2018

AP CollegeBoard

AP Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free Response Question 5

- ☑ **Scoring Guideline**
- ☑ **Student Samples**
- ☑ **Scoring Commentary**

AP® STATISTICS
2018 SCORING GUIDELINES

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 $\frac{189}{221} \approx 0.8552$.

AP® STATISTICS
2018 SCORING GUIDELINES

Question 5 (continued)

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. States that the median is 6 for High School B and the median is 7 for High School A.
2. Provides a reasonable explanation of how the decision was made.
3. Provides the definition of the median or explicitly applies the definition of a median as a criterion in reaching their decision.

OR

Essentially correct (E) if the response satisfies the following three components:

1. States that the median is 6 for High School B and the median is 7 for High School A.
2. States that High School B shows a skewed distribution (or High School A shows a less skewed distribution).
3. Provides a reasonable explanation of how the more skewed distribution (High School B) would be the one with a larger separation between the mean and median.

Partially correct (P) if the response satisfies the first component and only one of the other two components required for E.

Incorrect (I) if the response does not meet the criteria for E or P.

Note: An incorrect statistical statement in the response will result in E being lowered to P, but not P being lowered to I. For example,

- If either distribution is described as left skewed, normal, or approximately normal;
- If the discussion would indicate a median different than 7 for High School A or a median different than 6 for High School B.

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

1. The correct answer that the mean is 7.73.
2. Enough work to show that the answer was obtained as a weighted average of the two individual means.

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) if the response does not satisfy the requirements for E or P.

Question 5 (continued)

Part (c) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. Calculates that the appropriate interval is 1 to 15.4 or 1 to 15 teaching years.
2. Correctly sums the counts of data values in the numerator based on the intervals provided.
3. Computes the probability using 221 as the denominator.

Partially correct (P) if the response satisfies only two of the three components;
OR
if the response reports the correct probability (0.8552) without supporting work.

Incorrect (I) if the response satisfies at most one of the three components.

Notes:

- If the response attempts to use the Empirical Rule or normal distribution to provide the desired probability, the response is scored I.
- If an incorrect count is shown in component 2, for instance by including the interval from 16 to 19, then component 3 is satisfied if that incorrect count is divided by 221 to find the reported probability.
- It is acceptable if the count is slightly off because of difficulty reading the exact heights of the bars in the histogram.
- If only one of component 2 or component 3 is missing, but the correct probability (0.8552) is reported, the response can be scored E.
- If the response recognizes that all values in the histogram bins up to 16 fall within one standard deviation of the mean and reports the interval as 1 to 16, component 1 is satisfied.

Question 5 (continued)

4 **Complete Response**

Three parts essentially correct

3 **Substantial Response**

Two parts essentially correct and one part partially correct

OR

Part (a) essentially correct and two parts partially correct

2 **Developing Response**

Two parts essentially correct and no parts partially correct

OR

Part (b) or part (c) essentially correct and one or two parts partially correct

OR

Three parts partially correct

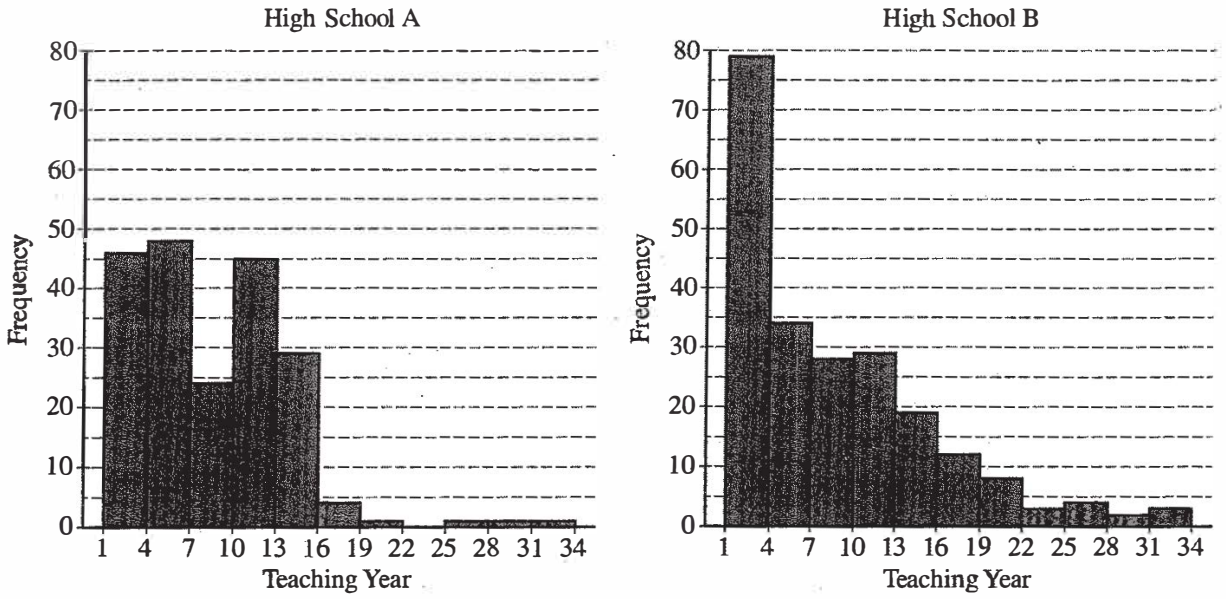
1 **Minimal Response**

One part essentially correct

OR

No parts essentially correct and one or two parts partially correct

5. 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.

High school A has 200 teachers so its median should be the mean of the 100th and 101st values when the values are ordered. Based on the chart, about <100 have values below 7 but >100 have values below 10 The median must be 7 at high school A.

Similarly in high school B the median of the 221 teachers will be the 101st value when ordered.

The chart shows there are <111 below 4 and >111 below 7.

The median must be 6 at high school B.

5A2

(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?

The sum of the 200 teachers in the given chart is $\mu_1 n_1 = 8.2 \cdot 200 = 1640$.

The added teachers sum would be $\mu_2 n_2 = 2.5 \cdot 18 = 45$

The total would be 1685 years for 218 teachers.

The mean would be

$$\mu_{1,2} = \frac{1640 + 45}{200 + 18} \approx \boxed{7.73 \text{ years}}$$

(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.

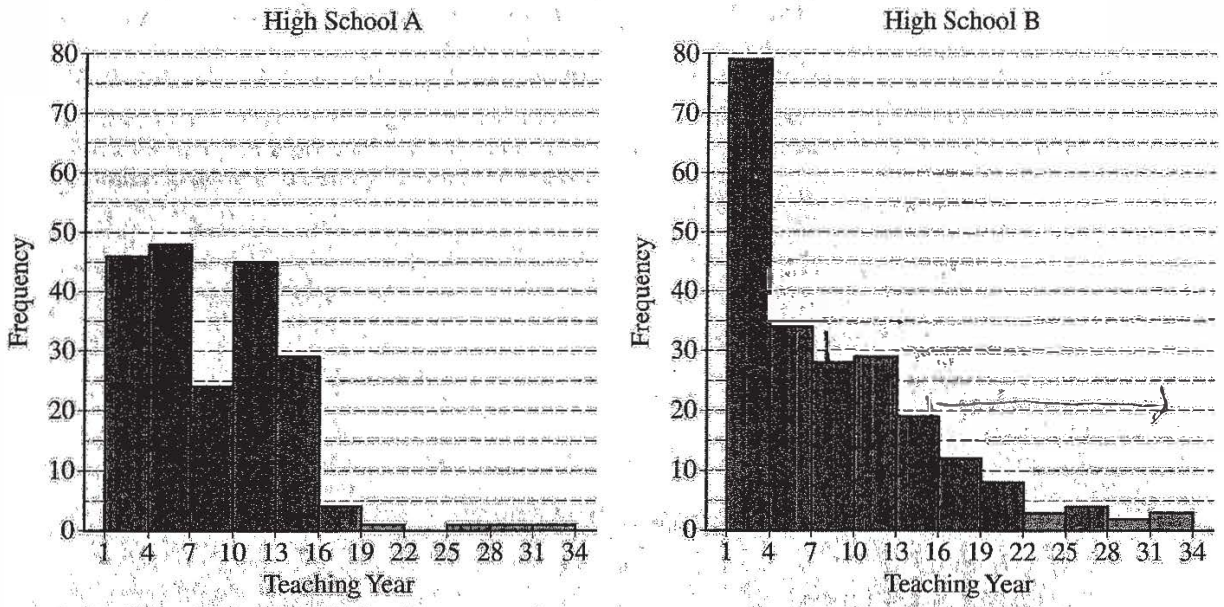
Falling with 7.2 years of 8.2 and being integral means the value must be between $[1, 15]$, or equivalent $[1, 16)$.

This is the sum of values in the first 5 bars: $79 + 34 + 26 + 29 + 19 = 189$ values < 16

The proportion is $\frac{189}{221} \approx \boxed{0.855}$

5B1

5. 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.

By looking at the graph, School B is skewed to the right more stronger than school A. When distributions are more skewed, then pull the mean towards the tail more. Since school B is more skewed, there a greater difference between mean and median in school B. And since they both have the same mean, school B will have the lower median of 6 and school A will have the higher median of 7.

GO ON TO THE NEXT PAGE.

(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?

$\frac{(200)(8.2) + (18)(2.5)}{218} = 7.73$

with the 18 additional teachers the mean teaching year decreases to 7.73.

(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.

8.2 ± 7.2

$(1, 15.4)$ ← interval the teacher must fall in.

$1 - P(\text{year} > 15.4) = 1 - \frac{51}{221}$ ← There are about 51 teachers with teaching year above 15.4

$= 1 - .2308$

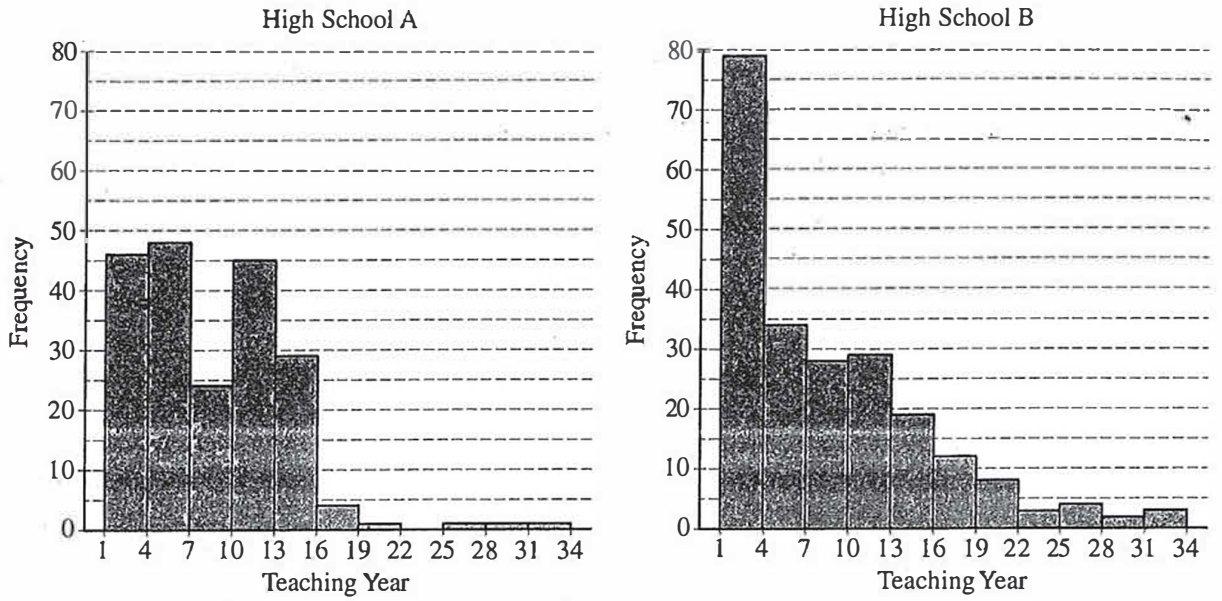
$= .7692$

There is about a 76.92% chance that a random teacher will have a teaching year within 1 standard deviation of the mean of 8.2

GO ON TO THE NEXT PAGE.

5C1

5. 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.

High School A has the median teaching year of 7 and High School B has the median teaching year of 6. According to the graphs. High School A has 46+48=94 teachers with 1 to 6 teaching year. Since High School A has 200 teachers recorded the median is between the 100th and 101th teacher which both fall in the range of 7-9 teaching years. so High School A has median teaching year of 7. According to the graphs. High School B has 79+34=113 teachers with 1 to 6 teaching year. Since High School B has 221 teachers recorded. the median is the 111th teacher which falls in the range of 4-6 teaching years. so High School B has median teaching year of 6.

GO ON TO THE NEXT PAGE.

5C2

(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?

$\mu_A = \frac{8.2 \times 200 + 18 \times 2.5}{218} = \frac{1685}{218} = 7.73$ years.

(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.

$\mu = 8.2$ yrs $\sigma = 7.2$ yrs

data within 1 standard deviation of the mean: (1, 15.4)

normalcdf (1, 15.4, 8.2, 7.2) = 0.683

Since the graph is strongly skewed to the right, the predicted probability may be inaccurate

AP® STATISTICS
2018 SCORING COMMENTARY

Question 5

Overview

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.

Sample: 5A
Score: 4

In part (a) the response correctly identifies the location of the median for High School A as "the mean of the 100^{th} and 101^{th} values when the values are ordered" and the location of the median for High School B as "the 111^{th} value when ordered," which satisfies component 3. For High School A, the response indicates "about < 100 have values below 7 and > 100 have values below 10," which satisfies component 2. The response correctly concludes that "the median must be 7 at high school A," which satisfies component 1. For High School B, the response correctly counts that there are "< 111 [values] below 4 and >111 [values] below 7," which again satisfies component 2. The response correctly concludes that "The median must be 6 at high school B," which again satisfies component 1. Because the response satisfies all three components, part (a) was scored as essentially correct. In part (b) the response uses the correct weights to compute the total number of teaching years for the initial 200 teachers (1,640) and for the 18 additional teachers (45), which satisfies component 2. The response adds these values, divides the total number of teaching years by 218, and reports the correct mean, which satisfies component 1. Because the response includes both components, part (b) was scored as essentially correct. In part (c) the response provides two appropriate intervals. The response points out that because the data values are recorded as integers, if a value is within one standard deviation of the mean, the value will be in the interval [1, 15] or equivalently [1, 16), which satisfies component 1. The response states that the appropriate number of data values "is the sum of values in the first 5 bars" and computes the correct sum, which satisfies component 2. The sum is divided by the correct denominator (221), and the correct probability is reported, which satisfies component 3. Because the response includes all three components, part (c) was scored as essentially correct. Because three parts were scored as essentially correct, the response earned a score of 4.

Sample: 5B
Score: 3

In part (a) the response correctly identifies the median for the distribution of teaching years for High School A as 7 and the median for the distribution of teaching years for High School B as 6, which satisfies component 1. The response bases the decision on the distribution of High School B being "skewed to the right much stronger than school A," which satisfies component 2. The response satisfies component 3 when it points out that in a skewed distribution the mean will be pulled towards the tail and states that "Since school B is more skewed there is a greater difference between the mean and median in school B." Because the response includes all three components, part (a) was scored as essentially correct. In part (b) the response provides an equation with the correct weighted sum of means in the numerator divided by the total number of teachers and reports the correct mean, which satisfies component 1 and component 2. Because the response includes both components, part (b) was scored as essentially correct. In part (c) the response reports the correct interval supported by appropriate calculations, which satisfies component 1. The probability that the randomly selected teaching year will be more than one standard deviation from the mean is calculated and subtracted from one. In this calculation the denominator is correct, which satisfies component 3. However, the numerator is incorrect (51 instead of 32) because when summing frequencies for observations with "year >15.4" the response includes those values in the interval from 13 to 16; therefore,

Question 5 (continued)

component 2 is not satisfied. Because the response satisfies only two of the three components, part (c) was scored as partially correct. Because two parts were scored as essentially correct, and one part was scored as partially correct, the response earned a score of 3.

Sample: 5C
Score: 2

In part (a) the response correctly identifies the median for the distribution of teaching years for High School A as 7 and the median for the distribution of teaching years for High School B as 6, which satisfies component 1. For High School A, the response correctly counts that there are "94 teachers with 1 to 6 teaching year" and that the "median is between the 100th and 101th teacher" and concludes that the median is 7, which satisfies component 2 and component 3. For High School B, the response correctly counts that there are "113 teachers with 1 to 6 teaching year" and that "the median is the 111th teacher" and concludes that the median is 6, which again satisfies component 2 and component 3. Because the response includes all three components, part (a) was scored as essentially correct. In part (b) the response provides an equation with the correct weighted sum of means in the numerator divided by the total number of teachers and reports the correct mean, which satisfies component 1 and component 2. Because the response includes both components, part (b) was scored as essentially correct. In part (c) the response computes the correct interval, which satisfies component 1. The response uses the normal distribution and computes an incorrect probability of 0.683. The response recognizes that the solution may not be correct and comments that "Since the graph is strongly skewed to the right the predicted probability may be inaccurate." However, because the response uses the normal distribution, part (c) was scored as incorrect. Because two parts were scored as essentially correct, and one part was scored as incorrect, the response earned a score of 2.