

## AP Statistics Vocabulary List

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

35. <b>Margin of Error</b>	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.	51. <b>Population</b>	The entire aggregation of items from which samples can be drawn.
36. <b>Marginal Frequency</b>	Row and column totals in a contingency table (cross-tabulation) that represent the univariate frequency distributions for the row and column variables.	52. <b>Probability</b>	The likelihood that a particular event will occur.
37. <b>Matched Pairs</b>	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.	53. <b>Qualitative</b>	Data identified by something other than numbers.
38. <b>Mean</b>	The arithmetic average of a distribution, obtained by adding the scores and then dividing by the number of scores.	54. <b>Quantitative</b>	Data or datum being numerically defined.
39. <b>Median</b>	The middle score in a distribution; half the scores are above it and half are below it.	55. <b>Quota Sample</b>	A sample deliberately constructed to reflect several of the major characteristics of a given population.
40. <b>Mode</b>	The datum which occurs the most in a set of data.	56. <b>Random Sample</b>	A sample in which every element in the sample has an equal chance of being selected.
41. <b>Mutually Exclusive</b>	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.	57. <b>Residual</b>	The difference between an observed value of the response variable and the value predicted by the regression line.
42. <b>Non-Response Bias</b>	A bias caused by a number of people who did not respond to the survey.	58. <b>Response Bias</b>	Anything in the survey design that influences the responses from the sample.
43. <b>Normal</b>	A sample which follows the Empirical Rule for distribution.	59. <b>Sample</b>	Items selected at random from a population and used to test hypotheses about the population.
44. <b>Null Hypothesis</b>	The hypothesis that states there is no difference between two or more sets of data in a significance test.	60. <b>Sample Space</b>	All possible outcomes of an experiment.
45. <b>Observational Study</b>	An experiment which observes individuals and measures variables of interest but does not attempt to influence the responses.	61. <b>Sampling Distribution</b>	The distribution of values taken by the statistic in all possible samples of the same size from the same population.
46. <b>Ogive</b>	A line graph that depicts cumulative frequencies.	62. <b>Scatterplot</b>	A graphed cluster of dots, each of which represents the values of two variables.
47. <b>Outlier</b>	An extreme deviation from the mean.	63. <b>Simple Random Sample</b>	Every member of the population has a known and equal chance of selection.
48. <b>p-Value</b>	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.	64. <b>Simulation</b>	The act of repeating an experiment to get more accurate statistical evidence.
49. <b>Parameter</b>	A numerical measurement describing some characteristic of a population.	65. <b>Snowball Sample</b>	Samples in which informants provide contact information about other people who share some of the characteristics necessary for a study.
50. <b>Placebo Effect</b>	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.	66. <b>Spread</b>	A descriptive feature in which describes the range of the data graphically.
		67. <b>Standard Deviation</b>	A measure of variability that describes an average distance of every score from the mean ( $\bar{x}$ ).
		68. <b>Standard Error</b>	The standard deviation of a sampling distribution.
		69. <b>Standardized Value</b>	A value found by subtracting the mean and dividing by the standard deviation.
		70. <b>Statistic</b>	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.