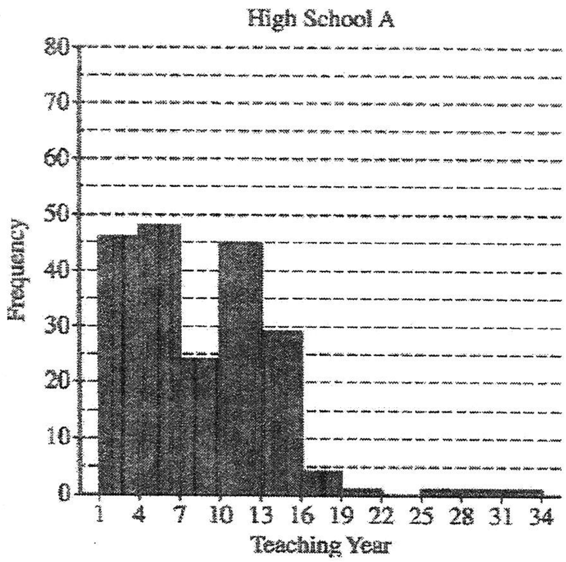
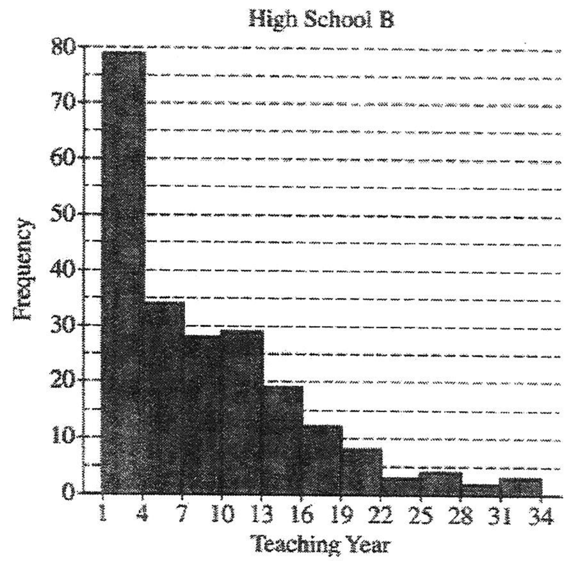
AP FRQ Review – Mr. Rich **Name:**______________________________ **Per:**_____ **Seat:**_____

AP Stats

Topic 1: Exploring and Understanding Data

. The following histograms summarize the teaching year for the teachers at two high schools, A and B.





Teaching year is recorded as an integer, with first-year teachers recorded as 1, second-year teachers recorded as 2, and so on. Both sets of data have a mean teaching year of 8.2, with data recorded from 200 teachers at High School A and 221 teachers at High School B. On the histograms, each interval represents possible integer values from the left endpoint up to but not including the right endpoint.

(a) The median teaching year for one high school is 6, and the median teaching year for the other high school is 7. Identify which high school has each median and justify your answer.

(b) An additional 18 teachers were not included with the data recorded from the 200 teachers at High School A. The mean teaching year of the 18 teachers is 2.5. What is the mean teaching year for all 218 teachers at High School A?

(c) The standard deviation of the teaching year for the 221 teachers at High School B is 7.2. If one teacher is selected at random from High School B, what is the probability that the teaching year for the selected teacher will be within 1 standard deviation of the mean of 8.2 ? Justify your answer.

Question 5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) determine which of two histograms represents data with a larger median; (2) calculate the mean of a combined data set when the separate means and sample sizes are known; and (3) calculate the probability that an individual randomly chosen from a finite population will have a value within one standard deviation of the mean, when provided with values for the mean, standard deviation, and all members of the population.

Solution

Part (a):

The median teaching year for High School A is any value with 100 data values at or below it and 100 data values at or above it. The median teaching year for High School B is the 111th value in the ordered list of values. For High School A the median is in the interval that starts at 7 and ends just before 10, because there are only 94 data values below 7 and 106 data values of at least 7. Therefore the median cannot be less than 7. For High School B the median is in the interval that starts at 4 and ends just before 7 because there are more than half (113) of the data values less than 7. Therefore the median must be less than 7. So High School A must be the one with a median of 7, and High School B must be the one with a median of 6.

Another way to determine which school has the median of 7 is to notice that the distribution for High School B is highly skewed to the right, whereas the distribution for High School A is bimodal with a few possible outliers on the right. A distribution that is highly right-skewed is likely to have a substantially larger mean than median. The mean of both distributions is given as 8.2 years, so it makes sense that the highly right-skewed distribution (High School B) is the one with the bigger gap between the mean and median and, therefore, the one with the lower median of 6.

Part (b):

The mean for the original 200 teachers was given as 8.2 years, and the mean for the additional 18 teachers is 2.5 years. Therefore the mean for the combined data set is:

$$\frac{(200)(8.2)+(18)(2.5)}{200+18}=\frac{1{,}640+45}{218}\approx 7.73 \text{ years.}$$

Part (c):

The interval mean plus or minus 1 standard deviation on either side of the mean is 8.2 ± 7.2, or from 1.0 year to 15.4 years. Because teaching year is recorded as an integer, the interval includes teaching years 1 to 15. The number of teachers in that interval can be found by adding the heights of the five bars in the histogram for the intervals from 1 to 16, which includes $79+34+28+29+19=189$. Therefore the probability is $\frac{189}{221} \approx 0.8552$.

A grocery store purchases melons from two distributors, J and K. Distributor J provides melons from organic farms. The distribution of the diameters of the melons from Distributor J is approximately normal with mean 133 millimeters (mm) and standard deviation 5 mm.

(a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm?

Distributor K provides melons from nonorganic farms. The probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137 mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor K.

(b) For a melon selected at random from the grocery store, what is the probability that the melon will have a diameter greater than 137 mm?

(c) Given that a melon selected at random from the grocery store has a diameter greater than 137 mm, what is the probability that the melon will be from Distributor J?

AP® STATISTICS
2017 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate a probability from a normal distribution; (2) calculate a weighted probability from two individual probabilities; and (3) calculate a conditional probability for dependent events when individual and joint probabilities are provided.

Solution

Part (a):

Let X denote the diameter of a randomly selected melon from Distributor J. X has an approximately normal distribution with mean 133 mm and standard deviation 5 mm.

The z-score for a diameter of 137 mm is $z = \frac{137-133}{5} = \frac{4}{5} = 0.8$.

Therefore, $P(X > 137) = P(Z > 0.8) = 1 - 0.7881 = 0.2119$.

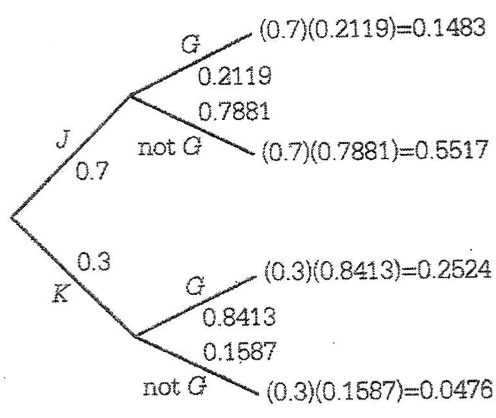
Part (b):

Define events:
J: melon is from Distributor J
K: melon is from Distributor K
G: melon diameter is greater than 137 mm

For a randomly selected melon from the grocery store,

$$\begin{aligned} P(G) &= P(G \mid J) \times P(J) + P(G \mid K) \times P(K) \\ &= (0.2119)(0.7) + (0.8413)(0.3) \\ &= 0.1483 + 0.2524 \\ &= 0.4007 \end{aligned}$$

OR

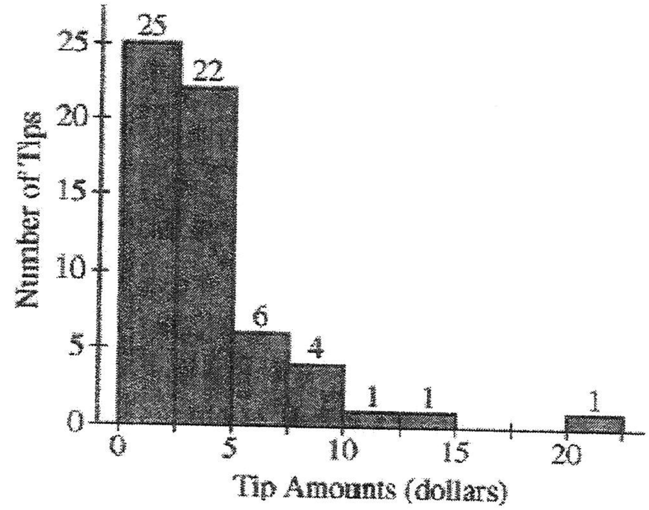


From the tree diagram, $P(G) = P(G \text{ and } J) + P(G \text{ and } K) = 0.1483 + 0.2524 = 0.4007$.

Part(c): $P(J \mid G) = \frac{P(J \text{ and } G)}{P(G)} = \frac{P(G \mid J)\,P(J)}{P(G)} = \frac{(0.2119)(0.7)}{0.4007}$

$= 0.3701$

Robin works as a server in a small restaurant, where she can earn a tip (extra money) from each customer she serves. The histogram below shows the distribution of her 60 tip amounts for one day of work.



(a) Write a few sentences to describe the distribution of tip amounts for the day shown.

(b) One of the tip amounts was $8. If the $8 tip had been $18, what effect would the increase have had on the following statistics? Justify your answers.

The mean:

The median:

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe the distribution of a quantitative variable based on a histogram and (2) determine the effect of changing one data value on the mean and the median.

Solution

Part (a):

The distribution of Robin's tip amounts is skewed to the right. There is a gap between the largest tip amount (in the \$20 to \$22.50 interval) and the second largest tip amount (in the \$12.50 to \$15 interval), and the largest tip amount appears to be an outlier. The median tip amount is between \$2.50 and \$5.00. Robin's tip amounts vary from a minimum of between \$0 and \$2.50 to a maximum of between \$20.00 and \$22.50. About 78 percent of the tip amounts are between \$0 and \$5.

Part (b):

The mean: If the \$8 tip had been \$18, the mean would increase by \$10 divided by 60, or $\$\frac{1}{6}$, or about 17 cents.

The median: If the \$8 tip had been \$18, the median would not change because the current median is between \$2.50 and \$5.00, and both \$8 and \$18 are greater than that.

Scoring

Parts (a) and (b) are scored as essentially correct (E), partially correct (P), or incorrect (I).

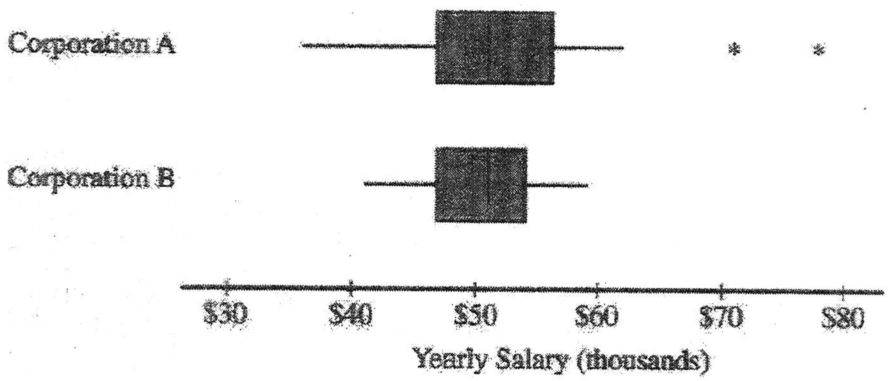
Part (a) is scored as follows:

Essentially correct (E) if the response includes reasonable comments on the following five components:
1. Shape (skewed right)
2. Outlier (at least one) *OR* gap (one tip amount greater than \$20, next highest at most \$15)
3. Center between \$2.50 and \$5.00 (median) or between \$2.62 and \$5.13 (mean)
4. Variability, by noting that the tip amounts vary from about \$0 to at most \$22.50, or that a majority of tip amounts are between \$0 and a value greater than or equal to \$5, or by providing a correct numerical approximation of a measure of variability
5. Context (tip amounts)

Partially correct (P) if the response includes only three or four of the five components.

Incorrect (I) if the response includes at most two of the five components.

Two large corporations, A and B, hire many new college graduates as accountants at entry-level positions. In 2009 the starting salary for an entry-level accountant position was $36,000 a year at both corporations. At each corporation, data were collected from 30 employees who were hired in 2009 as entry-level accountants and were still employed at the corporation five years later. The yearly salaries of the 60 employees in 2014 are summarized in the boxplots below.



(a) Write a few sentences comparing the distributions of the yearly salaries at the two corporations.

(b) Suppose both corporations offered you a job for $36,000 a year as an entry-level accountant.

(i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.

(ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) compare features of two distributions of data displayed in boxplots and (2) identify statistical measures that are important in making decisions based on data sets.

Solution

Part (a):

The median salary is approximately the same for both corporations. The range and interquartile range of the salaries are greater for Corporation A than for Corporation B. The two highest salaries at Corporation A are outliers while Corporation B has no outliers.

Part (b):

(i) Five years after starting, at least 3 out of 30 (10%) of the salaries at Corporation A are greater than the maximum salary at Corporation B. If I accept the offer from Corporation A, I might be able to make a higher salary at Corporation A than at Corporation B.

(ii) Five years after starting, the minimum salary at Corporation B is greater than at Corporation A. In fact, at Corporation A it looks like some people are still making the starting salary of $36,000 and never received a raise in the five years since they were hired. So if I work at Corporation A, I might never receive a raise in salary.

Scoring

Parts (a) and (b) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

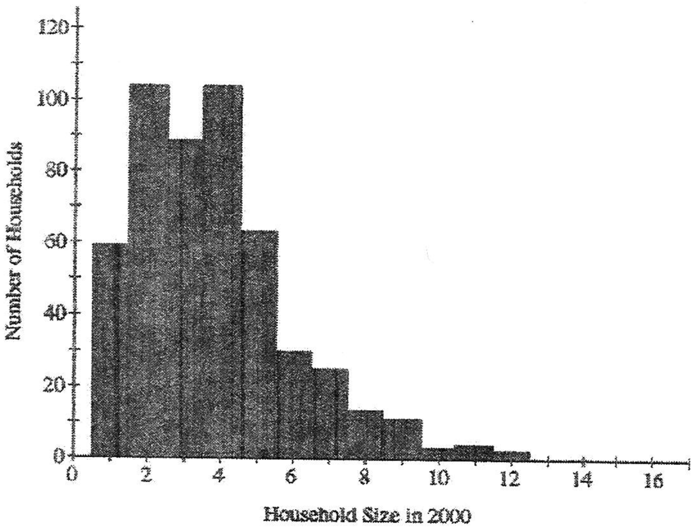
Essentially correct (E) if the response includes the following four components:
1. A correct comparison of center.
2. A correct comparison of spread.
3. A discussion of the outliers for Corporation A.
4. The response is in context.

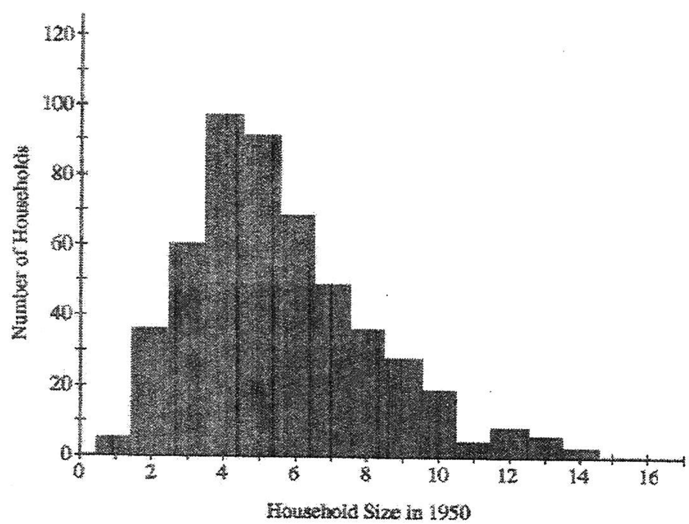
Partially correct (P) if the response includes only three of the four components.

Incorrect (I) if the response includes at most two of the four components.

Note: Any mention of shape should be ignored because complete shape information cannot be determined from a boxplot.

Independent random samples of 500 households were taken from a large metropolitan area in the United States for the years 1950 and 2000. Histograms of household size (number of people in a household) for the years are shown below.





(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

(b) A researcher wants to use these data to construct a confidence interval to estimate the change in mean household size in the metropolitan area from the year 1950 to the year 2000. State the conditions for using a two-sample t-procedure, and explain whether the conditions for inference are met.

Question 3

Intent of Question

The primary goals of this question were to assess students' ability to (1) compare two distributions presented with histograms; (2) comment on the appropriateness of using a two-sample *t*-procedure in a given setting.

Solution

Part (a):

Household size tended to be larger in 1950 than in 2000. The histograms reveal a much larger proportion of small (1-, 2-, and 3-person) households in 2000 than in 1950. Similarly, the histograms reveal a much smaller proportion of large (5-person and larger) households in 2000 than in 1950. Also, the median household sizes can be calculated to be 5 people per household in 1950 compared with 3 or 4 people per household in 2000. The year 1950 displayed slightly more variability in household sizes than the year 2000. Although the interquartile ranges for both years are the same (3 people), the standard deviation (1950: about 2.6 people; 2000: about 2.1 people) and the range (1950: 13 people; 2000: 11 people) are larger for 1950 than for 2000. Both distributions of household size are skewed to the right. In both years, there are a few households with very large families, as large as 14 people in 1950 and 12 people in 2000.

Part (b):

The conditions for applying a two-sample *t*-procedure are:

1. The data come from independent random samples or from random assignment to two groups;
2. The populations are normally distributed, or both sample sizes are large;
3. The population sizes are at least 10 (or 20) times the sample sizes.

The first condition is satisfied because independent random samples were selected for the years 1950 and 2000. The second condition is satisfied because the sample sizes (500 in each group) are quite large, despite the right skewness of the distributions of household sizes in the sample data. The third condition is satisfied because the number of households in the large metropolitan area in both 1950 and 2000 would easily exceed $10 \times 500 = 5{,}000$.

Scoring

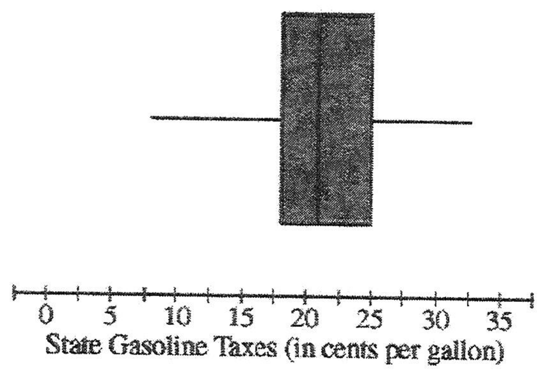
This question is scored in four sections. Part (a) has three components: (1) comparing the centers of the two distributions; (2) comparing variability for the two distributions; (3) identifying the shapes of both distributions and including context related to the variable of interest. Section 1 consists of part (a), component 1; section 2 consists of part (a), component 2; section 3 consists of part (a), component 3. Section 4 consists of part (b). Sections 1 and 2 are scored as essentially correct (E) or incorrect (I). Sections 3 and 4 are scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response correctly compares center (or location) for both distributions.

Incorrect (I) otherwise.

1. As gasoline prices have increased in recent years, many drivers have expressed concern about the taxes they pay on gasoline for their cars. In the United States, gasoline taxes are imposed by both the federal government and by individual states. The boxplot below shows the distribution of the state gasoline taxes, in cents per gallon, for all 50 states on January 1, 2006.



(a) Based on the boxplot, what are the approximate values of the median and the interquartile range of the distribution of state gasoline taxes, in cents per gallon? Mark and label the boxplot to indicate how you found the approximated values.

(b) The federal tax imposed on gasoline was 18.4 cents per gallon at the time the state taxes were in effect. The federal gasoline tax was added to the state gasoline tax for each state to create a new distribution of combined gasoline taxes. What are approximate values, in cents per gallon, of the median and interquartile range of the new distribution of combined gasoline taxes? Justify your answer.

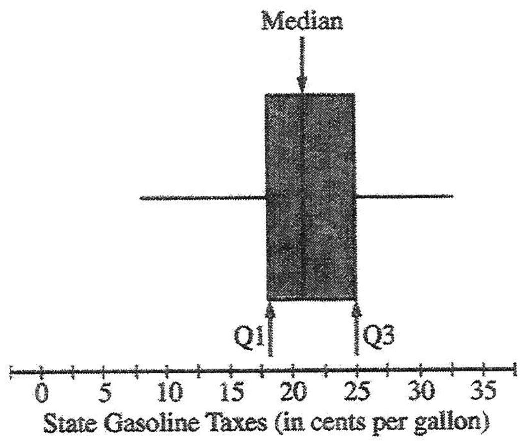
Question 1

Intent of Question

The primary goals of this question were to assess students' ability to (1) approximate the median and the IQR of a distribution from the boxplot and (2) recalibrate the values of the median and the IQR of the distribution if the same constant value is added to each observation in the distribution.

Solution

Part (a):



The median and quartiles are marked and labeled on the boxplot above. The median is approximately 21 cents per gallon.

The first and third quartiles are approximately 18 cents per gallon and 25 cents per gallon, respectively. The IQR is Q3 – Q1, which is approximately 25 – 18 = 7 cents per gallon.

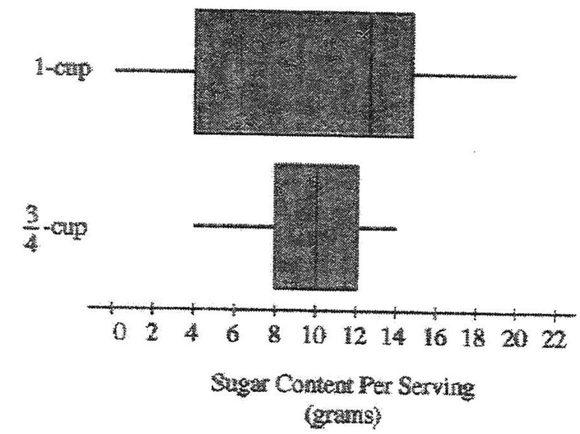
Part (b):

After adding 18.4 cents per gallon to each of the state taxes, the median of the combined gasoline taxes would be the median of the state tax plus the federal tax, which is approximately 21 + 18.4 = 39.4 cents per gallon.

Although the quartiles of the combined gasoline taxes will change (Q1 = 18 + 18.4 = 36.4 cents per gallon and Q3 = 25 + 18.4 = 43.4 cents per gallon), the IQR will remain the same as it was for the state taxes at 7 cents per gallon (43.4 – 36.4 = 7).

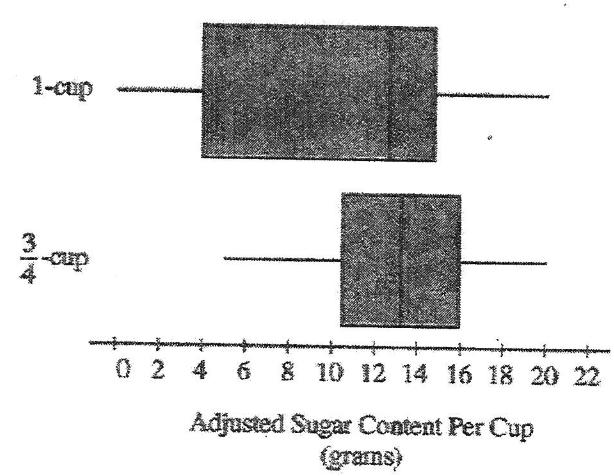
1. To determine the amount of sugar in a typical serving of breakfast cereal, a student randomly selected 60 boxes of different types of cereal from the shelves of a large grocery store.

The student noticed that the side panels of some of the cereal boxes showed sugar content based on one-cup servings, while others showed sugar content based on three-quarter-cup servings. Many of the cereal boxes with side panels that showed three-quarter-cup servings were ones that appealed to young children, and the student wondered whether there might be some difference in the sugar content of the cereals that showed different-size servings on their side panels. To investigate the question, the data were separated into two groups. One group consisted of 29 cereals that showed one-cup serving sizes; the other group consisted of 31 cereals that showed three-quarter-cup serving sizes. The boxplots shown below display sugar content (in grams) per serving of the cereals for each of the two serving sizes.



(a) Write a few sentences to compare the distributions of sugar content per serving for the two serving sizes of cereals.

After analyzing the boxplots on the preceding page, the student decided that instead of a comparison of sugar content per recommended serving, it might be more appropriate to compare sugar content for equal-size servings. To compare the amount of sugar in serving sizes of one cup each, the amount of sugar in each of the cereals showing three-quarter-cup servings on their side panels was multiplied by $\frac{4}{3}$. The bottom boxplot shown below displays sugar content (in grams) per cup for those cereals that showed a serving size of three-quarter-cup on their side panels.



(b) What new information about sugar content do the boxplots above provide?

(c) Based on the boxplots shown above on this page, how would you expect the mean amounts of sugar per cup to compare for the different recommended serving sizes? Explain.

AP® STATISTICS
2008 SCORING GUIDELINES

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) compare two distributions; (2) reevaluate shape, center, and spread for comparing the two distributions after one of the distributions is transformed by multiplying each of the data points by a constant; and (3) make a prediction about the means of the two distributions based on information derived about the behavior of the distributions from the boxplots.

Solution

Part (a):

The cereals that list a serving size of one cup have a median sugar amount larger than the median for the cereals that list a serving size of three-quarters of a cup. There is more variability (larger range and larger IQR) for the one-cup cereals. The shapes of the two distributions differ. The distribution of sugar content for three-quarter-cup cereals is reasonably symmetric: notice that the median is in the middle of the box. The distribution of sugar content for one-cup cereals is clearly skewed to the left (skewed toward the lower values): notice that the median is pulled to the right side of the central box closer to the third quartile.

Part (b):

The distribution of sugar content in the cereals that list one-cup serving sizes remains the same as in part (a) because no transformations were applied to this distribution. There is a noticeable shift toward higher sugar content for the cereals that list three-quarter-cup servings after the transformation was applied to this distribution. The two types of cereals (one-cup and three-quarter-cup) now have similar medians, and the two distributions now show similar maximum values. In addition, the variability in the sugar content for cereals with a three-quarter-cup serving size increased by a factor of $\frac{4}{3}$ after the transformation was applied to the data in this distribution.

Part (c):

Judging from the boxplots in part (b), we would expect the mean amounts of sugar per serving to be different. By the symmetry of the boxplot for the three-quarter-cup cereals, we would expect the mean and median to be similar. Because the boxplot for the one-cup cereals is skewed to the left, we would expect the mean to be lower than the median. Thus, because both types of cereal have similar medians, we would expect the mean amount of sugar per cup for cereals with a one-cup serving size to be lower than the mean amount of sugar per cup for cereals with a three-quarter-cup serving size.

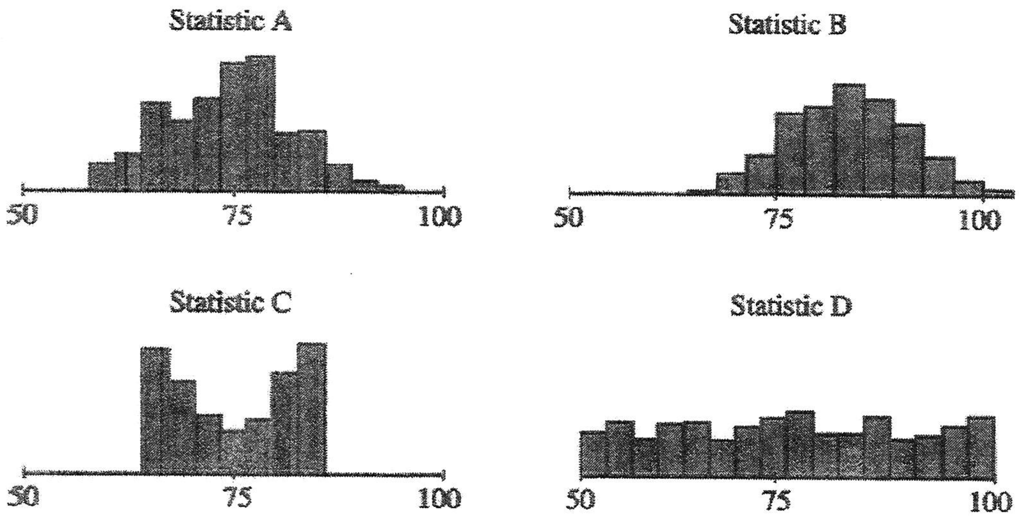
Scoring

Parts (a), (b), and (c) are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the student correctly <u>compares</u> center, shape, and spread of the two distributions. Specific numerical values are not required.

Four different statistics have been proposed as estimators of a population parameter. To investigate the behavior of these estimators, 500 random samples are selected from a known population and each statistic is calculated for each sample. The true value of the population parameter is 75. The graphs below show the distribution of values for each statistic.



(a) Which of the statistics appear to be unbiased estimators of the population parameter?

How can you tell?

b) Which of statistics A or B would be a better estimator of the population parameter?

Explain your choice.

c) Which of statistics C or D would be a better estimator of the population parameter?

Explain your choice.

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) recognize an unbiased estimator and explain why the estimator is unbiased and (2) compare two estimators with respect to center and variability.

Solution

Part (a):

Statistics A, C, and D appear to be unbiased. This is indicated by the fact that the mean of the estimated sampling distribution for each of these statistics is about 75, the value of the true population parameter.

Part (b):

Statistic A would be a better choice because it appears to be unbiased. Although the variability of the two estimated sampling distributions is similar, statistic A would produce estimates that tend to be closer to the true population parameter value of 75 than would statistic B.

Part (c):

Statistic C would be a better choice because it has smaller variability. Although both statistic C and statistic D appear to be unbiased, statistic C would produce estimates that tend to be closer to the true population parameter value of 75 than would statistic D.

Scoring

Parts (a), (b), and (c) are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

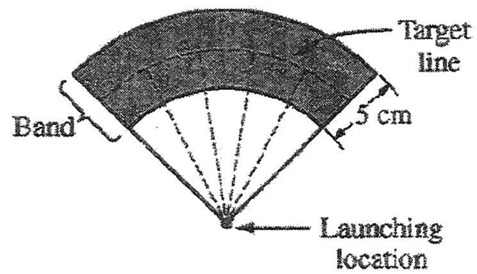
Essentially correct (E) if the response contains the following two components.

Component 1: Identifies statistics A, C, and D as the unbiased estimators.
Component 2: Clearly demonstrates an understanding of the meaning of the term *unbiased*. That is, states that the mean (or center) of each distribution is about 75. No other characteristic (e.g., shape, spread) should be mentioned in the response unless it clearly is discounted as a criterion for being unbiased.

Partially correct (P) if the response contains just one of these components. That is, the response either identifies statistics A, C, and D as the unbiased estimators but gives a weak or no explanation or includes some discussion of another characteristic (e.g., shape, spread) as of some importance in judging bias *OR* demonstrates clear understanding of the meaning of the term *unbiased* but identifies only one or two of statistics A, C, and D as unbiased estimators.

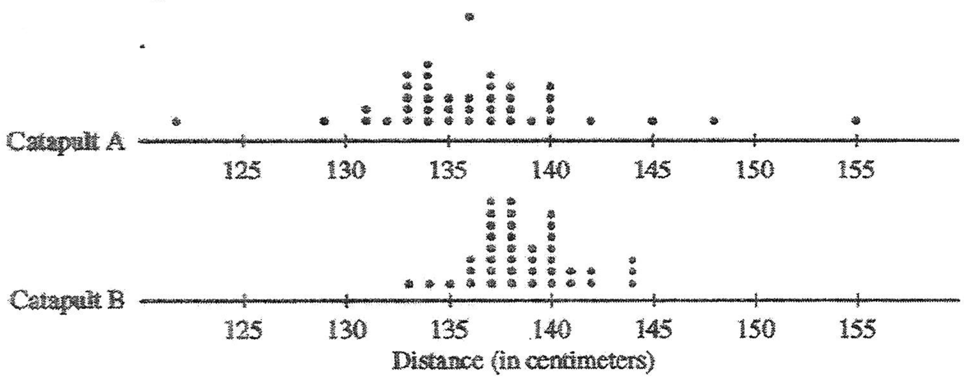
Incorrect (I) otherwise.

Two parents have each built a toy catapult for use in a game at an elementary school fair. To play the game, students will attempt to launch Ping-Pong balls from the catapults so that the balls land within a 5-centimeter band. A target line will be drawn through the middle of the band, as shown in the figure below. All points on the target line are equidistant from the launching location.



If a ball lands within the shaded band, the student will win a prize.

The parents have constructed the two catapults according to slightly different plans. They want to test these catapults before building additional ones. Under identical conditions, the parents launch 40 Ping-Pong balls from each catapult and measure the distance that the ball travels before landing. Distances to the nearest centimeter are graphed in the dotplots below.



(a) Comment on any similarities and any differences in the two distributions of distances traveled by balls launched from catapult A and catapult B.

(b) If the parents want to maximize the probability of having the Ping-Pong balls land within the band, which one of the two catapults, A or B, would be better to use than the other? Justify your choice.

(c) Using the catapult that you chose in part (b), how many centimeters from the target line should this catapult be placed? Explain why you chose this distance.

Question 1

Intent of Question

The primary goals of this question are: (1) to assess a student's ability to use simple graphical displays (dotplots in this case) to compare and contrast two distributions; and (2) to evaluate a student's ability to recognize what statistical information is most useful in making different practical decisions.

Solution

Part (a):

Both distributions of distances are roughly symmetric and somewhat mound-shaped. The center of the distances for catapult A (median A = 136 cm) is slightly lower than the center of the distances for catapult B (median B = 138 cm). There is more variability in the distances traveled by the Ping-Pong balls launched with catapult A. There are distances that are extreme enough to be called (potential) outliers in the catapult A distribution, but there are no outliers among the catapult B distances.

Part (b):

Catapult B would be best because the distances vary less about the center of the distribution for catapult B. If catapult B is properly placed, the balls launched will have a higher probability of landing in the narrow (only 5 cm wide) target band.

Part (c):

The catapult should be placed 138 cm from the target line. Since the distribution of distances for catapult B seems to be fairly symmetric and somewhat mound-shaped, the median (138 cm) is a good representation of the center of the distribution. Placing catapult B at this location would have resulted in a high proportion (30/40 = 0.75) of Ping-Pong balls from this sample of launches landing in the target band.

Scoring

Parts (a), (b), and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if the student correctly identifies similarities and differences in center, spread, and shape for the two distributions.

Part (a) is partially correct (P) if the student correctly identifies similarities and differences in two of the three characteristics (center, shape, and spread) for the two distributions.

Part (a) is incorrect (I) if the student correctly identifies no more than one similarity or difference of the three characteristics (center, shape, and spread) for the two distributions.

Notes:

- Correct comments regarding outliers should be viewed as a positive. However, comments about outliers do not count as one of the three required characteristics.

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.

Urban		Rural
99998876	2	
44310	3	2334
97665	3	56667
20	4	02224
	4	56889
	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.

Question 1

Solution

Part (a):

The mean (40.45 cal/kg) and median (41 cal/kg) daily caloric intake of ninth-grade students in the rural school are higher than the corresponding measures of center, mean (32.6 cal/kg) and median (32 cal/kg), for ninth-graders in the urban school. There is also more variability or spread in the daily caloric intake for students in the rural school (Range=19, SD=6.04, IQR=10) than in the daily caloric intake for students in the urban school (Range=16, SD=4.67, IQR=7). The shapes of the two distributions are also different. The distribution of daily caloric intake for rural students is more uniformly distributed (symmetric) between 32 cal/kg and 51 cal/kg while the distribution of daily caloric intake for urban students appears to be skewed toward the larger values.

Part (b):

No, the samples include students from only one rural and one urban high school so it is not reasonable to generalize the findings from these two schools to all rural and urban ninth-grade students in the United States.

Part (c):

Since we are assuming that students keep accurate records, Plan II will do a better job of comparing the daily caloric intake of adolescents living in rural areas with the daily caloric intake of adolescents living in urban areas. Both plans take body weight into account by converting to food consumed per kilogram of body weight. Plan II includes a 7-day period (possibly days in school and days at home on the weekend), and there are differences in caloric intake among days. It would therefore be better to average over the 7-day period rather than considering only the food consumed in one day, as is the case with Plan I. Plan II would provide a more precise estimate of the average daily intake.

Scoring

Parts (a) and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I). Part (b) is scored as essentially correct (E) or incorrect (I).

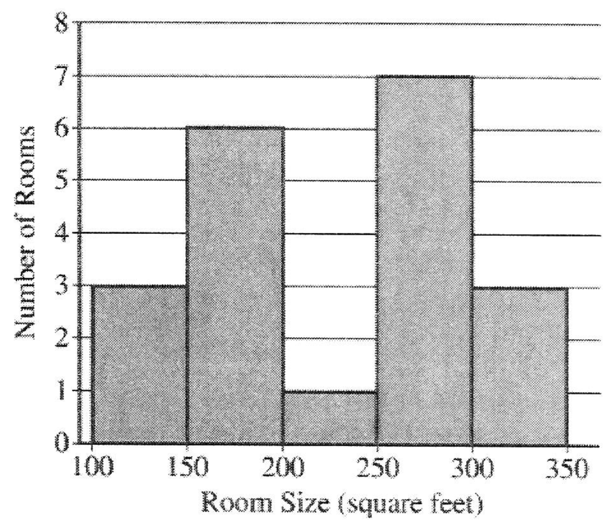
Part (a) is essentially correct (E) if the student correctly compares center, shape, and spread of the two distributions. Specific numerical values are not required.

Part (a) is partially correct (P) if the student correctly compares any two of the three characteristics (center, shape, or spread) of the two distributions.

Part (a) is incorrect (I) if the student correctly compares no more than one characteristic.

AP Statistics Free Response Questions – 2019

The sizes, in square feet, of the 20 rooms in a student residence hall at a certain university are summarized in the following histogram.

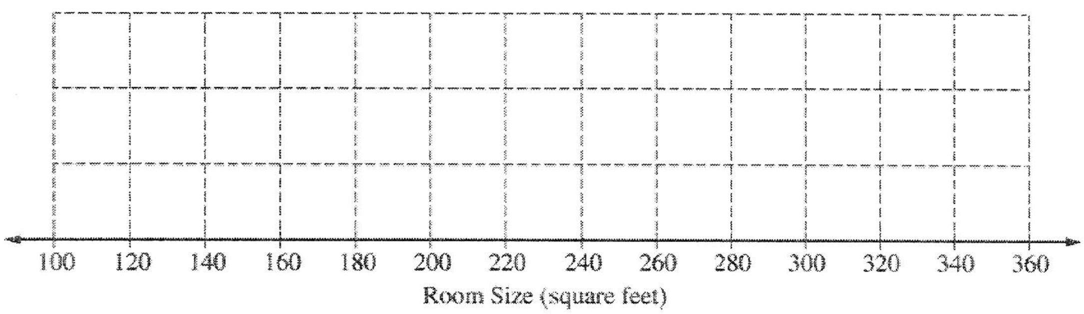


(a) Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.

(b) Summary statistics for the sizes are given in the following table.

Mean	Standard Deviation	Min	Q1	Median	Q3	Max
231.4	68.12	134	174	253.5	292	315

Determine whether there are potential outliers in the data. Then use the following grid to sketch a boxplot of room size.



(c) What characteristic of the shape of the distribution of room size is apparent from the histogram but not from the boxplot?

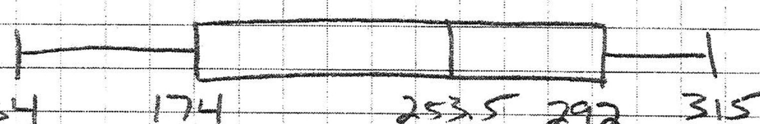
AP Stats 2019 FRQs

1. a. The distribution of room sizes is bimodal, with centers at about 150-200 ft^2, and 250-300 ft^2. There are no outliers, and distribution is symmetric. The median of the distribution is around 250 ft^2. The distribution of room sizes ranges from around 100 ft^2 to 350 ft^2.

b. To check for outliers, we will calculate the upper and lower fences.

$$\text{Upper Fence} = Q3 + 1.5\,IQR$$
$$= 292 + 1.5(292-174) = 469\ ft^2$$
$$\text{Lower Fence} = Q1 - 1.5\,IQR$$
$$= 174 - 1.5(292-174) = -3\ ft^2$$

The minimum room size of 134 ft^2 is higher than the lower fence, and the maximum room size of 315 ft^2 is less than the upper fence. Therefore there are no outliers in the distribution of room sizes.



c. The bimodal nature of the distribution of room sizes is apparent in the histogram but not the boxplot.