**AP Statistics Ch 5 - Extra Practice** **Name:** Solutions **Per:** ____

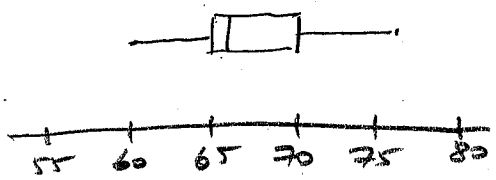
**8. Singers.** The frequency table shows the heights (in inches) of 130 members of a choir.

| L1 Height | L2 Count | L1 Height | L2 Count |
|---|---|---|---|
| 60 | 2 | 69 | 5 |
| 61 | 6 | 70 | 11 |
| 62 | 9 | 71 | 8 |
| 63 | 7 | 72 | 9 |
| 64 | 5 | 73 | 4 |
| 65 | 20 | 74 | 2 |
| 66 | 18 | 75 | 4 |
| 67 | 7 | 76 | 1 |
| 68 | 12 | | |

a) 1Varstats L1,L2

| min | Q1 | Med | Q3 | Max |
|---|---|---|---|---|
| 60 | 65 | 66 | 70 | 76 |

b)

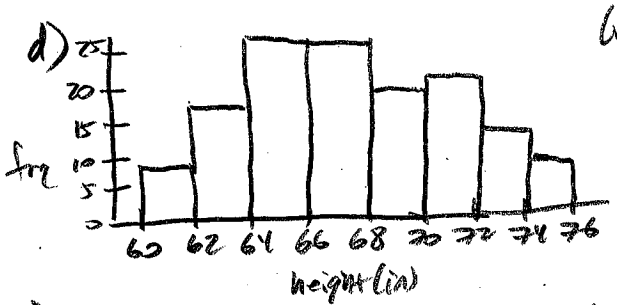



c) $\bar{x} = 67.115$ in

$s = 3.792$ in

a) Find the 5-number summary for these data.
b) Display these data with a boxplot.
c) Find the mean and standard deviation.
d) Display these data with a histogram.
e) Write a few sentences describing the distribution of heights.

d) (Use calc w/ Freq: L2) and sketch




e) The choir heights are unimodal, and fairly symmetrical with a mean of 67.115 in and standard deviation of 3.8 in. There appear to be no outliers.

**12. Wayne Gretzky.** In Chapter 4 (Exercise 12) you examined the number of games played by hockey great Wayne Gretzky during his 20-year career in the NHL. Here is the stem-and-leaf display:

| | |
|---|---|
| 8 | 000000122 |
| 7 | 8899 |
| 7 | 0344 |
| 6 | |
| 6 | 4 |
| 5 | |
| 5 | |
| 4 | 58 |
| 4 | |

Games Played (4 | 5 = 45 games)

a) Would you use the median or the mean to describe the center of this distribution? Why?
b) Find the median.
c) Without actually finding the mean, would you expect it to be higher or lower than the median? Explain.

a) median due to skew and possible outliers

b) (counting values) 79 games is the median

c) The mean will be lower because the low skew outliers will pull the mean down.

**14. Gretzky returns.** Look once more at the stem-and-leaf display of hockey games played each season by Wayne Gretzky, seen in Exercise 12.
a) Find the range.
b) Find the interquartile range.
c) Using the Outlier Rule, explain why the two seasons when Gretzky played only 45 and 48 games could be considered outliers.
d) Do you consider the 64-game season an outlier, too? Explain.

a) range = max − min
= 82 − 45 = 37 games

b) Q1 = 73.5
Q3 = 80
IQR = 80 − 73.5 = 6.5 games

c) UF = Q3 + 1.5(IQR) = 80 + 1.5(6.5) = 89.75
no values above this (no high outliers)
LF = Q1 − 1.5(IQR) = 73.5 − 1.5(6.5) = 63.75
so only the 45 and 48 values are below this.

d) Possibly, it seems separated from the 'bulk' of the data and is right at the edge of numerically being an outlier. I would say yes, it is an outlier.

**16. Camp sites.** Shown below are the histogram and summary statistics for the number of camp sites at public parks in Vermont.

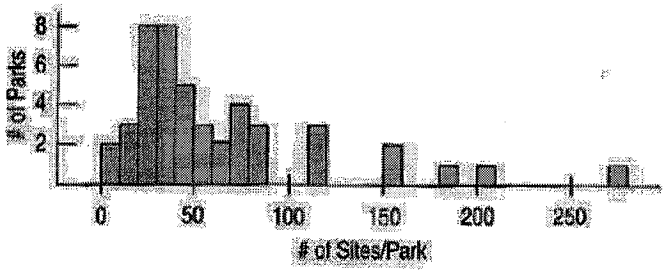
a) Which statistics would you use to identify the center and spread of this distribution? Why?

b) How many parks would you classify as outliers? Explain.

c) Create a boxplot for these data.

d) Write a few sentences describing the distribution.

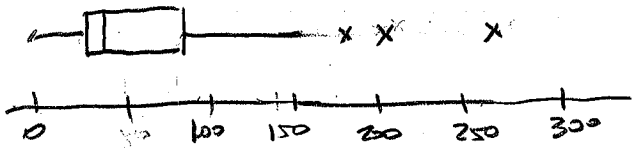
| | |
|---|---|
| Count | 46 |
| Mean | 62.8 sites |
| Median | 43.5 |
| StdDev | 56.2 |
| Min | 0 |
| Max | 275 |
| Q1 | 28 |
| Q3 | 78 |




a) use median of 43.5 due to skew/outliers

b) IQR = 78 − 28 = 50
no low outliers, Checking UF:
UF = 78 + 1.5(50) = 153
Definitely 3, possibly 2 more (right at edge)

c)

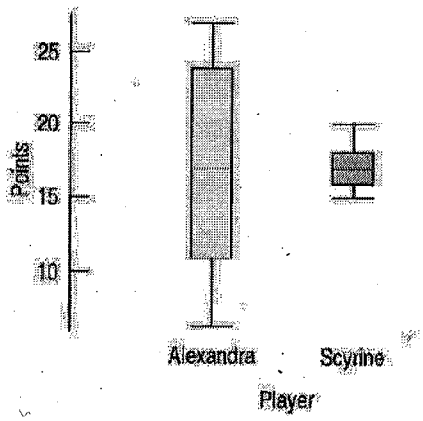



d) Site distribution is unimodal, skewed right, with a median of 43.5 sites and IQR of 50 sites. There are at least 3 high outliers: around 185, 205, and 275 sites.

**19. Women's basketball.** Here are boxplots of the points scored during the first 10 games of the season for both Scyrine and Alexandra.

a) Summarize the similarities and differences in their performance so far.

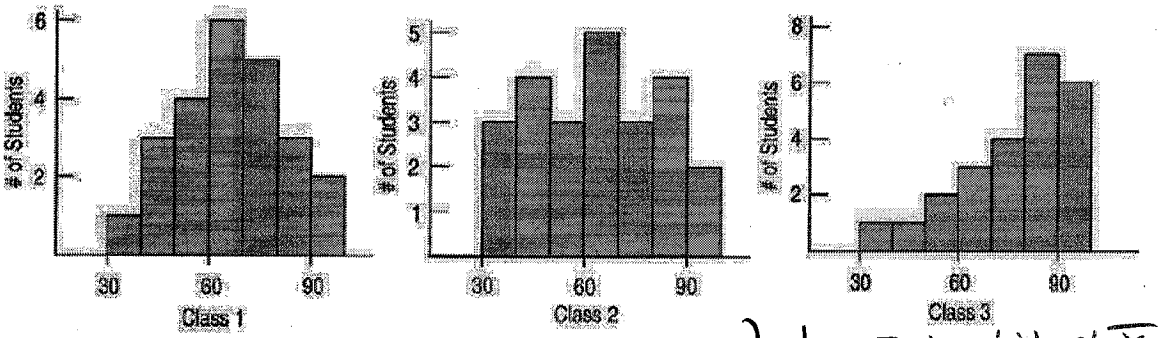
b) The coach can take only one player to the state championship. Which one should she take? Why?




a) Both players typically score around 16 points, but Scyrine is much more consistent (IQR of 3 points, vs. 13 for Alexandra). But Alexandra sometimes has very high scoring games, 28 points compared to max of 20 for Scyrine.

b) Depends upon the situation. If 15–20 points is good enough for a win, take Scyrine for consistency. If an underdog and need a 28 point game to win, take Alexandra.

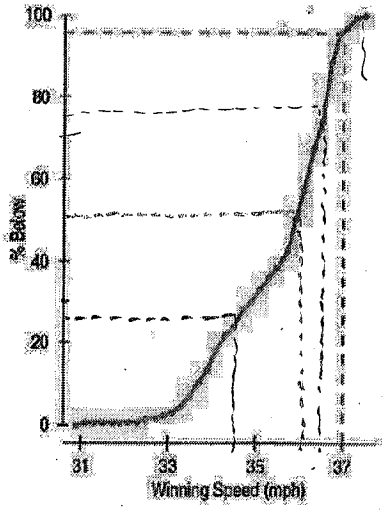
**27. Test scores.** Three Statistics classes all took the same test. Histograms of the scores for each class are shown below.




a) Which class had the highest mean score?
b) Which class had the highest median score?
c) For which class are the mean and median most different? Which is higher? Why?
d) Which class had the smallest standard deviation?
e) Which class had the smallest IQR?

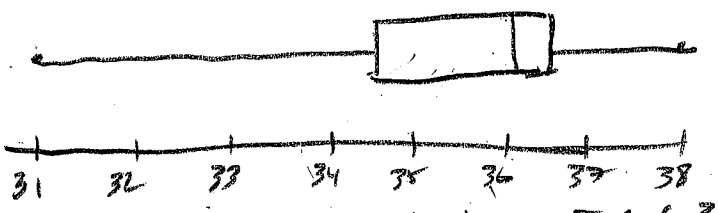
a) class 3 has highest $\bar{x}$.
b) class 3 has highest median.
c) Class 3. The others are fairly symmetrical, so median and $\bar{x}$ close. For class 3, mean pulled lower by skewed toward tail.
d) class 1 has most data near the mean. (smallest avg dist from mean)
e) class 1 (but class 3 might be close)

**35. Derby speeds.** How fast do horses run? Kentucky Derby winners top 30 miles per hour, as shown in the graph on the next page. In fact, this graph shows the percentage of Derby winners that have run *slower* than a given speed. Note that few have won running less than 33 miles per hour, but about 95% of the winning horses have run less than 37 miles per hour. (A cumulative frequency graph like this is called an "ogive.")



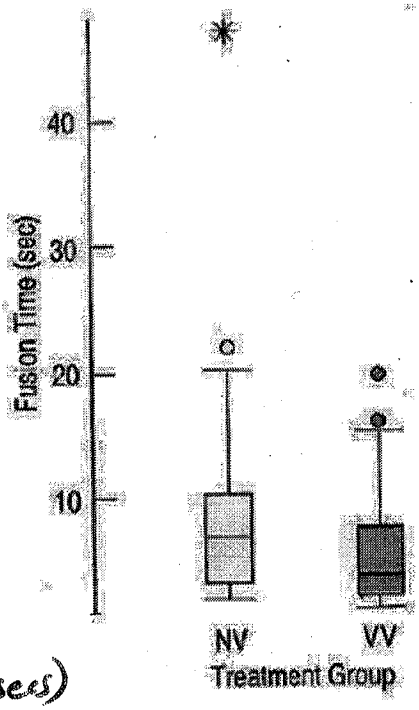

a) Estimate the median winning speed.
b) Estimate the quartiles.
c) Estimate the range and the IQR.
d) Create a boxplot of these speeds.
e) Write a few sentences about the speeds of the Kentucky Derby winners.

a) 36mph
b) Q1 = 34.5mph
Q3 = 36.5mph
LF = 34.5 - 1.5(2) = 30
c) range = 38 - 31 = 7mph
IQR = 36.5 - 34.5 = 2mph
d)




e) Winning speeds are skewed left, with a median of 36mph and IQR of 2mph. There appear to be no outliers.

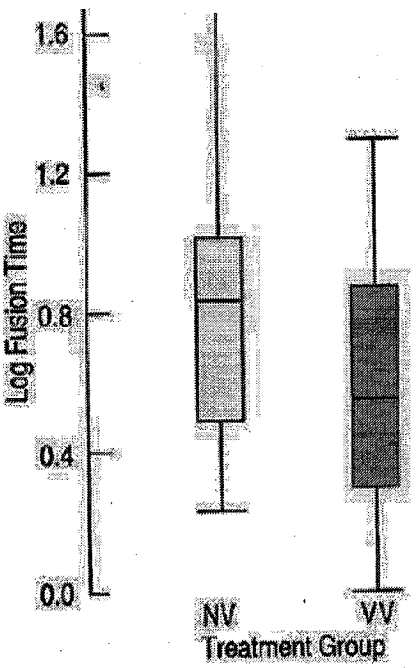
**48. Stereograms.** Stereograms appear to be composed entirely of random dots. However, they contain separate images that a viewer can "fuse" into a three-dimensional (3D) image by staring at the dots while defocusing the eyes. An experiment was performed to determine whether knowledge of the embedded image affected the time required for subjects to fuse the images. One group of subjects (group NV) received no information or just verbal information about the shape of the embedded object. A second group (group VV) received both verbal information and visual information (specifically, a drawing of the object). The experimenters measured how many seconds it took for the subject to report that he or she saw the 3D image.




a) What two variables are discussed in this description?
b) For each variable, is it quantitative or categorical? If quantitative, what are the units?
c) Here are boxplots comparing the fusion times for the two treatment groups. Write a few sentences comparing these distributions. What does the experiment show?

a/b) knowledge (categorical), fusion time (numerical, secs)
c) Both distrib. are skewed right, w/ high outliers. VV group has lower times typically (median 4 sec vs. 6 sec for NV), VV group is also more consistent (IQR 6 sec vs. 9 sec for NV)

**49. Stereograms, revisited.** Because of the skewness of the distributions of fusion times, we might consider a re-expression. Here are the boxplots of the *log* of fusion times. Is it better to analyze the original fusion times or the log fusion times? Explain.




The log fusion times make it easier to see the difference in medians, and reduce skew and outliers.