ID# ___________________________ Final Exam PS217, Spring 1995

First of all, be sure to enter your correct student ID# above (not your telephone number, social security number, license plate, or whatever). I have the correct list of ID#s on the roster, so check to ensure that you're using the correct number. As always, read over the exam carefully and answer each question completely. Good Luck!! Have a relaxing summer...it's been a pleasure!!

1a. Suppose that you are a member of a psychology department who serves as an advisor to students. You have student evaluations for four of your colleagues (Professors 1, 2, 3, and 4). The student evaluations are measured on a 5-point scale. Each professor taught three courses the past semester (Courses 1, 2 and 3). Below is the StatView analysis of the data. Analyze the data as completely as you can. What would you recommend to your advisees about these four professors? Any comments on this study? [15 pts]

ANOVA Table for Evaluation

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Professor	3	65.633	21.878	34.142	<.0001	102.426	1.000
Course	2	53.733	26.867	41.927	<.0001	83.855	1.000
Professor * Course	6	43.467	7.244	11.305	<.0001	67.833	1.000
Residual	228	146.100	.641				

Means Table for Evaluation
Effect: Professor * Course

	Count	Mean	Std. Dev.	Std. Err.
Prof 1, Course 1	20	3.650	1.040	.233
Prof 1, Course 2	20	2.500	.827	.185
Prof 1, Course 3	20	3.950	.887	.198
Prof 2, Course 1	20	4.650	.489	.109
Prof 2, Course 2	20	2.800	.696	.156
Prof 2, Course 3	20	4.250	.716	.160
Prof 3, Course 1	20	4.100	.718	.161
Prof 3, Course 2	20	4.150	.745	.167
Prof 3, Course 3	20	4.300	.733	.164
Prof 4, Course 1	20	2.200	.834	.186
Prof 4, Course 2	20	2.350	.933	.209
Prof 4, Course 3	20	3.900	.852	.191

First of all, you might note that this is surely not an experiment. Thus, the design is lousy (at best). The faculty are not likely to have been teaching the same courses, so Course 1 for Prof 1 is likely different from Course 1 for Prof 2. As such, you're comparing apples and oranges. It's also really a mixed design, with Course a repeated factor. Thus, the analysis is actually inappropriate.

All that said, however, it's clear from the source table that there is an interaction. You would compute HSD = .83 (q = 4.62). Thus, for Prof 1 and Prof 2, Course 1 and Course 3 don't differ, but both receive better ratings than Course 2. However, for Prof 3, there were no differences in ratings for the three courses. Finally, a different pattern appeared for Prof 4, where Course 1 and Course 2 did not differ, but both were significantly lower than Course 3. All in all, it appears that one might gravitate to Prof 3 and avoid Prof 4.

1b. Suppose that the preceding analysis had been conducted as a single factor ANOVA on Professor (i.e., ignoring the classes as a factor and analyzing the same data as a one-way ANOVA). Complete a source table that would show how these results would come out, analyze the data completely, then tell me specifically what sorts of information are lost in computing a one-way ANOVA instead of the two-way ANOVA in 1a. [5 pts]

Source	SS	df	MS	F	
Professor	**65.63**	**3**	**21.88**	**21.2**	**$F_{Crit}(3,236) = 2.64$**
Error	**243.3**	**236**	**1.03**		
Total	**308.93**	**239**			

You would reject H, and conclude that the professors differed in their ratings. Next, you would compute HSD = .48 (q = 3.65). The means are:

Prof 1	Prof 2	Prof 3	Prof 4
3.37	3.9	4.18	2.82

Thus, Prof 2 and Prof 3 are statistically equivalent and both receive better ratings than Prof 1 and Prof 4. Prof 1 receives better ratings than Prof 4.

As you will note, your F is smaller for the one-way ANOVA (although still significant). That's because there is a lot of variability in the courses. When Course is included as a factor, the variability is compartmentalized within each course. However, for the one-way ANOVA the variability within each Prof is larger, leading to a larger MS_{Error}. Needless to say, computing the one-way ANOVA also loses the ability to assess the impact of Course and the interaction between Course and Prof.

2. A professor is interested in assessing the difficulty of her exams. She looks at the performance of one class of 20 students over the four exams given during the semester. Shown below is a partially completed output from her StatView analysis of the data. First, complete the source table, then analyze and interpret the data as completely as possible. What could you tell this professor about the difficulty of the exams. Would that be your preferred explanation of the results? **NB This output is from an early version of StatView. You can simply ignore the second line (Within subjects). The other lines are just like those found in the output from the current version of StatView.** [10 pts]

One Factor ANOVA-Repeated Measures for X_1 ... X_4

Source:	df:	Sum of Squares:	Mean Square:	F-test:	P value:
Between subjects	19	4597			
Within subjects	60	486			
treatments	3	288	96	27.4	
residual	57	197	3.5		
Total	79	5083			

Reliability Estimates for- All treatments: .967 Single Treatment: .878

One Factor ANOVA-Repeated Measures for X_1 ... X_4

Group:	Count:	Mean:	Std. Dev.:	Std. Error:
Exam1	20	79.85	9.615	2.15
Exam2	20	82.35	7.721	1.726
Exam3	20	83.95	7.265	1.625
Exam4	20	84.85	6.892	1.541

First of all, you should note that the study is not really an experiment, because of the lack of random assignment. Furthermore, there is no counterbalancing. Thus, the differences in performance across exams could be due to practice effects. That's a major confound if you're interested in making an assessment of exam difficulty.

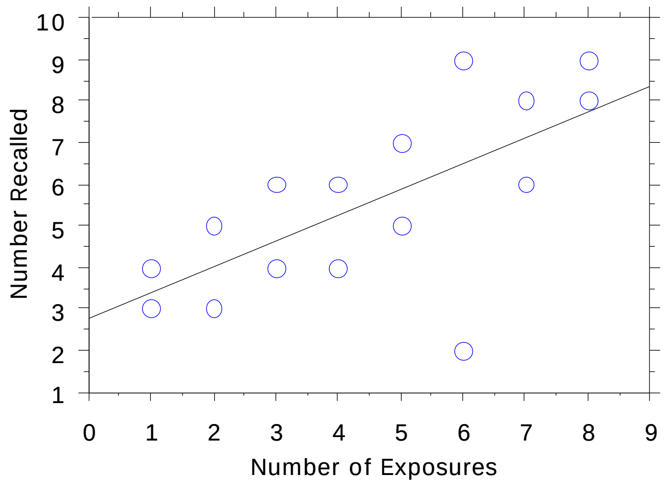
Nonetheless, you could conclude that you would reject H_0, because F_{Crit} (3,57) = 2.77. You could then compute HSD = 1.57 (q = 3.75). That would allow you to say that Exam 3 and Exam 4 are significantly higher than Exam 1 and Exam 2, but performance on Exam 4 is equal to Exam 3. Exam 2 is significantly higher than Exam 1. However, you wouldn't be able to say that it was the exam, per se, that is producing the observed differences.

3. A researcher is interested in the relationship between the number of times a word is seen in a list and a person's ability to recall the words. Sixteen people watch a computer screen on which appear 72 words. The words are presented one at a time in a random order, with two of the words appearing once, two twice, two three times, etc., up to the two words presented eight times each. Below is the StatView analysis of the data. Tell me as much as you can about the results from your interpretation of the analyses. What kinds of predictions would you feel comfortable making on the basis of this analysis. [10 pts]

Regression Summary
Number Recalled vs. Number of Exposures

Count	16
Num. Missing	0
R	.675
R Squared	.456
Adjusted R Squared	.417
RMS Residual	1.672

Regression Plot



Y = 2.75 + .625 * X; R^2 = .456

ANOVA Table
Number Recalled vs. Number of Exposures

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	32.812	32.812	11.741	.0041
Residual	14	39.125	2.795		
Total	15	71.937			

Regression Coefficients
Number Recalled vs. Number of Exposures

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	2.750	.921	2.750	2.986	.0098
Number of Exposures	.625	.182	.675	3.427	.0041

There is a positive linear relationship between exposures and recall ($p < .05$). Thus, for exposures between 1 and 8, it makes sense to use the regression equation for prediction. In fact, even though a correlational analysis is used, you should recognize that the study involves an experimental design. Thus, you would be comfortable making causal claims.

4a. A researcher is interested in whether or not males and females differ in spatial abilities. To that end, he gives a spatial abilities test (SPAT) to a group of 10 males and 10 females. SPAT scores range from 1-10, with higher SPAT scores indicating better spatial abilities. Analyze the data below to test the hypothesis and tell the researcher what he should conclude. [10 pts]

	Male	Female
	10	8
	9	7
	6	8
	7	8
	8	6
	10	5
	4	8
	6	7
	9	6
	8	6
$\Sigma X=$	77	69
$\Sigma X^2=$	627	487

Source	**SS**	**df**	**MS**	**F**	
Gender	**3.2**	**1**	**3.2**	**1.28**	**FCrit(1,18) = 4.41**
Error	**45**	**18**	**2.5**		
Total	**48.2**	**19**			

On the basis of these data, you would retain H_0. That is, you would have to conclude that there is insufficient evidence to claim that males and females differed. Noting that there is a difference, even though it is not significant, you might be inclined to pursue the hypothesis with a more powerful study. That is, you might want to run more participants, use a more sensitive test of spatial abilities, etc.

4b. A different researcher is interested in whether or not there is a relationship between gender and spatial abilities. Using the same data, compute the most appropriate statistic for this question. (Be careful. Think. You can't approach this analysis mindlessly.) [5 pts]

The trick here is to think of gender as one variable with two levels (0 = male, 1 = female). That would lead you to re-order your data for an analysis that is basically a point biserial correlation.

	Gender	Score	XY
	0	10	0
	0	9	0
	0	6	0
	0	7	0
	0	8	0
	0	10	0
	0	4	0
	0	6	0
	0	9	0
	0	8	0
	1	8	8
	1	7	7
	1	8	8
	1	8	8
	1	6	6
	1	5	5
	1	8	8
	1	7	7
	1	6	6
	1	6	6
Sum	10	146	69

$$\boldsymbol{r} = \frac{69 - \frac{(10)(146)}{20}}{\sqrt{\left(10 - \frac{10^2}{20}\right)\left(1114 - \frac{146^2}{20}\right)}} = \frac{-4}{\sqrt{(5)(48.2)}} = -.258$$

Because $r_{Crit}(18) = .444$, you would retain H_0. Thus, you would conclude that there is no evidence that there is a linear relationship between gender and spatial abilities.

4c. Test the null hypothesis that the males were sampled from a population of SPAT scores with a mean of 10. [5 pts]

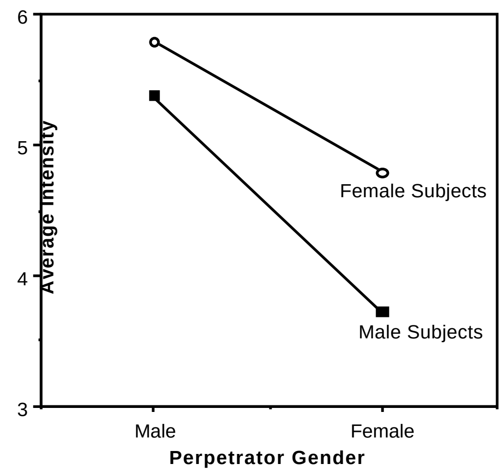
Mean	7.7
SS	34.1
s^2	3.79
s	1.95

$$\boldsymbol{t} = \frac{7.7 - 10}{1.95 / \sqrt{10}} = \frac{-2.3}{.616} = -3.73$$

With $t_{Crit}(9) = 2.262$, you would reject H_0 and conclude that it's likely that the sample of males was drawn from a population with $\mu < 10$.

5. In a study by Baron, Burgess, and Kao (1991), male and female participants read accounts of stories that included a description of a sexist act perpetrated by either a male or a female against a female. The 193 participants described the perpetrator in a way that could be scored for intensity of sexist behavior. The displayed sexist behavior was rated 1 for *slightly displayed* to 7 for *extremely displayed*. Part of their *Results* section reads:

> Perpetrator gender and subject gender main effects were both significant. Female subjects, compared with male subjects, gave more intense ratings to both male and female perpetrators...: $F(1,189) = 5.06$, $p < .03$....Furthermore, male perpetrators were seen as displaying more intense gender bias than female perpetrators: $F(1,189) = 15.97$, $p < .0001$. The interaction between subject gender and perpetrator gender was nonsignificant...$p < .34$...These results can be seen in the figure below:



Briefly describe the meaning of these results. [10 pts]

With no significant interaction, the focus should be on the two main effects. First of all, you could conclude that Females thought that more sexist behavior was displayed than Males (main effect for Gender). Next, you could conclude that when the perpetrator was Male, participants rated the behavior as more sexist than when the perpetrator was Female.

6. Blanchard, Lilly, and Vaughn (1991) conducted a study of influences on the expression of reaction to racism. In this study, white undergraduate women were approached while walking between classes and asked to participate in a survey. The survey asked what the college should do in response to anonymous racial notes. Half of the students were asked to respond privately (on paper) and half were asked to respond publicly (orally for the researcher and anyone else present to hear). Within each of these groups, a third of the subjects were in the "no influence" condition, in which they simply completed the survey with no one other than the experimenter present. For the other two-thirds of the subjects, before they could begin answering, another student was invited to participate. This other student was actually a confederate of the experimenter, and it was arranged so that she always gave her responses to the questions first. The confederate's opinions were either very antiracist or not at all antiracist, making the other two levels of this factor "antiracist influence" and "nonantiracist influence." Analyze the summary data below as completely as possible. [20 pts]

	Public Response	Private Response	Marginal
Antiracist Influence	$\Sigma X = 92$ SS = 14 n = 4	$\Sigma X = 84$ SS = 14 n = 4	$\Sigma X = 176$ **Mean = 22**
No Influence	$\Sigma X = 84$ SS = 6 n = 4	$\Sigma X = 84$ SS = 20 n = 4	$\Sigma X = 168$ **Mean = 21**
Nonantiracist Influence	$\Sigma X = 64$ SS = 18 n = 4	$\Sigma X = 72$ SS = 14 n = 4	$\Sigma X = 136$ **Mean = 17**
Marginal	$\Sigma X = 240$	$\Sigma X = 240$	$\Sigma X = 480$

Source	**SS**	**df**	**MS**	**F**	
Response	**0**	**1**	**0**	**0**	**$F_{Crit}(1,18) = 4.41$**
Conf Attitude	**112**	**2**	**56**	**11.72**	**$F_{Crit}(2,18) = 3.55$**
R x C	**16**	**2**	**8**	**1.67**	**$F_{Crit}(2,18) = 3.55$**
Error	**86**	**18**	**4.78**		
Total	**214**	**23**			

{Here's the "tough" part of this question. You will ultimately need to know ΣX^2. It's not provided. Thus, you'll need to remember the formula for SS and recognize that you can extract ΣX^2 from SS. For example, for the first cell (Antiracist, Public), knowing that SS = 14 and that $\Sigma X = 92$, you should be able to compute that $\Sigma X^2 = 2130$. Compute ΣX^2 for each of the cells and then sum to get that overall $\Sigma X^2 = 9814$.}

For this study, you would investigate the main effect for Confederate Influence, because none of the other effects were significant. To determine which of the means differed, you'd need to compute HSD = 2.79 (q = 3.61). Thus, there was no difference between Antiracist and No Influence, both of which were significantly higher than Nonantiracist Influence.

7. A researcher was interested in investigating the relationship between motivation (measured on a 9-point scale, with 1 = low and 9 = high) and GPA (0 = low and 4 = high). She obtained the data below from a sample of members of Phi Beta Kappa. Analyze the data as completely as possible. If a person got a score of 7 on the motivation scale, what would be your best prediction of GPA? Do you see any limitations to this study? [10 pts]

	Motivation	GPA	**XY**
	6	4	**24**
	8	4	**32**
	7	4	**28**
	6	3	**18**
	9	4	**36**
	8	4	**32**
	7	4	**28**
Sum	51	27	**198**
SS =	**7.43**	**.86**	

$$\boldsymbol{r} = \frac{198 - \frac{(51)(27)}{7}}{\sqrt{(7.43)(.86)}} = \frac{1.3}{2.53} = .51$$

Because $r_{Crit}(5) = .754$, you would retain H_0 and conclude that there does not appear to be a positive linear relationship between GPA and Motivation. However, you should note that the participants were all members of PBK, which means that they are smarter than the average bear. Thus, you are probably dealing with a truncation of range problem. It would be more appropriate to sample more widely than members of PBK, unless you had some hypotheses specific to that group. Collecting data from more participants would also make lots of sense.