**Link for ALL categories of statistic topics:** [**http://www.statisticslectures.com/topics/statistics/**](http://www.statisticslectures.com/topics/statistics/)

**Statistics:**

1. **Descriptive Statistics** 🡪 Link: <http://statisticslectures.com/topics/descriptiveinferential/>

Meaning: Organizing and Summarizing Data

Purpose: Organizing and Summarizing Data

1. **Inferential Statistics** 🡪 Link: <http://statisticslectures.com/topics/descriptiveinferential/>

Meaning: Consists of using data you’ve collected to form conclusions.

1. Population: The **ENTIRE** group we are interested in studying
2. Sample: A **subset of the population**. A select group of information taken from a population.

Example: Let’s say there are 20 statistics classes at your university, and you’ve collected the ages of all the students in one class.

Ages of students in your statistics class: 19, 21, 18, 18, 34, 30, 25, 26, 24, 24, 19, 18, 21, 49, 27

Descriptive Statistic: “What is the most common age of student in your statistics class?" The answer in this case would be 18.

Inferential Statistic: "Are the ages of the students in this classroom similar to what you would expect in a normal statistics class at this university?"

1. **Sampling Methods** 🡪 Link: <http://statisticslectures.com/topics/samplingmethods/>

* Purpose: To create a sample that is representative of the population it is being drawn from.

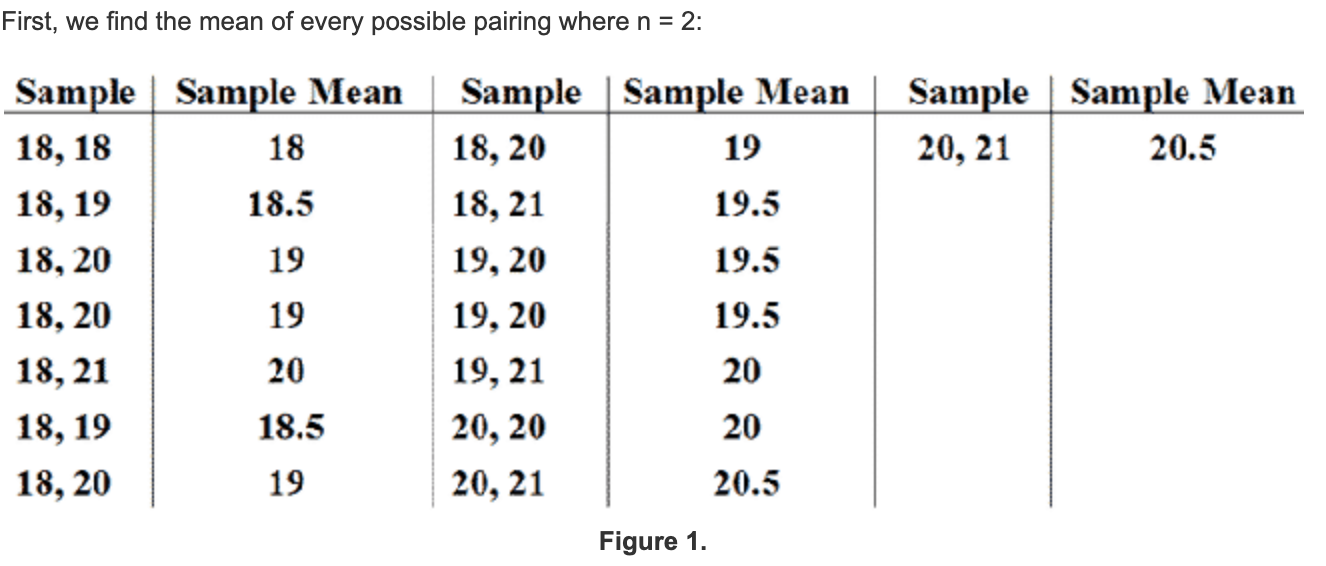
1. Simple Random Sampling:
   1. Meaning: Every member of the population (N) has an equal chance of being of being selected for your sample (N).

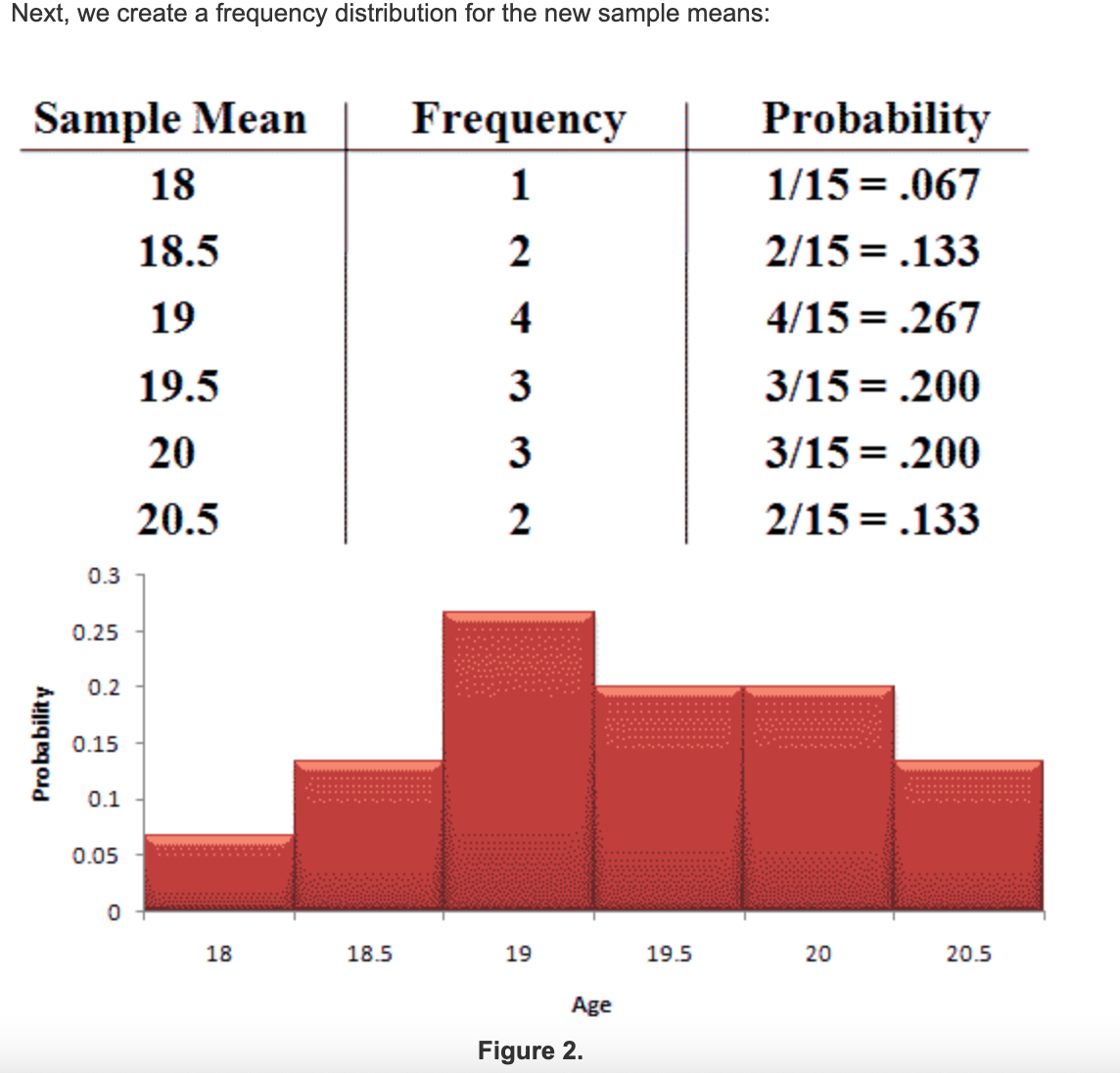
* This is considered the **BEST** sampling method. Your samples almost guaranteed to be representative of your population. However it is **rarely used** due to it being **IMPRACTICAL**.

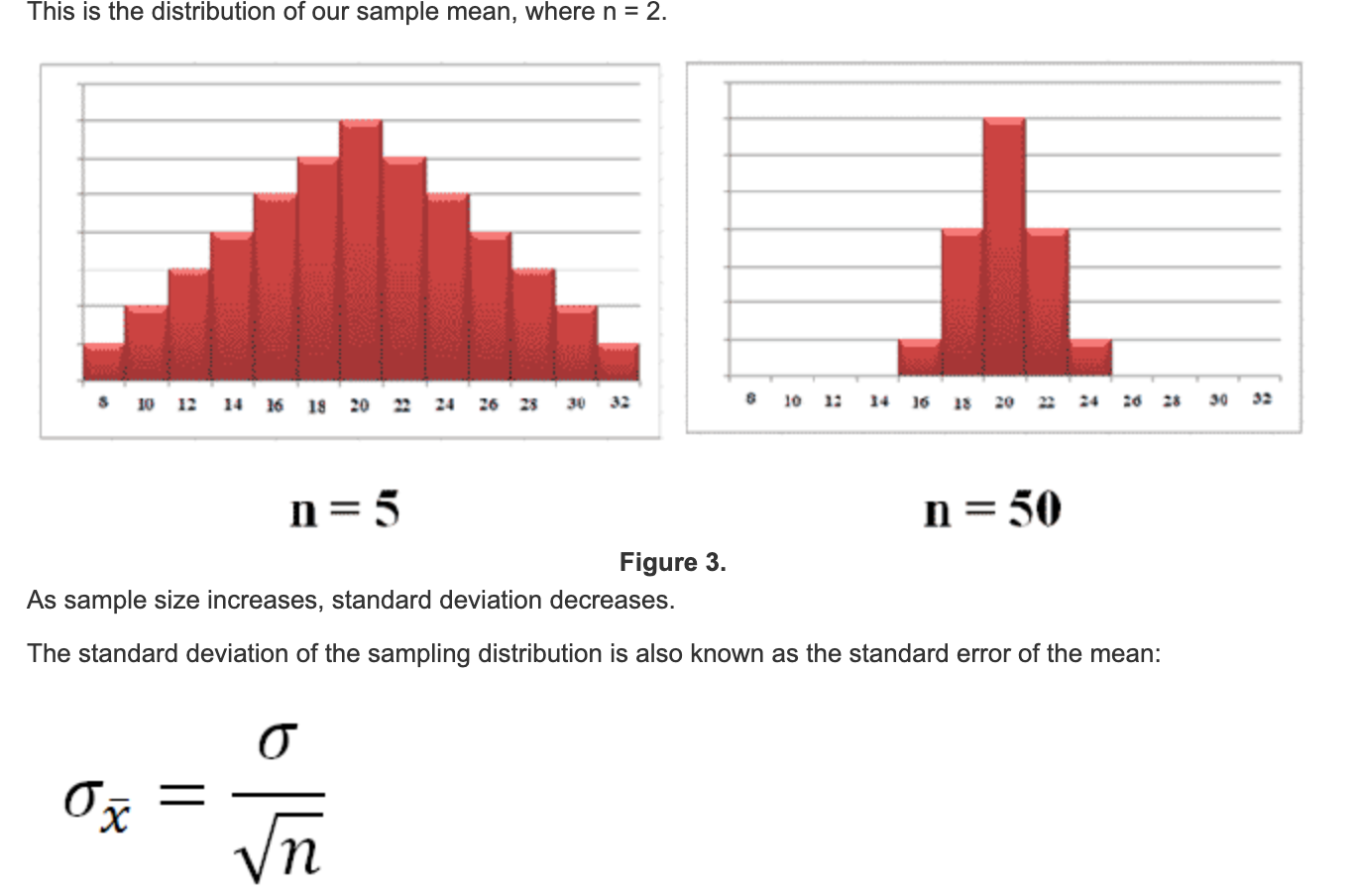
1. Stratified Sampling:
   1. Meaning: The population (N) is split into non-overlapping groups (“strata”), then simple random sampling is done on each group to form a sample (N).
   2. Example: Splitting a population of students into men and women, then sampling from each of the two groups. This may allow us to collect the same amount of information as simple random sampling, but use less people.
2. Systematic Sampling:
   1. Meaning: Every nth individual from the population is placed in the sample (N).
   2. Example: Every 7th individual to walk out of a supermarket to your sample, you are performing systematic sampling.
3. Convenience Sampling:
   1. Meaning: Easily obtained individuals from the population (N) are placed in the sample (N).
   2. Example: you pick the easiest way of getting your sample. This type of sampling is sometimes called voluntary response sampling, because individuals often select to be a part of the sample. This can be a problem, because there **may be a difference between people who choose to participate and people who don’t**.
4. **Parameter Statistics** 🡪 link: <http://statisticslectures.com/topics/parametersstatistics/>
   1. Parameter:
      1. Meaning: A characteristic that describes a **population** is called a **parameter**. Because it is often difficult (or impossible) to measure an entire population, **parameters are most often estimated**.
   * Population Mean Notation: μ
   * Population Standard Deviation: σ
     1. Example:
   1. Statistic:
      1. Meaning: A characteristic that describes a **sample** is called a **statistic**. **Statistics** are most often used to **estimate the value of unknown parameters**.
   * Sample Mean Symbol: x̄
   * Sample Variance Symbol: s² or σ²
     1. Example: For example, if I were to measure the height of 5000 randomly selected individuals, then find the mean of the heights I collected, the resulting value would be a statistic. I could then use the value of this statistic to make an estimation of the mean height of the population, which is a parameter.
   1. Sampling Error:
      1. Meaning: Any difference that exists between a statistic and its corresponding parameter.
      2. Example: Imagine that after measuring the heights of 5000 individuals, I calculate a statistic which estimates the population mean (a parameter) to be 68 inches. However, my estimate is off, and the actual mean of individuals in the population in 70 inches. This discrepancy is known as sampling error.
5. **Sample Mean:** Link 🡪 <http://statisticslectures.com/topics/distributionsamplemean/>
   1. Meaning: Distribution of the sample mean is the probability distribution for all possible sample mean, computed from a sample of size N.
   2. Purpose: To find the **frequency of each sample mean** **within a sample** N.
   3. Example: A statistics class has six students, ages displayed below. Construct a sampling distribution of the mean of age for samples (n = 2).

Ages: 18, 18, 19, 20, 20, 21

First, we find the mean of every possible pairing where n = 2:

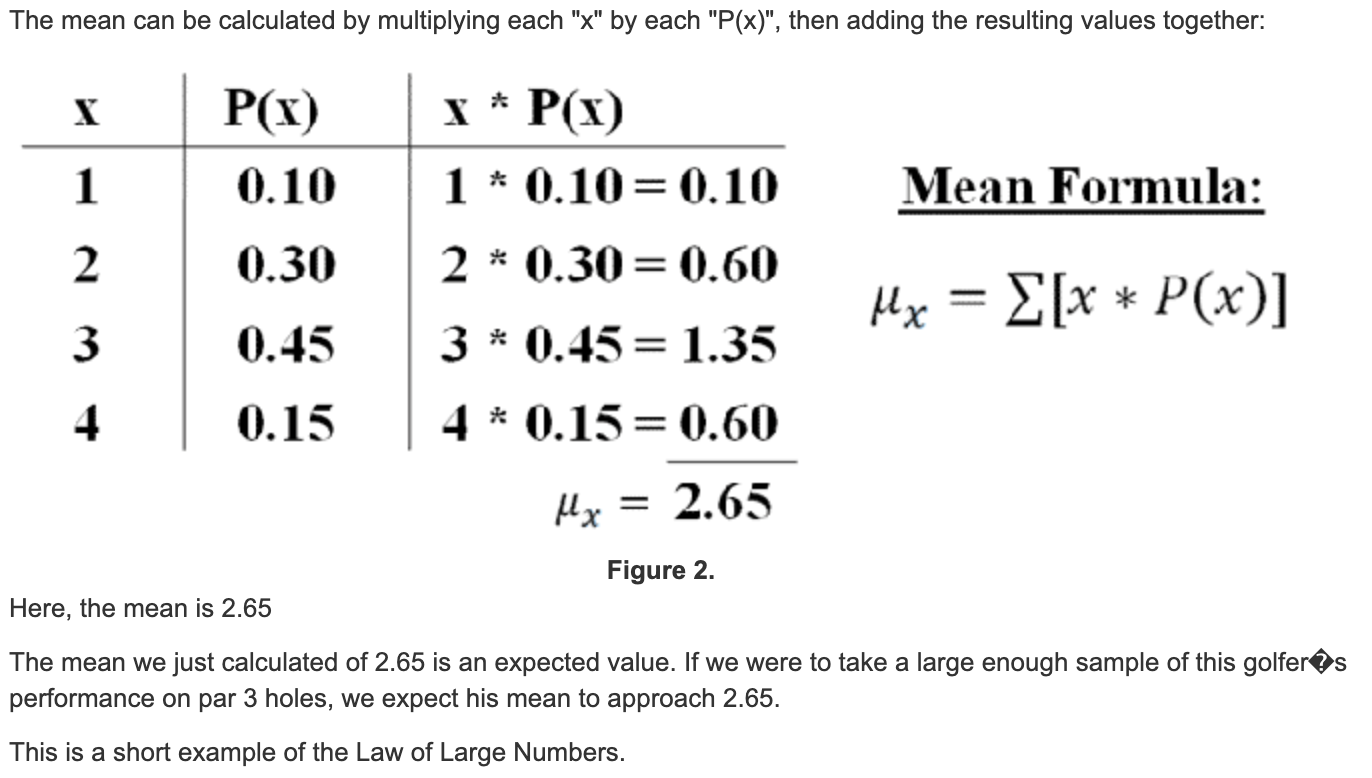






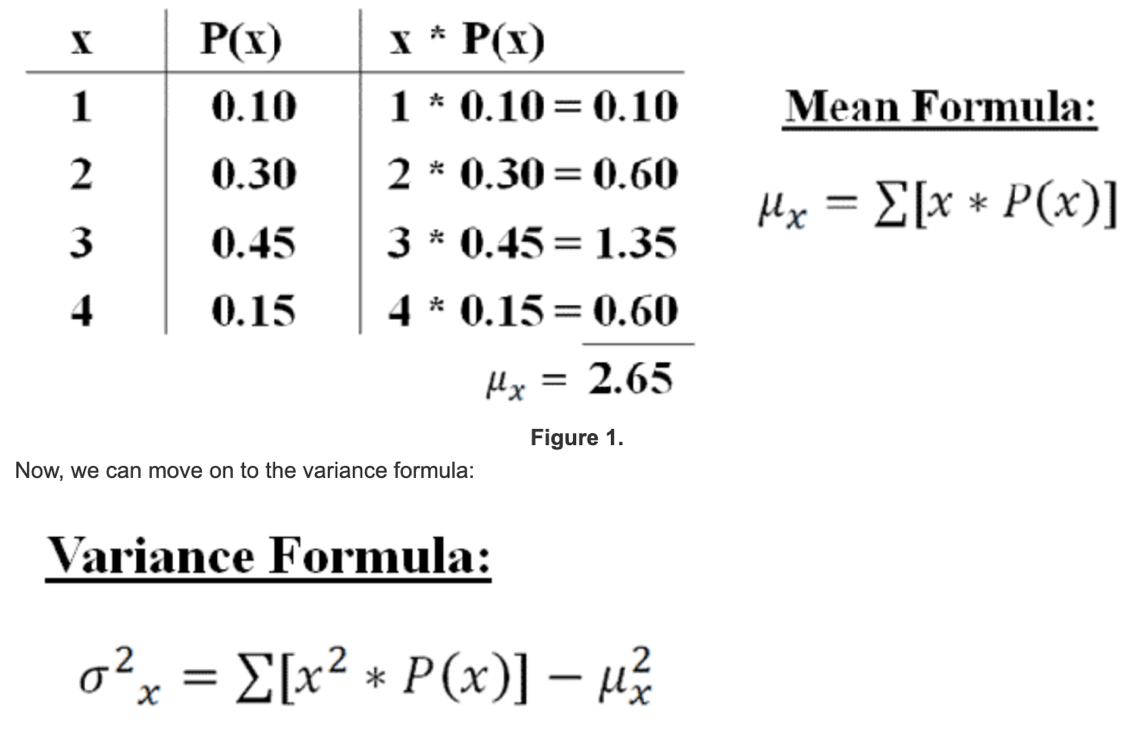
* **Note:** As **sample size** N **increases**, **standard deviation** (how wide the data is spread) **DECREASES**.

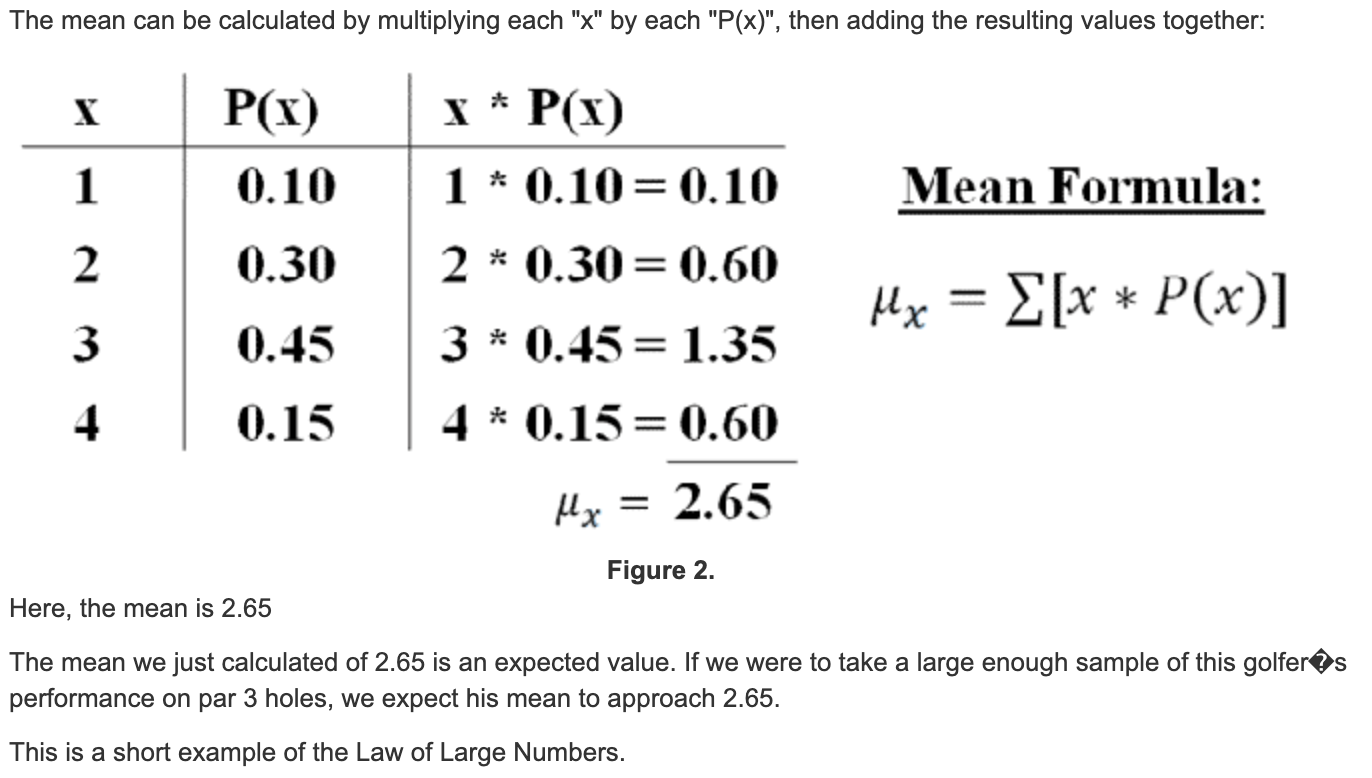
1. **Mean and Expected Value Discrete:** link 🡪 <http://statisticslectures.com/topics/meanexpectedvaluediscrete/>
   1. Meaning: The mean calculated is the expected value for the probability experiment.
   2. Example:

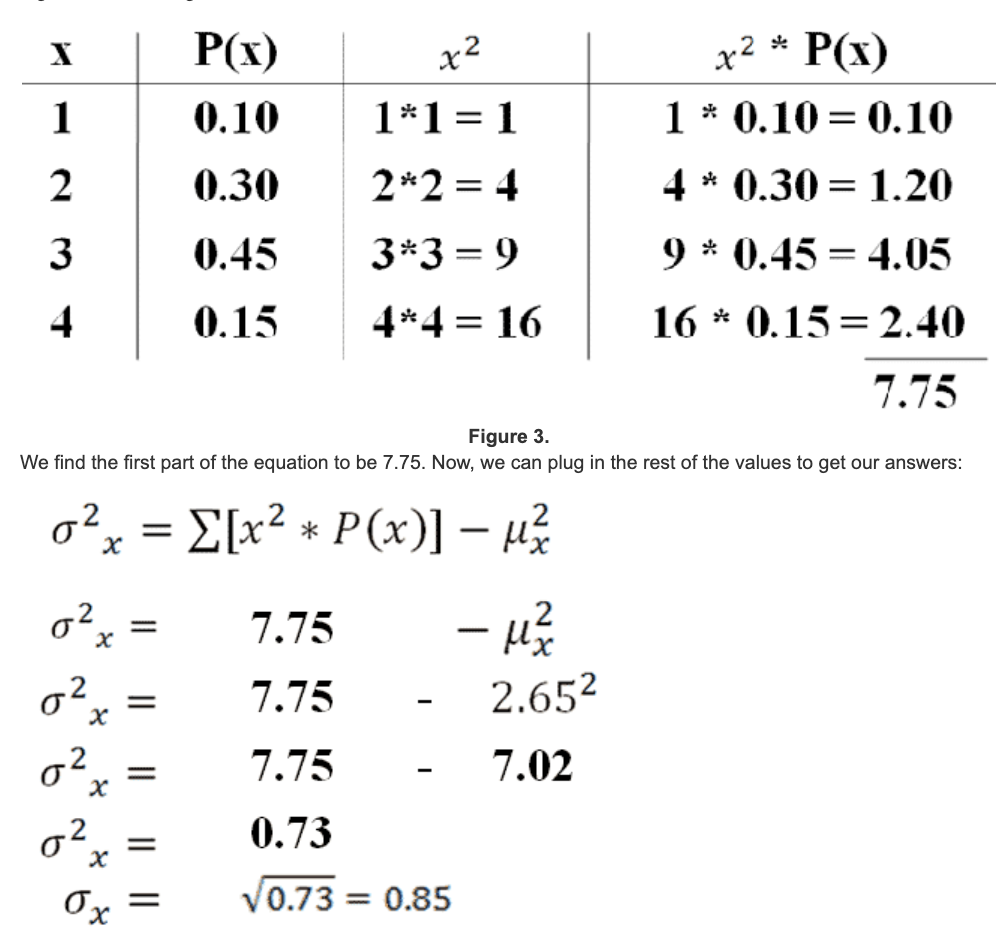


* If we were to take a large enough sample of this golfers performance on par 3 holes, we **expect** his **mean** to approach 2.65. This is an example of **Law of Large Numbers**.

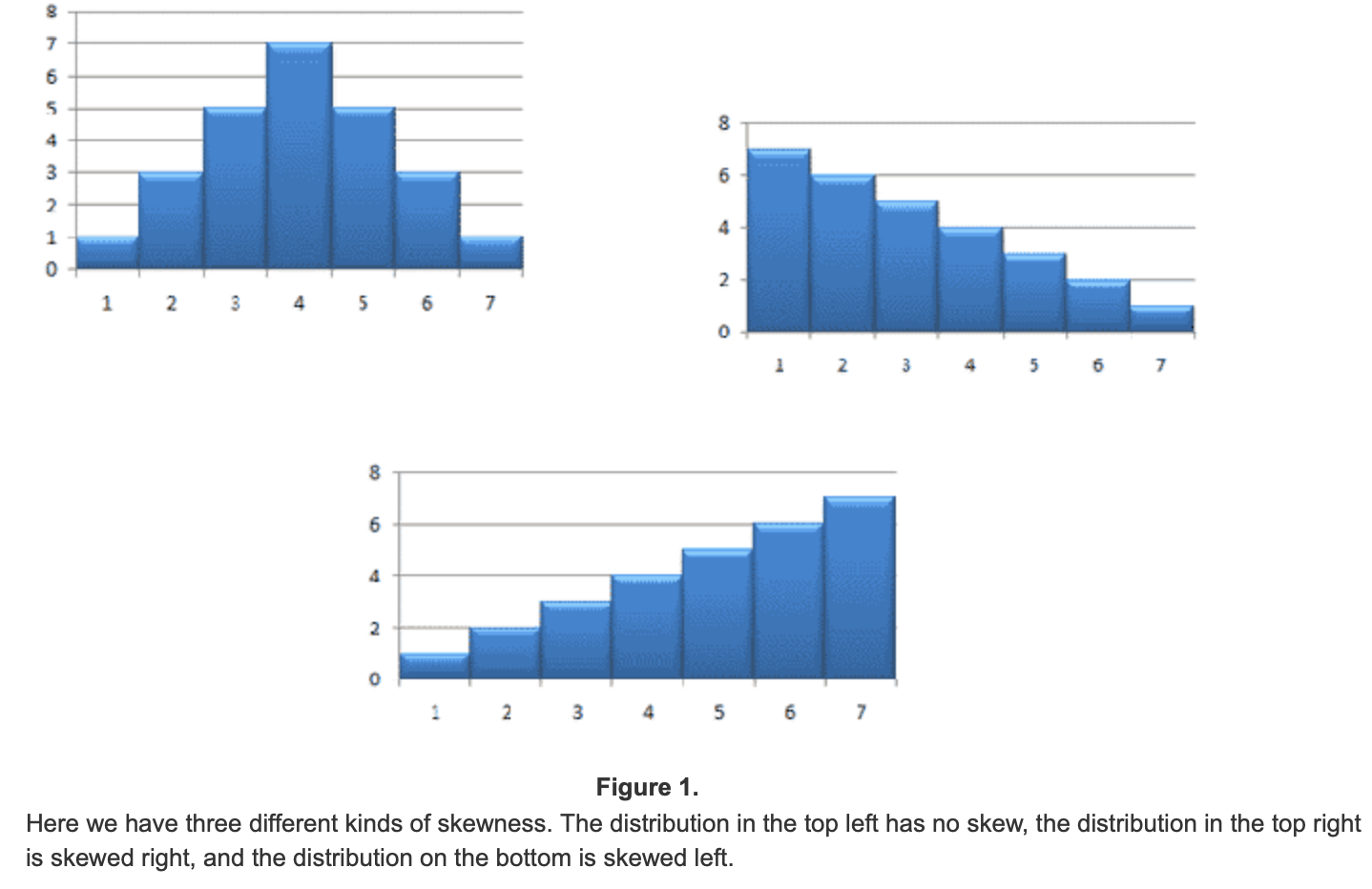
1. **Variance and the Standard Deviation of a Random Variable**: Link 🡪 <http://statisticslectures.com/topics/variancestandarddeviationdiscrete/>
   1. Meaning:
   2. Purpose: To calculate the variance of a discrete random variable, we must first calculate the mean.
   3. Example:







1. **Central Limit Theorem**: Link 🡪 <http://statisticslectures.com/topics/centrallimittheorem/>
   1. Meaning: The Central Limit Theorem states that regardless of the shape of the population distribution, the distribution of sample means will be approximately normal.



From the central limit theorem, the following is true:

1. Population distributions that have no skew will lead to distributions of sample means that have no skew.

2. Population distributions that are skewed right will lead to distributions of sample means that have no skew.

3. Population distributions that are skewed left will lead to distributions of sample means that have no skew.

The distribution of sample means will become more normal as its sample size increases.

Good rule of thumb: sample distributions will usually be approximately normal if their sample size is n = 30 or larger.

1. **Null Hypothesis = H0 (DONE IN MATH 242** 🡪 **Check notes)**

Meaning: Assumption you’re beginning with. The opposite of what you are testing.

Example: School District A states that its high school have 85% passage rate on the High School Exit Exam. A new school was recently opened in the district and it was found that a sample of 150 students had a passage rate of 88%, with a standard deviation of 4%. Does this new school have a different passing rate than the rest of school district A?

* If the standard deviation was 0.1% and it was a 3% difference. Then that’s a significant amount.
* Null Hypothesis: Starting with assumption that it’ll be 85%.

1. **Alternative Hypothesis = Ha Or H1**

**Meaning: Claim that you’re testing**

* Claim that the school has an 88% passing percentage

EX: **H0 = μ = 85% Ha = μ != 85%**

* (WATCH VIDEO ON THIS TO UNDERSTAND MORE)

|  |  |  |
| --- | --- | --- |
|  | **Reject** **H0** | **Do not reject** **H0** |
| **H0 is True** | Type I error | Correct Decision |
| **H0 is false** | Correct Decision | Type II error |

* **Done in Math 242 (11/20/20):**
  + Z-Test if we have the sample mean (do this test).
  + T-Test if we do not know what the sample mean is.
* EX:

**H0 = 100 H1 != 100**

**alpha = 0.05**

Split the graph with 95% in the middle, 2.5% on the left side and 2.5% on the right side. Take the exact value from the Z-test chart and plug all the values into the equation.

If the Z is less than -1.96 or greater than 1.96 reject the null hypothesis.

Z = 14.60 so accept the **alternate** hypothesis meaning that \_\_\_\_ (find what accepting alternate hypothesis means)

1. **Pearson Correlation**: Link 🡪 <http://statisticslectures.com/topics/pearsonr/>

Meaning: Measures the **strength of the linear relationship** between two variables.

1. **Spearman Correlation**: Link 🡪 <http://statisticslectures.com/topics/spearman/>

Meaning:

Purpose:

1. Measuring the relationship between two ordinal variables.

2. Measuring the relationship between two variables that are related, but not linearly.