Statistics

Assignment 5

1. **How are you going to figure out the average heights of all the trees in Karnataka?**
   Ans: First of all, I'd try to know how many trees there are in Karnataka. That is my population size, expressed by N. I'd also take a sample size that should fairly represent the population. This sample size is expressed by n and it is always smaller than the population size. Then, I'll measure the heights of the sample trees and divide the summation of all trees by the sample size. That is the average height of the sample that is expressed by $\bar{x}$. However, it may not accurately represent the population mean, expressed by μ. But we can get a fairly accurate answer.
2. **What is hypothesis testing and how does it work?**
   Ans: Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution. First, a null hypothesis (H0) and an alternative hypothesis (Ha) are proposed. Alternative hypothesis is just the opposite of the null hypothesis. If the sample obtained has a probability of occurrence less than the pre-specified threshold probability, the significance level, given the null hypothesis is true, the difference between the sample and the null hypothesis is deemed statistically significant. The hypothesis test may then lead to the rejection of the null hypothesis and acceptance of the alternate hypothesis.
3. **Explain the differences between Alpha and Beta errors. Which inaccuracy is the most hazardous?**
   Ans: Alpha error, also known as Type I error, takes place when the null hypothesis is erroneously rejected, and beta error, also known as Type II error, takes place when the null hypothesis is wrongly accepted. Alpha error is called False Positive and Beta error is called False Negative. Which error is more hazardous usually depends on the context. For example, Beta error is more hazardous as it accepts what is not there. In pharmaceutical research, believing an ineffectual drug works can be harmful. However, Alpha error is more hazardous in my opinion. In a trial, it can punish an innocent person. So, if I have to choose one for being "more hazardous" I would say Alpha error or Type I error.
4. **What is the significance of the p-value?**
   Ans: A p-value is used to measure the randomness of observation. It is the probability that an observed difference could have occurred just by random chance. A p value is the evidence against a null hypothesis. It is used in hypothesis testing to help the researcher support or reject the null hypothesis.
5. **What is the Probability Distribution Function and how does it work?**
   Ans: Probability Distribution Function (PDF), also known as Probability Density Function, is the function for mapping random variables to real numbers. PDF is used to determine the likelihood of the random variable falling within a particular range of values, as opposed to taking on any one value.