AP Statistics Vocabulary List

1. **Alternative Hyothesis** — The hypothesis which states the Null Hypothesis is incorrect in a significance test.
2. **Binomial** — An experiment in which a set number of trials is used.
3. **Block Design** — The random assignment of subjects to treatments is carried out separately within each block.
4. **Boxplot** — Displays the 5-number summary as a central box with whiskers that extend to the non-outlying data values.
5. **Causation** — A cause and effect relationship in which one variable controls the changes in another variable.
6. **Center** — A descriptive feature which describes the placement and relation of the median to the other parts of the graphic representation.
7. **Central Limit Theorem** — The sampling distribution of the mean will approach the normal distribution as n increases (n>30).
8. **Chi-Squared Goodness of Fit** — uses sample data to test hypotheses about the shape or proportions of a population distribution. The test determines how well the obtained sample proportions fit the population proportions specified by the null hypothesis.
9. **Cluster Sample** — A sampling design in which entire groups are chosen at random.
10. **Coefficient of Determination** — Measures the percentage of variation in a dependent variable explained by one or more independent variables (r^2).
11. **Conditional Probability** — The probability that a particular event will occur, given that another event has already occurred.
12. **Confidence Interval** — The range of values within which a population parameter is estimated to lie.
13. **Confounded Variable** — An unintended difference between the conditions of an experiment that could have affected the dependent variable.
14. **Convenience Sample** — A sample that includes members of the population that are easily accessed.
15. **Correlation** — The measure of a relationship between two variables or sets of data.
16. **Data** — Information gathered from observations.
17. **Degrees of Freedom** — A parameter of the t distribution. When the t distribution is used in the computation of an interval estimate of a population mean, the appropriate t distribution has n-1 degrees of freedom, where n is the size of the simple random sample.
18. **Descriptive Statistics** — Statistical procedures used to describe characteristics and responses of groups of subjects.
19. **Discrete Random Variable** — Variable where the number of outcomes can be counted and each outcome has a measurable and positive probability.
20. **Dotplot** — Graphs a dot for each case against a single axis.
21. **Empirical rule** — States that, in a normal distribution, about 68% of the terms are within one standard deviation of the mean, about 95% are within two standard deviations, and about 99.7% are within three standard deviations (normal curve).
22. **Experiment** — The act of conducting a controlled test or investigation.
23. **Experimental Probability** — Probability based on what happens when an experiment is actually done.
24. **Extrapolation** — Estimating a value outside the range of measured data.
25. **Geometric** — An experiment in which there is no set number of trials but is ended by achieving an outcome.
26. **Histogram** — A bar graph that shows the frequency of data within equal intervals.
27. **Independent** — A relationship between two sets of data or two datum which states the outcome of one has no effect on the outcome of the other.
28. **Inferential Statistics** — Statistics that are used to interpret data and draw conclusions.
29. **Interpolation** — Using the Least Squares Regression Line to predict a y-value for an x-value within the x-data set.
30. **IQR** — Range of the middle 50% of the values; Q3-Q1 = 75th percentile - 25th percentile.
31. **Joint Frequency** — The number of responses for a given characteristic.
32. **Law of Large Numbers** — Law stating that a large number of items taken at random from a population will (on the average) have the population statistics.
33. **Least Squares Regression Line** — The line that minimizes the sum of squared residuals.
34. **Lurking Variable** — A variable other than x and y that simultaneously affects both variables, accounting for the correlation between the two.

35. **Margin of Error** — The range of percentage points in which the sample accurately reflects the population, the range surrounding a sample's response within which researchers are confident the larger population's true response would fall.

36. **Marginal Frequency** — Row and column totals in a contingency table (cross-tabulation) that represent the univariate frequency distributions for the row and column variables.

37. **Matched Pairs** — Either two measurements are taken on each individual such as pre and post OR two individuals are matched by a third variable (different from the explanatory variable and the response variable) such as identical twins.

38. **Mean** — The arithmetic average of a distribution, obtained by adding the scores and then dividing by the number of scores.

39. **Median** — The middle score in a distribution; half the scores are above it and half are below it.

40. **Mode** — The datum which occurs the most in a set of data.

41. **Mutually Exclusive** — Each event or variable is independent from one another. No event or variable will have an effect on the probability of outcome for any other event or variable.

42. **Non-Response Bias** — A bias caused by a number of people who did not respond to the survey.

43. **Normal** — A sample which follows the Empirical Rule for distribution.

44. **Null Hypothesis** — The hypothesis that states there is no difference between two or more sets of data in a significance test.

45. **Observational Study** — An experiment which observes individuals and measures variables of interest but does not attempt to influence the responses.

46. **Ogive** — A line graph that depicts cumulative frequencies.

47. **Outlier** — An extreme deviation from the mean.

48. **p-Value** — The probability of getting a result at least as extreme as the result given from the test. The lower the value the stronger the evidence.

49. **Parameter** — A numerical measurement describing some characteristic of a population.

50. **Placebo Effect** — Experimental results caused by expectations alone; any effect on behavior caused by the administration of an inert substance or condition, which is assumed to be an active agent.

51. **Population** — The entire aggregation of items from which samples can be drawn.

52. **Probability** — The likelihood that a particular event will occur.

53. **Qualitative** — Data identified by something other than numbers.

54. **Quantitative** — Data or datum being numerically defined.

55. **Quota Sample** — A sample deliberately constructed to reflect several of the major characteristics of a given population.

56. **Random Sample** — A sample in which every element in the sample has an equal chance of being selected.

57. **Residual** — The difference between an observed value of the response variable and the value predicted by the regression line.

58. **Response Bias** — Anything in the survey design that influences the responses from the sample.

59. **Sample** — Items selected at random from a population and used to test hypotheses about the population.

60. **Sample Space** — All possible outcomes of an experiment.

61. **Sampling Distribution** — The distribution of values taken by the statistic in all possible samples of the same size from the same population.

62. **Scatterplot** — A graphed cluster of dots, each of which represents the values of two variables.

63. **Simple Random Sample** — Every member of the population has a known and equal chance of selection.

64. **Simulation** — The act of repeating an experiment to get more accurate statistical evidence.

65. **Snowball Sample** — Samples in which informants provide contact information about other people who share some of the characteristics necessary for a study.

66. **Spread** — A descriptive feature in which describes the range of the data graphically.

67. **Standard Deviation** — A measure of variability that describes an average distance of every score from the mean (r).

68. **Standard Error** — The standard deviation of a sampling distribution.

69. **Standardized Value** — A value found by subtracting the mean and dividing by the standard deviation.

70. **Statistic** — A numerical measurement describing some characteristic of a sample.

71. **Statistical Significance**	When your discovered p-value is less than your alpha (.05 if not given). States that chance alone would rarely produce an equally extreme result.
72. **Stemplot**	A graphical representation of a quantitative data set. Leading values of each data point are presented as stems and second digits are given as leaves.
73. **Stratified Sample**	The population is divided into strata and a random sample is taken from each stratum.
74. **Systematic Sample**	A sample drawn by selecting individuals systematically from a sampling frame.
75. **t-Test**	A parametric inferential statistical test of the null hypothesis for a single sample where the population standard deviation is unknown.
76. **Theoretical Probability**	The ratio of the number of favorable outcomes to the number of possible outcomes if all outcomes have the same chance of happening.
77. **Two-way Table**	A table containing counts for two categorical variables. It has r rows and c columns.
78. **Type I Error**	An error that occurs when a researcher concludes that the independent variable had an effect on the dependent variable, when no such relation exists; a false positive.
79. **Type II Error**	An error that occurs when a researcher concludes that the independent variable had no effect on the dependent variable, when in truth it did; a false negative.
80. **Undercoverage**	Occurs when some groups in the population are left out of the process of choosing the sample.
81. **Voluntary Response Bias**	Bias introduced to a sample when individuals can choose on their own whether to participate in the sample.
82. **Wording Bias**	A type of response bias where the question is posed to achieve a desired result.
83. **z-Test**	A parametric inferential statistical test of the null hypothesis for a single sample where the population standard deviation is known.