

# BUS 322

## Quiz 1

Due by midnight on 2/21/21

### Study Guide and Class Notes

The following concepts will be on this quiz:

- 1. Statistical inference tools:** tools that allow a decision maker to reach a conclusion about a population of data based on a subset of data from population.
- 2. Experiment:** a process that produces a single outcome whose result cannot be predicted with certainty.
- 3. Sample:** a subset of the population.
- 4. Census:** an enumeration of the entire set of measurements taken from the whole population.
- 5. Non-statistical sampling techniques:** those methods of selecting samples using convenience, judgement, or other non-chance processes.
- 6. Quantitative data:** measurements whose values are inherently numerical.
- 7. Qualitative data:** data whose measurement scale is inherently categorical.
- 8. Parameter:** a measure computed from the entire population. As long as the population does not change, the value of the parameter will not change.
- 9. Statistic:** a measure computed from a sample that has been selected from a population. The value of the statistic will depend on which sample is selected.
- 10. Dependent variable:** a variable whose values are thought to be a function of, or dependent on, the values of another variable called the independent variable.
- 11. Independent variable:** a variable whose values are thought to impact the values of the dependent variable. The independent variable, or explanatory variable, is often within the direct control of the decision maker.

**12. Right-skewed data:** a data distribution when the mean of the data is larger than the median.

**13. Left-skewed data:** a data distribution where the mean for the data is smaller than the median.

**14. Population Variance:** is the average of the squared distances of the data values from the mean. It is a measure of variability that utilizes all data. The variance is based on the deviation about the mean.

**15. Standard deviation:** is the positive square root of the variance. It is a distance from the mean on the x axis.

**16. Standardization of data values (standard normal distribution):** a normal distribution that has a mean = 0 and a standard deviation = 1,

Standardized data values are referred to as z-scores.

**17. Coefficient of variation:** the ratio of the standard deviation to the mean expressed as a percentage. The coefficient of variation is used to measure the relative variation in data.

**18. Decision tree:** a diagram that illustrates the correct ordering of actions and events in a decision-analysis problem. Each act or event is represented by a branch on the decision tree.

**19. Axioms:** givens and assumptions that are necessary in analytical work. In logic, an indemonstrable first principle, rule, or maxim, that has found general acceptance or is thought worthy of common acceptance whether by virtue of a claim to intrinsic merit or on the basis of an appeal to self-evidence. An example would be: "Nothing can both be and not be at the same time and in the same respect", or the "80-20 rule."

**20. Consumer Research process (Burke video):** is never fully reliable. Prospective consumers are irrational in their behavior.

**21. Forensic research:** a retrospective, descriptive research based on historical data. Typically, it is applied in GAPS analysis (explaining differences between stated goals and the actual results).

**22. Predictive analysis** - decisions making process about the future when the future is uncertain. Techniques that use models constructed from past data to predict the future or to ascertain the impact of one variable on another.

**23. The concept of "self":** it is a product of all external influences on an individual and all internal characteristics of that individual.

**24. Mean:** a numerical measure of the center of a set of the quantitative measures computed by dividing the sum of the values by the number of values in the data.

**25. Weighted mean:** the mean value of data values that have been weighted according to their relative importance.

**26. Median:** is the center value that divides a data array into two halves.

**27. Mode:** is the value on a data set that occurs most frequently.

**28. Empirical Rule:** The empirical rule states that for a normal distribution, nearly all of the data will fall within three standard deviations of the mean. The empirical rule can be broken down into three parts:

- 68.28% of data falls within the first standard deviation from the mean.
- 95.44% fall within two standard deviations from the mean.
- 99.72% fall within three standard deviations from the mean.

**29. Bias:** is a disproportionate weight in favor of or against an idea or thing, usually in a way that is closed-minded, prejudicial, or unfair. Biases can be innate or learned. People may develop biases for or against an individual, a group, or a belief. In science and engineering, a bias is a systematic error. In statistics, it is the tendency of a predictive model to overestimate or underestimate the value of a continuous outcome.

**30. Business Statistics:** This is a form of mathematical analyses that are used to convert data into meaningful information. It uses quantified models, representations and synopses for a given set of experimental data or real-life studies. The discipline of Business Statistics studies methodologies to gather, review, analyze and draw conclusions from data.

**31. Factorials:** i.e. :  $0! = 1$

**32. Utility:** In economics and finance, risk aversion (one's utility) is the behavior of humans (especially consumers and investors), who, when exposed to uncertainty, attempt to lower that uncertainty.

It is the hesitation of a person to agree to a situation with an unknown payoff rather than another situation with a more predictable payoff but possibly lower expected payoff.

For example, a risk-averse investor might choose to put their money into a bank account with a low but guaranteed interest rate, rather than into a stock that may have high expected returns, but also involves a chance of

losing value. It is a measure of the total worth or relative desirability of a particular outcome.

**33. Normal distribution:** is a bell-shaped distribution with the following properties. It is:

a) unimodal: that is, the normal distribution peaks at a single value.

b) symmetrical: this means that the two areas under the curve between the mean and any two points equidistant on either side of the mean are identical.

c) is asymptotic to the x-axis (the normal curve approaches the x-axis as it moves toward the + or- infinity but never crosses them).

also,

d) one side of the distribution is the mirror image of the other.

e) the mean, median, and mode are equal.

f) the amount of variation in the random variable determines the width of the normal distribution.

**34. Probability:** the chance that a particular event will occur. The probability of an event will have a value in the range from 0 to 1.

**35. Random variable:** a variable that assigns a numerical value to each outcome of a random experiment or trial.

**36. Discrete random variable:** a random variable that can only assume a countable number of values.

**37. Continuous random variable:** random variables that can assume any value in an interval.

**38. Outliers:** observations with unusually large or unusually small values. Typically located above or below  $\pm$  two standard deviations from the mean. These extreme values are called outliers.

**39. Frequency distribution:** is a tabular summary of data showing the number (frequency) of data values in each of several categories.

**40. Variation:** captures differences in values of a variable of interest.

**41. Cumulative relative frequency distribution:** is a running total of the proportions of observations with a given array of data.

**42. Uncertainty:** no one can predict the future, yet business decisions about the future have to be made. Good decisions will create value and bad decisions will destroy it.

**43. Advanced analytics:** predictive and prescriptive analytics

**44. Big data:** any set of data that is so large or too complex to be handled by standard data processing techniques and typical desktop software.

**45. Business analytics:** the scientific process of transforming data into insight for making better decisions.

**46. Data dashboard:** a collection of table, charts, and maps to help management monitor selected aspects of company's performance.

**47. Data mining:** the use of analytical techniques for better understanding patterns and relationships that exist in large data sets.

**48. Data query:** a request for information with certain characteristics from a database.

**49. Data scientists:** analysts trained in both computer science and statistics who know how to effectively process and analyze massive amounts of data.

**50. Data security:** protecting stored data from destructive forces or unauthorized use.

**51. Decision analysis:** a technique used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events.

**52. Descriptive analytics:** analytical tools that describe what has happened.

**53. Hadoop:** an open-source programming environment that supports bigdata processing through distributed storage and distributed processing of data clusters.

**54. Internet of Things (IofT):** the technology that allows data collected from sensors in all types of machines to be sent over the Internet to repositories where it can be stored and analyzed.

**55. Map reduce programing:** programming model used with Hadoop that performs the two major steps for which it is named: the map step and the reduce step.

**56. The map step:** divides the data into manageable subsets and distributes it to the computers in the cluster for storing and processing.

**57. The reduce step:** collects answers from the nodes and combines them into an answer to the original problem.

**58. Operations decision:** a decision concerned with how the organization is run from day-to-day.

**59. Optimization model:** a mathematical model that gives the best decision, subject to decision's constraints.

**60. Prescriptive analytics:** techniques that analyze input data and yield a best course of action.

**61. Simulation:** the use of probability and statistics to construct a computer model to study the impact of uncertainty on the decision at hand.

**62. Simulation optimization:** the use of probability and statistics to model uncertainty, combined with optimization techniques, to find good decision in highly complex and highly uncertain settings.

**63. Strategic decision:** a decision that involves higher-level issues and that is concerned with the overall direction of the organization, defining the overall goals and aspirations for the organization's future.

**64. Tactical decision:** a decision concerned with how the organization should achieve the goals and objectives set by its strategy.

**65. Utility theory:** the study of the total worth or relative desirability of a particular outcome that reflects the decision maker's attitude toward a collection of factors such as profit, loss, and risk.

**66. Decision-making:** involves the following steps. 1. Identify and define the problem, 2. Determine the criteria that will be used to evaluate alternative solutions. 3. Determine the set of alternative solutions, 4. Evaluate the alternatives, 4. Choose best alternative.

**67. Four V's of big data:** volume, velocity, variety, veracity.

**68. Correlation coefficient:** a standardized measure of linear association between two variables that takes on values between -1 and +1. Values near -1 indicate a strong negative linear relationship, values near +1 indicate a strong positive linear relationship, and the values near zero indicate the lack of linear relationship.

**69. Covariance:** a measure of linear association between two variables. Positive values indicate positive relationship, negative values indicate the lack of a linear relationship.

**70. Cross-sectional data:** data collected at the same or approximately the same point in time at different locations.

**71. Time-series data:** data that are collected over a period of time (i.e. hours, minutes, years, etc.)

**72. Variable:** a characteristic or quantity of interest that can take of different values.

**73. z-score:** denotes the number of standard deviations from the mean.

**74. Observation:** a set of values corresponding to a set of variables.

**75. Population:** the set of all elements of interests in a particular study.

**76. Binomial probability distribution:** a probability distribution for a discrete random variable showing the probability of  $x$  “successes” in  $n$  trials.

**77. Exponential probability distribution:** a continuous probability distribution that is useful in computing probabilities for the time it takes to complete a task or the time between two points in time.

**78. Normal probability distribution:** a continuous probability distribution in which the probability density function is bell-shaped and determined by its mean and standard deviation.

**79. Poisson probability distribution:** a probability distribution for a discrete random variable showing the probability of  $x$  occurrences of an event over a specified interval of time or space.

**80. Uniform probability distribution:** a continuous probability distribution for which the probability that the random variable will assume a value in any interval is the same for each interval of equal length.