CS 1173: Hand computation of statistical indicators

In the following, **A** = [2, 4, 2, 3, 5, 1, 4, 3] and **x** = $[x_1, x_2, x_3, \ldots, x_n]$ is a general set of n measurements used in the formulas.

1. The average or mean measures the central tendency of a set of numbers. To calculate the average, add the values and divide by the number of values. The average of **x** is:

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

Example 1: The average or mean of **A** is 3:

a) Sum the elements of **A**: 2 + 4 + 2 + 3 + 5 + 1 + 4 + 3 = 24.

b) Divide by the number of elements in **A**: 24/8 = 3

2. The median is another measure of central tendency. In contrast to the average, the median is not heavily influenced by outliers. When the average and the median are far apart in value, you can surmise that there must be some outliers in the data. To calculate the median, sort the values and take the middle one. If the number of items is even, average the middle two values.

Example 2: The median of **A** is 3:

a) Sort **A** in increasing order: [1, 2, 2, 3, 3, 4, 4, 5].

b) Average of the middle two values since **A** has an even number of elements: (3 + 3)/2 = 3.

3. The mode is the value most frequently appearing in the data. This value may not be unique.

Example 3: A has three most frequent values: 2, 3 and 4. The mode may not be a single value.

a) Compute the frequencies of the elements of **A**: 1:1, 2:2, 3:2, 4:2, 5:1

b) Find the element(s) with the largest count: 2, 3, and 4 all appear twice.

4. The variance measures the mean-squared error (MSE) of the values from the mean. The variance is the average of the squares of the distance of each value from the mean. The variance **x** is:

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2$$

where $\bar{x}$ is the mean of **x**.

Example 4: The variance of **A** is 1.5:

a) First compute the sum of the squares of the differences with the mean:

$(2 - 3)^2 + (4 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (5 - 3)^2 + (1 - 3)^2 + (4 - 3)^2 + (3 - 3)^2 =$

$1 + 1 + 1 + 0 + 4 + 4 + 1 + 0 = 12$

b) Divide by n to obtain the variance: 12/8 = 1.5

5. The standard deviation measures the spread or dispersion of values from the mean. The standard deviation is calculated as the square root of the average of the squares of the distance of each value from the mean. In other words, the standard deviation is the square root of the variance. The standard deviation of **x** is:

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}$$

where $\bar{x}$ is the mean of **x**.

Example 5: The standard deviation **A**:

a) First compute the sum of the squares of the differences with the mean:
$(2 - 3)^2 + (4 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (5 - 3)^2 + (1 - 3)^2 + (4 - 3)^2 + (3 - 3)^2 =$
$1 + 1 + 1 + 0 + 4 + 4 + 1 + 0 = 12$

b) Divide by n and take the square root: $\sqrt{12/8} = 1.225$

6. The p^{th} percentile is the data value such that p% of the data fall below that value. The 50% percentile is the median. The 25^{th} percentile is sometimes called the first quartile.

Example 6: The 25^{th} percentile of **A** is 2, and the 75^{th} percentile of **A** is 4.

a) Sort **A** in increasing order: [1, 2, 2, 3, 3, 4, 4, 5]

b) Compute the percentiles for the data: 0.5*(100/8), 1.5*(100/8), ..., 7.5*(100/8) =
1:6.25, 2:18.75, 2:31.25, 3:43.75, 3:56.25, 4:68.75, 4:81.25, 5:93.75

c) Linearly interpolate the two closest values. For the 25th percentile, we interpolate between 2 at the 18.75^{th} percentile and 2 at the 31.25^{th} percentile, so the answer is 2.

7. The inter-quartile range (IQR) is the 75^{th} percentile minus the 25^{th} percentile.

Example 7: What is the inter-quartile range of **A**?

Since the 25^{th} percentile of **A** is 2 and the 75^{th} percentile of **A** is 4, the inter-quartile range is 2.

Example 8: What is the position of the fence for a box plot of **A**?

The top of the fence is at the 75^{th} percentile plus $1.5 \times IQR = 4 + 1.5 \times 2 = 4 + 3 = 7$

The bottom of the fence is at the 25^{th} percentile minus $1.5 \times IQR = 2 - 3 = -1$

Example 8: What are the positions of the whiskers for a box plot of **A**?

All of the values of **A** are inside the fence, so the top whisker is at 5 and the bottom whisker is at 1.