Section 2.4, Measures of Variation

In this section, we will learn several ways to describe how spread out the data is.

1 Range, Variance, and Standard Deviation

The **range** is the difference between the maximum and minimum entries in the data set.

Range = (Maximum data entry) − (Minimum data entry)

The **deviation** of an entry x in a sample data set is the difference between the entry and the mean $\bar{x}$:

Deviation of $x = x - \bar{x}$.

The sample variance of a data set is:

$$\text{Sample variance} = s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

The sample standard deviation is the square root of the variance:

$$\text{Sample standard deviation} = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

Similar calculations can be done with a population data set, but then you divide by the size of the population (N) instead of one less than the size of the sample ($n-1$).

For grouped data from a frequency distribution, we can approximate the standard deviation with:

$$\text{Sample standard deviation} = s \approx \sqrt{\frac{\sum(x-\bar{x})^2 f}{n-1}},$$

where x is the midpoint of each class.

Example
Calculate the range, sample variance, and sample standard deviation for the following data from a sample:
10 12 12 17 19 20
First, we will calculate the range. Since the data is in order, we just subtract the first from the last: Range $= 20 - 10 = 10$.

Now, to calculate the variance, we first need the mean:

$$\bar{x} = \frac{10+12+12+17+19+20}{6} = 15$$

Now, we use the formula:

$$s^2 = \frac{(10-15)^2+(12-15)^2+(12-15)^2+(17-15)^2+(19-15)^2+(20-15)^2}{5}$$
$$= \frac{88}{5} = 17.6$$

Now, to find the standard deviation, we only need to take the square root of the variance:

$$s = \sqrt{17.6} = 4.195$$

2 The Empirical Rule

Using the standard deviation, we can learn a lot about symmetric bell-shaped distributions:

- About 68% of the data will lie within one standard deviation of the mean.
- About 95% of the data will lie within two standard deviations of the mean.
- About 99.7% of the data will lie within three standard deviations of the mean.

Using this information we can predict where data will lie.

Example
The weight of 64 female college athletes is roughly bell-shaped with a mean $\bar{x} = 133$ and standard deviation $s = 17$. In what range do about 95% of the weights fall?
By the empirical rule, about 95% of the weights will fall within $2s = 34$ of 133, so between 99 and 167.