5. Sleep researchers know that some people are early birds (E), preferring to go to bed by 10 P.M. and arise by 7 A.M., while others are night owls (N), 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. One hundred people of each of the two types were selected at random and asked to record their dreams for one week. Some of the results are presented below.

	Number of Dreams Recalled During the Week			Proportion Who Recalled	
Group	Mean	Median	Standard Deviation	No dreams	5 or more dreams
Early birds	7.26	6.0	6.94	0.24	0.55
Night owls	9.55	9.5	5.88	0.11	0.69

(a) The researchers believe that night owls may have better dream recall than do early birds. One parameter of interest to the researchers is the mean number of dreams recalled per week with μ_E representing this mean for early birds and μ_N representing this mean for night owls. The appropriate hypotheses would then be $H_0: \mu_E - \mu_N = 0$ and $H_a: \mu_E - \mu_N < 0$. State two other pairs of hypotheses that might be used to test the researchers' belief. Be sure to define the parameter of interest in each case.

(b) Use the data provided to carry out a test of the hypotheses about the mean number of dreams recalled per week given in the statement of part (a). Do the data support the researchers' belief ?

STATISTICS

SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

6. A survey given to a random sample of students at a university included a question about which of two well-known comedy shows, S or F, students preferred. The students were asked the question, "Do you prefer S or F ?" The responses are shown below.

Preference		
S	F	Total
185	139	324

(a) Based on the results of this survey, construct and interpret a 95% confidence interval for the proportion of students in the population who would respond S to the question, "Do you prefer S or F ?"

(b) What is the meaning of "95% confidence" in part (a) ?

(c) In a follow-up survey, a separate group of randomly selected students was asked "Do you prefer F or S ?" The responses are shown below.

Preference		
S	F	Total
68	88	156

Based on these two surveys, is there evidence that the stated preference depends on the order in which the comedy shows were listed in the survey question? Justify your answer.

(d) Suppose the test in part (c) indicates that the order in which the shows were listed does make a difference.

Is the pooled value $\frac{185 + 68}{324 + 156} = 0.527$ a reasonable estimate for the proportion of students at the university who would respond S ? If so, justify your answer. If not, what would be a more reasonable estimate? Explain why.

END OF EXAMINATION

Question 5

Solution

Part (a):

Possibilities include:

1. p_E = proportion of early birds who recall dreams
   p_N = proportion of night owls who recall dreams

   $H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N < 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E < p_N$

   OR

   p_E = proportion of early birds who do not recall dreams
   p_N = proportion of night owls who do not recall dreams

   $H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N > 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E > p_N$

   NOTE: Either of these, BUT NOT BOTH, can be used as one of the possibilities for part (a).

2. p_E = proportion of early birds who recall 5 or more dreams
   p_N = proportion of night owls who recall 5 or more dreams

   $H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N < 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E < p_N$

   OR

   p_E = proportion of early birds who do not recall 5 or more dreams
   p_N = proportion of night owls who do not recall 5 or more dreams

   $H_0: p_E - p_N = 0$ vs. $H_a: p_E - p_N > 0$ OR $H_0: p_E = p_N$ vs. $H_a: p_E > p_N$

   NOTE: Either of these, BUT NOT BOTH, can be used as one of the possibilities for part (a).

3. M_E = median number of dreams early birds recall
   M_N = median number of dreams night owls recall

   $H_0: M_E - M_N = 0$ vs. $H_a: M_E - M_N < 0$ OR $H_0: M_E = M_N$ vs. $H_a: M_E < M_N$

NOTE: 1. A complete response for part (a) requires two pairs of hypotheses.
2. Hypotheses for a chi-square test of homogeneity are not correct since this is a one-sided test.

Question 5 (cont'd.)

Part (b):

Part 1: States a correct pair of hypotheses, identifies a correct test (by name or by formula) and checks appropriate requirements.

μ_E = mean number of dreams early birds recall
μ_N = mean number of dreams night owls recall

$H_0 : \mu_E = \mu_N$ OR $H_0 : \mu_E - \mu_N = 0$
$H_a : \mu_E < \mu_N$ $H_a : \mu_E - \mu_N < 0$

Two-sample *t*-test $$t = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\frac{s_E^2}{n_E} + \frac{s_N^2}{n_N}}}$$

Requirements:
1. Problem states that independent random samples were taken.
2. Normal population distributions or large samples. Since these are not normal, we need to note that $n_E = 100$ and $n_N = 100$ are both large in order to perform the *t*-test.

OR

Two-sample *z*-test $$z = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\frac{s_E^2}{n_E} + \frac{s_N^2}{n_N}}}$$

Requirements:
1. Problem states that independent random samples were taken.
2. Since $n_E = 100$ and $n_N = 100$ are both large, it is OK to perform the approximate *z*-test.

OR

Pooled *t*-test $$t = \frac{\bar{x}_E - \bar{x}_N - 0}{s_p\sqrt{\frac{1}{n_E} + \frac{1}{n_N}}}$$ where s_p^2 is the pooled variance.

Requirements:
1. Problem states that independent random samples were taken.
2. Normal population distributions or large samples. Since these are not normal, we need to note that $n_E = 100$ and $n_N = 100$ are both large in order to perform the *t* test.
3. The sample standard deviations $s_E = 6.94$ and $s_N = 5.88$ are reasonably close, so it is OK to assume that the two population variances are equal, i.e. $\sigma_E^2 = \sigma_N^2$.

Question 5 (cont'd.)

Part 2: Correct mechanics, including the value of the test statistic, df, and P-value (or rejection region)

$$t = \frac{\bar{x}_E - \bar{x}_N - 0}{\sqrt{\frac{s_E^2}{n_E} + \frac{s_N^2}{n_N}}} = \frac{7.26 - 9.55}{\sqrt{\frac{(6.94)^2}{100} + \frac{(5.88)^2}{100}}} = -2.52$$

So, t = -2.52 df = 192 P-value = 0.006

- It is OK to use conservative df of 99.
- Using t-tables: P-value < 0.01
- Using calculator: $t = -2.517578$, P-value = 0.006304, df = 192.799
- Using z-test: P-value = 0.005908
- Using z-table: 0.0059 < P-value < 0.0060
- The pooled t-test results in the same value of t, but a df of 198.

Part 3: Stating a correct conclusion in the context of the problem, using the result of the statistical test.

Because the P-value is small (or less than an α selected and stated by the student), reject H_0. There is convincing evidence that the mean number of dreams night owls recall is greater for than the mean number of dreams early birds recall.

Question 5 (cont'd.)

Scoring

Part (a):
Essentially correct (E) if two distinct pairs of hypotheses are given and the parameters in the hypotheses are defined.

Partially correct (P) if only one pair of correct hypotheses is given and the parameters in these hypotheses are defined or if two "approved" pairs of hypotheses are given but the parameters are poorly defined.

Part (b):
Each of the 3 parts of the hypothesis test is scored either as correct (E) or incorrect (I).
- Because the hypotheses are given in the statement of part (a), they need not be restated here. However, if wrong hypotheses are stated, then part 1 is scored as incorrect.
- Because the problem states that samples are random, it is OK if a student doesn't repeat this.
- Some reference to both samples being large is essential.
- For the pooled *t*- test, some comment on the reasonableness of such an assumption is necessary. It is not sufficient just to say that population variances are equal.

4 Complete Response
Part (a) essentially correct and all three parts of the hypothesis test correct.

3 Substantial Response
Part (a) essentially correct and two parts of the hypothesis test correct.
OR
Part (a) partially correct or incorrect and three parts of the hypothesis test correct.

Part (a) partially correct and two parts of the hypothesis test correct may be scored either as a 2 or a 3, depending on the overall strength of the paper.

2 Developing Response

Part (a) essentially correct and one part of the hypothesis test correct.
OR
Part (a) incorrect and two parts of the hypothesis test correct.

Part (a) partially correct and two parts of the hypothesis test correct may be scored either as a 2 or a 3, depending on the overall strength of the paper.

1 Minimal Response

Part (a) essentially correct.
OR
Part (a) partially correct and zero or one parts of the hypothesis test correct.
OR
Part (a) is incorrect and one part of the hypothesis test correct.