

Review Session – AP Statistics – Unit 1

1. The five-number summary for scores on a statistics exam is 11, 35, 61, 70, 79. In all, 380 students took the test. About how many had scores between 35 and 61?
(a) 26 (b) 76 (c) 95 (d) 190 (e) none of these
2. Mr. Yates picked up a dozen items in the grocery store with a mean cost of \$3.25. Then he added an apple pie for \$6.50. The new mean for all 13 items is
(a) \$3.00 (b) \$3.50 (c) \$3.75 (d) \$4.88 (e) none of these
3. A sample of 99 distances has a mean of 24 feet and a median of 24.5 feet. Unfortunately, it has just been discovered that a value was recorded as “30”, but actually has a value of “35.” If we make this correction to the data, then
(a) The mean remains the same, but the median is increased.
(b) The mean and median remain the same.
(c) The median remains the same, but the mean is increased.
(d) The mean and median are both increased.
(e) We do not know how the mean and median are affected without further calculations.
4. A clothing and textiles student is trying to assess the effect of a jacket’s design on the time it takes preschool children to put the jacket on. In a pretest, she times 7 children as they put on her prototype jacket. The times (in seconds) are provided below.

n	n	65	39	n	43	102
---	---	----	----	---	----	-----

The n’s represent children who had not put the jacket on after 120 seconds (in which case the children were allowed to stop). Which of the following would be the best value to use as the “typical” times required to put on the jacket?

- (a) The mean time, which was 62.25 seconds.
(b) The mean time, which was 85.6 seconds.
(c) The median time, which was 54 seconds.
(d) The median time, which was 102 seconds.
(e) The missing times (the n’s) mean we can’t calculate any useful measures of center.
5. A scientist is weighing each of 30 fish. She obtains a mean of 30 g and a standard deviation of 2 g. After completing the weighing, she finds that the scale was misaligned and always under reported every weight by 2 g that is, a fish that really weighed 26 g was reported to weigh 24 grams. What are the mean and standard deviation after correcting for the error in the scale?
(a) 28 g, 2 g (b) 30 g, 4 g (c) 32 g, 2 g (d) 32 g, 4 g (e) 28 g, 4 g
6. What are all the values that a standard deviation can possibly take?
(a) $0 \leq s$ (b) $0 \leq s \leq 1$ (c) $-1 \leq s \leq 1$ (d) $s \leq 0$ (e) any real number
7. A particularly common question in the study of wildlife behavior involves observing contests between “residents” of a particular area and “intruders.” In each contest, the “residents” either win or lose the encounter (assuming there are no ties). Observers might record several variables. Which of the following variables is categorical?
 - a. The duration of the contest (in seconds).
 - b. How long the “intruder” lives in the area before it is accepted as a “resident”.
 - c. Whether the “residents” win or lose.
 - d. The total number of contests won by the “residents”.
 - e. The number of animals involved in the contest.
8. The rental values (in dollars) of a sample of four available apartments close to the university are

470	600	580	550
-----	-----	-----	-----

The standard deviation of the sample is

a. \$30.31

b. \$35

c. \$49.50

d. \$53.33

e. \$57.15

During the early part of the 1994 baseball season, many sports fans and baseball players noticed that the number of home runs being hit seemed to be unusually large. Here are the data on the number of home runs hit by American and National League teams:

American League	35, 40, 43, 49, 51, 54, 57, 58, 58, 64, 68, 68, 75, 77
National League	29, 31, 42, 46, 47, 48, 48, 53, 55, 55, 55, 63, 63, 67

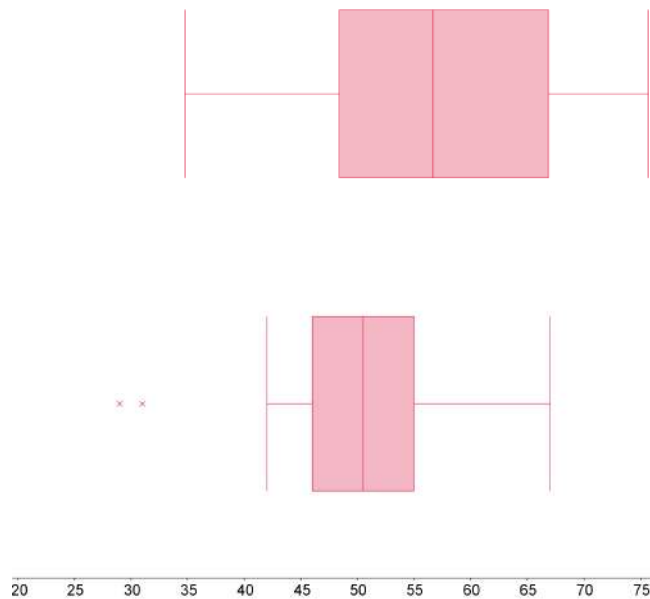
(a) Construct a single graph appropriate for comparing the number of home runs hit in the two leagues.

(b) Are there any outliers in either of the two data sets? Justify your answer numerically. Show all work.

(c) Write a few sentences comparing the distributions of home runs in the two leagues.

ANSWERS

1. **C.** 35 and 61 are the 1st quartile and median, respectively, which means that 25% of all the data is between them. So 25% of all the students are between 35 and 61. The problem says there are 380 students, and 25% of that is **95**.
 2. **B.** Remember that $\text{mean} = \text{sum} / \#$. If a dozen items have a mean of \$3.25, then their sum must be $12 \times 3.25 = 39$. Then Mr. Yates adds another 6.50, making the new sum $39 + 6.50 = 45.5$. There are now 13 items, so the new mean would be $45.5 / 13 = 3.5$.
 3. **C.** If a number has a value of "30", it must be at the high end of the data set, because the mean and median are 24 and 24.5, respectively. So if we change that high number (30), to an even higher number (35), it would pull the mean even further up towards that new higher number. The median would stay the same, though, because the median is resistant to outliers. So the answer is C.
 4. **D.** If "n" represents children who couldn't put on a jacket after 120 seconds, then "n" must have a value greater than 120, making it a big outlier. So for starters, we should use the median, because there are outliers. Cross out A and B. Now, there are 7 total numbers, so the 4th number would be the median. If you arrange the numbers from lowest to highest, with "n" at the high end because it's a big outlier, then **102** is the median.
 5. **C.** To begin with, the answer must be A or C. If you change all the numbers in a data set, in this case by adding 2, you don't change the standard deviation, because all the numbers are still just as spread out as they were before. So the standard deviation is still 2, as stated in the first line of the problem. Now, if the machine underreports data, then in the example it says a fish is 24 when the truth is it's 26. So if the machine is saying fish are 30, then the truth must be 32. The answer is C.
 6. **A.** Standard deviation is a measure of all the numbers' average distance to the mean. So it's a measure of distance. And distance can't be negative. So the answer is A.
 7. **C.** Categorical means it can't be counted with numbers. C is the only viable answer. In A, you're counting in seconds. B, you're counting in time (days maybe). D, number of contests. E, number of animals. If you're counting, it's quantitative.
 8. **E.** Type the numbers into L_1 of your calculator, then do Stat:Calc:#1 to get all the numerical calculations for your data. The standard deviation is either S_x or σ_x . In this case, the problem states it's a "sample" of apartments, so use S. You should use σ if it's a population.
- a. See below. You could also make a side-by-side stem plot. Make sure to label your axes.



- b. There are two outliers at the low end of the National League, 29 and 31. $1.5 \times \text{IQR} = 1.5(9) = 13.5$, so the lower limit is 32.5.
- c. The shapes seemed fairly symmetric, although the National League was skewed left with two low outliers, as identified by the $1.5 \times \text{IQR}$ rule. The American League has a higher median of approximately 56, compared to National League's median of approximately 51. There was also greater spread in the American League, with a standard deviation of 12.69 compared to 11.13. The American League teams seemed to hit many more home runs in 1994, with more variability.