**Inferential statistics**

Inferential statistics is a branch of statistics that deals with drawing conclusions, inferences, or generalizations about a larger population based on the analysis of a smaller sample. It allows researchers to make predictions or estimations about population parameters without having to study the entire population. Some key concepts and techniques in inferential statistics include:

1. Probability: Probability is the measure of the likelihood that a particular event will occur. It plays a crucial role in inferential statistics, as it provides a framework for understanding the randomness and uncertainty in data.
2. Sampling: Sampling is the process of selecting a subset of individuals or observations from a population to make inferences about the whole population. Proper sampling methods are essential for ensuring that the sample is representative of the population.
3. Sampling distribution: The sampling distribution is the distribution of a sample statistic (e.g., mean, proportion, or difference between means) obtained from all possible samples of a given size taken from the population. It provides the basis for making inferences about the population parameter.
4. Confidence intervals: Confidence intervals provide a range of values within which the population parameter is likely to fall, with a certain level of confidence (e.g., 95% confidence interval). They are a way of expressing the uncertainty associated with a sample statistic.
5. Hypothesis testing: Hypothesis testing is a method for comparing sample statistics to make inferences about population parameters. It involves stating a null hypothesis (H0) and an alternative hypothesis (H1), calculating a test statistic, and using a probability distribution to determine the probability of observing the test statistic under the null hypothesis.
6. P-value: The p-value is the probability of obtaining a test statistic as extreme as or more extreme than the one observed, assuming the null hypothesis is true. A small p-value (typically less than 0.05) indicates that the observed result is unlikely to have occurred by chance alone, leading to the rejection of the null hypothesis.
7. Type I and Type II errors: Type I error (false positive) occurs when the null hypothesis is rejected when it is actually true, while Type II error (false negative) occurs when the null hypothesis is not rejected when it is actually false. The balance between these two types of errors is an important consideration in hypothesis testing.
8. Power analysis: Power analysis is used to determine the likelihood of correctly rejecting the null hypothesis when it is false, which is known as the statistical power of a test. It is often used in study design to estimate the sample size required to detect an effect of a given magnitude.

Inferential statistics is widely used in various fields, including social sciences, economics, health sciences, and business, to draw conclusions and make decisions based on data collected from samples.

Hypothesis testing

Hypothesis testing is a fundamental statistical method used in inferential statistics to make decisions or inferences about population parameters based on sample data. It helps researchers test the validity of a claim or assess the evidence in favor of a certain hypothesis. The process of hypothesis testing involves the following key steps:

1. State the hypotheses: Formulate the null hypothesis (H0) and the alternative hypothesis (H1). The null hypothesis usually represents the status quo or no effect (e.g., no difference between two groups or no relationship between variables). The alternative hypothesis represents the claim that the researcher wants to test (e.g., a difference between groups or a relationship between variables).
2. Choose the significance level: The significance level (alpha, α) is the probability of making a Type I error (rejecting the null hypothesis when it is true). Commonly used significance levels are 0.05 (5%) or 0.01 (1%). The smaller the significance level, the more stringent the test and the less likely it is to make a Type I error.
3. Select the appropriate test statistic: Depending on the data type, distribution, and the specific hypothesis being tested, choose a suitable test statistic (e.g., t-test, chi-square test, or ANOVA). The test statistic is a numerical value that helps to compare the observed data with what would be expected under the null hypothesis.
4. Calculate the test statistic and p-value: Using the sample data, compute the test statistic and the corresponding p-value. The p-value represents the probability of observing a test statistic as extreme as or more extreme than the one obtained, assuming the null hypothesis is true.
5. Make a decision: Compare the p-value to the chosen significance level (α):
   * If the p-value is less than or equal to α, reject the null hypothesis (H0) and conclude that there is sufficient evidence in favor of the alternative hypothesis (H1). This means that the observed result is statistically significant and unlikely to have occurred by chance alone.
   * If the p-value is greater than α, fail to reject the null hypothesis (H0) and conclude that there is not enough evidence to support the alternative hypothesis (H1). This does not necessarily mean that the null hypothesis is true, but rather that the data does not provide strong enough evidence against it.

It is important to note that hypothesis testing provides a decision-making framework based on probabilities and is subject to errors (Type I and Type II errors). The results should be interpreted cautiously and in the context of the research question, study design, and other relevant information.

**Primary Errors**

Type I and Type II errors are concepts related to hypothesis testing in the field of inferential statistics. Hypothesis testing is a method for making inferences about population parameters based on sample data. It involves stating a null hypothesis (H0) and an alternative hypothesis (H1) and using a test statistic to compare the sample data to what would be expected under the null hypothesis.

1. Type I error (false positive): This error occurs when the null hypothesis is rejected when it is actually true. In other words, a Type I error is when we mistakenly conclude that there is a significant effect or relationship when there is none in the population. The probability of making a Type I error is denoted by the significance level (alpha, α), which is usually set at 0.05 or 5% in most research studies.
2. Type II error (false negative): This error occurs when the null hypothesis is not rejected when it is actually false. In other words, a Type II error is when we fail to detect a significant effect or relationship that actually exists in the population. The probability of making a Type II error is denoted by beta (β), and the power of a test, which is the probability of correctly rejecting a false null hypothesis, is calculated as 1 - β.

Balancing Type I and Type II errors is important in hypothesis testing, as both errors can have significant consequences in research studies and decision-making processes. The choice of the significance level, sample size, and effect size are some factors that can influence the rates of these errors.

**Other types of errors**

In the context of hypothesis testing and inferential statistics, Type I and Type II errors are the primary errors of concern. However, there are other types of errors that can occur in research studies or data analysis. Some of these errors include:

1. Sampling error: This occurs when the sample selected for a study does not accurately represent the population from which it is drawn. Sampling error can lead to biased estimates of population parameters and affect the generalizability of the results.
2. Measurement error: This error arises from inaccuracies in the instruments or methods used to measure variables in a study. Measurement error can be random (unsystematic) or systematic (biased), and can affect the validity and reliability of the results.
3. Non-sampling error: This is a broad category of errors that are not related to the process of sampling. Examples include errors in data collection, data entry, and data processing, as well as errors due to nonresponse or missing data.
4. Confounding: Confounding occurs when the effect of an independent variable on the dependent variable is mixed with the effect of another variable. Confounding can lead to biased estimates of the relationship between variables and can be addressed through proper study design and statistical techniques, such as stratification, matching, or multivariate regression.
5. Bias: Bias refers to any systematic deviation from the true value that affects the results of a study. There are various types of bias, including selection bias, information bias, and confounding, which can compromise the internal validity of a study.

These errors highlight the importance of careful study design, data collection, and analysis to minimize potential sources of error and ensure the validity and reliability of research findings.

**Common statistical tests**

There are several common statistical tests used in research to test hypotheses and analyze relationships between variables. Here are some of the most frequently used tests, including when to use z and t statistics:

1. Z-test: The z-test is used when comparing a sample mean (or proportion) to a known population mean (or proportion) when the population standard deviation is known. The z-test assumes that the data follow a normal distribution. It is most suitable for large sample sizes (n > 30) or when the population is normally distributed.
2. T-test: The t-test is used when comparing a sample mean to a known population mean when the population standard deviation is unknown. The t-test relies on the t-distribution, which is similar to the normal distribution but with thicker tails. There are three main types of t-tests:

a. One-sample t-test: Used to compare a sample mean with a known or hypothesized population mean. b. Independent two-sample t-test: Used to compare the means of two independent groups (e.g., experimental and control groups). c. Paired t-test: Used to compare the means of two related or dependent groups (e.g., before and after measurements on the same subjects).

1. Chi-square test: The chi-square test is used to test the relationship between two categorical variables in a contingency table. The test compares the observed frequencies in the table to the expected frequencies under the assumption of independence between the variables. It is a non-parametric test that does not assume a specific distribution for the data.
2. Analysis of Variance (ANOVA): ANOVA is used to compare the means of three or more groups. It tests the null hypothesis that all group means are equal, against the alternative hypothesis that at least one group mean is different. There are several types of ANOVA, including one-way ANOVA (for one independent variable) and two-way ANOVA (for two independent variables).
3. Pearson's correlation coefficient: This test measures the strength and direction of a linear relationship between two continuous variables. Pearson's correlation coefficient (r) ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation.
4. Spearman's rank correlation coefficient: Spearman's rank correlation coefficient (rho) measures the strength and direction of a monotonic relationship between two ordinal or continuous variables when data is non-normally distributed or the relationship is non-linear. It is a non-parametric test that assesses the correlation between the ranks of the data rather than the raw data values.
5. Regression analysis: Regression analysis is used to model the relationship between a dependent variable and one or more independent variables. The most common type of regression is linear regression, which models the relationship as a straight line. Other types of regression include logistic regression (for binary outcomes), Poisson regression (for count data), and Cox regression (for time-to-event data).

**Left, Right and Two tailed test**

In summary, z-tests are used when comparing a sample mean or proportion to a population mean or proportion with a known population standard deviation. T-tests are used when comparing a sample mean to a population mean with an unknown population standard deviation, or when comparing means of two independent or paired groups. Other tests, such as chi-square, ANOVA, correlation, and regression, are used to analyze relationships between variables or compare means across multiple groups.

In hypothesis testing, the choice between a left-tailed, right-tailed, or two-tailed test depends on the research question and the alternative hypothesis (H1). Here's a description of when to use each type of test with mathematical notations:

1. Left-tailed test: A left-tailed test is used when the research question is focused on detecting a decrease or a value less than a specified value. The alternative hypothesis (H1) states that the population parameter is less than the hypothesized value.

Mathematically, for a population mean (μ) and a hypothesized value (μ₀): H0: μ ≥ μ₀ H1: μ < μ₀

For a population proportion (p) and a hypothesized value (p₀): H0: p ≥ p₀ H1: p < p₀

In a left-tailed test, the null hypothesis is rejected if the test statistic is smaller than the critical value, which corresponds to the chosen significance level (α) on the left side of the distribution.

1. Right-tailed test: A right-tailed test is used when the research question is focused on detecting an increase or a value greater than a specified value. The alternative hypothesis (H1) states that the population parameter is greater than the hypothesized value.

Mathematically, for a population mean (μ) and a hypothesized value (μ₀): H0: μ ≤ μ₀ H1: μ > μ₀

For a population proportion (p) and a hypothesized value (p₀): H0: p ≤ p₀ H1: p > p₀

In a right-tailed test, the null hypothesis is rejected if the test statistic is larger than the critical value, which corresponds to the chosen significance level (α) on the right side of the distribution.

1. Two-tailed test: A two-tailed test is used when the research question is focused on detecting a difference, but the direction of the difference is not specified. The alternative hypothesis (H1) states that the population parameter is not equal to the hypothesized value.

Mathematically, for a population mean (μ) and a hypothesized value (μ₀): H0: μ = μ₀ H1: μ ≠ μ₀

For a population proportion (p) and a hypothesized value (p₀): H0: p = p₀ H1: p ≠ p₀

In a two-tailed test, the null hypothesis is rejected if the test statistic is either smaller than the critical value on the left side or larger than the critical value on the right side of the distribution. The significance level (α) is split between the two tails, so each tail has a significance level of α/2.

The choice between a left-tailed, right-tailed, or two-tailed test depends on the specific research question and the direction of the effect that the researcher is interested in detecting.

**difference between using p value and rejection region method**

Both the p-value and the rejection region methods are used in hypothesis testing to make a decision about the null hypothesis (H0). However, they approach this decision-making process differently:

1. P-value method: The p-value is the probability of obtaining a test statistic as extreme as or more extreme than the one observed, assuming the null hypothesis is true. The p-value is calculated for the specific sample data, and it is compared to the chosen significance level (α) to make a decision about the null hypothesis.
   * If the p-value ≤ α, then the null hypothesis is rejected, and the result is considered statistically significant. This indicates that the observed data is unlikely to have occurred by chance alone if the null hypothesis is true.
   * If the p-value > α, then the null hypothesis is not rejected, and there is insufficient evidence to support the alternative hypothesis.
2. Rejection region method: The rejection region method, also known as the critical value method, defines a region in the sample space (the critical region) where the null hypothesis will be rejected if the test statistic falls within that region. The critical region is determined based on the chosen significance level (α) and the distribution of the test statistic under the null hypothesis.
   * In a left-tailed test, the critical region is the left tail of the distribution, and the null hypothesis is rejected if the test statistic is smaller than the critical value (which corresponds to the α significance level on the left side of the distribution).
   * In a right-tailed test, the critical region is the right tail of the distribution, and the null hypothesis is rejected if the test statistic is larger than the critical value (which corresponds to the α significance level on the right side of the distribution).
   * In a two-tailed test, the critical region is both tails of the distribution, and the null hypothesis is rejected if the test statistic is either smaller than the critical value on the left side or larger than the critical value on the right side. In this case, the significance level (α) is split between the two tails, so each tail has a significance level of α/2.

The main difference between the p-value and rejection region methods is the way they approach hypothesis testing decision-making. The p-value method calculates the probability based on the specific sample data and compares it to the significance level, while the rejection region method determines a critical region based on the significance level and compares the test statistic to the critical values.

In practice, both methods lead to the same conclusion when making a decision about the null hypothesis. The p-value method is more commonly used in modern statistical software and research, as it provides a continuous measure of the evidence against the null hypothesis and allows for more nuanced interpretations of the results.

**Population Mean known and unknown**

In hypothesis testing, the choice of statistical test depends on whether the population mean is known or unknown. Here, we will discuss two common scenarios: when the population mean is known, and when it is unknown.

1. Population mean is known:

If the population mean (μ) is known, a z-test can be used to compare a sample mean (x̄) to the known population mean. However, this scenario is quite rare in practice because it is unusual to know the population mean without knowing the population standard deviation (σ).

In this case, the null hypothesis (H0) is typically formulated as H0: x̄ = μ₀, where μ₀ is the known population mean. The test statistic (z) for the z-test is calculated as:

z = (x̄ - μ₀) / (σ / √n)

where n is the sample size, and σ is the known population standard deviation. The z-test assumes that the sample data follow a normal distribution or that the sample size is large enough (n > 30) for the central limit theorem to apply.

1. Population mean is unknown:

When the population mean (μ) is unknown, a t-test is used to compare a sample mean (x̄) to a hypothesized population mean (μ₀). The t-test relies on the t-distribution, which is similar to the normal distribution but with thicker tails, and is most suitable when the sample size is small (n < 30) or the population is not normally distributed.

For a one-sample t-test, the null hypothesis (H0) is formulated as H0: μ = μ₀, where μ₀ is the hypothesized population mean. The test statistic (t) for the t-test is calculated as:

t = (x̄ - μ₀) / (s / √n)

where n is the sample size, and s is the sample standard deviation. The t-test assumes that the population data follow a normal distribution, although it is more robust to violations of normality than the z-test, particularly when the sample size is small.

In summary, when the population mean is known, a z-test is used, while a t-test is employed when the population mean is unknown. In practice, it is much more common to encounter situations where the population mean is unknown, and the t-test is widely used in various research settings.

**Likelihood Ratio Test: Neyman Pearson Lemma**

The Neyman-Pearson Lemma provides a criterion for constructing the most powerful test for simple hypotheses, where both the null and alternative hypotheses have completely specified probability distributions. According to the lemma, the most powerful test for a given significance level (α) is based on the likelihood ratio, which is the ratio of the probability density functions (pdfs) or probability mass functions (pmfs) of the two hypotheses:

Likelihood Ratio Test (LRT) = (pdf or pmf of H1) / (pdf or pmf of H0)

To find the most powerful size α test using the Neyman-Pearson Lemma:

1. Calculate the likelihood ratio for each possible outcome in the sample space.
2. Rank the outcomes by their likelihood ratios in descending order.
3. Define the critical region as the set of outcomes with the highest likelihood ratios, such that the sum of the probabilities of those outcomes under H0 is less than or equal to the significance level (α).

A test with the highest power for a given significance level (α) minimizes the probability of making a Type II error (failing to reject H0 when it is false) while controlling the probability of making a Type I error (rejecting H0 when it is true) at the desired level.

In practice, finding the most powerful size α test can be computationally intensive, and the Neyman-Pearson Lemma is primarily applicable to simple hypotheses. However, the concept of the most powerful size α test provides a foundation for understanding the trade-offs between the power of a test and the probabilities of making Type I and Type II errors. In many real-world applications, researchers rely on well-established statistical tests (e.g., t-test, chi-square test, or ANOVA) that have been developed based on the principles of the Neyman-Pearson approach and other statistical frameworks.

**Best Critical Region**

The Best Critical Region (BCR) is a concept in statistical hypothesis testing related to the Neyman-Pearson Lemma, which is a fundamental result in the decision theory approach to hypothesis testing. The BCR is a region in the sample space in which the null hypothesis (H0) is rejected in favor of the alternative hypothesis (H1) with the highest power (i.e., the highest probability of correctly rejecting H0 when it is false) for a given significance level (α).

The Neyman-Pearson Lemma provides a criterion for constructing the most powerful test for simple hypotheses, where both the null and alternative hypotheses have completely specified probability distributions. The lemma states that the most powerful test for a given significance level (α) is based on the likelihood ratio, which is the ratio of the probability density functions (pdfs) or probability mass functions (pmfs) of the two hypotheses.

The likelihood ratio test (LRT) is defined as:

LRT = (pdf or pmf of H1) / (pdf or pmf of H0)

To construct the best critical region using the Neyman-Pearson Lemma:

1. Calculate the likelihood ratio for each possible outcome in the sample space.
2. Rank the outcomes by their likelihood ratios in descending order.
3. Define the critical region as the set of outcomes with the highest likelihood ratios such that the sum of the probabilities of those outcomes under H0 is less than or equal to the significance level (α).

In practice, finding the best critical region can be computationally intensive, and the Neyman-Pearson Lemma is primarily applicable to simple hypotheses. However, the concept of the BCR provides a foundation for understanding the trade-offs between the power of a test and the probability of making Type I and Type II errors. In many real-world applications, researchers rely on well-established statistical tests (e.g., t-test, chi-square test, or ANOVA) that have been developed based on the principles of the Neyman-Pearson approach and other statistical frameworks.