

Student ID# _____
(You must use your actual student ID #)

Final Exam

PS 306, Fall 2000

Well, here it is...the last exam that you'll take in this course. I trust that you come to the exam well prepared. The exam should contain relatively few surprises for you. Be sure to comment on any design flaws that you spot as you go through the exam. To ensure that you get as much credit as possible, show all your work. The exam totals 120 points, so it may only take you two hours to complete. ☺ Good luck and may you have a pleasant break.

Peace,

1. As clearly and carefully as you can, articulate the difficulties of removing deception by means of a debriefing. What is a process debriefing and what is the evidence that process debriefing may be useful in removing the effects of deception? [10 pts]

For this answer, you need to first articulate why it is difficult to remove the effects of deception. Then, you need to use the Ross, et al. article to describe process debriefing. Finally, you need to discuss the differences between Exp 1 and Exp2 in the Ross, et al. article to highlight the extent to which process debriefing is effective.

2. Participants were asked to take multiple-choice tests under each of three different conditions of crowding. In one condition (alone), the participant worked alone at a table. In another (three other participants) the same person was at the same table with three other individuals, all working at the same task. In the third condition (seven other participants), seven other people were also at the table, all working on the same task. The dependent variable was the number of items answered correctly out of a total of 60 questions. Does crowding have an effect on the number of items answered correctly? Complete the source table below and interpret the results of the study as completely as possible. [15 pts]

ANOVA Table for Number present

	DF	Sum of Squ...	Mean Sq...	F-Value	P-Value	Lambda	Power
Subject	14	860.578	61.470				
Category for Number present	2	1500.978	750.489	19.868	<.0001	39.735	1.000
Category for Number present * S...	28	1057.689	37.775				

Means Table for Number present

Effect: Category for Number present

	Count	Mean	Std. Dev.	Std. Err.
Alone	15	34.600	8.601	2.221
Three Others	15	28.067	6.088	1.572
Seven Others	15	20.467	5.097	1.316

First of all, it should be clear to you that this is a repeated measures design (the same participant provides data in all three conditions). As such, you would need to counterbalance. Because you have only 3 conditions, you should use complete counterbalancing, which means that you need to run in multiples of 6 participants. With 15 participants, you could not have properly counterbalanced the experiment.

With the significant result ($p < .05$), you would need to compute HSD to determine which means differed. You would get $q = 3.5$, so $HSD = 5.55$. With that HSD you would find that people performed best when alone at the table, next best when 3 others were present, and worst when 7 others were present (as seen below).

	Alone	Three others	Seven Others
Alone			
Three Others	6.4*		
Seven Others	7.6*	14.0*	

3. The following data represent the salaries *and* benefits (total compensations) for United Way executives in 10 cities with the highest compensations for one year during the early 1990s. Also included are the per-capita contributions to the United Way in each of those cities. Is there a relationship between the compensation received by these 10 best paid executives and the amount of contributions raised from their constituents? Suppose that an executive were paid \$200,000 in total compensation. What would expect that chapter of United Way to receive in per-capita contributions? Suppose that an executive were paid \$300,000 in total compensation. What would you expect that chapter to United Way to receive in per-capita contributions? [15 pts]

Regression Summary

Contributions vs. Executive Compensation

Count	10
Num. Missing	0
R	.633
R Squared	.401
Adjusted R Squared	.326
RMS Residual	5.347

ANOVA Table

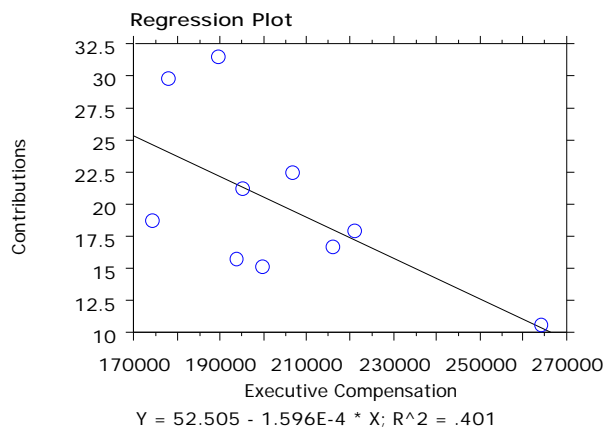
Contributions vs. Executive Compensation

	DF	Sum of Squares	Mean Square	F-Value	P-Value
Regression	1	153.192	153.192	5.357	.0493
Residual	8	228.755	28.594		
Total	9	381.946			

Regression Coefficients

Contributions vs. Executive Compensation

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	52.505	14.147	52.505	3.711	.0059
Executive Compensation	-1.596E-4	6.893E-5	-.633	-2.315	.0493



There is a weak negative linear relationship between executive compensation and contributions ($p < .05$). That is, as compensation increases, contributions go down. This would not be a strong recommendation to pay these executives large salaries! 😊 Of course, it should also occur to you that only the 10 executives with the highest compensation are in this survey, so the relationship may well be stronger (and positive) if a wider range of compensations were considered. Thus, there is a real truncation of range problem. (Oh, you need to note that the scientific notation means that the slope is $-.000159$) You should predict that paying an executive \$200k would yield contributions of \$20.59. You couldn't readily predict the contributions for an executive paid \$300k, because you didn't observe any executive paid that much.

4. An educational psychologist was interested in testing the effectiveness of two different means of teaching reading (phonics vs. sight methods) at three different age grades (Kindergarten, First, and Second Grades). To that end, classes of 20 students from each grade level were selected and 10 were assigned to learn to read by the Phonics method and 10 were assigned to learn to read by the Sight method. The DV is scores on a reading test (higher scores = better reading). Complete the course table below and interpret the results of this study as completely as you can. [20 pts]

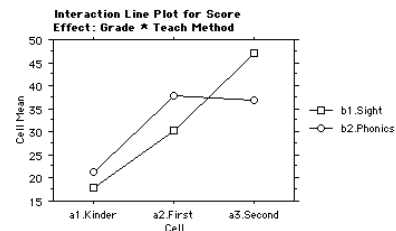
ANOVA Table for Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Grade	2	5113.733	2556.867	217.503	<.0001	435.006	1.000
Teach Method	1	1.667	1.667	.142	.7080	.142	.065
Grade * Teach Method	2	872.133	436.067	37.095	<.0001	74.189	1.000
Residual	54	634.800	11.756				

Means Table for Score

Effect: Grade * Teach Method

	Count	Mean	Std. Dev.	Std. Err.
a1.Kinder, b1.Sight	10	17.800	3.521	1.114
a1.Kinder, b2.Phonics	10	21.400	3.471	1.097
a2.First, b1.Sight	10	30.200	3.490	1.104
a2.First, b2.Phonics	10	37.800	3.490	1.104
a3.Second, b1.Sight	10	47.000	3.091	.978
a3.Second, b2.Phonics	10	36.800	3.490	1.104



Because the interaction is significant, that's where you should focus your attention. First, look at the figure. It appears that for Second Graders, Sight is better than Phonics, but for children in Kindergarten and First Graders, Phonics is better than Sight. Now, I need to compute HSD to see if it confirms what my eyes are telling me. I would use $q = 4.2$, so $HSD = 4.55$. So, using HSD I would conclude that for children in Kindergarten both methods are equally good (though I would recognize that with greater power that difference might become significant). However, for First Graders Phonics leads to better learning than Sight. The Second Graders were different still, with Sight leading to better learning than Phonics.

5. During the penultimate (next-to-last) class meeting, you broke into groups to discuss the ethics of three particular studies. Using only one of the *specific studies* that your group discussed that day, tell me here how you responded to that particular study. If you thought it was unethical, tell me why. If you thought that the study was ethical, tell me why. [10 pts]

This question paid dividends for the students who came to class on the day of the particular ethics discussion, because they could talk about specific studies with no difficulty. For them, I was simply looking to see if they could apply the APA guidelines to the evaluation of the study (as they had done in the class).

6. In the Fine & Kurdek (1993) article, the authors use four different hypothetical cases to make the point that determining authorship in faculty-student collaborative research projects is often complex. They outline a number of principles that they think should guide such decisions. Briefly describe those principles, then briefly describe some recommendations for determining authorship that emerge from those principles. Then, consider the following scenario. Suppose that you completed a senior thesis under the supervision of one of the faculty in our department. Suppose that your thesis work later was incorporated in an article written by the faculty person, along with several other studies. Using the principles and recommendations articulated by Fine & Kurdek, tell me the conditions under which you think that your contribution to the paper should be acknowledged with authorship. [15 pts]

For this question it was critical that the answer include a discussion of justice, beneficence, and parentalism in determining authorship. Then, the answer needed to include using the three principles in determining the authorship in the scenario provided.

7. The state superintendent of instruction asks the director of educational research to investigate differences in scores on a standardized teacher examination for senior education students majoring in the following subject areas: English, Mathematics, Physical Education, and Vocational Education. The following results are from a random sample of 32 graduating seniors (16 males and 16 females). Complete the source table and interpret the results as completely as you can. [15 pts]

ANOVA Table for Score

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Major	3	647.344	215.781	21.159	<.0001	63.478	1.000
Gender	1	3.781	3.781	.371	.5483	.371	.088
Major * Gender	3	37.344	12.448	1.221	.3238	3.662	.278
Residual	24	244.750	10.198				

Means Table for Score

Effect: Major * Gender

	Count	Mean	Std. Dev.	Std. Err.
English, Female	4	37.750	1.708	.854
English, Male	4	37.750	2.630	1.315
Math, Female	4	43.000	5.598	2.799
Math, Male	4	39.000	5.292	2.646
Phys-Ed, Female	4	35.500	2.082	1.041
Phys-Ed, Male	4	34.750	.957	.479
Voc-Ed, Female	4	27.750	1.708	.854
Voc-Ed, Male	4	29.750	2.062	1.031

In this case, the interaction is not significant, so I would focus my attention on the main effects. Only the main effect for major is significant, so that's what I would need to analyze with a post-hoc test. In this case, $q = 3.9$ and $HSD = 4.4$. Thus, Math majors score higher than Phys-Ed and Voc-Ed majors, but don't differ from English majors. English majors score higher than Voc-Ed majors, but don't differ from Phys-Ed majors. Phys-Ed majors score higher than Voc-Ed majors. Of course, you need to recognize that people are not randomly assigned to majors (a non-manipulated characteristic of the participant), so you couldn't make a claim that being in the major caused the difference.

8. Suppose that you were interested in the extent to which participants like particular color combinations. You first create 9 different color combinations (e.g., mauve-chartreuse). You are also interested in the extent to which lighting conditions will influence the judgments, so you also have three different lighting conditions (incandescent light, fluorescent light, sunlight). Thus, you are using a two-factor design (9x3). Your dependent variable is the participant's rating on a 9-point scale (1=yuck, 9=yum). Suppose, further, that you expect that (for reasons of power) you need to have about 30 scores in each of your cells. [15 pts]

a. If you ran this study as a completely between (independent groups) design, how many total participants would you need?

$$9 \times 3 \times 30 = 810$$

b. If you ran this study as a completely within (repeated measures) design, how many total participants would you need?

Too many conditions (27) for complete counterbalancing, so I would use incomplete counterbalancing. Because 27 is an odd number, I would need to mirror-image the counterbalancing scheme, leading to 54 orders. Thus, I would need to run 54 participants to have more than 30 scores per cell.

c. If you ran this study as a mixed design, with the three different lighting conditions as the between (independent groups) factor and the nine different color combinations as the within (repeated measures) factor, how many total participants would you need?

With 9 levels of color combinations, you could use incomplete counterbalancing. Thus, you would need to run in multiples of 18 participants. To get up to 30 scores per cell, you would need to run 36 participants per level of the between-groups factor. Thus, you would need to run $3 \times 36 = 108$ total participants.

d. Which design would you choose to use? Why?

I would choose the mixed design, because it seems to provide a more easily executed study (the completely repeated measures design would be a real imposition on the participants) while still providing a fair amount of power and efficiency.

9. Throughout the semester, we've talked about power. Define power, then label where power occurs in the figure below. While you're at it, label where Type I and Type II errors are to be found as well. [5 pts]

Power is the probability of correctly rejecting H_0 . Thus, when your statistic is drawn from a distribution other than H_0 (in this case labeled "Real"), AND your statistic is sufficiently large that you would reject H_0 (the area with blue dots), then that's where you'd find power. Type II errors occur when your statistic comes from a distribution other than H_0 , but its value is not large enough to lead you to reject H_0 .

