ADVANCED PLACEMENT STATISTICS

Description: This course is intended for students who have a thorough knowledge of algebra, geometry, trigonometry, and elementary functions. The purpose of the course is to prepare the student for advanced placement into college statistics. The content of the course is driven by The College Board Advanced Placement Course Development Syllabus. Students are required to have completed an Honors Pre Calculus class. Exceptions, on a case by case, may be made for exceptionally motivated students who have completed a PreCalculus course with a grade of at least A-. Students are required to complete a summer reading and problem solution assignment prior to admission into the course. The assignment is due on the first day of class in September. The graphing calculator is used extensively throughout the course for calculations, displays, and simulations. Computer output, in the form of TI-83 screenshots, Minitab displays or Excel displays, are used to enhance the development of statistical understanding through data analysis and assessment of models. Problem solving is introduced early and integrated throughout as connections to other subject areas are made through practical applications. Models and technology are used when appropriate. Throughout the course, students are encouraged to use the language and symbology of statistics.

Student Summer Preparation: During the summer preceding enrollment in AP Statistics, students are required to read Statistics for the Utterly Confused, listed below, and to complete the problems in the first eight chapters. These chapters introduce measures of central tendency, measures of variability, probability, distributions, hypothesis testing, and confidence intervals.

Daily Schedule: Quizzes are held several days every week. The quizzes are modeled after the multiple choice questions on the AP Statistics Exam, although they are frequently presented in "calculation form" with no answer choices given. The quizzes are presented in a form which resembles the AP Statistics Exam, to ensure that the students are accustomed to the Exam format. Daily quizzes last 5-10 minutes and are completed in mechanical lead pencil. At the completion of each quiz, students use a pen or colored lead pencil to mark their papers. We review the quizzes immediately and questions are answered. The class discusses any errors they have made so we can all learn from each other. Students submit their work and I give the solutions to the day's quiz with the solutions worked out step-by-step, printed on blue paper, to each student. Students are advised to keep a file of their "blue sheets" to use in preparation for the AP Statistics Exam

Technology:

a. Each student is expected to have a TI-83, TI-83 Plus or TI-83 Silver calculator and is required to bring the calculator to class every day. The calculator is used extensively for calculations, simulations, graphing histograms, boxplots, and distributions, normal, skewed, geometric, and binomial. We use the ProbSim simulations package.

b. I use the TI-Presenter to display my graphing calculator on the classroom television screen. I use wet-erase colored markers to write on the television screen to accentuate distribution shapes, outliers, etc.
c. I use an overhead projector with transparencies to review old AP exam questions, solutions and rubrics.

Objectives: The student will be able to:

- Understand, interpret and create statistical graphs and displays
- Describe distributions with numbers, mean, standard deviation, variance, etc.
- Understand and recognize normal distributions and approximately normal distributions
- Calculate z-scores, z-sub scores, and normal scores
- Find the area under the standard normal curve between any two z-scores
- Use normal distributions to determine probabilities
- Plot two-variable data in scatterplots
- Calculate linear correlation between two variables and interpret the results
- Apply the least square regression method to determine the best-fit linear predictor
- Understand the limitations of linear regression
- Understand relationships between categorical variables
- Differentiate among experiments, simulations, and observational studies.
- Collect useable, valid data for experiments and simulations
- Design samples and experiments; simulate experiments
- Understand and apply the concepts and rules of statistical probability
- Understand discrete and continuous random variables
- Calculate discrete random variable probabilities
- Calculate and interpret the means and variances of discrete random variables
- Understand and interpret binomial and geometric distributions
- Understand and apply the concepts and rules of statistical inference
- Construct and interpret confidence intervals; apply tests of significance
- Conduct hypothesis testing
- Calculate the probability different types of errors in hypothesis testing
- Perform appropriate statistical testing for normal distributions and Poisson distributions
- Perform appropriate statistical testing for Chi-Square distributions, including testing for goodness of fit and inferences for two-way tables
- Use regression analysis to make predictions and interpret correlation
- Express hypotheses, designs, conclusions, etc. in the vocabulary of statistics
- Interpret computer outputs
- Use the TI-83 calculator to compute statistical calculations, such as five-number summaries, P-values, standard deviation, etc
- Use the TI-83 calculator to conduct one-sample and two-sample z-tests and t-tests, one-proportion and two-proportion z tests, and χ^2 tests
- Use the TI-83 calculator to find one-sample and two-sample z-intervals, one-sample and two-sample t-intervals, and one-proportion and two-proportion z-intervals
- Use the TI-83 calculator to create statistical displays, such as boxplots, histograms, scatterplots, regression models, and rejection and non-rejection regions for a variety of hypothesis tests

Introduction: The course is divided into four major sections: (1) collection and organization of data, including patterns in data; (2) data production, including sampling, experiments and simulations; (3) probability and distributions, including the normal, binomial and geometric distributions; and, (4) inference, conclusions and confidence, including z-tests, t-tests, and chi-square tests. As we progress through the course, the relationships among sampling design, analysis, inference, conclusions, and confidence are continually tied together. Students do not learn concepts in a vacuum; they see how they are interrelated. At least once per week, I bring copies of our local newspaper to class. In the beginning of the course, students enjoy finding, interpreting, and criticizing graphs, sports statistics and other statistics of local interest. As we progress in the course, they relish pinpointing statistical errors, such as misleading or incorrect graphs; discovering how statistical statements are used to favor a particular point of view; criticizing the "randomness" of man-on-the-street surveys; to question the sample size of reports with a particular margin of error; and, to correct incorrect wording of confidence interval statements. It is a joy to see the students integrate all they have learned and to make accurate assessments and applications in their lives. The terminology and symbology of statistics are stressed throughout the course. Starting on the first day of class, students are introduced to the notion that some words in the English language have specific meanings in Statistics. Students use the term "*measure of central tendency*" not "average." They learn that "*expected value*" may actually be a value which is not achievable, such 2.3 people. And, the words *significant*, *normal, discrete,* and *confidence* are used appropriately. Occasionally, a student will slip and use a word incorrectly. The other students revel in pointing out the mistake. ☺ Another example is the stressing of the meaning of the confidence interval. Each year, in the media, we can usually find a plethora of incorrect statements concerning confidence intervals. The format of the erroneous statements follow the pattern: "There is a 95% chance that the mean is between 20 and 30." Students delight in correcting the statement to read, "There is a 95% chance that the interval from 20 to 30 actually contains the mean." Use of the proper symbology is required throughout the course. So, the mean of a sample is $\bar{x}$ and the mean of a population is μ. Using them interchangeably is not acceptable. Similarly, students must distinguish among p, $\bar{p}$, $\hat{p}$, p_1 , p_2 , $\hat{p}_1$, $\hat{p}_2$, *etc.* and always use the correct term. When identifying critical values and test statistics, students distinguish between test *z* and critical *z* as well as between test *t* and critical *t*.

Course Content:

Part I: Organizing Data: Looking for Patterns and Departures from Patterns. Students begin their exploration of data analysis. On the first day of class, I ask the students to remain quiet for one minute and take their pulse. After one minute, I ask each student to report not only his or her pulse rate, but where the student measured the pulse (wrist, neck) and how long he or she measured it (10 seconds, 15 seconds, 60 seconds). From this simple exercise, the class learns that measurements must be taken in a consistent manner in order to be compared. To illustrate the normal distribution, we conduct an experiment with grid paper and grains of salt. We watch as the dropping salt forms a bell-shaped curve. The students are then asked to make the salt form a different shape. As they experiment with tilting the paper, they realize that outliers on one side will form a skewed graph. The class constructs and interprets graphical displays of distributions of univariate and bivariate data. The class becomes quite adept at producing graphs on the calculator, analyzing the data, and insightfully interpreting the results. The graphing calculator is utilized in class on a daily basis to strengthen statistical comprehension and assimilation of concepts. After students become adept at graphing data in a variety of formats, we then focus on using the calculator to analyze the data. For example, for a given data set, I will ask the students to tell me what they expect the mean value to be and how far away from the mean are the other data. Then, we calculate the mean and the standard deviation and compare them to our expectations. The students learn form this that they must do a "reality check" on their answers. Entering one value incorrectly, can result in very inaccurate results. We do these "reality minutes" throughout the course. Students are introduced to the vocabulary and symbology of Statistics. Topics are:

Chapter 1: Exploring Data

- Displaying Distributions with Graphs
  - Types of Data
  - Dotplots, stemplots, histograms, ogives, time plots, bar charts, pie charts, Pareto charts
  - Graphical Measures of Center
  - Symmetry and Skewness
  - Outliers, trends, seasonal variations
- Describing Distributions with Numbers
  - Numeric Measures of Center
  - Quartiles and Outliers
  - Five Number Summaries, Boxplots, Modified Boxplots
  - Range, IQR, Variance and Standard Deviation

Chapter 2: The Normal Distributions

- Density Curves and the Normal Distributions
  - Density curves, shapes , mean and standard deviation

    - Normal distributions and normal curves
    - The 68-95-99.7 Rule
- Standard Normal Calculations
    - The Standard Normal Distribution
    - Standardized observations, *z*-scores
    - Assessing normality, normal probability plots
    - Area under the normal curve
    - Normal areas and probability

Chapter 3: Examining Relationships

- Scatterplots
    - Explanatory and Response Variables
    - Examining scatterplots for pattern, form, direction, outliers and strength of relationship
    - Linear relationships
- Correlation
    - Linear Correlation Coefficient
    - Correlations versus Causation
    - Correlation versus Association
- Least-Squares Regression
    - Regression Lines
    - Least Squares Regression Line
    - Correlation and Regression
    - Coefficient of Determination
    - Influential Observations

Chapter 4: More on Two-Variable Data

- Transforming Relationships
    - Monotonic relationships
    - Transforming Linear Relationships
    - Power and Exponential Transformations, including fractional exponents
    - Power and Exponential Growth Models
- Cautions about Correlation and Regression
    - Interpretations of correlation and regression
    - Cautions about correlation and Causation
    - Avoiding extrapolations and correlations based on averages
    - Confounding variables and lurking variables
    - Effective experimentation
- Relations in Categorical Data
    - Two Way Tables, Row Variables and Column Variables
    - Conditional Distributions

- Simpson's Paradox

Part II: Producing Data: Samples, Experiments, and Simulations. Students study sampling and experimentation through planning and conducting studies. Methods of data collection are investigated. The class plans, designs and conducts surveys; compare samples, and verify randomness. Students plan and conduct experiments, treatments, control groups, replication, random design, and generalization of results. At the beginning of this section, the class designs a survey to gather information from a random sample from the school population. They also decide on a data collection method. At the end of this section, I return their plan to them and they are amazed at how many errors they made. The class critiques its own plan and rewrites it using their newfound knowledge. When reading our newspapers articles, students offer observations such as, "where was this survey done," "what time of day was it," how did they choose who they were going to ask," "did they consider such-and-such lurking variable," etc. Here, the students are demonstrating not only their grasp of statistical processes but are also tying together design, analysis and conclusions. We continue to build on our use of statistics vocabulary and symbology. We make frequent use of the ProbSim capabilities of the TI-83 calculator. Students enjoy devising ways in which we can use these simulations to represent real-world problems. Topics are:

Chapter 5: Producing Data

- Designing Samples
  - Statistical Inference
  - Sampling, Experiments, and Observational Studies
  - Samples and Populations
  - Probability Sampling Designs: SRS with and without replacement, stratified, cluster, census, and systematic
  - Multistage Sampling
  - Random Number Generation
  - Voluntary Responses
  - Bias, Undercoverage, Nonresponse and Poorly Worded Questions
- Designing Experiments
  - Subjects, Factors, Levels, Treatments and Response Variables
  - Control, Randomization and Replication
  - Double Blind
  - Blocking
  - Matched Pairs, Independent Samples
- Simulating Experiments
  - Simulation Design
  - Running Simulations on the Calculator
  - Repetition of Simulations

Part III: Probability, Foundations of Inference. Students focus on patterns and pattern recognition by studying random phenomena using probability and simulation. The class interprets probability, conditional probability, independence and mutual exclusivity of events; discrete random variables, simulation of random behavior, mean and standard deviation of random variables. We then progress to combining independent random variables and mean and standard deviation for sums and differences of independent random variables. The normal distribution is studied in depth and we discover the properties of the normal distributions, investigate normal distribution tables, and use the normal distribution as a model for measurements. We make copious use of the graphing calculator to graph the distributions; calculate factorials, permutations, and combinations; and, calculate probabilities and cumulative probabilities for different distributions. We also use the ProbSim functions to conduct numerous simulations. We also use large foam dice and decks of cards to conduct live simulations. The students quickly learn that the ProbSim simulations are as valid, and less time-consuming, than the 'hands-on" experiments. However, one "hands-on" experiment is always a hit. I give each student in the class a small sealed bag of M&Ms candies. Each student opens their bag and determines the distribution of the colors. We then choose a color at random, and while the samples are being eaten, we study the sampling distribution of the sample proportion of that color and then compare it with the published population proportion. Again, students are firming their grasp of the interrelationships among design, analysis and conclusions. Again, we continue to add to our statistics vocabulary and symbology, especially the symbols of probability. Topics include:

Chapter 6: Probability, The Study of Randomness

- The Idea of Probability
  - Randomness, Chance and Probability
  - Random Phenomena
  - Uses of Probability
- Probability Models
  - Probability and Probability Models
  - Sample Space, Events and Complement of an Event
  - Disjoint and Independent Events
  - Multiplication and Addition Rules of Probability
- General Probability Rules
  - Tree Diagrams
  - Unions and Intersections of Events
  - Conditional and Marginal Probabilities
  - Venn Diagrams
  - Bayes' Rule
  - Joint Probabilities

Chapter 7: Random Variables

- Discrete and Continuous Random Variables
  - Random Variables
  - Probability Distributions for Random Variables
  - Discrete Random Variables
  - Continuous Random Variables
  - Density Curves
  - Probability Histograms
- Means and Variances of Random Variables
  - Mean, μ_x, and standard deviation σ_x, of discrete random variables
  - Mean, μ_x, and standard deviation s_x, of discrete random variables
  - The variance, s_x^2, of discrete random variables
  - The Law of Large Numbers
  - Rules for Means and Variances of Linear Combinations of Discrete Random Variables

Chapter 8: The Binomial and Geometric Distributions

The Binomial Distribution

- Requirements for Binomial Distributions
  - Binomial Probabilities
  - Binomial Coefficient, Factorials

Binomial Probability Distribution Function and Cumulative Distribution Function

  - Mean and Standard Deviation of a Binomial Distribution
  - Normal Approximation to the Binomial Distribution
  - Corrections for Continuity
- The Geometric Distribution
  - Requirements for Geometric Distributions
  - Geometric Probabilities
  - Mean and Standard Deviation of a Geometric Distribution

Chapter 9: Sampling Distributions

- Sampling Distributions
  - Parameters and Statistics
  - Variability of a Statistic
  - Biased and Unbiased Estimators
- Sample Proportions
  - Sampling Distribution of Proportion, $\hat{p}$
  - Mean and Standard Deviation of the Sampling Distribution of $\hat{p}$
  - The Normal Approximation of the Mean and Standard Deviation of the Sampling Distribution of $\hat{p}$
- Sample Means
  - Sampling Distribution of the Mean, $\bar{x}$

- Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$
- Sampling Distribution of a Sample Mean from a Normal Distribution
- The Central Limit Theorem

Part IV: Inference, Conclusions with Confidence. Students study statistical inference by estimating population parameters and by hypothesis testing procedures. To get this section off on an energetic start, I show an excellent PBS video on Probability. From Las Vegas casinos to political elections to research laboratories to sports analysis, the applications of probability and inference are presented in an interesting and animated fashion. This sparks the interest of the students. Throughout this section, we refer back to the examples in the video as illustrations of the concepts. At the beginning of chapter 11, we conduct a paper airplane experiment using two different paper airplane pattern sheets, one of which supposedly has superior flight capabilities. We state the null and alternative hypotheses and then conduct the experiment. Students are very adept at pointing out possible complications of bias in this experiment, such as one student being taller than another or one student applying more force to the plane before releasing it. After we iron out any bias as well as we can, we conduct an interesting and fun experiment. It is useful, and memorable, even though we do not know the population standard deviation for either prototype. This experiment, as well as other examples in Part IV, reinforce the interrelationships among design, analysis and conclusions. Later, we develop a method for calculating a test statistic. We use the calculator to assess various models both graphically and quantitatively. When assessing the "fit" of a model, students are cautioned to pay attention to the scales. Alternatively, when calculating r^2, we consider not only the numeric value but also the application. Students learn that exploration and analysis of data does not result merely in a number, but must also include human scrutiny. What does the deviation not explained by regression mean in a particular model. What does the deviation which is explained by regression mean in a particular model. These may be more meaningful in one application or model than in another. Critical thinking is included in our daily discussions. Topics are:

Chapter 10: Introduction to Inference

- Estimating with Confidence
  - Confidence Intervals and Confidence Levels
  - Confidence Interval for a Population Mean, μ
  - Choosing a Sample Size
  - Sample Size and Margin of Error
  - Critical Value, z^*
- Tests of Significance
  - Null and Alternative Hypotheses
  - Hypothesis Testing Procedures
  - Left-Tailed, Right-Tailed and Two-Tailed Tests
  - P-Values and Statistical Significance
  - One and Two Sample z Tests

    - Tests with a Fixed Significance Level
    - Tests for Confidence Intervals
- Making Sense of Statistical Significance
    - Choosing a Level of Significance
    - Statistical Significance and Practical Significance
    - Statistical Inference
- Inference as Decision
    - Type I and Type II Errors
    - Error Probabilities
    - Significance and Type I Error
    - Power and Type II Error

Chapter 11: Inference for Distributions

- Inference for the Mean of a Population
    - *t* Distributions
    - One-Sample *t* Statistic
    - Degrees of Freedom
    - Standard Error
    - The *t* Confidence Intervals and Tests
    - One-Sample *t* Procedures
    - Matched Pairs *t* Procedures
- Comparing Two Means
    - Conditions for Comparing Two Means
    - The Sampling Distribution of $\bar{x}_1 - \bar{x}_2$
    - The Two Sample *t* Statistic
    - Two-Sample *t* Procedures
    - Significance Tests

Chapter 12: Inference for Proportions

- Inference for a Population Proportion
    - Conditions for Inference
    - Inference for a Population Proportion
    - Choosing a Sample Size for a Desired Margin of Error
    - Confidence Level C for a Confidence Interval
    - Tests Based on the *z* statistic
- Comparing Two Proportions
    - The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$
    - Confidence Intervals for $\hat{p}_1 - \hat{p}_2$
    - Significance Tests for $\hat{p}_1 - \hat{p}_2$
    - Significance Tests for the Pooled Sample Proportion

Chapter 13: Inference for Tables: Chi-Square Procedures

- Test for Goodness of Fit
  - Chi-Square Test for Goodness of Fit
  - Conducting Inference by Simulation
  - Expected Count
  - The Chi-Square Statistic, χ^2
  - Components of Chi-Square
  - Chi Square Procedures
- Inference for Two-Way Tables
  - Two-Way Tables and Problems with Multiple Comparisons
  - The Chi-Square Test for Homogeneity of Populations
  - The Chi-Square Test of Association/Independence
  - Computing Expected Cell Counts
  - Performing the Chi-Square Test
  - The -Square Test and the z Test
  - Interpreting Chi-Square Tests

Chapter 14: Inference for Regression

- Inference About the Model
  - The Regression Model
  - The True Regression Line
  - The Standard Error about the Line
  - Residuals
  - Confidence Intervals for the Regression Line Slope
  - Testing the Hypothesis on No Linear Relationship
  - Least Squares Regression
  - Explained and Unexplained Variation
- Predictions and Conditions
  - Confidence for the Mean
  - Regression Conditions
  - Residual Plots

Part V: Post AP Exam In this after AP Exam topic, we revisit our paper airplane experiment, adding a third design. We conduct a one-way ANOVA using flight distance data from all three airplane models and we perform inference on the mean distances flown of the three models. We also compare the three means graphically and determine if the means appear to be the same or if at least one mean is different from the other means. Without a practical experiment, I have

found that ANOVA is difficult to conceptualize. Once we have conducted the experiment, it makes a lot more sense to the students. Topics are:

Chapter 15: Analysis of Variance

- Inference for Population Spread
  - The F Test for Comparing Two Standard Deviations
  - The F Statistic and F Distributions
- One-Way ANOVA
  - The Analysis of Variance F Test
  - The Idea of Analysis of Variance
  - The Analysis of Variance F Statistic
  - Degrees of Freedom for the F Test
  - Conditions for performing ANOVA
  - Checking Standard Deviations in ANOVA
  - Reading ANOVA Tables

Evaluation: All students are required to participate in the nationwide AP Statistics Exam. No other Final Exam in the course will be administered. Evaluation is based on homework, quizzes, chapter tests, projects, and the semester examination.

Text and Supplemental Materials:

1. Yates, Daniel S., Moore, David S., and Starnes, Daren S. (2003). The Practice of Statistics, Second Edition. New York: W. H. Freeman and Company
2. Jaisingh, L., (2005). Statistics for the Utterly Confused, 2nd Ed. New York: McGraw-Hill
3. TI-83 or TI-83 Plus or TI-83 Silver Graphing Calculator
4. Previously released AP Statistics exam questions
5. Weekly copies of local newspaper
5. Supplementary books and materials
6. Graph paper
7. Mechanical Pencil
8. Colored pencil (not red) or pen with colored ink (not red).
9. Small ruler
10. A pencil pouch is recommended but not required.