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-Sq = 96.7% R-Sq(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.

Question 4

Solution

This question is divided into four parts.

Part (a): State a correct pair of hypotheses.

Let p = the proportion of boxes of this brand of breakfast cereal that include a voucher for a free video rental.

$H_0 : p = 0.2$
$H_a : p < 0.2$

Part (b): Identify a correct test (by name or by formula) and check appropriate conditions.

One-sample z-test for a proportion $\quad OR \quad z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$

Conditions:

1. $np_0 = 65 \times 0.2 = 13 > 10$ and $n(1 - p_0) = 65 \times 0.8 = 52 > 10$.
2. It is reasonable to assume that the company produces more than $65 \times 10 = 650$ boxes of this cereal ($N > 10n$).
3. The observations are independent because it is reasonable to assume that the 65 boxes are a random sample of all boxes of this cereal.

Part (c): Use correct mechanics and calculations, and provide the p-value (or rejection region).

The sample proportion is $\hat{p} = \dfrac{11}{65} = 0.169$. The test statistic is $z = \dfrac{0.169 - 0.2}{\sqrt{\dfrac{0.2(1 - 0.2)}{65}}} = -0.62$ and the p-value is

$P(Z < -0.62) = 0.2676$.

Part (d): State a correct conclusion, using the result of the statistical test, in the context of the problem.

Since the p-value $= 0.2676$ is larger than any reasonable significance level (e.g., $\alpha = 0.05$), we cannot reject the company's claim. That is, we do not have statistically significant evidence to support the student's belief that the proportion of cereal boxes with vouchers is less than 20 percent.

Question 4 (continued)

Scoring

The question is divided into four parts. Each part is scored as essentially correct (E) or incorrect (I).

Part (a) is essentially correct (E) if the student states a correct pair of hypotheses.

Notes:
1. Since the proportion was defined in the stem, standard notation for the proportion (p *or* π) need not be defined in the hypotheses.
2. Nonstandard notation must be defined correctly.
3. A two-sided alternative is incorrect for this part.

Part (b) is essentially correct (E) if the student identifies a correct test (by name or by formula) and checks for appropriate conditions.

Notes:
1. $np_0 > 5$ and $n(1 - p_0) > 5$ are OK as long as appropriate values are used for n and p_0.
2. Since students cannot check the actual population size, they do not need to mention it.
3. The stem of the problem indicates this is a random sample so it (or a discussion of independence) does not need to be repeated in the solution.

Part (c) is essentially correct (E) if no more than one of the following errors is present in the student's work:
- Undefined, nonstandard notation is used; OR
- The correct z-value $= -0.62$ is given with no setup for the calculation; OR
- The incorrect z-value $= -0.67$ is given because $\hat{p}$ was used in the calculation of the standard error. For this incorrect z-value, the p-value $= 0.2514.$; OR
- The incorrect z-value is calculated because of a minor arithmetic error.

Part (c) is incorrect (I) if:
- Inference for a lower tail alternative is based on either two-tails p-value $= 0.535$ or the upper tail p-value $= 0.734$; OR
- An unsupported z-value other than -0.62 or -0.67 is given; OR
- The correct z-value $= -0.62$ is given but equated to an incorrect formula.

Question 4 (continued)

Notes:

1. Students using a rejection region approach should have critical values appropriate for a lower tail test, e.g., for $\alpha = 0.05$ the rejection region is $z < -1.645$.
2. Other possible correct mechanics include:

   - Exact Binomial
     X ~ Binomial(n=65, p=0.2). The exact *p*-value is $P(X \leq 11) = 0.33$.

   - Normal Approximation to Binomial (with or without a continuity correction)
     X is approximately Normal(13, 3.225). The approximate *p*-value using the continuity correction is $P\left(Z \leq \frac{11 + 0.5 - 13}{3.225}\right) = P(Z \leq -0.4651) = 0.3209$.

   - Confidence interval approach – provided there is a reasonable interpretation tied to a significance level. For example if $\alpha = 0.05$, and $p = 0.20$ is within a 95% upper confidence bound (0, 0.2457) or a two-tailed 90% confidence interval (0.0927, 0.2457).

Part (d) is essentially correct (E) if the student states a correct conclusion in the context of the problem, using the result of the statistical test.

Notes:

1. If both an α and a *p*-value (or critical value) are given, the linkage is implied.
2. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the *p*-value or explaining how the conclusion follows from the *p*-value.
3. If the *p*-value in part (c) is incorrect but the conclusion is consistent with the computed *p*-value, part (d) can be considered as essentially correct (E).
4. If a student accepts the null hypothesis and concludes the proportion really is 0.20, this part is incorrect (I).

Each essentially correct (E) response counts as 1 point, each partially correct (P) response counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

Note: If a response is in between two scores (for example, 2 ½ points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.