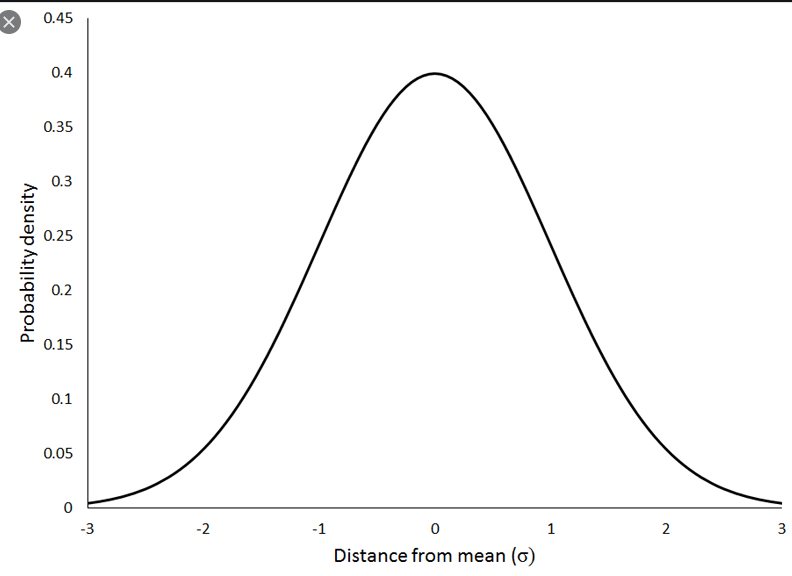
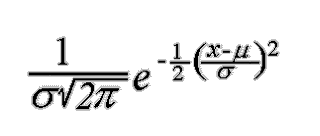
1. What do you understand by the term Normal Distribution?

* It is a type of continuous probability distribution.
* It is also referred to as Bell Curve sometimes due to the shape
* 
* The density curve is symmetrical and centered about the mean of data.
* A normal curve can be represented as 
* Most simple form of a Normal distribution is standard normal distribution where mean is 0 and SD is 1.

1. How do you handle missing data? What imputation techniques do you recommend?

* Missing data can be handled by 2 ways
* Dropping null values
* If there are very less(less than 7-8%) null values in the dataset we can directly drop the null values
* Imputing null values
* If the amount of null values is more it is better to impute values rather than dropping values as dropping will lead to loss of data.
* Imputation is of 2 types
* Mean imputation for continuous data.
* Mode imputation for categorical data.

1. What is A/B testing?

* A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment
* A/B testing includes 3 steps:
* Make a hypothesis
* H0(null hypothesis): this states that there will be no difference between results of control and variant groups
* Ha(alternative hypothesis): this states that there will be either positive or negative difference between results of control and test groups.
* Create control group and test group
* Control group receives a product or service without any changes
* Test group will receive a product or service with the proposed changes
* Conduct A/B test and collect data.
* The daily conversion rates are then measured for both control and test groups to study the impact of the changes made

1. Is mean imputation of missing data acceptable practice?

* Mean imputation is done when some values are missing in continuous columns.
* Mean imputation can be risky incase there are many null values as the mean may be biased, it is better to drop the column on whole if its correlation with the target is less.

1. What is linear regression in statistics?

* Linear regression is a basic and commonly used type of predictive analysis.
* The simplest form of linear regression with one dependent and one independent variable is y=mx+c where y is dependent variable, x is independent variable, m is slope of the regression line, c is the y intercept.
* This regression technique help us establish direct relation between the independent and dependent variable.
* Following are the types of regressions:
* **Simple linear regression :** 1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous)
* [**Multiple linear regression**](https://www.statisticssolutions.com/data-analysis-plan-multiple-linear-regression/)**:** 1 dependent variable (interval or ratio) , 2+ independent variables (interval or ratio or dichotomous)
* [**Logistic regression**](https://www.statisticssolutions.com/data-analysis-plan-logistic-regression/)**:** 1 dependent variable (dichotomous), 2+ independent variable(s) (interval or ratio or dichotomous)
* [**Ordinal regression**](https://www.statisticssolutions.com/data-analysis-plan-ordinal-regression/)**:** 1 dependent variable (ordinal), 1+ independent variable(s) (nominal or dichotomous)
* [**Multinomial regression**](https://www.statisticssolutions.com/data-analysis-plan-multinominal-logistic-regression/)**:** 1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio or dichotomous)
* [**Discriminant analysis**](https://www.statisticssolutions.com/discriminant-analysis-independent-variables/)**:** 1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio)

1. What are the various branches of statistics?

* 2 main branches of statistics are
* Descriptive Statistics
* [Descriptive statistics](https://explorable.com/descriptive-statistics) deals with the presentation and collection of data
* This is usually the first part of a statistical analysis
* This includes EDA, preprocessing and visualization normally
* Inferential statistics
* This involves drawing right conclusion from the statistical analysis that has been performed using descriptive analysis.
* Results for a sample data are studied to check how the data behaves in general.
* This includes the prediction part of the process.