MUTHAYAMMAL ENGINEERING COLLEGE

(An Autonomous Institution)

(Approved by AICTE, New Delhi, Accredited by NAAC & Affiliated to Anna University)

Rasipuram - 637 408, Namakkal Dist., Tamil Nadu.

Estd. 2000

IQAC

MUST KNOW CONCEPTS

MKC

MBA

2020-21

Course Code & Course Name : 19MBA02 & Business Statistics

Year/Sem/Sec : I/I/-

S.No.	Term	Notation (Symbol)	Concept / Definition / Meaning / Units / Equation / Expression	Units
Unit-I : Introduction				
1.	Statistic		A characteristic of a sample	I
2.	Statistics		The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions	I
3.	Mean		The sum of the scores in a distribution divided by the number of scores in the distribution	I
4.	Median		The midpoint of the values after they have been ordered from the smallest to the largest, or the largest to the smallest	I
5.	Mode		The most commonly occurring value in a distribution.	I
6.	Range		Largest value - smallest value	I
7.	Mean Deviation		the arithmetic mean of the absolute values of the deviations from the arithmetic mean	I
8.	Variance		Variance is the statistical average of the dispersion of scores in a distribution	I
9.	Standard Deviation		the square root of the variance	I
10.	Sampling methods		Non probability and probability sampling	I
11.	Population		the entire set of individuals or objects of interest or the measurements obtained from all individuals or	I

			objects of interest	
12.	Sample		a portion, or part, of the population of interest	I
13.	Probability		a value between zero and one, inclusive, describing the relative possibility (chance or likelihood) an event will occur	I
14.	Simple Random Sample		a sample selected so that each item or person in the population has the same chance of being included	I
15.	Sampling Error		the difference between a sample statistic and its corresponding population parameter	I
16.	Null Hypothesis		a statement about the value of a population parameter developed for the purpose of testing numerical evidence	I
17.	Alternate Hypothesis		a statement that is accepted if the sample data provide sufficient evidence that the null hypothesis is false	I
18.	Type I Error		rejecting the null hypothesis when it is true	I
19.	Type II Error		accepting the null hypothesis when it is false	I
20.	Normal distribution		a frequency distribution that fits a bell curve, typical of parametric data	I
21.	Sample size		refers to the number of test specimens	I
22.	Hypothesis		A tentative statement about the value of one or more population parameters	I
23.	Level of Significance		the probability of rejecting the null hypothesis when it is true	I
24.	Critical Value		the dividing point between the region where the null hypothesis is rejected and the region where it is not rejected.	I
25.	Significance Test		Measures the strength of evidence against the null hypothesis	I

Unit-II : Sampling Distribution & Estimation				
26.	Degrees of freedom:		A technical term reflecting the number of independent elements comprising a statistical measure.	II
27.	Frequency Distribution		a grouping of data into mutually exclusive classes showing the number of observations in each class.	II
28.	F-test (F		: A parametric statistical test of the equality of the means of two or more samples	II
29.	t-test		A data analysis procedure that assesses whether the means of two related groups are statistically different from each other	II
30.	Z Value		the signed distance between a selected value, designated X, and the mean, divided by the standard deviation	II
31.	Correlation Analysis		a group of techniques to measure the association between two variables.	II
32.	Coefficient of Correlation		a measure of the strength of the linear relationship between two variables.	II
33.	Regression Equation		an equation that expresses the linear relationship between two variables	II
34.	Spearman's rank order correlation		A nonparametric test used to measure the relationship between two rank ordered scales.	II
35.	Discrete Variable		limited number of values	II
36.	Positive correlation		one variable rises or falls, the other does as well	II
37.	Negative correlation		indicates that two variables move in opposite directions	II
38.	Chi-square (^{2}):		A nonparametric test of statistical significance appropriate when the data are in the form of frequency counts.	II
39.	Analysis of variance (ANOVA)		A statistical technique for determining the statistical significance of differences among means	II
40.	Multivariate analysis of covariance		An extension of ANOVA that incorporates two or more	II

			dependent variables in the same analysis.	
41.	Trend analysis		Trend analysis (or trend-line analysis) is a special form of simple regression in which time is the explanatory variable.	II
42.	Time series		An ordered sequence of values of a variable observed at equally spaced time intervals is referred to as a time series	II
43.	Seasonal index		A seasonal index is a number that indicates the seasonality for a given time period.	II
44.	Forecasting		Forecasting is the prediction of values of a variable based on known past values of that variable or other related variables	II
45.	Index numbers		These numbers are frequently used as summary indicators of the level of economic activity and/or corporate performance. F	II
46.	Lead time		This term refers to the time interval between two events	II
47.	Moving average		Time series we can define the moving average of order K as the average (mean) value of K consecutive observations.	II
48.	Seasonal variation		The change that seasonal factors cause in a data series is frequently called seasonal variation	II
49.	Σ		Sigma summation	II
50.	H 0		Null hypothesis	II
Unit-III : Testing of Hypothesis				
51.	Additive effect		Statistical Glossary Additive effect: An additive effect refers to the role of a variable in an estimated model	III
52.	Aggregate Mean		Aggregate Mean: In ANOVA and some other techniques used for analysis of several samples	III
53.	Analysis of Commonality		Analysis of Commonality: Analysis of commonality is a method for causal	III

			modeling .	
54.	ARIMA		ARIMA as an acronym for Autoregressive Integrated Moving Average Model	III
55.	Boosting		In predictive modeling, boosting is an iterative ensemble method	III
56.	Canonical Discriminate Analysis		Canonical Discriminate Analysis: See Multiple discriminate analysis. Browse Other Glossary Entries	III
57.	Central Limit Theorem		In a census survey, all units from the population of interest are analyzed.	III
58.	Chi-Square Statistic		The chi-square statistic (or -statistic) measures agreement between the observed and hypothetical frequencies.	III
59.	Co linearity		In regression analysis, co linearity of two variables means that strong correlation	III
60.	Estimator		A statistic, measure, or model, applied to a sample, intended to estimate some parameter of the population that the sample came from.	III
61.	General Linear Model		In contrast to linear models, allow you to describe both additive and non-additive relationship between a dependent variable and N independent variables.	III
62.	Beta Distribution		Beta Distribution: Suppose x1, x2, xn are n independent values of a random variable uniformly distributed within the interval [0,1].	III
63.	Causal analysis		See causal modeling. Browse Other Glossary Entries.	III
64.	Cochran´s Q Statistic		Cochran´s Q statistic is computed from replicated measurements data with binary responses.	III
65.	Coefficient of variation		The coefficient of variation is the standard deviation of a data set, divided by the mean of the same data set. Browse Other Glossary Entries	III
66.	Linear Filter		A linear model specifies a linear relationship between a dependent	III

			variable and n independent variables.	
67.	Logistic Regression		Logistic regression is used with binary data when you want to model the probability that a specified outcome.	III
68.	Mean Values (Comparison)		Statistical Glossary Mean Values (Comparison): The numerical example below illustrates basic properties of various descriptive statistics with "mean"	III
69.	Network Analytics		Network analytics is the science of describing and, especially, visualizing the connections among objects.	III
70.	Time Series		Time series data are measurements of a variable taken at regular intervals over time.	III
71.	Time-series data		Time-series data: See longitudinal data Browse Other Glossary Entries	III
72.	Uniform Distribution		The uniform distribution describes probabilistic properties of a continuous random variable	III
73.	Validation Set		A validation set is a portion of a data set used in data mining to assess the performance	III
74.	Variance		Variance is a measure of dispersion. It is the average squared distance between the mean and each item in the population or in the sample.	III
75.	Census Survey		In a census survey, all units from the population of interest are analyzed. A related concept is the sample survey	III
Unit-IV : Correlation & Regression				
76.	Chi-Square Statistic		The chi-square statistic (or -statistic) measures agreement between the observed and hypothetical frequencies.	IV
77.	Fixed Effects		The term "fixed effects" (as contrasted with "random effects") is related to how particular coefficients in a model	IV
78.	Correlation matrix		Most computer programs designed to perform multiple regression analysis	IV

			include the computation of the correlation coefficients between each pair of variables.	
79.	Descriptive statistics		Descriptive statistics are used to describe the main features of a collection of data in quantitative terms.	IV
80.	Dummy variable		In regression analysis, a dummy variable is one that takes the values 0 or 1 to indicate the absence	IV
81.	Hypothesis testing		The procedure for deciding if a null hypothesis should be accepted or rejected in favor of an alternate hypothesis.	IV
82.	Mean absolute percentage error		The mean absolute percentage error is the mean or average of the sum	IV
83.	Nonlinear estimation		If parameters have to be estimated for nonlinear functions, then ordinary least squares estimation may not apply.	IV
84.	Null hypothesis		The null hypothesis is a hypothesis that the researcher tries to disprove, reject, or nullify.	IV
85.	Polynomial fitting		It is possible to fit a polynomial of any number of terms to a set of data.	IV
86.	Regression coefficients I		In regression, a forecast variable Y is modeled as a function of explanatory variables X1 through Xk.	IV
87.	Standard deviation		The standard deviation is the average deviation, or degree of dispersion, about the mean of any data set.	IV
88.	Time-series analysis		Time series analysis is a statistical technique that deals with time series data, or trend analysis.	IV
89.	Beginning of observation		In survival analysis, the moment in time when subjects begin to be followed by the researcher.	IV
90.	Center of a distribution		The typical or average value in a variable’s distribution.	IV
91.	Variance of a variable		The average of the squared deviation scores.	IV
92.	Sensitivity analysis		An alternative analysis using a	IV

			different model	
93.	Simple random sample		A sample in which every member of the population has the same chance of being selected into the sample.	IV
94.	Research hypothesis		The hypothesis that the researcher is trying to marshal evidence for; this is usually the hypothesis	IV
95.	Sensitivity of classification		In logistic regression, the probability of a case being classified as a case by the prediction equation.	IV
96.	Survival analysis		The analysis of time-to- event data, i.e., the length of time until an event occurs to subjects.	IV
97.	Test statistic		A sample statistic measuring the discrepancy between what is observed in the sample	IV
98.	Two-tailed test		A test of hypothesis for which the research hypothesis is not directional	IV
99.	Type I error		The probability of rejecting a true null hypothesis in a statistical test	IV
100.	Type II error		The probability of failing to reject a false null hypothesis in a statistical test.	IV
Unit-V : Time Series & Forecasting				
101.	Φ2		A measure of the strength of association for two qualitative variables that are each binary variables.	V
102.	Survival function		The probability of surviving to a particular point in time without experiencing the event of interest	V
103.	Symmetric		Said of distributions that exhibit no skewness, and for which exactly 50 % of cases lie above and below the mean of the distribution.	V
104.	Within-subjects variable		A variable in repeated-measures ANOVA or linear mixed modeling that takes on different values over time for the same subject.	V
105.	Reverse causation		The situation in which the study endpoint in a regression model	V

106.	Specifi city of classification		In logistic regression, the probability of a control being classified as a control by the prediction equation.	V
107.	Strength of association		The degree to which knowledge of one's status on one variable enables prediction of one's status on another variable that it is associated with.	V
108.	Symmetric		Said of distributions that exhibit and for which exactly 50 % of cases lie above and below the mean of the distribution.	V
109.	Right skewed		Said of distributions where most cases have low values of the variable, and a few outliers have very high values.	V
110.	Scatterplot		A graphical display of the association between two quantitative variables achieved by plotting points representing the intersection of each variable's values.	V
111.	Chi-square curve		The chi-square curve is a family of curves that depend on a parameter called degrees of freedom	V
112.	Cluster Sample		In a cluster sample, the **SAMPLING UNIT** is a collection of population units, not single population units.	V
113.	Consequent		In a **CONDITIONAL p → q**, the consequent is **q**	V
114.	Asymptotic Property		An asymptotic property is a property of an estimator that holds as the sample size approaches infinity. Browse Other Glossary Entries	V
115.	Biased estimator		A sample statistic that is an inaccurate estimator of the corresponding population parameter	V
116.	Hazard function		Approximately the instantaneous probability of experiencing the event of interest at any given time	V
117.	Linear regression		A type of analysis in which a quantitative study endpoint is posited to be determined by one or more explanatory variables	V

118.	Multiple-comparison procedure		A statistical procedure for comparing group means that avoids capitalization on chance.	V
119.	Non-probability sample		A sample that is not a probability sample, i.e., a handpicked sample, a convenience sample, a "snowball sample,etc.	V
120.	Non-parametric test		A statistical test that makes very few assumptions about population distributions.	V
121.	Parameter		A summary measure of some characteristic for the population, such as the population mean or proportion.	V
122.	Poisson regression		A type of regression analysis in which the study endpoint is a count of the number of occurrences	V
123.	R2		A measure of the strength of association between a quantitative study endpoint and one or more quantitative explanatory variables.	V
124.	Scatter plot		A graphical display of the association between two quantitative variables achieved by plotting points representing the intersection of each variable's values.	V
125.	Standard error		The standard deviation of the sampling distribution of a statistic.	V
Placement Questions				
126.	What are the various branches of statistics?		• Descriptive Statistics • Inferential Statistics	
127.	Enumerate various fields where statistics can be used?		• Providing comparison • Explaining action which has already occurred	
128.	Difference between Data Science and Statistics?		Data Science is a science fields of scientific methods, algorithms, and even the process for extracting the insights from the data. statistics with the collection, analysis, interpretation,	

			organization, and presentation of data.	
129.	What is Bayesian?		Bayesian rests on the data which is observed in reality and further considers the probability distribution on the hypothesis.	
130.	What is Frequentist?		Frequentists rest on the hypothesis of choice and further consider the probability distribution on the data, whether it is observed or not.	
131.	What is the Likelihood?		In terms of statistical significance testing, the p-value represents the probability of obtaining a test value, which is as extreme as the one which had been observed originally.	
132.	Explain P-value with the help of an example?		• Null hypothesis (HO): a fair coin • Observation 0: 14 heads out of 20 flips • P-value of observation 0 given HO= Prob (? 14 heads or ? 14 tails) = 0.115	
133.	What do you mean by sampling?		• Randomly or in a simple yet random method • Systematically or taking every kth member of the population	
134.	What do you mean by Mode?		The mode is defined as that element of the data sample, which appears most often in the collection. X= [1 5 5 6 3 2]	
135.	What do you mean by Median?		Median is often described as that numerical value that separates the higher half of the sample, which can be either a group or a population or even a probability distribution from the lower half.	
136.	What do you mean by skewness?		Skewness is described as the data asymmetry, which is centered around a mean. If skewness is negative, the data is spread more on the left of the mean to the right. If skewness is seen as positive, then the data is moving more	

			to the right.	
137.	What is the meaning of Covariance?		Covariance is a measure of how two variables move in sync with each other. y 2= [1 3 4 5 6 7 8]	
138.	What is One Sample test?		T-test refers to any statistical hypothesis test in which the statistic of the test follows a Student's t distribution if the null hypothesis is supported.	
139.	What do you mean by Alternative Hypothesis?		The Alternative-hypothesis, which is represented by H1 is the statement that holds true if the null hypothesis is false.	
140.	What do you mean by Significance Level?		The probability of rejection of the null hypothesis when it is known as the significance level a, and very common choices are ?=0.05 and ?=0.01.	
141.	What is the meaning of normal distribution?		• Unimodal or one-mode. • Both the left and right halves are symmetrical and are mirror images of each other. • It is bell-shaped with a maximum height at the center.	
142.	What is Standard Deviation?		Standard Deviation is a measure of how much your data is spread out in statistics.	
143.	What is a Sample in Statistics and list the Sampling Methods?		➢ Cluster Sampling ➢ Simple Random ➢ Stratified ➢ Systematical	
144.	What is the difference between type I vs type II error?		"A type I error occurs when the null hypothesis is true, but is rejected. A type II error occurs when the null hypothesis is false, but erroneously fails to be rejected."	
145.	What is a statistical interaction?		"Basically, an interaction is when the effect of one factor (input variable) on the dependent variable (output	

			variable) differs among levels of another factor.”	
146.	What is an outlier?		Outliers are determined by using two methods: • Standard deviation/z-score • Interquartile range (IQR)	
147.	How is missing data handled in statistics?		• Prediction of the missing values • Assignment of individual (unique) values	
148.	What are the types of selection bias in statistics?		• Observer selection • Attrition • Protopathic bias • Time intervals	
149.	What is exploratory data analysis?		Exploratory data analysis is the process of performing investigations on data to understand the data better.	
150.	Where are long-tailed distributions used?		A long-tailed distribution is a type of distribution where the tail drops off gradually toward the end of the curve.	

Faculty Team Prepared

Signatures

1. P.Keerthi

HoD