Chapter 12 AP Statistics Practice Test

Printed Page 796

[Notes/Highlighting]

Section I: Multiple Choice *Select the best answer for each question.*

T12.1. Which of the following is *not* one of the conditions that must be satisfied in order to perform inference about the slope of a least-squares regression line?

(a) For each value of x, the population of y-values is Normally distributed.

(b) The standard deviation σ of the population of y-values corresponding to a particular value of x is always the same, regardless of the specific value of x.

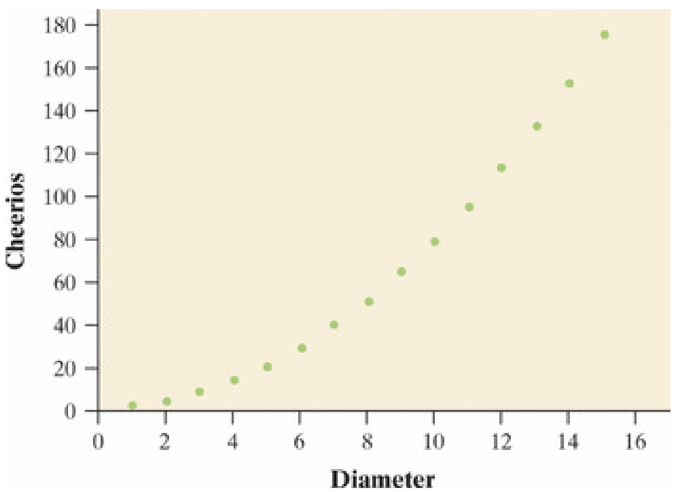
(c) The sample size—that is, the number of paired observations (x, y) — exceeds 30.

(d) There exists a straight line $y = \alpha + \beta x$ such that, for each value of x, the mean μ_y of the corresponding population of y-values lies on that straight line.

(e) The data come from a random sample or a randomized experiment.

Show Answer

T12.2. Students in a statistics class drew circles of varying diameters and counted how many Cheerios could be placed in the circle. The scatterplot shows the results.



The students want to determine an appropriate equation for the relationship between diameter and the number of Cheerios. The students decide to transform the data to make it appear more linear before computing a least-squares regression line. Which of the following single transformations would be reasonable for them to try?

I. Take the square root of the number of Cheerios.

II. Cube the number of Cheerios.

III. Take the log of the number of Cheerios.

IV. Take the log of the diameter.

(a) I and II

(b) I and III

(c) II and III

(d) II and IV

(e) I and IV

Show Answer

T12.3. Inference about the slope β of a least-squares regression line is based on which of the following distributions?

(a) The t distribution with $n - 1$ degrees of freedom

(b) The standard Normal distribution

(c) The chi-square distribution with $n - 1$ degrees of freedom

(d) The t distribution with $n - 2$ degrees of freedom

(e) The Normal distribution with mean μ and standard deviation σ

Show Answer

Exercises T12.4 through T12.8 refer to the following setting. An old saying in golf is "You drive for show and you putt for dough." The point is that good putting is more important than long driving for shooting low scores and hence winning money. To see if this is the case, data from a random sample of 69 of the nearly 1000 players on the PGA Tour's world money list are examined. The average number of putts per hole and the player's total winnings for the previous season are recorded. A least-squares regression line was fitted to the data. The following results were obtained from statistical software.

Predictor	Coef	SE Coef	T	P
Constant	7897179	3023782	6.86	0.000
Avg. Putts	−4139198	1698371	****	****

S = 281777 R-Sq = 8.1% R-Sq(adj) = 7.8%

T12.4. The correlation between total winnings and average number of putts per hole for these players is

(a) −0.285.

(b) −0.081.

(c) −0.007.

(d) 0.081.

(e) 0.285.

Show Answer

T12.5. Suppose that the researchers test the hypotheses $H_0 : \beta = 0$, $H_a : \beta < 0$. The value of the t statistic for this test is

(a) 2.61.

(b) 2.44.

(c) 0.081.

(d) −2.44.

(e) −20.24.

Show Answer

T12.6. The P-value for the test in Question T12.5 is 0.0087. A correct interpretation of this result is that

(a) the probability that there is no linear relationship between average number of putts per hole and total winnings for these 69 players is 0.0087.

(b) the probability that there is no linear relationship between average number of putts per hole and total winnings for all players on the PGA Tour's world money list is 0.0087.

(c) if there is no linear relationship between average number of putts per hole and total winnings for the players in the sample, the probability of getting a random sample of 69 players that yields a least-squares regression line with a slope of −4139198 or less is 0.0087.

(d) if there is no linear relationship between average number of putts per hole and total winnings for the players on the PGA Tour's world money list, the probability of getting a random sample of 69 players that yields a least-squares regression line with a slope of −4139198 or less is 0.0087.

(e) the probability of making a Type II error is 0.0087.

Show Answer

T12.7. A 95% confidence interval for the slope β of the population regression line is

(a) 7,897,179 ± 3,023,782.

(b) 7,897,179 ± 6,047,564.

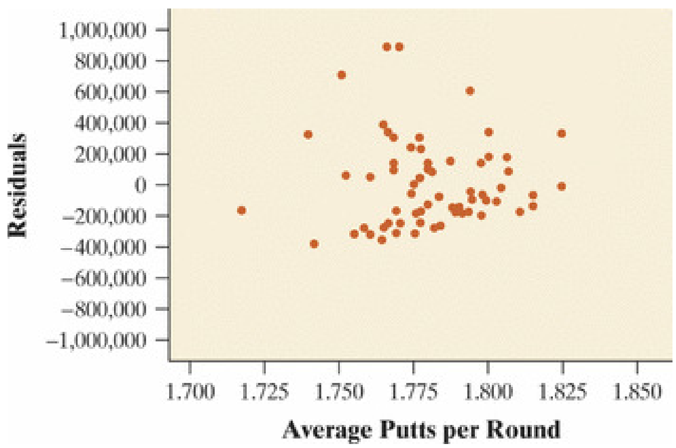
(c) −4,139,198 ± 1,698,371.

(d) −4,139,198 ± 3,328,807.

(e) −4,139,198 ± 3,396,742.

Show Answer

T12.8. A residual plot from the least-squares regression is shown below. Which of the following statements is supported by the graph?



(a) The residual plot contains dramatic evidence that the standard deviation of the response about the population regression line increases as the average number of putts per round increases.

(b) The sum of the residuals is not 0. Obviously, there is a major error present.

(c) Using the regression line to predict a player's total winnings from his average number of putts almost always results in errors of less than $200,000.

(d) For two players, the regression line underpredicts their total winnings by more than $800,000.

(e) The residual plot reveals a strong positive correlation between average putts per round and prediction errors from the least-squares line for these players.

Show Answer

T12.9. Which of the following would provide evidence that a power law model of the form $y = ax^b$, where $b \neq 0$ and $b \neq 1$, describes the relationship between a response variable y and an explanatory variable x?

(a) A scatterplot of y versus x looks approximately linear.

(b) A scatterplot of $\ln y$ versus x looks approximately linear.

(c) A scatterplot of y versus $\ln x$ looks approximately linear.

(d) A scatterplot of $\ln y$ versus $\ln x$ looks approximately linear.

(e) None of these

Show Answer

T12.10. We record data on the population of a particular country from 1960 to 2010. A scatterplot reveals a clear curved relationship between population and year. However, a different scatterplot reveals a strong linear relationship between the logarithm (base 10) of the population and the year. The least-squares regression line for the transformed data is

$$\widehat{\log(\text{population})} = -13.5 + 0.01(\text{year})$$

Based on this equation, the population of the country in the year 2020 should be about

(a) 6.7.

(b) 812.

(c) 5,000,000.

(d) 6,700,000.

(e) 8,120,000.

Show Answer

Section II: Free Response *Show all 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 and completeness of your results and explanations.*

T12.11. Growth hormones are often used to increase the weight gain of chickens. In an experiment using 15 chickens, five different doses of growth hormone (0, 0.2, 0.4, 0.8, and 1.0 milligrams) were injected into chickens (3 chickens were randomly assigned to each dose), and the subsequent weight gain (in ounces) was recorded. A researcher plots the data and finds that a linear relationship appears to hold. Computer output from a least-squares regression analysis for these data is shown below.

Predictor	Coef	SE Coef	T	P
Constant	4.5459	0.6166	7.37	<0.0001
Dose	4.8323	1.0164	4.75	0.0004

S = 3.135 R-Sq = 38.4% R-Sq(adj) = 37.7%

(a) What is the equation of the least-squares regression line for these data? Define any variables you use.

(b) Interpret each of the following in context:

(i) The slope

(ii) The *y* intercept

(iii) *s*

(iv) The standard error of the slope

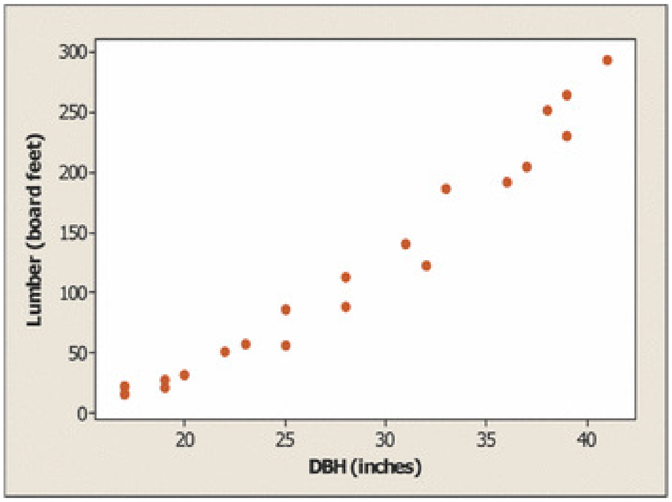
(v) r^2

(c) Assume that the conditions for performing inference about the slope β of the true regression line are met. Do the data provide convincing evidence of a linear relationship between dose and weight gain? Carry out a significance test at the $\alpha = 0.05$ level.

(d) Construct and interpret a 95% confidence interval for the slope parameter.

Show Answer

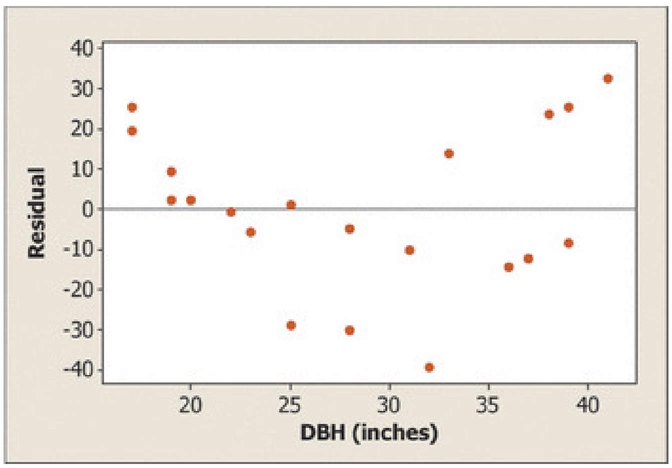
T12.12. Foresters are interested in predicting the amount of usable lumber they can harvest from various tree species. They collect data on the diameter at breast height (DBH) in inches and the yield in board feet of a random sample of 20 Ponderosa pine trees that have been harvested. (Note that a board foot is defined as a piece of lumber 12 inches by 12 inches by 1 inch.) A scatterplot of the data is shown below.



(a) Some computer output and a residual plot from a least-squares regression on these data appear below. Explain why a linear model may not be appropriate in this case.

Predictor	Coef	SE Coef	T	P
Constant	−191.12	16.98	−11.25	0.000
DBH (inches)	11.0413	0.5752	19.19	0.000

S = 20.3290 R-Sq = 95.3% R-Sq(adj) = 95.1%

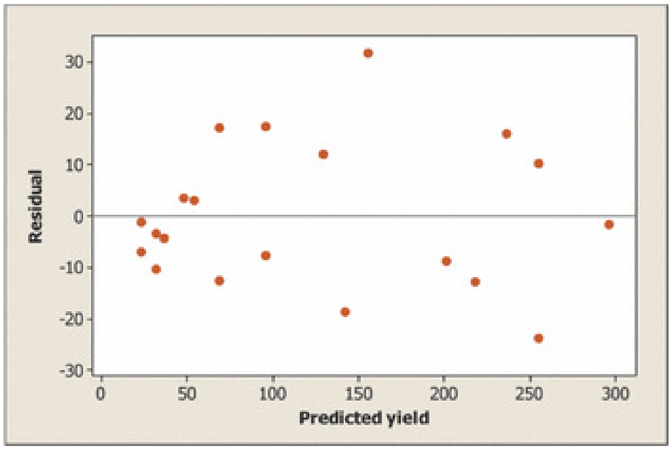


The foresters are considering two possible transformations of the original data: (1) cubing the diameter values or (2) taking the natural logarithm of the yield measurements. After transforming the data, a least-squares regression analysis is performed. Some computer output and a residual plot for each of the two possible regression models follow.

Option 1: Cubing the diameter values

Predictor	Coef	SE Coef	T	P
Constant	2.078	5.444	0.38	0.707
DBH^3	0.0042597	0.0001549	27.50	0.000

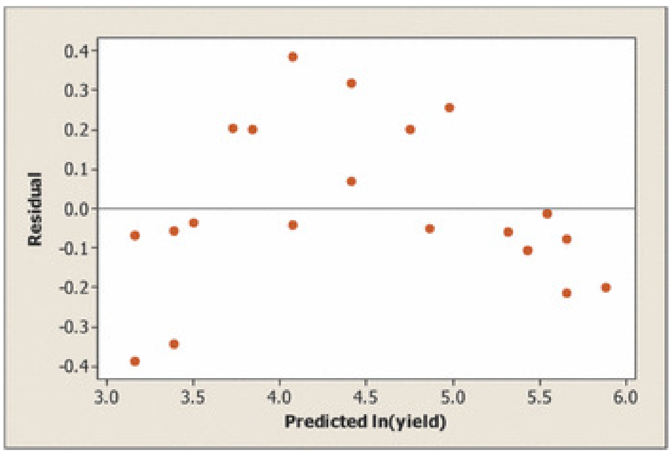
S = 14.3601 R-Sq = 97.7% R-Sq(adj) = 97.5%



Option 2: Taking natural logarithm of yield measurements

Predictor	Coef	SE Coef	T	P
Constant	1.2319	0.1795	6.86	0.000
DBH (inches)	0.113417	0.006081	18.65	0.000

S = 0.214894 R-Sq = 95.1% R-Sq(adj) = 94.8%



(b) Use both models to predict the amount of usable lumber from a Ponderosa pine with diameter 30 inches. Show your work.

(c) Which of the predictions in part (b) seems more reliable? Give appropriate evidence to support your choice.

Show Answer

Chapter 12 AP Statistics Practice Test