PSTAT 5E Final Examination, Winter 2006
12:00pm - 3:00pm

DO NOT TURN TO THE QUESTIONS UNTIL TOLD TO DO SO!

NAME: _____

PERM #: _____

NAME OF YOUR TA: _____

HOURS OF YOUR DISCUSSION SECTION: _____

INSTRUCTIONS: You have 3 hours to complete this exam. You should attempt ALL questions. You are allowed to consult your textbook and class notes. In order to receive full credit, you need to show your work. Anyone found copying another students' work will be given zero score.

Questions	1	2	3	4	5	6	total
Points available	15	15	15	15	20	20	100
Points awarded							

GOOD LUCK!!!

1. Each year, billions of dollars are spent at theme parks owned by Disney, Universal Studios, Sea World and others. A management consultant claims that 20% of trips include a theme park visit. A survey of 1233 randomly selected people who took trips revealed that 111 of them visited a theme park.

   (i) Construct a 95% confidence interval for the proportion of trips that include a theme park visit.

   (ii) Do these data support the consultant's claim?

2. A mathematical proficiency test were given to randomly selected 13-year-old male and female students. The following tables gives the sample mean scores and standard deviations:

	Sample Size	Mean	Standard Deviation
Male Students	905	474.6	192.5
Female Students	905	473.2	153.4

   (i) Estimate the difference in mean scores between male student and female students and construct the 95% confidence interval.

   (ii) Can you conclude that the mean scores are different for male and female students?

3. The paper "The association of marijuana use with outcome of pregnancy" (*Amer. J. Public Health, 1983, pp.1161-1164*) reported the following data on incidence of major malfunctions among newborns both for mothers who were marijuana users and for mothers who did not use marijuana.

	User	Nonuser
Sample size	1,246	11,178
Number of major malfunctions	42	294

   (i) Construct a 99% confidence interval for the difference between the incidence rate among all mothers who use marijuana and the incidence rate among all mothers who do not use marijuana.

   (ii) Do these data indicate that the incidence rate is higher for mothers who use marijuana?

4. A new program has been developed to enrich the kindergarten experience of children in preparation for the first grade. Pupils in each classroom are tested at the beginning of the school year (pretest) and again at the end of the school year (posttest). The following table gives the scores of 9 randomly selected students exposed to the new curriculum (high score=better performance).

Pupil	1	2	3	4	5	6	7	8	9
Pretest	9	6	14	12	9	8	12	8	11
Posttest	16	11	14	10	14	12	15	11	14

(i) Apply an appropriate test to decide at the 5% level if the new curriculum significantly increased pupil's performance. Follow five steps in the lecture note.

(ii) Specify assumptions for the above test.

(iii) Suppose that further study establishes that, in fact, the population mean score at the beginning is 12.4 and the mean score at the end of the year is 12.3. Refer back to part (i). Did your analysis lead to a

(a) Type I Error; (b) Type II error;
(c) Correct Decision; (d) None of (a)-(c).

Circle the correct response.

(iv) Do you change your conclusion in (i) if $\alpha = .01$?

5. An automobile manufacture recommends that any purchaser of one of its new cars bring it in to a dealer for a 3000-mile checkup. The company wishes to know whether the true average mileage for initial servicing differs from 3000.

   (i) A random sample of 20 recent purchasers resulted in a sample average mileage of 3108 and a sample standard deviation of 273 miles. Does the data suggest that true average mileage for this checkup is something other than the recommended value? Use $\alpha = .01$ and follow five steps in the lecture note.

   (ii) In (i), instead of 20, suppose that the manufacture selected 50 recent purchasers, and gets the same sample mean and standard deviation as in (i). Does the data suggest that true average mileage for this checkup is something other than the recommended value? Use $\alpha = .01$.

   (iii) In (ii), what is the smallest significance level that you will reject the null hypothesis?

   (iv) Specify assumptions for the tests in (i) and (ii).

6. In planning for a meeting with accounting majors, the head of the Accounting Program wants to emphasize the importance of doing well in the major courses to get better-paying jobs after graduation. To support this point, he plans to show that there is a strong relationship between starting salaries for recent accounting graduates and their grade-point average (GPA) in the major courses. Records for seven of last year's accounting graduates are selected at random:

GPA in major courses	Starting salary (in thousands dollars)
2.58	16.5
3.27	18.8
3.85	19.5
3.50	19.2
3.33	18.5
2.89	16.6
2.23	15.6

$$\sum x_i = 21.65, \quad \sum x_i^2 = 68.84, \quad \sum y_i = 124.70, \quad \sum y_i^2 = 2235.75,$$
$$\sum x_i y_i = 390.69, \quad S_{xx} = 1.88, \quad S_{yy} = 14.31, \quad S_{xy} = 5.01$$

(i) What are dependent and independent variables?

(ii) Find and report the least-square regression line.

(iii) How much of the variability in starting salary is explained by the GPA in major courses?

(iv) Find 95% confidence interval for the slope. Interpret the point and interval estimates of the slope.

(v) Obtain a 95% confidence interval for the expected starting salary of all graduates with major GPA 3.0.

(vi) Obtain a 95% confidence interval for a graduate with major GPA 3.0.

(vii) Suppose 5 graduates each has major GPA 3.0. Do you expect these 5 graduates to have exactly the same starting salary?