STATISTICS CURRICULUM

INTRODUCTION

Course Description: The course is a study of descriptive statistics (how to properly interpret and represent categorical and quantitative data), sampling and experimentation (principles of experimentation design and random selection), anticipating patterns using the rules of probability (includes the normal distribution curve) and using statistical inferences to justify conclusions. This course is designed to teach students to become critical consumers of data and to prepare them for post-secondary courses in statistics.

The curriculum is designed for a full-block schedule with 90 class periods in the school year. It is typical that a few class periods may be missed due to scheduling conflicts, therefore pacing is written for a school year of 88 days, with approximately 22 per quarter.

Students having completed Algebra 1 and Geometry are encouraged to take this course as an elective to complement their math program. The curriculum closely follows the outline and sequence of the supporting textbook, *Stats in Your World, 2nd edition*. Additional instructional resources may include: CollegeBoard PSAT and SAT release problems and preparation exercises, Khan Academy – Topics in Introductory Statistics, US Census Bureau, Gallup Poll, Nielsen Ratings, regional and national newspapers, other web sources.

Units of Study:

Unit Title	Textbook Chapters
Unit ONE: Exploring and Understanding Univariate Data	1 through 5
Unit TWO: Exploring Relationships Between Variables	6 through 8
Unit THREE: Gathering Data	9 through 11
Unit FOUR: Probability and Randomness	12 through 15
Unit FIVE: Statistical Inference	16 through 20

Alignment with Common Core Standards: This course meets or exceeds all common core standards for statistics in the K-12 continuum. The early course material revisits many of the grade 6-8 standards for learning (graphical representation, arithmetic mean, etc.) but in a more robust fashion. These standards are purposely not referenced in the final curriculum document although they are a significant part of the curriculum. Only the common core standards in the high school curriculum are referenced in the final curriculum document.

STATISTICS CURRICULUM

INTRODUCTION

Alignment with 21st Century Skills: Outcomes for 21st century skills in math include *Information, Communication and Technology Literacy*: Students use digital technologies to manage, integrate, evaluate and create information, and to apply technology effectively, using it as a tool to research, organize, evaluate and communicate.

Post-secondary Majors requiring a Statistics Course:

Agricultural Business
Agricultural Economics
Agronomy and Crop Science
Applied Mathematics
Chemical Engineering
Civil Engineering
Computer forensics
Fishing and fisheries
Food and Nutrition Studies
Forensic Science
Health care administration
Health information management
Human Resources Management
Liberal Art and Sciences
Neuroscience
Occupational therapy
Physical therapy
Pre-veterinary
Psychology
Real Estate
Software Engineering
Special Education
Statistics
Surveying technology
Wildlife and Wilderness management

Approved June 2016

Statistics
Unit 1: Exploring and Understanding Univariate Data

Pacing: 6 weeks – Block Schedule

Mathematical Practices
Mathematical Practices #1 and #3 describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. *Practices in bold are to be emphasized in the unit.* *1. Make sense of problems and persevere in solving them.* **2. Reason abstractly and quantitatively.** ***3. Construct viable arguments and critique the reasoning of others.*** **4. Model with mathematics.** **5. Use appropriate tools strategically.** **6. Attend to precision.** 7. Look for and make use of structure. 8. Look for and express regularity in repeated reasoning.

Essential Questions
How can technology display and create graphical representations? • How are univariate data best represented for different types of data and various distributions? • Which calculations provide the most appropriate characterization of a distribution? • How can the properties of a normal distribution be used to analyze a data set?

Statistics
Unit 1: Exploring and Understanding Univariate Data

Corresponding Big Ideas
A system of equations is an algebraic way to compare two or more functions that model a situation. Matrices can be used to facilitate the solving of systems involving two- and three-variables.

Vocabulary
Population, sample, variable, quantitative variable, categorical variable, relative frequency, contingency table, marginal distribution, conditional distribution, histogram, symmetrical data, skewed data, outlier, mean, median, interquartile range, standard deviation, Normal model, z-score.

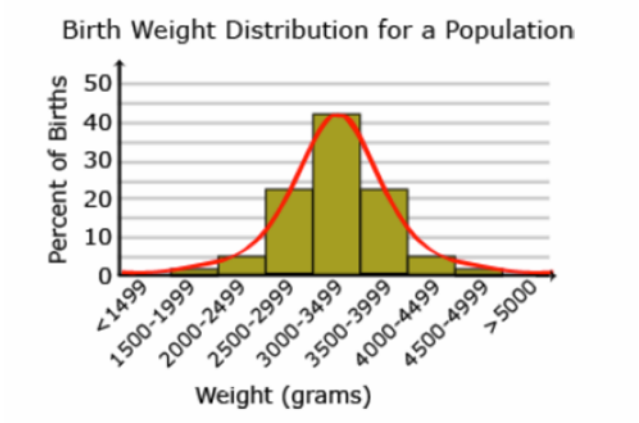
Standards Overview
Describe patterns of data in one variable. Differentiate between categorical and quantitative data and methods used to represent this data. Organizing sets of data into graphs and calculating numerical summaries. Critically analyzing data by describing, interpreting, and comparing important features. Generate conjectures about the variable.

Unit 1: Exploring and Understanding Univariate Data

Priority and Supporting **CCSS**	**Explanations and Examples**
HSS.ID.A.1 – Represent data with plots on the real number line (dot plots, histograms, and box plots).	Students will learn how to create graphs (pie charts, bar graphs, dot plots, histograms, box and whisker plots) that are appropriate for the data. Students will use the graphing calculator for quantitative data and spreadsheet software for qualitative data.
HSS.ID.A.2 – Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.	Students may use spreadsheets, graphing calculators and statistical software for calculations, summaries, and comparisons of data sets. Students will learn how to best describe central tendencies and spread of data for different distributions. Students will learn methods to compare sets of data, for example, by using side-by-side box-and-whisker plots, side-by-side histograms, by comparing 5-number summaries or by comparing mean and standard deviation. Example: Given a set of test scores: 99, 96, 94, 93, 90, 88, 86, 77, 70, 68, find the mean, median and standard deviation. Explain how the values vary about the mean and median. What information does this give the teacher?

Unit 1: Exploring and Understanding Univariate Data

HSS.ID.A.3 – Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).	Example: After the 2009-2010 NBA season LeBron James switched teams from the Cleveland Cavaliers to the Miami Heat, and he remained the top scorer (in points per game) in his first year in Miami. Compare team statistics for Cleveland (2009-2010) and Miami (2010-2011) for all players who averaged at least 10 minutes per game. Using the 1.5 X IQR rule, determine for which team and year James's performance may be considered an outlier.
HSS.ID.A.4 – Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.	Students may use spreadsheets, graphing calculators, statistical software and tables to analyze the fit between a data set and normal distributions and estimate areas under the curve. Students will learn to recognize distributions that are not normally distributed and develop strategies to analyze this data. Examples:



- The bar graph above gives the birth weight of a population of 100 chimpanzees. The line shows how the weights are normally distributed about the mean, 3250 grams. Estimate the percent of baby chimps weighing 3000-3999 grams.
- Determine which situation(s) is best modeled by a normal distribution. Explain your reasoning.
  - Annual income of a household in the U.S.
  - Weight of babies born in one year in the U.S.

Unit 1: Exploring and Understanding Univariate Data

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Measures of central tendency: mean, median and mode	• Calculate • Interpret • Solve (data analysis, use of equations)
• Measures of dispersion: interquartile range, range, standard deviation	• Calculate • Interpret • Solve (data analysis, use of equations)
• Effect of an outlier on a data set	• Define • Interpret
• Weighted mean	• Calculate • Interpret (data analysis)
• Graphic representation of data	• Create
• Graphs and charts (pie, frequency, histogram, bar and double bar, stem-and-leaf, box-plots)	• Classify (Categorical, Quantitative) • Interpret • Identify
• Percentiles and quartiles (mean, median, mode, interquartile range)	• Calculate (mean, median, mode, interquartile range) • Interpret

Unit 1: Exploring and Understanding Univariate Data

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Qualitative and quantitative data	• Identify (distinguish between) • Use (statistics to infer)
• Standard deviation Thumb rules	• Define • Understand • Calculate (with technology) • Use to interpret
• Normally distributed data	• Identify • Understand • Interpret
Outliers Population percentages Area under normal curve	• Identify • Estimate • Calculate (with technology) • Analyze

Standardized Assessment Correlations (State, College and Career)
College Board PSAT, SAT and ACT

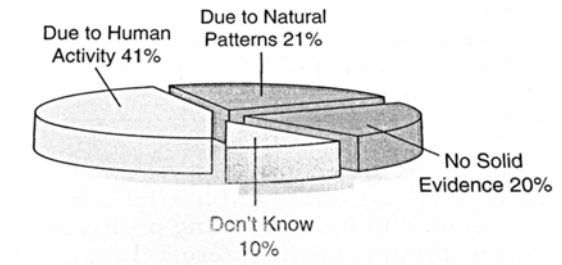
Unit 1: Exploring and Understanding Univariate Data

Learning Activities		
Topics	**Reference** **Stats in Your World Text**	**CCS**
Define Data – 5 W's	**Chapter 1** www.learner.org/courses/againstallodds/ www.census.gov/data.html	HSS.ID.A.1
Categorical Data Create a relative frequency table Analyze data displays	**Chapter 2** www.learner.org/courses/againstallodds/	
Calculate marginal distributions from a contingency table		HSS.ID.A.1
Calculate conditional distribution from a contingency table		HSS.ID.A.2
Determine whether variables are dependent/independent		HSS.ID.A.3
Quantitative Data Create histogram by hand and with calculator Discuss shape, center, spread Calculate range, median, quartiles and interquartile range Calculate population percentages based on quartile info	**Chapter 3** lib.stat.cmu.edu/DASL education.ti.com/en/84activitycentral	HSS.ID.A.4
Compare Quantitative Data of Groups Mean, Median, Standard Deviation, IQR (Interquartile range)	**Chapter 4** www.seanlahman.com/baseball-archive/statistics	HSS.ID.A.4

Unit 1: Exploring and Understanding Univariate Data

Learning Activities		
Topics	**Reference** **Stats in Your World Text**	**CCS**
Boxplots Determine most appropriate measure based on data info Timeplots Shift and Rescale Data Normal Distribution Determine whether a data set is normal Z-Scores Calculate Use scores to compare data Normal Model 68-95-99.7 model Sketch with deviations identified Calculate percentages of a population Calculate z-scores for a given population percentage	**Chapter 5** education.ti.com/en/84activitycentral www.ncssdm.edu/courses/math/Talks/index.htm	HSS.ID.A.4 HSS.ID.A.4

Unit Assessments
Section quizzes, Project(s) – see appendix, Applications, End-of-Unit Test

Applications
The Pew Research Center for the People and the Press (http://people-press.org) has asked a representative sample of U.S. adults about global warming, repeating the question over time. In January, 2007 the responses reflected an increased belief that global warming is real and due to human activity. Below is a display of the percentages of respondents choosing each of the major alternatives offered, list two errors in the display. 
A town's January high temperatures average 36°F with a standard deviation of 8°, while in July the mean high temperature is 75°F and the standard deviation is 10°. In which month is it more unusual to have a day with a high temperature of 55°? Show work and explain your answer.

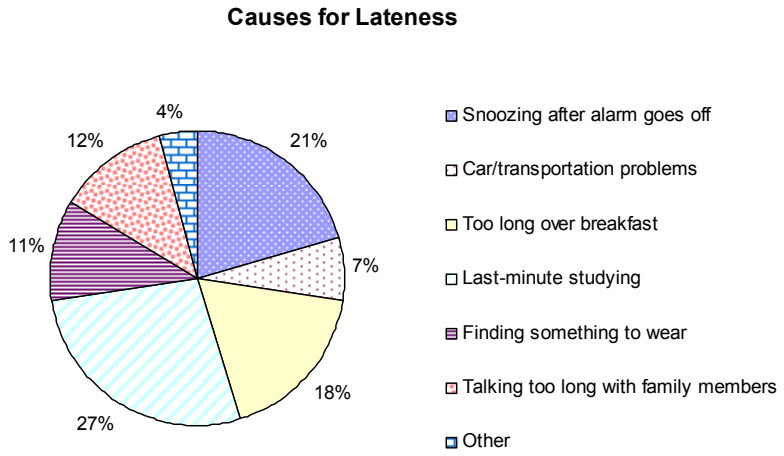
Applications
Companies who design furniture for elementary school classrooms produce a variety of sizes for kids of different ages. Suppose the heights of kindergarten children can be described by a Normal model with a mean of 39.2 inches and a standard deviation of 1.9 inches. a. Calculate the z-score for a child's height of 42 inches. b. Calculate the z-score for a child's height of 35 inches. c. Use your calculator to determine the percent of children whose heights fall 35 and 42 inches. Write down the calculator "operation" as your work. d. What percent of kindergarten children should the company expect to be less than 35 inches tall? Again use your calculator and write down the "operation" as your work. e. Use your calculator and find the z-score for the lowest 20% of the children's heights. Find the height that corresponds to this z-score, then sketch and shade in the normal curve showing this 20%. Clearly label your graph. f. Use your calculator and find the z-score for the highest 4% of the children's heights. Find the height that corresponds to this z-score, then sketch and shade in the normal curve showing this 4%. Clearly label your graph.

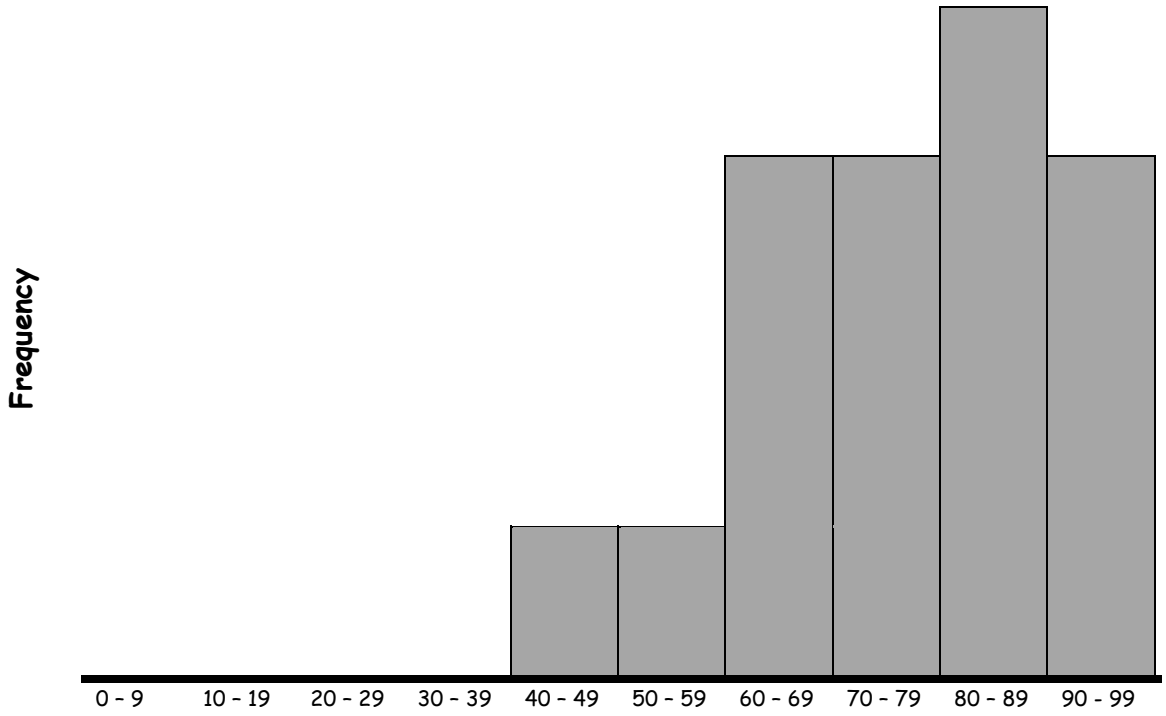
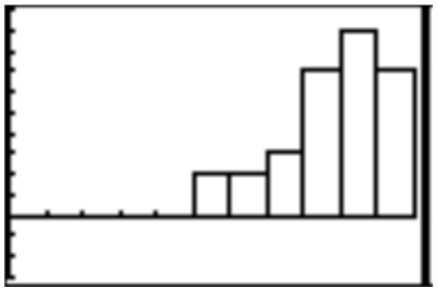
Applications

Obtain a set of categorical data and create an appropriate chart using MS Excel. An example of a suitable activity is to use the following data to create a pie chart in MS Excel.

Causes for Lateness

Cause	Frequency
Snoozing after alarm goes off	15
Car/transportation problems	5
Too long over breakfast	13
Last-minute studying	20
Finding something to wear	8
Talking too long with family members	9
Other	3



Applications
Students will gather, or be provided, a set of quantitative data, similar to that shown below in the example, and graph the data using appropriate means (histogram, dot plot, etc.) and use the TI-84 graphing calculator (shown below) to analyze the data in more detail. Example. Shown below are the final exam grades for students in a recent Algebra 2 class. 72, 81, 42, 87, 99, 67, 84, 68, 86, 53, 91, 98, 75, 96, 70, 94, 83, 93, 89, 40, 80, 72, 68, 88, 53, 72, 72, 70, 89, 99      

Unit 1: Exploring and Understanding Univariate Data

Applications

Explore a variety of methods to compare two similar data sets. Methods include side-by-side stem-and-leaf plots, side-by-side histograms, and side-by-side box-and-whisker plots. Measures of central tendency and spread will be compared to better understand the two data sets.

Example. Here are the numbers of home runs that Babe Ruth hit in his 15 years with the New York Yankees, 1920-1934:

54 59 35 41 46 25 47 60 54 46 49 46 41 34 22

Babe Ruth's home run record for a single season was broken by another Yankee, Roger Maris, who hit 61 home runs in 1961. Here are Maris' home run totals for his 10 years in the American League:

13 23 26 16 33 61 28 39 14 8

Babe Ruth		Roger Maris
	0	8
	1	3 4
5 2	2	3 6 8
5 4	3	3 9
9 7 6 6 6 1 1	4	
9 4 4	5	
0	6	1

Approved June 2016

Statistics
Unit 2: Exploring Relationships Between Variables

Pacing: 3 weeks – Block Schedule

Mathematical Practices
Mathematical Practices #1 and #3 describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. *Practices in bold are to be emphasized in the unit.* ***1. Make sense of problems and persevere in solving them.*** **2. Reason abstractly and quantitatively.** ***3. Construct viable arguments and critique the reasoning of others.*** **4. Model with mathematics.** 5. Use appropriate tools strategically. **6. Attend to precision.** 7. Look for and make use of structure. 8. Look for and express regularity in repeated reasoning.

Essential Question
How can the properties of data be communicated to illuminate its important features? • Which are the best types of displays to use for comparing data sets? • What determines sets of data as Normal models?

Approved June 2016

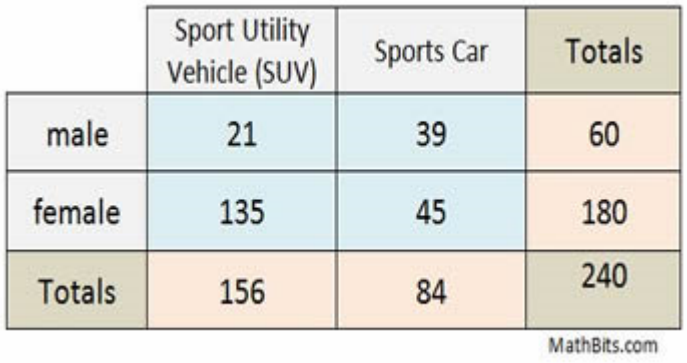
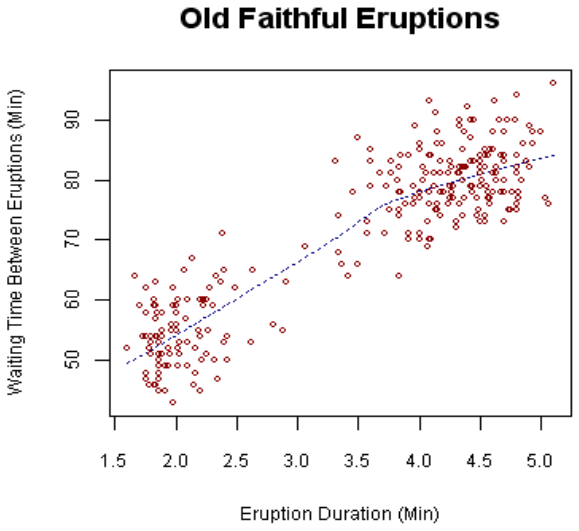
Statistics
Unit 2: Exploring Relationships Between Variables

Corresponding Big Ideas
Statisticians summarize, represent and interpret categorical and quantitative data in multiple ways since one method can reveal or create a different impression than another.

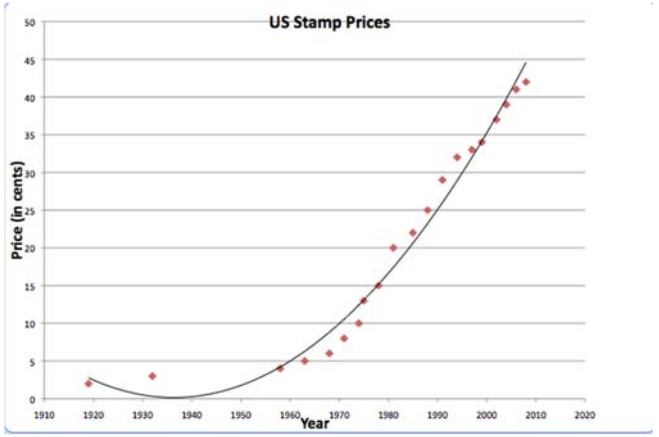
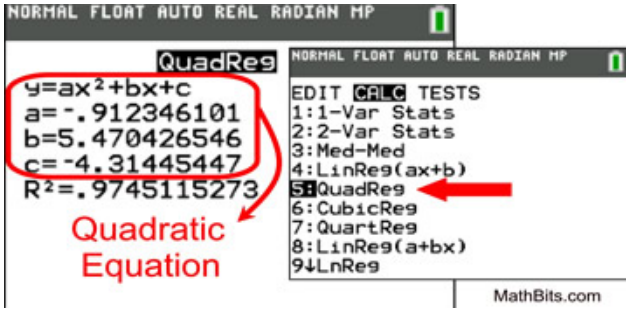
Vocabulary
Data, two-way frequency table, contingency table, percentages, ratios, relative frequencies, joint relative frequency, marginal relative frequency, conditional relative frequency, patterns, associations, variables, scatter plot, quantitative variable, **independent variable, dependent variable, scale, direction, form, strength, outliers, linear, quadratic, exponential, models, data set, function of best fit, line of best fit, residuals, y = mx + b, slope, y-intercept, correlation coefficient, significance, correlation, causation**

Standards Overview
Describe patterns of categorical data in two variables and recognize possible associations and trends. Organize sets of data into scatter plot graphs and describe how the variables are related. Fit functions to data sets and use to solve problems in the context of the data. (Linear, Quadratic, and Exponential Models) Distinguish between correlation and causation.

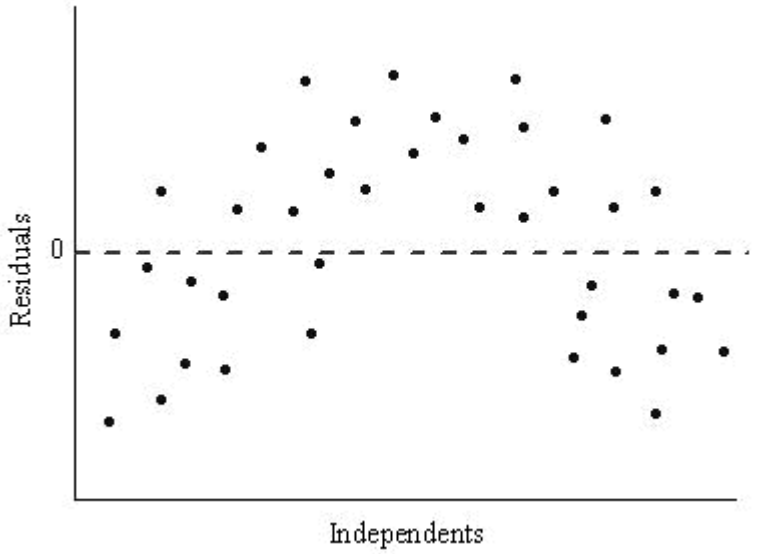
Statistics
Unit 2: Exploring Relationships Between Variables

Priority and Supporting **CCSS**	**Explanations and Examples**
HSS.ID.B.5 – Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in data.	
HSS.ID.B.6 – Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.	

Statistics
Unit 2: Exploring Relationships Between Variables

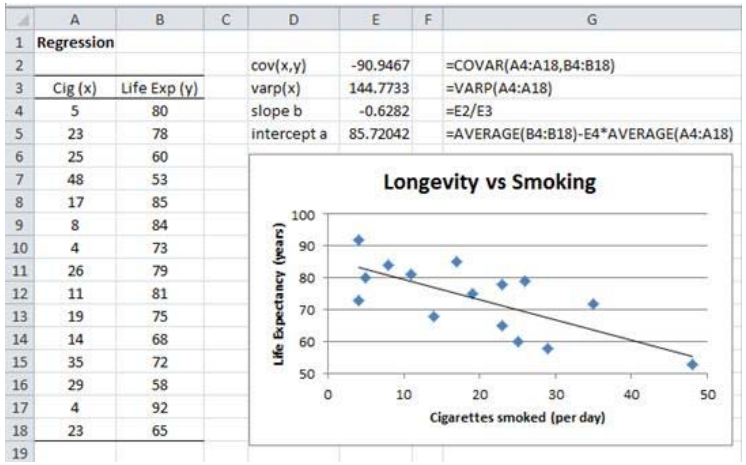
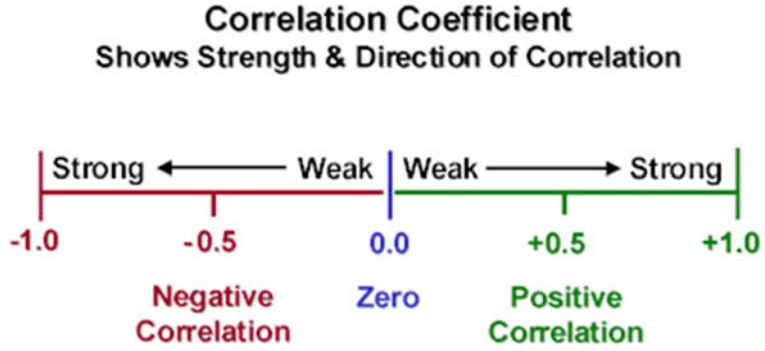
	Direction (positive or negative), Form (straight, curved, no pattern), Strength (tightness of cluster), Unusual features (outliers or subgroups) 
HSS.ID.B.6.A – Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.	

Statistics
Unit 2: Exploring Relationships Between Variables

HSS.ID.B.6.B – Informally assess the fit of a function by plotting and analyzing residuals.	 Residual = Data - Model

Approved June 2016

Statistics
Unit 2: Exploring Relationships Between Variables

HSS.ID.B.6.C - Fit a linear function for a scatter plot that suggests a linear association.	
HSS.ID.C.7 – Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.	
HSS.ID.C.8 – Compute (using technology) and interpret the correlation coefficient of a linear fit.	

	A	B	C	D	E	F	G
1	Regression						
2				cov(x,y)	-90.9467		=COVAR(A4:A18,B4:B18)
3	Cig (x)	Life Exp (y)		varp(x)	144.7733		=VARP(A4:A18)
4	5	80		slope b	-0.6282		=E2/E3
5	23	78		intercept a	85.72042		=AVERAGE(B4:B18)-E4*AVERAGE(A4:A18)
6	25	60					
7	48	53					
8	17	85					
9	8	84					
10	4	73					
11	26	79					
12	11	81					
13	19	75					
14	14	68					
15	35	72					
16	29	58					
17	4	92					
18	23	65					
19							

Statistics
Unit 2: Exploring Relationships Between Variables

HSS.ID.C.9 – Distinguish between correlation and causation.	Two variables may have a high correlation without being related or connected. For example…You might find a strong correlation between depth and urchin density at Onekahakaha when possibly there is little true causation (cause-effect relationship). In actuality the relationship is probably driven by salinity being very low in shallow, nearshore waters and higher in deeper waters further from the freshwater outflow.

Statistics
Unit 2: Exploring Relationships Between Variables

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Data displayed in a two-way frequency table	• Read • Interpret • Write summaries
• Ratios/percentages in a two-way frequency table	• Calculate
• Joint, Marginal, Conditional relative frequencies	• Calculate • Interpret • Explain meaning • Create displays
• Patterns observed in the data	• Describe
• Conditional and Marginal percentages	• Compare • Analyze
• Independent and Dependent Variables	• Identify
• Scatter Plot	• Construct
• Outliers	• Identify

Statistics
Unit 2: Exploring Relationships Between Variables

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Models to represent a data set (linear, quadratic, exponential) • Residuals	• Predict effect • Determine • Sketch • Predict • Compute • Plot
• Slope & intercept of linear model	• Analyze • Interpret
• Correlation Coefficient of linear fit	• Compute (using technology) • Interpret
• Correlation vs. Causation	• Distinguish

Approved June 2016

Statistics
Unit 2: Exploring Relationships Between Variables

Standardized Assessment Correlations (State, College and Career)
College Board PSAT, SAT and ACT

Learning Activities		
Topics	**Reference** **Stats in Your World Text**	**CCS**
Two-way Frequency Tables (Contingency Tables) Marginal Distribution Conditional Distributions Associations vs. Independence	**Chapter 3** lib.stat.cmu.edu/DASL education.ti.com/en/84activitycentral	HSS.ID.B.5
Scatterplots Describe direction, form, strength, outliers Define variables Create using technology	**Chapter 6** www/learner/org/courses/againstallodds/ www.stat.sc.edu/~west/javahtml/Regression.html	HSS.ID.B.6
Compute Correlation coefficient	education.ti.com/en/84activitycentral	HSS.ID.C.8

Statistics
Unit 2: Exploring Relationships Between Variables

Correlation Conditions Correlation vs. Causation Linear Models Predicted value Residuals Interpret Slope and Intercept Least Squares Line (Line of Best Fit/Regression Line) Lurking Variable and Causation Exponential and Power (ex. Quadratic) Models Create using technology Assess the model fit Models Justify choice of model Describe and interpret model Use model to make predictions	**Chapter 7** Illuminations.nctm.org Sambaker.com/courses/J716/demos/LeastSquares/LeastSquaresDemo.html **Chapter 8** education.ti.com/en/84activitycentral	HSS.ID.C.9 HSS.ID.B.6.B HSS.ID.C.7 HSS.ID.B.6.C HSS.ID.C.9 HHS.ID.B.6.A HHS.ID.B.6.B HSS.ID.B.6.A

Statistics
Unit 2: Exploring Relationships Between Variables

Unit Assessments
Section quizzes, Project(s) – see appendix, Applications, End-of-Unit Test

Statistics
Unit 2: Exploring Relationships Between Variables

Applications

1.) Use the contingency table below to answer the following questions.

	Height			
	Short	Medium	Tall	Total
Left Handed	3	5	2	10
Right Handed	19	16	15	50
Total	22	21	17	60

a. Find the marginal distribution of height.

b. What percent of the right-handed respondents are short?

c. What percent of the respondents are short?

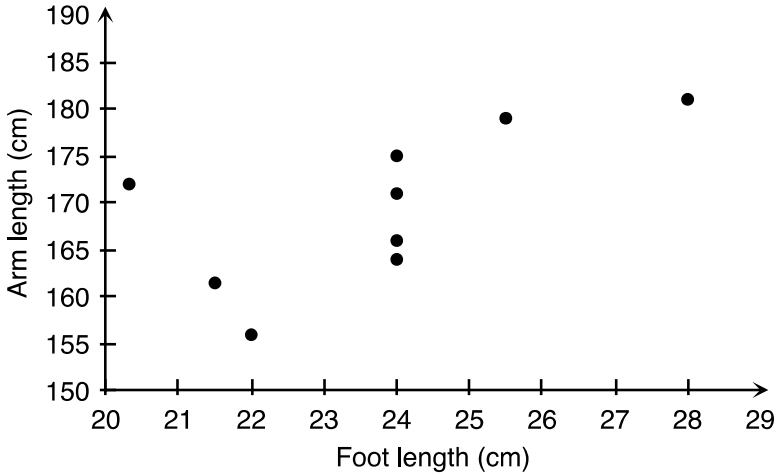
d. What percent of the short respondents are right-handed?

e. Determine the conditional distribution of left- or right-handedness for height.

Inference question: Is there evidence to suggest that height and left- or right-handedness are dependent?

Statistics
Unit 2: Exploring Relationships Between Variables

2.) The table and scatterplot below display the arm length and foot length data for a random sample of nine high school students in a local high school.



Foot length (cm)	Arm length (cm)
24	164
24	166
24	171
25.5	179
24	175
22	156
21.5	161.5
28	181
20.32	172

1. Summarize the sample data by calculating the correlation coefficient and the slope and *y*-intercept of the least-squares regression line. Round your answers to two decimal places.
2. Interpret the sample statistics. What do they tell about the relationship between these two variables?

Inference question: Does the sample provide evidence that a linear relationship exists between foot length and arm length for *all* students in the high school?

Statistics
Unit 2: Exploring Relationships Between Variables

Exponential Model

3.) The data below shows the cooling temperatures of a freshly brewed cup of coffee after it is poured from the brewing pot into a serving cup. The brewing pot temperature is approximately 180º F.

Task: a.) Determine an exponential regression model equation to represent this data.

b.) Graph the new equation.

c.) Decide whether the new equation is a "good fit" to represent this data.

d.) Based upon the new equation, what was the initial temperature of the coffee?

e.) Interpolate data: When is the coffee at a temperature of 106 degrees?

f.) Extrapolate data: What is the predicted temperature of the coffee after 1 hour?

g.) In 1992, a woman sued McDonald's for serving coffee at a temperature of 180º that caused her to be severely burned when the coffee spilled. An expert witness at the trial testified that liquids at 180º will cause a full thickness burn to human skin in two to seven seconds. It was stated that had the coffee been served at 155º, the liquid would have cooled and avoided the serious burns. The woman was awarded over 2.7 million

dollars. As a result of this famous case, many restaurants now serve coffee at a temperature around 155º. How long should restaurants wait (after pouring the coffee from the pot) before serving coffee, to ensure that the coffee is not hotter than 155º ?

h.) If the temperature in the room is 76° F, what will happen to the temperature of the coffee, after being poured from the pot, over an extended period of time?

Time (mins)	Temp (º F)
0	179.5
5	168.7
8	158.1
11	149.2
15	141.7
18	134.6
22	125.4
25	123.5
30	116.3
34	113.2
38	109.1
42	105.7
45	102.2
50	100.5

Approved June 2016

Statistics
Unit 3: Gathering Data

Pacing: 3 weeks – Block Schedule

Mathematical Practices
Mathematical Practices #1 and #3 describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. *Practices in bold are to be emphasized in the unit.* ***1. Make sense of problems and persevere in solving them.*** 2. Reason abstractly and quantitatively. ***3. Construct viable arguments and critique the reasoning of others.*** **4. Model with mathematics.** **5. Use appropriate tools strategically.** **6. Attend to precision.** 7. Look for and make use of structure. 8. Look for and express regularity in repeated reasoning.

Essential Question
How can a population be accurately described when it is very large? • How can bias be avoided in sampling a larger population?

Approved June 2016

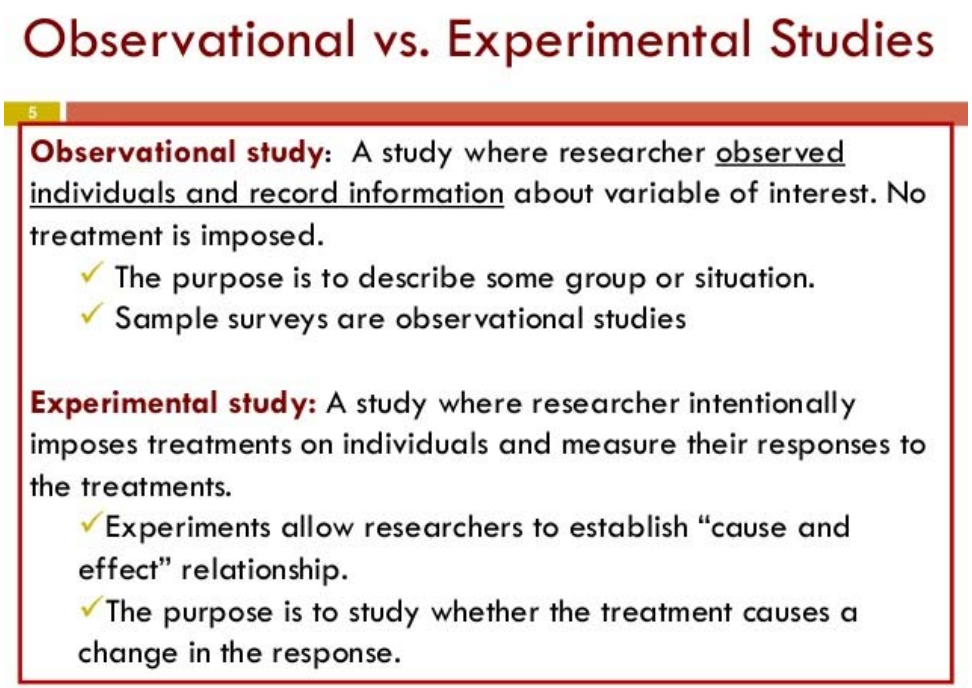
Statistics
Unit 3: Gathering Data

Corresponding Big Ideas
Statisticians design experiments based on random samples and analyze the data to estimate the important properties of a population and make informed judgments.

Vocabulary
Population, sample, **sample survey, biased, randomizing, census, population parameter, representative, simple random sample, sampling frame, sampling variability, stratified random sampling, cluster sampling, multistage samples, systematic sampling, pilot study, voluntary response bias, convenience sample, undercoverage, nonresponse bias, response bias, observational study, experiment, factor, principles of experimental design, control group, blinding, placebo, blocking, matching, confounding, simulation, trial, response variable,** statistical significance

Standards Overview
Make inferences and justify conclusions from sample surveys, experiments, and observational studies.

Statistics
Unit 3: Gathering Data

Priority and Supporting **CCSS**	**Explanations and Examples**
HSS.IC.B.3 - Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.	

Statistics
Unit 3: Gathering Data

Bias in Sampling

- The design of a study is *biased* if it systematically *favors* certain outcomes. *There has to be randomness*!
  - A **voluntary response** sample is biased in that it favors negative outcomes regardless of the question.
  - A **convenience sample** is usually biased in that it favors the opinions of people in a certain location at a certain time.
    - There is no guarantee that such opinion is representative of the population as a whole.
  - In both cases a conscious *choice* is made to include/exclude a respondent
    - We want a method in which the choice is random and does not depend on any individual
    - This is what we strive for as scientists, phycologists, and statisticians.

Statistics
Unit 3: Gathering Data

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Sample types	• Identify • Interpret
• Bias	• Identify • Interpret • Evaluate
• Randomization	• Identify • Evaluate
• Surveys (survey claims)	• Evaluate • Design • Summarize
• Observational Studies	• Understand
• Experiments – Randomized and Comparative	• Understand • Design • Summarize

Statistics
Unit 3: Gathering Data

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Control Treatments Blinding Placebos Blocking Matching Confounding	• Identify • Understand
• Random selection • Random assignment	• Create (using technology) • Create (using technology)
• Simulation	• Design • Conduct • Summarize results
• Statistical Significance	• Assess

Statistics
Unit 3: Gathering Data

Standardized Assessment Correlations (State, College and Career)
College Board PSAT, SAT and ACT

Learning Activities		
Topics	**Reference** **Stats in Your World Text**	**CCS**
Representative Samples Simple Random Sample (SRS) Stratified Samples Cluster Samples Systematic Samples Bias Nonresponse Response Sampling method Surveys	**Chapter 9** www.learner.org/courses/aginstallodds/ www.gallup.com www.nielsen.com	HSS.IC.B.3

Statistics
Unit 3: Gathering Data

Report methods Evaluate fairness & accuracy Observational Studies Retrospective study Prospective study Experiments Random assignment Experimental Design Experiments vs. Samples Control Treatments Placebos Random Selection and Assignment Simulation Components Outcomes Trial Response variable Statistical Significance	**Chapter 10** www.learner.org/courses/againstallodds/ **Chapter 11** www.random.org www.randomnumbers.info	HSS.IC.B.3 HSS.IC.B.3

Statistics
Unit 3: Gathering Data

Unit Assessments
Section quizzes, Project(s) – see appendix, Applications, End-of-Unit Test

Applications
Work with Ledyard Police department to gather and organize data about vehicle infractions – speeding/accidents. Categorize the types of data that are collected using sample vocabulary from the unit. Discuss the anticipated statistical significance of the information gathered and develop a plan for analyses. Design a safety campaign to address areas of concern determined by results of analyses.

Approved June 2016

Statistics
Unit 4: Probability and Randomness

Pacing: 6 weeks – Block Schedule

Mathematical Practices
Mathematical Practices #1 and #3 describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. *Practices in bold are to be emphasized in the unit.* ***1. Make sense of problems and persevere in solving them.*** **2. Reason abstractly and quantitatively.** ***3. Construct viable arguments and critique the reasoning of others.*** **4. Model with mathematics.** 5. Use appropriate tools strategically. **6. Attend to precision.** **7. Look for and make use of structure.** **8. Look for and express regularity in repeated reasoning.**

Essential Question
In what ways does one event impact the probability of another event occurring? • How is probability used to make informed decisions about uncertain events?

Approved June 2016

Statistics
Unit 4: Probability and Randomness

Corresponding Big Ideas
Probability provides a process to determine the likelihood of events and determine whether the occurrence of one event makes some other result more or less likely. The rules of probability can lead to more valid and reliable predictions about the likelihood of an event occurring.

Vocabulary
Event, sample space, subset, union, intersection, complement, independent events, dependent events, probability, conditional probability, product, two-way frequency table display, data, variable, **category, random sample, formula, Addition Rule, General Multiplication Rule, fundamental counting principle, outcomes, factorial, permutation, combination, compound event, expected value, mean, random variable, probability distribution, theoretical probability, simulation**

Standards Overview
Understand independence and conditional probability and use them to interpret data. Calculate expected values and use them to solve problems.

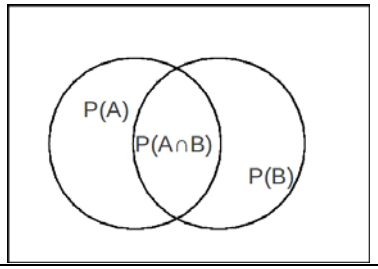
Statistics
Unit 4: Probability and Randomness

Priority and Supporting **CCSS**	**Explanations and Examples**
HSS.CP.A.1 Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events ("or," "and," "not").	
HSS.CP.A.2 Understand that two events *A* and *B* are independent if the probability of *A* and *B* occurring together is the product of their probabilities. Use this characterization to determine if they are independent.	*Note:* When two events, A and B, are independent, the probability of both occurring is P(A and B) = $P(A) \cdot P(B)$
HSS.CP.A.3 Understand the conditional probability of *A* given *B* as *P*(*A* and *B*)/*P*(*B*), and interpret independence of *A* and *B* as saying that the conditional probability of *A* given *B* is the same as the probability of *A*, and the conditional probability of *B* given *A* is the same as the probability of *B*.	The conditional probability of an event B, in relation to event A, is the probability that event B will occur given the knowledge that an event A has already occurred. Example: You toss two pennies. The first penny shows HEADS and the other penny rolls under the table and you cannot see it. What is the probability that they are both heads? Since you already know that one is HEADS, the probability of getting HEADS on the second penny is 1 out of 2. Notation for conditional probability: *P*(*B*|*A*) *read ... the probability of B given A*.

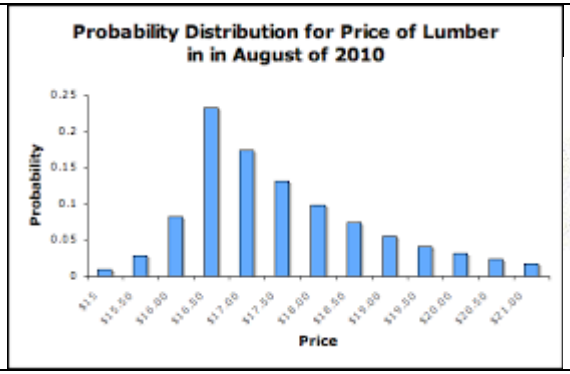
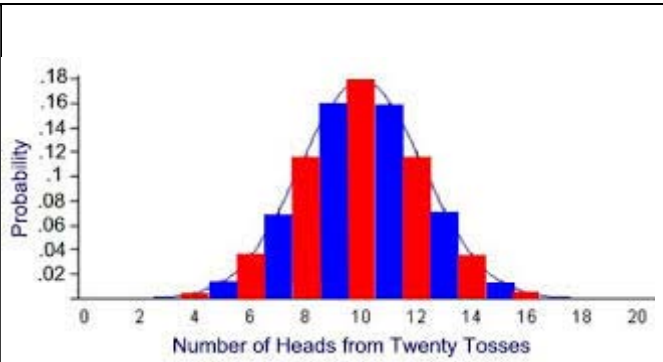
Statistics
Unit 4: Probability and Randomness

HSS.CP.A.4 Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities.	Example: Collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.
HSS.CP.A.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations.	Example: Compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.
HSS.CP.B.6 Find the conditional probability of *A* given *B* as the fraction of *B*'s outcomes that also belong to *A*, and interpret the answer in terms of the model.	Example: A jar contains black and white marbles. Two marbles are chosen without replacement. The probability of selecting a black marble and then a white marble is 0.34, and the probability of selecting a black marble on the first draw is 0.47. What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black? Solution: $P(\text{White}\mid\text{Black}) = \frac{P(\text{Black and White})}{P(\text{Black})} = \frac{0.34}{0.47} = 0.72 = 72\%$
HSS.CP.B.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answer in terms of the model.	Example: You are going to roll two dice. Find P(sum that is even or sum that is a multiple of 3). The addition rule says we need to find P(even) + P(multiple of 3) - P(both).

Statistics
Unit 4: Probability and Randomness

	
HSS.CP.B.8 (+) Apply the general Multiplication Rule in a uniform probability model, P(A and B) = P(A)P(B|A) = P(B)P(A|B), and interpret the answer in terms of the model.	
HSS.CP.B.9 (+) Use permutations and combinations to compute probabilities of compound events and solve problems.	A permutation is the choice of *r* things from a set of *n* things without replacement and where the order matters. $${}_nP_r = \frac{n!}{(n-r)!}$$ A combination is the choice of *r* things from a set of *n* things without replacement and where order does **not** matter. *(Notice the two forms of notation.)* $${}_nC_r = \binom{n}{r} = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}$$ Example: There are 4 men and 5 women. Find the probability of selecting 3 of which (i) exactly two are women, (ii) no woman, (iii) at least one women, (iv) at most one women, (v) no men

Statistics
Unit 4: Probability and Randomness

HSS.MD.A.1 (+) Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions.	 
HSS.MD.A.2 (+) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution.	
HSS.MD.A.3 (+) Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value.	Example: Find the theoretical probability distribution for the number of correct answers obtained by guessing on all five questions of a multiple-choice test where each question has four choices, and find the expected grade under various grading schemes.
HSS.MD.A.4 (+) Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value.	Example: Find a current data distribution on the number of TV sets per household in the United States, and calculate the expected number of sets per household. How many TV sets would you expect to find in 100 randomly selected households?
HSS.MD.B.5 (+) Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding	

expected values.	
HSS.MD.B.5.A Find the expected payoff for a game of chance.	Example: Find the expected winnings from a state lottery ticket or a game at a fast-food restaurant.
HSS.MD.B.5.B Evaluate and compare strategies on the basis of expected values.	Example: Compare a high-deductible versus a low-deductible automobile insurance policy using various, but reasonable, chances of having a minor or a major accident.
HSS.MD.B.6 (+) Use probabilities to make fair decisions.	Draw by lots or use a random number generator
HSS.MD.B.7 (+) Analyze decisions and strategies using probability concepts	Example: Analyze product testing, medical testing, pulling a hockey goalie at the end of a game

Concepts **What Students Need to Know**	**Skills** **What Students Need To Be Able To Do**
• Law of Averages • Probability	• Interpret • Define • Calculate • Interpret

Statistics
Unit 4: Probability and Randomness

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Counting Principles	• Apply
• Permutations	• Define • Calculate • Use • Interpret
• Combinations	• Define • Calculate • Use • Interpret
• Probability Rules	• Apply • Interpret results • Calculate • Use
• Conditional Probability • Independence	• Define • Calculate • Interpret

Approved June 2016

Statistics
Unit 4: Probability and Randomness

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Probability Model	• Construct • Interpret
• Statistical Significance	• Define • Evaluate

Standardized Assessment Correlations (State, College and Career)
College Board PSAT, SAT and ACT

Approved June 2016

Statistics
Unit 4: Probability and Randomness

Learning Activities		
Topics	**Reference** **Stats in Your World Text**	**CCS**
Probability	**Chapter 12**	HSS.CP.A.1
Theoretical	www.learner.org/courses/againstallodds/	HSS.CP.A.2
Equally Likely Condition		
Counting Principles		HSS.CP.B.9
Permutations		HSS.MD.A.3
Factorials		HSS.MD.A.4
Combinations		
Combination and Probability		
Formal Probability	**Chapter 13**	
Complement Rule		
Addition Rule		HSS.CP.B.7
Multiplication Rule		HSS.CP.B.8
Probability Rules	**Chapter 14**	HSS.CP.B.7
Addition Rule	www.census.gov/data.html	HSS.CP.B.6

Statistics
Unit 4: Probability and Randomness

Venn Diagrams		HSS.CP.A.5
Conditional Probability		HSS.CP.A.4
Independence		HSS.CP.B.8
Contingency Tables		
Multiplication Rule		
Tree Diagrams		
Probability Models	**Chapter 15**	HSS.MD.A.1
Random variables	www.learner.org/courses/againstallodds/	HSS.MD.A.2
Expected Value		
Spread – standard deviation		HSS.MD.A.4
		HSS.MD.B.5.B
Binomial Probability/Binomial Model		
Normal model		
Statistical Significance		

Unit Assessments
Section quizzes, Project(s) – see appendix, Applications, End-of-Unit Test

Statistics
Unit 4: Probability and Randomness

Applications

1. **Mutual Fund Sales**. Suppose a mutual fund sales person has a 50% chance of closing a sale on each call she makes. Suppose further that she made four calls in the last hour. Consider "closing a sale" a success and "not closing a sale" a failure. Then, we will study the variables: X = total number of successes Y = number of successes before first failure. How would the distribution of Y vary for different values of X?

2. **Lottery**. Let X be a random variable, and let h(X) be a function of X; then, the expected value of h(X), written as E(h(X)), is defined by: E(h(X)) ≡ X all x h(x) P(x). In a lottery where the buyer of a ticket picks 6 numbers out of 50, X can be the number of matches out of the picked numbers and the actual payoff is a function of X.

3. **Managing Investments**. Managing risk is an important part of life. This is particularly true when we are assessing the desirability of an investment portfolio.

   Consider the following two investments:

	Investment #1	Investment #2
Mean Rate of Return	0.06	0.08
Standard Deviation	0.02	0.03

   Which would you choose?

Approved June 2016

Statistics
Unit 5: Statistical Inference

Pacing: 6 weeks – Block Schedule

Mathematical Practices
Mathematical Practices #1 and #3 describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. *Practices in bold are to be emphasized in the unit.* ***1. Make sense of problems and persevere in solving them.*** **2. Reason abstractly and quantitatively.** ***3. Construct viable arguments and critique the reasoning of others.*** **4. Model with mathematics.** **5. Use appropriate tools strategically.** **6. Attend to precision.** **7. Look for and make use of structure.** **8. Look for and express regularity in repeated reasoning.**

Essential Question
How can the data from a randomized experiment be used to compare two treatments? • What is the meaning of significant differences? • Why should reports on data be evaluated?

Approved June 2016

Statistics
Unit 5: Statistical Inference

Corresponding Big Ideas
Statisticians design experiments based on random samples and analyze the data to estimate the important properties of a population and make informed judgments.

Vocabulary
Sampling distribution model, sampling variability/sampling error, standard error, confidence interval, margin of error, critical value, null hypothesis, alternative hypothesis, two-tailed alternative, one-tailed alternative, z-test, statistically significant**, Type I error, Type II error, significance level, P-value, Central Limit Theorem, t-models, one-sample t-interval, one-sample t-test, paired data, paired t-test, paired-t confidence interval, Pythagorean Theorem of Statistics, two-proportion z-interval, two-proportion z-test, two-sample t-interval, two-sample t-test**

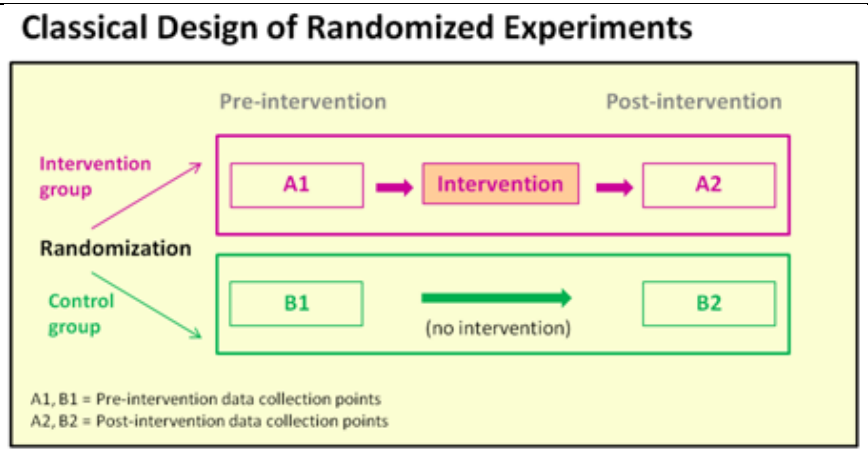
Standards Overview
Make inferences and justify conclusions from ample surveys, experiments, and observational studies Evaluate reports based on data.

Statistics
Unit 5: Statistical Inference

Priority and Supporting **CCSS**	**Explanations and Examples**
HSS.IC.A.1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population.	
HSS.IC.A.2 Decide if a specified model is consistent with results from a given data-generating process, such as using simulation.	Example: A model says a spinning coin falls heads up with probability 0.5. Would a result of 5 tails in a row cause you to question the model?
HSS.IC.B.4 Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.	Example: Suppose that the Gallup Organization's latest poll sampled 1,000 people from the United States, and the results show that 520 people (52%) think the president is doing a good job, compared to 48% who don't think so. First, assume you want a 95% level of confidence, so $z^* = 1.96$. The number of Americans in the sample who said they approve of the president was found to be 520. This means that the sample proportion, $\hat{p}$, is 520 / 1,000 = 0.52. The margin of error for this polling question is calculated in the following way: $z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{(0.52)(0.48)}{1{,}000}}$ $= (1.96)(0.0158) = 0.0310$ According to this data, you conclude with 95% confidence that 52% of all Americans approve of the president, plus or minus 3.1%.

Approved June 2016

Statistics
Unit 5: Statistical Inference

HSS.IC.B.5 Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.	
HSS.IC.B.6 Evaluate reports based on data.	

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Sampling Distribution Models	• Define • Create • Interpret
• Confidence Interval	• Create using technology • Interpret

Statistics
Unit 5: Statistical Inference

Concepts What Students Need to Know	Skills What Students Need To Be Able To Do
• Margin of Error	• Calculate • Evaluate
• Hypothesis Testing	• Create • Apply • State conclusion
• Errors	• Analyze • Describe
• Central Limit Theorem	• Apply
• Confidence Interval for Means	• Calculate • Interpret
• Paired Data	• Identify • Apply
• Expected Counts	• Find • Apply

Approved June 2016

Statistics
Unit 5: Statistical Inference

Standardized Assessment Correlations (State, College and Career)
College Board PSAT, SAT and ACT

Learning Activities		
Topics	**Reference** Stats in Your World Text	**CCS**
Variability in Sample Proportions	**Chapter 16**	HSS.IC.A.1
Independence Assumption	www.gallup.com	
Randomization Condition	education.ti.com/en/84activitycentral	
Sampling Distribution Model for a Proportion		HSS.IC.B.4
Confidence Interval		HSS.IC.A.2
Margin of error		
Standard error		
Critical Values		
Testing Hypotheses	**Chapter 17**	
4-step procedure		

Statistics
Unit 5: Statistical Inference

• Null hypothesis • Model • Mechanics • Conclusion Making Errors Alternative Decision Rules Inferences About a Mean Simulating the Sampling Distribution of a Mean The Fundamental Theorem of Statistics Confidence Interval Hypothesis testing Paired Data Comparing Proportions or Means Standard Deviation of the difference between two proportions Assumptions & Conditions Confidence Interval Comparing Two Means Confidence Interval for the difference of two means Testing the difference between two means Testing Hypotheses by Simulation	**Chapter 18** lib.stat.cmu.edu/DASL www.learner.org/courses/againstallodds/ education.ti.com/en/84activitycentral **Chapter 19** www.fastfoodnutrition.org/	HSS.IC.B.4 HSS.IC.A.2 HSS.IC.B.5 HSS.IC.B.5

Statistics
Unit 5: Statistical Inference

Comparing Counts Hypothesis Test - "goodness-of-fit" Assumptions & Conditions Chi-Square Model Significance Two-way frequency tables Examine Components	**Chapter 20** www.mms.com	HSS.IC.B.5

Unit Assessments
Section Quizzes, Project(s) – see Appendix, Applications, End-of-Unit Test

Statistics
Unit 5: Statistical Inference

Applications

1. **Coin Activity.** Suppose we want to understand how the sample "mean year on pennies" is distributed. The population of pennies we have available for investigation is a collection of 1002 pennies which were obtained from the UMass Five College Credit Union on August 25, 2010. What do you think the distribution of year is for the population of pennies? Explain.

   Obtain a sample of 30 pennies, and compute the sample mean year. What value do you get? We are not sampling with replacement. Compare your mean value with the class (class graph). Are the values very different? What does the distribution of sample mean year look like based on the graph? Do you think looking at roughly 30 samples of size 30 is good enough to tell us about the distribution of sample mean year when n is 30?

2. **Pine Trees Related to Soil Type.** You have noticed that pine trees grow well in some parts of the woods, but not others. You speculate that the distribution of pines is related to drainage, that is, that pines prefer a very well-drained soil, while they do poorly in wet areas. You sample soil from evenly spaced plots throughout the forest, two days after a heavy rain. You find that you can describe each plot as belonging to one of three *categories* of soil: dry (sample falls apart in your hand), loamy (holds shape if you squeeze it, falls apart if you drop it), and wet (muddy - you can squeeze lots of water out, soil tends to run through your fingers).

   Now, if soil drainage has no bearing on the distribution of pines, then you would expect half of the plots of each soil type to have pine trees, provided you sampled enough plots. That is, the *expected frequency* of soil types in plots with pine trees is 50% dry, 30% loamy, and 20% wet. An expected frequency assumes that catagories have no effect on the variable being measured (in this case, whether or not a plot has pines) and assumes that you sample enough times so that you have a representative sample.

Let's say you had 100 plots, and you found that 50 were dry, 30 loamy, and 20 were wet. Let's also say that 50 plots had pine trees on them. Among the 50 plots with pines, then, the *expected distribution* of soil types would be 25 dry, 15 loamy, and 10 wet. Suppose now that you *observed* that of the 50 plots with pine trees, 31 were dry, 17 loamy, and only 2 were wet. It looks like there was a tendency for pines to grow in dry soils.

Determine the probability that your observation would hold up if you were to take an infinite number of samples. The following method gives you a probability that your conclusion is accurate. For each category take the observed frequency (O) and subtract the expected frequency (E). Square the difference and divide by E. Add up the results for the three categories. The total is the Chi-Square statistic.

Calculation of the chi-square statistic

31 observed dry minus 25 expected dry = 6

6 squared = 36

36 divided by expected frequency E = 36/25 = 1.44

The other two categories gave values of 0.27 and 6.4. The total adds up to 8.11, which is the chi-square value.