**Activity 7.2.2 – Measuring Variability in Data Distributions**

This activity focuses on data from an observational study of life expectancies in countries throughout the world. The dotplot below summarizes life-expectancy data for 188 countries. Each dot details a country’s average life expectancy for the year 2015. For example, the 2015 average life expectancy for the United States is 79 years. This means that the average person born in the United States in 2015 is predicted to live for 79 years.



The stacked dotplot below compares life expectancies for people in 53 countries in Africa and 43 countries in Asia.



1. Compare the two data sets in terms of their center, shape, and spread. What conclusions can we make about life expectancies of people in African and Asian countries based on these data?

**Measuring Variability**

We will explore two important measures of variability: *mean absolute deviation* and *standard deviation*, by focusing on a subset of the African and Asian countries. The stacked dotplot compares life expectancies from a random sample of countries from each region. The tables below display the actual data values.



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample of African Countries**

|  |  |
| --- | --- |
| **Country** | **Life Expectancy** |
| Algeria | 76.5 |
| Djibouti | 64.6 |
| Kenya | 66.6 |
| Senegal | 66.1 |
| Swaziland | 51.5 |
| Zambia | 59.0 |

 | **Sample of Asian Countries**

|  |  |
| --- | --- |
| **Country** | **Life Expectancy** |
| Afghanistan | 57.6 |
| Malaysia | 75.1 |
| Nepal | 71.2 |
| Philippines | 70.2 |
| Turkmenistan | 67.9 |
| Uzbekistan | 70.1 |

 |

**Mean Absolute Deviation**

The mean absolute deviation (MAD) is a useful measure for understanding how data values in a data set vary. The mean absolute deviation is the average distance of data values from the mean. Calculating this measure requires that we determine each data value’s *deviation* from the mean.

**Procedure for Calculating Mean Absolute Deviation**

1. Find the mean of the data set
2. Find each data value’s deviation from the mean
3. Take the absolute value of each deviation
4. Find the mean of the absolute values of the deviations
5. Complete the following table to calculate the mean absolute deviation for the sample of African countries. The first column contains the data values, ordered from least to greatest. Fill in the middle column with the deviation of each data value; this is found by subtracting the mean from the data values. Fill in the third column with the absolute value of each deviation. Round to the nearest tenth.

**Sample:** African Countries’ Life Expectancies, *n* = 6

|  |  |  |
| --- | --- | --- |
| $$x$$ | $$x-\overbar{x}$$ | $$\left|x-\overbar{x}\right|$$ |
| 51.5 |  |  |
| 59.0 |  |  |
| 64.6 |  |  |
| 66.1 |  |  |
| 66.6 |  |  |
| 76.5 |  |  |

Notice that the deviations contain positive and negative values. The mean of a data set has the unique property that the sum of deviations from the mean add up to 0. Taking the absolute value of the deviations makes all deviations positive; the mean of these positive deviations represents the typical *deviation* (or distance) from the mean.

1. Find the mean absolute deviation for the sample of African countries based on the table above.
2. Complete the following tables to calculate the mean absolute deviation for the sample of Asian countries. Round to the nearest tenth.

**Sample:** Asian Countries’ Life Expectancies, *n* = 6

|  |  |  |
| --- | --- | --- |
| $$x$$ | $$x-\overbar{x}$$ | $$\left|x-\overbar{x}\right|$$ |
| 57.6 |  |  |
| 67.9 |  |  |
| 70.1 |  |  |
| 70.2 |  |  |
| 71.2 |  |  |
| 75.1 |  |  |

1. Find the mean absolute deviation for the sample of Asian countries based on the table above.
2. Compare the mean absolute deviations for these two samples. What does this tell us?

**Standard Deviation**

The standard deviation (SD) is the most commonly used measure for describing the variability of a data set. Similar to the mean absolute deviation, the standard deviation describes the typical deviation from the mean of a set if data. The key difference is in how the measure is calculated. Rather than taking the mean of the absolute values, the standard deviation is the square root of the “average” of the squared deviations.

**Procedure for Calculating Standard Deviation**

1. Find the mean of the data set
2. Find each data value’s deviation from the mean
3. Square each deviation
4. Find the sum of the squared deviations (this is called the sum of squares)
5. Find the average of the squared deviations. For a **sample** data set, divide the sum of squares by the sample size minus 1 ($n-1)$. For a **population** data set, divide the sum of squares by the population size (*N*).
6. Take the square root of the previous result from Step 5.
7. The following tables provide the data values and deviations from the mean for each data value. Fill in the third column in each table by squaring the deviations.

|  |  |  |
| --- | --- | --- |
| $$x$$ | $$x-\overbar{x}$$ | $$\left(x-\overbar{x}\right)^{2}$$ |
| 57.6 | -11.1 |  |
| 67.9 | -0.8 |  |
| 70.1 | 1.4 |  |
| 70.2 | 1.5 |  |
| 71.2 | 2.5 |  |
| 75.1 | 6.4 |  |

**Sample:** Life Expectancies – Africa **Sample:** Life Expectancies – Asia

 $\overbar{x}=64.1$ years $\overbar{x}=68.7$ years

|  |  |  |
| --- | --- | --- |
| $$x$$ | $$x-\overbar{x}$$ | $$\left(x-\overbar{x}\right)^{2}$$ |
| 51.5 | -12.6 |  |
| 59 | -5.1 |  |
| 64.6 | 0.5 |  |
| 66.1 | 2 |  |
| 66.6 | 2.5 |  |
| 76.5 | 12.4 |  |

1. Find and interpret the sample standard deviation for the sample of African countries.
2. Find and interpret the sample standard deviation for the sample of Asian countries.

**Formulas for Standard Deviation**

|  |  |  |
| --- | --- | --- |
| **Sample** Standard Deviation$$s=\sqrt{\frac{\sum\_{}^{}\left(x-\overbar{x}\right)^{2}}{n-1}}$$where $\sum\_{}^{}\left(x-\overbar{x}\right)^{2}$ represents the sum of the squared deviations. |  | **Population** Standard Deviation$$σ=\sqrt{\frac{\sum\_{}^{}\left(x-μ\right)^{2}}{N}}$$where $\sum\_{}^{}\left(x-μ\right)^{2}$ represents the sum of the squared deviations. |

**Range Rule of Thumb**

The standard deviation of a data set can be approximated by the formula: $\frac{range}{4}$. This formula is most appropriate when you do not know the actual data values in a data set.