PRACTICE QUESTIONS FOR FINAL EXAM

1) A company bakes computer chips in two ovens, oven A and oven B. The chips are randomly assigned to an oven and hundreds of chips are baked each hour. The percentage of defective chips coming from these ovens for each hour of production throughout a day is shown below.

Percentage of Defective Chips

Hour	Oven A	Oven B
1	45	36
2	32	37
3	34	33
4	31	34
5	35	33
6	37	32
7	31	33
8	30	30
9	27	24

a) Find the equation of the least-squares regression line for use in predicting the percentage of defective chips produced in an hour by oven B (y) from the percentage produced by oven A (x).

b) Find the correlation coefficient between the percentage defective in Oven A and Oven B. How much of the variability in percentage defective chips from Oven B is explained by the regression line?

c) Use least squares regression to predict the percentage defective from Oven B in an hour when Oven A has a 40 percent defective rate.

2) Once again consider the computer chips data from the preceding problem. Does there appear to be a difference between oven A and oven B with respect to the mean percentages of defective chips produced? Give appropriate statistical evidence to support your answer. Be sure to define any notation you use, justify your statistical procedures, state and check any assumptions (if possible), and write your conclusion clearly in the context of the data.

3) Sleep researchers know that some people are early birds (EB), preferring to go to bed by 10 p.m. and arise by 7 a.m., while others are night owls (NO), preferring to go to bed after 11 p.m. and arise after 8 a.m. A study was done to compare dream recall for early birds and night owls. Twenty-five (25) of *each* of the two types were selected at random from the adult population and asked to record their dreams for one week. Use SPSS to access the dreams data (**dreams.sav**).

Number of Dreams Recalled for One Week					
Early Birds (EB)			Night Owls (NO)		
0	1	5	6	25	5
0	12	1	0	16	25
0	6	0	21	3	5
7	19	24	0	2	5
0	12	14	0	2	13
0	0	10	2	1	5
1	3	1	28	2	2
8	2	1	7	22	11
6			0		

a) The researchers carrying out the study believe that night owls might have better dream recall than early birds. Specifically, they theorize that night owls recall more dreams on average than do early birds. Write down in words and symbols the appropriate null and alternative hypotheses for testing their belief.

b) Now, perform an appropriate significance test to address whether the data constitute evidence against the null hypothesis. Identify the test you are using, report your test statistic and *p*-value, and write a one-sentence conclusion with specific reference to the researchers' theory.

c) Discuss in a brief paragraph the reasons for and against the appropriateness of the inference procedures you used in parts (b) and (c). Give at least one argument *supporting* using the procedures as well as one argument *opposing* them. You may find it helpful to refer to statistical plots but you need not sketch them.

4) For the questions that follow, again refer to the dream data set from Question 2. A second team of researchers suggested that another way of measuring dream recall is to look at the proportion in each group that recalled 5 or more dreams.

   a) For each group (early birds and night owls), how many out of 25 recalled 5 or more dreams in a week?

   b) For each group, obtain a 95% confidence interval for the chance that each recalls 5 or more dreams in a week?

   c) Test at the .05 level of significance whether the proportions of early birds and night owls who recall 5 or more dreams are significantly different.

   d) In order for your inference in b) and c) to be valid, is it necessary that the number of dreams recalled by each group have a normal distribution? Briefly explain why or why not.

5) A pharmaceutical firm would like to show that taking a particular mixture of vitamins and natural herbs will raise the IQ of a teenager. A sample of 15 teenagers is chosen and their IQs are measured before and after taking the vitamin-herb mixture for 1 year. You may wish to enter the data in SPSS for this analysis.

IQ before Vitamin-herb Mixture	103	98	95	107	112	125	111	89	100	121	114	107	95	106	115
IQ after Vitamin-herb Mixture	107	100	96	104	110	129	110	87	97	121	117	105	98	108	119

   a) The pharmaceutical firm claim that after taking the vitamin-herb mixture, teenagers have a higher than average IQ. (Average IQ is 100.) Do the data support their claim? Explain using appropriate statistical procedures.

   b) The firm also claims that teenagers can significantly improve their IQs by taking the mixture. Do the data support this claim? Explain using appropriate statistical procedures.

   c) Is there evidence in these data that teenagers taking this mixture are more likely than not to raise their IQs?

   d) Is there evidence in these data that the firm did not draw a simple random sample of teenagers for this study? Explain using appropriate statistical procedures.

   e) Can the change in IQs be predicted by before-mixture IQs? Explain using appropriate statistical procedures.

6) The statistics department at a large university is trying to determine if it is possible to predict whether an applicant will successfully complete the Ph.D. program or will leave before completing the program. The department is considering whether GPA (grade point average) in undergraduate statistics and mathematics courses (a measure of performance) and mean number of credit hours per semester (a measure of workload) would be helpful measures. To gather data, a random sample of 20 entering students from the past 5 years is taken. The data are given below. Use SPSS to access the data (**phdprog.sav**).

Successfully Completed Ph.D. Program

Student	A	B	C	D	E	F	G	H	I	J	K	L	M
GPA	3.8	3.5	4.0	3.9	2.9	3.5	3.5	4.0	3.9	3.0	3.4	3.7	3.6
Credit Hours	12.7	13.1	12.5	13.0	15.0	14.7	14.5	12.0	13.1	15.3	14.6	12.5	14.0

Did Not Complete Ph.D. Program

Student	N	O	P	Q	R	S	T
GPA	3.6	2.9	3.1	3.5	3.9	3.6	3.3
Credit hours	11.1	14.5	14.0	10.9	11.5	12.1	12.0

a) The average GPA of students completing Ph.D. programs in all fields for the 5-year period under study was 3.1. Do the data provide evidence that GPAs were higher on average for students completing statistics Ph.Ds than for Ph.D. programs in general? Use appropriate statistical procedures to support your answer.

b) Do these data provide evidence that those completing the Ph.D. program had a heavier workload than those that did not? Use appropriate statistical procedures to support your answer.

c) During your first year as a graduate student in statistics at this university, you meet two students who have just completed their Ph.D. programs. One has a GPA that is 0.5 points higher than the other. Using these data, by how much would you expect their mean credit hours per semester to differ? Use appropriate statistical procedures to support your answer.

7) In a test of braking performance, a tire manufacturer measured the stopping distance for one of its tire models. On a test track, a car made repeated stops from 60 miles per hour. The experiment was run on two days, Day 1 on dry pavement and Day 2 on wet pavement, with results shown in the table. Enter and analyze the data with SPSS.

Stopping Distance (ft)	
Dry Pavement	Wet Pavement
145	211
152	191
141	220
143	207
131	198
148	208
126	206
140	177
135	183
133	223

a) Calculate a 95% confidence interval for the mean dry pavement stopping distance. Identify the procedure you are using, check the appropriate assumptions, and explain what your interval means.

b) Is there evidence that average stopping distances are longer on wet than dry pavement? Explain using an appropriate statistical procedure. Identify the statistical procedure you use, check the appropriate assumptions, carry out the procedure, and interpret your results.