Probability and Statistics Vocabulary List
(Definitions for Middle School Teachers)

B

- **Bar graph** – a diagram representing the frequency distribution for nominal or discrete data. It consists of a sequence of bars, or rectangles, corresponding to the possible values, and the length of each is proportional to the frequency.
    - o For more info:
        - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=50

Bayes' formula – Let $U_1, U_2, ..., U_n$ be n mutually exclusive events whose union is the sample space S. Let E be an arbitrary event in S such that $P(E) \neq 0$. Then

$$P(U_1|E) = \frac{P(U_1 \cap E)}{P(U_1 \cap E) + P(U_2 \cap E) + ... P(U_n \cap E)}.$$

Corresponding results hold for $U_2, U_3, ..., U_n$.
    - o For more info:
        - http://mathworld.wolfram.com/BayesTheorem.html

- **Binomial distribution** – the discrete probability distribution for the number of successes when *n* independent experiments are carried out, each with the same probability *p* of success.
    - o For more info:
        - http://mathworld.wolfram.com/BinomialDistribution.html

- **Binomial theorem** –the formulae $(x+y)^2 = x^2 + 2xy + y^2$ and $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$ are used in elementary algebra. The Binomial theorem gives an expansion like this for $(x+y)^n$, where n is any positive integer: $(x+y)^n = {}^nC_0x^n + {}^nC_1x^{n-1}y + {}^nC_2x^{n-2}y^2 + ... + {}^nC_{n-1}xy^{n-1} + {}^nC_ny^n$.
    - o For more info:
        - http://www.mathwords.com/b/binomial_theorem.htm

- **Bins** – a term used to describe class intervals on a histogram.
    - o For more info:
        - http://online.math.uh.edu/MiddleSchool/Modules/Module_5_Prob_Stat/Content/Stat/Ch4_3.doc

- **Bivariate data** – Data involving two random variables, such as height and weight, or amount of smoking and measure of health; often graphed in a scatter plot.
    - o For more info:
        - http://cnx.rice.edu/content/m10949/latest/
        - http://www.introductorystatistics.com/escout/main/Glossary.htm

- **Box-and-whisker plot** – a diagram constructed from a set of numerical data showing a box that indicates the middle 50% of the marked observations together with lines, sometime called 'whiskers', that go out from the quartile to the most extreme data value in that direction which is not more than 1.5 times the Inter Quartile Range from the quartile.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=57
    - ▪ http://www.introductorystatistics.com/escout/main/Glossary.htm

C
- **Categorical data** – data that fits into a small number of discrete categories. Categorical data is either non-ordered (nominal) such as gender or city, or ordered (ordinal) such as high, medium, or low temperature.
  - o For more info:
    - ▪ http://www.stat.yale.edu/Courses/1997-98/101/catdat.htm

- **Central limit theorem** – it pertains to the convergence in distribution of (normalized) sums of random variables. The distribution of the mean of a sequence of random variables tends to a normal distribution as the number in the sequence increases indefinitely. The most general version of the C.L.T states: Let $X_1, X_2, ...$ be a sequence of independent, identically distributed random variables with mean μ and finite variance σ^2. Let $\overline{X_n} = \dfrac{\sum\limits_{i=1}^{n} X_i}{n}, Z_n = \dfrac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma}$.
  Then as n increases indefinitely, the distribution of Z_n tends to the standard normal distribution.
  - o For more info:
    - ▪ http://mathworld.wolfram.com/CentralLimitTheorem.html

- **Circle graph** – a graph for categorical data. The proportion of elements belonging to each category is proportionally represented as a pie-shaped sector of a circle. Sometimes called a pie chart.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=68

- **Class intervals** – a subdivision within a range of values. In a histogram, the range of values is divided into sections, known as class intervals, also referred to as "bins."
  - o For more info:
    - ▪ http://online.math.uh.edu/MiddleSchool/Modules/Module_5_Prob_Stat/Content/Stat/Ch4_3.doc
    - ▪ http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Clusters of data** – a portion of high concentration in a data set.
  - o For more info:
    - ▪ http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Combination** – the number of ways of picking *k unordered* outcomes from *n* possibilities. The number of combinations of *n* distinct objects taken *r* at a time is $C(n,r) = \dfrac{n!}{(n-r)!r!}$.

  For example: The number of ways in which a committee of 2 people can be selected out of 4 people is $C(4,2) = \dfrac{4!}{(4-2)!2!} = 6\ ways$. Let these people be denoted as A, B, C, and D. Then the set of all possible combinations is {AB, AC, AD, BC, BD, CD}. Note that the choice AB is equivalent to BA, i.e. ordering doesn't matter.
  - For more info:
    - http://www.pballew.net/permute.html

- **Complement of an event** – suppose A is an event in the universal set U, the complement of *A* ("not *A*") consists of all the outcomes in U that are not in *A*. For example, if *A* is the event that two of three children are boys, then A^c (complement of A) is the event that there are either zero, one, or three boys.
  - For more info:
    - http://www.mathgoodies.com/lessons/vol6/complement.html
    - http://www.mathwords.com/c/complement_event.htm

- **Compound events** - an event made of two or more simple events.
  - For more info:
    - http://www.harcourtschool.com/glossary/math2/define/gr6/compound_event6.html

- **Conditional probability** – let A and B be two events. The probability that A will occur given that B has already occurred is the 'conditional probability of A given B' and is denote by $P(A|B)$.
  - For more info:
    - http://www.mathwords.com/c/conditional_probability.htm
    - http://www.cut-the-knot.org/fta/Buffon/ConditionalProbability.shtml

- **Confidence interval** – an interval, calculated from a sample, which contains the value of a certain population parameter with a specified probability.

- **Confidence level** - the probability that the statistician's confidence interval contains the true, unknown population parameter.

- **Correlation** – the correlation between two variables x and y is a measure of how closely related they are, or how linearly related they are. Correlation is the measure of the extent to which a change in one random variable tends to correspond to change in the other random variable. For example, height and weight have a moderately strong positive correlation.
  - For more info:
    - http://www.mathwords.com/c/correlation.htm

- **Correlation coefficients** – a measure of how close two random variables are to being perfectly linearly related; computed by dividing the covariance of the random variables by the product of their standard deviation. The correlation coefficient denoted by ρ takes values between -1 and 1; -1 represents a perfect *negative* correlation while 1 represents a perfect *positive* correlation.
  - For more info:
    - http://www.mathwords.com/c/correlation_coefficient.htm

- **Counting Principle** – a method used to compute the number of possible outcomes of an experiment. If each outcome has independent parts, the total number of possible outcomes can be found by multiplying the number of choices for each part.
  - For more info:
    - http://www.richland.cc.il.us/james/lecture/m116/sequences/counting.html
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=94

- **Cumulative frequency** – the sum of the frequencies of all the values up to a given value. If the values $x_1, x_2, ..., x_n$, in ascending order, occur with frequencies $f_1, f_2, ..., f_n$ respectively, then the cumulative frequency at x_i is $f_1 + f_2 + ... f_i$.
  - For more info:
    - http://www.harcourtschool.com/glossary/math2/define/gr4/cumulative_frequency4.html
    - http://www.gcseguide.co.uk/cumulative_frequency.htm
    - http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Cumulative relative frequency** (relative cumulative frequency) –
  - The cumulative frequency in a frequency distribution divided by the total number of data points.
  - For more info:
    - http://www.netnam.vn/unescocourse/statistics/28.htm
    - http://mathworld.wolfram.com/RelativeCumulativeFrequency.html

D
- **Data** – the observations gathered from an experiment, survey or observational study.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=102

- **Density function** – a mathematical function used to determine probabilities for a continuous random variable. For example, the bell-shaped curve corresponding to a normal distribution.
  - For more info:
    - http://www.google.com/search?hl=en&lr=&ie=UTF-8&oi=defmore&q=define:density+function

- **Dependent event** – two events are dependent if the occurrence of either affects the probability of the occurrence of the other.
  - o For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=109
    - http://www.harcourtschool.com/glossary/math2/define/gr6/dependent_events6.html
    - http://www.gomath.com/htdocs/lesson/probability_lesson9.htm

- **Designed experiment** – the process of planning an experiment or evaluation so that appropriate data will be collected, which may be analyzed by statistical methods resulting in valid and objective conclusions. Examples include: complete random design, random design, and randomized block design.

- **Deterministic experiment** –a process in which the outcome is known in advance. For example tossing a two headed coin.

- **Disjoint** – sets are disjoint if they have no elements in common. For example, the sets A = {1,2,3} and B = {5,6,7} are disjoint.

- **Dispersion** – a way of describing how scattered or spread out the observations in a sample are. Common measures of dispersion are the range, inter quartile range, variance, and standard deviation.

- **Distribution** – the distribution of a random variable is the way in which the probability of it taking a certain value, or a value within a certain interval is described. It may be given by the cumulative distribution function, the probability mass function (discrete random variable) or the probability density function (continuous random variable).

E
- **Element** – an object in a set is an element of that set.

- **Empirical probability** – the probability of an event determined by repeatedly performing an experiment. It may be determined by dividing the number of times the event occurred by the number of times the experiment was repeated.
  For more info:
    - http://www.oswego.org/mtestprep/math8/g/empprobprac.cfm

- **Equality** (of sets) – sets A and B are equal if they consist of the same elements. In order to establish A=B, a technique that can be useful is to show that each is contained in the other.

- **Equally likely outcomes** – every outcome of an experiment has the same probability. For example: rolling a fair die has equally likely outcomes.
  - o For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=127

- **Event** – a subset of the sample space. For example, the sample space for an experiment in which a coin is tosses twice is given by {HH, HT, TH, TT} and let A= {HT, HH}, then A is an event in which Head occurs at the first place.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=140

- **Expected value** – it is the average value of a random quantity that has been repeatedly observed in replications of an experiment. For example, if a fair 6-sided die is rolled, its expected value is 3.5.
  - o For more info:
    - ▪ http://www.mathwords.com/e/expected_value.htm

- **Experiment** – processes in which there are an observable set of outcomes are called experiments. For example, the following are all experiments: tossing a coin, rolling a die, or selecting a ball from a bag.
  - o For more info:
    - ▪ http://www.mathwords.com/e/experiment.htm

- **Experimental probability** – The estimated probability of an event; obtained by dividing the number of successful trials by the total number of trials.

F
- **Fair coin** – a fair coin is defined as coin where the probability of landing heads up or tails up are the same (0.5).

- **Fair game** – a game is fair if each player has an equal chance of winning.

- **Finite sample space** – a sample space which contains a finite number of possible outcomes.

- **Frequency** – the number of times that a particular value occurs as an observation.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=396

- **Frequency distribution** – the information consisting of the possible values/groups and the corresponding frequencies is called the frequency distribution.
  - o For more info:
    - ▪ http://infinity.sequoias.cc.ca.us/faculty/woodbury/Stats/Tutorial/Data_Freq.htm

- **Frequency table** – a table giving the number of data points in a data set falling in each of a set of given intervals.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=396

G

- **Geometric distribution** – a discrete probability distribution for the number of trials required to achieve the first success in a sequence of independent trials, all with the same probability 'p' of success. Its probability mass function is given by $P[X = x] = p(1-p)^{x-1}$. For example, the number of times one must toss a fair coin until the first time the coin lands heads has a geometric distribution with parameter $p = 50\%$.

- **Geometric probability** – the probability of an event as determined by comparing the areas (or perimeters, angle measures, etc.) of the regions of success of an event to the total area of the figure (sample space).

H
- **Histogram** – A bar graph presenting the frequencies of occurrence of data points. Sometimes called a frequency histogram
  - For more info:
    - http://www.shodor.org/interactivate/activities/histogram/

I
- **Independent event** – two events are independent if the outcome of one event has no effect on the outcome of the other.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=173

- **Infinite sample space** – a sample space containing an infinite number of possible outcomes.

- **Inter-quartile range** – the difference between the first quartile and third quartile of a set of data, (IQR).
  - For more info:
    - http://www.mathwords.com/i/interquartile_range.htm

- **Intersection of Sets** – the intersection of sets A and B, denoted by $A \cap B$, is the set of elements that are in both A and B.
  - For more info:
    - http://www.mathwords.com/i/intersection.htm
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=182

- **Interval scale** –a scale of measurement where the distance between any two adjacent units of measurement (or 'intervals') is the same but the zero point is arbitrary.
  - For more info:
    - http://eksl-www.cs.umass.edu/eis/pages/glossary/entries/interval-scale.html

- **Interval variable** – an interval variable is similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced.
  - For more info:
    - http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm

L

- **Least squares** – is used to estimate parameters in statistical models such as those that occur in regression. Estimates for the parameter are obtained by minimizing the sum of the squares of the differences between the observed values and the predicted values under the model.

- **Linear regression** – a method for finding an equation for the line that best fits the data set. The method is based on minimizing the sum of the squared vertical distances from the data points to the line of best fit.

- **Line-of-best-fit** – the line that best represents the trend that the points in a scatter plot follow.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=198

- **Line plot** – a line graph that orders the data along a real number line. Also called a dot plot. For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=200

M

- **Maximum value** – The maximum is highest point on a graph or the largest number in a data set.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=218

- **Mean** – the mean is an appropriate location measure for interval or ratio variables. Suppose there are N individuals in the population and x denotes an interval or ratio variable. Let values for the ith individual be denoted by x_i. The mean of x is the

  number $\overline{x} = \dfrac{x_1 + x_2 + .. + x_N}{N}$.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=210

- **Measures of central tendency** –measures of the location of the middle or the center of a distribution. The definition of "middle" or "center" is purposely left somewhat vague so that the term "central tendency" can refer to a wide variety of measures. The three most common measures of central tendency are the mean, median, and mode.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=212

- **Median** – suppose the observations in a set of numerical data are ranked in ascending order. The median is the middle observation if there are an odd number of observations, and is the average of the two middlemost observations if there are an even number of observations.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=213

- **Minimum value** – The minimum is the lowest point on a graph, or the smallest number in a data set.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=217

- **Mode** –
  - The mode is the most frequently occurring value in a set of discrete data. There can be more than one mode if two or more values are equally common.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=219

- **Mutually exclusive** – events that have no outcomes in common. Events A and B are mutually exclusive if $A \cap B = \phi$.
  - For more info:
    - http://www.mathwords.com/m/mutually_exclusive.htm

N
- **N factorial** – for a positive integer n, the notation n! is used for the product $n \times (n-1) \times (n-2) .... \times 2 \times 1$. For example $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$. 0! is defined to be 1.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=148
    - http://www.ilovemaths.com/3factorial.htm

- **Nominal data** – a data set is said to be nominal if the observations belonging to it can be assigned a code in the form of a number where the numbers are simply labels. You can count but not order or measure nominal data. For example, in a data set males could be coded as 0, females as 1; marital status of an individual could be coded as Y if married, N if single.

- **Normal distribution** – a continuous probability distribution (with parameters μ and σ) whose probability density function f is given by $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(\dfrac{-(x-\mu)^2}{2\sigma^2} \right)$. Normal distributions are symmetric about their mean and have bell-shaped density curves.

- **Notation** – notations are symbols denoting quantities, operations, etc.

O

- **Observational study** – a study where data is collected through observation.

- **Odds** – a way of representing the likelihood of an event's occurrence. The odds m:n in favor of an event means we expect the event will occur *m* times for every *n* times it does not occur.
  - o For more info:
    - ▪ http://www.mathwords.com/o/odds.htm

- **Ogive** – any continuous cumulative frequency.
  - o For more info:
    - ▪ http://mathworld.wolfram.com/Ogive.html

- **One-variable data** – a collection of related behaviors that are associated in one meaningful way.

- **One-variable data table** – a table showing a collection of related behaviors that are associated in one meaningful way.

- **Ordinal data** – a set of data is said to be ordinal if the values / observations belonging to it can be ranked (put in order) or have a rating scale attached. For example: Questionnaire responses coded: 1 strongly disagree, 2 disagree, 3 indifferent, 4 agree, 5 strongly agree**.**

- **Outcome** – an outcome is the result of an experiment. The set of all possible outcomes of an experiment is the sample space.

- **Outliers of data** – an observation that is deemed to be unusual and possibly erroneous because it does not follow the general pattern of the data in the sample.
  - o For more info:
    - ▪ http://www.mathwords.com/o/outlier.htm

P

- **Pascal's triangle** – Pascal's triangle is a number triangle with numbers arranged in staggered rows such that $a_{nr} \equiv {}^n C_r$ starting with $n = 0 \, and \, r = 0,1,..n$. It looks like :

$$
\begin{array}{ccccccc}
 & & & 1 & & & \\
 & & 1 & & 1 & & \\
 & 1 & & 2 & & 1 & \\
1 & & 3 & & 3 & & 1
\end{array}
$$

  - o For more info:
    - ▪ http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Percentile** – the n-th percentile is the value $x_{n/100}$ such that n percent of the population is less than or equal to $x_{n/100}$. The 25th, 50th and 75th percentiles are called quantiles.
  - o For more info:
    - ▪ http://www.mathwords.com/p/percentile.htm

- **Permutation** – a permutation of n objects can be thought of as all the possible ways of their arrangement or rearrangement. The number of 'permutations of n objects taken r at a time' is denoted by ^nP_r $which\,equals$ $\dfrac{n!}{(n-r)!}$. For example there are 6 permutations of A, B, C taken two at a time: AB, AC, BA, BC, CA, and CB.
  - o For more info:
    - http://www.mathwords.com/p/permutation.htm

- **Pie chart (also known as circle graph)** – a graph for categorical data. The proportion of elements belonging to each category is proportionally represented as a pie-shaped sector of a circle.
  - o For more info:
    - http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Population** – the entire set of items from which data can be selected. For example, a poll given to a sample of voters is designed to measure the preferences of the population of all voters.
  - o For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=269

- **Population variable** – collection of related behaviors of a group that are associated in a meaningful way.

- **Principle** – a basic generalization that is accepted as true and that can be used as a basis for reasoning.

- **Probability** – the chance/likelihood that a particular event (or set of events) will occur expressed on a scale from 0 (impossibility) to 1 (certainty), also expressed as a percentage between 0 and 100%.
  - o For more info:
    - http://www.mathwords.com/p/probability.htm
    - http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Probabilistic experiment** – a probabilistic experiment is an occurrence such as the tossing of a coin, rolling of a die, etc. in which the complexity of the underlying system leads to an outcome that cannot be known ahead of time.

Probability (density) function – the probability function *f(x)* of a continuous distribution is defined as the derivative of its cumulative distribution function F(x).

- **Probability models** – A *probability model* is a mathematical representation of a random phenomenon. It is defined by its *sample space*, *events* within the sample space, and *probabilities* associated with each event.

Q

- **Quartile** – for numerical data ranked in ascending order, the quartiles are values derived from the data which divide the data into four equal parts.

R

- **Random experiment** –an experiment whose outcome cannot be predicted with certainty, before the experiment is run**.**
  o For more info:
  ▪ http://www.ds.unifi.it/VL/VL_EN/prob/prob1.html

- **Random number generator** – a program that will generate random numbers (as output).

- **Random sample** – a set of data chosen from a population in such a way that each member of the population has an equal probability of being selected
  o For more info:
  ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=292

- **Random variable** – a variable that takes different numerical values according to the result of a particular experiment. Another definition: A numerically valued function X of ω with domain Ω such that $\omega \in \Omega : \omega \to X(\omega)$ is called a random variable [on Ω].

- **Range** – the range of a sample (or a data set) is a measure of the spread or the dispersion of the observations. It is the difference between the largest and the smallest observed value.
  o For more info:
  ▪ http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Ratio variable** –a comparison of two quantities, expressed as a fraction where the quantity can assume any of a set of values.

- **Regression line** - a straight line used to estimate the relationship between two variables, based on the points of a scatter plot; often determined by a least squares analysis. When it slopes down (from top left to bottom right), this indicates a negative or inverse relationship between the variables; when it slopes up (from bottom left to top right), a positive or direct relationship is indicated.

- **Relative frequency** – Relative frequency is another term for proportion; it is the value calculated by dividing the number of times an event occurs by the total number of times an experiment is carried out. The probability of an event can be thought of as its long-run relative frequency when the experiment is carried out many times.
  o For more info:
  ▪ http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- **Replacement** – replacing/ returning an item back into the sample space after an event and thus allowing an item to be chosen more than once.

- **Residual** – residual represents unexplained (or residual) variation after fitting a regression model. It is the difference between the observed value of the variable and the value suggested by the regression model.
  - For more info:
    - http://www.mathwords.com/r/residual.htm

- **Residual variance** – the square of the standard error of estimate.

S
- **Sample** – a subset of a population that is obtained through some process, possibly random selection or selection based on a certain set of criteria, for the purposes of investigating the properties of the underlying parent population.
  - For more info:
    - http://www.teachers.ash.org.au/jeather/maths/dictionary.html
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=316

- **Sample size** – the sample size is the number of items in a sample.

- **Sample space** – the set of all possible outcomes of a probability experiment.
  - For more info:
    - http://www.mathwords.com/s/sample_space.htm
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=317

- **Sample variance** – Sample variance is a measure of the spread or dispersion within a set of sample data.
  - For more info:
    - http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- **Sampling** – the process of selecting a proper subset of elements from the full population so that the subset can be used to make inference to the population as a whole.
  - For more info:
    - http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- **Scatter-plot** – a graph of two-variable (bivariate) data in which each point is located by its coordinates (X, Y).
  - For more info:
    - http://www.mathwords.com/s/scatterplot.htm
- **Set** – a set is a well-defined collection of objects. Sets are written using set braces {}. For example, {1,2,3} is the set containing the elements 1, 2, and 3.
  - For more info:
    - http://www.mathwords.com/s/set.htm
    - http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Simple event** – an event which is a single element of the sample space.

- **Simulation** – an experiment that models a real-life situation**.**
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=335

- **Single-variable data –** data that uses only one unknown.

- **Skewness** – the degree of asymmetry of a distribution.
  - For more info:
    - http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- **Spread of data**- the degree to which data are spread out around their center. Measures of spread include the mean deviation, variance, standard deviation, and interquartile range

- **Standard deviation** – standard deviation is a measure of the spread or dispersion of a set of data. It is defined as the square root of the variance.
  - For more info:
    - http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- **Standard normal distribution** – a normal distribution with parameters 0(mean) and 1(variance).
  - For more info:
    - http://www.ruf.rice.edu/~lane/hyperstat/A75494.html

- **Statistical inference** – statistical inference makes use of information from a sample to draw conclusions (inferences) about the population from which the sample was taken.
  - For more info:
    - http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- **Statistics** – the branch of mathematics that deals with the collection, organization, and interpretation of data.
  - For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=346

- **Stem-and-leaf plot** – a semi-graphical method used to represent numerical data, in which the first (leftmost) digit of each data value is a stem and the rest of the digits of the number are the leaves.
  - For more info:
http://online.math.uh.edu/MiddleSchool/Modules/Module_5_Prob_Stat/Content/Stat/Ch4_3.pdf

- **Subset** – set A is a subset of set B if all of the elements of set A are contained in set B. It is written as A $\subseteq$ B.
  - For more info:

- ▪ http://www.mathwords.com/s/subset.htm
- ▪ http://mathworld.wolfram.com/Subset.html

T
- **Table** – mathematical information organized in columns and rows.
  - o For more info:
    - ▪ http://www.teachers.ash.org.au/jeather/maths/dictionary.html

- **Theoretical probability** – the theoretical probability of an event, *P* (event), is the ratio of the number of outcomes in the event to the number of outcomes in the sample space, if all outcomes are equally likely.

- **Theoretical regression line** – The line of best fit drawn through a scatter-plot before values are actually calculated.

- **Tree diagram** – A tree diagram displays all the possible outcomes of an event.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=363

U
- **Unbiased estimator** – for an estimator to be unbiased it is required that *on average* the estimator will yield the true value of the unknown parameter. An estimator X is an unbiased estimator of the parameter θ if $E(X) = \theta$.
  - o For more info:
    - ▪ http://mathworld.wolfram.com/UnbiasedEstimator.html

- **Unfair game** –a game in which all players do not have the same probability of winning.

- **Union** – combining the elements of two or more sets. Union is indicated by the $\cup$ (cup) symbol.
  - o For more info:
    - ▪ http://www.mathwords.com/u/union.htm

- **Union of sets** – the union of two sets *A* and *B* is the set obtained by combining the members of each the set. If A={1,2.3} and B={2,4,6} , then $A \cup B$ ={ 1,2,3,4,6}.
  - o For more info:
    - ▪ http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=376

- **Universal set** – a set that contains all the elements being considered in the given discussion or problem.

V

- **Variable** – a quantity that varies. For example, the weight of a randomly chosen member of a football team is such a variable. Variables are usually represented by letters.
  - o For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=379

- **Variance** – a measure of the amount of spread (variation) in a set of data; the larger the variance, the more scattered the observations on average.
  - o For more info:
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=449

- **Venn diagram** – a graphic means of showing intersection and union of sets by representing them as bounded regions.
  - o For more info:
    - http://www.teachers.ash.org.au/jeather/maths/dictionary.html
    - http://www.intermath-uga.gatech.edu/dictnary/descript.asp?termID=380

W
- **Weighted arithmetic mean** (Weighted Average) – a method of computing a kind of arithmetic mean of a set of numbers in which some elements of the set carry more importance (weight) than others.
  - o For more info:
    - http://www.mathwords.com/w/weighted_average.htm