**Capstone Project: Applied Statistics Interview Grind**

**Contribution – Individual**

**Name – Monti Kumar**

**1. What is a vector in mathematics?**

**Answer: -** A vector is a fundamental mathematical object that represents both magnitude(length) and direction. Vectors are used to describe quantities that have both size and orientation in space. They are often depicted as arrows, with the length of the arrow representing the magnitude of the vector, and the direction of the arrow indicating its orientation.

One of the best examples of a vector is velocity. Velocity is a vector because it has both magnitude and direction.

**2. How is a vector different from a scalar?**

**Answer: -** The difference between vectors and scalars is that vectors have both magnitude and direction, while scalars have magnitude only. Vectors are often used to represent physical quantities in which direction matters, such as motion and forces, whereas scalars represent quantities that are fully described by their size alone.

| **Vector** | **scalar** |
| --- | --- |
| It has both magnitude and direction. | It has only the magnitude. |
| It is multidimensional | Only one dimensional |
| This changes with magnitude and direction | This quantity changes with the change in magnitude |
| One vector cannot divide another vector | One scalar quantity can divide another scalar |
| Examples: Displacement, Acceleration, Velocity, etc | Examples: Distance, mass, time, etc |

**3. What are the different operations that can be performed on vectors?**

**Answer: -** Vectors can undergo various fundamental operations in mathematics, including:

1) Vector Addition: Adding two or more vectors to get a resultant vector.

2) Scalar Multiplication: Multiplying a vector by a scaler.

3) Dot Product: A binary operation that takes two vectors and returns a scaler.

4) Cross Product: The magnitude of the resulting vector is the product of the magnitudes of the input vectors and the sine of the angle between them.

5) Vector Subtraction: Subtracting one vector from another vector.

6) Unit Vector: Creating a unit vector from a given vector by dividing the vector by its magnitude.

**4. How can vectors be multiplied by a scalar?**

**Answer: -** Though vectors and scalars represent different varieties of physical quantities, at times both of them must interact. The addition of a scalar to a vector quantity is highly impossible because of their differences in dimensions. However, a vector quantity can be multiplied by a scalar. At the same time, the converse of this is not possible. i.e. A scalar can never be multiplied by a vector.

When a vector multiply by a vector

Ø If the scalar is greater than 1, the vector gets longer.

Ø If the scalar is between 0 and 1, the vector gets shorter.

Ø If the scalar is negative, the direction of the vector flips, but it still gets longer or shorter based on the magnitude of the scalar.

**5. What is the magnitude of a vector?**

**Answer:** -The vector is an object which has magnitude as well as direction. To find the magnitude of a vector. For any vector A its magnitude is represented as |A|. Suppose a vector is defined as xi + yj then its magnitude is defined as the square root of the sum of squares of the individual terms. The magnitude of the vector represents the length of the vector i.e. the value or impact the vector has.

**6. How can the direction of a vector be determined?**

**Answer: -** The direction of a vector is the angle made by the vector with the horizontal axis, which is often represented by the positive direction of the X-axis. The direction of a vector is typically measured counterclockwise from the positive X-axis. Thus, if we consider east as the positive X-axis, the direction of a vector would be the angle of rotation counterclockwise from the eastward direction.

**7. What is the difference between a square matrix and a rectangular matrix?**

**Answer: -** A square matrix is a matrix that contains the same number of rows and the same number of columns. Then, we should know the order of a matrix. An order of matrices refers to the number of rows and columns it has. Based on the orders, matrices can be If a matrix is not a square matrix, then it is known as a rectangular matrix. We can also say that the matrices which have different numbers of rows and columns are called rectangular matrices.

For example,

A 2x2 matrix (2 rows, 2 columns) is a square matrix.,

A 3x3 matrix (3 rows, 3 columns) is a square matrix.

A 4x4 matrix (4 rows, 4 columns) is a square matrix. is a square matrix. As we see, it has the same number of rows and columns.

But if we look at Examples:

A 2x3 matrix (2 rows, 3 columns) is a rectangular matrix.

A 3x2 matrix (3 rows, 2 columns) is a rectangular matrix.

A 4x5 matrix (4 rows, 5 columns) is a rectangular matrix, then we can say that it is a rectangular matrix. Here, the number of rows and columns is different. The key difference between square and rectangular matrices is that square matrices have an equal number of rows and columns (i.e. n×n), while rectangular matrices have a different number of rows and columns (i.e. m×n where m does not equal to n).

Square matrices are often associated with specialized properties and applications in linear algebra, while rectangular matrices are more general and versatile, often used to represent various types of data.

**8. What is a basis in linear algebra?**

**Answer: -** In linear algebra, a basis vector refers to a vector that forms part of a basis for a vector space. A basis is a set of linearly independent vectors that can be used to represent any vector within that vector space. Basis vectors play a fundamental role in describing and analysing vectors and vector spaces. The basis of a vector space provides a coordinate system that allows us to represent vectors using numerical coordinates.

**9. What is a linear transformation in linear algebra?**

**Answer: -** A linear transformation is a mathematical operation that maps one geometric figure, matrix, or vector to another, using a formula with a specific structure. This formula must involve a linear combination, where the original components (such as the x and y coordinates of each point in the original figure) are manipulated through expressions like ax + by to generate the coordinates of the transformed figure. Linear transformations encompass a variety of changes to geometric objects, matrices, or vectors, including:

scaling, rotation, reflection, shearing, transformation.

**10. What is an eigenvector in linear algebra?**

**Answer: -** eigenvector is a vector that is associated with a set of linear equations. They are also known as characteristic vectors, latent vectors, or proper vectors.

Eigenvectors are non-zero vectors that do not change direction when a linear transformation is applied. They are only stretched by a constant factor, with no rotation or shear. The corresponding eigenvalue is the factor by which an eigenvector is stretched or squished. If the eigenvalue is negative, the eigenvector's direction is reversed.

**11. What is the gradient in machine learning?**

**Answer: -** Gradient is a commonly used term in optimization and machine learning. For example, deep learning neural networks are fit using stochastic gradient descent, and many standard optimization algorithms used to fit machine learning algorithms use gradient information. In order to understand what a gradient is, you need to understand what a derivative is from the field of calculus. This includes how to calculate a derivative and interpret the value. An understanding of the derivative is directly applicable to understanding how to calculate and interpret gradients as used in optimization and machine learning.

**12. What is backpropagation in machine learning?**

**Answer: - Backpropagation** is the essence of neural network training. It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization.

Backpropagation in neural network is a short form for “backward propagation of errors.” It is a standard method of training artificial neural networks. This method helps calculate the gradient of a loss function with respect to all the weights in the network.

**13. What is the concept of a derivative in calculus?**

**Answer: -** In calculus, the derivative is a fundamental concept that measures the rate at which a function's output (dependent variable) changes with respect to a change in its input (independent variable). It essentially tells you how the function's value is changing as you move along its graph. The derivative provides important information about the slope or steepness of a function at a given point. The derivative in calculus is a mathematical tool that provides information about how a function changes as its input varies. It is a fundamental concept with broad applications in mathematics and various fields of science and engineering.

**14. How are partial derivatives used in machine learning?**

**Answer: -** Partial derivatives and gradient vectors are used very often in machine learning algorithms for finding the minimum or maximum of a function. Gradient vectors are used in the training of neural networks, logistic regression, and many other classification and regression problems.

**15. What is probability theory?**

**Answer: -** Probability means possibility. Probability theory, is a branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event cannot be determined before it occurs, but it may be any one of several possible outcomes. Probability can be defined as the number of favourable outcomes divided by the total number of possible outcomes of an event. Probability theory makes use of some fundamentals such as sample space, probability distributions, random variables, etc.

And the formula is,

Probability of events P(E) = Number of favourable outcomes

Total number of possible outcomes

**16. What are the primary components of probability theory?**

**Answer: -** The primary components of probability theory are:

Ø Set theory: This is the study of sets, which are collections of objects.

Ø Probability space: This is a mathematical structure that describes the possible outcomes of an experiment and their associated probabilities.

Ø Random variable: This is a variable whose value depends on the outcome of an

experiment.

Ø Distribution: This is a function that describes the probability of a random variable taking

on different values.

Ø Expectation: This is the average value of a random variable.

Ø Variance: This is a measure of how spread out the values of a random variable are.

Ø Covariance: This is a measure of how two random variables are related to each other.

Ø Correlation coefficient: This is a standardized measure of covariance.

**17. What is conditional probability, and how is it calculated?  
 Answer: -** Conditional probability is the probability of event A happening given that event B has already happened. It is calculated as follows:

P (A|B) = P (A and B) / P (B)

Where, P(A|B) is the conditional probability of event A happening given that event B has already happened, P (A and B) is the probability of event A and event B both happening, and P(B) is the probability of event B happening.

**18. What is Bayes theorem, and how is it used?  
 Answer: -** Bayes theorem is a formula that can be used to update our beliefs about the probability of an event happening based on new information. It is calculated as follows:

P (A|B) = P (B|A) P (A) / P (B)

Where P(A|B) is the posterior probability of event A happening given that event B has

already happened, P(B|A) is the likelihood of event B happening given that event A has

already happened, P(A) is the prior probability of event A happening, and P(B) is the

probability of event B happening.

Bayes theorem is used in a variety of applications, such as medical diagnosis, weather

forecasting, and finance.

**19. What is a random variable, and how is it different from a regular variable?  
 Answer: -** A random variable is a variable whose value depends on the outcome of an experiment. A regular variable is a variable whose value does not depend on the outcome of an experiment.

For example, the number of heads you get when you flip a coin is a random variable. The number of students in a class is a regular variable.

**20. What is the law of large numbers, and how does it relate to probability theory?**

**Answer: -** The law of large numbers states that as the number of trials in an experiment increase, the average of the results tends to get closer to the expected value of the random variable. This concept is fundamental in probability theory and statistics, providing a basis for making inferences about populations based on sample data. There are two main versions: the weak and strong laws of large numbers.

**21. What is the central limit theorem, and how is it used?  
 Answer: -** The central limit theorem**,** which is a statistical theory, states that when a large sample size has a finite variance, the samples will be normally distributed, and the mean of samples will be approximately equal to the mean of the whole population. It is used in a variety of statistical applications, such as hypothesis testing and confidence interval estimation.

**22. What is the difference between discrete and continuous probability distributions?  
 Answer: -**

| **Aspect** | **Discrete Probability Distribution** | **Continuous Probability Distribution** |
| --- | --- | --- |
| **Definition** | Deals with countable outcomes | Deals with uncountable outcomes |
| **Sample Space** | Consists of distinct, separate values | Forms a continuous range of values |
| **Probability Function** | Often represented byprobability mass function (PMF) | Often represented by a probability density function (PDF) |
| **Sum of Probabilities** | Sum of probabilities over all possible outcomes equals 1 | Integral of the probability density function over the entire range equals 1 |
| **Examples** | Coin toss, dice roll, number of students in a class | Height of individuals, temperature, time taken to complete a task |

**23. What are some common measures of central tendency, and how are they calculated?**

**Answer: -** The mean, median, and mode are some common measures of central tendency.

Ø The mean is the average of all the values in a set of data. It is calculated by adding all the values and dividing by the number of values.

Ø The median is the middle value in a set of data when the data is arranged in order from least to greatest.

Ø The mode of a dataset is the value or values that appear with the highest frequency

**24. What is the purpose of using percentiles and quartiles in data summarization?  
 Answer: -**

Percentiles and quartiles are used to summarize data by dividing it into equal parts.

Percentiles divide the data into 100 equal parts, while quartiles divide the data into 4 equals

parts. The first quartile is the 25th percentile, the second quartile is the 50th percentile (also

known as the median), and the third quartile is the 75th percentile.

Percentiles and quartiles are useful for identifying outliers in a dataset. An outlier is a data

point that is much larger or smaller than the other data points. Outliers can be caused by

errors in data collection or measurement, or they can be genuine data points.

**25. How do you detect and treat outliers in a dataset?**

**Answer: -** There are a few ways to detect outliers in a dataset. One way is to use the interquartile

range (IQR). The IQR is the difference between the third quartile and the first quartile.

Outliers are any data points that are more than 1.5 times the IQR away from the first or

third quartile.

Another way to detect outliers is to use boxplots. A boxplot is a graphical representation of

a dataset that shows the distribution of the data, including the median, quartiles, and

outliers.

Once outliers have been detected, there are a few ways to treat them. One way is to simply

remove them from the dataset. However, this can be problematic if the outliers are genuine

data points. Another way to treat outliers is to transform the data. This can be done by

using a logarithmic transformation, a square root transformation, or another type of

transformation.

**26. How do you use the central limit theorem to approximate a discrete probability distribution?**

**Answer: -** The central limit theorem states that the sampling distribution of the mean of a large sample will approach a normal distribution, regardless of the shape of the population distribution. This means that we can use the normal distribution to approximate the probability distribution of the mean, even if the population distribution is not normal. To use the central limit theorem, we need to know the sample size. The larger the sample size, the more closely the sampling distribution of the mean will approximate a normal distribution.

**27. How do you test the goodness of fit of a discrete probability distribution?**

**Answer: -** Test the goodness of fit of a discrete probability distribution, we can use two common methods:

**A.** **Chi-squared test:** This test compares the observed frequencies in your dataset with the expected frequencies according to the specified probability distribution. If there's a significant difference between them, it suggests that the data might not follow the expected distribution.

**B.** **Kolmogorov-Smirnov test:** This test compares the overall shape of the cumulative distribution function of your data with the expected cumulative distribution function from the specified probability distribution. If they're significantly different, it indicates that the data might not fit the expected distribution well.

**28. What is a joint probability distribution?**

**Answer: -** A joint probability distribution is a statistical function that describes the likelihood of multiple random variables occurring simultaneously. It assigns probabilities to each possible combination of values for those variables. In other words, it specifies the probability of every possible outcome for all the variables in consideration. This distribution is commonly represented in the form of a table, matrix, or equation.

**29. How do you calculate the joint probability distribution?**

**Answer: -** The joint probability distribution can be calculated by adding up the probabilities of all the possible combinations of the events. For example, if there are two events, A and B, and each event can occur in two ways, then the joint probability distribution would have four entries. Each entry would show the probability of event A occurring and event B occurring.

**30. What is the difference between a joint probability distribution and a marginal probability distribution?**

**Answer: -** A joint probability distribution deals with the probability of two or more events occurring simultaneously, providing probabilities for all possible combinations of the variables involved.

On the other hand, a marginal probability distribution focuses on the probability of a single event occurring, regardless of the outcomes of other variables. It is obtained by summing or integrating over all possible outcomes of the other variables in the joint distribution to obtain the probabilities associated with the single variable of interest.

**31. What is the covariance of a joint probability distribution?**

**Answer: -** The covariance of a joint probability distribution of two random variables X and Y is a measure of how much they tend to vary together. It is calculated as the expected value of the product of the deviations of X and Y from their respective means.

**32. How do you determine if two random variables are independent based on their joint probability distribution?**

**Answer: -** Two random variables are independent if their joint probability distribution can be factorized into the product of the marginal probability distributions of X and Y. In other words, the probability of X taking on a particular value and Y taking on a particular value is the same as the probability of X taking on that value multiplied by the probability of Y taking on that value.

**33. What is sampling in statistics, and why is it important?**

**Answer: -** A **sample** is a subset of individuals from a larger [population](https://www.scribbr.com/methodology/population-vs-sample/). [**Sampling**](https://www.scribbr.com/methodology/sampling-methods/) means selecting the group that you will actually collect data from in your research. Sampling is important because it can provide insights when it's not feasible to measure the entire population. Sampling is used in a variety of statistical methods, such as hypothesis testing, confidence interval estimation, and regression analysis.

**34. What is the central limit theorem, and why is it important in statistical inference?**

**Answer: -**  The central limit theorem (CLT) states that the [distribution of a sample](https://www.investopedia.com/terms/s/sampling-distribution.asp) variable approximates a normal distribution (i.e., a “bell curve”) as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population's actual distribution shape. Central limit theorem is useful when analysing large data sets because it allows one to assume that the sampling distribution of the mean will be normally-distributed in most cases. This allows for easier statistical analysis and inference.

**35. What is the difference between parameter estimation and hypothesis testing?**

**Answer: -** Parameter estimation is the process of estimating the values of population parameters, such as the mean and standard deviation. Hypothesis testing is the process of testing hypotheses about population parameters.

**36. What is the p-value in hypothesis testing?**

**Answer: -** A p-value in hypothesis testing is the probability of observing a test statistic as extreme as, or more extreme than, the one actually observed, assuming that the null hypothesis is true. In simpler terms, it tells you the likelihood of obtaining the observed results if the null hypothesis is correct. A smaller p-value indicates stronger evidence against the null hypothesis, suggesting that you should reject it in favor of the alternative hypothesis.

**37. What is confidence interval estimation?**

**Answer: -** A confidence interval is the mean of your estimate plus and minus the variation in that estimate. This is the range of values you expect your estimate to fall between if you redo your test, within a certain level of confidence.

**38. What are Type I and Type II errors in hypothesis testing?**

**Answer: -** A Type I error is rejecting the null hypothesis when it is actually true. A Type II error is failing to reject the null hypothesis when it is actually false.

**Type I Error (False Positive)**

Ø A Type I error occurs when a null hypothesis that is actually true is incorrectly rejected. it's a false alarm or a "false positive" result.

Ø The probability of making a Type I error is denoted as α (alpha) and is known as the significance level of the test.

Ø A lower significance level (α) reduces the probability of a Type I error but may increase the probability of a Type II error.

**Type II Error (False Negative)**

Ø A Type II error occurs when a null hypothesis that is actually false is not rejected. It's failing to detect an effect or a "false negative" result.

Ø The probability of making a Type II error is denoted as β (beta).

Ø The power of a statistical test is equal to (1−β) and represents the probability of correctly rejecting a false null hypothesis.

Ø A lower Type II error rate (higher power) is desirable, but it may require larger sample sizes or more sensitive tests.

**39. What is the difference between correlation and causation?  
 Answer: -** Correlation is a statistical measure (expressed as a number) that describes the size and direction of a relationship between two or more variables. A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable.

Causation indicates that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events. This is also referred to as cause and effect.

**40. How is a confidence interval defined in statistics?  
 Answer: -** A confidence interval is a range of values that is likely to contain the true population parameter. It is typically calculated as the mean plus or minus a margin of error. The confidence level is the probability that the confidence interval will contain the true population parameter. For example, a 95% confidence interval means that there is a 95% probability that the true population parameter is within the confidence interval.

**41. What does the confidence level represent in a confidence interval?  
 Answer: -** The confidence level is the percentage of times you expect to reproduce an estimate between the upper and lower bounds of the confidence interval, and is set by the alpha value. For example, a 95% confidence interval means that there is a 95% probability that the true population parameter is within the confidence interval.

**42. What is hypothesis testing in statistics?  
 Answer: -** Hypothesis testing can be defined as a statistical tool that is used to identify if the results of an experiment are meaningful or not. It involves setting up a null hypothesis and an alternative hypothesis. These two hypotheses will always be mutually exclusive. This means that if the null hypothesis is true then the alternative hypothesis is false and vice versa. An example of hypothesis testing is setting up a test to check if a new medicine works on a disease in a more efficient manner.

**43. What is the purpose of a null hypothesis in hypothesis testing?  
 Answer: -** The null hypothesis is a concise mathematical statement that is used to indicate that there is no difference between two possibilities. In other words, there is no difference between certain characteristics of data. This hypothesis assumes that the outcomes of an experiment are based on chance alone. It is denoted as H0. Hypothesis testing is used to conclude if the null hypothesis can be rejected or not. Suppose an experiment is conducted to check if girls are shorter than boys at the age of 5. The null hypothesis will say that they are the same height.

**44. What is the difference between a one-tailed and a two-tailed test?  
 Answer: -** A one-tailed test is a hypothesis test in which the alternative hypothesis specifies that the population parameter is either greater than or less than a certain value.

A two-tailed test is a hypothesis test in which the alternative hypothesis specifies that the population parameter is different from a certain value.

**45. What is experiment design, and why is it important?  
 Answer: -** Experiment design is the process of planning and conducting an experiment. It is important because it ensures that the experiment is conducted in a way that minimizes bias and maximizes the chances of obtaining accurate results.

**46. What is the geometric interpretation of the dot product?  
 Answer: -** The dot product has the following geometric interpretation: Let α be the angle between A and B. Then A · B = |A|·| B| · cos(α). A number of important properties of the dot product should be noted. Most of these are obvious consequences, either of the definition or of the above geometric formula or both.

**47. What is the geometric interpretation of the cross-product?  
 Answer: -** Cross product is a binary operation on two vectors in three-dimensional space. It results in a vector that is perpendicular to both vectors. The [Vector product](https://byjus.com/maths/calculate-vector-product/) of two vectors, **A** and **B**, is denoted by **A** × **B**. Its resultant vector is perpendicular to **A** and **B**. Vector products are also called cross products. Cross product of two vectors will give the resultant a vector and calculated using the Right-hand Rule.

**48. What is kurtosis?**

**Answer: -** Kurtosis is a statistical measure that describes the shape of a distribution's tails in relation to its overall shape. Specifically, kurtosis quantifies whether the tails of a given distribution contain extreme values ([outliers](https://deepai.org/machine-learning-glossary-and-terms/outlier)) that are either more or less common than those of a [normal distribution](https://deepai.org/machine-learning-glossary-and-terms/normal-distribution). It is a crucial tool for understanding the [probability](https://deepai.org/machine-learning-glossary-and-terms/probability) and likelihood of extreme deviations in datasets.

**49. What is the difference between Descriptive and Inferential Statistics?  
 Answer: -** Descriptive statistics is all about summarizing and organizing data to make it easy to understand. It gives us a snapshot of the data, allowing us to see patterns, trends, and basic characteristics. Inferential statistics goes beyond just describing the data. It’s about making inferences or predictions about a larger population based on a smaller sample of data.

**50. What is the meaning of degrees of freedom (DOF) in statistics?  
 Answer: -** Degrees of freedom (DOF) are the maximum number of logically independent values, which may vary in a data sample. Degrees of freedom are calculated by subtracting one from the number of items within the data sample.