ELEMENTS OF STATISTICS I
COMMON FINAL
FALL 1996

PLEASE PRINT THE FOLLOWING INFORMATION:

NAME: _____ INSTRUCTOR: _____

STUDENT ID: _____ SECTION/TIME: _____

THIS EXAM HAS TWO PARTS.

PART I.Consists of 30 multiple choice questions (4 points each). Read all questions carefully. You may use the test to do calculations. Mark the number of the opscan sheet corresponding to the test questions number with a Number 2 pencil or a mechanical pencil with an HB lead. Mark only the answer; otherwise the answer will be counted as incorrect. In case there is more than one answer, mark the best answer. Please make sure that your name appears on the opscan sheet in the spaces provided.

PART II.

This part consists of 3 questions (Total of 40 pts). You must show all work for each question in the space provided to receive full credit for that question. If you write your explanations in another part of the test, please indicate accordingly.

FOR DEPARTMENTAL USE ONLY:

Part II

Questions	1	2	3	4
Score				

Part I	Part II	TOTAL

STAT 122X
Common Final Examination
Fall 1996

PART I

The following data gives the number of children of 10 randomly sampled female workers, beyond child-bearing age:

2, 3, 2, 1, 1, 2, 1, 4, 0, 1

1. The mean of this data is

   a) 17
   b) 1.9
   c) 1.8
   d) 1.7
   e) None of these

2. The standard deviation of the above data is closest to

   a) 1.951
   b) 1.095
   c) 1.240
   d) 1.159
   e) 1.303

3. Ten presumably trained rats were released in a maze. Their time to escape (in seconds) are recorded below. M and N represent the time for two rats that had still not escaped by the end of the experiment.

   100 38 M 122 95 116 56 135 N 104

   Calculate the median for these data.

   a) 102
   b) 110
   c) 105.5
   d) 108
   e) cannot be determined

   The next two questions refer to the following figure concerning examination scores of a group of students (possible score was 40 to 100 and every score appeared).

4. The proportion of all scores at 70 or above was:

   a) .45
   b) .65
   c) .35
   d) .10
   e) .40

5. The 90th percentile on the exam was:

   a) 70

b) 75

c) 80

d) 85

e) 90

6. A professor believes that if a class is allowed to work on an exam as long as desired, the times spent by the students would be approximately mound-shaped with mean 40 minutes and standard deviation 5 minutes. Using the Empirical Rule, approximately how long should be allotted for the exam if the professor wants 97.5% of the class to finish?

   a) 45

   b) 50

   c) 38

   d) 35

   e) 40

7. A sample space consists of 5 mutually exclusive events E_1, E_2, E_3, E_4, and E_5. If $P(E_1) = \frac{1}{2}$, $P(E_2) = \frac{1}{4}$, $P(E_3) = \frac{1}{8}$, and $P(E_4) = P(E_5)$, find $P(E_5)$.

   a) $\frac{1}{2}$

   b) $\frac{1}{8}$

   c) $\frac{1}{4}$

   d) $\frac{1}{32}$

   e) $\frac{1}{16}$

   Use the following to answer questions 8 and 9.

   In a survey, 200 shoppers at a large mall were asked two questions: (1) Did you see a TV ad for sale at department store X? (2) Did you shop at department store X? The responses to the questions are summarized as follows:

	Shopped at X	Did not shop at X
Saw ad	90	35
Did not see ad	30	45

8. Find the probability that a randomly selected person in the survey saw ad or did not shop at X.

   a) 0.175

   b) 0.45

   c) 1.025

   d) 0.85

   e) 0.225

9. Find the conditional probability that a randomly select person in the survey shopped at store X, given that the person saw the ad.

   a) 0.45

   b) 0.72

   c) 0.60

   d) 0.25

e) .15

Use the following to answer questions 10 - 11.

A discrete random variable x has the following probability distribution:

x	0	2	4	6
$p(x)$	.1	.2	?	.6

10. Find $p(4)$.

a) 0.2
b) 0.4
c) 0.7
d) 0.1
e) 0.9

11. The mean μ and standard deviation σ of the random variable x are

a) $\mu = 4$ $\sigma = 2.154$
b) $\mu = 4$ $\sigma = 4.64$
c) $\mu = 4.4$ $\sigma = 1.71$
d) $\mu = 4.4$ $\sigma = 2.154$
e) $\mu = 3$ $\sigma = 1.71$

12. An automobile manufacturer has determined that 20% of all gas tanks that were installed on its 1992 compact model are defective. If 10 of the cars are recalled by a particular dealer, what is the probability that more than four of the ten will need new gas tanks?

a) 0.201
b) 0.088
c) 0.967
d) 0.033
e) .20

13. Consider the following probability distribution:

$$p(x) = \frac{5!}{x!(5-x)!}\,(.7)^x\ (.3)^{5-x}\,, \quad x = 0, 1, \ldots, 5.$$

Find the mean of the random variable x.

a) 5
b) 35
c) 3.5
d) 1.5
e) .21

14. Let z be the standard normal random variable, find $P\,(1.13 \leq z \leq 1.85)$.

a) 0.0322
b) .4678
c) .3708
d) .8386
e) 0.097

15. According to The World Almanac, the mean travel time to work in New York State is 29 minutes. Let x be the time, in minutes, that it takes a randomly selected New Yorker to get to work on a randomly selected day. If the travel times are normally distributed with a standard deviation of 8 minutes, find $P(x < 45)$.

a) .4225
b) .0228
c) .4772
d) .9772
e) .1235

16. According to the Bureau of Labor Statistics, the mean weekly earning of workers in the trucking industry is \$456 with a standard deviation of \$63. Suppose 81 workers in the trucking industry are to be selected at random. Find the probability that the mean weekly earnings of these 81 workers is between \$442 and \$470.

a) 0.0456
b) 0.4772
c) 0.0228
d) 0.0871
e) 0.9544

17. A major airline wants to estimate its average number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected, and the number of unoccupied seats is noted for each of the sampled flights. The sample mean and standard deviation are $\bar{x} = 11.6$ seats and $s = 4.1$ seats. Set up a 95% confidence interval for μ, the mean number of unoccupied seats per flight during the past year.

a) $(11.15, 12.45)$
b) $(11.06, 12.13)$
c) $(12.18, 11.02)$
d) $(11.63, 11.56)$
e) $(11.51,\ 12.54)$

18. How large a sample size should one use to estimate p, the population proportion from a binomial population to within .01 at a 95% confidence level?

a) 83
b) 98
c) 6766
d) 9604

e) 2401

19. Assume that the duration time for long distance phone calls is normally distributed but with unknown mean and standard deviation. A random sample of 15 calls were timed and yielded the statistics $\bar{x} = 12.5$ minutes and $s = 4.1$ minutes. A 95% confidence interval for the mean duration of all long-distance calls is

    a) $12.5 \pm 2.131 \frac{4.1}{\sqrt{15}}$
    b) $12.5 \pm 1.96 \frac{4.1}{\sqrt{15}}$
    c) $12.5 \pm 1.761 \frac{4.1}{\sqrt{15}}$
    d) $12.5 \pm 2.145 \frac{4.1}{\sqrt{15}}$
    e) $12.5 \pm 1.64 \frac{4.1}{\sqrt{15}}$

20. Which of the following sets of hypotheses have a correct set-up for H_0 and H_a ?

    I. $H_0 : \mu = 90$, $H_a : \mu < 100$
    II. $H_0 : \bar{x} = 100$, $H_a : \bar{x} \neq 100$
    III. $H_0 : \mu \geq 90$, $H_a : \mu < 90$
    IV. $H_0 : \mu = 45$, $H_a : \mu \neq 45$
    V. $H_0 : \mu \leq 45$, $H_a : \mu > 45$

    a) I, II, and IV only
    b) III, IV, and V only
    c) IV and V only
    d) II and IV only
    e) II, III, IV, and V only

Use the following statement to answer questions 21-22.

The Chronicle of Higher Education Almanac (Sep. 1990) reported that for the 1989-1990 academic year, 4-year private colleges charged students an average of \$8,450. Suppose that for 1990-1991 a random sample of 36 colleges yielded the following data on tuition and fees: $\bar{x} = \$9,192$ and $s = \$1,843$.

21. Specify the null and alternative hypotheses you would use to investigate whether the mean amount for tuition and fees in 1990-1991 was larger than it was in 1989-1990.

    a) $H_0 : \mu = 9192$
       $H_a : \mu > 9192$
    b) $H_0 : \bar{x} = 8450$
       $H_a : \bar{x} > 8450$
    c) $H_0 : \mu = 9192$
       $H_a : \mu \neq 9192$
    d) $H_0 : \mu = 8450$
       $H_a : \mu \neq 8450$
    e) $H_0 : \mu = 8450$
       $H_a : \mu > 8450$

22. The p-value for the hypothesis test described in question 21 is closest to

a) .05

b) .4920

c) .008

d) .9920

e) .01

23. A random sample of size 20 is taken from a normally distributed population of unknown variance, in order to test $H_0 : \mu = 40$ vs. $\mu \neq 40$. If the value of the test statistic is 1.42, then for a test at the $\alpha = .10$ level of significance, the only correct statement is:

    a) Reject H_0, since 1.42 is larger than the critical value 1.328

    b) Reject H_0, since 1.42 is larger than the critical value 1.282

    c) Do not reject H_0, since 1.42 is less than the critical value 1.645

    d) Do not reject H_0, since 1.42 is less than the critical value 1.729

    e) Do not reject H_0, since 1.42 is less than 2.539

Use the following to answer questions 24-26.

A manufacturer of automobile shock absorbers was interested in comparing the durability of its shocks with that of the main competitor. To make the comparison, one of the manufacturer's and one of the competitor's shocks were randomly selected and installed on the rear wheels of six cars. After the cars had been driven 20,000 miles, the strength of each test shock was measured, coded, and recorded. The following are the results of the examination:

Car	1	2	3	4	5	6
Manufacturer's	8.8	10.5	12.5	9.7	9.6	13.2
Competitor's	8.4	10.1	12.0	9.3	9.0	13.0

Assume that the population of differences has a normal distribution.

24. Let μ_1 be the population mean strength of the manufacturer and let μ_2 be the population mean strength of the competitor after 20,000 miles. Specify the appropriate null and alternative hypotheses to test whether there is a difference in the mean strength of the two types of shocks after 20,000 miles.

    a) $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 > 0$

    b) $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 < 0$

    c) $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 \neq 0$

    d) $H_0 : \mu_1 - \mu_2 = 20,000$, $H_a : \mu_1 - \mu_2 > 20,000$

    e) $H_0 : \mu_1 - \mu_2 = 20,000$, $H_a : \mu_1 - \mu_2 \neq 20,000$

25. The value of the test statistics is

    a) 2.14

    b) 5.61

    c) 7.66

    d) -2.41

    e) 3.13

26. Based on the test statistic, we conclude that

    I. At $\alpha = 0.05$, the sample evidence suggests a significant difference between the strength of the two types of shocks.

    II. At $\alpha = 0.01$, the data present sufficient evidence to conclude there is a difference in the mean strength of the two types of shocks.

    III. At any significance level greater than or equal to .001, the data present sufficient evidence to conclude there is a difference in the mean strength of the two types of shocks.

    Choose one of the following:

    a) Only I is correct
    b) Only II is correct
    c) Only III is correct
    d) Only I and III are correct
    e) I, II, and III are all correct

27. The correlation coefficient for the accompanying scatter gram is approximately equal to

    a) 1
    b) -.30
    c) .80
    d) -.95
    e) 0

28. Regarding a simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$, which of the following statements are correct?

    I. The coefficient of determination is the square root of the correlation coefficient.

    II. The correlation coefficient r satisfies $-1 \leq r \leq 1$.

    III. The confidence interval for mean of y at $x = x_p$ is always contained in the prediction interval for an individual value of y , at $x = x_p$.

    a) I only
    b) I and II only
    c) I and III only
    d) II and III only
    e) I, II and III

    Use the following statement for questions 29-30.

    A fire insurance company is interested in relating the amount of fire damage, y , in major residential fires to the distance, x , between the residence and the nearest fire station.

    Based on a sample of recent fires in a large suburb of a major city, the following linear regression mode is obtained:

    $\widehat{y} = 10.278 + 4.919x$, $r^2 = .92$

29. Interpret the value of the coefficient of determination.

    a) 92% of the observed damages will fall within two standard deviation of the least squares line.

b) There is a strong negative correlation between the fire damage y and the distance x.

c) 92% of the total sample variability around $\bar{y}$ is explainable by the linear relation between y and x.

d) We are 92% confident that the fire damage is related linearly to the distance x.

e) None of the above.

30. The estimated fire damage to a residence 3.5 miles from the nearest fire station is closest to (assuming 3.5 miles is within the range of x values)

a) 17.216

b) 10.278

c) 27.495

d) 22.3

e) 32.7

PART II

1. A pharmaceutical company wants to test Dozenol, a new cold medicine intended for night use. Tests for such products often include a "treatment group" of people who use the drug, and a "control group" of people who don't use the drug. Fifty people with colds are given Dozenol, and 100 others are not. The systolic blood pressure is measured for each subject, and the sample statistics are summarized below:

Treatment group	Control group
$n_1 = 50$	$n_2 = 100$
$\bar{x}_1 = 203.4$	$\bar{x}_2 = 189.4$
$s_1 = 39.4$	$s_2 = 39.0$

a) Specify the null and alternative hypotheses the pharmaceutical company should use in testing whether Dozenol increases the mean blood pressure. (4 pts)

b) Find the test statistics for the problem. (5pts)

c) At 5% significance level, do the data provide sufficient evidence to conclude that Dozenol increases blood pressure? Show your work. (6 pts)

2. A rock concert producer has scheduled an outdoor concert for Saturday, April 24th. If it does not rain, the producer expects to make $20,000 profit from the concert. If it does rain, the producer will be forced to cancel the concert and will lose $12,000. The producer has learned from the National Weather Service that the probability of rain on April 24th is is .4 .

a) Set-up a probability distribution table to determine the producer's expected profit from the concert. (4 pts)

b) For a fee of $1,000 an insurance company offers to insure the producer against all losses resulting from the rained-out concert. If the producer buys the insurance, find his expected profit from the concert. (4 pts)

3. One of the many variables that influences the sales of existing single-family homes is the interest rate charged for mortgage loans. The following table shows the average annual mortgage interest rate, x , and the total number of existing single-family homes sold annually, y , for 1982-1991.

Interest rate x (%)	Homes sold y (thousands)	
14.8	1,990	
12.3	2,719	$\bar{x} = 10.68$, $\bar{y} = 3125.3$
12.0	2,868	$SS_{xx} = 32.116$
11.2	3,214	$SS_{xy} = -7926.54$
9.8	3,565	$SS_{yy} = 2,166,654.1$
8.9	3,526	
9.0	3,594	
9.8	3,346	
9.8	3,211	
9.2	3,220	

a) Find the least squares line relating y to x. (4 pts)

b) Calculate the correlation coefficient, and interpret its value. (4 pts)

c) Do the data provide sufficient evidence to indicate that mortgage interest rates affect the annual number of existing single-family homes sold? Test using $\alpha = .05$. (4pts)

d) Find a 90% confidence interval for the mean annual number of existing single-family homes sold if the average annual mortgage interest rate is 10.0%. (5 pts)