

2005 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

The regression equation is
Fuel Consumption = $10.7 + 2.15$ Railcars

Predictor	Coef	StDev	T	P
Constant	10.677	5.157	2.07	0.072
Railcar	2.1495	0.1396	15.40	0.000

$S = 4.361$ $R\text{-}Sq = 96.7\%$ $R\text{-}Sq(\text{adj}) = 96.3\%$

- (a) Is a linear model appropriate for modeling these data? Clearly explain your reasoning.
- (b) Suppose the fuel consumption cost is \$25 per unit. Give a point estimate (single value) for the change in the average cost of fuel per mile for each additional railcar attached to a train. Show your work.
- (c) Interpret the value of r^2 in the context of this problem.
- (d) Would it be reasonable to use the fitted regression equation to predict the fuel consumption for a train on this route if the train had 65 railcars? Explain.
4. Some boxes of a certain brand of breakfast cereal include a voucher for a free video rental inside the box. The company that makes the cereal claims that a voucher can be found in 20 percent of the boxes. However, based on their experiences eating this cereal at home, a group of students believes that the proportion of boxes with vouchers is less than 0.2. This group of students purchased 65 boxes of the cereal to investigate the company's claim. The students found a total of 11 vouchers for free video rentals in the 65 boxes.
- Suppose it is reasonable to assume that the 65 boxes purchased by the students are a random sample of all boxes of this cereal. Based on this sample, is there support for the students' belief that the proportion of boxes with vouchers is less than 0.2? Provide statistical evidence to support your answer.
5. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:
- Does not have a high school diploma
 - Has a high school diploma
- The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.
- (a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.
- (b) A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.
- (c) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

2005 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II grade—25

Directions: 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 of your results and explanation.

6. Lead, found in some paints, is a neurotoxin that can be especially harmful to the developing brain and nervous system of children. Children frequently put their hands in their mouth after touching painted surfaces, and this is the most common type of exposure to lead.

A study was conducted to investigate whether there were differences in children's exposure to lead between suburban day-care centers and urban day-care centers in one large city. For this study, researchers used a random sample of 20 children in suburban day-care centers. Ten of these 20 children were randomly selected to play outside; the remaining 10 children played inside. All children had their hands wiped clean before beginning their assigned one-hour play period either outside or inside. After the play period ended, the amount of lead in micrograms (mcg) on each child's dominant hand was recorded.

The mean amount of lead on the dominant hand for the children playing inside was 3.75 mcg, and the mean amount of lead for the children playing outside was 5.65 mcg. A 95 percent confidence interval for the difference in the mean amount of lead after one hour inside versus one hour outside was calculated to be $(-2.46, -1.34)$.

A random sample of 18 children in urban day-care centers in the same large city was selected. For this sample, the same process was used, including randomly assigning children to play inside or outside. The data for the amount (in mcg) of lead on each child's dominant hand are shown in the table below.

Urban Day-Care Centers

Inside	6	5	4	4	4.5	5	4.5	3	5
Outside	15	25	18	14	20	13	11	22	20

- (a) Use a 95 percent confidence interval to estimate the difference in the mean amount of lead on a child's dominant hand after an hour of play inside versus an hour of play outside at urban day-care centers in this city. Be sure to interpret your interval.

AP[®] STATISTICS
2005 SCORING GUIDELINES

Question 5

Solution

Part (a):

Since random-digit dialing will be used, individuals without phones will not be included in the sample. People without a high school diploma are more likely to have lower-paying jobs and therefore may not be able to afford a telephone. Thus, the estimated proportion of adult heads of households in the United States without a high school diploma may be less than the true population proportion.

Part (b):

The sample size necessary to estimate the proportion of the population that does not have a high school diploma, p , within 0.03 with 95% confidence is:

$$0.03 = z^* \left(\sqrt{\frac{p^*(1-p^*)}{n}} \right), \text{ or}$$
$$0.03 = 1.96 \left(\sqrt{\frac{0.22(1-0.22)}{n}} \right), \text{ so}$$
$$n = 0.22(0.78) \left[\frac{1.96}{0.03} \right]^2 = 732.4651.$$

Thus, 733 respondents would be needed.

Part (c):

To achieve this additional goal, the agency should use stratified random sampling by taking samples within each state. Each state would be a stratum. Within each state, a random sample of adult heads of households would be selected and surveyed. The sample size within each state will be based on the desired precision. Data from the individual states should be combined to obtain the national estimate.