ID# ____________________ Final Exam PS217, Fall 1996

OK, the trick is to work carefully yet quickly through this exam. Answer each question as completely as you can. Show ALL your work. You have exactly 3 hours to complete the exam (no extra time). Thus, you have to be careful to watch your time to avoid leaving questions unanswered. If you're having trouble deciding what analysis to conduct, you might check out the Statistics Organizer (A-57 to A-64). Good Luck! And have a relaxing break...I'll see you next semester!

1. Memory for a list of items can be improved if you make meaningful associations with each item. In one demonstration of this phenomenon, Craik and Tulving (1975) presented 6 people with a list of 60 words to be remembered. Each word was presented in the context of a sentence. For 30 of the words, the sentences were very simple, and for other 30 words, the sentences were more elaborate. (The order of the type of sentence was randomized.) For the word *rabbit*, for example, a simple sentence would be "She cooked the rabbit." An elaborate sentence would be "The great bird swooped down and carried off the struggling rabbit." The researchers recorded the number of words recalled for each type of sentence to determine whether the more elaborate sentences produced better memory. Hypothetical data, similar to experimental results, are seen below. Do these data indicate a significant difference between the two types of sentence? [25 pts]

	Simple Sentence	Elaborate Sentence	**P**
	14	22	**36**
	15	17	**32**
	19	24	**43**
	12	19	**31**
	17	28	**45**
	15	20	**35**
$\Sigma X=$	92	130	**222**
$\Sigma X^2=$	1440	2894	**4334**
SS=	**29.3**	**77.3**	

Source	**SS**	**df**	**MS**	**F**	
Sentence	**120.33**	**1**	**120.33**	**25.44**	$F_{Crit}(1,5) = 6.61$
Within	**106.67**	**10**			
Subject	**83**	**5**			
Error	**23.67**	**5**	**4.73**		
Total	**227**	**11**			

Because $F_{Obt} \geq F_{Crit}$, I would reject H_0. Note that with only two levels of the IV, no HSD would be necessary. Thus, I would conclude that people recall significantly more words in elaborate sentences than in simple sentences.

2. A researcher in interested in the possible relationship between two subtests of the WAIS (Wechsler Adult Intelligence Scale). A random sample of 8 army recruits is selected; they are given both the Vocabulary subtest and the Digit Span (a test of short-term memory) subtest. Their weighted, scaled scores are seen below. Is there a relationship between the two subtests? If a person receives a score of 11 on the Digit Span subtest, what is your best estimate of the subtest score that person will receive on Vocabulary? [25 pts]

	Digit Span	Vocabulary	**XY**
	9	11	**99**
	6	8	**48**
	12	13	**156**
	7	6	**42**
	10	10	**100**
	5	6	**30**
	9	11	**99**
	10	9	**90**
$\Sigma X=$	68	74	
$\Sigma X^2=$	616	728	
Mean =	**8.5**	**9.25**	

H_0: $\rho = 0$

H_1: $\rho \neq 0$

$$r = \frac{664 - \frac{(68)(74)}{8}}{\sqrt{(38)(43.5)}} = \frac{35}{40.66} = .86$$

With $r_{Crit}(6) = .707$, because $r_{Obt} \geq r_{Crit}$ you would reject H_0 and conclude that there is a significant positive linear relationship between Digit Span and Vocabulary.

Because of the significant linear relationship, you should compute the regression equation:

$Y = .92X + 1.43$

Thus, if a person scored 11 on the Digit Span subtest, you would predict that that person would get a Vocabulary score of 11.55.

3. Given the data from problem 2, test the H_0 that $\mu = 10$ for the Digit Span subtest. [15 pts]

H_0: $\mu = 10$

H_1: $\mu \neq 10$

ss = 38
$s^2 = 5.43$
s = 2.33
$s_X = .82$

t = -1.83 and $t_{Crit}(7) = 2.365$

Thus, you would retain H_0 and conclude that the sample could have been drawn from a population with $\mu = 10$.

4. Suppose that you conduct an experiment using a repeated measures design, but when it comes time to analyze the data you realize that you've forgotten how to compute a repeated measures ANOVA. All that you can remember is how to compute an independent groups ANOVA. You decide that you'll go ahead and compute the independent groups ANOVA on the data. You find that your results are significant. Explain (in sufficient detail) why you would expect that the treatment really did have an effect and that had you computed the appropriate ANOVA you would have been led to reject H_0 as well. [10 pts]

Unless the individual differences are minor, the repeated measures ANOVA will produce a larger F-ratio than an independent groups ANOVA. (The repeated measures ANOVA is more powerful than the independent groups ANOVA.) Thus, if your results are significant with an independent groups ANOVA, you would expect that the results would also be significant with a repeated measures ANOVA.

5. In a study of hyperactivity among elementary school boys, 63 students were randomly selected from a school population of ADHD, 7-year-old boys are randomly assigned to one of 9 groups (n=7). (ADHD is Attention Deficits with Hyperactivity, and left untreated, it can prevent a child from attending to incoming learning stimuli and may also create major disruptions in the classroom.) The researcher wanted to study the classroom effects of both the drug Ritalin as well as a behavior modification program on the activity levels of the students. The drug administered at three levels: 0 mg Ritalin (placebo), 10 mg Ritalin, and 20 mg Ritalin. The behavior modification program consisted of giving each student 10 tokens to start the day and then taking away a token for each hyperactive infraction. The tokens that were saved could then be exchanged for some valued prize. The behavior modification program was varied across three levels: no program, program every other day, and program every day. After four weeks, all the children were evaluated for hyperactivity and were assigned scaled scores ranging from a possible low of 0 (no indication of hyperactivity) to a high of 40 (extreme hyperactivity). A partially completed StatView output of the data is seen below. Complete the source table and then analyze the data as completely as possible, providing a complete interpretation. [20 pts]

ANOVA Table for Hyperactivity Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Behavior Mod	2	1283.143	641.571	26.292	<.0001	52.583	1.000
Ritalin	2	657.238	328.619	13.467	<.0001	26.934	.999
Behavior Mod * Ritalin	4	128.762	32.190	1.319	.2746	5.277	.376
Residual	54	1317.714	24.402				

Means Table for Hyperactivity Score
Effect: Behavior Mod * Ritalin

	Count	Mean	Std. Dev.	Std. Err.
Every Day, 10 mg	7	12.286	3.352	1.267
Every Day, 20 mg	7	8.714	3.592	1.358
Every Day, 5 mg	7	14.429	3.645	1.378
Every Other Day, 10 mg	7	19.000	5.477	2.070
Every Other Day, 20 mg	7	17.143	5.429	2.052
Every Other Day, 5 mg	7	22.429	7.413	2.802
None, 10 mg	7	22.000	3.958	1.496
None, 20 mg	7	16.429	4.276	1.616
None, 5 mg	7	29.143	5.815	2.198

Because the interaction is not significant, I would focus my attention on the two main effects (both of which are significant).

Because both main effects have identical df, the same HSD would be used for testing the two main effects. HSD = 3.67 (q = 3.42).

For Behavior Mod: Administering the Behavior Mod every day leads to less hyperactivity than administering Beh Mod every other day or not using Beh Mod (and those two levels produce statistically similar results).

For Ritalin: Dosage 20 leads to significantly less hyperactivity than Dosage 10 or Placebo. Dosage 10 leads to significantly less hyperactivity than Placebo.

6. According to Milton Rokeach, there is a positive correlation between dogmatism and anxiety. Dogmatism is defined as rigidity of attitude that produces a closed belief system (or a closed mind) and a general attitude of intolerance. In the following study, dogmatism was measured on the basis of the Rokeach D scale (Rokeach, 1960), and anxiety is measured on the 30-item Welch Anxiety Scale, an adaptation taken from the MMPI (Welch, 1952). A random sample of 30 undergraduate students from a large western university was selected and given both the D scale and the Welch Anxiety test. The data analyses are as seen below.

Regression Summary
Anxiety Test vs. D-Scale

Count	30
Num. Missing	0
R	.659
R Squared	.434
Adjusted R Squared	.414
RMS Residual	17.492

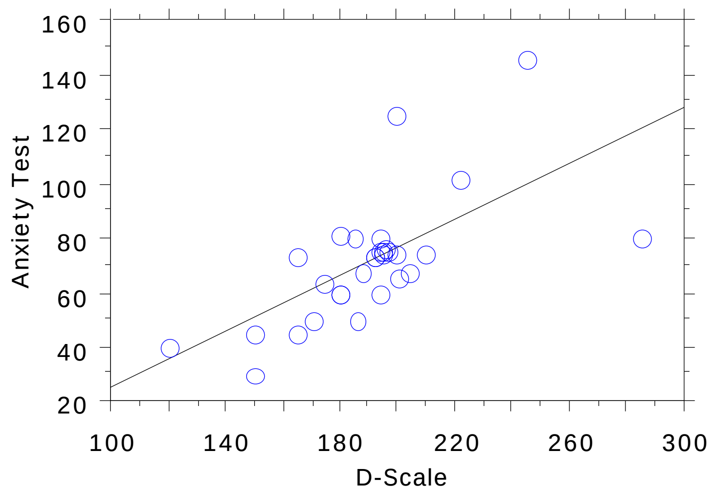
ANOVA Table
Anxiety Test vs. D-Scale

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	6567.946	6567.946	21.466	<.0001
Residual	28	8567.021	305.965		
Total	29	15134.967			

Regression Coefficients
Anxiety Test vs. D-Scale

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	-27.081	21.487	-27.081	-1.260	.2180
D-Scale	.517	.112	.659	4.633	<.0001

Regression Plot



Y = -27.081 + .517 * X; R^2 = .434

Explain what these results tell you about Rokeach's initial hypothesis. Do you find these results compelling in light of the hypothesis? If a person received a score of 220 on the D-Scale, what would you predict that that person would receive on the Anxiety Test? Suppose that a person received a score of 360 on the D-Scale, what would you predict that that person would receive on the Anxiety Test? [10 pts]

The results seem to support the initial hypothesis, because there is a significant positive linear relationship between the D-scale and Anxiety ($p < .05$). However, you might note that very few people have high D-scale scores. With D = 220, the regression equation would yield a predicted Anxiety score of 86.66. Because you didn't observe a person with a D score of 360, you should probably avoid making a prediction of that person's Anxiety score.

7. A researcher is interested in investigating the impact of the drug magnesium pemoline (MgPe) on retention of learned material. A group of 16 people is randomly selected from the population of students at a large university. The people are then randomly assigned to one of four conditions: Placebo, 10cc of MgPe, 20cc of MgPe, or 30cc of MgPe. All the people are given the same material to read and four hours later are tested for retention (higher scores indicate greater retention). Analyze the data as completely as you can, including very clear advice about what to do next. [25 pts]

	Placebo	10cc MgPe	20cc MgPe	30cc MgPe
	8	10	11	10
	6	7	6	8
	7	8	8	7
	5	6	9	9
$\Sigma X=$	26	31	34	34
$\Sigma X^2=$	174	249	302	294
SS =	**5**	**8.75**	**13**	**5**

F_{Max} = 2.6 and $F_{MaxCrit}$ = 20, so I would not be concerned about heterogeneity of variance.

Source	SS	df	MS	F
Drug	**10.69**	**3**	**3.56**	**1.34**
Error	**31.75**	**12**	**2.65**	
Total	**42.44**	**15**		

$F_{Crit}(3,12) = 3.49$

Because $F_{Obt} < F_{Crit}$, you would retain H_0. At this point you have no evidence that MgPe has any impact. So, it's back to the drawing board. You could reasonably conclude that with greater power, you might find some effect of MgPe. To make your experiment more powerful, one of the first changes you could institute would be to include more than 4 participants within each condition. You might have larger dosages of MgPe. You might change the nature of the reading material to make it more difficult, or change the retention interval from 4 to 8 hours.

8. A study was conducted to compare the effectiveness of different rewards that might be used in teaching retarded children. Forty retarded children, ages 5 to 7, were randomly assigned to one of four independent groups. Each child was shown 5 common objects and five cards, each showing the printed name of one of the objects. The child's task was to match each object correctly with its name card. Whenever a correct match was made, the experimenter rewarded the child. Children in Group 1 were rewarded with candy; children in Group 2 were rewarded with tokens that could later be exchanged for candy; children in Group 3 were rewarded with tokens that could be exchanged for attention from the experimenter (playing games, reading to the child, etc.); children in Group 4 were rewarded with verbal praise. The experimenter recorded the number of trials before a child could correctly match all five pairs. (Thus, larger scores indicate less effective learning.) Analyze these data as completely as you can. [20 pts]

ANOVA Table for Trials

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Reward	3	101.075	33.692	3.496	.0252	10.489	.733
Residual	36	346.900	9.636				

Means Table for Trials
Effect: Reward

	Count	Mean	Std. Dev.	Std. Err.
Group 1	10	7.800	1.398	.442
Group 2	10	8.400	1.506	.476
Group 3	10	10.000	5.477	1.732
Group 4	10	11.900	2.079	.657

$F_{Max} = \frac{30}{1.95} = 15.38$ $\qquad F_{MaxCrit} = 6.31$

Thus, because F_{Max} exceeds the critical value, you would conclude that you have heterogeneity of variance. Because you are likely to have violated the homogeneity of variance assumption, you would be concerned about an inflated chance of a Type I error. To compensate for that likelihood, you could change your α from .05 to .025.

Because the P-value is roughly .025, I would treat the results as significant. Thus, I would reject H_0 and compute post-hoc analyses to determine which of the means differed. HSD = 3.72 (q = 3.79).

	Candy	Tokens: Candy	Tokens: Attention	Praise
Candy				
Tokens:Candy	.6			
Tokens: Attention	2.2	1.6		
Praise	4.1*	3.5	1.9	

The results of this study indicate that the children learned best (more quickly) with a candy reward as compared to a praise reward. None of the other differences were significant.

9. It is at least part of the folklore that repeated experience with the Graduate Record Examination (GRE) leads to better scores, even without any intervening study. We obtain eight people and give them the GRE verbal exam every Saturday morning for three weeks. The analyses of these data appear below. Analyze and interpret these data as completely as you can. What advice would you give a prospective taker of the GRE verbal exam? [10 pts]

ANOVA Table for GRE Order

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	7	189112.500	27016.071				
Category for GRE Order	2	1918.750	959.375	3.858	.0463	7.716	.595
Category for GRE Order * Subject	14	3481.250	248.661				

Means Table for GRE Order
Effect: Category for GRE Order

	Count	Mean	Std. Dev.	Std. Err.
GRE1	8	552.500	103.613	36.633
GRE2	8	564.375	91.395	32.313
GRE3	8	574.375	91.785	32.451

You might note that this is the rare repeated measures design in which you would not be able to counterbalance conditions because the order is what's being examined. Nonetheless, you might note that you wouldn't use the exact same test each time, so you would probably counterbalance the actual tests used to ensure that any effect is not due to the fact that Test 1 (for example) always was given as the first test.

Because the F-ratio is significant ($p < .05$), you would need to conduct post hoc analyses to determine which means were different. You would compute HSD = 20.6 ($q = 3.7$).

	GRE1	GRE2	GRE3
GRE1			
GRE2	11.9		
GRE3	21.88*	10	

Thus, the third time one takes the GRE leads to improved performance relative to the first time one takes the GRE. However, no other differences were significant.