Statistics

Mathematics Curriculum Framework

Revised 2004

Course Title: Statistics (Fourth-Year Course)
Course/Unit Credit: 1
Course Number:
Teacher Licensure: Secondary Mathematics
Pre-requisite: Algebra II
Grades: 9-12

Statistics

This two-semester fourth-year course is designed for students who have completed Algebra II. Statistics may be taken before, in place of, or concurrently with other fourth-year mathematics. Statistics provides students with an understanding of the concepts of mathematics analysis and normal curve distribution and measures of variability. This course satisfies the proficiency requirement for those students who need to develop their numeric and statistical literacy. Statistics information about your city or town, gathered from local news sources, social agencies, and the U. S. Census, is used as a vehicle to investigate ways of collecting, displaying, and analyzing data. Basic statistical topics such as percentiles, the normal distribution, standard deviation, and fitting curves to data will be studied through analysis of the city's statistics. To develop a student's critical sense in reading numerical information, the course will discuss ways of expressing quantitative data numerically, via percentages and proportions, and visually, using tables, graphs, and charts.

Strand	Standard
Descriptive Statistics	
	1. Students will create, compare, and evaluate data displays using such methods as *histograms*, cumulative distribution functions, and scatter plots. For these data, they calculate measures of central tendency (various kinds of *means*, the *median*, and the *mode*) and their derivatives (*range*, *variance*, and *standard deviation*).
Data Collection	
	2. Students will describe the method of data collection in a census, *sample survey*, *experiment*, and *observational study*, and identify an appropriate method of solution for a given familiar or unfamiliar contextual problem. Students will plan and conduct a survey. The plan will address sampling techniques (simple random and stratified) and methods to reduce bias.
	3. Students will construct and interpret display of data to solve problems.
Data Analysis	
	4. Students will collect and analyze data to solve problems
	5. Students will use statistical models to describe and analyze sets of data.
Probability	
	6. Students will compute and distinguish between *permutations* and *combinations* and use technology for application.
	7. Students will identify random variables as independent or dependent and find mean and standard deviations for sums and differences of independent random variables.
	8. Students will find probabilities, including conditional probabilities for events that are either dependent or independent, by applying the *law of large numbers*, the *addition rule*, and the
	9. Students will develop, interpret, and apply the binomial probability distribution for discrete random variable, including computing the mean and standard deviation for the binomial variable.
Statistical Inference	
	10. Students will use probability distributions to make statistical inferences.
	11. Students will use *confidence intervals* and hypothesis tests, fit curves to data, and calculate *correlation coefficients*.

Descriptive Statistics

Content Standard 1. Students will create, compare, and evaluate data displays using such methods as *histograms*, cumulative distribution functions, and scatter plots. For these data, they calculate measures of central tendency (various kinds of *means*, the *median*, and the *mode*) and their derivatives (*range*, *variance*, and *standard deviation*).

DS.1. S.1	Create, compare, and evaluate different graphic displays of the same data, using histograms, *frequency* polygons, cumulative distribution functions, pie charts, scatter plots, stem-and-leaf plots, and box-and-whisker plots and draw these by hand or use a computer spread sheet program (Ex: Gather data to answer the question: Which area of the country has the highest school dropout rate? Display your dropout data in various forms.)
DS.1.S.2	Compute and use mean, mode, *weighted mean*, *geometric mean*, *harmonic mean*, *range*, *quartiles*, variance, and standard deviation (Ex: Use spreadsheet formulas to compute measures that summarize your dropout data by state.)

Key: DS.1.S.1 = Descriptive Statistics. Content Standard 1. Statistics. 1st Student Learning Expectation

Data Collection

Content Standard 2. Students will describe the method of data collection in a census, *sample survey*, *experiment*, and *observational study*, and identify an appropriate method of solution for a given familiar or unfamiliar contextual problem. Students will plan and conduct a survey. The plan will address sampling techniques (simple random and stratified) and methods to reduce bias.

DC.2.S.1	Compare and contrast *controlled experiments* and observational studies and the conclusions one can draw from each
DC.2.S.2	Compare and contrast *population* and *sample*, and *parameter* and *statistic*
DC.2.S.3	Identify biased sampling methods
DC.2.S.4	Describe simple *random sampling*
DC.2.S.5	Select a data collection method appropriate for a given context
DC.2.S.6	Investigate and describe sampling techniques, such as simple random sampling, *stratified sampling*, and *cluster sampling*
DC.2.S.7	Determine which sampling technique is best, given a particular context
DC.2.S.8	Plan and conduct a survey to answer a question or address an issue, identify possible sources of *bias,* and describe ways to reduce bias

Key: DC.2.S.1 = Data Collection. Content Standard 2. Statistics. 1st Student Learning Expectation

Data Collection

Content Standard 3. Students will construct and interpret display of data to solve problems.

DC.3.S.1	Analyze categorical data
DC.3.S.2	Use and compare methods of data collection
DC.3.S.3	Apply statistical principles and methods in sample surveys; identify difficulties
DC.3.S.4	Apply concepts of probability to solve familiar and unfamiliar contextual problems
DC.3.S.5	Use simulations to develop an understanding of the *Central Limit Theorem* and its importance in *confidence intervals* and tests of significance
DC.3.S.6	Recognize, construct and interpret results using confidence intervals in the context of a problem

Key: DC.3.S.1 = Data Collection. Content Standard 3. Statistics. 1st Student Learning Expectation

Data Analysis

Content Standard 4: Students will collect and analyze data to solve problems

DA.4.S.1	Summarize distributions of *univariate* data by determining and interpreting measures of center, spread, position, boxplot, and effects of changing units on summary measures.
DA.4.S.2	Analyze distribution of continuous univariate data (both normal and non-normal)
DA.4.S.3	Construct and interpret graphical display of data
DA.4.S.4	Compare *distributions* among sets of data.

Key: DA.4.S.1 = Data Analysis. Content Standard 4. Statistics. 1st Student Learning Expectation

Data Analysis

Content Standard 5: Students will use statistical models to describe and analyze sets of data.

DA.5.S.1	Investigate and solve relevant problems, using technology to collect, organize, display, and analyze data in tabular, graphical, and symbolic forms
DA.5.S.2	Use linear and nonlinear models to formulate predictions from data
DA.5.S.3	Recognize the limitations of mathematical models based on sample data as representations of real world situations
DA.5.S.4	Identify possible *correlations* between variables in a data set
DA.5.S.5	Develop, use, and explain application and limitations of linear models and line of best fit (linear regression) in a variety of contexts
DA.5.S.6	Use data from samples to make inferences about a population and determine whether claims are reasonable or unreasonable
DA.5.S.7	Determine and use measures of central tendency and dispersion to describe and compare sets of data
DA.5.S.8	Design, conduct, interpret, and justify the results of a probability experiment, sample, or statistical simulation

Key: DA.5.S.1 = Data Analysis. Content Standard 5. Statistics. 1st Student Learning Expectation

Probability

Content Standard 6: Students will compute and distinguish between *permutations* and *combinations* and use technology for application.

P.6.S.1	Understand the counting principle, permutations and combinations and use them to solve problems
P.6.S.2	Compare and contrast permutations and combinations
P.6.S.3	Calculate the number of permutations of *n* objects taken *r* at a time
P.6.S.4	Calculate the number of combinations of *n* objects taken *r* at a time
P.6.S.5	Calculate *relative frequency* and expected frequency
P.6.S.6	Find *conditional probabilities* for *dependent*, *independent*, and *mutually exclusive* events

Key: P.6.S.1 = Probability. Content Standard 6. Statistics. 1st Student Learning Expectation

Probability

Content Standard 7: Students will identify random variables as independent or dependent and find mean and standard deviations for sums and differences of independent random variables.

P.7.S.1	Compare and contrast independent and dependent random variables
P.7.S.2	Find the standard deviation for sums and differences of independent random variables

Key: P.7.S.1 = Probability. Content Standard 7. Statistics. 1st Student Learning Expectation

Probability

Content Standard 8: Students will find probabilities, including conditional probabilities for events that are either dependent or independent, by applying the *law of large numbers*, the *addition rule*, and the *multiplication rule*.

P.8.S.1	Understand and use the addition rule to calculate probabilities for mutually exclusive and other events
P.8.S.2	Understand and use the multiplication rule to calculate probabilities for independent and dependent events
P.8.S.3	Develop the *binomial distribution* within a real world context
P.8.S.4	Calculate the mean and standard deviation for a binomial variable
P.8.S.5	Use the binomial distribution to calculate probabilities associated with experiments for which there are only two possible outcomes

Key: P.8.S.1 = Probability. Content Standard 8. Statistics. 1st Student Learning Expectation

Probability

Content Standard 9: Students will develop, interpret, and apply the binomial probability distribution for discrete random variable, including computing the mean and standard deviation for the binomial variable.

P.9.S.1	Design and conduct an experiment that simulates a binomial distribution.
P.9.S.2	Design and conduct an experiment that simulates a *geometric distribution*.
P.9.S.3	Simulate probability distributions, including binomial and geometric.

Key: P.9.S.1 = Probability. Content Standard 9. Statistics. 1st Student Learning Expectation

Statistical Inference

Content Standard 10: Students will use probability distributions to make statistical inferences.

SI.10.S.1	Explore the characteristics and applications of the *normal distribution* and *standardized* scores
SI.10.S.2	Explore a variety of statistical tests such as *chi-squares* and *t-tests* and understand the meaning of *hypothesis testing*
SI.10.S.3	Use relative frequency and expected values to represent and solve problems involving uncertainty

Key: SI.10.S.1 = Statistical Inference. Content Standard 10. Statistics. 1st Student Learning Expectation

Statistical Inference

Content Standard 11: Students will use *confidence intervals* and hypothesis tests, fit curves to data, and calculate *correlation coefficients*.

SI.11.S.1	Compute and use confidence intervals to make an estimate
SI.11.S.2	Understand hypothesis tests of means and differences between means and use them to reach a conclusion
SI.11.S.3	Use the principle of least squares to find the curve of best fit for a set of data
SI.11.S.4	Calculate and interpret the correlation coefficient of a set of data

Key: SI.11.S.1 = Statistical Inference. Content Standard 11. Statistics. 1st Student Learning Expectation

Addition Rule	The sum of the probability of events A and B minus the probability of the intersection of events A and B
Bias	The average (expected) difference between the measurement and the truth
Binomial Distribution	A random variable that has the number of "successes" in a fixed number *n* of independent random trials, all of which have the same probability *p* of resulting in "success"
Central Limit Theorem	States that the probability histograms of the sample mean and sample sum of *n* draws with replacement from a box of labeled tickets converge to a normal curve as the sample size *n* grows
Chi-square curve	An approximation to the probability histogram of the chi-square statistic for multinomial model if the expected number of outcomes in each category is large
Cluster Sample	The sampling unit is a collection of population units, not single population units
Combinations	The number of combinations of *n* things taken *k* at a time is the number of ways of picking a subset of *k* of the *n* things, without replacement, and without regard to the order in the elements of the subset are picked.
Conditional Probability	When event A is expected to occur but event B occurs, this knowledge should affect the probability that A occurred quantitatively.
Confidence Interval	An estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data
Controlled Experiment	An experiment that uses the method of comparison to evaluate the effect of a treatment by comparing treated subjects with a control group, who do not receive the treatment
Correlation	A measure of linear association between two (ordered) lists where two variables can be strongly correlated without having any causal relationship, and two variables can have a causal relationship and yet be uncorrelated.
Correlation Coefficient	A measure of how nearly a scatter plot falls on a straight line (The correlation coefficient is always between –1 and +1.)
Dependent Events	Two events are dependent if they are not independent.
Dependent Variable	In regression, the variable whose values are supposed to be explained by changes in the other variable (the independent or explanatory variable)
Distribution	How the values of a set of numerical data are distributed over the real numbers
Experiment	What distinguishes an experiment from an observational study is that in an experiment, the experimenter decides who receives the treatment.
Frequency	The number, fraction, or percentage of observations in different ranges, called class intervals, used to compare outcomes
Geometric Distribution	Describes the number of trials up to and including the first success, in independent trials with the same probability of success

Geometric Mean	In *n* numbers, it is the *n*th root of their product (The geometric mean of 6 and 24 is 6/x = x/24, where x^2 = 144, and x = 12.) Ex: 6 x 2 = 12, 12 x 2 = 24
Harmonic Mean	$If\ a_1, a_2, ..., a_n$ are positive numbers, we define their harmonic mean as $H.M. = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + ... + \frac{1}{a_n}}$
Histogram	A graphic representation of the frequency distribution of a continuous variable (Rectangles are drawn in such a way that their bases lie on a linear scale representing different intervals (bin width), and their heights are proportional to the frequencies of the values within each of the intervals.)
Hypothesis Testing	Statistical hypothesis testing is formalized as making a decision between rejecting or not rejecting a null hypothesis on the basis of a set of observations.
Independent	Two events A and B are (statistically) independent if the chance that they both happen simultaneously is the product of the chances that each occurs individually.
Independent Variable	In regression, the independent variable is the one that is supposed to explain the other; the term is a synonym for "explanatory variable".
Law of Large Numbers	The percentage of successes in repeated, independent trials with the same probability *p* of success in each trial, is increasingly likely to be close to the chance of success as the number of trials increases.
Mean	The sum of a list of numbers divided by the number of numbers
Median	The "middle value" of a list of numbers after the numbers have been ordered or the average of the two numbers in the middle.
Mode	A most common (frequent) value (A list can have more than one mode. For histograms, a mode is a relative maximum ("bump").)
Multiplication Rule	The chance that events A and B both occur is the conditional probability that A occurs given that B occurs, times the unconditional probability that B occurs.
Mutually Exclusive	In two events, if the occurrence of one is incompatible with the occurrence of the other, (they can't both happen at once) and two sets are mutually exclusive if they have no element in common
Normal Distribution	A random variable X has a normal distribution with mean *m* and standard error *s* if for every pair of numbers P(a < (X-m)/s < b) = area under the normal curve between a and b.
Observational Study	Where the researcher observes the experimental units in their natural setting and records the variable(s) of interest
Parameter	A numerical property of a population, such as its mean
Permutation	An arrangement of the elements of a set in some order
Population	A collection of units being studied
Quartiles	Values that partition the data set into four groups, each containing 25% of the measurements
Principle of Least Squares	The least squares line is one with the properties of the sum of the errors equals 0 and the sum of squared

	errors is smaller than that for any other straight-line model.
Random Sample	A sample whose members are chosen at random from a given population in such a way that the chance of obtaining any particular sample can be computed
Range	The largest value in the set minus the smallest value in the set (As a statistical the range is a single number, not a range of numbers.)
Relative Frequency	The proportion of one variable compared to the total term
Sample	A collection of units from a population
Sample Survey	A survey based on the responses of a sample of individuals, rather than the entire population
Significance, Significance level, Statistical significance.	The significance level of an hypothesis test is the chance that the test erroneously r rejects the null hypothesis when the null hypothesis is true.
Standard Deviation	In a set of numbers, the Root-Mean-Square of the set of deviations between each element of the set and the mean of the set
Standardize	To transform into standard units
Statistic	A number that can be computed from data involving no unknown parameters
Stratified Sampling	The act of drawing subsets of sampling units from different strata rather than from the frame as a whole is stratified sampling.
T-test	A hypothesis test based on approximating the probability histogram of the test statistic by a student's t-curve
Univariate	Having or having to do with a single variable
Variable	A numerical value or a characteristic that can differ from individual to individual
Variance	The square of the standard deviation of a list (the average of the squares of the deviations of the numbers in the list from their mean)
Venn Diagram	A pictorial way of showing the relations among sets or events
Weighted Mean	The pooled variance is simply a weighted average of the two sample variances.