Estimate

In statistics, an **estimate** is a value or range of values calculated from sample data and used to infer the value of an unknown population parameter. Estimates are crucial for making inferences about populations based on sample data, especially when it is impractical or impossible to collect data from the entire population.

Here are some key concepts related to estimates in statistics:

Types of Estimates

1. **Point Estimate**:
   * A single value that serves as an estimate of a population parameter.
   * **Example**: The sample mean ((\bar{x})) is a point estimate of the population mean ((\mu)).
2. **Interval Estimate**:
   * A range of values within which the population parameter is expected to lie, typically expressed with a certain level of confidence.
   * **Example**: A 95% confidence interval for the population mean.

Point Estimates

Point estimates are derived from sample data and provide a single best guess for a population parameter. Common point estimates include:

* **Sample Mean ((\bar{x}))**: Estimates the population mean ((\mu)).
* **Sample Proportion ((p))**: Estimates the population proportion ((P)).
* **Sample Variance ((s^2))**: Estimates the population variance ((\sigma^2)).

## What is a Proportion in Statistics?

In statistics, a [**proportion**](http://study.com/learn/lesson/what-is-proportion-math.html) is simply the ratio of the number of successful outcomes of an event to the number of total outcomes of the event. As a simple example, consider the flipping of a fair two-sided coin. Let a successful flip be a flip that lands on heads. If the coin is flipped ten times and lands on heads six times, then the proportion of successes is 610, or 0.6.

Example:

<https://digfir-published.macmillanusa.com/psbe4e/psbe4e_ch8_2.html>

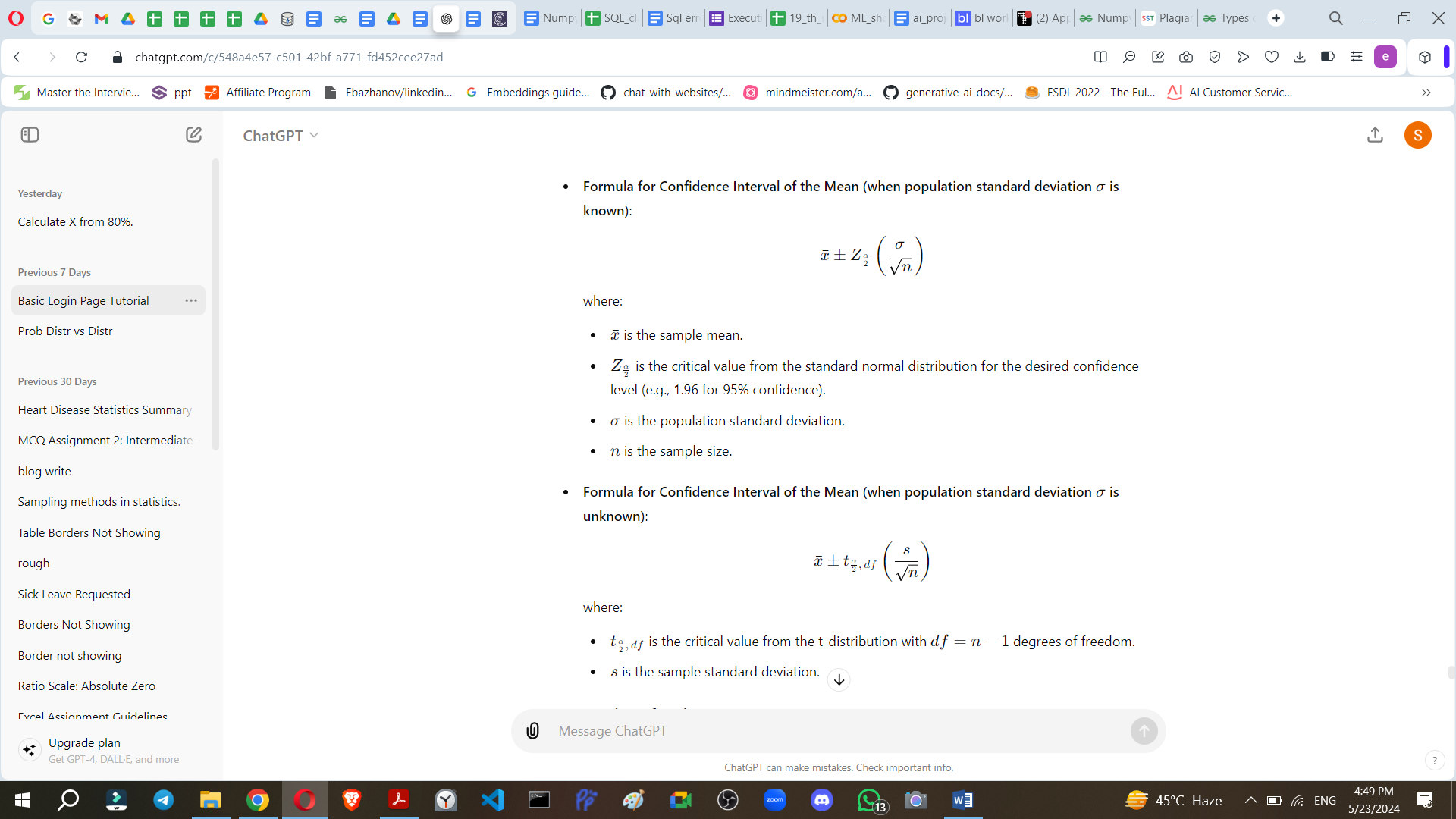
Interval Estimates

Interval estimates provide a range within which the parameter is expected to fall. The most common form of interval estimate is the confidence interval.

Confidence Interval

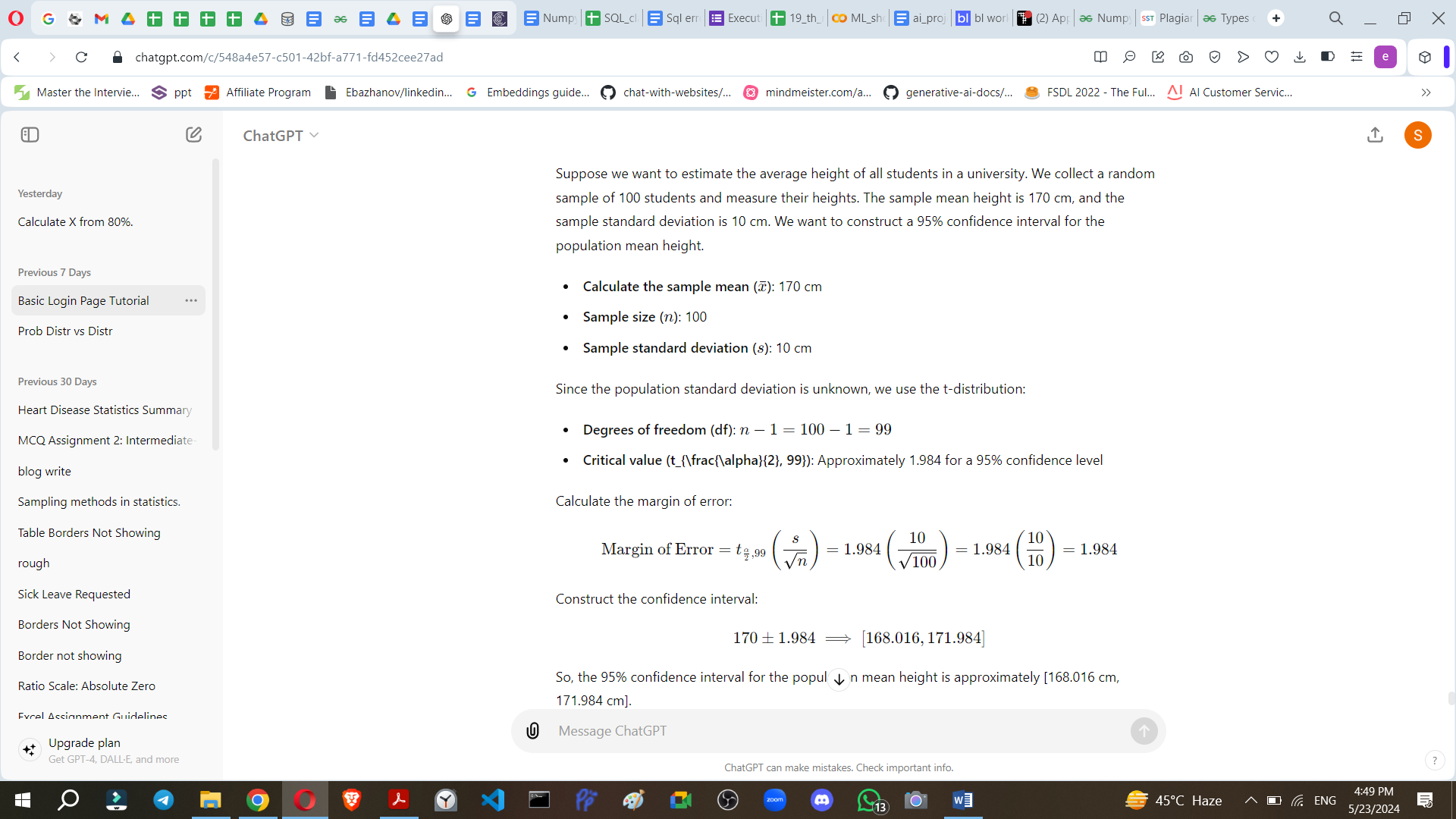
A confidence interval gives an estimated range of values which is likely to include the population parameter, calculated from the sample data. It is expressed with a certain level of confidence, such as 95% or 99%.

* Properties of Estimates



1. **Unbiasedness**:
   * An estimator is unbiased if its expected value is equal to the true parameter value.
   * **Example**: The sample mean (bar{x}) is an unbiased estimator of the population mean (mu).
2. **Consistency**:
   * An estimator is consistent if it converges to the true parameter value as the sample size increases.
   * **Example**: The sample proportion (p) is a consistent estimator of the population proportion (P).
3. **Efficiency**:
   * An estimator is efficient if it has the smallest variance among all unbiased estimators.
   * **Example**: Under certain conditions, the sample mean (\bar{x}) is the most efficient estimator of the population mean (\mu).
4. **Sufficiency**:
   * An estimator is sufficient if it uses all the information in the sample relevant to estimating the parameter.
   * **Example**: The sample mean is a sufficient statistic for the population mean in a normal distribution.

Example of Estimation



In summary, estimates in statistics are crucial tools for making inferences about population parameters based on sample data. They come in two main forms: point estimates and interval estimates, each with its own properties and applications.

1. Hypothesis testing and mechanism
2. P – value
3. Z-test with example
4. Student-t Distrubution test
5. T- Stats and t-test
6. When use T-test and Z-Test
7. Type 1 error and type 2 error
8. Confidence Intervel and Margin of error
9. CHI- square test
10. Chi-square for goodness of fit
11. Chi-square in python
12. F-Distribution
13. Variance Ratio test(F test)
14. Anovo
15. Assumptions in ANOVo
16. Types of ANOVA
17. Partition of Variance in anova

<https://chatgpt.com/share/705e42c8-fd2c-4690-9c60-326896da0be6>

1. **Hypothesis Testing and Mechanism:**
   * **Definition:** Hypothesis testing is a statistical method used to make inferences about a population based on sample data. It involves setting up a hypothesis (null and alternative), collecting data, and using statistical techniques to determine whether the observed differences are statistically significant.
   * **Use Case:** Testing whether a new drug treatment is effective compared to a placebo, evaluating the impact of a marketing campaign on sales, etc.
   * **Example:** Suppose we want to test if the mean height of students in a school is different from a national average of 65 inches. We collect a sample of heights and perform hypothesis testing to draw conclusions.
2. **P-Value:**
   * **Definition:** The p-value is the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is true.
   * **Use Case:** It helps determine the significance of results in hypothesis testing.
   * **Example:** In our height example, if the p-value is 0.03, it suggests a 3% probability of observing the sample data if the true mean height were the national average.
3. **Z-Test with Example:**
   * **Definition:** A Z-test is used when the population mean and standard deviation are known. It tests whether a sample mean differs significantly from the population mean.
   * **Use Case:** Comparing a sample mean to a known population mean.
   * **Example:** Testing if the mean score of students in a class (sample) differs significantly from the national average score (population).
4. **Student-t Distribution and t-Test:**
   * **Definition:** The t-distribution is used when the sample size is small or the population standard deviation is unknown. A t-test is used to determine if there is a significant difference between the means of two groups.
   * **Use Case:** Comparing the effectiveness of two different teaching methods on student performance.
   * **Example:** Testing if the mean exam scores of two groups (Group A and Group B) are significantly different.
5. **When to Use T-Test and Z-Test:**
   * **T-Test:** When the sample size is small (typically less than 30) or the population standard deviation is unknown.
   * **Z-Test:** When the population standard deviation is known and the sample size is large (typically greater than 30).
6. **Type 1 Error and Type 2 Error:**
   * **Type 1 Error:** Rejecting a true null hypothesis (false positive).
   * **Type 2 Error:** Failing to reject a false null hypothesis (false negative).
7. **Confidence Interval and Margin of Error:**
   * **Confidence Interval:** A range of values within which we are confident the population parameter lies.
   * **Margin of Error:** The amount of random sampling error in a survey's results.
8. **Chi-Square Test:**
   * **Definition:** A statistical test used to determine if there is a significant association between categorical variables.
   * **Use Case:** Testing independence in contingency tables, goodness of fit tests.
   * **Example:** Testing if there is a relationship between gender and voting preference in a survey.
9. **Chi-Square for Goodness of Fit:**
   * **Definition:** Chi-square test used to compare observed frequencies with expected frequencies to determine if they differ significantly.
10. **Chi-Square in Python:**
    * Python libraries like scipy and numpy provide functions for chi-square tests.
11. **F-Distribution and Variance Ratio Test (F Test):**
    * **F-Distribution:** Used for comparing the variances of two populations.
    * **Variance Ratio Test (F Test):** Determines if the variances of two populations are equal.
12. **ANOVA (Analysis of Variance):**
    * **Definition:** A statistical technique used to analyze differences among means of two or more groups.
13. **Assumptions in ANOVA:**
    * Independence of observations, normality of residuals, and homogeneity of variances.
14. **Types of ANOVA:**
    * One-way ANOVA, Two-way ANOVA, and N-way ANOVA.
15. **Partition of Variance in ANOVA:**
    * Total variance is partitioned into components attributed to different sources (e.g., between groups and within groups).

Each of these topics can delve quite deeply into statistical theory and application. If you'd like more detailed explanations or examples with datasets and calculations, let me know which specific topic you're interested in!