

1. The goal of a nutritional study was to compare the caloric intake of adolescents living in rural areas of the United States with the caloric intake of adolescents living in urban areas of the United States. A random sample of ninth-grade students from one high school in a rural area was selected. Another random sample of ninth graders from one high school in an urban area was also selected. Each student in each sample kept records of all the food he or she consumed in one day.

The back-to-back stemplot below displays the number of calories of food consumed per kilogram of body weight for each student on that day.

<u>Urban</u>		<u>Rural</u>
9 9 9 9 8 8 7 6	2	
4 4 3 1 0	3	2 3 3 4
9 7 6 6 5	3	5 6 6 6 7
2 0	4	0 2 2 2 4
	4	5 6 8 8 9
	5	1

Stem: tens

Leaf: ones

- (a) Write a few sentences comparing the distribution of the daily caloric intake of ninth-grade students in the rural high school with the distribution of the daily caloric intake of ninth-grade students in the urban high school.
- (b) Is it reasonable to generalize the findings of this study to all rural and urban ninth-grade students in the United States? Explain.
- (c) Researchers who want to conduct a similar study are debating which of the following two plans to use.

Plan I: Have each student in the study record all the food he or she consumed in one day. Then researchers would compute the number of calories of food consumed per kilogram of body weight for each student for that day.

Plan II: Have each student in the study record all the food he or she consumed over the same 7-day period. Then researchers would compute the average daily number of calories of food consumed per kilogram of body weight for each student during that 7-day period.

Assuming that the students keep accurate records, which plan, I or II, would better meet the goal of the study? Justify your answer.

2. Let the random variable X represent the number of telephone lines in use by the technical support center of a software manufacturer at noon each day. The probability distribution of X is shown in the table below.

x	0	1	2	3	4	5
$p(x)$	0.35	0.20	0.15	0.15	0.10	0.05

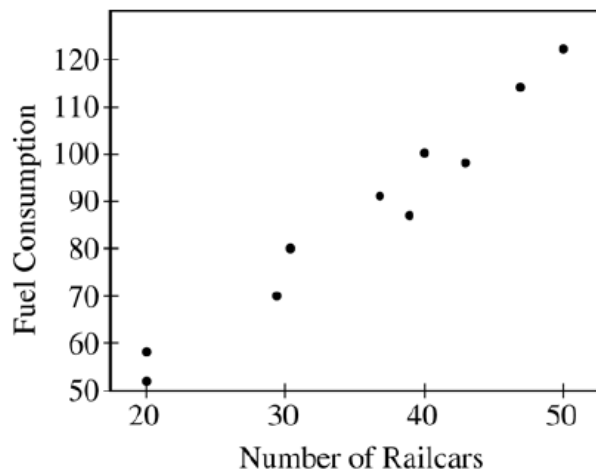
- (a) Calculate the expected value (the mean) of X .
- (b) Using past records, the staff at the technical support center randomly selected 20 days and found that an average of 1.25 telephone lines were in use at noon on those days. The staff proposes to select another random sample of 1,000 days and compute the average number of telephone lines that were in use at noon on those days. How do you expect the average from this new sample to compare to that of the first sample? Justify your response.
- (c) The median of a random variable is defined as any value x such that $P(X \leq x) \geq 0.5$ and $P(X \geq x) \geq 0.5$. For the probability distribution shown in the table above, determine the median of X .
- (d) In a sentence or two, comment on the relationship between the mean and the median relative to the shape of this distribution.

3. The Great Plains Railroad is interested in studying how fuel consumption is related to the number of railcars for its trains on a certain route between Oklahoma City and Omaha.

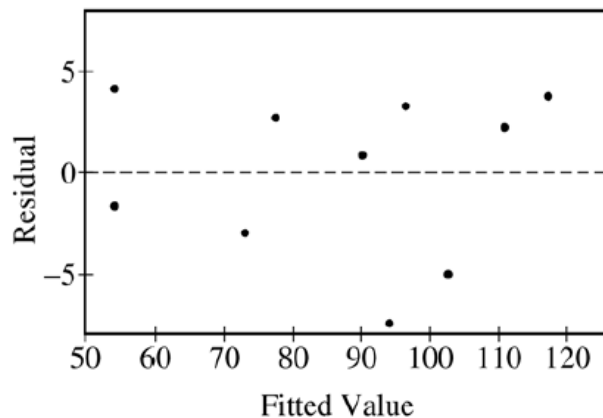
A random sample of 10 trains on this route has yielded the data in the table below.

Number of Railcars	Fuel Consumption (units/mile)
20	58
20	52
37	91
31	80
47	114
43	98
39	87
50	122
40	100
29	70

A scatterplot, a residual plot, and the output from the regression analysis for these data are shown below.



RESIDUALS VERSUS THE FITTED VALUES



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.

6. Every year, each student in a nationally representative sample is given tests in various subjects. Recently, a random sample of 9,600 twelfth-grade students from the United States were administered a multiple-choice United States history exam. One of the multiple-choice questions is below. (The correct answer is C.)

In 1935 and 1936 the Supreme Court declared that important parts of the New Deal were unconstitutional. President Roosevelt responded by threatening to

- (A) impeach several Supreme Court justices
- (B) eliminate the Supreme Court
- (C) appoint additional Supreme Court justices who shared his views
- (D) override the Supreme Court's decisions by gaining three-fourths majorities in both houses of Congress

Of the 9,600 students, 28 percent answered the multiple-choice question correctly.

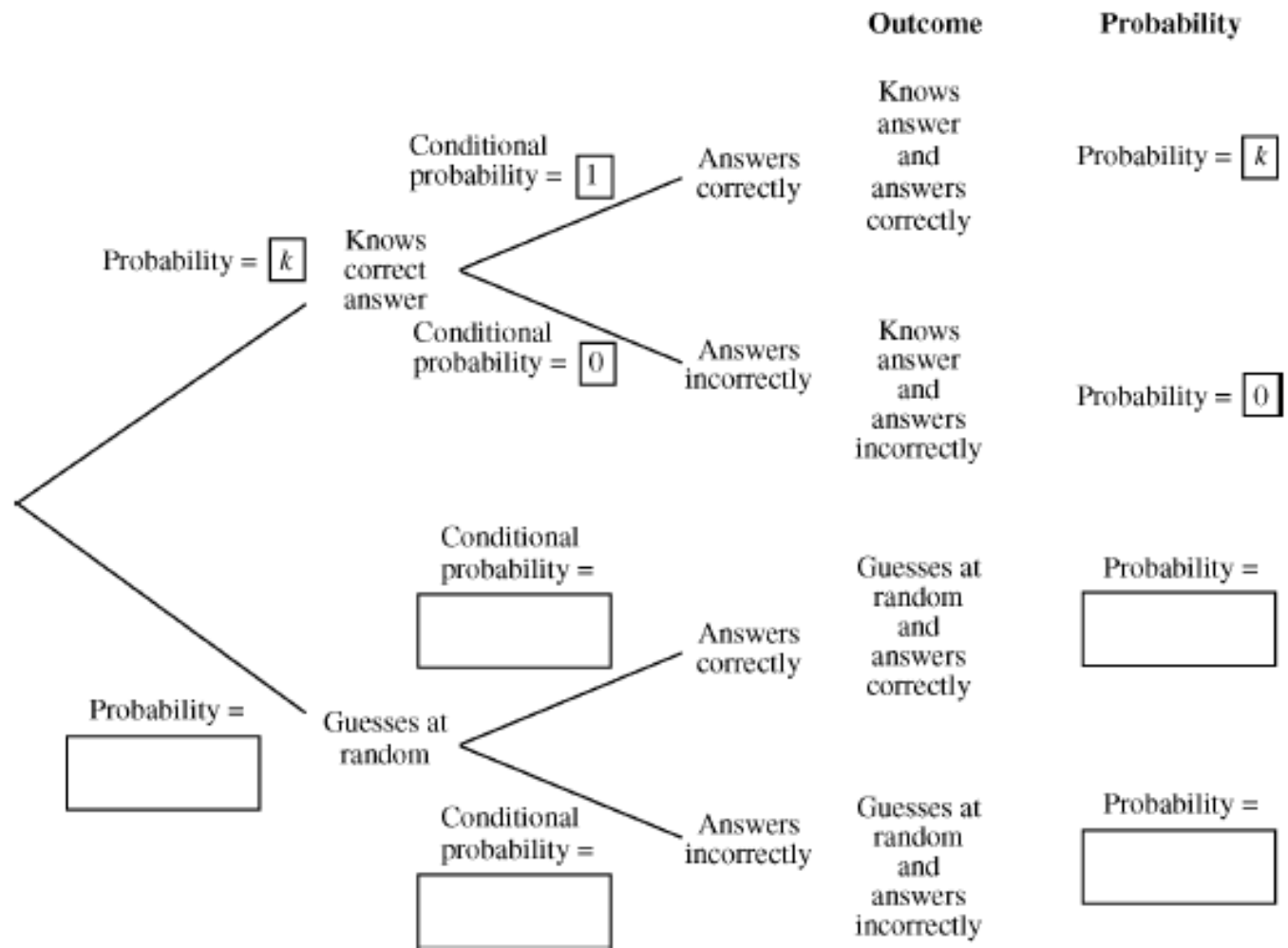
- (a) Let p be the proportion of all United States twelfth-grade students who would answer the question correctly. Construct and interpret a 99 percent confidence interval for p .

Assume that students who actually know the correct answer have a 100 percent chance of answering the question correctly, and students who do not know the correct answer to the question guess completely at random from among the four options.

Let k represent the proportion of all United States twelfth-grade students who actually know the correct answer to the question.

- (b) A tree diagram of the possible outcomes for a randomly selected twelfth-grade student is provided below. Write the correct probability in each of the five empty boxes. Some of the probabilities may be expressions in terms of k .

TREE DIAGRAM OF OUTCOMES FOR A
RANDOMLY SELECTED TWELFTH-GRADE STUDENT



- (c) Based on the completed tree diagram, express the probability, in terms of k , that a randomly selected twelfth-grade student would correctly answer the history question.
- (d) Using your interval from part (a) and your answer to part (c), calculate and interpret a 99 percent confidence interval for k , the proportion of all United States twelfth-grade students who actually know the answer to the history question. You may assume that the conditions for inference for the confidence interval have been checked and verified.