

---

AP Statistics Diagnostic Test

---

## SECTION I

Time: 1 hour and 30 minutes

Number of questions: 40

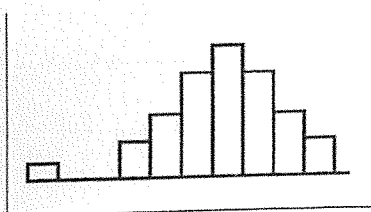
Percent of total grade: 50

**Directions:** Use the answer sheet provided on the previous page. All questions are given equal weight. There is no penalty for unanswered questions, but  $\frac{1}{4}$  of the number of incorrect answers will be subtracted from the number of correct answers. The use of a calculator is permitted in all parts of this test. You have 90 minutes for this part of the test.

1. Eighteen trials of a binomial random variable  $X$  are conducted. If the probability of success for any one trial is 0.4, write the mathematical expression you would need to evaluate to find  $P(X = 7)$ . Do not evaluate.
  - a.  $\binom{18}{7}(0.4)^{11}(0.6)^7$
  - b.  $\binom{18}{11}(0.4)^7(0.6)^{11}$
  - c.  $\binom{18}{7}(0.4)^7(0.6)^{11}$
  - d.  $\binom{18}{7}(0.4)^7(0.6)^{18}$
  - e.  $\binom{18}{7}(0.4)^{18}(0.6)^7$
2. Two variables,  $x$  and  $y$ , seem to be exponentially related. The natural logarithm of each  $y$  value is taken and the least-squares regression line of  $\ln(y)$  on  $x$  is determined to be  $\ln(y) = 3.2 + 0.42x$ . What is the predicted value of  $y$  when  $x = 7$ ?
  - a. 464.05
  - b. 1380384.27
  - c. 521.35
  - d. 6.14
  - e. 1096.63
3. You need to construct a large sample 94% confidence interval for a population mean. What is the upper critical value of  $z$  to be used in constructing this interval?
  - a. 0.9699
  - b. 1.96
  - c. 1.555
  - d. -1.88
  - e. 1.88

GO ON TO THE NEXT PAGE

4.



Which of the following best describes the shape of the histogram at the left.

- Approximately normal
  - Skewed left
  - Skewed right
  - Approximately normal with an outlier
  - Symmetric
5. The probability is 0.2 that a term selected at random from a normal distribution with mean 600 and standard deviation 15 will be above what number?
- 0.84
  - 603.80
  - 612.6
  - 587.4
  - 618.8
6. Which of the following are examples of continuous data?
- The speed your car goes
  - The number of outcomes of a binomial experiment
  - The average temperature in San Francisco
  - The wingspan of a bird
  - The jersey numbers of a football team
- I, III, and IV only
  - II and V only
  - I, III, and V only
  - II, III, and IV only
  - I, II, and IV only

Use the following computer output for a least-squares regression in Questions 7 and 8.

The regression equation is				
Predictor	Coef	St Dev	t ratio	P
Constant	22.94	11.79	1.95	0.088
x	-0.6442	0.5466	-1.18	—
s = 2.866		R-sq = 14.8%		R-sq(adj) = 4.1%

7. What is the equation of the least-squares regression line?

- $\hat{y} = -0.6442x + 22.94$
- $\hat{y} = 22.94 + 0.5466x$
- $\hat{y} = 22.94 + 2.866x$
- $\hat{y} = 22.94 - 0.6442x$
- $\hat{y} = -0.6442 + 0.5466x$

8. Given that the analysis is based on 10 datapoints, and using Table B (see Appendix), what is the  $P$ -value for the  $t$ -test of the hypothesis  $H_0: \beta = 0$  versus  $H_A: \beta \neq 0$ ?
- $0.02 < P < 0.03$
  - $0.20 < P < 0.30$
  - $0.01 < P < 0.05$
  - $0.15 < P < 0.20$
  - $0.10 < P < 0.15$
9. "A hypothesis test yields a  $P$ -value of 0.20." Which of the following best describes what is meant by this statement?
- The probability of getting a finding at least as extreme as obtained by chance alone if the null hypothesis is true is 0.20.
  - The probability of getting a finding as extreme as obtained by chance alone from repeated random sampling is 0.20.
  - The probability is 0.20 that our finding is significant.
  - The probability of getting this finding is 0.20.
  - The finding we got will occur less than 20% of the time in repeated trials of this hypothesis test.
10. A random sample of 25 men and a separate random sample of 25 women are selected to answer questions about attitudes toward abortion. The answers were categorized as "pro-life" or "pro-choice." Which of the following is the proper null hypothesis for this situation?
- The variables "gender" and "attitude toward abortion" are related.
  - The proportion of "pro-life" men is the same as the proportion of "pro-life" women.
  - The proportion of "pro-life" men is related to the proportion of "pro-life" women.
  - The proportion of "pro-choice" men is the same as the proportion of "pro-life" women.
  - The variables "gender" and "attitude toward abortion" are independent.
11. A sports talk show asks people to call in and give their opinion of the officiating in the local basketball team's most recent loss. What will most likely be the typical reaction?
- They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.
  - They would most likely call for the team to get some new players to replace the current ones.
  - The team probably wouldn't have lost if the officials had been doing their job.
  - Because the team had been foul-plagued all year, the callers would most likely support the officials.
  - They would support moving the team to another city.
12. A major polling organization wants to predict the outcome of an upcoming national election (in terms of the proportion of voters who will vote for each candidate). They intend to use a 95% confidence interval with margin of error of no more than 2.5%. What is the minimum sample size needed to accomplish this goal?
- 1536
  - 39
  - 1537
  - 40
  - 2653
13. A sample of size 35 is to be drawn from a large population. The sampling technique is such that every possible sample of size 35 that could be drawn from the population is equally likely. What name is given to this type of sample?
- Systematic sample
  - Cluster sample
  - Voluntary response sample
  - Random sample
  - Simple random sample

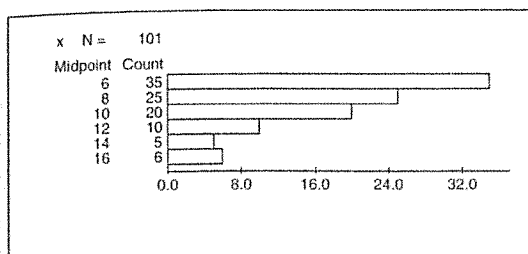
14. A teacher's union and a school district are negotiating salaries for the coming year. The teachers want more money, and the district, claiming, as always, budget constraints, wants to pay as little as possible. The district, like most, has a large number of moderately paid teachers and a few highly paid administrators. The salaries of all teachers and administrators are included in trying to figure out, on average, how much the professional staff currently earn. Which of the following would the teachers' union be most likely to quote during negotiations?
- The mean of all the salaries.
  - The mode of all the salaries.
  - The standard deviation of all the salaries.
  - The interquartile range of all the salaries.
  - The median of all the salaries.
15. Alfred and Ben don't know each other but are each considering asking the lovely Charlene to the school prom. The probability that at least one of them will ask her is 0.72. The probability that they both ask her is 0.18. The probability that Alfred asks her is 0.6. What is the probability that Ben asks Charlene to the prom?
- 0.78
  - 0.30
  - 0.24
  - 0.48
  - 0.54
16. A significance test of the hypothesis  $H_0: p = 0.3$  against the alternative  $H_A: p > 0.3$  found a value of  $\hat{p} = 0.35$  for a random sample of size 95. What is the  $P$ -value of this test?
- 1.06
  - 0.1446
  - 0.2275
  - 0.8554
  - 0.1535
17. Which of the following best describes the Central Limit Theorem?
- The mean of the sampling distribution of  $\bar{x}$  is the same as the mean of the population.
  - The standard deviation of the sampling distribution of  $\bar{x}$  is the same as the standard deviation of  $\bar{x}$  divided by the square root of the sample size.
  - If the sample size is large, the shape of the sampling distribution of  $\bar{x}$  is approximately normal.
- I only
  - I & II only
  - II only
  - III only
  - I, II, and III
18. If three fair coins are flipped,  $P(0 \text{ heads}) = 0.125$ ,  $P(\text{exactly 1 head}) = 0.375$ ,  $P(\text{exactly 2 heads}) = 0.375$ , and  $P(\text{exactly 3 heads}) = 0.125$ . The following results were obtained when three coins were flipped 64 times:

# Heads	Observed
0	10
1	28
2	22
3	4

What is the value of the  $X^2$  statistic used to test if the coins are behaving as expected, and how many degrees of freedom does the determination of the  $P$ -value depend on?

- 3.33, 3
- 3.33, 4
- 11.09, 3
- 3.33, 2
- 11.09, 4

19.



For the histogram pictured above, what is the class interval (boundaries) for the class that contains the median of the data?

- (5, 7)
- (9, 11)
- (11, 13)
- (15, 17)
- (7, 9)

20. Thirteen large animals were measured to help determine the relationship between their length and their weight. The natural logarithm of the weight of each animal was taken and a least-squares regression equation for predicting weight from length was determined. The computer output from the analysis is given below:

The regression equation is				
$\ln(\text{wt}) = 1.24 + 0.0365 \text{ length}$				
Predictor	Coef	St Dev	t ratio	P
Constant	1.2361	0.1378	8.97	0.000
Length	0.036502	0.001517	24.05	0.000
s = 0.1318		R-sq = 98.1%		R-sq(adj) = 98.0%

Give a 99% confidence interval for the slope of the regression line. Interpret this interval.

- (0.032, 0.041); the probability is 0.99 that the true slope of the regression line is between 0.032 and 0.041.
  - (0.032, 0.041); 99% of the time, the true slope will be between 0.032 and 0.041.
  - (0.032, 0.041); we are 99% confident that the true slope of the regression line is between 0.032 and 0.041.
  - (0.81, 1.66); we are 99% confident that the true slope of the regression line is between 0.032 and 0.041.
  - (0.81, 1.66); the probability is 0.99 that the true slope of the regression line is between 0.81 and 1.66.
21. What are the mean and standard deviation of a binomial experiment that occurs with probability of success 0.76 and is repeated 150 times?
- 114, 27.35
  - 100.5, 5.23
  - 114, 5.23
  - 100.5, 27.35
  - The mean is 114, but there is not enough information given to determine the standard deviation.
22. Which of the following is the primary difference between an experiment and an observational study?
- Experiments are only conducted on human subjects; observational studies can be conducted on nonhuman subjects.
  - In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.
  - Experiments must use randomized treatment and control groups; observational studies also use treatment and control groups, but they do not need to be randomized.
  - Experiments must be double-blind; observational studies do not need to be.
  - There is no substantive difference—they can both accomplish the same research goals.

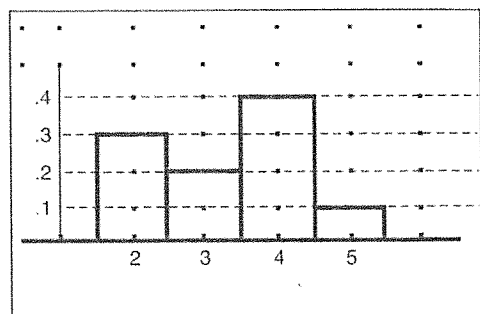
23. The regression analysis of question 20 indicated that “R-sq = 98.1%.” Which of the following is (are) true?

I. There is a strong positive linear relationship between the explanatory and response variables.  
 II. There is a strong negative linear relationship between the explanatory and response variables.  
 III. About 98% of the variation in the response variable can be explained by the regression on the explanatory variable.

- a. I and III only  
 b. I or II only  
 c. I or II (but not both) and III  
 d. II and III only  
 e. I, II, and III
24. A hypothesis test is set up so that  $P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) = 0.05$  and  $P(\text{failing to reject } H_0 \text{ when } H_0 \text{ is false}) = 0.26$ . What is the power of the test?
- a. 0.26  
 b. 0.05  
 c. 0.95  
 d. 0.21  
 e. 0.74
25. For the following observations collected while doing a chi-square test for independence between the two variables  $A$  and  $B$ , find the expected value of the cell marked with “XXXX.”

5	10(XXXX)	11
6	9	12
7	8	13

- a. 4.173  
 b. 9.00  
 c. 11.56  
 d. 8.667  
 e. 9.33
26. The following is a probability histogram for a discrete random variable  $X$ .



What is  $\mu_x$ ?

- a. 3.5  
 b. 4.0  
 c. 3.7  
 d. 3.3  
 e. 3.0

27. A psychologist believes that positive rewards for proper behavior is more effective than punishment for bad behavior in promoting good behavior in children. A scale of "proper behavior" is developed.  $\mu_1$  = the "proper behavior" rating for children receiving positive rewards, and  $\mu_2$  = the "proper behavior" rating for children receiving punishment. If  $H_0: \mu_1 - \mu_2 = 0$ , which of the following is the proper statement of  $H_A$ ?
- $H_A: \mu_1 - \mu_2 > 0$
  - $H_A: \mu_1 - \mu_2 < 0$
  - $H_A: \mu_1 - \mu_2 \neq 0$
  - Any of the above is an acceptable alternative to the given null.
  - There isn't enough information given in the problem for us to make a decision.
28. Estrella wants to become a paramedic and takes a screening exam. Scores on the exam have been approximately normally distributed over the years it has been given. The exam is normed with a mean of 80 and a standard deviation of 9. Only those who score in the top 15% on the test are invited back for further evaluation. Estrella received a 90 on the test. What was her percentile rank on the test, and did she qualify for further evaluation?
- 13.35; she didn't qualify.
  - 54.38; she didn't qualify.
  - 86.65; she qualified.
  - 84.38; she didn't qualify.
  - 88.69; she qualified.
29. Which of the following statements is (are) true?
- In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5.
  - In order to use  $\chi^2$  procedures, you must have at least 2 degrees of freedom.
  - In a  $4 \times 2$  two-way table, the number of degrees of freedom is 3.
- I only
  - I and III only
  - I and II only
  - III only
  - I, II, and III
30. When the point (15,2) is included, the slope of regression line ( $y = a + bx$ ) is  $b = -0.54$ . The correlation is  $r = -0.82$ . When the point is removed, the new slope is  $-1.04$  and the new correlation coefficient is  $-0.95$ . What name is given to a point whose removal has this kind of effect on statistical calculations?
- Outlier
  - Statistically significant point
  - Point of discontinuity
  - Unusual point
  - Influential point
31. A one-sided test of a hypothesis about a population mean, based on a sample of size 14, yields a  $P$ -value of 0.075. Using Table B (p. 360), which of the following best describes the range of  $t$  values that would have given this  $P$ -value?
- $1.345 < t < 1.761$
  - $1.356 < t < 1.782$
  - $1.771 < t < 2.160$
  - $1.350 < t < 1.771$
  - $1.761 < t < 2.145$

32. Use the following excerpt from a random digits table for assigning six people to treatment and control groups: 98110 35679 14520 51198 12116 98181 99120 75540 03412 25631  
The subjects are labeled: Arnold: 1; Betty: 2; Clive: 3; Doreen: 4; Ernie: 5; Florence: 6. The first three subjects randomly selected will be in the treatment group; the other three in the control group. Assuming you begin reading the table at the extreme left digit, which three subjects would be in the *control* group?
- Arnold, Clive, Ernest
  - Arnold, Betty, Florence
  - Betty, Clive, Doreen
  - Clive, Ernest, Florence
  - Betty, Doreen, Florence
33. A null hypothesis,  $H_0: \mu = \mu_0$  is to be tested against a two-sided hypothesis. A sample is taken,  $\bar{x}$  is determined and used as the basis for a  $C$ -level confidence interval (e.g.,  $C = 0.95$ ) for  $\mu$ . The researcher notes that  $\mu_0$  is not in the interval. Another researcher chooses to do a significance test for  $\mu$  using the same data. What significance level must the second researcher choose in order to guarantee getting the same conclusion about  $H_0: \mu = \mu_0$  (that is, reject or not reject) as the first researcher?
- $1 - C$
  - $C$
  - $\alpha$
  - $1 - \alpha$
  - $\alpha = 0.05$
34. Which of the following is *not* required in a binomial setting?
- Each trial is considered either a success or a failure.
  - Each trial is independent.
  - The value of the random variable of interest is the number of trials until the first success occurs.
  - There is a fixed number of trials.
  - Each trial succeeds or fails with the same probability.
35.  $X$  and  $Y$  are independent random variables with  $\mu_X = 3.5$ ,  $\mu_Y = 2.7$ ,  $\sigma_X = 0.8$ , and  $\sigma_Y = 0.65$ . What are  $\mu_{X+Y}$  and  $\sigma_{X+Y}$ ?
- $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y} = 1.03$
  - $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y} = 1.0625$
  - $\mu_{X+Y} = 3.1$ ,  $\sigma_{X+Y} = 0.725$
  - $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y} = 1.45$
  - $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y}$  cannot be determined from the information given.
36. A researcher is hoping to find a predictive linear relationship between the explanatory and response variables in her study. Accordingly, as part of her analysis she plans to generate a 95% confidence interval for the slope of the regression line for the two variables. The interval is determined to be (0.45, 0.80). Which of the following is (are) true?
- She has good evidence of a predictive linear relationship between the variables.
  - It is likely that there is a non-zero correlation ( $r$ ) between the two variables.
  - It is likely that the true slope of the regression line is 0.
- I and II only
  - I and III only
  - II and III only
  - I only
  - II only



37. In the casino game of roulette, there are 38 slots for a ball to drop into when it is rolled around the rim of a revolving wheel: 18 red, 18 black, and 2 green. What is the probability that the first time a ball drops into the red slot is on the 8th trial (in other words, suppose you are betting on red every time—what is the probability of losing 7 straight times before you win the first time?)?

- a. 0.0278
- b. 0.0112
- c. 0.0053
- d. 0.0101
- e. 0.0039

38. You are developing a new strain of strawberries (say, Type X) and are interested in its sweetness as compared to another strain (say, Type Y). You have four plots of land, call them A, B, C, and D, which are roughly four squares in one large plot for your study (see the figure below). A river runs alongside of plots C and D. Because you are worried that the river might influence the sweetness of the berries, you randomly plant type X in either A or B (and Y in the other) and randomly plant type X in either C or D (and Y in the other). Which of the following terms best describes this design?

A	C
B	D



- a. A completely randomized design
- b. A randomized study
- c. A randomized observational study
- d. A block design, controlling for the strain of strawberry
- e. A block design, controlling for the effects of the river

39. Grumpy got 38 on the first quiz of the quarter. The class average on the first quiz was 42 with a standard deviation of 5. Dopey, who was absent when the first quiz was given, got 40 on the second quiz. The class average on the second quiz was 45 with a standard deviation of 6.1. Grumpy was absent for the second quiz. After the second quiz, Dopey told Grumpy that he was doing better in the class because they had each taken one quiz, and he had gotten the higher score. Did he really do better? Explain.

- a. Yes.  $z_{\text{Dopey}}$  is more negative than  $z_{\text{Grumpy}}$ .
- b. Yes.  $z_{\text{Dopey}}$  is less negative than  $z_{\text{Grumpy}}$ .
- c. No.  $z_{\text{Dopey}}$  is more negative than  $z_{\text{Grumpy}}$ .
- d. Yes.  $z_{\text{Dopey}}$  is more negative than  $z_{\text{Grumpy}}$ .
- e. No.  $z_{\text{Dopey}}$  is less negative than  $z_{\text{Grumpy}}$ .

40. A random sample size of 45 is obtained for the purpose of testing the hypothesis  $H_0: p = 0.80$ . The sample proportion is determined to be  $\hat{p} = 0.75$ . What is the value of the standard error of  $\hat{p}$  for this test?

- a. 0.0042
- b. 0.0596
- c. 0.0036
- d. 0.0645
- e. 0.0055

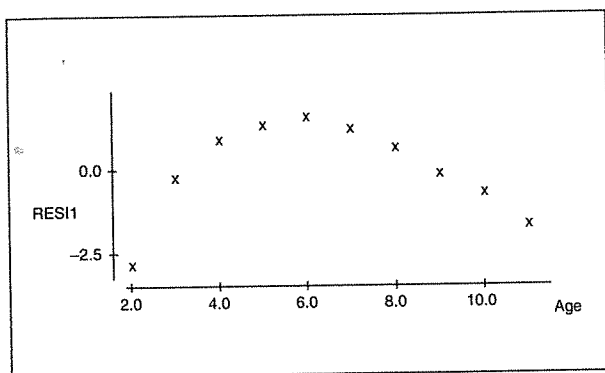
## SECTION II—PART A, QUESTIONS 1–5

Spend about 65 minutes on this part of the exam. Percentage of Section II grade—75.

**Directions:** Show all of 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.

1. The ages (in years) and heights (in cm) of 10 girls, ages 2 through 11, were recorded. Part of the regression output and the residual plot for the data are given below.

The regression equation is				
Predictor	Coef	St Dev	t ratio	P
Constant	76.641	1.188	64.52	0.000
Age	6.3661	0.1672	38.08	0.000
$s = 1.518$		R-sq = 99.5%		R-sq(adj) = 99.4%



- What is the equation of the least-squares regression line for predicting height from age?
  - Interpret the slope of the regression line in the context of the problem.
  - Suppose you wanted to predict the height of a girl 5.5 years of age. Would the prediction made by the regression equation you gave in (a) be too small, too large, or is there not enough information to tell?
2. You want to determine whether a greater proportion of men or women purchase vanilla latte (regular or decaf). To collect data, you hire a person to stand inside the local Scorebucks for 2 hours one morning and tally the number of men and women who purchase the vanilla latte as well as the total number of men and women customers: 63% of the women and 59% of the men purchase a vanilla latte.
- Is this an experiment or an observational study? Explain.
  - Based on the data collected, you write a short article for the local newspaper claiming that a greater proportion of women than men prefer vanilla latte as their designer coffee of choice. A student in the local high school AP Statistics class writes a letter to the editor criticizing your study. What might the student have pointed out?
  - Suppose you wanted to conduct a study less open to criticism. How might you redo the study?
3. Sophia is a nervous basketball player. Over the years she has had a 40% chance of making the first shot she takes in a game. If she makes her first shot, her confidence goes way up, and the probability of her making the second shot she takes rises to 70%. But if she misses her first shot, the probability of her making the second shot she takes doesn't change—it's still 40%.
- What is the probability that Sophia makes her second shot?
  - If Sophia does make her second shot, what is the probability that she missed her first shot?

4. A random sample of 72 seniors taken 3 weeks before the selection of the school Homecoming Queen identified 60 seniors who planned to vote for Buffy for queen. Unfortunately, Buffy said some rather cat things about some of her opponents, and it got into the school newspaper. A second random sample of 80 seniors taken shortly after the article appeared showed that 56 planned to vote for Buffy. Does this indicate a serious drop in support for Buffy? Use good statistical reasoning to support your answer.
5. Some researchers believe that education influences IQ. One researcher specifically believes that the more education a person has, the higher, on average, will be his or her IQ. The researcher sets out to investigate this belief by obtaining eight pairs of identical twins reared apart. He identifies the better educated twin as Twin A and the other twin as Twin B for each pair. The data for the study are given in the table below. Do the data give good statistical evidence, at the 0.05 level of significance, that the twin with more education is likely to have the higher IQ? Give good statistical evidence to support your answer.

Pair	1	2	3	4	5	6	7	8
Twin A	103	110	90	97	92	107	115	102
Twin B	97	103	91	93	92	105	111	103

### SECTION II—PART B, QUESTION 6

Spend about 25 minutes on this part of the exam. Percentage of Section II grade—25.

**Directions:** Show all of 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 paint manufacturer claims that the average drying time for its best-selling paint is 2 hours. A random sample of drying times for 20 randomly selected cans of paint are obtained to test the manufacturer's claim. The drying times observed, in minutes, were: 123, 118, 115, 121, 130, 127, 112, 120, 116, 136, 131, 128, 139, 110, 133, 122, 133, 119, 135, 109.
- Obtain a 95% confidence interval for the true mean drying time of the paint.
  - Interpret the confidence interval obtained in part (a) in the context of the problem.
  - Suppose, instead, a significance test, at the 0.05 level, of the hypothesis  $H_0: \mu = 120$  was conducted against the alternative  $H_A: \mu \neq 120$ . What is the  $P$ -value of the test?
  - Are the answers you got in part (a) and part (c) consistent? Explain.
  - At the 0.05 level, would your conclusion about the mean drying time have been different if the alternative hypothesis had been  $H_A: \mu > 120$ ? Explain.

END OF DIAGNOSTIC EXAM