

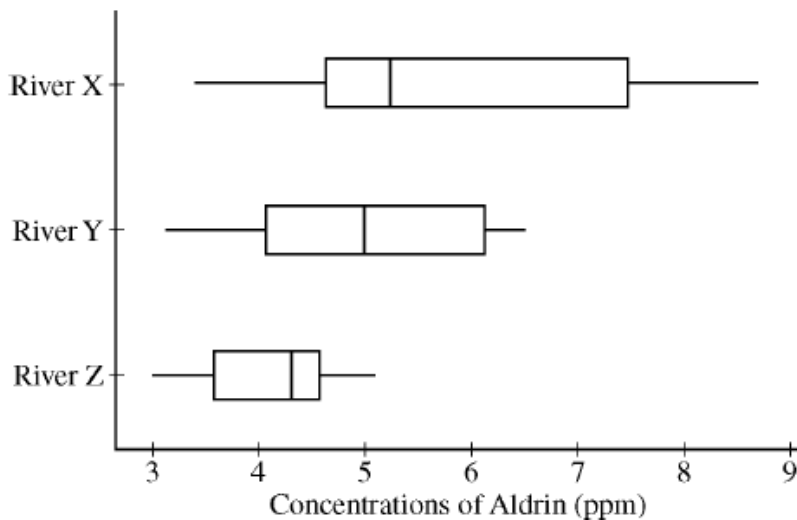
AP Statistics Exam Review

Comparing Boxplots

NAME _____

As a part of the United States Department of Agriculture's Super Dump cleanup efforts in the early 1990s, various sites in the country were targeted for cleanup. Three of the targeted sites—River X, River Y, and River Z—had become contaminated with pesticides because they were located near abandoned pesticide dump sites. Measurements of the concentration of aldrin (a commonly used pesticide) were taken at twenty randomly selected locations in each river near the dump sites.

The boxplots shown below display the five-number summaries for the concentrations, in parts per million (ppm) of aldrin, for the twenty locations that were sampled in each of the three rivers.



(a) Compare the distributions of the concentration of aldrin among the three rivers.

(b) The twenty concentrations of aldrin for River X are given below.

3.4	4.0	5.6	3.7	8.0	5.5	5.3	4.2	4.3	7.3
8.6	5.1	8.7	4.6	7.5	5.3	8.2	4.7	4.8	4.6

Construct a stemplot that displays the concentrations of aldrin for River X.

(c) Describe a characteristic of the distribution of aldrin concentrations in River X that can be seen in the stemplot but cannot be seen in the boxplot.

Solution

Part (a):

Comparing the medians reveals that the concentration of aldrin tends to be highest for River X and lowest for River Z. About 50 percent of the concentrations of aldrin for Rivers X and Y are higher than all of the concentrations for River Z. River X also displays the most variability in aldrin concentrations, as seen by the largest range and largest IQR, and River Z has the least variability, as judged by both IQR and range. The shapes of the three distributions differ, in that the distribution appears to be skewed to the right for River X, roughly symmetric for River Y and slightly skewed to the left for River Z.

Part (b):

Aldrin concentrations (in ppm) for River X
Leaf unit = 0.1 (for example, 3 | 4 represents 3.4 ppm)

```
3 | 47
4 | 0236678
5 | 13356
6 |
7 | 35
8 | 0267
```

Part (c):

The stemplot shows a gap in the distribution of aldrin concentrations for River X between the values of 5.6 and 7.3 ppm of aldrin. This gap is not apparent in the boxplot.

Scoring

Part (a) is scored as follows:

Essentially correct (E) if the response correctly describes, in context, the center, spread and shape (all three characteristics) of the three distributions *AND* makes a *comparative* statement involving all three distributions for at least one characteristic. Specific numerical values are not required.

Partially correct (P) if the response does not warrant an E, but it includes all three of the following components:

1. Mentions all three distributions
2. Correctly describes at least two of the characteristics (center, spread, shape) of at least two distributions
3. Includes a correct comparison of at least two distributions for at least one characteristic *OR* if the response describes all three characteristics of the three distributions but does not make a *comparison* across distributions.

Incorrect (I) otherwise.

Note: Context is required to earn an E but not to earn a P.

Part (b) is scored as follows:

Essentially correct (E) if a reasonable stemplot that includes a leaf unit key is provided. It is *not* necessary for the key to include measurement units (ppm).

Partially correct (P) if a reasonable stemplot without a leaf unit key is provided.

Incorrect (I) if an unreasonable stemplot or a graph other than a stemplot is provided.

Part (c) is scored as follows:

Essentially correct (E) if the response includes a recognition of the gap in the stemplot *AND* gives an indication of where the gap occurs, *OR* if the response comments on bimodality *AND* specifies where the modes/clusters occur.

Partially correct (P) if the response indicates there is a gap or bimodality in the stemplot but does not give an indication of where the gap occurs.

Incorrect (I) otherwise. For example, the response might indicate that the numerical values can be seen in the stemplot but not the boxplot, or that the mean and standard deviation can be computed with the stemplot but not the boxplot, or only that the distribution is skewed to the right.

Note: The scoring system counts part (a) at *double weight*. In other words, an E counts as 2 points in part (a) and as 1 point in each of parts (b) and (c). Similarly, a P counts as 1 point in part (a) and as 1/2 point in parts (b) and (c).