

ID# \_\_\_\_\_

Final Exam

PS217, Spring 1994

First of all, be sure to enter your correct 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 insure 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!!

1. You are hired to determine if students at Skidmore College have SAT scores that differ from those at institutions with which Skidmore likes to be compared (e.g., Colby, Bates, Bowdoin, Williams, Union, etc.), where the mean combined Math and Verbal SAT is *already* known to be 1150. Tell me *explicitly* how you would complete this job for which you were hired and what statistic you would compute to analyze your data. [10 pts]

2. On your next job (were you fired from the first one???) you are asked to determine if there is a relationship between the number of packs of cigarettes per day smoked and the number of days a year that the person is out sick. You look at a random sample of 9 executives of this company and collect the following data:

	<u>Packs</u>	<u>Days Absent</u>
Exec 1	0	0
Exec 2	2	0
Exec 3	0	0
Exec 4	0	1
Exec 5	1	0
Exec 6	1	1
Exec 7	0	1
Exec 8	2	1
Exec 9	0	0
Sum	6	4
SS	6	2.22

What could you conclude from these data and how would you interpret the results? Given these data, if a person smoked 2 packs a day, how many days would you predict that she or he would be absent? [5 pts]

3. You now sample 9 workers from the assembly line at the same company. [15 pts]

(1) Is there a linear relationship between the number of packs of cigarettes smoked and the number of days absent?

(2) If an assembly line worker smokes 9 packs a day, how many days is that person likely to miss in a year?

(3) If an assembly line worker smokes 4 packs a day, how many days is that person likely to miss in a year?

(4) Assume that the results of this study indicate a significant linear relationship between the two variables, would you then be safe in concluding that a good way to cut down on the number of worker absences would be to have a worker cut down on the number of packs of cigarettes smoked? Why or why not? Why might these results differ from those obtained with executives?

	<u>Packs</u>	<u>Days Absent</u>
Worker 1	3	7
Worker 2	4	9
Worker 3	0	1
Worker 4	4	8
Worker 5	2	2
Worker 6	5	10
Worker 7	0	2
Worker 8	1	3
Worker 9	3	6
Sum	22	48
SS	26.22	92

4. Is there a significant difference in the number of days missed by an executive at this company and the number of days missed by an assembly worker? (Use the data from the previous problems to compute the answer to this problem.) [15 pts]

5. You are now asked to determine if there is an effect of three different noise levels in the working environments of the workers on the assembly line. You decide to use a repeated measures design, but you need to explain to the executives at this company why you've chosen to do so. Lay out the pros and cons of such a design now. [5 pts]

6. Here are the results of the StatView analysis for the study you conduct with three different noise levels (80, 90 and 100 decibels). You decide to use the number of errors made in a 3 hour time period as your dependent variable. Complete the source table and interpret the data as completely as you can. [10 pts]

Source:	df:	Sum of Squares:	Mean Square:	F-test:	P value:
Between subjects		7.467			
Within subjects	20	62			
treatments		46.067			.0001
residual		15.933			
Total		69.467			

Group:	Count:	Mean:	Std. Dev.:	Std. Error:
80 dB	10	.5	.707	.224
90dB	10	1.6	.966	.306
100dB	10	3.5	1.08	.342

7. Suppose that you decide to pursue your study of noise levels, but now choose to use an independent groups design with two factors (Gender and Noise Level). {Gender: level 1 = Male, level 2 = Female || Noise Level: level 1 = Lowest Noise...level 5 = Highest Noise} Given the StatView output seen below, complete the source table and fully interpret the results. Oh, you need to know one *important* piece of information to complete the analysis...your best estimate is that  $\sigma^2 = 1$ . To solve this problem you must know what part of the source table indicates your best estimate of  $\sigma^2$ ...and fill it in with the value 1. This isn't a trick question!! We spoke about this (estimating  $\sigma^2$ ) several times in the context of t-tests and ANOVA. Check your notes if you are uncertain. [15 pts]

Source:	df:	Sum of Squares:	Mean Square:	F-test:	P value:
Gender (A)				50	.0001
Noise Level (B)		176			.0001
AB		12			.0244
Error					

Noise Level:		level 1	level 2	level 3	level 4	level 5	Totals:
Gender	level 1	6 1.333	6 3.333	6 4.5	6 6.5	6 7.333	30 4.6
	level 2	6 .833	6 2	6 2.833	6 3.667	6 4.5	30 2.767
Totals:		12 1.083	12 2.667	12 3.667	12 5.083	12 5.917	60 3.683

8. Suppose that you had run an equal number of males and females in the earlier study (problem 7), but had chosen not to include Gender as a factor in the analysis. That is, if all the data remained intact as seen in the earlier problem, but you chose to analyze the data as a single factor independent groups ANOVA with Noise as the only factor, what would your source table look like?? Please explain to me (in words) why the components of the source table would appear as you've shown them below. [5 pts]

Source	SS	df	MS	F
--------	----	----	----	---

---

9. Dr. Lurid is interested in investigating the extent to which different material in erotic films is arousing, so she selected four types of films to show to subjects. Films A and B were both very explicit and graphic in their portrayal of sexual acts, but Film A showed only heterosexual relationships, while Film B showed some homosexual and some heterosexual relationships. Films C and D both had themes with strong sexual content, but the sex acts portrayed were far less graphic. Film C showed only heterosexual relationships, while Film D showed some homosexual and some heterosexual relationships. Dr. Lurid has decided to use an independent groups design, so a subject sees only one of the four films and the dependent variable is the amount of arousal that the subject experiences. Dr. Lurid has collected the data, but having spent all her college hours pursuing sex, she missed out on that most erotic of experiences — a statistics class! Of course, you can help her. First *carefully* determine what sort of design she's used (i.e., read description carefully and think), then conduct the appropriate analysis. Be sure to be very explicit and graphic in your interpretation of her results. [20 pts]

	<u>FILM A</u>	<u>FILM B</u>	<u>FILM C</u>	<u>FILM D</u>	
	6	7	4	1	
	8	7	5	2	
	9	6	3	2	
	8	8	5	3	
	8	6	4	2	
	7	6	3	1	
	8	5	6	1	
	7	4	7	2	
	7	5	8	1	
	6	3	6	1	
	8	5	7	2	
	9	5	7	1	
					SUM
X	91	67	65	19	242
X <sup>2</sup>	701	395	383	35	1514
Mean	7.6	5.6	5.4	1.6	