Statistics Vocab List 2022

NG

- 1. Box Plot a type of chart often used in explanatory data analysis
- 2. **Frequency** the number of times the observation occurred/recorded in an experiment or study
- 3. Histogram an approximate representation of the distribution of numerical data
- 4. **Interval** a data type which is measured along a scale, in which each point is placed at equal distance from one another
- IQR (Inter-Quartile Range) a measure of statistical dispersion, which is the spread of the data. Also known as the midspread, which is the difference between the 75th and 25th percentiles of data
- 6. **Mean** the average of a data set
- 7. **Median** the middle of the set of numbers
- 8. Mode the most common number in a data set
- 9. **Outlier** an observation that lies an abnormal distance from other values in a random sample from a population
- 10. **Quartile** term that describes a division of observations into four defined intervals based on the values of the data and how they compare to the entire set of observations
- 11. **Relative Frequency** describes the number of times a particular value for a variable has been observed to occur in relation to the total number of values for that variable
- 12. **Skewed** when the curve appears distorted to the left or right in a statistical distribution
- 13. **Standard Deviation** the average amount of variability in your dataset
- 14. **Variance** the expectation of the squared deviation of a random variable from its population mean or sample mean
- 15. **Z-Score** tells you where the score lies on a normal distribution curve

FΙ

- 1. **Bias** the difference between the estimators expected value and the true value of the parameter being estimated
- Blind the research participant is not told of the treatment assignment
- 3. Categorical collection of information that is divided into groups
- 4. **Compile** process of condensing information by classifying and tabulating vital statistical data into various categories or groups with the object of producing vital statistics according to a determined tabulation programme
- 5. **Constant** an identifier for a simple value. The value cannot be changed during the script
- 6. **Control Group** statistically significant portion of participants in an experiment that are shielded from exposure to variables
- 7. **Convenience Sample** a method of non-probability sampling where researchers will choose their sample based solely on the convenience
- 8. **Data** individual pieces of factual information recorded and used for the purpose of analysis
- 9. **Datum** a measurement of something on a scale understood by both the recorder (a person or device) and the reader (another person or device)
- 10. Double-Blind the research participant, investigator, study coordinator/nurse, study sponsor, and in some cases the data analyst are kept unaware of the treatment assignment
- 11. **Explicit Bias** the traditional conceptualization of bias. Individuals are aware of their prejudices and attitudes toward certain groups
- 12. **Implicit Bias** involve association outside conscious awareness that lead to a negative evaluation of a person on the basis of irrelevant characteristics such as race or gender
- 13. **NA** abbreviation in tables and lists for the phrases not applicable, not available, not assessed, or no answer
- 14. Numerical of or relating to numbers the numerical superiority of the enemy. Expressed in or involving numbers or a number system numerical standing in a Class A numerical code

- 15. **Observational Unit** the entity on which information is received and statistics are compiled in the process of collecting statistical data
- 16. **Parameter** any measured quantity of a statistical population that summarizes or describes an aspect of the population, such as a mean or standard deviation
- 17. **Placebo** appears to the participant to be an active treatment, but does not actually contain the active treatment
- 18. **Population** a set of similar items or events which is of interest for some question or experiment
- 19. **Qualitative (factor)** information that cannot be counted, measured or easily expressed using numbers
- 20. Quantitative (numeric) data expressing a certain quantity, amount or range
- 21. Random if each member in the population has an equal chance of being chosen
- 22. **Representative Sample** sample from a larger group that accurately represents the characteristics of a larger population
- 23. **Run** a sequence of similar or like events, items or symbols that is preceded by and followed by an event, item or symbol of a different type, or by none at all
- 24. Sample the specific group that you will collect data from
- 25. **Sampling Bias** a bias in which a sample is collected in such a way that some members of the intended population have a lower or higher sampling probability than others
- 26. **Simple Random Sample** a subset of individuals chosen from a larger set in which a subset of individuals are chosen randomly, all with the same probability
- 27. **Statistic** any quantity computed from values in a sample which is considered for a statistical purpose
- 28. **Stratified Sampling** a method of sampling from a population which can be partitioned into subpopulations
- 29. **Systematic Sample** a probability sampling method where researchers select members of the population at a regular interval

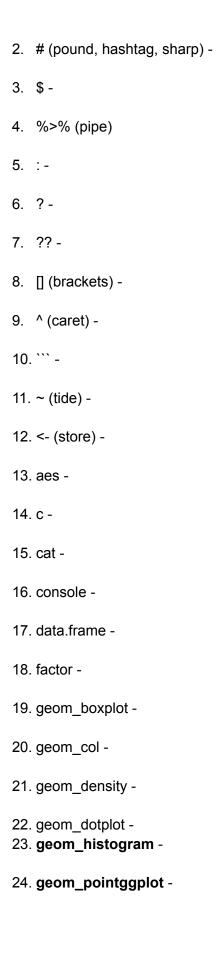
- 30. **Variability** describes how far apart data points lie from each other and from the center of a distribution
- 31. Variable any characteristics, number, or quantity that can be measured or counted
- 32. Whimsical subject to erratic behavior or unpredictable change

OE

- 1. **Causality** refers to a relationship between two events, or variables, which has a valid explanation
- 2. Causation indicates that one event is the result of the occurrence of the other event
- 3. Cause the identified variable that brings about a subsequent outcome
- Correlation a statistical measure that expresses the extent to which two variables are linearly related
- 5. **Effect** the difference between the true population parameter and the null hypothesis value
- 6. **Experimental** procedure carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law
- 7. **Linear** term used to describe a straight-line relationship between two variables
- 8. Observational facts or figures that's been collected about a given variable
- 9. **Regression** a statistical technique that relates a dependent variable to one or more independent (explanatory) variables
- 10. **Scatterplot** type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data
- Statistical Model model that embodies a set of statistical assumptions concerning the generation of sample data. It also represent (in idealized form) the data-generating process

R Commands

1. ! (not) -



25. ggplot -
26. knit -
27. library -
28. lm -
29. mapping -
30. read.csv -
31. sample -
32. sep -
33. size -
34. skim -
35. str -
36. summary -
37. table -
38. wordcloud2 -
39. write_csv -

PND

1. Binomial distribution -