1. What is a Cumulative Distribution Function, and how does it work?

The cumulative distribution function is used to describe the probability distribution of random variables. It can be used to describe the probability for a discrete, continuous or mixed variable. It is obtained by summing up the probability density function and getting the cumulative probability for a random variable

2. When should we use a t-test vs a z-test?

As mentioned, a t-test is primarily used for research with limited sample sizes whereas a z-test is deployed for hypothesis testing that requires researchers to look at a population size that's larger than 30

3. How do we examine two category characteristics?

This test is used to determine if two categorical variables are independent or if they are in fact related to one another. If two categorical variables are independent, then the value of one variable does not change the probability distribution of the other.

4. Explain the concept of Chebyshev's Inequality.

Chebyshev's inequality states that within two standard deviations away from the mean contains 75% of the values, and within three standard deviations away from the mean contains 88.9% of the values. It holds for a wide range of probability distributions, not only the normal distribution.

5. Explain the concept of Pareto Distribution.

The Pareto Principle, named after economist Vilfredo Pareto, specifies that 80% of consequences come from 20% of the causes, asserting an unequal relationship between inputs and outputs. This principle serves as a general reminder that the relationship between inputs and outputs is not balanced.