2017 AP STATISTICS: Question 4

The intent of this question was to assess:

1) **Students' ability to apply descriptive statistics to a real life scenario, and compare similarities and differences.**
2) **Use graphs and statistical analyses to determine which is the most likely site from which a piece of pottery came and justify why it came from that site AND not the other two.**
3) **Choose the best chemical and justify why you chose the chemical AND not the other two.**

To access the question, please visit https://secure-media.collegeboard.org/ap/pdf/ap-statistics-frq-2017.pdf. Be sure to read the College Board Solutions as well as this commentary.

a. Part a required the students to look at the information for chemical Z and describe how the percentages found in the pieces of pottery are similar and how they are different by looking at the three sites.

A correct solution indicates the median value for the percent of chemical Z in the pottery pieces is similar for all three sites and is about 7%. The ranges for the percents of chemical Z at the 3 sites are very different with the smallest range being 2% (from 6% to 8%), at site II, and the largest range of about 8% at site III (from about 3% to 11%) with site I being the middle range of 6% (from about 4 % to 10%). So the ranges of chemical Z at the sites differed a lot, but the median value at the sites was about the same.

b. Part b (part i) required a student to determine the most likely location of a piece of pottery, based on the sum of the percentages of the three chemicals being 20.5%. Justification was required.

	Site I		Site II		Site III	
Chemical	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
X	**6**	**8**	**5**	**7**	**5**	**7.5**
Y	**11**	**15**	**1.9**	**4**	**6**	**8**
Z	**4**	**10**	**6**	**8**	**3**	**11**
Sum	**21**	**33**	**12.9**	**19**	**14**	**26.5**

If the minimum and maximums of each of the chemicals are placed in a chart and added, it can be determined that 20.5% most likely will fall between the min and maximum at Site III.

a. Part b (part ii) also asks the student to identify the chemical, X, Y o1 Z that would be most helpful in identifying the site at which the pottery originated. Justification for the answer was required also.
Distribution Y would be the most helpful because the sites' distributions of the percentages of chemical Y do not overlap. Both chemicals X and Z have significant overlap in the distributions of percentages.

SCORING

There were 3 sections. Each section was essentially correct (E), partially correct (P) or incorrect (I).

Section 1: Part a:

Essentially correct if all 3 components are met:

1. Identifies that the medians or centers are almost the same for all three sites.
2. Recognizes that the variability (ranges, IQRS, spread) is different across all three sites
3. Context is used

Partially correct (P) if only 2 of the 3 components are used
Incorrect (I) if at most one component is used.

Notes:

- Comments about shape should be ignored because complete shape information cannot be obtained from a boxplot
- Responses are not required to give numerical value. Any reasonable approximations from the boxplots are acceptable.
- Because the boxplots are symmetrical, it is acceptable to use means instead of medians.
- Any discussion of chemical X and chemical Y is considered extraneous
- Context may be obtained by using one of the words: site, chemical, weight, X, Y or Z.

Section 2: Part b i.

Essentially correct if all 3 components are met:

1. Site III is the chosen site.
2. Sums of the minimums and maximums for each of the sites and each of the chemicals are computed
3. A reasonable numerical justification is given involving the sums of a statistical measure (obtained in part 2) of the 3 chemicals in order to choose Site III.

Partially correct (P) if only 2 of the 3 components are used.
Incorrect (I) if at most one component is used.

Notes:

- If the response computes the sum of the minimums for Site I and the maximums for Site II and recognizes that this is enough justification, the response is an E.
- If alternative measures are calculated, like the sum of the medians, the second component is not satisfied, but the third component may be satisfied. It could be a P.
    - Ex: A response compares the alternate measure to the other two sites by saying the sum is closest to 20.5 → scored as a P
    - If the response does not have an implicit or explicit comparison, it is scored an I.

- Choosing Site I or Site II as the correct choice is scored as an I.
- Sums of the medians are: Site 1: 27.5; Site 2: 16, Site 3: 20

Section 3: Part b ii.

Essentially correct (E) if the response chooses chemical Y and gives a reasonable justification based on chemical Y's distribution being different/distinctive across the sites

Partially correct (P) if the response chooses chemical and provides justification based on the boxplots, but does not clearly explain that the distributions are distinctively different across the sites.
Or discusses that distribution Y is distinctively different across sites, but never chooses chemical Y as the best choice.

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

Notes:

- To get credit for distinctive, justification must address location and variability of the boxplots (stating they do not overlap for chemical Y does this).
- Choosing Chemical X or Chemical Z or chooses Y with no reasonable justification, the response is an I.
- See more commentary below

COMMENTARY AND TEACHER TIPS

1. **Work needs to go in the answer, not on the graph.**
2. **Know correct terms. When discussing range, it is a value. Range = maximum minus minimum. It is not an interval of numbers and many students interpreted it as such.**
   Part 2
3. **Drawing numbers on the graph does not help with the solution if the numbers are just listed.**
4. **It is best to do the work in a table in the section for which the student is working. It is not appropriate for a student to expect the reader to do work (adding/subtracting etc.) to determine if one site is better than another.**
5. **Some students chose site 3 only based on medians instead of the minimum and maximum. If you only look at the medians, you don't really get side 3. At most the median argument is a P.**
6. **Reasoning must include the reason for the choice. This reason must be statistical in nature and a rationale for not including the other choices must be given. (Many students chose Site III and did not justify why Sites I and II would not be the preferred choice).**

   Part 3
7. **Only 1 chemical could be used, so one might say choose Chemical Y because the percentages don't overlap, and Chemical X and Chemical Y have substantial overlap. This solution has a reason for the choice and reasonable justification for excluding the two not chosen.**
8. **A common problem/difficulty was using terminology indicating that the boxplots did not overlap. They must have indications of different locations and small variability---saying different means alone was not enough.**

9. **Words that were acceptable for boxplots that did not overlap were and were scored an "E":**
    - **Completely different percentages at each site**
    - **Absolutely different percentages at each site**
    - **Distinctly different percentages at each site.**
    - All values of Site I are high, all values of site II are low and all values of site III are in the middle
    - The ranges never intersect for the three sites
    - The boxplots share no data.

    - **However, words like drastically, extremely, radically, vastly or very different could earn at most a "P".**
        - Boxplots vary.
        - Chemical Y varies the most or has the most variation.
        - The variation among/between the sites is the largest.
        - Boxplots are different.
        - Means/medians differ.
        - Means/medians are most variable.
        - There is a difference in percentages of chemical Y at each site.
        - The distributions of percentages is greatest among the sites for chemical Y.

10. **The distributions are different can be interpreted a lot of ways --- no overlap, different means, different ranges, or different shapes. That is why clarification was needed.**
11. **Boxplots of Chemical X and Z overlap while the boxplot of Chemical Y does not (acceptable).**
12. **An incomplete response simply indicates the median/means vary (without further explanation) and could be <u>at most a P</u>.**
    - **Boxplots vary**
    - **Chemical Y varies the most**
    - **Distribution of percentages differs the most among site "A"**
    - **Chemical Y has the most variation between sites**
    - **The variation among/between the sites is the largest.**
    - **Boxplots are different.**
    - **Means/medians differ.**
    - **Means/medians are most variable.**
    - **There is a difference in percentages of chemical Y at each site.**
    - **The distribution of percentages is greatest among the sites for chemical Y.**
    - **Each of the above is incomplete because you must give which one you choose and why, as well as why you did not choose the others.**

OTHER TEACHING TIPS:

1. **Read the question carefully --- ANSWER THE QUESTION. If it asks for similarities and differences, organize your thoughts and be sure to give both.**
2. **When asked to pick a choice, justify your choice and why you didn't choose the others (FOR THIS AND ALL AP QUESTIONS!)**
3. **Use statistics terms correctly.**
4. **Use clear communication within the questions.**
5. **Put appropriate work in each part→ don't expect the reader to go back or hunt for your work.**