Simple Overview of Key Statistical Concepts Used in the Subject

Point estimation. Statistical point estimation is a process used to estimate unknown parameters of a population based on a sample. It involves selecting a single value, called a point estimate, that represents the best guess for the quantity of interest.

Point estimation relies on statistical techniques to calculate the estimate. The goal is to find an estimate that is:

- unbiased, meaning it is not systematically too high or too low on average,

- with minimum variance, meaning it is not too much to deviate from the quantity of interest.

However, point estimates are subject to random variation, so they are not guaranteed to be exactly equal to the true population parameter.

Confidence intervals. Statistical confidence intervals provide a range of values (L,U) within which we can reasonably expect the true population parameter to lie. They help to quantify the uncertainty associated with sample point estimates. A wider confidence interval indicates greater uncertainty, while a narrower interval suggests more precise estimates.

The confidence level associated with a confidence interval represents the probability that the interval will contain the true population parameter in repeated sampling. For example, a 95% confidence level means that if we were to construct many intervals from different samples using the same method, approximately 95% of those intervals would contain the true parameter, while about 5% would not.

Hypothesis testing. Hypothesis testing is a statistical method used to draw conclusions about a population based on a sample of data. Some key concepts in hypothesis testing:

- Hypothesis: A hypothesis is a statement about a population parameter, such as the mean or proportion. It is often either the null hypothesis (H_0) or the alternative hypothesis (H_1), which is the opposite of the null hypothesis.

- Test Statistic: A test statistic is a numerical value calculated from the sample data that is used to assess the likelihood of observing the data under the null hypothesis. The choice of test statistic depends on the type of data and the specific hypothesis being tested.

- Level of Significance (α): The level of significance α is the predetermined probability used to determine whether to reject the null hypothesis. Commonly used values are 0.05 or 0.01, representing a 5% or 1% chance of making a Type I error, respectively.

- P-value: The p-value is a measure of evidence against the null hypothesis. It represents the probability of observing a test statistic as extreme as, or more extreme than, the one obtained from the sample data, assuming the null hypothesis is true. A smaller p-value suggests stronger evidence against the null hypothesis.

- Decision Rule: Based on the p-value and the level of significance, a decision rule is established. Usually, if the p-value is smaller than the significance level (0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, if the p-value is greater than or equal to the significance level, the null hypothesis is not rejected.

- Type I Error: A Type I error occurs when the null hypothesis is rejected, but it is actually true in the population. It represents a false positive result, indicating that a difference or relationship is detected when it does not exist.

- Type II Error: A Type II error occurs when the null hypothesis is not rejected, but it is actually false in the population. It represents a false negative result, indicating that a difference or relationship is not detected when it actually exists.