Assignment-6

1. The outcome of a flip of a coin.

2. Discrete

3. PDF

4. Mean

5. Variance

6. Standard deviation

7. 0 and 10

8. Bootstrap

9. summarized

10.- Histograms and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability before using advanced statistical analysis techniques.Histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space although histograms are better in displaying the distribution of data, you can use a box plot to tell if the distribution is symmetric or skewed. You can use histograms and box plots to verify whether an improvement has been achieved by exploring the data before and after the improvement initiative. Both tools can be helpful to identify whether variability is within specification limits, whether the process is capable, and whether there is a shift in the process over time. You can use histograms and box plots to verify whether an improvement has been achieved by exploring the data before and after the improvement initiative.

11.- Good metrics are important to your company growth and objectives your key metrics should always be closely tied to your primary objective. A good metric example might be month-on-month revenue growth or LTV: CAC ratio. ‘Important’ is somewhat subjective since growth for one company may be centered around revenue while another company may focus more on user growth. Good metrics measure progress, which means there needs to be room for improvement. For example, reducing churn by 0.8% or increasing your activation rate by 3%. One exception to this might be customer satisfaction - if you’re already at 100%, your team will be focused on maintaining that level instead of improving it.

1.- Pirate Metrics: The five metrics in this framework for online startups represent the different stages of customer behavior - Acquisition, Activation, Retention, Referral, and Revenue (AARRR). Check out this Beginner’s Guide to Pirate Metrics or these examples to learn more. USEFUL FOR: online startups who need a framework to structure their key metrics around.

2.IPA: (Important, Potential Improvement, Authority): The IPA rule ensures you’re focusing on only important metrics that have potential for improvement and that you have the authority to improve. USEFUL FOR: deciding what metrics to display on your dashboard.

3.TIE: (Trackable, Important, Explainable): Good metrics are always trackable, important, and explainable. This simple framework makes the complicated task of choosing metrics straightforward and easy to understand. USEFUL FOR: validating and revising existing metrics you may be tracking.

4. KPIs: (Key Performance Indicators): Use this option to select only the most important metrics that help you understand how you’re performing against your goals. KPIs are given context in terms of a target or benchmark which relates back to a business’s objectives and goals.USEFUL FOR: determining a select few, high-level metrics for the entire team to focus on.

All good metrics need goals without context or parameters, our ‘good metrics’ aren’t nearly as good. The parameters for our metrics can be either benchmarks (an industry average or internal reference point) or goals (growth targets we set for ourselves), or a combination of the two.

Goals can be defined as the specific target (i.e. dollar amount, number, percentage, etc.) you wish to reach. Goals should always align with your overall objective and are most effective when time bound. For example, if one of your key metrics is Churn Rate - currently at 4.2%, then your goal might be to reduce churn to 3.5% by the end of this quarter.

Goals can be defined as the specific target (i.e. dollar amount, number, percentage, etc.) you wish to reach. Goals should always align with your overall objective and are most effective when time bound. For example, if one of your key metrics is Churn Rate - currently at 4.2%, then your goal might be to reduce churn to 3.5% by the end of this quarter.

12. ans- Statistical significance is calculated using a p-value, which tells you the probability of your result being observed, given that a certain statement (the null hypothesis) is true. If this p-value is less than the significance level set (usually 0.05), the experimenter can assume that the null hypothesis is false and accept the alternative hypothesis. Using a simple t-test,

you can calculate a p-value and determine significance between two different groups of a dataset.

1-DEFINE YOUR HYPOTHESE= The hypothesis is a statement about your experimental data and the differences that may be occurring in the population. For any experiment, there is both a null and an alternative hypothesis. Generally, you will be comparing two groups to see if they are the same or different.

The null hypothesis (H0) generally states that there is no difference between your two data sets. For example: Students who read the material before class do not get better final grades.

The alternative hypothesis (Ha) is the opposite of the null hypothesis and is the statement you are trying to support with your experimental data. For example: Students who read the material before class do get better final grades.

2…Set the significance level to determine how unusual your data must be before it can be considered significant. The significance level (also called alpha) is the threshold that you set to determine significance. If your p-value is less than or equal to the set significance level, the data is considered statistically significant.

As a general rule, the significance level (or alpha) is commonly set to 0.05, meaning that the probability of observing the differences seen in your data by chance is just 5%.

A higher confidence level (and, thus, a lower p-value) means the results are more significant.

If you want higher confidence in your data, set the p-value lower to 0.01. Lower p-values are generally used in manufacturing when detecting flaws in products. It is very important to have high confidence that every part will work exactly as it is supposed to.

For most hypothesis-driven experiments, a significance level of 0.05 is acceptable.

3…Decide to use a one-tailed or two-tailed test. One of the assumptions a t-test makes is that your data is distributed normally. A normal distribution of data forms a bell curve with the majority of the samples falling in the middle. The t-test is a mathematical test to see if your data falls outside of the normal distribution, either above or below, in the “tails” of the curve.

A one-tailed test is more powerful than a two-tailed test, as it examines the potential of a relationship in a single direction (such as above the control group), while a two-tailed test examines the potential of a relationship in both directions (such as either above or below the control group).

If you are not sure if your data will be above or below the control group, use a two-tailed test. This allows you to test for significance in either direction.

If you know which direction you are expecting your data to trend towards, use a one-tailed test. In the given example, you expect the student’s grades to improve; therefore, you will use a one-tailed test.

4… Determine sample size with a power analysis. The power of a test is the probability of observing the expected result.

Researchers usually do a small pilot study to inform their power analysis and determine the sample size needed for a larger, comprehensive study.

If you do not have the means to do a complex pilot study, make some estimations about possible means based on reading the literature and studies that other individuals may have performed. This will give you a good place to start for sample size.

14. Ans- “The mean is typically better when the data follow a symmetric distribution. When the data are skewed, the median is more useful because the mean will be distorted by outliers.”

MEAN = The average taken for a set of numbers is called a mean.

2-Add all of the numbers together and divide the sum by the total number of values.

3-The result is the mean or average score.

Example= To find the average of four numbers 2,4,6 and 8 we need to add the number first.

2+4+6+8=20

Divide the sum by the total number of numbers.

20/4=5 is the average or mean

MEDIAN = The middle value in the data set is called the median.

2- Place all the given number in an ascending order

3-The next step is to find the middle number on the list it is called the median.

Example= 4,2,8,10,19.

* Arrange the number in ascending order. 2,4,8,10,19.
* As the total numbers are 5, so the middle number 8 is the median here.

15.Ans- Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample. The likelihood function itself is not probability (nor density) because its argument is the parameter T of the distribution, not the random (vector) variable X itself. For example, the sum (or integral) of the likelihood function over all possible values of T should not be equal to 1.

1