Section G
Measures of Spread

Another type of descriptive summary is called measures of spread or measures of variation. Measures of spread summarize the data in a way that shows how scattered the values are from each other and how much they differ from the mean value. Just as there are different measures of center there are different measures of spread such as range, variance, and standard deviation. Note, two data sets can have the same mean, median or mode, but be very different in their measure of spread, therefore it important to summarize the data using both a measure of center and a measure of spread.

1) The **range** of a data set is the difference between its largest value and its smallest value.

Range = largest value – smallest value

In using the range, a great deal of information is ignored since only the largest and smallest values are used to calculated the range, so the range is not a measure of spread used often in summarizing data.

2) **Variance –** is a measure of how far the values in a data set are from the mean, on average

Deviation – the difference between a value and the mean, μ. deviation = $x - \mu$

If the deviation is positive the value lies above the mean.
If the deviation is negative the value lies below the mean.

The sum of the deviations equals zero: $\sum(x-\mu) = 0$. Recall, the fact that the mean is the value where the data would balance, it makes sense that the deviations on either side of the mean would cancel each other out.

Formula for **population variance**: $\sigma^2 = \frac{\sum(x-\mu)^2}{N} = \frac{\sum x^2}{N} - \mu^2$ where μ is the population mean and N is the population size

When the data values come from a *sample* rather than a population, the variance is called the **sample variance**. The procedure for computing the sample variance is a bit different from the one used to compute a population variance. In the formula, the population mean μ is replaced by the sample mean $\bar{x}$ and the denominator is $n-1$ (n is the sample size) instead of N. The sample variance is denoted by s^2.

Formula for **sample variance**: $s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$ where $\bar{x}$ is the sample mean and n is the sample size

When computing the sample variance, s^2, we use the sample mean, $\bar{x}$, to compute the deviations. For the population variance, σ^2, we use the population mean, μ, for the deviations. It turns out calculating the deviations using the sample mean tend to be a bit smaller than the deviations using the population mean. If we were to divide by n when computing a sample variance, the value would tend to be a bit smaller than the population variance.

It can be shown mathematically that the appropriate correction is to divide the sum of the squared deviations by $n - 1$ rather than n.

Because the variance is computed using squared deviations, the units of the variance are the squared units of the data. In most situations, it is better to use a measure of spread that has the same units as the data.

We do this simply by taking the square root of the variance. This quantity is called the **standard deviation**.

The standard deviation of a population is denoted, σ,
and the standard deviation of a sample is denoted, s.

population standard deviation: $\sigma = \sqrt{\sigma^2}$ **sample standard deviation:** $s = \sqrt{s^2}$

In other words, the computational formula for **sample standard deviation** is as follows:

$$s = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

Note: $\sum x^2 \neq (\sum x)^2$

For example: Given: 1, 2, 3, 4

$\sum x = 1 + 2 + 3 + 4 = 10$ so $(\sum x)^2 = 10^2 = 100$

$\sum x^2 = 1^2 + 2^2 + 3^2 + 4^2 = 30$ and $30 \neq 100$

1) For the following set of numbers: 10 12 14 17 18 20

Find:

a) n = **6** b) $\sum x =$ **10 + 12 + 14 + 17 + 18 + 20 = 91** c) $\sum x^2 =$ **$10^2 + 12^2 + 14^2 + 17^2 + 18^2 + 20^2 = 1453$**

d) s^2 = **14.57** $s^2 = \frac{6(1453) - 91^2}{6(6-1)} = \frac{437}{30} = \mathbf{14.57}$

e) s = **3.82** $s = \sqrt{14.57} = \mathbf{3.82}$

2) For the following set of numbers: 45 34 29 31 54 42 37 32

Find:

a) n = **8** b) $\sum x = \underline{\mathbf{45 + 34 + 29 + 31 + 54 + 42 + 37 + 32 = 304}}$

c) $\sum x^2 =$ **$45^2 + 34^2 + 29^2 + 31^2 + 54^2 + 42^2 + 37^2 + 32^2$ = 12056**

d) s^2 = **72** $$s^2 = \frac{8(12056) - 304^2}{8(8-1)} = \frac{4032}{56} = \mathbf{72}$$

e) s = **8.49** $s = \sqrt{72} = \mathbf{8.49}$

3) For the following set of numbers: 15 16 25 29 32 39 41 48 46 47

Find:

a) n = **10** b) $\sum x =$ **15 + 16 + 25 + 29 + 32 + 39 + 41 + 48 + 46 + 47 = 338**

c) $\sum x^2 =$ **$15^2 + 16^2 + 25^2 + 29^2 + 32^2 + 39^2 + 41^2 + 48^2 + 46^2 + 47^2$ = 12802**

d) s^2 = **153.07** $$s^2 = \frac{10(12802) - 338^2}{10(10-1)} = \frac{13776}{90} = \mathbf{153.07}$$

e) s = **12.37** $s = \sqrt{153.06667} = \mathbf{12.37}$

f) Add 20 to each of the data values above, what is the new variance, s^2 = **153.07**

what is the new standard deviation, s = **12.37**

g) Multiply each data value by 12, what is the new variance, s^2 = **12^2(153.07) = 22042.08**

what is the new standard deviation, s = **12(12.37) = 148.44**

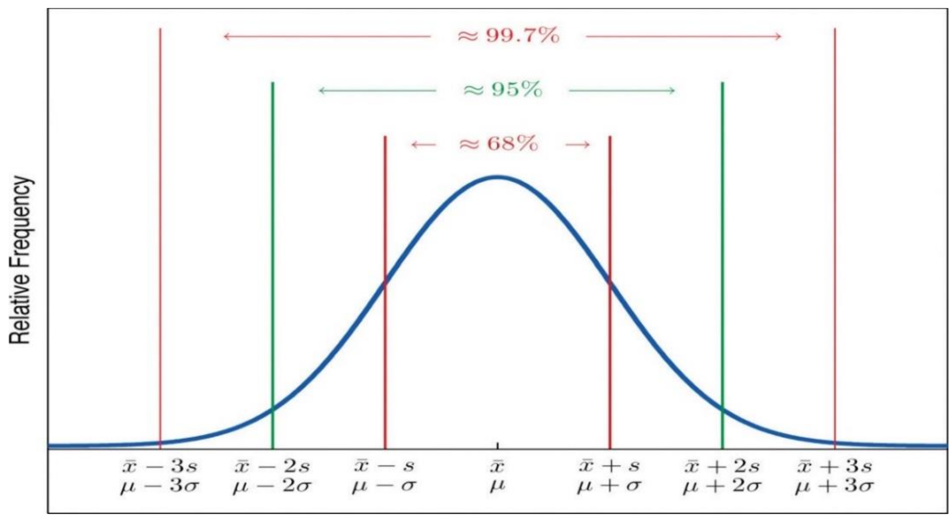
If a data set is approximately bell-shaped, the mean and standard deviation together can provide an approximate description of the data using the following rule:

The Empirical Rule: When a population has a histogram that is approximately bell-shaped then

- Approximately **68%** of the data will be within **one standard deviation** of the mean.
  $(\mu - \sigma, \mu + \sigma)$ or $(\bar{x} - s, \bar{x} + s)$

- Approximately **95%** of the data will be within **two standard deviations** of the mean.
  $(\mu - 2\sigma, \mu + 2\sigma)$ or $(\bar{x} - 2s, \bar{x} + 2s)$

- Approximately **99.7%** of the data will be within **three standard deviations** of the mean.
  $(\mu - 3\sigma, \mu + 3\sigma)$ or $(\bar{x} - 3s, \bar{x} + 3s)$



Examples:

1) IQ scores are approximately bell-shaped with a mean of 100 and a standard deviation of 15.

a) Between what two values will approximately 95% of the IQ scores be within? **(70, 130)**
(100 – 2(15), 100 + 2(15)) → (70, 130)

b) About what percent of the IQ scores is between 85 and 115? **68%**

c) About what percent of the IQ scores is between 55 and 145? **99.7%**

2) The heights of 2-year old girls are approximately bell-shaped with a mean of 34 inches and a standard deviation of 2.5 inches.

a) About what percent of the heights is between 29 and 39 inches? **95%**

b) About what percent of the heights is between 31.5 and 36.5 inches? **68%**

c) Between what two values will approximately 99.7% of the heights be within? **(26.5, 41.5)**
(34 – 3(2.5), 34 + 3(2.5)) → (26.5, 41.5)