AP Biology
Stats Notes

Mean, Standard Deviation and Standard Error

Mean

The mean formula should be used when you need to calculate the average of a set of values (data points). You are finding the average – don't be intimidated by the crazy formula below.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Standard Deviation

The standard deviation formula is used to determine the amount by which your values (data points) typically differ from the mean value. In other words, the standard deviation determines the amount of variation in your data.

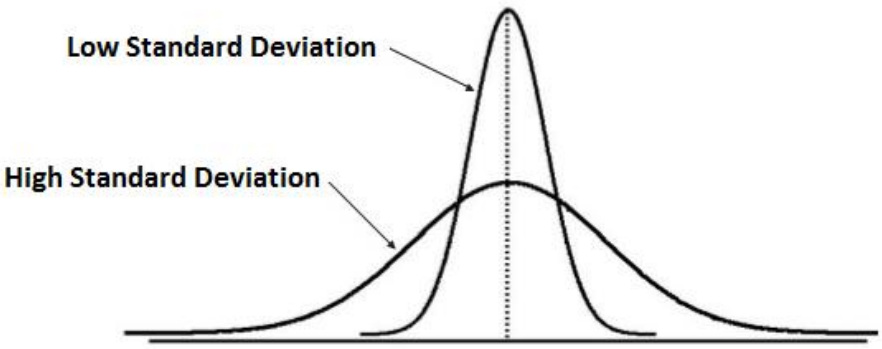
$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

A **low standard deviation** (less than 2):

- Tells you that your data is closer to the **mean** (most of the numbers are close to the **mean**)
- Increases the likelihood that the independent variable is causing the change in the dependent variable

A **high standard deviation** (greater than 2):

- Tells you that your data is spread out (numbers are further from the **mean**)
- Additional factors, other than the independent variable, are affecting the dependent variable



Standard Error

The standard error formula is used to determine the precision of the mean value. In other words, you are determining how confident you are in your mean value by considering both the standard deviation (s) and the number of data points (n). Typically, when we have more data points, we can be more confident in our data (i.e., a lower standard error).

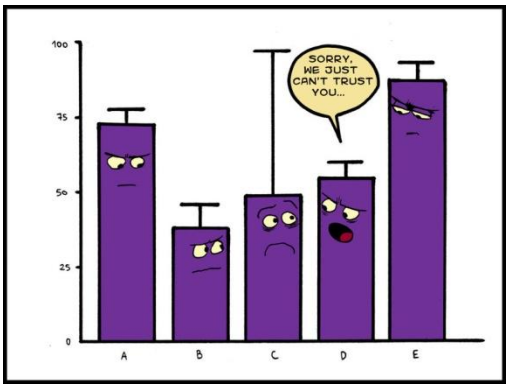
$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

A **low standard error**:

- Tells you that the **calculated mean** is closer to the **true mean** of the population

Standard error allows you to:

- Infer how well the sample mean represents the true population mean
- Helps you to determine confidence in the data collected in a sample
- Set the **95% confidence interval = ± 2 SE (and place Error Bars on the data sets)**
- Random sampling of the population should produce a mean that falls within ± 2 SE 95% of the time



Error Bars

- If the error bars/confidence intervals **DO NOT** overlap, then you could claim that the data is statistically significant
- If the error bars/confidence internal **DO** overlap, you could not claim a statistically significant difference