Independence of Random Variables

- An **independent random variable** is a random variable that doesn’t have an effect on the other random variables in your experiment.
- In other words, it doesn’t affect the probability of another event happening.
- For example, let’s say you wanted to know the average weight of a bag of sugar so you randomly sample 50 bags from various grocery stores.
- You wouldn’t expect the weight of one bag to affect another, so the variables are independent. The opposite is a **dependent random variable**, which *does* affect probabilities of other random variables.

- If X and Y are two random variables and the distribution of X is not influenced by the values taken by Y, and vice versa, the two random variables are said to be independent.
- Mathematically, two discrete random variables are said to be independent if:
- $P(X=x, Y=y) = P(X=x)\, P(Y=y)$, for all x,y.

- P(x|y) = P(x), for all values of X and Y.
- **This statement, P(x|y) = P(x), for all values of X and Y,** is stating “the probability of x, given y, is x.” In other words, knowing y should make no difference on the probability, x — it’s still going to be just x no matter what the value of y.

- As a **simple example**, let's say you have two random variables X and Y. X can equal 0, 1, or 2 and Y can equal 0 or 1. First, let's take a look at their probabilities:

The probability that X = 0 is 20%: Or, more formally — $P(X = 1) = 0.2$.

The probability that X = 1 is 30%: $P(X = 3) = 0.3$.

The probability that X = 2 is 50%: $P(X = 5) = 0.5$.

The probability that Y is 0 is 40%: $P(Y = 0) = 0.4$.

The probability that Y is 1 is 60%: $P(Y = 1) = 0.6$

- But what happens to the probabilities when the two happen at the same time?

For each possible combination of X, given that Y has happened (in notation, that's (X|Y)), the probabilities are:

- $P(x = 0 \mid y = 0) = 0.2$;
- $P(x = 1 \mid y = 0) = 0.3$;
- $P(x = 2 \mid y = 0) = 0.5$;
- $P(x = 0 \mid y = 1) = 0.2$;
- $P(x = 1 \mid y = 1) = 0.3$;
- $P(x = 2 \mid y = 1) = 0.5$;
- The changing y-values have no effect on the x probabilities. Assuming the reverse is also true (that changing x-values would have no effect on the y-values), these are **independent random variables.**

Covariance

- Covariance is the measure of the joint variability of two random variables .
- It shows the degree of linear dependence between two random variables.
- Positive covariance implies that there is a direct linear relationship i.e. increase in one variable corresponds with greater values in the other.
- Negative covariance implies the greater the values of one random variable the lower the values for the other. Thus, the sign of covariance shows the nature of the linear relationship between two random variables.
- Finally, a covariance is zero for two independent random variables. However, a zero covariance does not imply that two random variables are independent.

- The formula for covariance is:

$$
\begin{aligned}
Cov(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\
&= \sum_{x,y} (X-\mu_X)(Y-\mu_Y) P_{X,Y}(x,y)
\end{aligned}
$$

Central Limit Theoram(CLT)

Why use it?

The Central Limit Theorem (CLT) is a foundation for parametric hypothesis testing.

Understanding this theorem provides knowledge of how to apply inferential statistics to data.

What does it do?

The Central Limit Theorem states that the means of random samples drawn from *any distribution with* mean **μ** and variance **σ^2** will have an approximately normal distribution with a mean equal to **μ** and a variance equal to **σ^2/n.**

μ is the population mean

σ is the population standard deviation

σ^2 is the population variance

n is the sample size

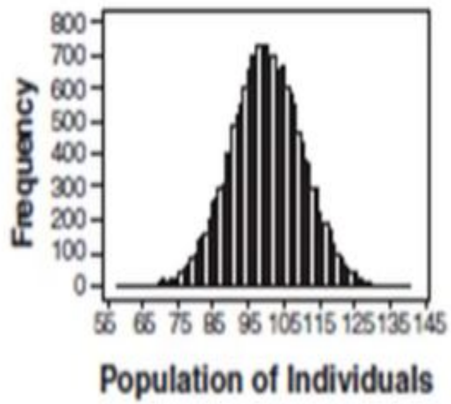
The CLT allows the use of confidence intervals, hypothesis testing, DOE, regression analysis, and other analytical techniques on data.

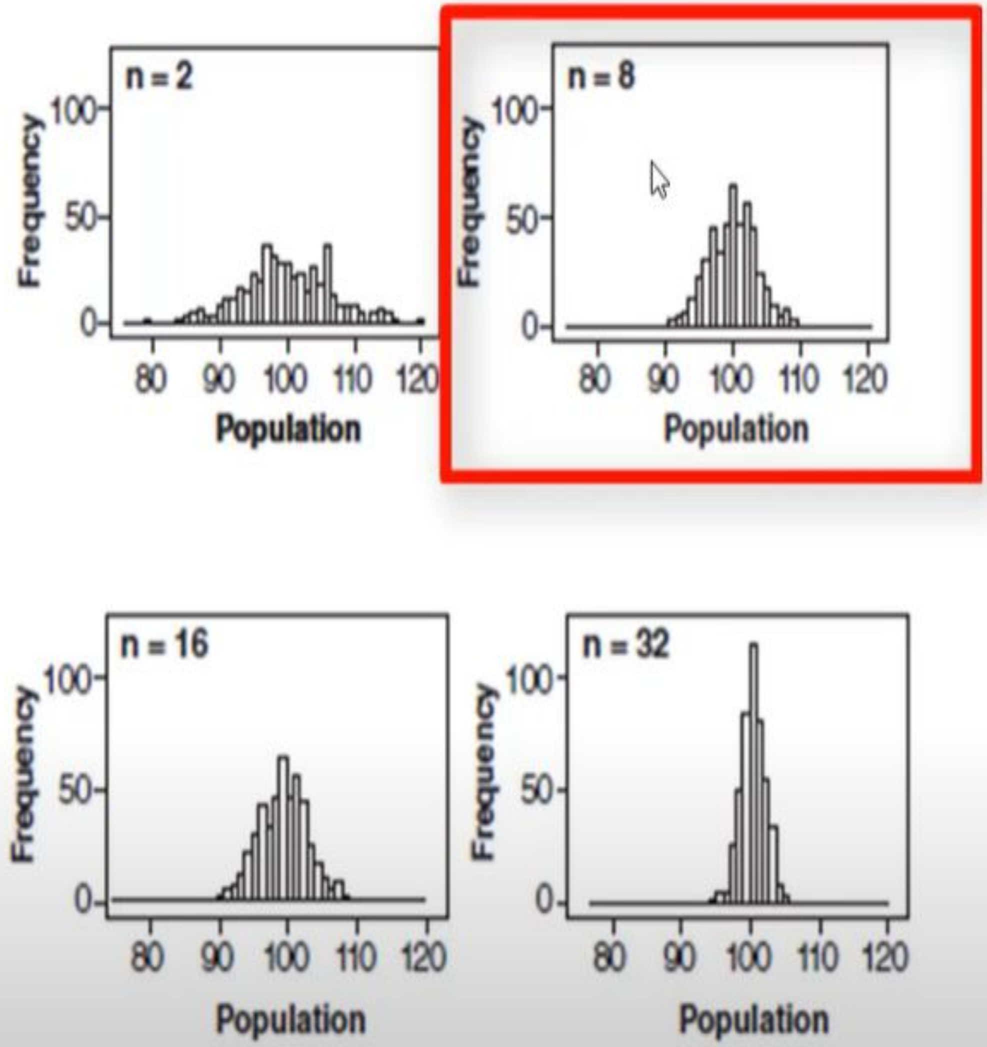
The CLT can be better understood by reviewing examples of its application.

- ***The first example takes samples from a normal distribution;***
- ***the second and third examples take samples from non-normal distributions.***

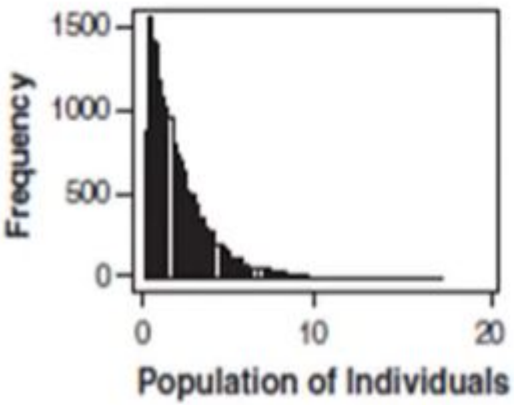
Case # 01

Normal Distribution

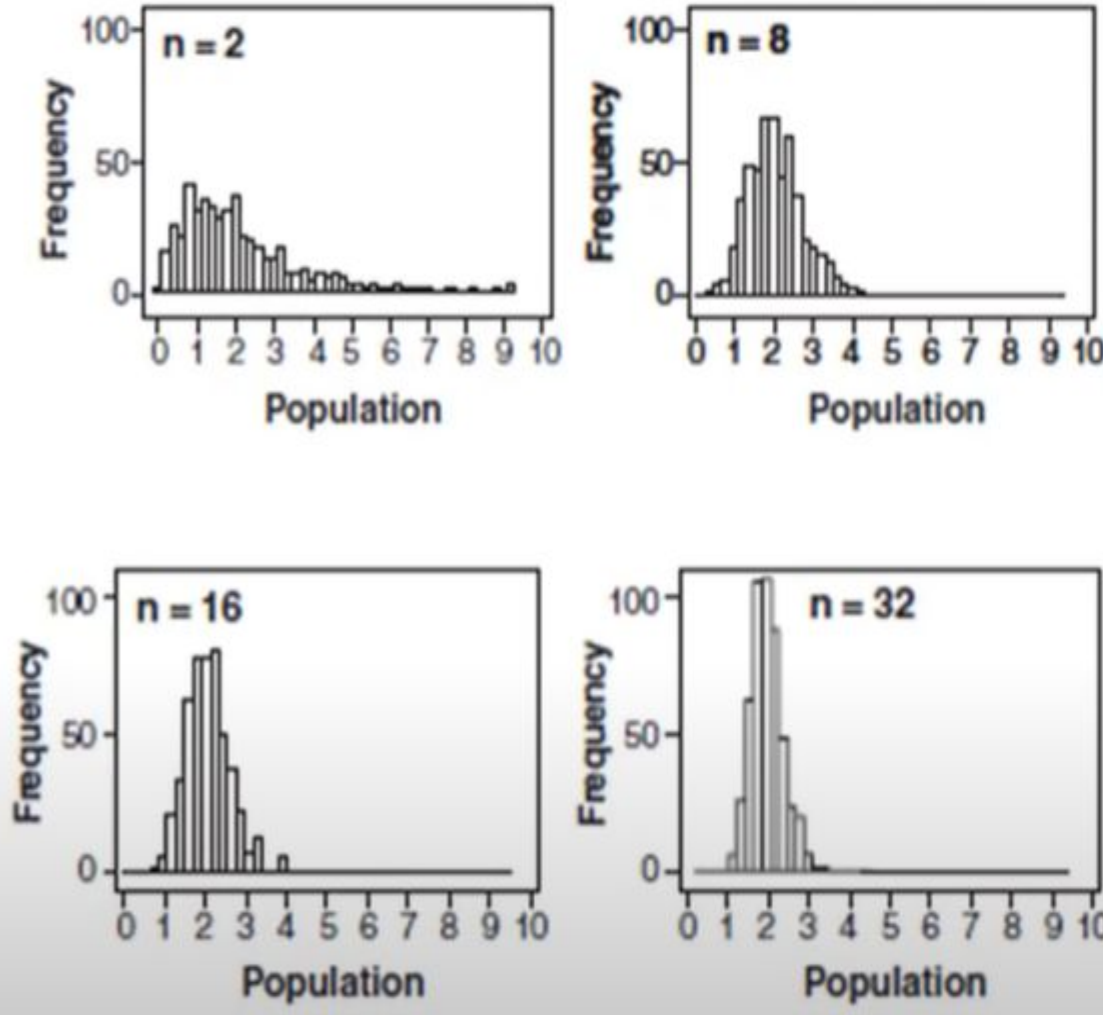




Non-normal Distribution

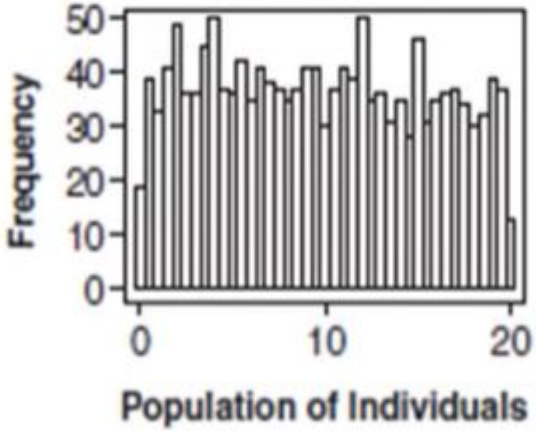


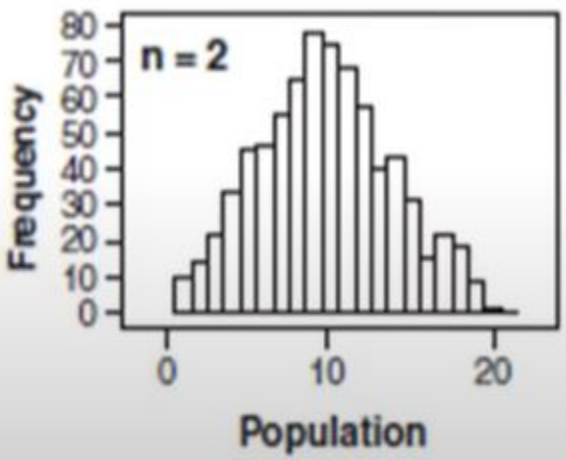
Case # 02

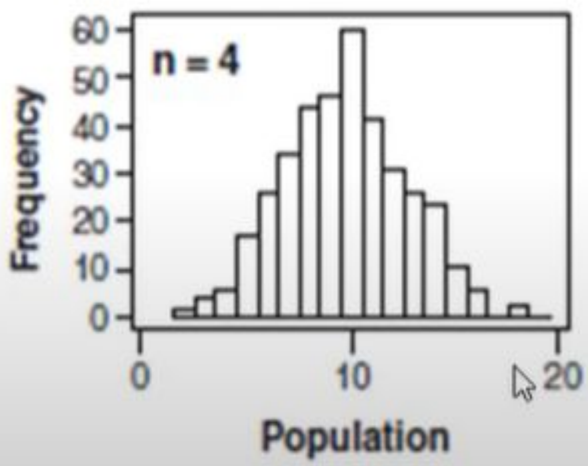


Case # 03

Non-normal Distribution







The **Central Limit Theorem** is exactly what the shape of the distribution of means will be when we draw repeated samples from a given population. Specifically, as the sample sizes get larger, the distribution of means calculated from repeated sampling will approach normality.

Three different components of the **central limit theorem**

(1) Successive sampling from a population

(2) Increasing sample size

(3) Population distribution.

Chebyshev's Inequality

What is Chebyshev's Inequality

- Chebyshev's inequality is a theory describing the maximum number of extreme values in a probability distribution.
- It states that no more than a certain percentage of values (1/k2) will be beyond a given distance (k standard deviations) from the distribution's average.
- The theorem is particularly useful because it can be applied to any probability distribution, even ones that violate normality, so long as the mean and variance are known.

- A big advantage of this theorem is that it is applicable to any probability distribution regardless of its normality, as long as the mean and variance are known. The theorem is as follows:

$$Pr(|X - \mu| \geq k \times \sigma) \leq \frac{1}{k^2}$$

How to calculate Chebyshev inequality with real-world data?

- Take the example of an insurance company that provides different types of claims for its customers at different times and in varying amounts. Based on what you have received so far, you want to know how large the claims will likely be in the future. In order to cover the claims of the upcoming calendar year, for example, you need to set aside enough reserves. So all claims are assumed to be independent and equally distributed. Therefore, each claim is considered to be based on a random drawing from a single unknown distribution.
- It is clear from Chebyshev's inequality that you can be at least 90% sure that future claims will not be more than three standard deviations from their means. This is without knowing anything about the underlying distribution.

Applications

- Chebyshev's inequality theorem is one of many (e.g., Markov's inequality theorem) helping to describe the characteristics of probability distributions.
- The theorems are useful in detecting outliers and in clustering data into groups.

A Numerical Example

Suppose a fair coin is tossed 50 times. The bound on the probability that the number of heads will be greater than 35 or less than 15 can be found using Chebyshev's Inequality.

Let X be the number of heads. Because X has a binomial distribution:

- The expected value will be: $\mu = n \times p = 50 \times 0.5 = 25$
- The variance will be: $\sigma^2 = n \times p \times (1 - p) = 50 \times 0.5 \times 0.5 = 12.5$
- The standard deviation will be $\sqrt{\sigma^2} = \sqrt{12.5}$
- The value 35 and 15 are 10 away from the average, which is 2.8284 standard deviations (ie, $10/\sqrt{12.5}$)

Using Chebyshev's Inequality we can write the following probability:

$$Pr(X < 15 \cup X > 35) = Pr(|X - \mu| \geq k \times \sigma) \leq \frac{1}{k^2} = \frac{1}{2.8284^2} = 0.125$$

In other words, chances of a fair coin coming up heads outside the range of 15 to 35 times is at most **0.125**.

What is Probability Distribution?

- Probability Distribution is a statistical function which links or lists all the possible outcomes a random variable can take, in any random process, with its corresponding probability of occurrence.

Need of Probability Distribution

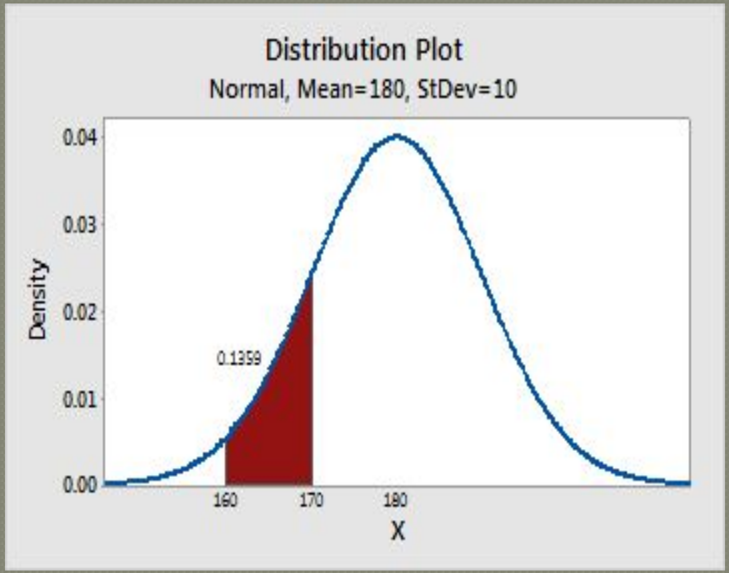
- According to the definition of random variable, it’s the variable which can hold different set of values from the outcome of any random process. However, it lacks the capability to capture the probability of getting those different values.
- So, probability distribution helps to create a clear picture of all the possible set of values with their respective probability of occurrence in any random process.

What is a continuous distribution?

- A continuous distribution describes the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.
- Probabilities of continuous random variables (X) are defined as the area under the curve of its PDF. Thus, only ranges of values can have a nonzero probability. The probability that a continuous random variable equals some value is always zero.

Example of the distribution of weights

- The continuous normal distribution can describe the distribution of weight of adult males. For example, you can calculate the probability that a man weighs between 160 and 170 pounds.



What is a discrete distribution?

- A discrete distribution describes the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.
- With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability. Thus, a discrete probability distribution is often presented in tabular form.

Example of the number of customer complaints

- With a discrete distribution, unlike with a continuous distribution, you can calculate the probability that X is exactly equal to some value. For example, you can use the discrete Poisson distribution to describe the number of customer complaints within a day. Suppose the average number of complaints per day is 10 and you want to know the probability of receiving 5, 10, and 15 customer complaints in a day.

x	P (X = x)
5	0.037833
10	0.125110
15	0.034718