

# Final Exam Review

AP Statistics  
Mr Murphy

## Multiple Choice (1 pt. each)

16. A 90 percent confidence interval is to be created to estimate the proportion of television viewers in a certain area who favor moving the broadcast of the late weeknight news to an hour earlier than it is currently. Initially, the confidence interval will be created using a simple random sample of 9,000 viewers in the area. Assuming that the sample proportion does not change, what would be the relationship between the width of the original confidence interval and the width of a second 90 percent confidence interval that is created based on a sample of only 1,000 viewers in the area?

- (a) The second confidence interval would be 9 times as wide as the original confidence interval.
- (b) The second confidence interval would be 3 times as wide as the original confidence interval.
- (c) The width of the second confidence interval would be equal to the width of the original confidence interval.
- (d) The second confidence interval would be  $\frac{1}{3}$  as wide as the original confidence interval.
- (e) The second confidence interval would be  $\frac{1}{9}$  as wide as the original confidence interval.

17. A local real estate magazine used the median instead of the mean when it reported the SAT score of the average student who attends Groveland High School. A graphical display of SAT scores of students who attend Groveland High School indicated that the data were strongly skewed to the right. Which of the following explains why, in this situation, the median is a more accurate indicator of the SAT score of the average student than the mean is?

- (a) The mean is affected by the skewness, whereas the median is not.
- (b) The median is always the preferred statistic.
- (c) The mean will be less than the median when the data are strongly skewed to the right.
- (d) The mean should be used only when data are strongly skewed to the left.
- (e) The median is equal to one-half the sum of the maximum and minimum SAT scores at Groveland High.

18. The director of a fitness center wants to examine the effects of two exercise classes (spinning and aerobics) on body fat percentage. A six-week spinning class and a six-week aerobics class are offered at the same time and on the same days, so that a person can enroll in only one of them. A new class of each is about to begin, and each class has 25 people in it. Ten people are randomly selected from each class. Each person's body fat percentage is measured at the beginning and again at the end of the six-week class. Using the change in body fat percentage as the response variable and conducting a test at the  $\alpha = 0.01$  level, the director determines that there is a significant difference between the treatment means. Which of the following is a confounding variable in the study?

- (a) The director's choice of spinning and aerobics classes as the types for use in the study
- (b) The random sample of 10 people from each class
- (c) The participants' choice of which class to take
- (d) The use of body fat percentage as the measure of effectiveness of the treatment
- (e) The fact that both classes were conducted three times each week at the same time of day and for the same amount of time each day

19. Raoul performed an experiment using 16 windup rubber band single-propeller airplanes. He wound up the propeller a different number of times and recorded the amount of time (in seconds) that the airplane flew for each number of rotations that the propeller was wound. A regression analysis was performed and the partial computer output is given below.

The regression equation is				
Time = 0.924 + 0.0462 Rotations				
Predictor	Coef	SE Coef	T	P
Constant	0.9241	0.6413	1.44	0.172
Rotation	0.04625	0.01565	2.96	0.10
S = 0.5426    R-Sq = 38.4%    R-Sq(adj) = 34.0%				

Which of the following is a 95 percent confidence interval for the slope of the regression line that relates the number of rotations the rubber band is wound and the plane's flight time?

- (a)  $0.0462 \pm (2.145)(0.01565)$
- (b)  $0.924 \pm (2.145)(0.6413)$
- (c)  $0.0462 \pm (2.96)(0.01565)$
- (d)  $0.924 \pm (2.96)(0.6413)$
- (e)  $0.0462 \pm (2.131)(0.01565)$

20. A recent study was conducted to investigate the duration of time required to complete a certain manual dexterity task. The reported mean was 10.2 seconds with a standard deviation of 16.0 seconds. Suppose the reported values are the true mean and standard deviation for the population of subjects in the study. If a random sample of 144 subjects is selected from the population, what is the approximate probability that the mean of the sample will be more than 11.0 seconds?

- (a) 0.1151
- (b) 0.2743
- (c) 0.7257
- (d) 0.8849
- (e) Based on the values of the true mean and true standard deviation, it can be concluded that the population distribution is not normal and therefore the probability cannot be calculated.

21. To determine whether employees at Site X have higher salaries, on average, than employees at Site Y of the same company do, independent random samples of salaries were obtained for the two groups. The data are summarized below. Based on the data, which of the following statements is true?

	Site X	Site Y
Mean	\$61,234	\$60,529
Standard Deviation	\$4,352	\$3,456
<i>n</i>	235	183

- (a) At the 5% significance level, employees at Site Y have a significantly higher mean salary than employees at Site X do.
- (b) At the 1% significance level, employees at Site Y have a significantly higher mean salary than employees at Site X do.
- (c) At the 5% significance level, employees at Site X have a significantly higher mean salary than employees at Site Y do.
- (d) At the 1% significance level, employees at Site X have a significantly higher mean salary than employees at Site Y do.
- (e) At the 10% significance level, there is no significant difference in salaries between the employees at the two sites.

22. On their birthdays, employees at a large company are permitted to take a 60-minute lunch break instead of the usual 30 minutes. Data were obtained from 10 randomly selected company employees on the amount of time that each actually took for lunch on his or her birthday. The company wishes to investigate whether these data provide convincing evidence that the mean time is greater than 60 minutes. Of the following, which information would NOT be expected to be a part of the process of correctly conducting a hypothesis test to investigate the question, at the 0.05 level of significance?

- (a) Being willing to assume that the distribution of actual birthday lunch times for all employees at the company is approximately normal
- (b) Knowing that there are no outliers in the data as indicated by the normal probability plot and boxplot
- (c) Using a  $t$ -statistic to carry out the test
- (d) Using 9 for the number of degrees of freedom
- (e) Given that the  $p$ -value is greater than 0.05, rejecting the null hypothesis and concluding that the mean time was not greater than 60 minutes

23. The statistics below provide a summary of the distribution of heights, in inches, for a simple random sample of 200 young children.

Mean: 46 inches

Median: 45 inches

Standard Deviation: 3 inches

First Quartile: 43 inches

Third Quartile: 48 inches

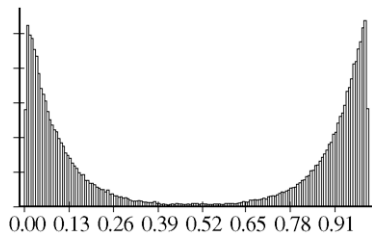
About 100 children in the sample have heights that are

- (a) less than 43 inches
- (b) less than 48 inches
- (c) between 43 and 48 inches
- (d) between 40 and 52 inches
- (e) more than 46 inches

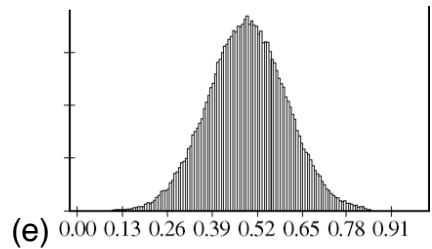
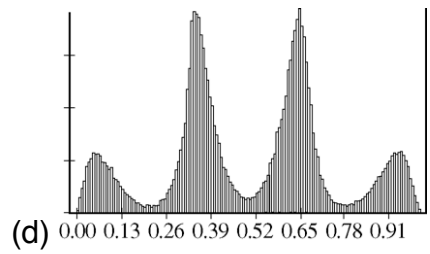
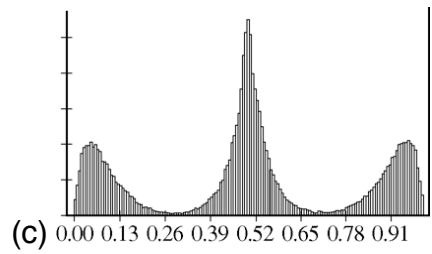
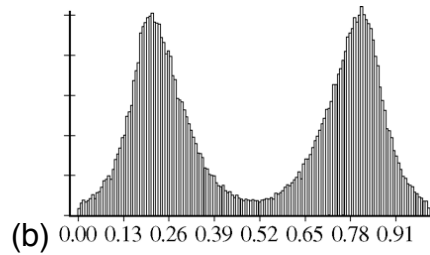
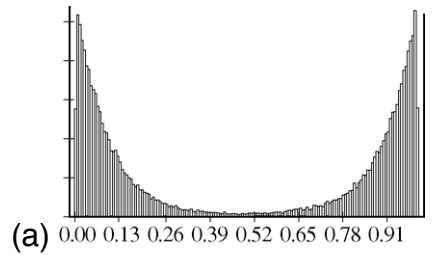
24. In which of the following situations would it be most difficult to use a census?

- (a) To determine what proportion of licensed bicycles on a university campus have lights.
- (b) To determine what proportion of students in a high school support wearing uniforms
- (c) To determine what proportion of registered students enrolled in a college are employed more than 20 hours each week
- (d) To determine what proportion of single-family dwellings in a small town have two-car garages
- (e) To determine what proportion of fish in Lake Michigan are bass

25. The histogram below represents data obtained after the census of an entire population was conducted.



The sampling distribution of the sample mean based on samples of size 2 for the population was simulated, and a histogram of the results was produced. Which of the following histograms is most likely the histogram of that sampling distribution?



26. The stemplot below shows the yearly earnings per share of stock for two different companies over a six-year period.

Company A		Company B
	0	58, 75, 96, 98
92, 91, 90, 82, 78, 43, 38, 26	1	01, 10, 17, 21, 43, 43, 53, 65, 73
49, 47, 44, 00	2	09, 27, 29
73, 27, 05, 02	3	

$3|47 = \$3.47$

Which of the following statements are true?

- (a) The median of the earnings of Company A is less than the median of the earnings of Company B.
- (b) The range of the earnings of Company A is less than the range of the earnings of Company B.
- (c) The third quartile of Company A is smaller than the third quartile of Company B.
- (d) The mean of the earnings of Company A is greater than the mean of the earnings of Company B.
- (e) The interquartile range of Company A is twice the interquartile range of Company B.

27. A television news editor would like to know how local registered voters would respond to the question, "Are you in favor of the school bond measure that will be voted on in an upcoming special election?" A television survey is conducted during a break in the evening news by listing two phone numbers side by side on the screen, one for viewers to call if they approve the bond measure, and the other to call if they disapprove. This survey method could produce biased results for a number of reasons. Which one of the following is the most obvious reason?

- (a) It uses a stratified sample rather than a simple random sample.
- (b) People who feel strongly about the issue are more likely to respond.
- (c) Viewers should be told about the issues before the survey is conducted.
- (d) Some registered voters who call might not vote in the election.
- (e) The wording of the question is biased.

28. A high school physics teacher was conducting an experiment with his class on the length of time it will take a marble to roll down a sloped chute. The class ran repeated trials in order to determine the relationship between the length, in centimeters, of the sloped chute and the time, in seconds, for the marble to roll down the chute. A linear relationship was observed and the correlation coefficient was 0.964. After discussing their results, the teacher instructed the students to convert all of the length measurements to meters but leave time in seconds. What effect will this have on the correlation of the two variables?

- (a) Because the standard deviation of the lengths in meters will be one hundredth of the standard deviation of the lengths in centimeters, the correlation will decrease by one hundredth to 0.954.
- (b) Because the standard deviation of the lengths in meters will be one hundredth of the standard deviation of the lengths in centimeters, the correlation will decrease proportionally to 0.00964.
- (c) Because changing from centimeters to meters does not affect the value of the correlation, the correlation will remain 0.964.
- (d) Because only the length measurements have been changed, the correlation will decrease substantially.
- (e) Because meters are a much more common measurement for length in determining speed, the linear relationship of the data will be stronger and thus the correlation will increase substantially.

29. The table below show the sample size, the mean, and the median for two samples of measurements. What is the median for the combined sample of 47 measurements?

	<i>n</i>	MEAN	MEDIAN
SAMPLE I	21	42.6	45.0
SAMPLE II	26	49.2	48.5

- (a)  $\frac{42.6 + 49.2}{2}$
- (b)  $\frac{45.0 + 48.5}{2}$
- (c)  $\frac{21(42.6) + 26(49.2)}{47}$
- (d)  $\frac{21(45.0) + 26(48.5)}{47}$
- (e) It cannot be determined from the information given.

30. A researcher wishes to test a new drug developed to treat hypertension (high blood pressure). A group of 40 hypertensive men and 60 hypertensive women is to be used. The experiment randomly assigns 20 of the men and 30 of the women to the placebo and assigns the rest to the treatment. The major reason for separate assignment for men and women is that

- (a) it is a large study with 100 subjects
- (b) the new drug may affect men and women differently
- (c) the new drug may affect hypertensive and non-hypertensive people differently
- (d) this design uses matched pairs to detect the new-drug effect
- (e) there must be an equal number of subjects in both the placebo group and the treatment group

31. A delivery service places packages into large containers before flying them across the country. These filled containers vary greatly in their weight. Suppose the delivery service's airplanes always transport two such containers on each flight. The two containers are chosen so their combined weight is close to, but does not exceed, a specified weight limit. A random sample of flights with these containers is taken, and the weight of each of the two containers on each selected flight is recorded. The weights of the two containers on the same flight

- (a) will have a correlation of 0
- (b) will have a negative correlation
- (c) will have a positive correlation that is less than 1
- (d) will have a correlation of 1
- (e) cannot be determined from the information given

32. Which of the following is NOT a characteristic of stratified random sampling?

- (a) Random sampling is part of the sampling procedure.
- (b) The population is divided into groups of units that are similar on some characteristic.
- (c) The strata are based on facts known before the sample is selected.
- (d) Each individual unit in the population belongs to one and only one of the strata.
- (e) Every possible subset of the population, of the desired sample size, has an equal chance of being selected.

33. Random variable  $X$  is normally distributed with mean 10 and standard deviation 3, and random variable  $Y$  is normally distributed with mean 9 and standard deviation 4. If  $X$  and  $Y$  are independent, which of the following describes the distribution of  $Y - X$  ?

- (a) Normal with mean 1 and standard deviation -1
- (b) Normal with mean -1 and standard deviation 1
- (c) Normal with mean -1 and standard deviation 5
- (d) Normal with mean 1 and standard deviation 7
- (e) Normal with mean -1 and standard deviation 7



34. The table below shows the height, in inches, and the arm span, in inches, for 10 randomly selected high school students. Which of the following significance tests should be used to determine whether a linear relationship exists between height and arm span, provided the assumptions of the test are met?

Student	1	2	3	4	5	6	7	8	9	10
Height	65	72	64	68	65	70	61	73	69	70
Arm Span	67	71	60	69	60	65	58	74	70	67

- (a) Two sample z-test
- (b) Two sample  $t$ -test
- (c) Chi-square test of independence
- (d) Chi-square goodness of fit test
- (e)  $t$ -test for slope of the regression line

35. An independent research firm conducted a study of 100 randomly selected children who were participating in a program advertised to improve mathematical skills. The results showed no statistically significant improvement in mathematics skills, using  $\alpha = 0.05$ . The program sponsors complained that the study had insufficient statistical power. Assuming that the program is effective, which of the following would be an appropriate method for increasing power in this context?

- (a) Use a two-sided test instead of a one-sided test.
- (b) Use a one-sided test instead of a two-sided test.
- (c) Use  $\alpha = 0.01$  instead of  $\alpha = 0.05$ .
- (d) Decrease the sample size to 50 children
- (e) Increase the sample size to 200 children.

36. An experiment has three mutually exclusive outcomes,  $A$ ,  $B$ , and  $C$ . If  $P(A) = 0.12$ ,  $P(B) = 0.61$  and  $P(C) = 0.27$ , which of the following must be true?

- I.  $A$  and  $C$  are independent
- II.  $P(A \text{ and } B) = 0$
- III.  $P(B \text{ or } C) = P(B) + P(C)$

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

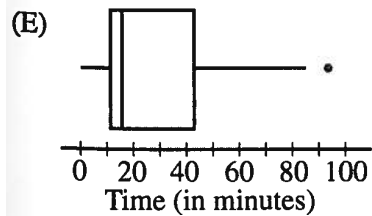
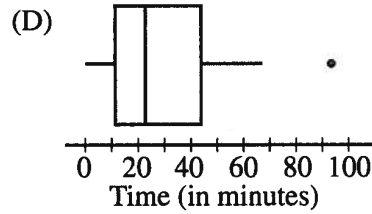
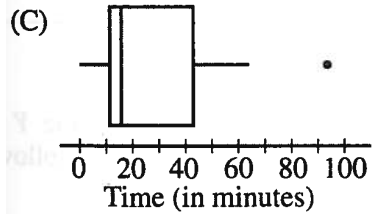
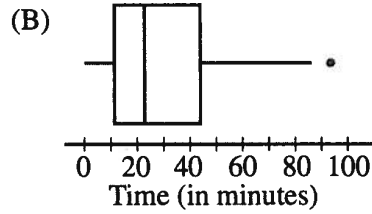
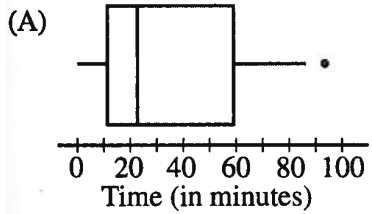
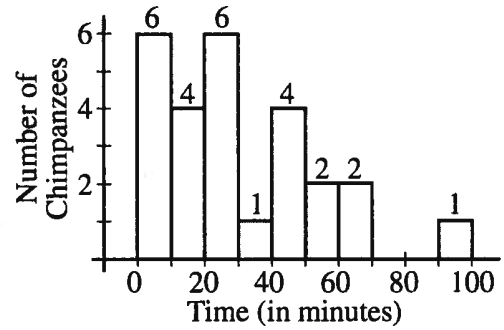
37. A planning board in Elm County is interested in estimating the proportion of its residents that are in favor of offering incentives to high-tech industries to build plants in that county. A random sample of Elm County residents was selected. All of the selected residents were asked, "Are you in favor of offering incentives to high-tech industries to build plants in your county?" A 95 percent confidence interval for the proportion of residents in favor of offering incentives was calculated to be  $0.54 \pm 0.05$ . Which of the following statements is correct?

- (a) At the 95% confidence level, the estimate of 0.54 is within 0.05 of the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county.
- (b) At the 95% confidence level, the majority of residents are in favor of offering incentives to high-tech industries to build plants in the county.
- (c) In repeated sampling, 95% of the sample proportions will fall in the interval (0.49, 0.59).
- (d) In repeated sampling, the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county will fall in the interval (0.49, 0.59).
- (e) In repeated sampling, the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county will be equal to 0.54.

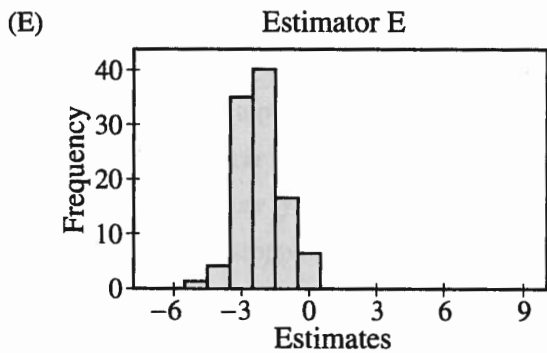
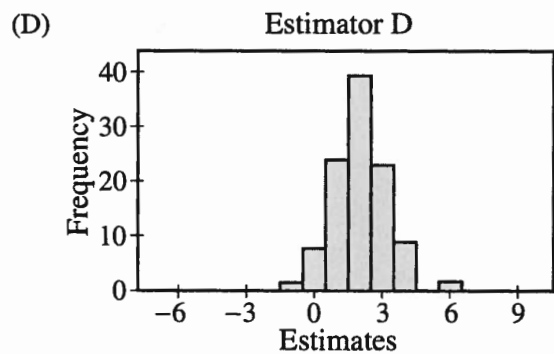
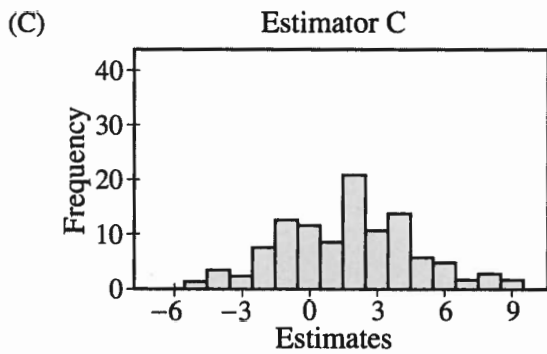
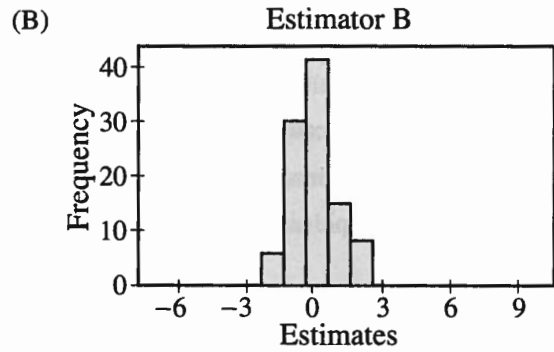
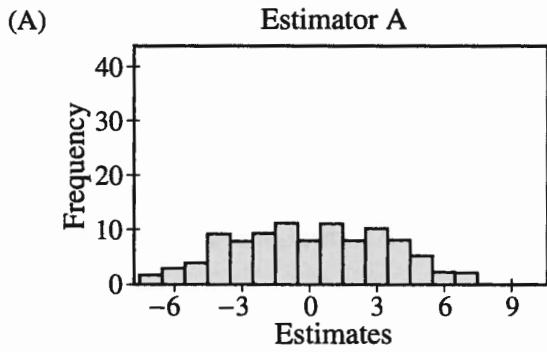
38. The histogram below displays the times, in minutes, needed for each chimpanzee in a sample of 26 to complete a simple navigation task.

It was determined that the largest observation, 93, is an outlier since  $Q_3 + 1.5(Q_3 - Q_1) = 87.125$ .

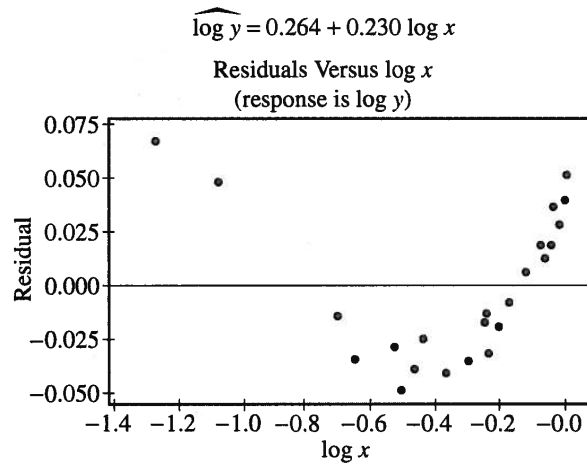
Which of the following boxplots could represent the information in the histogram?



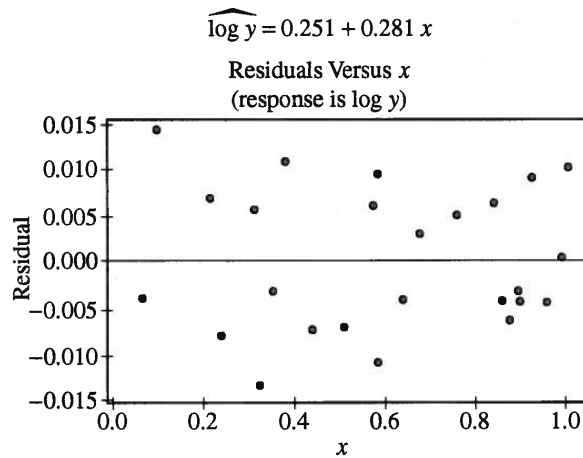
39. Five estimators for a parameter are being evaluated. The true value of the parameter is 0. Simulations of 100 random samples, each of size  $n$ , are drawn from the population. For each simulated sample, the five estimates are computed. The histograms below display the simulated sampling distribution for the five estimators. Which simulated sampling distribution is associated with the best estimator for this parameter?



40. Two variables,  $x$  and  $y$ , were measured for a random sample of 25 subjects, and two separate regression models were fit to the data. Least squares estimation of the parameters in Model A yielded the following equation and residual plot.



Least squares estimation of the parameters in Model B yielded the following equation and residual plot.



Which of the following conclusions is correct?

- (a) Model A is appropriate, since the relationship between  $x$  and  $y$  is linear.
- (b) Model B is appropriate, since the relationship between  $x$  and  $y$  is linear.
- (c) Model A is appropriate, since the relationship between  $\log x$  and  $\log y$  is linear.
- (d) Model A is appropriate, since the relationship between  $\log x$  and  $y$  is linear.
- (e) Model B is appropriate, since the relationship between  $x$  and  $\log y$  is linear.