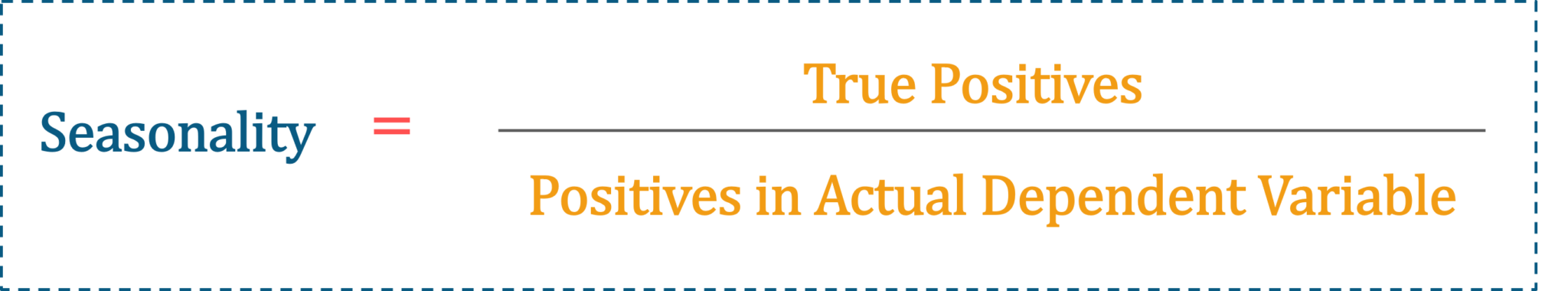
**A/B Testing :**

* A/B testing is the statistical hypothesis testing for a randomized experiment with two variables A and B. Also known as the split testing, it is an analytical method that estimates population parameters based on sample statistics. This test compares two web pages by showing two variants A and B, to a similar number of visitors, and the variant which gives better conversion rate wins.
* The goal of A/B Testing is to identify if there are any changes to the web page. For example, if you have a banner ad on which you have spent an ample amount of money. Then, you can find out the return of investment i.e. the click rate through the banner ad.

**Statistical power of sensitivity :**

* The statistical power of sensitivity is used to validate the accuracy of a classifier. This classifier can be either Logistic Regression, Support Vector Machine, Random Forest etc.
* If I have to define sensitivity, then sensitivity is nothing but the ratio of Predicted True Events to Total Events. Now, True Events are the events which were true and the model also predicts them as true.



|  |  |  |
| --- | --- | --- |
| **Univariate** | **Bivariate** | **Multivariate** |
| * A descriptive statistical technique that can be differentiated based on the count of variables involved at a given instance of time. | * This analysis is used to find the difference between two variables at a time. | * The study of more than two variables is nothing but multivariate analysis. * This analysis is used to understand the effect of variables on the responses. |

|  |  |
| --- | --- |
| **Eigenvectors** | **Eigenvalue** |
| * Eigenvectors are basically used to understand linear transformations. * These are calculated for a correlation or a covariance matrix. * For definition purposes, you can say that Eigenvectors are the directions along which a specific linear transformation acts either by flipping, compressing or stretching. | * Eigenvalues can be referred to as the strength of the transformation or the factor by which the compression occurs in the direction of eigenvectors. |

|  |  |  |
| --- | --- | --- |
| **Descriptive** | **Predictive** | **Prescriptive** |
| It provides insights into the past to answer “what has happened” | Understands the future to answer “what could happen” | Suggest various courses of action to answer “what should you do” |
| Uses data aggregation and data mining techniques | Uses statistical models and forecasting techniques | Uses simulation algorithms and optimization techniques to advise possible outcomes |
| Example: An ice cream company can analyze how much ice cream was sold, which flavors were sold, and whether more or less ice cream was sold than the day before | Example: An ice cream company can analyze how much ice cream was sold, which flavors were sold, and whether more or less ice cream was sold than the day before | Example: Lower prices to increase the sale of ice creams, produce more/fewer quantities of a specific flavor of ice cream |
|  |  |  |

**What are the different types of sampling techniques used by data analysts?**

Sampling is a statistical method to select a subset of data from an entire dataset (population) to estimate the characteristics of the whole population.

There are majorly five types of sampling methods:

* Simple random sampling
* Systematic sampling
* Cluster sampling
* Stratified sampling
* Judgmental or purposive sampling

|  |  |
| --- | --- |
| Treemaps | Heatmaps |
| Treemaps are used to display data in nested rectangles. | Heat maps can visualize measures against dimensions with the help of colors and size to differentiate one or more dimensions and up to two measures. |
| You use dimensions to define the structure of the treemap, and measures to define the size or color of the individual rectangles. | The layout is like a text table with variations in values encoded as colors. |
| Treemaps are a relatively simple data visualization that can provide insight in a visually attractive format. | In the heatmap, you can quickly see a wide array of information. |

**Properties of a normal distribution**

* The mean, mode and median are all equal.
* The **curve** is symmetric at the center (i.e. around the mean, μ).
* Exactly half of the values are to the left of center and exactly half the values are to the right.
* The total area under the **curve** is 1.

https://mathbitsnotebook.com/Algebra1/StatisticsData/STSD.html