SECTION 3.2 – Measures of Dispersion

Dispersion – The degree to which the data are spread out. There are four measures of dispersion:

Range, Standard Deviation, Variance, & Interquartile Range

Range – The **range**, *R*, of a variable is the difference between the largest and the smallest data value. That is,

Range = *R* = largest data value – smallest data value.

Standard Deviation

The **standard deviation** measures variation by indicating how far, on average, the observations are from the mean. For a data set with a large amount of variation, the observations will, on average, be far from the mean; so the standard deviation will be large. For a data set with a small amount of variation, the observations will, on average, be close to the mean; so, the standard deviation will be small.

Population Standard Deviation, σ

The **population standard deviation**, σ (lowercase Greek sigma) of a variable is the square root of the sum of squared deviations about the population mean divided by the number of observations in the population, *N*. That is, it is the square root of the mean of the squared deviations about the population mean.

$$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{N}} = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2}{N}}$$

where $x_1, x_2, ..., x_N$ are the *N* observations in the population and μ is the population mean.

Conceptual Formula & Computational Formula

There are two ways to calculate the population standard deviation, σ.

Conceptual Formula

$$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{N}}$$

Computational Formula

$$\sigma = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{N}}{N}}$$

Sample Standard Deviation, *s*

The **sample standard deviation**, *s*, of a variable is the square root of the sum of squared deviations about the population mean divided by *n* – 1, where *n* is the sample size.

$$s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1}}$$

where $x_1, x_2, ..., x_n$ are the *n* observations in the sample and $\bar{x}$ is the sample mean.

Conceptual Formula & Computational Formula

There are two ways to calculate the sample standard deviation, *s*.

Conceptual Formula

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

Computational Formula

$$s = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}}$$

SECTION 3.2 – Measures of Dispersion

Variance

The **variance** of a variable is the square of the standard deviation. The population variance is σ^2 and the sample variance is s^2.

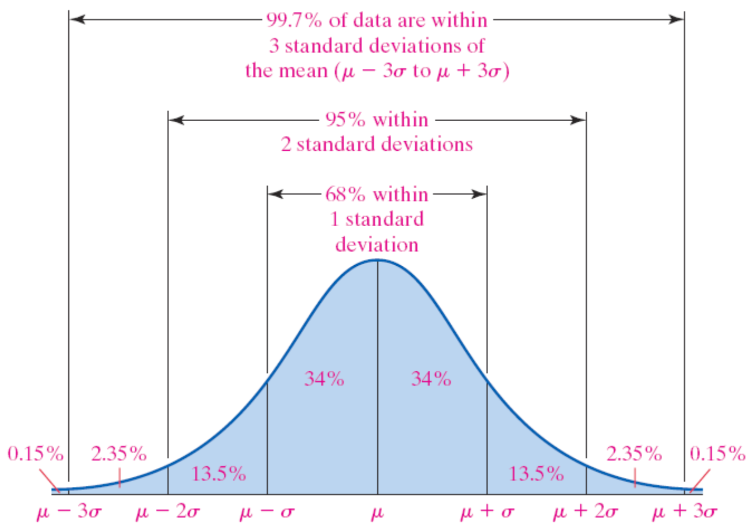
The Empirical Rule or (The 68-95-99.7 Rule)

If a distribution is roughly bell shaped, then

Approximately 68% of the data will lie within 1 standard deviation to either side of the mean. That is, approximately 68% of the data lie between $\mu - 1\sigma$ and $\mu + 1\sigma$.

Approximately 95% of the data will lie within 2 standard deviations to either side of the mean. That is, approximately 95% of the data lie between $\mu - 2\sigma$ and $\mu + 2\sigma$.

Approximately 99.7% of the data will lie within 3 standard deviations to either side of the mean. That is, approximately 99.7% of the data lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.



Final Thought: The sample variance is obtained using the square of the formula for sample standard deviation. Suppose we divide by n instead of $n - 1$ in this formula to obtain the sample variance, as we might expect. Then the sample variance will consistently underestimate the population variance. Whenever a statistic consistently underestimates a parameter, it is said to be **biased**. To obtain an unbiased estimate of the population variance, we divide the sum of squared deviations about the sample mean by $n - 1$.

SECTION 3.2 – Measures of Dispersion

☺ **Exercises**:

1) Using the conceptual formula, find the sample standard deviation of the following data set: Round your answer to the nearest hundredth.

7, 6, 8, 5, 6, 7, 8, 9

x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$

2) Using the computational formula, find the sample standard deviation of the same data set in the previous example. Round your answer to the nearest hundredth.

7, 6, 8, 5, 6, 7, 8, 9

x_i	x_i^2

56 404

SECTION 3.2 – Measures of Dispersion

☺ **Exercises**:

3) Article by D. Schaefer et al. (*Journal of Tropical Ecology*, Vol. 16, pp. 189─207) reported on a long-term study of the effects of hurricanes on tropical streams of the Luquillo Experimental Forest in Puerto Rico. The study showed that Hurricane Hugo had a significant impact on stream water chemistry. The following table shows a sample of 10 ammonia fluxes in the first year after Hugo. Data are in kilograms per hectare per year. Find the sample standard deviation of the data set using both the conceptual and computational formulas.

96	66	147	147	175
116	57	154	88	154

x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$

x_i	x_i^2

Answer: **#3)** $s = 41.23$

SECTION 3.2 – Measures of Dispersion

☺ **Exercises**:

4) Find the range of the following data set: 2, 5, 5, 5, 6, 8, 12, 15

5) Find the range of the following data set: 5, 4, 3, 11, 12, 13

6) If the population variance of a data set is known to be 4, then what is the population standard deviation?

7) If the sample standard deviation of a data set is known to be 9, then what is the sample variance?

8) The mean lifetime of a particular LED light bulb is 3000 hours with a standard deviation of 700 hours. Assume that bulb life has a bell-shaped distribution.

a) 68% of all manufactured light bulbs will last between __________ and __________ hours.

b) 95% of all manufactured light bulbs will last between __________ and __________ hours.

c) 99.7% of all manufactured light bulbs will last between __________ and __________ hours.

9) Suppose the pulse rates (in terms of beats per minute) of 200 college men are bell-shaped with a mean of 72 bpm and standard deviation of 6 bpm.

a) What percentage of college men have pulse rates between 54 and 90?

b) What percentage of college men have pulse rates between 66 and 78?

c) What percentage of college men have pulse rates between 60 and 84?

d) What percentage of college men have pulse rates over 72?

10) SAT Math scores have a bell-shaped distribution with a mean of 515 and a standard deviation of 114.
Source: College Board, 2010

a) What percentage of SAT scores is between 401 and 629?

b) What percentage of SAT scores is less than 401 or greater than 629?

c) What percentage of SAT scores is greater than 743?