Module	Topic	Usage in Machine Learning		
Descriptive Statistics	What is Stats/Types of Stats			
	Population Vs Sample	In machine learning, a population might refer to the entire set of data relevant to a problem, while a sample would be a subset of that data. Training a model typically happens on a sample of the total data (the training set), which is assumed to be representative of the population. This concept is used to perform inferential statistics and to estimate the model's performance on unseen data.		
	Types of Data	Understanding the type of data you're working with helps in selecting the appropriate preprocessing techniques, feature engineering methods, and machine learning models.		
	Measures of Central Tendency	These measures provide the central value of a data distribution. They are used to understand the 'typical' value in a dataset and are used in various areas of machine learning including exploratory data analysis, outlier detection, and data imputation.		
	- Mean		Yellow	Important
	- Median		Red	Extremely Important
	- Mode		[L]	Later
	- Weighted Mean [L]			
	- Trimmed Mean [L]			
	Measure of Dispersion	These measures provide insights into the spread or variability of the data distribution. They help in understanding the consistency in the data and are also used in exploratory data analysis, outlier detection, feature normalization, etc.		
	- Range			
	- Variance			
	- Standard Deviation			
	- Coefficient of Variation			
	Quantiles and Percentiles	These help in understanding the distribution of data and are used in descriptive statistics, outlier detection, and setting up thresholds for decision-making.		
	5 number summary and BoxPlot	These are used in the exploratory data analysis phase to understand the data distribution and identify outliers. Boxplots graphically depict the minimum, first quartile, median, third quartile, and maximum of a dataset.		
	Skewness	These are used to understand the asymmetry and tailedness of the data distribution, respectively. They're particularly useful in exploratory data analysis, informing data transformations needed to meet the assumptions of some machine learning algorithms.		
	Kurtosis [L]			
	Plotting Graphs	Graphical analysis is crucial in the exploratory phase of machine learning. It helps in understanding the distributions of individual variables (univariate), relationships between two variables (bivariate), or complex interactions among multiple variables (multivariate).		
	- Univariate Analysis			
	- Bivariate Analysis			
	- Multivariate Analysis			
Correlation	Covariance	Covariance is a measure that indicates the extent to which two variables change in tandem. The covariance matrix, on the other hand, gives the covariance between each pair of features in a dataset. These concepts are used in many machine learning algorithms, such as Principal Component Analysis (PCA) for dimensionality reduction, or Gaussian Mixture Models for clustering.		
	Covariance Matrix			
	Pearson Correlation Coefficient	This statistic measures the linear relationship between two datasets. It's used in feature selection, where highly correlated input features can be identified and reduced, to improve the performance and interpretability of the model.		
	Spearman Correlation Coefficient [L]	This measures the monotonic relationship between two datasets. It's useful when the data doesn't meet the assumptions of Pearson's correlation (linearity, normality). It can be used in the same contexts as Pearson's correlation coefficient.		
	Correlation and Causation	Correlation measures association between variables, while causation indicates a cause-effect relationship. In machine learning, it's important to remember that correlation doesn't imply causation, and algorithms based purely on correlation might fail to generalize well.		
Probability Distributions	Random Variables	Random variables and their distributions form the mathematical basis of probabilistic machine learning algorithms. They help us understand the data's inherent randomness and variability, and guide the choice and behavior of algorithms.		
	What are Probability Distributions			
	Why are Probability Distributions important			
	Probability Distribution Functions and it's types			
	Probability Mass Function (PMF)	These concepts are critical in understanding and manipulating discrete random variables, often used in algorithms like Naive Bayes, Hidden Markov Models, etc.		
	CDF of PMF			
	Probability Density Function(PDF)	These are used for continuous random variables. For instance, in the Gaussian Mixture Model, each cluster is modeled as a Gaussian distribution with its PDF.		
	CDF of PDF			
	Density Estimation [L]	Density estimation is the construction of an estimate of the probability distribution that generated a dataset. It's used in unsupervised learning for tasks such as anomaly detection. Kernel Density Estimation (KDE), a non-parametric way to estimate the PDF of a random variable, is particularly useful when no suitable parametric form of the data is known.		
	Parametric Density Estimation [L]			
	Non-Parametric Density Estimation [L]			
	Kernel Density Estimation(KDE) [L]			
	How to use PDF/PMF and CDF in Analysis	hese concepts are used for data analysis and visualization, to understand and communicate the distribution and trends in the data. In machine learning, these analyses can inform the choice of model, preprocessing steps, and potential feature engineering.		
	2D Density Plots	These plots are a useful tool in exploratory data analysis for visualizing the relationship and density between two numerical variables. They can reveal patterns and associations in the data that can guide subsequent modeling steps. For instance, they could help identify clusters for a clustering algorithm in unsupervised learning.		
Types of Probability Dist	Normal Distribution	This distribution is fundamental to many machine learning algorithms, including linear regression, logistic regression, and any algorithm that uses these as a base, such as neural networks. Also, many statistical methods require the assumption of normally distributed errors.		
	- Properties of Normal Distribution			
	- CDF of Normal Distribution			

Module	Topic	Usage in Machine Learning
	- Standard Normal Variate	
	Uniform Distribution	This distribution is used in random forest algorithms for feature splits, and also in initialization of weights in neural networks. It is also used in methods like grid search where you need to randomly sample parameters.
	Bernaulli Distribution	Used in algorithms that model binary outcomes, such as the Bernoulli Naive Bayes classifier and logistic regression.
	Binomial Distribution	Used in modelling the number of successes in a fixed number of Bernoulli trials, often applied in classification problems.
	Multinomial Distribution	Text Classification, topic modelling, deep learning and word embeddings
	Log Normal Distribution	Useful in various contexts, such as when dealing with variables that are the multiplicative product of other variables, or when working with data that exhibit skewness.
	Pareto Distribution [L]	Often used in the realm of anomaly detection or for studying phenomena in the domain of social, quality control, and economic sciences.
	Chi-square Distribution	Chi-square tests use this distribution extensively to test relationships between categorical variables. The chi-square statistic is also used in the context of feature selection.
	Student's T Distribution	Plays a crucial role in formulating the confidence interval when the sample size is small and/or when the population standard deviation is unknown.
	Poisson Distribution [L]	Used for modeling the number of times an event might occur within a set time or space. It's often used in queuing theory and for time-series prediction models.
	Beta Distribution [L]	This is a versatile distribution often used in Bayesian methods, and is also the conjugate prior for the Bernoulli, binomial, negative binomial and geometric distributions.
	Gamma Distribution [L]	The Gamma distribution is used in a variety of fields, including queuing models, climatology, and financial services. It's the conjugate prior of the Poisson, exponential, and normal distributions.
	Transformations	These are used to make the data conform to the assumptions of a machine learning algorithm, enhance the performance of the algorithm, or help visualize the data. Common examples are the logarithmic, square root, and z-score standardization transformations.
Confidence Intervals	Point Estimates	Point estimates are used to provide a single predicted value for a variable of interest. They are used in a wide range of machine learning algorithms to make predictions. Confidence intervals, on the other hand, give us a range of possible values within which we can expect the true population parameter to lie, with a given level of confidence. They are used to understand the reliability of point estimates and are often used to report the results of models.
	Confidence Intervals	
	Confidence Interval(Sigma Known)	
	Confidence Interval(Sigma Unknown)	
	Interpreting Confidence Interval	
	Margin of Error and factors affecting it	
Central Limit Theorem	Sampling Distribution	The concept of a sampling distribution is used to make inferences about a population from a sample. The Central Limit Theorem (CLT) is a fundamental theorem in statistics that states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the shape of the population distribution. This is the foundation for many machine learning methods and is often used in hypothesis testing and in creating confidence intervals.
	What is CLT	
	Standard Error	This is used to understand the variability in a point estimate. In machine learning, it's often used in constructing confidence intervals for model parameters and in hypothesis testing.
Hypothesis Tests	What is Hypothesis Testing?	Hypothesis testing is used extensively in machine learning, especially in model selection, feature selection, and in checking assumptions related to specific models. For instance, a t-test might be used to determine if the means of two sets of results (like two algorithms) are significantly different.
	Null and Alternate Hypothesis	These are fundamental components of all hypothesis tests. The null hypothesis typically represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved.
	Steps involved in a Hypothesis Test	
	Performing Z-test	These are all components of hypothesis testing, and they're used to make decisions about whether the observed effect in our sample is real or happened due to chance. These concepts are used in feature selection, model validation, and comparisons between models.
	Rejection Region Approach	
	Type 1 Vs Type 2 Errors	
	One Sided vs 2 sided tests	
	Statistical Power	This is the ability of a hypothesis test to detect an effect, if the effect actually exists. In machine learning, power analysis can be used to estimate the minimum number of observations required to detect an effect.
	P-value	The p-value is used in hypothesis testing to help support or reject the null hypothesis. It represents the probability that the results of your test occurred at random. If p-value is small (typically ≤ 0.05), it indicates strong evidence to reject the null hypothesis. In machine learning, p-values are often used in feature selection where the null hypothesis is that the feature has no effect on the target variable.
	How to interpret P-values	
Types of Hypothesis Tes	Z-test	Z-tests are statistical calculations that can be used to compare population means to a sample's. The Z-test is used in machine learning when the data is normally distributed and the population variance is known. It's often used in A/B testing to decide whether two groups' mean outcomes are different.
	T-test	T-tests are used when the data is normally distributed but the population variance is unknown.
	- Single Sample T-test	compares the mean of a single sample to a known population mean.
	- Independent 2 sample t-test	compares the means of two independent samples.
	- Paired 2 sample t-test	compares the means of the same group at two different times (say, before and after a treatment). In machine learning, t-tests are often used in experiments designed to compare the performance of two different algorithms on the same problem.
	Chi-square Test	The Chi-square test is used when dealing with categorical variables. It helps to establish if there's a statistically significant relationship between categorical variables.
	Chi-square Goodness of Fit Test	determines if a sample data matches a population.
	Chi-square Test of Independence	checks the relationship between two categorical variables.

Module	Topic	Usage in Machine Learning			
	ANOVA	ANOVA tests the hypothesis that the means of two or more populations are equal. ANOVAs assess the importance of one or more factors by comparing the response variable means at the different factor levels. The null hypothesis states that all population means are equal while the alternative hypothesis states that at least one is different.			
	One Way Anova	It's used to test for differences among at least three groups, as they relate to one factor or variable.			
	Two Way Anova	It's used to compare the mean differences between groups that have been split on two independent variables.			
	F-test				
	Levene Test [L]	This test assesses the equality of variances for a variable calculated for two or more groups. It's often used in feature selection where the null hypothesis is that the variances are equal.			
	Shapiro Wilk Test [L]	This test is used to check the normality of a distribution. Many machine learning algorithms assume normal distribution, making this test quite useful.			
	K-S Test [L]	he K-S test is a non-parametric test that compares a sample with a reference probability distribution, or two samples with each other. It's used in goodness-of-fit tests.			
	Fisher's Test [L]	Fisher's test is used to determine if there are nonrandom associations between two categorical variables.			
Miscellaneous Topics	Chebyshev's Inequality [L]	This mathematical theorem provides a universal boundary on the spread of data, irrespective of the shape of the distribution. It's useful for understanding the range within which most data points lie and can be applied for outlier detection. Chebyshev's inequality is also used in the analysis and proof of convergence of some machine learning algorithms.			
	QQ Plot	A QQ (Quantile-Quantile) plot is a graphical tool to help us assess if a dataset is distributed in a certain way. It's often used to check the assumption of normality in data. Normality of residuals is an assumption in certain statistical and machine learning models, so this can help in diagnostic analysis of these models.			
	Sampling	Sampling is the technique of selecting a subset of individuals from a statistical population to estimate characteristics of the population. It's widely used in machine learning, especially in the context of large datasets, where it may be computationally infeasible to use the entire population. Techniques such as train-test split, k-fold cross-validation, and stratified sampling all involve sampling principles.			
	Resampling Techniques	Cross Validation			
	Bootstraping [L]	Bootstrapping is a powerful statistical method for estimating the sampling distribution of an estimator by resampling with replacement from the original sample. It's used for hypothesis testing and to construct confidence intervals for generalizing results from a sample to the population. In machine learning, it's used in ensemble methods like Bagging and Random Forests to generate diverse models by creating different datasets.			
	Standardization	This is a scaling technique where the values are centered around the mean with a unit standard deviation. It's used to bring data to a common scale without distorting differences in the ranges of values or losing information. Many machine learning algorithms perform better with standardized input features.			
	Normalization	Similar to standardization, normalization is a scaling technique that modifies the values of numeric columns in the dataset to a common scale, but without distorting differences in the ranges of values or losing information. It's also known as Min-Max scaling.			
	Statistical Moments [L]	The statistical moments (mean, variance, skewness, and kurtosis) capture different aspects of the distribution shape. They are used in machine learning to describe, understand, and compare variable distributions. In particular, skewness and kurtosis can be used in feature engineering to create new features or to select features.			
	Bayesian Statistics	Hyperparameter Tuning			