mean	Std. Dev.	Std. Err.
High, Ambiguous	10	8.000	1.155	.365
High, Strong	10	5.900	1.101	.348
High, Weak	10	3.000	1.155	.365
Low, Ambiguous	10	5.400	.699	.221
Low, Strong	10	6.600	1.174	.371
Low, Weak	10	5.100	1.197	.379



First, I would compute Hartley's F_{Max}, to determine if I should be concerned about heterogeneity of variance:

$F_{Max} = 1.432 / .489 = 2.93$, **so, with $F_{Max/Crit} = 7.8$, I would conclude that there is no reason to be concerned about heterogeneity of variance and I would use $\alpha = .05$.**

There is a significant main effect for Endorsement and a significant interaction between Source Credibility and Endorsement. To explain the interaction, I would look at the graph and see where the simple effects appear to differ, then I would compute HSD and determine if I could interpret the interaction.

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 4.18\sqrt{\frac{1.196}{10}} = 1.45$$

With a Weak argument, people are more persuaded (more willing to buy) when the argument is made by a low-credibility source than when made by a high-credibility source. However, when the argument is Ambiguous, people are more willing to buy when the argument comes from a high-credibility source than when it comes from a low-credibility source. When the argument is Strong, however, participants are persuaded equally by the weak and strong sources.

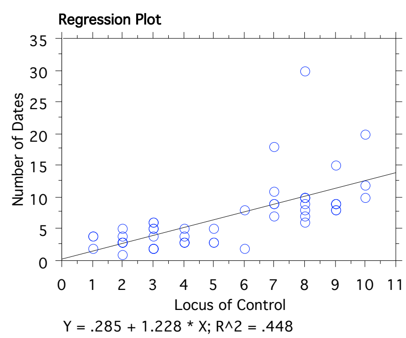
If you were interested in assessing the effect size, you would focus on the interaction:

$$\eta^2 = \frac{58.233}{58.233 + 64.6} = .47$$

3. 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? If a person had a Locus of Control score of 12, 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

There is a significant positive linear relationship between Locus of Control (Loc) score and Number of Dates, $r(48) = .669$, $p < .001$. If a person had a LoC score of 7, I would predict that the person would have had 8.88 (~9) dates. If a person had a LoC score of 12, I would have to say that I couldn't safely predict the number of dates, because I didn't observe anyone with that high a LoC score. If the trend continued, however, I would predict 15 dates. I wouldn't be willing to conceive of the relationship as a causal one because it could well be that as a person has more dates, her or his LoC scores changes (causal arrow problem). It could also be that a "third" variable may be creating the relationship. That is, maybe self-esteem affects both a person's LoC score and the number of dates that person has had.

4. 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. [10 pts]

ANOVA Table for Prob Solved

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Depression	2	221.58	110.79	131.89	<.0001	263.781	1.000
Residual	42	35.28	.84				

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

First, let's ensure that there is no concern that the homogeneity of variance assumption has been violated.

$F_{Max} = {}^{.972}\!/_{.601} = 1.62$**, so, with $F_{Max/Crit} = 3.6$, I wouldn't be concerned, so I would use $\alpha = .05$.**

There is a significant effect of depression (reject H_0). To determine which levels differed, I would need to compute HSD:

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 3.44\sqrt{\frac{.84}{15}} = .81$$

So, people with low levels of depression solve significantly more problems than those who are moderately or severely depressed. Those who are moderately depressed solve

significantly more problems than those who are severely depressed.

$$\eta^2 = \frac{221.58}{221.58 + 35.28} = .86$$

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.

With four levels, he should use complete counterbalancing, which means that he needs to run multiples of 24 participants. The orders in which they would be exposed to the messages would be H->DMC->DDC->K; H->DMC->K->DDC; H->DDC->DMC->K; H->DDC->K->DMC; etc.

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 fairly similar, producing an F-ratio that is very small and indicative of no effect.

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. [10 pts]

ANOVA Table for Type of Affect

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	29	27.55	.95				
Category for Type of Affect	2	161.4	80.7	80.7	<.0001	161.436	1.000
Category for Type of Affect * Subject	58	58.0	1.0				

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

The description is not completely clear, but let's presume that the words are presented in some random fashion, so that there is no reason to think that anything is contributing to differences except the nature of the words. (Note that you wouldn't necessarily need to counterbalance the items, as long as they were presented in a random fashion.)

There is a significant effect of type of word, so you would reject H_0. To determine which of the means differed, you'd need to compute HSD:

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 3.41\sqrt{\frac{1.0}{30}} = .62$$

Thus, it appears that people recall significantly more happy and sad words than neutral words. Happy and sad words are recalled about equally.

7. Dr. Michael T. Hedd suspects that LSD affects the speech center in the brain. Specifically, he believes that repeated use of LSD is related to a person's ability to retrieve verbal information. To see if he can obtain any evidence regarding his hunch, Dr. Hedd advertises for participants who have taken LSD at least once. Nine people of comparable IQ and education are selected from the applicants. All participants are given a 50-item test. Each item consists of a definition of a low-frequency English word; the participant's task is to produce the target word. Below are two sample items:

Definition	**Target Word**
to make things thinner or weaker by the addition of water	dilute
patronage bestowed in consideration of family relationship and not merit	nepotism

The dependent variable is the number of failures to produce the target words (errors). Here are the relevant data:

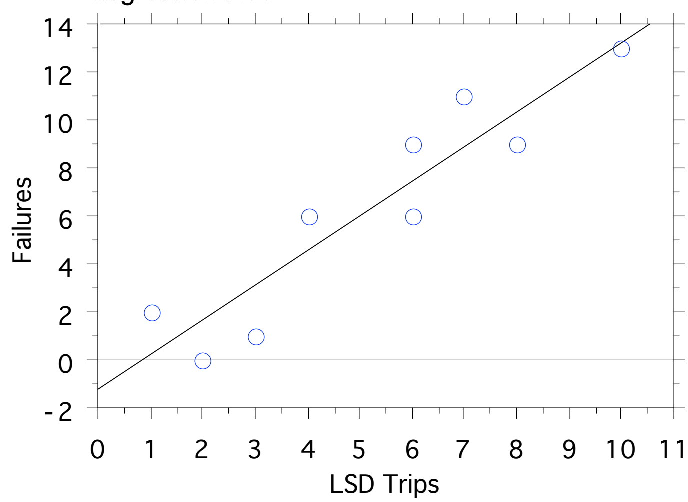
Participant	**Number of LSD trips**	**Number of failures**	**XY**
1	6	6	36
2	7	11	77
3	4	6	24
4	2	0	0
5	10	13	130
6	1	2	2
7	3	1	3
8	8	9	72
9	6	9	54
ΣX	47	57	398
ΣX^2	315	529	
s^2	8.69	21	
SS	69.52	168	

Analyze these data as completely as you can. [20 pts] {Johnson}

Regression Summary
Failures vs. LSD Trips

Count	9
Num. Missing	0
R	.928
R Squared	.861
Adjusted R Squared	.842
RMS Residual	1.823

Regression Plot



Y = -1.2 + 1.442 * X; R^2 = .861

ANOVA Table
Failures vs. LSD Trips

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	144.730	144.730	43.537	.0003
Residual	7	23.270	3.324		
Total	8	168.000			

Regression Coefficients
Failures vs. LSD Trips

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	-1.200	1.293	-1.200	-.928	.3845
LSD Trips	1.442	.219	.928	6.598	.0003

I produced the analysis in StatView, but note that you could get SS_X and SS_Y by multiplying the s^2 for each variable by df. Then, the only value you'd need to actually compute is the sum of the XY scores. So:

$$r = \frac{100.33}{\sqrt{(69.52)(168)}} = .928$$

Basically, there is a positive linear relationship between the number of LSD trips and memory failures, because $p < .001$ (or you could look up $r_{Crit}(7) = .666$).

8. 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}

	Males				Females				
	None	Low	Med	High	None	Low	Med	High	
	10	8	4	3	12	9	3	5	
	11	7	3	4	8	6	6	2	
	8	10	5	5	10	7	5	3	
	7	9	7	2	9	5	2	1	
	12	8	6	7	7	6	3	2	
	4	5	5	1	5	4	4	4	
	8	4	3	3	4	5	2	4	
	6	7	2	1	5	6	3	2	
	8	6	4	4	3	7	3	1	
	9	8	4	2	8	8	5	1	Sum
ΣX (T)	83	72	43	32	71	63	36	25	425
ΣX^2	739	548	205	134	577	417	146	81	2847
SS	50.1	29.6	20.1	31.6	72.9	20.1	16.4	18.5	259.3

ANOVA Table for Activity

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Dosage	3	313.738	104.579	29.039	<.0001	87.116	1.000
Gender	1	15.312	15.312	4.252	.0428	4.252	.518
Dosage * Gender	3	.837	.279	.078	.9720	.233	.063
Residual	72	259.300	3.601				

Means Table for Activity
Effect: Dosage

	Count	Mean	Std. Dev.	Std. Err.
High	20	2.850	1.663	.372
Low	20	6.750	1.682	.376
Med	20	3.950	1.432	.320
None	20	7.700	2.618	.585

Means Table for Activity
Effect: Gender

	Count	Mean	Std. Dev.	Std. Err.
Female	40	4.875	2.633	.416
Male	40	5.750	2.790	.441

Again, I used StatView for the analysis, but you should get the same values (except for rounding error). I would first determine whether or not I needed to be concerned about heterogeneity of variance by computing F_{Max}:

$F_{Max} = 72.9/16.4 = 4.45$**, so with $F_{Max/Crit} = 8.95$ I would have no concern about heterogeneity of variance, and I would use $\alpha = .05$.**

The interaction is not significant, so I would focus on the two main effects, both of which are significant. With only two levels of Gender, no post hoc test would be necessary, so I could say that Male activity scores were significantly higher than Female activity scores. For the Dosage factor, I would need to compute HSD:

$$HSD = q\sqrt{\frac{MS_{Error}}{n}} = 3.7\sqrt{\frac{3.6}{20}} = 1.57$$

So activity levels were highest in the Placebo (None) condition than Medium and High doses of the drug (but activity levels for None and Low were not different). Activity levels in the Low dosage condition were also higher than with Medium and High doses. Activity levels were higher for Medium dosage than High dosage.