Livermore Valley Joint Unified School District

Course Title:	Advanced Placement (AP) Statistics A/B
Grade Level(s):	10 - 12
Length of Course:	Two semesters or equivalent term
Credit:	10 units
Prerequisite:	Completion of Algebra II with a C- or better or consent from instructor
Co-requisite:	none

Course Overview:
The purpose of *AP Statistics A/B* is to introduce students to the major concepts and tools for collecting, analyzing, and drawing conclusions from data. Students are exposed to four broad conceptual themes:

1. Exploring Data: Describing patterns and departures from patterns
2. Sampling and Experimentation: Planning and conducting a study
3. Anticipating Patterns: Exploring random phenomena using probability and simulation
4. Statistical Inference: Estimating population parameters and testing hypotheses

Technology and statistical literacy will be integrated throughout the course. Students who successfully complete the course and exam may receive credit, advanced placement, or both for a one-semester introductory college statistics course.

Schools Offering:	Del Valle High School Granada High School Livermore High School Vineyard High School
Meets University of California Entrance Requirements:	*Update seeking "c" Approval* Prior UC Approval Feb. 12, 2018
Board Approval:	*Pending Board Approval* Prior Board Approval March 6, 2001
Course Materials:	*The Practice of Statistics, 6th edition;* Starnes, Daren and Tabor, Josh; Bedford, Freeman & Worth, 2019 ISBN-13: 978-1319269296 http://www.highschool.bfwpub.com
Supplemental Materials:	Sapling Plus Strive for a 5: Preparing for AP Statistics Statsmedic

AP STATISTICS A/B

COURSE CONTENT:

Unit 1: Analyzing One-Variable Data

Unit 1 introduces students to how to organize and visually display data from a single variable. Students then analyze data using numerical and graphical summaries. Students also learn how to calculate and describe the location of an individual within the distribution of a quantitative variable using the numerical data. They also use graphical models of density curves, including the Normal distribution, to estimate an individual's location in a distribution.

Summary of Key Assignments and/or Activities

At the end of chapter 1 within this unit, students will complete an assignment called the American Community Survey (ACS) Project. For this project, students will access the U.S. Census Bureau database where over 300,000 households provided information to the ACS. Data from the ACS are used to determine how the federal government allocates over $400 billion in funding for local communities. Students will produce a spreadsheet of the data via the help of bfdpub.com and then use the files to summarize a report answering various questions related to the topics covered in this unit. Topics address include how many variables were recorded and of what type, constructing appropriate graphical displays and describing them, and accurately comparing distributions.

Also, in this unit, students will complete an activity called “Where Do I Stand.” In this activity, students will explore ways to describe where they stand (literally) within a distribution. The teacher will mark out a number line on the floor with a scale running ~58 to ~78 inches and students will make a human dot plot along the number line in regard to their height. They will then convert this into a dot plot on their paper and complete a report analyzing the class distribution. Within the student height distribution, students will be able to see in real life concepts being learned such as percentiles, the mean and standard deviation, how the distribution would be adjusted if units were converted into centimeters. Students can then access the cdc.gov/nchs to apply these ideas to other measurements such as weight and body mass.

Unit 2: Analyzing Two-Variable Data

In this unit, students investigate relationships between two quantitative variables by making scatterplots of paired data. They determine the strength of the linear relationship between two variables, calculate the least-square regression equation, how well the linear model fits the data, and use the linear model to make predictions. Students will also learn how to transform curved data to achieve a linear model and use that to help calculate an exponential or power model that best fits the curved data.

Summary of Key Assignments and/or Activities

At the end of this unit, students will complete a project called Investigating Relationships in Baseball. In this project, students will use technology to help them answer questions such as, What is a better predictor of the number of wins for a baseball team (the number of runs scored or the number of runs they allow to be scored), What variables can we use to predict the number of runs a team scores, etc. This activity will have students access the MLB Team Data 2012-2016 excel file and import it into statistical software of their choice (online, graphing calculator). They will then construct two scatterplots to investigate this relationship, calculate summary statistics such as correlation, the coefficient of determination, and LSRL for each set of variables. Then, compare the two associations.

Unit 3: Collecting Data
In this unit, students learn how to identify populations and samples, good and bad sampling methods for observational studies, and how to take a simple random sample. They expand their knowledge to other random sampling methods and how sampling errors and non-sampling errors can occur. Students then learn good and bad methods of experiments, how to identify the experimental units and treatments and methods of random allocation to treatment groups. They finish by learning how to use studies wisely.

Summary of Key Assignments and/or Activities
A key assignment within this unit is called "Who wrote the Federalist Papers?" This assignment is key to help students understand the value in randomization when it comes to sampling data. Students will start the activity with a passage from the Federalist Paper No. 51, which is one of the papers where the author has been disputed. Students will start by choosing 5 words from this passage; any 5 words they would like. They then count the number of words and find the average word length and construct a dot plot of the class data. Students are then introduced to a random number generator (for the first time!) and then use it to randomly select 5 words. They will then repeat the process of finding the average and constructing a dotplot. Students will then see and discuss how the dotplots compare, why they may look the way that they do, and how random selection changes the types of samples we ended up finding data on. Random sampling is a KEY element of conducting well-designed experiments and this activity will help students see why.

Unit 4: Probability
In this unit, students will learn about randomness, probability, and simulation to estimate probabilities along with probability rules and how to organize information into tables to make probability calculations and conditional probability calculations. They will also learn about discrete and continuous random variables including binomial and geometric random variables, how to transform and combine random variables and make corresponding probability calculations. Connections will be made between binomial random variables and the Normal distribution as well as making and comparing probabilities from a binomial distribution and the Normal approximation to the binomial distribution.

Summary of Key Assignments and/or Activities
The activity "1 in 6 wins" helps students understand an important element of what probability is. Most people understand probability as the "odds" or "chances" of something happening, however, in Statistics, probability is related to a long run proportion of many trials occurring. This activity helps students see (literally, with a dotplot) that there will be variability in the outcomes within a game of chance (like winners under a soda bottle cap) and that simply saying "1 in 6 wins" doesn't truly mean that 1 in every group of 6 people will be a winner. In this assignment, students will roll a 6-sided dice and simulate a group of 30 people playing this soda bottle game and how many people within the class win. Students will then produce a dotplot and analyze the probability of whether the class should be suspicious if only 2 or fewer students win in a class of 30 people.

Unit 5: Sampling Distributions
In this unit, students will learn about the sampling distribution of a statistic or the difference of statistics and how its variability is affected by the size of the sample. Students will calculate the mean and standard deviation for the sampling distribution of a statistic or difference of statistics, interpret the standard deviation, and determine if the sampling distribution can be approximated by a Normal distribution. They will then use the sampling distribution and its normal approximation

along with probability calculations to make a conclusion or evaluate a claim about a population parameter or the difference of population parameters. This includes the Central Limit Theorem, how it applies to the sampling distribution of sample means, and how to apply it to determining the Normal approximation to the sampling distribution of sample means or the difference of sample means.

Summary of Key Assignments and/or Activities
In this activity, students will investigate how the mean year and the proportion of pennies from the 2000s vary from sample to sample using a large population of pennies of various ages. Each member of the class will randomly select a penny from the population and record the year of the penny with an "X" on a class dotplot. The class will then return their pennies to the population and repeat this until they have at least 100 pennies selected. This graph gives students a visual for the population distribution of penny years. Then, each student will select a random sample of 5 pennies from the population and note the year on each penny, but now, they will also record the mean year. Instead of plotting all 5 pennies, each student will plot one xbar on a second class dotplot. Students will then return their pennies to the population and repeat this process until at least 100 samples of size 5 have been plotted. Now students will have the opportunity to compare their population distribution to the distribution of sample statistics. How are the distributions different? Similar? Students will compare the shape, center, and spread of the two distributions and generalize their findings to determine how sample size impacts these key features on a distribution. Students will produce the dotplots as a class but formalize their analysis on individual submissions.

Unit 6: Inference for Categorical Data
In this unit, students will create and interpret confidence intervals for proportions and the difference between proportions as well as determine the sample size necessary for a confidence interval to have a specified margin of error. Students will also state the appropriate hypotheses for significance tests and perform significance tests for proportions and the difference between proportions. Students will be required to check the conditions to determine the validity of creating a confidence interval or performing a significance test. They will calculate and interpret P-values for significance tests, interpret Type I and Type II errors, as well as calculate the power of a test. Learning and using technology will play a critical role in all of these calculations.

Summary of Key Assignments and/or Activities
In this unit, students will complete a key assignment called The Confidence Intervals for Proportions applet. In this activity, students will use the Confidence Intervals for Proportions applet to learn what it means to say that we are "95% confident" that our confidence intervals capture the parameter value. Students will start by simulating a sample of size n=11 and display the resulting confidence intervals. They will then simulate another, and another, and another, so that they can see the variability of the intervals as they are constructed around the sample proportion. They will identify if each interval was successful at capturing the parameter value and record what proportion of the intervals did so. They will then adjust the confidence level to 98% and then 99%. Students will summarize what they notice changes about the display of intervals. Students will then repeat this process and record their results for sample sizes of 1000 instead. This is a key activity because interpreting what the confidence level means is typically an area that students struggle with. This activity helps students construct a visual of what this phrase means that they can then apply moving forward with confidence intervals.

Unit 7: Inference for Quantitative Data

In this unit, students will learn how to check conditions for constructing confidence intervals and performing significance tests for a population mean or a difference in means. If conditions are met they will construct and interpret C% confidence intervals for estimating a population mean, estimating the difference between two means, and estimating a mean difference. They will also check conditions for performing a significance test and determine when it is appropriate to use paired t procedures versus two-sample t procedures. Students will use formulas and tables as well as technology to calculate a test statistic and P-value when performing a significance test about a population mean, difference between means and a mean difference. After constructing C% confidence intervals or performing significance tests students will then make inferences about the population or a claim about a population mean.

Summary of Key Assignments and/or Activities
Students will finish this unit with a project called "Which Costs More: Diesel or Unleaded?" For this project, students will produce a report analyzing the two types of fuel and the price of each considering various factors. To begin, students will access some research that has been collected on the price per gallon of gasoline of a random sample of gas stations in six states: Colorado, Illinois, Indiana, Kansas, Missouri, and Ohio. Students will then start by calculating the difference (Diesel - Unleaded) in gas prices for all 82 stations and store these values in a list either on their calculator or in statistical software. Students will construct a graph to display the distribution of difference in gas prices and thoroughly describe it (shape, center, spread, and outliers). They will then construct and interpret a 95% confidence interval for the true mean difference and answer the question, does the interval provide convincing evidence of a difference in the mean price per gallon of diesel fuel and regular unleaded gasoline? Students will then extend their report by comparing two adjacent states to see if they have similar prices and carry out an appropriate significance test to see if the data provide convincing evidence to contradict this prediction.

Unit 8: Inference for Distributions & Relationships
In this unit, students will state hypotheses, check conditions for and perform a chi-square test for goodness of fit for a categorical variable with two or more categories. They will state hypotheses, check conditions for and perform a chi-square test for homogeneity to determine if the distribution of a categorical variable differs for two or more populations or treatments or perform a chi-square test for independence to determine whether there is convincing evidence of an association between two categorical variables in a population. Finally, students will check conditions for and perform a t test for the slope of a least-squares regression line to determine if there is convincing evidence of a linear association between two quantitative variables.

Summary of Key Assignments and/or Activities
A key activity used to our Chi-Square unit specifically involves an analysis of the color distribution of M&M candies. For this activity, each student will need a regular size bag of plain M&Ms as their sample. Students will open their bag and compare their sample's color distribution to the one claimed by the Mars Chocolate Factory (found publicly on their website). This will involve carrying out a Chi-Square Goodness of FIT Test (calculating a X^2 statistic and p-value). The class will then combine all of their individual samples to create one large sample and repeat this process. Is there any difference in difference in the conclusion made now versus the smaller sample? Why/why not? If students find a significant difference between the class color distribution and the distribution claimed by the Mars Candy Company, students are encouraged to reach out via letter or email to ask for an explanation. An added feature that makes this activity even better is that students are then asked to recall their previous knowledge of 1-sample proportion confidence intervals and

significance tests and do an analysis of the color that differed the most within the class distribution (which color had the highest Chi-Squared component).

California Content Standards and Standards for Mathematical Practice
Higher Mathematics Standards, 2013

Standards for Mathematical Practice

MP1 Make sense of problems and persevere in solving them.
MP2 Reason abstractly and quantitatively.
MP3 Construct viable arguments and critique the reasoning of others.
MP4 Model with mathematics.
MP5 Use appropriate tools strategically.
MP6 Attend to precision.
MP7 Look for and make use of structure.
MP8 Look for and express regularity in repeated reasoning.

Statistics and Probability
Interpreting Categorical and Quantitative Data
Summarize, represent, and interpret data on a single count or measurement variable

1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
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.
3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
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.

Summarize, represent, and interpret data on two categorical and quantitative variables

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 the data.
6. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
    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.
    b. Informally assess the fit of a function by plotting and analyzing residuals.
    c. Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
8. Compute (using technology) and interpret the correlation coefficient of a linear fit.
9. Distinguish between correlation and causation.

Making Inference and Justifying Conclusions

Understand and evaluate random processes underlying statistical experiments

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
2. Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. For 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?

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

3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
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.
5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.
6. Evaluate reports based on data.

Conditional Probability and The Rules of Probability-

Understand independence and conditional probability and use them to interpret data

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").
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, and use this characterization to determine if they are independent.
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.
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. For 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.
5. Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. For 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.

Use the rules of probability to compute probabilities of compound events.

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.
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.
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.
9. (+) Use permutations and combinations to compute probabilities of compound events and solve problems.

Using Probability to Make Decisions

Calculate expected values and use them to solve problems

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.
2. (+) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution.
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. For 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.
4. (+) Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value. For 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?

Use probability to evaluate outcomes of decisions

5. (+) Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values.
   a. Find the expected payoff for a game of chance. For example, find the expected winnings from a state lottery ticket or a game at a fast-food restaurant.
   b. Evaluate and compare strategies on the basis of expected values. For 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.
6. (+) Use probabilities to make fair decisions (e.g., drawing by lots, using a random number generator).
7. (+) Analyze decisions and strategies using probability concepts (e.g., product testing, medical testing, pulling a hockey goalie at the end of a game).

Instructional Methods and/or Strategies

Direct Instruction (Lecture)

- Direct instruction as an instructional method is used to provide students with clear understanding of new curriculum concepts through lecture and guided discussion. This can be used to help simplify technical language and connect current topics being learned to earlier information that has already been covered.

Flipped Classroom

- Pre-recorded videos are used to support students in one of two ways:
  1) provide students with review of previous material that will be used in the upcoming topics
  2) provide initial exposure to new material that students can/will then apply to practice/applications in upcoming classes
- Both of these options support students in successfully learning new material.

Activity based instruction

- Activity based instruction is used to help students see the topics being learned in the real world. This strategy gives students the opportunity to extend their knowledge beyond a textbook solution and problem solve using real data and collaboration with peers.

Assessment Methods and/or Tools

Practice Problem Sets

- Intent: Check understanding of small sections of curriculum on specific concept.
- Goal: The goal is to provide specific feedback on detailed skills as students are learning the material to recognize common pitfalls and avoid forming bad habits.

Lab Activities

- Intent: Show real life applications of content. Students will apply concepts learned in the course to real life situations including data that they themselves collect.
- Goal: Help students display an understanding of the content beyond a structured practice problem.

Assessments

- Intent: Provide students with a more summative assessment of their understanding in the form of Collegeboard style questions and expose them to the scoring guidelines used on the exam in May.
- Goal: Measure student understanding and retention on larger breath of concepts than individual one at a time and prepare them for the formatting on the AP exam in May.

Projects: Probability Unit Project and End of Course Project

- Intent: Give students an opportunity to apply what they have learned in a creative way.
- Goal: Show students how what they have learned can be applied to a topic of their choosing.

AP Exam (administered by Collegeboard) typically given in May of each year.

Honors Courses

The AP Statistics curriculum is designed to prepare students for the AP Statistics Exam which involves higher level critical thinking questions and free response questions which require students to correctly use statistical terminology and more comprehensive analysis of one and two variable statistics and probability. Students will do in depth checks of conditions for inference, interpret computer output to create and interpret regression models, write formal conclusions based on inference methods, and write formal comparative analyses of distributions of samples and sampling distributions. Projects and Application problems that span multiple units are used for review and to assess students' understanding of how topics tie together to answer real world questions. Students are also required to do full investigations by designing appropriate observational or experimental studies, producing data, performing comprehensive analysis of the data, and performing inferential statistics to answer questions about a population or a claim about a population. The comprehensive final examination is modeled after the AP Statistics exam which has two sections: a multiple-choice section and a free response section. The final exam will cover all units as described in the College Board AP Statistics description. Students are also given an end of year project in which they must design their own study to answer a question of their choice to demonstrate what they have learned throughout the course.