STAT 1220 **Fall 2011**

Common Final Exam **December 9, 2011**

PLEASE PRINT THE FOLLOWING INFORMATION:

Name: ______________________ Instructor: ______________________

Student ID #: ______________ Section/Time: ______________

THIS EXAM HAS TWO PARTS.

PART I.
Part I consists of 30 multiple choice questions. Each correct answer is scored 2 points; each incorrect (or blank) answer is scored 0, so there is no penalty for guessing. You may do calculations on the test paper, but your answers must be marked on the OPSCAN sheet with a soft lead pencil (HB or No. 2 lead). Any question with more than one choice marked will be counted as incorrect. If more than one choice seems correct, choose the one that is most complete or most accurate. Make sure that your name and ID number are written and correctly bubbled on the OPSCAN sheet.

PART II.
Part II consists of 3 free response questions, with values as indicated. You must show all work in the space provided or elsewhere on the exam paper in a place that you clearly indicate. Work on loose sheets will not be graded.

FOR DEPARTMENT USE ONLY:

Part II.

Question	1	2	3
Score			

Part I	Part II	TOTAL

Part I

Problems 1 through 3 pertain to the following situation:

5-lb Bags of fresh peaches are sold at a farmers market. Nine such bags are randomly sampled and the number of peaches in each bag is noted as follows:

6, 7, 7, 7, 6, 8, 6, 7, 8

1. The mean of this data set is about:

   (a) 6.00 (b) 6.89 (c) 7.00 (d) 7.55 (e) 8.00

2. The median of this data set is about:

   (a) 8.00 (b) 7.00 (c) 6.00 (d) 6.50 (e) 7.50

3. The sample standard deviation of this data set is about:

   (a) 6.89 (b) 8.00 (c) 0.61 (d) 7.00 (e) 0.78

Problems 4 and 5 pertain to the following situation:

The annual 2-mile fun-run is a traditional fund-raising event to support local arts and sciences activities. It is known that the mean and standard deviation of finish times for this event are respectively $\mu = 30$ and $\sigma = 5.5$ minutes. Suppose the distribution of finish times is approximately bell-shaped and symmetric.

4. Find the approximate proportion of runners who finish in under 19 minutes.

   (a) 16% (b) 32% (c) 5% (d) 2.5% (e) 97.5%

5. Which one of the statements is correct?

   (a) The proportion of runners finished in under 24.5 minutes is approximately 68%.
   (b) The proportion of runners finished in under 24.5 minutes is approximately 95%.
   (c) The proportion of runners finished in under 30 minutes is greater than the proportion of runners finished above 30 minutes.
   (d) The proportion of runners finished in under 30 minutes is approximately equal to the proportion of runners finished above 30 minutes.
   (e) The proportion of runners finished in under 30 minutes is less than the proportion of runners finished above 30 minutes.

6. A sample data set contains $n = 2$ observations. Suppose one of the two observations is 1 and the sample standard deviation is $s = 0$. Identify the correct statement(s) among the following .

I. Each of the two observations is 1.
II. It is not possible to have a data set containing 2 observations and yet to have a variance equal to 0.
III. The other observation must be -1.

(a) I, II and III (b) I only (c) II only (d) III only (e) None

7. The variability of a sample data set is measured by which of the following statistics?

I. most frequent value
II. sample size
III. range
IV. standard deviation
V. median

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

Problems 8 and 9 pertain to the following situation:

Professor Jackson is in charge of a program to prepare students for a high school equivalency exam. Records show that, in the program, 80% of the students need work in mathematics, 70% need work in English, and 55% need work in both areas. One person is to be randomly selected from this population of all students in the program. Let

M = the selected person needs help in Mathematics
E = the selected person needs help in English

8. The probability that the selected person needs help in English and in Mathematics, *i.e.*, $P(E \text{ and } M)$ is

(a) 0.56 (b) 0.55 (c) 0.45 (d) 0.44 (e) 0.95

9. The probability that the selected person needs help in English or in Mathematics, *i.e.*, $P(E \text{ or } M)$ is

(a) 1.00 (b) 0.55 (c) 0.95 (d) 0.56 (e) 0.45

Problems 10 and 11 pertain to the following situation:

The proportions of families with various numbers of children age 18 or under in a small town are given in the following table. One family is randomly selected from this town.

x	0	1	2	3	4	5
$P(x)$	.10	.40	.30	.10	.05	.05

10. Find the probability that the selected family has at least 3 children age 18 or under.

    (a) .10 (b) .20 (c) .30 (d) .40 (e) .50

11. Find the average number of children age 18 or under per family in this town.

    (a) 1.75 (b) 1.0 (c) 1.5 (d) 2.0 (e) 2.5

Problems 12 and 13 pertain to the following situation:

The heights of fully grown English oak (also known as Brown oak) trees are normally distributed with a mean of 95 feet and a standard deviation of 10 feet.

12. What proportion of fully grown English oak trees are taller than 110 feet?

    (a) 0.1587 (b) 0.8413 (c) 0.9332 (d) 0.1321 (e) 0.0668

13. If three fully grown English oak trees are randomly selected, what is the probability that all three are over 110 feet tall?

    (a) 0.0040 (b) 0.5955 (c) 0.8127 (d) 0.0023 (e) 0.0003

14. At the end of each semester, Professor Mann calculates an overall score for each of his students in large sections of an Introductory Statistics course. The overall score is calculated based on each student's performance on homework, attendance, tests, quizzes and a final exam. A final grade is then assigned based on the overall score for the course. In a particular semester, the scores are normally distributed with a mean score of $\mu = 78$ and a standard deviation $\sigma = 6$. Professor decides to give "A" to the top 14% of the students. What is the the minimum score a student can get and still get an "A"?

    (a) 84.5 (b) 90.0 (c) 89.8 (d) 85.7 (e) 87.9

Problems 15 and 16 pertain to the following situation:

A blood bank catalogs the types of blood, including positive or negative Rh-factor, given by donors during the last five days. The number of donors who gave each blood type is listed below. Suppose a donor is selected at random from the this group of 409 donors.

	Blood Type				
	O	**A**	**B**	**AB**	Total
Rh-Positive	156	139	37	12	344
Rh-Negative	28	25	8	4	65
Total	184	164	45	16	409

15. Find the probability that the donor has blood type **O** or type **A** blood.

(a) 0.3124 (b) 0.8411 (c) 0.1296 (d) 0.8509 (e) 0.7213

16. Given that the selected donor has negative Rh-factor, find the probability that the donor has type **B** blood.

(a) 0.2162 (b) 0.8222 (c) 0.1231 (d) 0.1778 (e) 0.1076

17. To estimate the mean retail price of designer jeans, a buyer sampled 19 retailers in New York City and recorded the selling price of the jeans. The following statistics were reported: $\bar{x} = 51.75$ dollars, and $s = 5.50$ dollars. Assuming the prices at different retailers are normally distributed, a 99% confidence interval for the mean is closest to:

(a) (48.12,55.38) (b) (48.50,55.00) (c) (47.12,56.38) (d) (47.49,56.01) (e) (50.25,53.00)

Problems 18 through 20 pertain to the following situation:

Economic theory suggests that quit rate (quits per 100 employees) and average hourly wage are related. The table below lists quit rates (y) and the average hourly wage (x) in a sample of 15 manufacturing industries.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	1.4	0.7	2.6	3.4	1.7	1.7	1.0	0.5	2.0	3.8	2.3	1.9	1.4	1.8	2.0
x	8.2	10.4	6.2	5.4	9.9	9.1	10.6	13.3	8.0	5.5	7.5	6.4	8.8	10.9	8.8

This data set produces the following statistics: $\bar{x} = 8.6$, $\bar{y} = 1.88$, $SS_{xx} = 68.82$, $SS_{xy} = -23.88$, and $SS_{yy} = 11.324$.

18. The regression equation is closest to:

(a) $\hat{y} = 1.31 - 0.298x$
(b) $\hat{y} = 1.31 + 0.298x$
(c) $\hat{y} = -1.31 - 0.298x$
(d) $\hat{y} = -0.347 + 0.486x$
(e) $\hat{y} = 4.86 - 0.347x$

19. At $x_0 = 10$ dollars per hour, the predicted value of y based on the regression equation above is closest to:

(a) $\hat{y} = 0.92$ (b) $\hat{y} = 3.11$ (c) $\hat{y} = 4.29$ (d) $\hat{y} = 4.51$ (e) $\hat{y} = 1.39$

20. The coefficient of correlation and the coefficient of determination are closest to:

(a) $(r = -0.8554, r^2 = 0.7317)$
(b) $(r = -9539, r^2 = 0.9100)$
(c) $(r = 0.8554, r^2 = 0.7317)$

(d) $(r = 0.9539, r^2 = 0.9100)$
(e) $(r = -0.8891, r^2 = 0.7905)$

Problems 21 and 22 pertain to the following situation:

Suppose that you have a sample of size $n = 100$ with mean $\bar{x} = 5$ and standard deviation $s = 2$, and that you are to construct a confidence interval for the population mean μ.

21. A 95% confidence interval for μ is closest to:

(a) (4.744,5.256) (b) (4.671,5.329) (c) (4.608,5.392) (d) (4.485,5.515) (e) (4.534,5.466)

22. If, after you obtain the confidence interval, you find it to be too wide, which of the following remedial steps can you take to reduce the width of the confidence interval?

I. To construct a 90% confidence interval instead of a 95% one.
II. To construct a 99% confidence interval instead of a 95% one.
III. To re-do the 95% confidence interval with only a half of the sample data.

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

23. In a situation of hypothesis testing, what happens when the null hypothesis H_0 is wrongly rejected?

(a) The Type I error probability is 1 or 100%.
(b) The Type I error probability is 0 or 0%.
(c) The Type II error probability is 0.5 or 50%.
(d) A Type I error occurs.
(e) A Type II error occurs.

Problems 24 through 27 pertain to the following situation:

A standard painkiller used for patients after minor surgeries is known to bring relief in 3.5 minutes on the average (μ). A new painkiller is hypothesized to bring faster relief to patients. To test this hypothesis, a sample of 19 patients with minor surgeries are selected and given the new painkillers. This sample yields a mean of 2.8 minutes and a standard deviation of 1.1 minutes. Do the data provide sufficient evidence to indicate that the new painkiller indeed works faster? Assume the times to relief are normally distributed and use $\alpha = 0.05$.

24. Which of the following pairs of hypotheses is appropriate for this study?

(a) $H_0 : \mu = 2.8$ vs. $H_a : \mu < 2.8$
(b) $H_0 : \mu = 3.5$ vs. $H_a : \mu < 2.8$
(c) $H_0 : \mu = 3.5$ vs. $H_a : \mu > 3.5$

(d) $H_0 : \mu = 3.5$ vs. $H_a : \mu < 3.5$

(e) $H_0 : \mu = 3.5$ vs. $H_a : \mu \neq 3.5$

25. The test statistic is closest to:

(a) $t = -2.77$ (b) $t = -0.64$ (c) $t = 0.64$ (d) $t = 2.77$ (e) $t = 3.5$

26. The rejection region for the test is:

(a) $(-\infty, -1.645]$ (b) $(-\infty, -1.729]$ (c) $(-\infty, -1.729] \cup [1.729, \infty)$ (d) $(-\infty, -1.734]$

(e) $(-\infty, -1.645] \cup [1.645, \infty)$

27. The correct decision and justification are:

(a) Reject H_0 because $2.8 < 3.5$.

(b) Reject H_0 because $z < -1.645$.

(c) Reject H_0 because the test statistic falls in the rejection region.

(d) Do not reject H_0 because the test statistic does not fall in the rejection region.

(e) Do not reject H_0 because the p-value is greater than $\alpha = 0.05$.

28. In testing $H_0 : p = 0.25$ vs. $H_a : p \neq 0.25$ with a sample proportion $\hat{p} = 0.30$ based on a sample of size $n = 400$. The observed significance (the p-value) of the test is closest to:

(a) 2.31 (b) 0.0104 (c) 0.0208 (d) 0.05 (e) 0.025

29. A college administrator would like to estimate the average grade point average (average of GPA) of all the currently registered students based on a random sample. He plans to estimate the average GPA to within 0.05 with a 90% confidence interval. At least how many students does he need to include in the sample in order to accomplish that? Assume the standard deviation of all GPA's in the student population is $\sigma = 0.4$.

(a) 174 (b) 196 (c) 96 (d) 361 (e) 97

30. The age distribution for all the residents of Alaska has a mean of $\mu = 31$ and a standard deviation $\sigma = 19.5$. If a sample of size $n = 140$ is randomly selected from that population and the sample mean is noted, the probability that the sample mean age will exceed 34 is closest to:

(a) 0.0344 (b) 0.9987 (c) 0.9656 (d) 0.0013 (e) 0.1401

Part II

1. An insurance company study shows that 60% of the auto insurance claims submitted for property damage were submitted by males under 25 years of age in a particular region of the country. Suppose 8 property damage claims involving automobiles are selected at random from that region. Let x be the number of claims (among the 8 selected) that are made by males under age 25.

   (a) Find the mean and standard deviation of x. [4 points]

   (b) Find the probability that all 8 claims are made by males under age 25. [3 points]

   (c) Find the probability that exactly 2 of the 8 claims are made by males under age 25. [3 points]

   (d) Find the probability that x is less or equal to 2. [3 points]

2. A breeder of thoroughbred horses wishes to model the relationship between the gestation period in days (x) and the life span in years (y) of a horse . The breeder believes that the two variables may follow a linear trend.The information in the following table was supplied to the breeder from various thoroughbred stables across the state.

$Horses$	1	2	3	4	5	6	7
y	24	25.5	20	21.5	22	23.5	21
x	416	279	298	307	356	403	265

This data set produces the following statistics:
$n = 7$, $\sum x = 2324$, $\sum x^2 = 793320$, $\sum y = 157.5$, $\sum y^2 = 3565.75$, $\sum xy = 52526.5$

(a) Compute SS_{xx}, SS_{xy}, and SS_{yy}. [3 points]

(b) Compute the slope of the regression equation b and s_e. [2 points]

(c) In testing $H_0 : B = 0$ vs. $H_a : B > 0$, compute the appropriate test statistic. [3 points]

(d) Identify the rejection region for the test at $\alpha = 0.05$. [2 points]

(e) Make a decision whether to reject the null hypothesis at $\alpha = 0.05$. [2 points]

(f) State your conclusion in the context of the problem. [2 points]

3. A new type of crank for bicycles called the Powercam relies on a pushing motion rather than a circular motion. The company that markets this crank plans a test of five riders who will travel a 100-mile route once with the Powercam and once with a conventional crank. The company claims that the new crank reduces the time needed to cover the 100 miles. The times in hours for the five riders are as follows:

Rider	1	2	3	4	5
Powercam	4.4	4.6	4.7	4.8	4.9
Conventional	5.1	5.3	5.1	4.9	5.2

Is there sufficient evidence in the data to support the claim that the new crank reduces the time needed to cover the 100 miles? Use a 5% significance level and assume that both population distributions are normal.

(a) State the null and alternative hypotheses for the test. [2 points]

(b) Identify the correct formula for the test statistic and calculate its value. [4 points]

(c) Identify the rejection region. [3 points]

(d) Make a decision whether to reject H_0. [2 points]

(e) State a conclusion in the context of the problem. [2 points]