Mr. Rogers' Syllabus: AP Statistics

What to Expect:

Reasoning based on probability and statistics gives modern society the ability to cope with uncertainty. It deals with the **design** of how data is collection, the **analysis** of the data, and the drawing of **conclusions** from the data.

Statistics has astonishing power to improve decision-making accuracy and test new ideas. It's a key analytical tool used in education, the social sciences, and business administration and is often a required college subject for majors in those areas. Statistics is frequently used for data analysis in the sciences and forms the mathematical basis for quality control in manufacturing.

AP Statistics is a college level class for students who have been highly successful in Algebra II. It covers the topics needed for the American College Board AP Statistics exam. Students passing this test may receive college credit.

Who Should Take This Class: Students with an interest in careers related to: **medicine, bioinformatics, genetics, business, mathematics, architecture, engineering, any of the sciences, or the social sciences including psychology, sociology, political science, and education.** In college, statistics is required for most majors with the possible exception of the fine arts..

Credit: One math unit

Summer Assignment: Students who will be taking AP Statistics in the fall should read and familiarize themselves with the online articles at ***Amazing Applications of Probability and Statistics*** .

Prerequisites: Algebra II Honors or Algebra II with teacher recommendation. A combined PSAT verbal and math score of 111 or higher is a good indication that you have the background to take the course.

Extra Credit Opportunity--Science Fair

Resources and Materials for Class

Calculators and Computers

For the purpose of understanding, students will learn all the traditional statistics techniques of using statistics equations and tables for problem solving. However, In today's world, most statistical analysis is done on calculators or computers and so resources will be provided as follows:

1. **TI-83 Calculator:** virtually every traditional calculation will also be done on a TI-83 .
2. **Excel Spreadsheet Software:** has become ubiquitous and so students will learn how to insert functions and write formulas in it. These will be used for simulations and regression analysis.
3. **Minitab Statistics Software:** a popular statistics software with more statistics features than Excel. Students will use Minitab for tasks such as regression analysis, chi squared tests, and parallel box plots.
4. **Fathom Software:**: will be used by the teacher for demonstrating the basis behind

Grading and Assignments

Grading

(For details see Mr. Rogers' Syllabus - Information Common to All Classes.) Tests will be the single largest item and will be written as close to AP exam standards as possible. The good news: practice tests for most units are available online. The link is at the top of each unit plan.

Additional grades will come from the Statistics Investigations and Enrichment Activity as shown below. Course work will generally be finished by sometime in March leaving the rest of the year for review and special projects.

The first semester exam will be taken directly

several of the statistics techniques such as regression and correlation.

Textbooks

Mr. Rogers will provide the following books:

The Practice of Statistics Advanced Placement Edition: TI-83 Graphing Calculator Enhanced (Hardcover)
by David S. Moore, George P. McCabe, Dan Yates

Cracking the AP Statistics Exam, 2004-2005 Edition (College Test Prep) (Paperback)
by Princeton Review

Web Page Resources

Amazing Applications of Probability and Statistics: Mr. Rogers has provided a series of online articles at http://www.intuitor.com/statistics/ both for your enrichment and use in class. These contain a number of applets supported in part by Clemson University and the National Science Foundation.

Syllabus and Unit Plans: a copy of this syllabus with links to a complete set of unit plans including home work assignments and practice tests can be found at http://www.intuitor.com/student/AP_Stats_Syl.php .

Materials for Class to be Provided by The Student

1. **A USB thumb drive** or other storage media for backing up spread sheet programs.
2. **A TI-83 graphing calculator.** Other calculators can be used, however, the TI-83 is specifically designed for statistics and will be the only calculator covered in class.
3. **A set of dry erase markers**. You will frequently be working problems at the front

from old AP tests or AP study books and is often an indication of success on the AP exam. To maximize your grade, you will have to begin using the Barron's AP test study book (see Materials for Class) no later than December. The fourth quarter grade will consist primarily of practice AP test grades. Generally, there is a high correlation between these practice exams and your grade on the AP test.

Classwork

You will be performing statistical analysis of various types both as an individual and as a member of a team on almost a daily basis. At times you will be asked to work on the white board at the front of the room. Your active participation will be required.

Statistics Investigations

A statistical study has 3 distinct elements:

1. **Design:** This involves the systematic approach for collecting the data needed for the other 2 steps.
2. **Analysis:** Here the data is analyzed using specific statistical tools such as box plots.
3. **Conclusions:** The goal of any statistical study is to draw conclusions but even here, statistics uses formalized methods.

Students will perform a number of statistics investigations involving all three of the above elements.

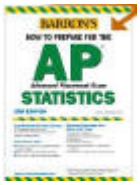
We will also be performing a number of simulations to help us verify theory as well as understand the use of and limitations on many of our statistical techniques. Again these will involve all 3 of the elements described above.

Each student will maintain a bound composition book (see Materials for Class at right). Complete specifications on the format for the investigations can be found here.

The data and analysis from the various investigations should be stored on a USB thumb drive or other media. We will be returning to some of the investigations later in the year for further analysis. If you do not save the data, you will have

of class on a white board.

4. **A bound composition book** for statistics investigations. Color is unimportant, but the size and type of book must match with those shown in the picture. Spiral bound books or loose sheets will not be graded.

5. 

Barron's How to Prepare for the Ap Statistics : Advanced Placement Test in Statistics * by Martin Sternstein, available from Amazon.com (follow the link) or through local book stores such as The Open Book or Barnes and Noble. To maximize you grade you will have to begin studying from this book no later than Dec.

to re-generate it

Enrichment Activity
Each student is to Review the statistical analysis in an article published in an academic, technical, or scientific journal once per quarter in order to see how "the professionals" use statistics (professionals being a very loose term). Popular magazines are not allowed. The paper reviewed must be at least 3 pages long.

The write up should be four concise paragraphs. The first should **summarize** the article's content (what the article is about). The 2nd briefly review how the article **designed** its data collection methods. The 3rd how it **analyzed** its data, and the 4th whether the **conclusions** drawn were statistically defensible. It's only necessary to choose one item or aspect from the article and briefly discuss the design, analysis, and conclusions relative to it.

All reviews are to be typewritten and have a photocopy of the first page of the article and subsequent pages with relevant statistics. Reviews should be less than 2 pages in length.

A given article can be reviewed by only one person. Students should sign up in advance for the article they want to review. All reviews are due one week before the end of the quarter.

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1 Distributions	2 N-Distribution	3 Regression	4 NL Regression	5 Data
Unit Plan		Practice Test		

AP Statistics Standards

I. Exploring Data: Describing patterns and departures from patterns (20% –30%)

A. Constructing and interpreting graphical displays of distributions of univariate data (dotplot, stemplot, histogram, cumulative frequency plot)

1. **Center and spread**
2. **Clusters and gaps**
3. **Outliers and other unusual features**
4. **Shape**

B. Summarizing distributions of univariate data

1. **Measuring center:** median, mean
2. **Measuring spread:** range, interquartile range, standard deviation
3. **Measuring position:** quartiles, percentiles, standardized scores (z-scores)
4. **Using boxplots**
5. **The effect of changing units on summary measures**

C. Comparing distributions of univariate data (dotplots, back-to-back stemplots, parallel boxplots)

1. **Comparing center and spread:** within group, between group variation
2. **Comparing clusters and gaps**
3. **Comparing outliers and other unusual features**
4. **Comparing shapes**

Objectives	Activities
Essential Question: How many numbers are needed to describe a complex event or object? 1. State the difference between categorical and quantitative variables and give examples of each. 2. Define distribution and state two key pieces of information require to produce a distribution. ***The pattern of variation of a single variable*** • ***Quantitative data (numbers)*** • ***How often various values are expected*** 3. State the 3 key ways a distribution can be described.	**Lesson 1** **Class Start Up** Distribute & discuss syllabi. **Warm Up:** **Purpose:** Can complex objects and phenomenon be described with a few numbers? **Design:** Have each person describe themselves with 2 words and 2 numbers and see if the class can determine who's who. **Analysis**: Record how many guesses were required for each person and plot them sequentially. **Conclusions:** Was it possible to beat random chance? Did the number of guesses improve with time? In real life

A. Center or central tendency
B. Spread
C. Shape

3. Make dot plots.
4. Make histograms using the TI-83 calculator and in Minitab.
5. Convert distribution data into a cumulative frequency plot.
6. State the key weakness of histograms.

Homefun: Read section 1.1 and 1.2, work exercises 1.3(p.9), 1.5, 1.7(pp.16, 17)

Essential Question: When using a number to describe a complex event or object what kind of number should be used ? Is there a difference between using a single number and using a single data point?

Ch1.2 Describing Distributions

7. Name and define the 3 key measure of central tendency.

   - Mean - numerical average
   - Median - midpoint, 50% above, 50% below
   - Mode - most common data point or highest peak

8. Given a set of data determine the mean, median and mode.
9. Define and ID outliers.

   - Data point not in distribution
   - Gaps

9. State which measure of central tendency is generally most influenced by outliers.
10. Using the Mr. Rogers Rat Tail Rule, state whether a distribution is skewed left or right, high or low.

are complex objects and phenomenon ever described using a few numbers? In those cases, how important is it to make sure the numbers are meaningful.

Key Concept: What is a distribution?
Purpose: Lay the foundation for describing a set of data.

Interactive Discussion: Objectives

Resources/Materials: Picture of histogram with various sized increments to illustrate the key weakness of histograms.

Lesson 2

Key Concept: Central tendency and shape of a distribution.

Purpose: Lay the foundation for describing a set of data.

Interactive Discussion: Objectives

Individual Work: Find the mean, median, and mode in simulated sets of data with both odd and even numbers of data points.

The Mr. Rogers Rat Tail Rule--FAQ

Skewed distributions often look like a rat with a long tail. The tail points in the direction of skew.

What gets skewed? The mean gets skewed or moved in the direction the rat tail points.

Why does skew matter? For a skewed distribution, the mean is less representative of the bulk of the data

11. Give examples of data which would tend to be symmetrical and data which would be skewed left or right.

    - Easy Test
    - Hard Test
    - Normal Test
    - Incomes

12. Explain the difference betaeen and outlier and skew.
13. Explain how skew and outliers can impact a study's **design, analysis**, and **conclusions**.

Homefun: read 1.2, 1.15, 1.17, 1.19 p. 27-28

points.

What gets skewed very little? The median. It will be more representative of the bulk of the data points than the mean.

Does the possibility that the distribution is skewed affect the design of a study? Yes, more data points usually need to be collected in order to characterize the central tendency of data from a skewed distribution.

Does the possibility of skew affect the analysis of a study? Yes, as we shall see in later chapters, some statistical tools should not be used with skewed data. With skewed data, the median is a better measure of central tendency than the mean.

Does the possibility of skew affect the reliability of conclusions drawn from a study? Yes and no. If the study has been properly designed and analyzed with enough data points then reliable conclusions can be drawn. If for some reason only a few data points are available, the conclusions will be much less reliable and repeatable if skew is present even with good data analysis.

Stats Investigation: Investigation School Evaluation - time approx 3 class periods (individual work)

Purpose: Determine if it is reasonable for 50% of all schools receiving a school report card to be scored below average.

Design: There are hypothetical schools in the simulation, each receiving a randomly determined integer score from 0 to 9. The Excel spread sheet formulas are partially completed. Finish the spread sheet set up as shown below:

1) Increase the number of schools to 30.
2) Add the median statistic below the Mean.

3) Run the simulation 100 times.

Analysis:

1. Record the lowest and highest values for mean and median. Keep track of how many times the mean is below the expected value.
2. Keep a running tally of the frequency of each score 0 through 9. Plot a distribution using this data.
3. Categorize the distribution as symmetrical, skewed right, skewed left, etc.

Conclusions: Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

1) What type of distribution of scores did you expect to get from the above simulation and did the data match with the expectation?
2) Did 50% of the schools typically score below average? Justify your answer using your data.
3) Were there times when more than 50 % scored below average. What would school officials be likely to do in these situations?
4) Were there times when less than 50 % scored below average. What would school officials be likely to do in these situations?
5) Are actions taken in response to the conditions in questions 2 and 3 likely to produce meaningful changes in school performance?
6) What is the minimum and maximum possible % of schools below average? What type of distribution or phenomena could cause this to happen?
7) Why would it be important to know the distribution before setting up a performance measurement for schools?

Stem and Leaf Plots

14. State why a time plot should always be used in an analysis of data.
15. Draw and interpret stem and leaf plots including back to back stem and leaf plots (p. 19).
16. Draw and interpret back to back stem and leaf plots .

Homefun: prob. 1.20

Essential Question: Is there a difference between skew and outliers?

Lesson 3

Key Concept: All data varies with time.
Stem and Leaf Plot

Purpose: Understand the reasons all data should be plotted against time.

Interactive Discussion: Objectives

Seat Work: Draw stem and leaf plots both on paper and with TI-83 calculators using simulated data.

Lesson 4

Box and Whiskers Plots

17. Calculate quartiles, Q1 and Q3.
18. Interpret 5 number summaries.
19. Find the IQR or interquartile range for a data set.
20. Draw a box and whiskers plot.
21. State the Mr. Rogers Rat Whisker Rule for determining skew using a box and whiskers plot.
22. State the % of the data expected in each whisker and in the box for a box and whiskers plot.

Homefun: prob. 1.49, 1.52:

Key Concept: Using box and whiskers plots to describe distributions

Purpose: Box and whiskers plots are an outstanding tool for communicating information about data in a straight forward manner

Interactive Discussion: Objectives

Individual Work: Draw box and whiskers plots both on paper and with TI-83 calculators using simulated data.

Essential Question: In designing a study, analyzing its data and drawing conclusions, why are outliers important?

Lesson 5

Key Concept: ID outliers using modified B&W plots

Modified Box and Whiskers Plot

23. Identify outliers using a modified box and whiskers plot.
    - Whisker's End = 1st data pt within 1.5 IQR of box edge
    - Outlier = data pt beyond the whisker's end
24. Create box and whisker plots on the TI-83.
25. Create and interpret parallel box and whisker plots on the TI-83 and in Minitab.

Purpose: The modified box and whiskers plots are an outstanding tool for identifying outliers.

Interactive Discussion: Objectives

Individual Work: Draw modified box and whiskers plots both on paper and with TI-83 calculators using simulated data.

Homefun: prob. 1.55, 1.59:

Essential Question: Ideally, how many data points in a set of data are needed to characterize spread?

Lesson 6

Key Concept: Measuring spread - Range, standard deviation, and IQR

Standard Deviation

26. Calculate range.
27. Write the formula for standard deviation from memory and explain its meaning.
28. State how standard deviation and variance are related.
29. Calculate standard deviations by hand and with a calculator
30. Note the difference between S and sigma.
31. Be as one with the 3 points about standard deviation in the magic box on page 49.
32. State why the standard deviation is a better indicator of spread than range.
33. State an approximate relationship between range and standard deviation.

Purpose: Understand the pros and cons of various spread measuring techniques.

Interactive Discussion: Objectives. Which measure of spread is least affected by skew and outliers? How would this affect you **analysis** of and **conclusions** drawn from data?

Seat Work: Calculate ranges, standard deviations, and variances both on paper and with TI-83 calculators using simulated data.

Homefun: work exercises

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1 Distributions	2 N-Distribution	3 Regression	4 NL Regression	5 Data
Unit Plan		Practice Test		Study Guide

AP Statistics Standards

I. Exploring Data: (continued).

II. Anticipating Patterns:

C. The normal distribution

1. **Properties of the normal distribution**
2. **Using tables of the normal distribution**
3. **The normal distribution as a model for measurements**

Objectives	Activities
Essential Question: Does all data fit a bell curve and why is this important in the design, analysis, and conclusions drawn from a study?	**Lesson 1** **Key Concept:** What is a normal

The Normal Distribution

1. Define density curve.

    - Models the a distribution
    - Above horizontal axis
    - Area under it = 1
    - Area under part of the curve = probability

2. Describe the normal distribution.

    - Symetrical
    - Bell Shaped
    - Range negative infinity to positive infinity

3. Be as one with the 68-95-99.7 Rule for normal distributions. ***Note: in the 1st part of the test for this chapter, you will not be permitted to use calculators or tables. The entire test will be based on the above rule.***

4. Be as one with the 0-75-89 Rule for *any* distribution. ***Note: this is the worst case situation for any distr.***

Chebyshev's Rule: p = (1 - 1/k^2)

5. State where the 2 inflection points fall on a normal distribution. *This will help you draw the norm. distr.*
6. Correctly use N(mean, sigma) notation.
7. State the effects on a normal distribution of increasing standard deviation.
8. According to the normal distribution, what is the probability of obtaining an exact value for a data point.

distribution?
Purpose: Lay the foundation for using the normal distribution.

Interactive Discussion: Objectives. Explain why the 68-95-99.7 Rule is critical for developing "normal distribution intuition".

Seat Work: Use the empirical rule to find various areas under the n-distribution.

Homefun: prob. 2.3, 2.5, 2.7, 2.9

Essential Question: What does the area under the "bell curve" mean?

Finding Areas Under the Normal Distribution

9. Calculate z-scores.

   z = distance from mean in std dev units

10. Estimate the probability of obtaining a range of values by using the normal distribution. *(find the area under the curve)*

    Probability = area under curve

11. Find z-scores corresponding to Q1 &Q3 of a box and whiskers plot for normally distributed data.
12. Using a normal distribution, estimate a critical value given the probability of finding a value as extreme or more extreme (a tail area).
13. Judge the normality of a distribution by examining histograms, stem plots, dot plots, or box and whiskers plots.

    - Bell shaped
    - Symmetrical
    - 68-95-99.7 Rule

14. Judge the normality of a distribution using a normal quantile plot on a TI-83.

Homefun: prob. 2.12, 2.13, 2.17

15. Work quality control reject rate problems.
16. Convert a normal distribution into a

Lesson 2

Key Concept: Area under the normal distribution and how it relates to a box and whiskers plot.

Purpose: Understand when a data set is normally distributed.

Interactive Discussion: Objectives

2-person teams: Derive an equation for the z-score

Seat Work: Find the area under n-distribution using tables and the TI-83 calculator.

2-person teams:

- Derive the z-scores of Q1, Q2, LL, UL
- Draw a box and whiskers diagram of an n-distribution

cumulative frequency plot.

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1 Distributions	2 N-Distribution	3 Regression	4 NL Regression	5 Data
Unit Plan		Practice Test		Study Guide

AP Statistics Standards

I. Exploring Data: (continued)

D. Exploring bivariate data

1. **Analyzing patterns in scatterplots**
2. **Correlation and linearity**
3. **Least-squares regression line**
4. **Residual plots, outliers, and influential points**

Objectives	Activities
Essential Question: How can we establish and quantify a cause and effect relationship between two variables?	**Lesson 1** **Key Concept:** How to represent 2 variable data.
Chap 3 2 Variable Relationships 1. Identify the response and explanatory variables from a plot. 2. Identify positive and negative associations from scatter plots.	**Purpose:** Lay the foundation for **designing, analyzing** and drawing **conclusions** from studies using correlation and regression.
Note: an association does not establish cause and effect 3. Detect linear and non-linear relationships using scatter plots. 4. Judge the relative strength of a relationship by the amount of scatter around the curve of best fit. 5. Identify outliers on scatter plots. • Within the expected range X-values • Outside the expected range of Y-	**Interactive Discussion:** Objectives--How to interpret scatterplots. How to identify outliers.

values for a given X-value

6. Identify "influential outliers" on scatter plots.

   - Outside the expected range X-values
   - Often outside the expected range of Y-values

7. Make scatter plots using the TI-83 calculator and Excel.
8. State why any analysis of 2 variable (bivariate) data should always begin with a scatter plot regardless of which tools are used to further analyze the data.

Homefun: 3.11, 3.13

Essential Question: Why is it important to quantify correlation instead of just estimating it by looking at a graph?

Lesson 2

Key Concept: Correlation - strength of the relationship

Correlation

9. State the meaning of correlation and how it is typically indicated.

   - Strength
   - Direction
   - Linear Relationship
   - r

10. Be as one with the 5 facts about correlation. on p. 132.
11. Calculate r using the formula.

Purpose: Develop the ability to evaluate the strength of the relationship.

Interactive Discussion: The dog barked and the tree fell down. Was there an association. Was there causation

2-person teams: Perform correlations on SAT data using a TI-83. (This will carry over into the regression section.)

Homefun: prob. 3.19, 3.23

Essential Question: Why would we need to find a mathematical relationship between variables? Isn't correlation enough?

Lesson 3

Key Concept: Regression - finding the mathematical relationship between two variables.

Regression

12. Explain the difference between correlation and regression.
13. Perform regression/correlation analysis

Purpose: Obtain and understand regression equations.

Demo: Using **Fathom** software, demonstrate the reasoning process

with the TI-83 calculator, Excel Spreads sheets.

14. What type of error does least squares regression minimize?
15. Interpret regression equations.

    - Single
    - Multiple

15. Calculate y-bar using a regression equation, given x-bar.
16. Properly state the meaning of slope according to the official statistics definition. (p142)
17. Properly interpret the intercept. (example sales vs. advertising dollars, what are the sales with no advertising?
18. Describe the region where a given regression equation will give a meaningful association.
19. Define and decry the use of extrapolation.
20. Be aware that the point (x-bar, y-bar) is in the center of the regression line.
21. Solve problems using the equation for b on p.144.

$$b = r\,(s_x / s_x)$$

behind least squares regression analysis. Interactive Discussion: On objectives

2-person teams: See above

Essential Question: Would changing the units of the variables affect the R-square value and would this change the design, analysis, and conclusions drawn from a statistical study?

Lesson 4

Key Concept: The use and misuse of R-square

The Meaning of R-Square

22. State the meaning of SSM and SSE. Use them to calculate R-square.
23. Give the official interpretation of r-squared.

    - Use the proper magic words p.149
    - Evaluates the entire equation

24. Explain why care must be taken in using the official interpretation of r-squared.

Purpose: To understand how R-square is often overused as a measure of regression analysis "goodness".

Interactive Discussion: Objectives

2-person teams: See Stats investigation below

- Susceptible to outliers
- Data points furthest from the center of the line have more influence
- There may be no causative relationship between explanatory and response variables.

Homefun: prob. 3.33, 3.35

Stats Investigation: Meaning of R-Square - time approx 2 class periods (individual work)

Purpose: Determine if a regression analysis using random numbers can yield an r-square value of 50% or more.

Design: Set up a 2 variable simulation in Excel. For x-values use the integers 0 to 9. Generate a random number from 0 to 10 for the y-values. Run this simulation 100 times.

Analysis: Perform regression and correlation analysis on each simulation. Calculate the average r-square and record the highest r-squared value. Make a box and whiskers plot using all the r-squared data. Record the three highest r-square values obtained in the class. **Save the data sets** from your 4 regression/correlation results with the highest R-square value. You will use it again at the end of the year.

Conclusions: Answer the questions below provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions

1. Based on your data, does a high r-square value by itself indicate a meaningful association or causation?
2. Is the random number generator used in this investigation truly random?
3. Is it possible to get a high r-squared value merely from random events?
4. What does it really mean when we say that r-square represents the fraction of the variation in the values of y that is "explained" by the least squares regression of y on x? Discuss things like the SSM and SSE.
5. What do you conclude about r-squared values by looking at the box and whiskers plot?

Essential Question: Can a regression equation with a high R-square be inappropriate?

Lesson 5
Key Concept: Residual Plots.

Residuals

25. Define y-hat.
26. Calculate residuals using a TI-83 calculator.
27. State 2 ways to plot residuals.
28. State a major assumption behind all regression lines.

    Variability around the line is constant

29. Interpret residual plot patterns.

    - Random
    - Smiley or Frowning Face (X's Terms)
    - Pattern in the scatter

30. Make residual plots using a TI-83.
31. State the sum of the residuals.

Purpose: Understand when a given regression equation is appropriate.

Interactive Discussion: Objectives.

Individual work: Perform residual plots on TI-83 calculators and with Excel software.

Homefun: prob. 3.43, 3.45, 3.49

Stats Investigation: Determining if a Regression Equation is Appropriate - time approx 1 class periods (individual work)

Purpose: Determine if a linear regression equation is appropriate for two different situations.

Background: Commercial resistors follow ohm's law while light bulbs, due to their high temperatures do not. Ohm's law is as follows:

I = (1/R) V

Where: I = current, V= voltage and R = resistance.

Plotting I vs. V will theoretically yield a straight line passing through the origin.

Design: For this investigation a data set collected by physics students will be

provided by the teacher. **Note: We will return to this data later in the year.**

Analysis: Set up a least-squares linear regression analysis in Excel to find the association between current (response variable) and voltage (explanatory variable) for a commercial resistor and for a light bulb. Remember that this means a scatter plot as well as finding the slope, intercept, and R-square for the data. Set up the formulas needed to plot a residual plot and make such a plot for the two sets of data.

Conclusions: Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions:

1. Based on your data, does a high r-square value by itself indicate a meaningful association or causation?
2. Find the resistance value in Ohms for the commercial resistor.
3. Is a linear equation appropriate for the commercial resistor? How about the light bulb. Explain your answers.

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1 Distributions	2 N-Distribution	3 Regression	4 NL Regression	5 Data
Unit Plan		Practice Test		Study Guide

AP Statistics Standards

I. Exploring Data: Observing patterns and departures from patterns (continued)

D. Exploring bivariate data

5. **Transformations to achieve linearity:** logarithmic and power transformations

Objectives	Activities
Essential Question: Is everything we'd like to study and model linear? How would you alter the design, analysis, and conclusions drawn from a statistical analysis if you suspected a non-linear relationship?	**Lesson 1** **Key Concept:** Transforms **Purpose:** Create a linear plot from

Chapter 4 : 2 Variable Data Continued

Modeling Non-Linear Data

1. Explain how both power and exponential functions can be transformed into linear forms.
2. Give examples where an exponential regression model would be appropriate.

Growth or decay situations (response variable multiplied by a fixed amount in each time interval) such as:

- bacteria growth
- the growth of computer power
- radioactive decay

3. Give examples where a power regression model would be appropriate.

Scaling problems:

- Volume & mass scale with the cube of the scaling factor
- Area scales with the square of the scaling factor

4. Explain how to determine if an non-linear model is appropriate.

- theoretical basis such as 2 & 3 above
- random residuals

5. Explain why a model should not be selected on the basis of optimizing r-square.
6. Perform linear regression on transformed data and convert the results to the appropriate power or exponential equations.

Homefun: Read section 4.1; prob. 4.1, 4.4, 4.13, 4.17

nonlinear data

Interactive Discussion: Objectives. Explain the terms concaved upward and downward. Review logarithms and explain why the common transforms use them.

Demo: Use **Fathom software** to show that any nonlinear data set can be transformed if the equation is known.

Individual Work: work example problems

Essential Question: What is the most common form of extrapolation?

Interpreting Correlation and Regression

7. Decry the evils of extrapolation.
8. Identify possible lurking variable. (An important variable which is not included in the study.)
9. Name the most common lurking variable. (time)
10. State the pitfall of using averaged data. (The results look artificially good.)

Lesson 2

Key Concept: Extrapolation and lurking variables

Purpose: Understand how conclusions drawn from data can be disastrously wrong

Interactive Discussion: Objectives. Explain real growth curves--usually sigmoidal. How would you **design, analyze,** and draw **conclusions** when conducting a study that might involve a growth curve?

Essential Question: Can we ever be completely sure that causation exists?

Causation

11. State 4 possible explanations for getting a strong association based on regression analysis.

    - **Causation** --Sometimes it's true
    - **Common response variables** (affect both x & y variables), example: rum and Methodist Ministers
    - **Confounding variables** (affect the y variables but not the x), example: The shaman chants an incantation and five days later the patient gets well.
    - **Random chance**

12. Explain 4 ways that causation can be established.

    - multiple independent studies of different types
    - Account for control or eliminate lurking variables

Lesson 3

Key Concept: Extrapolation and lurking variables

Purpose: Understand how conclusions drawn from data can be disastrously wrong

Interactive Discussion: Objectives.

Video: use video describing the Surgeon Generals eventual ban on smoking

Establishing Causation

How would you go about **designing** studies, **analyzing** the results, and drawing **conclusions** if you wanted to establish causation?

How was this done to establish a

- Turn the causative variable on and off
- Develop a plausible theory

Homefun: Read section 4.2; prob. 4.25, 4.27, 4.29

link between smoking and cancer?

Why is it important for the results of a study to be repeatable by others?

How do things like outliers, skew, and lurking variables affect a study's repeatability?

AP Statistics Standards

I. Exploring Data: Observing patterns and departures from patterns (continued)

E. Exploring categorical data

1. Frequency tables and bar charts

2. Marginal and joint frequencies for two-way tables

3. Conditional relative frequencies and association

4. Comparing distributions using bar charts

Essential Question: Can data sets be added together to obtain a larger sample size and hence more meaningful conclusion?

Lesson 1
Key Concept: Simpson's Paradox

Categorical Data & Simpson's Paradox

13. Create frequency tables for categorical data.
14. Convert the above tables into bar charts.
15. Use conditional distributions based on relative frequencies to establish associations.
16. Compare distributions using bar charts.
17. Interpret 2-way tables.

- 2 variables
- Convert to %

15. Interpret marginal distributions.

- 2 for each table, horizontal & vertical
- Histogram like
- Single variable only

Purpose: Understand how conclusions drawn from data can be disastrously wrong

Interactive Discussion: Objectives. Work through hospital example of Simpson's paradox.

Individual work:

- Work through Simpson's paradox worksheet provided by teacher.
- Use Titanic data to determine if the class of one's ticket had an association with the chances of survival.

Materials: Simpson's Paradox

15. Calculate and interpret conditional distributions
16. Analyze data for Simpson's paradox.

    - Conclusions based on parts can be reversed when considering the whole
    - Conclusions based on parts is more likely to be valid.

17. State two conditions which must exist for Simpson's Paradox to occur.

    - One or more lurking variables
    - Data from unequal sized groups being combined into a single group.

18. State how Simpson's paradox can be prevented.

    - Avoid combining data from unequal groups into a single study
    - Identify and include lurking variables in the study

Worksheet and Titanic data.

How would avoiding Simpson's Paradox impact the **design, analysis,** and **conclusions** drawn from a study? How should data be analyzed if it's collected from numerous disparate sources?

Homefun: Read Simpsons's Paradox - When Big Data Sets Go Bad

prob. 4.37, 4.39, 4.45

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1 Distributions	2 N-Distribution	3 Regression	4 NL Regression	5 Data
Unit Plan		Practice Test		Study Guide

AP Statistics Standards

II. Sampling and Experimentation: Planning and conducting a study (10% –15%)

A. Overview of methods of data collection

1. **Census**
2. **Sample survey**
3. **Experiment**
4. **Observational study**

B. Planning and conducting surveys

1. **Characteristics of a well-designed and well-conducted survey**
2. **Populations, samples, and random selection**
3. **Sources of bias in sampling and surveys**
4. **Sampling methods, including simple random sampling, stratified random sampling, and cluster sampling**

C. Planning and conducting experiments

1. **Characteristics of a well-designed and well-conducted experiment**
2. **Treatments, control groups, experimental units, random assignments, and replication**
3. **Sources of bias and confounding, including placebo effect and blinding**
4. **Completely randomized design**
5. **Randomized block design, including matched pairs design**

D. Generalizability of results from observational studies, experimental studies, and surveys

Objectives	Activities
Essential Question: Can bad data form a poorly designed study be corrected with good statistical analysis in order to draw meaningful conclusions? **Chapter 5 : Producing Data** **Designing Samples** 1. Distinguish between population and samples and tell which one forms the basis of statistics. 2. Define bias (p246). 3. Define voluntary response and explain why it invariably produces biased results. 4. Identify when confounding is present. 5. Explain why confounding and bias make statistical inference impossibly. (Inference implies that there is no other reasonable explanation for the data. 6. State the key difference between a statistical study and a non-statistical study. 7. State the two basic forms of statistical	**Lesson 1** **Key Concept:** Designing systematic, statistically sound ways to collect data **Interactive Discussion:** Objectives **Seat Work:** generate random numbers using a random number table.

studies.

- Observational / Survey
- Experiment

8. Describe an SRS and state what it seeks to eliminate. (With a sample size of n, every ***set*** of n individuals has an equal chance of being chosen.)
9. State how an SRS is formed.

- Label
- Table

8. Use a table of random digits to create an SRS.
9. State the problem which the magic word "randomization" solves.
10. State the primary weakness of an SRS.

- Variability from study to study

Homefun : prob. 5.1, 5.3, 5.7

Essential Question: What is the difference between preventing variability and preventing bias in a statistical study?

12. Describe the key method for preventing variability in observational studies and surveys.
13. Describe how a multistage sample design is used for preventing variability.
14. Describe 4 ways to do a perfectly good job of sampling and still get worthless results (p. 257).

- **Under coverage** - Leaving groups out of the sample selection process
- **Non response** - Mr. Rogers Syndrome
- **Response Bias** - a) Intimidating

Lesson 2

Key Concept: Variability and bias--surveys
Purpose: How to prevent the above in surveys

Interactive Discussion: Objectives. How does the prevention of variability and bias affect the **design, analysis**, and **conclusions** drawn from a study? How does it impact attempts to establish causation?

Group Work: Correctly design a survey for Y high. Consider how the data will be collected and analyzed.

interviewer b) Intimidating question
- **Wording Effects** - Asking the wrong question or biasing the result.

Homefun: 5.11, 5.12, 5.15, 5.20, 5.23

Essential Question: Why are experiments considered more convincing than observational studies?

Lesson 3

Key Concept: Variability and bias--experiments

Experiments

15. Correctly use the following terms:
    - Experimental unit/subject
    - Treatment
    - Factor/level
    - Placebo effect
    - Control group
    - Completely randomized design
16. State the magic word which is used in all experiments and state why and how it is used
17. Explain the conditions which make an effect **statistically significant** (p. 276).
18. Be as one with the three basic principles of experimental design.
    - **Control** - effects of lurking variables
    - **Randomization** - prevents sampling bias
    - **Replication** - collect numerous data points
19. Describe how double blind testing is used.
20. Discuss the ethical considerations of double blind testing.
21. Correctly use blocking in an experimental design.
22. Explain why blocking reduces study to study variability.

Purpose: How to prevent the above in experiments.

Interactive Discussion: Objectives. What are the differences in **design, analysis,** and the certainty of **conclusions** between observational studies and experiments?

When we say blocking and stratification reduce variability, what are we talking about and why is it important?

Group Work: Correctly design an experiment for the AP test fish tank example.

23. State the problem that blocking does not solve.
24. Set up matched pairs designs.

Homefun: 5.31, 5.38, 5.41, 5.47

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
	Chap 6 Probability	7&8 Binomal Distr	9 Sampling Distr	10 Conf Intervals
Unit Plan	Practice Test		Study Guide	

III. Anticipating Patterns: Exploring random phenomena using probability and simulation (20% –30%)

A. Probability

1. **Interpreting probability, including long-run relative frequency interpretation**
2. **"Law of Large Numbers" concept**
3. **Addition rule, multiplication rule, conditional probability, and independence**
4. **Discrete random variables and their probability distributions, including binomial and geometric**
5. **Simulation of random behavior and probability distributions**

B. Combining independent random variables

1. **Notion of independence versus dependence**
2. **Mean and standard deviation for sums and differences of independent random variables**

Objectives	Activities
Essential Question: Can you win money in Las Vegas? **Ch 6.1, 6.2 -- Randomness and Probability Models** 1. State the basis for all predictions based on probability models.	**Lesson 1** **Key Concept:** Predictions from probability models tend to match results from real data if the sample size is large enough. **Purpose:** Learn the basic principles and vocabulary of probability.

The Law of Large Numbers

2. Name the two factors which exist in a random phenomenon (p. 314).

    - Uncertain outcome
    - Regular distribution of outcomes with a large number of trials. (Note that even randomness follows a pattern.)

3. State the difference between an outcome, an event (p. 324), and a sample space (p. 318).
4. Use tree diagrams to identify sample spaces.
5. Use the multiplication rule to calculate the number of outcomes.

omefun: 6.8 -- Read section 6.1, 6.2

Interactive Discussion: Objectives. How could you **design** a simulation of Las Vegas gaming such as roulette? How would you **analyze** it? Would your **conclusions** be valuable in the real world?

Problem solving (individual):

- Draw the tree diagram of outcomes for rolling a pair of dice.
- Draw a tree diagram for the sample space obtained by throwing one coin and one die. Identify an event and calculate the probability that it will occur. Are the various events independent?

Essential Question: Is anything in nature truly random?

Ch 6.2 -- Probability Models

6. Calculate the number of outcomes using sampling without replacement. (Hint: use a tree diagram.)
7. **Draw a Venn diagram for disjointed events** and give examples.
8. **Draw a Venn diagram for independent events** and give examples.
9. Correctly apply the 5 probability rules (p. 324-325 and p.341)

    - Range of values: 0 to 1
    - Sum of probabilities for all outcomes = 1.
    - Prob. of not happening = 1-(prob. of happening)
    - **Addition Rule - 2 disjointed events**, prob. of one **or** the other occurring is the sum of individual probabilities.

Lesson 2

Key Concept: Probabilities for independent events can be calculated using a few simple rules.

Purpose: Be able to Calculate Probabilities for independent events

Interactive Discussion: 10 monkeys are in a barrel, 3 are dead and 7 are alive. What is the probability of removing a dead monkey? After 3 dead monkeys have been removed (and not replaced), what is the probability of removing an additional dead monkey. Run the same mind experiment with replacement.

P(A **or** B) = P(A) + P(B)

- **Multiplication Rule- 2 independent events,** the probability of one **and** the other occurring is the probabilities of both multiplied

P(A **and** B) = P(A) P(B)

Problem Solving (Teams of two): Calculate the probability of hopeless failure for small groups of 2-7. Assume that the odds of a single person being right is 60%.

11. **Draw a Venn diagram showing the complement (A^c)** of an event (A).

Homefun: 6.8 6.9, 6.11, 6.13, -- Read section 6.3

Stats Investigation: The Spinning Wheel (p. 310 Teams of two)
Purpose: Determine if actual results match predicted results better with large sample sizes. **Design:** Read the instructions on p. 310. Use a TI-83 and a die to generate groups of 3 random numbers. Record 20 groups of random numbers for each type of randomization. **Analysis:** Calculate the probability of getting at least one number in the correct order for each set of 20 experiments and record the probabilities on the board. Calculate average and range for the two data sets from individual teams. Note: the two averages are also calculated probabilities. Use a tree diagram to list the possible outcomes. From the tree diagram, calculate the theoretical probability of having at least one of 3 numbers in order for the above experiment. **Conclusions:** Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section. **Questions:** 1. Was there a significant difference (in other words one probably not due to random chance) in the calculated probabilities for the two different types of random number generators? 2. How closely did the calculated probability of the entire group match the predicted? 3. How did the calculated probabilities from individual data sets (n=20) compare to the three probabilities calculated by combining data from all the groups? **Resources/Materials:** TI-83's, 10 dice

Ch 6.2 -- Probability Models

12. Calculate the probabilities of events for equally likely outcomes.

    - drawing a random integer from 0-9
    - flipping coins
    - rolling a die

    $$P(A) = \frac{\text{count of outcome in A}}{\text{count of outcomes in S}}$$

13. Create distributions representing the sample space of a random process both for outcomes with equal probability and outcomes with unequal probabilities (see example 6.6, p. 320 and see **How to Design Small Decision Making Groups**).
14. Define a union (A **or** B) or intersection (A **and** B) for a collection of events (p.341).

Homefun: 6.31, 6.35

Solving Probability Problems With the Equations

16. Calculate the probabilities for unions (unions correspond to **or**-statements, p.341-343) with independent and disjointed events.

Homefun: 6.37, 6.39

Key Concept: Sample spaces can be represented using distributions.

Purpose: Relate distributions to tree diagrams and sample spaces.

Warm up (Individual): calculate the probabilities of getting 3 heads with 3 coin tosses, four heads with 4 coin tosses and 5 with 5.

Interactive Discussion: Objectives

Problem Solving (Teams of two): create 2 distributions for the decision accuracy of 4 person groups assuming individual accuracies of 50% for the first distribution and 60% for the second.

Lesson 4

Key Concept: Probabilities with Unions

Purpose: Solve probability problems with both disjointed and non-disjointed independent events.

Interactive Discussion: Objective 16. Use coin and election (control of House, Senate) examples

P(A **or** B) = P(A) + P(B) - P(A and B)

Disjointed: 2 coins both heads **or** both tails

Not Disjointed: 2 coins at least one head **or** at least one head tail

Problem Solving (Teams of two): Work problems 6.39 and 6.40

18. Calculate conditional probabilities for unions (contain **or**-statements, p.348-349).
19. Calculate conditional probabilities for intersections (contain **and**-statements, p.348-349).
20. Determine if 2 events are independent using the test P(B|A) = P(B).

Key Concept: Conditional Probability for unions and intersections

Purpose: Solve probability problems when the events are not independent. In other words they involve events which are conditional.

Interactive Discussion: Objective 17. poor, not poor; go to college, not go to college.

P(A **and** B) = P(A) * P(B|A)	
P(B/A) = P(A **and** B)/P(A)	Used as test for independence. When P(B/A) = P(B) the events are independent.

Problem Solving (Teams of two): Work problems 6.41, 6.46, 6.48

Essential Question: What are the dangers of massive drug abuse testing programs ?

Solving Probability Problems With Diagrams

- Why you should be a tree hugger

18. Use trees for probability calculations (see **The Probability of Penalizing the Innocent Due to Bad Test Results**) with independent events. P(B|A) = P(B)
19. Use trees for probability calculations with dependent events. P(B|A) ≠ P(B)
20. **Use Venn Diagrams and probability equations** to explain the differences between the following types of events

    - disjointed
    - independent
    - conditional

24. **Relate Venn Diagrams to tree diagrams**. (the equations and trees give the same information. Both can be related to Venn Diagrams)

Key Concept: Predictions from probability models tend to match results from real data if the sample size is large enough.

Purpose: Learn the basic principles and vocabulary of probability.

Interactive Discussion: Objectives. How do probabilities relate to areas under the normal distribution curve?

How can the n-distribution be used in probability problems?

When calculating probabilities, when might it be important to know if the data used comes from a skewed distribution?

Board Work (individual): Solve tree

25. Work problems that combine the normal distribution with probability relationships. diagram problems and problems combining the n-distribution with probability.

Homefun: 6.41, 6.43, 645, 6.49,6.53, 6.55

Mr. Rogers - AP Statistics Objectives

AP Statistics Standards

III. Anticipating Patterns: (continued)

A. Probability

1. **Interpreting probability, including long-run relative frequency interpretation**
2. **"Law of Large Numbers" concept**
3. **Discrete random variables and their probability distributions, including binomial and geometric**
4. **Discrete random variables and their probability distributions, including binomial and geometric**
5. **Simulation of random behavior and probability distributions**
6. **Mean (expected value) and standard deviation of a random variable, and linear transformation of a random variable**

B. Combining independent random variables

1. **Notion of independence versus dependence**
2. **Mean and standard deviation for sums and differences of independent random variables**

Objectives	Activities
Essential Question: Can humans simulate a random process and why is this an important issue?	**Lesson 1** **Key Concept:** Random processes have predictable probability distributions yet true randomness is hard to simulate.
Ch 7.1, 7.2 -- Random Variables	
1. Be as one with the following vocabulary: random variable, discrete random variable,	**Purpose:** Learn the basic principles of probability distributions.

continuous random variable, density curve.
2. Plot discrete probability distributions for simple systems such as flipping coins.
3. Be as one with the law of large numbers (p. 389).

Average results of many independent observations are stable and predictable (p.392)

4. Describe the law of small numbers (p.392).
5. Find the mean of a discrete probability distribution (p. 387).

Homefun: 7.5, 7.7, 7.19, 7.23 -- Read section 7.1, 7.2

Warm up: Draw the probability distribution for flipping four coins. Find the area under the density curve for values of 1 through 3. Find the mean of the distribution.

Interactive Discussion: Objectives 1-2. Discuss the results of the warm up.

Stats Investigation (Teams of two):

Stats Investigation: Simulation of a Random Process (Teams of two)

Purpose: Determine if actual results of flipping a coin match results simulated by humans.

Design: Each partner in a two person team will record 100 fake coin toss tosses on a sheet of paper using H for heads and T for Tails. The team will then record 200 actual tosses. The experimenters will then circle the runs of 2 or more heads within the data both for the fake and real tosses. For example the following data contains 2 runs of 3, 1 run of 6, and 3 runs of 2 heads:

HHHTT**HHH**THTHTTT**HHHHHH**TTTT**HH**TTTT**HH**T**HH**

The experimenters will post the number of runs of each length on the board.

Analysis: When all results are posted for the class, the number of runs vs. size of runs is to be plotted for the classes fake and real data (two separate plots).

Conclusions: Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions: Answer with short paragraphs.

1. What is the biggest difference between the real and fake data. Speculate about why this is the way it is.
2. Are humans reliable at generating random numbers? Discuss this both in terms of the law of large numbers and the law of small numbers (p. 392)

Resources/Materials: pennies for flipping

Essential Question: Can standard deviations be added?

Ch 7.2 -- Random Variables

3. Apply the rules for means on p.396.
4. Calculate the standard deviation of a discrete random variable. (p.398).
5. Apply the rules for variance on page 400.

Homefun: 7.25, 7.31, 7.39 --

Lesson 2

Key Concept: Rules for working with random variables
Purpose: Apply the above rules

Interactive Discussion: Objectives. Could you **design** a simulation to see if standard deviation could be added? How would you **analyze** the results and would your **conclusions** be more valid than simply mathematically determining the answer from the equations?

Problem Solving (individual): Practice using the above rules

Essential Question: Do we live in a binary world?

Ch 8.1 -- The Binomial Distribution

6. Be as one with the Binomial Setting on page 416, ***SNIP***.
7. Calculate binomial distributions with a TI - 83 (npk - **n**o **p**igs **k**illed).
8. Use the binomial coefficient or combinations.
9. Calculate means and standard deviations for the binomial distr.

Homefun: 8.1 to 8.4, 8.9, 8.15, 817, 8.19

Lesson 3

Key Concept: The binomial distribution

Purpose: Understand how the world can be modeled as a binary system and how the binomial distribution can be used to analyze it.

Warm up: List various ways the world can be viewed as binary

Interactive Discussion: Objectives. Note that the binomial distribution is often used for **designing** and **analyzing** surveys resulting in **conclusions** (predictions) about election. Note that knowing the mean and probability of success for a binomial fixes both the mean and standard deviation as compared to the normal distribution where these are independent. Why does this make the binomial distribution so useful in **designing** studies?

Individual Work: Calculate the binomial coefficient first by hand and next using the combinations feature of the TI-83 calculator

- 3 coins in the fountain

analysis

Lesson 3

Essential Question: When is estimating how many tries before success an issue?

Key Concept: The geometric distribution

Purpose: Use the geometric distribution

Warm up: List various situation when estimating the first success is important

Interactive Discussion: Objectives.

Individual Work: Solve geometric distribution problems first by hand and next using the combinations feature of the TI-83 calculator.

- batter problem

Ch 8.2 -- The Geometric Distribution

10. Be aware of the key difference between binomial and geometric distributions.
11. Be as one with the Geometric Setting on page 435, ***SPIT***
12. Correctly use the geometric distribution for calculating probabilities with the TI-83 calculator.

omefun: 8.25 to 8.45

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
	Chap 6 Probability	7&8 Binomal Distr	9 Sampling Distr	10 Conf Intervals
Unit Plan	Practice Test		Study Guide	

AP Statistics Standards

III. Anticipating Patterns: (continued)

D. Sampling distributions

1. **Sampling distribution of a sample proportion**
2. **Sampling distribution of a sample mean**
3. **Central Limit Theorem**
4. **Sampling distribution of a difference between two independent sample proportions**
5. **Sampling distribution of a difference between two independent sample means**
6. **Simulation of sampling distributions**

Objectives	Activities
Essential Question: Can we describe	**Lesson 1**

the shape of a distribution of many samples of the same size even without knowing what the population's distribution's shape is like?

Key Concept: Characteristics of the sampling distribution

Ch 9

Purpose: Form the foundation for tests of significance

Interactive Discussion: Objectives.

Stats Investigation (Teams of two, see below):

1. Describe the difference between a parameter and a statistic and give examples.
2. Explain the use of p and p-hat.
3. Given a sampling distribution, explain its meaning (p. 459).
4. Determine if a statistic is unbiased (p. 468).

   Unbiased: sampling distr. mean = pop. mean (p.464)

5. Compare variability to bias (targets p.465).

Which is worse, variability or bias?

6. State how the variability of a statistic changes relative to population size.
7. State which distribution has more variability, a sampling distribution or the population distribution it is based on?
8. Calculate standard error (p. 587).

(**standard error)=(s of samp. distr.)**

=(s of pop.) / (n^.05)

Homefun: 9.1 to 9.4 -- Read section 9.1 to 9.2

Stats Investigation: Central Limit Theorem
Purpose: Does the variability in the sampling distribution actually decrease as predicted by the central limit theorem?

Design: Go to http://intuitor.com/statistics/CentralLim.html and read the write up. Open the *Central Limit Theorem Applet* and set the number of samples slider to its maximum (max=2010). Run at least 10 simulations using a variety of sample sizes from 1 to 100. From the sampling distribution plot, record the sample size, standard deviation, and standard error.

Analysis:

1. Make a scatter plot of standard deviation vs. sample size. Perform linear regression and power regression on the data in the plot. Also make a residual plot for both forms of regression.
2. Make a second scatter plot of standard deviation vs. standard error. Perform linear regression, report the r-square value, and make a residual plot for this data.

Conclusions: Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions:

1. Which regression equation best fits the data in Analysis 1? Explain why?
2. Explain the r-square value for the most appropriate equation in analysis 1.
3. What is the predicted equation in analysis 2 and how does it compare to the regression equation.
4. In analysis 2, what is the difference between the standard deviation of the sampling distribution and the standard error. Why do they have almost identical values?
5. What is the difference between sample size and number of samples. Describe how increasing them influences the sampling distribution.
6. What is the difference between the central limit theorem and the law of large numbers.

Ch 9

9. Calculate the mean and standard deviations of a binomial distribution for proportions.
10. Compare the binomial distributions for proportions to the binomial distributions for counts.
11. State why the binomial distribution is basically always a sampling distribution.
12. State the two rules of thumb which must be met before using the normal

Lesson 2

Key Concept: The normal approximation of the binomial distribution
Purpose: Apply the above

Warm up (Teams of two):

1. Match sample size to the distribution
2. Given sets of sample and population means for determine which ones are biased

approximation of the binomial distribution (pp. 473 & 475).
13. Calculate the mean and standard deviation of a non-binomial sampling distribution (p. 483).
14. State how the central limit theorem applies to sampling distributions.
15. State how the law of large numbers applies to sampling distributions.

Homefun: 9.15, 9.17, 9.19, 9.25, 9.27, 9.31, 9.37 -- Read section 9.3

Interactive Discussion: Objectives. Note that with a large enough sample size the normal distribution of the sampling distribution starts looking spike-like. Why is knowing that the sampling distribution is approximately normally distributed so important to the **design, analysis,** and **conclusions** drawn from a study? Why would it be a problem if the sampling distribution were instead highly skewed?

Problem Solving (Teams of two): Use the central limit theorem applet to understand the effects of sample size on normality.

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
	Chap 6 Probability	7&8 Binomal Distr	9 Sampling Distr	10 Conf Intervals
Unit Plan	Practice Test		Study Guide	

AP Statistics Standards

IV. Statistical Inference: Estimating population parameters and testing hypotheses (30% –40%) *Statistical inference guides the selection of appropriate models.*

A. Estimation (point estimators and confidence intervals)

1. **Estimating population parameters and margins of error**
2. **Properties of point estimators, including unbiasedness and variability**
3. **Logic of confidence intervals, meaning of confidence level and confidence intervals, and properties of confidence intervals**
4. **Large sample confidence interval for a proportion**
5. **Large sample confidence interval for a difference between two proportions**
6. **Confidence interval for a mean**
7. **Confidence interval for a difference between two means** (unpaired and paired)

Objectives	Activities

Essential Question: Is there a way to get any researcher in the world to agree agree that your conclusions are reasonable even if they disagree with them?

Lesson 1

Key Concept: Statistical inference

Purpose: Gain an intuitive understanding of what statistical inference is.

Stats Investigation (Teams of two):

Interactive Discussion: Objectives What two things does a confidence interval give? What parts of the **design, analysis,** and **conclusion** making processes are related to statistical inference?

Resources/Materials: two decks of cards

Ch 10.1 Statistical Inference

14. Correctly use the term statistical inference.

    To draw a conclusion from data in a formal manner using well defined procedures

15. Give the quick and dirty explanation of a confidence interval.

    An estimate of a parameter based on a statistic accounting for uncertainty.

Homefun: -- Read section 10.1

Stats Investigation 10.1: Estimating a Proportion

Purpose: Determine a reasonable way to estimate a proportion for a population.

Design: Write a rational hypothesis for what the proportion of red cards is in a deck of cards.

Draw 10 samples with n=2 and 10 samples with n=20 from a deck of cards. Remove a card one at a time. Replace and shuffle the deck each time a card is drawn. Do this twice for each sample of two and twenty times for each sample of 20. Record the proportion of red cards for each sample.

Analysis: Create an interval around each sample that you feel has a high chance of containing the mean. Record your reasons for making the interval. Make two plots of all the intervals. One for n=2 and one for n=20. The plots should look like the one on page 511 minus the normal distribution picture.

Repeat the process for a second deck of cards.

Conclusions: Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions:

1. Which sample size tended to have the widest interval?

2. Did every interval contain the true mean?

Essential Question: How can you express a measurement in an internationally accepted manner?

Lesson 2
Key Concept: Confidence intervals

Ch 10.1 Confidence Intervals

16. State the type of distribution which confidence intervals are based on and sketch the appropriate picture of a confidence interval.
17. State the 3 parts of a confidence interval and explain their meaning.
    - Confidence level
    - Estimate
    - Margin of error
18. Describe what happens to the margin of error as confidence level is increased.
19. Formally state the meaning of a level C confidence interval. (magic box p. 514)
20. Tell why the margin of error is not a measure of inaccuracy in the data.
21. State which form of error a margin of error represents.
    - Sampling Error
22. Be as one with the cautions listed on page 524.

Purpose: Gain an intuitive understanding of what a confidence interval is and how it gives more information than just an estimate.

Interactive Discussion: Objectives.

Homefun: 10.1, 10.3, 10.9, 10.11, --

Stats Investigation 10.2: Margin of Error

Purpose: Determine the effects that changes in confidence level, sample size, and standard deviation have on margin of error.

Design: Use the equation for margin of error to make three different plots as follows:

1. **Margin of error vs confidence level** with sigma = 1 and n = 100. Vary confidence level from 60% to 99%.
2. **Margin of error vs sigma** with confidence level = 95%, and n = 100. Vary sigma from 1 to 5.
3. **Margin of error vs sample size** with confidence level = 95% and sigma = 1. Vary n from 100 to 500.

Analysis: Look for trends in the above plots. Write a brief explanation of why this

investigation is not a true statistical simulation.

Questions /Conclusions: Answer the questions the purpose by answering the following questions:

1. What is the effect of doubling the sigma and why?
2. What is the effect of doubling the sample size and why?
3. What is the effect of increasing the confidence level from 90% to 95%?
4. Of the three items mentioned above, which one(s) are under the control of the experimenter?
5. What type of error does the margin of error represent?
6. Why is margin of error not a measure of accuracy?
7. What sampling technique must be used if valid confidence intervals are calculated?
8. Is a high confidence level necessarily more meaningful than a low confidence level?

AP Statistics Standards

IV. Statistical Inference: Confirming models (continued)

B. Tests of significance

1. **Logic of significance testing, null and alternative hypotheses; p-values; one- and two-sided tests; concepts of Type I and Type II errors; concept of power**
2. **Large sample test for a proportion**
3. **Large sample test for a mean**

Essential Question: How does the US justice system compare to statistical analysis?

Lesson 3

Key Concept: Significance Testing or Hypothesis Testing

Ch 10.2 Significance Testing

1. State the question asked by a significance test and the two possible answers. *(Is there clear evidence of an effect?)*
2. State a generic null hypothesis. *(There is no evidence of an effect.)*
3. State a generic alternative hypothesis. *(There is clear evidence of an effect.)*
4. Give the null and alternative hypothesis for the American justice system.
5. Define P-value.

Purpose: Learn the format for one of the key inference tools in statistics.

Interactive Discussion: Objectives. Relate the various parts of a test of significance to **design,**

6. State the statistic used for indicating the level of significance.
7. State the type of distribution used for tests of significance.
8. Describe one tail and two tail tests from the standpoint of the null hypothesis and the p-values.
9. Be as one with "z-test for a population mean" on page 546
10. Perform "z-test for a population mean" in the following ways:

    - by hand (using the calculator only for basic mathematics) using z-tables.
    - by hand, using the calculator only for basic mathematics and finding areas
    - using the hypothesis testing features of the TI-83
    - using Minitab software

Homefun: 10.29 - 10.32, 10.35, 10.37

analysis, and the drawing of **conclusions**.

Web Page Resource: Type I and II Errors-Making Mistakes in the Justice System

Essential Question: Can a statistically significant hypothesis have no practical value?

Lesson 4

Key Concept: Statistical standards of judgment

Ch 10.2 Alpha Levels

11. Define statistical significance in terms of alpha.
12. Use one and two tailed tests of significance.
13. Use alpha to evaluate statistical significance.(p.542)
14. Use a confidence interval for a significance test.
15. Describe the difference between statistical and practical significance.
16. State when statistical inference is not valid.

Homefun: 10.45, 10.53, 10.55

Purpose: learn how to use alpha levels in hypothesis tests.

Interactive Discussion: Objectives. What parts of the **design, analysis,** and **conclusion** drawing processes does the selection of alpha most impact?

Essential Question: How many ways can you be wrong?

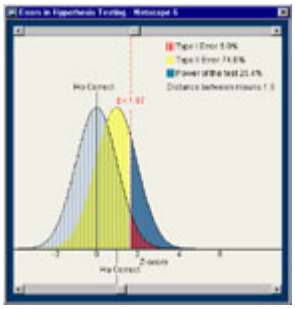
Lesson 5

Key Concept: Type I and type II errors, power of the

test

Ch 10.3 - 10.4 Types of Errors

16. Pass the ultimate test of true statistics nerdhood: (Explain the difference between type 1 and type 2 errors.)
17. Explain what a type 1 and type 2 error is for the American Justice System.
18. Generate a truth table for a hypothesis test.
19. State how alpha relates to the type 1 error.
20. Name the hypothesis which is considered true when calculating alpha.
21. Name the hypothesis which is considered true when calculating beta.
22. Identify the areas representing the probabilities of type 2 and type 1 errors on a diagram of a hypothesis test showing a hypothetical sampling distribution.



23. Calculate alpha and beta.
24. Calculate the power of a hypothesis test.
25. Plot and interpret a power curve for a hypothesis test.
26. State how power can be applied to quality testing in manufacturing.

Purpose: develop an intuition for the elements that affect the above.

Interactive Discussion: Objectives. Is there anything in either the **design** or **analysis** process that could minimize both the possibilities of a type I or II error in the **conclusion**?

Web Page Resource: Type I and II Errors-Making Mistakes in the Justice System .Use the applet provided to simulate various types of errors

Homefun: 10.67, 10.69, 10.71

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
11 t-Test	12 Inf for Prop	13 Chi Test	14 Regession HT	
Unit Plan	Practice Test		Study Guide	

AP Statistics Standards

IV. Statistical Inference: Confirming models

A. Confidence intervals

IV. Statistical Inference: Confirming models (continued)

B. Tests of significance

C. Special case of normally distributed data

1. **t-distribution**
2. **Single sample t procedures**
3. **Two sample (independent and matched pairs) t procedures**

Objectives	Activities
Essential Question: How can we account for the greater uncertainty of analyzing data when sample size is small and little is known about the population? **Ch 11** 1. State the 2 assumptions for drawing inferences about a population mean when sigma is not known. • SRS • Normal Distribution of the population 2. Calculate standard error (standard deviation of the sampling distr. = s/(n^.5) for a sample. 3. Explain when a t statistic is used rather than a z score. 4. Calculate t statistics. 5. State the degrees of freedom for a t test. 6. Construct confidence intervals using the t statistic. 7. Perform one sample t-procedures: • by hand (using the calculator only for basic mathematics) using t-tables. • by hand, using the calculator only for basic mathematics and finding areas • using the hypothesis testing features of the TI-83 • using Minitab software	**Lesson 1** **Key Concept:** How is t-Distribution used? **Warm up:** What conditions must be met to use a z-test? How would you analyze your data to make this determination? **Interactive Discussion: Objectives.** What conditions must be met to use a t-test and how does this differ from a z-test? How would you deal with these requirements in the **design** and **analysis** parts of a study to help insure the **conclusions** were meaningful? Assuming a t-test is called for, what kinds of tests should be performed before using a t-test?

Homefun: 11.7, 11.9, -- Read section

Essential Question: What is the single most powerful form of hypothesis testing and why ?

Ch 11

3. Apply the t-test of significance to matched pairs situations.

8. Explain what is meant by a robust test or confidence interval. (p-value or confidence interval changes little if assumptions violated.)
9. Be as one with the information in the "using t procedures" box on p.606.
10. Be aware that outliers are very harmful to the t-test.

Homefun: 11.13, 11.17, 11.21, 11.26

Lesson 2

Key Concept: How are Matched Pairs t-test used.

Warm up questions (individuals):

1. What two assumptions must be met to use a t-test?
2. How do you know you have normally distributed data?
3. What step must you take before using a t-test of any kind?

Interactive Discussion: Set up confidence intervals and hypothesis tests for matched pairs. When does the decision to use a matched pairs test need to be made, the **design** or **analysis** part? How does this decision affect the **conclusions**?

Problem Solving (Teams of two): Make a hypothesis test with matched pairs for a pre and post test situation--Spanish Camp

Essential Question: How can you determine if 2 populations differ if at the start you have no information?

Ch 11

11. State the assumptions made for two sample tests. (p. 619)

    - normally distributed
    - independent

12. Create confidence intervals and hypothesis

Lesson 3

Essential Question: How are two sample t-procedures -- no assumption made about the standard deviation being equal(p.624)--used?

Warm up questions (Individual):

1. What two types of tools are used for inference?
2. What type of distribution is used for making inferences?

test using two sample t procedures assuming that the sigmas of the two populations are unequal. This is the most conservative assumption.
13. Create confidence intervals and hypothesis test using two sample t procedures and the most conservative method of determining df. (the lower of n_1 -1 or n_2 - 1)
14. Be aware of the more accurate way to calculate df as shown on page 633.
15. Perform two sample hypothesis t-procedures on the TI-83.
16. State the key assumption required for using the pooled two-sample t-procedures.

    - The sigmas of the two populations are the same

Interactive Discussion: Set up a hypothesis test for two samples and point out how it differs from Z-test

Problem Solving (Teams of two): Make a hypothesis test with 2 sample and matched pairs for a pre and post test situation--return to the Spanish Camp data.

Homefun 11.33, 11.35, 11.37, 11.49:

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
11 t-Test	12 Inf for Prop	13 Chi Test	14 Regression	
Unit Plan	Practice Test		Study Guide	

AP Statistics Standards

IV. Statistical Inference: Confirming models

A. Confidence intervals

1. **The meaning of a confidence interval**
2. **Large sample confidence interval for a proportion**

B. Tests of significance

1. **Logic of significance testing, null and alternative hypotheses; p-values; one- and two-sided tests; concepts of Type I and Type II errors; concept of power**
2. **Large sample test for a proportion**
3. **Large sample test for a mean**
4. **Large sample test for a difference between two proportions**

Objectives	Activities
Essential Question: How does a confidence interval for proportions compare to one for means? **Ch. 12.1 Inference for Proportions** 1. State the meaning of p-hat. 2. Create a large sample confidence level for a proportion (p.665). 3. Perform a hypothesis test comparing a large sample proportion (p-hat) against a know population proportion (p). **Homefun: -- 12.7, 12.9 -- Read section 12.1**	**Lesson 1** **Essential Question:** How can inferences be draw for single sample proportions? **Warm up (individuals):** 1. What type of distribution is most useful for evaluating surveys for voting or yes or no questions? 2. Why does the binomial distribution start to look like a normal distribution when the sample size is large? **Interactive Discussion:** Objectives **Individual problem solving:** create confidence intervals for proportions
Essential Question: Assuming an SRS and given equal sized margins of error, is the sample size required to survey the entire United States substantially larger than the one for conducting the same survey in Greenville SC? How would this affect the design, analysis, and conclusion parts of a study? **Ch. 12.1, 12.2** 4. Calculate the desired sample size for a given margin of error in a proportion (p. 670). • Sample size is a key element in the **design** of data collection.	**Lesson 2** **Essential Question:** How can inferences be draw for two sample proportions? **Warm up questions (individuals):** 1. What's the difference between p and p-hat? 2. What are the two rules of thumb for modeling the binomial distribution with a normal distribution?

- The binomial distribution enables the calculation of a needed sample size.
- Being able to calculate needed sample size is a good reason to **design** survey questions in a binomial type format.

5. Create a confidence interval for comparing two sample proportions (p.681).
6. State the Ho used for comparing two sample proportions.
7. Calculate the pooled portion of successes using both samples.
8. Perform a hypothesis test for comparing two sample proportions (p. 684).

Interactive Discussion: Objectives

Individual problem solving: proportion problems

Homefun: 12.11, 12.25, 12.27, 12.35 -- Read Section12.2

Mr. Rogers - AP Statistics Objectives

Syllabus	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
11 t-Test	12 Inf for Prop	13 Chi Test	14 Regression	
Unit Plan	Practice Test	Study Guide		

AP Statistics Standards

III. Anticipating Patterns: (continued)

D. Sampling distributions

8. Chi-square distribution

IV. Statistical Inference:

B. Tests of significance

6. **Chi-square test for goodness of fit, homogeneity of proportions, and independence** (one- and two-way tables)

Objectives	Activities

Essential Question: Why would it be useful to have a quantitative way to test if data fits a particular distribution instead of merely relying on histograms, box plots, or normal quantile plots?

Ch. 13.1 Inference for Tables

1. Name 2 instances when a chi-squared test can be used. Note: chi is pronounced kie (rhymes with pie).

   - Goodness-of-fit (does data match a type of distribution?)
   - Inference for 2-way tables (tests the Ho that there is no relationship between row and column variables)

2. Describe the shape and range of the chi-squared distribution.
3. Determine degrees of freedom for a goodness-of-fit chi squared calculation. (df = n-1)
4. Calculate the chi squared statistic.

$$X^2 = \Sigma (O - E)^2 / E$$

5. Perform chi squared goodness of fit hypothesis tests.
6. Note that the hypotheses for a chi-squared test cannot readily be stated mathematically. They are as follows:

- **Null hypothesis:** The data's distribution and the reference distribution are not the same.
- **Alternative hypothesis:** The data's distribution and the reference distribution are the same.

Homefun: -- Exercises 13.9, 13.11 -- Read section 13.1

Lesson 1

Key Concept: Significance tests for distributions

Warm up: What would a normal distribution look like if it were squared?

Interactive Discussion: Objectives. Applying a Chi squared analysis to a table of data would tell if there is and relationship between rows and columns. If a chi squared test came back positive, what further **analysis** would need to be done to identify the specific relationship? What **conclusions** could be drawn if only the chi squared test were run?

Individual work: perform a chi-squared test on age data to determine if the population is indeed aging as a whole.

Essential Question: How can statistics be applied to genetic analysis in the real world?

7. Describe how the chi-squared test can be

Lesson 2

Key Concept: determining if data is random

used for determining if a set of data is not randomly distributed assuming that all events are equally probable.

Stats Investigation: statistical Analysis of Genome Data - computer lab using Minitab.

Essential Question: How can tables containing massive amounts of data be rapidly screened for relationships between the rows and columns?

Ch. 13.2 Inference for Two-Way Tables

8. Calculate expected results for tables (p.720).

expected = (row total X column total) / (table total)

9. Calculate chi-squared statistics for tables (p. 723).
10. Determine the degrees of freedom for a chi-squared (p. 724).

    df = (rows-1)(columns-1)

11. Perform hypothesis tests using chi-squared statistics.
12. Be able to read chi-squared computer print outs.

Homefun: -- Exercises 13.15, 13.17, 13.18 -- Read section 13.2

Interactive Discussion: Objectives.

2 person teams: see link to stats investigation

Lesson 3

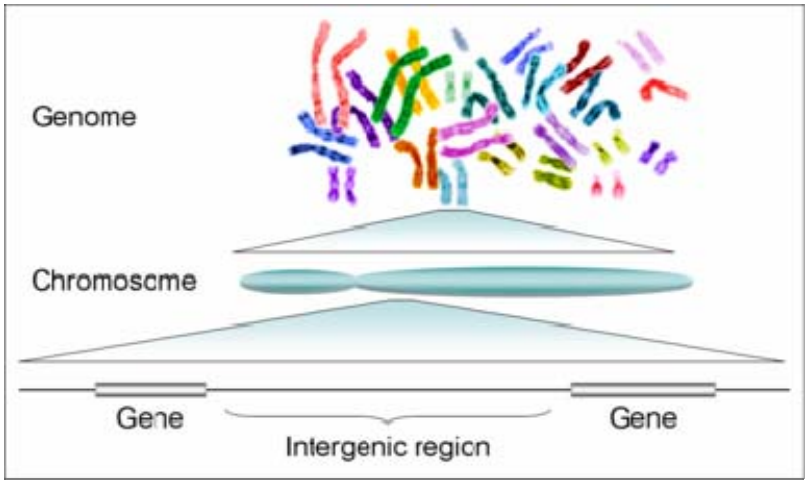
Key Concept: Analysis of large tables for relationships

Interactive Discussion: Objectives.

Individual work: perform a chi-squared test on a table containing drug data to determine if any of the drugs differ from the placebo.

- perform a chi-squared test by hand (using a calculator only for low level math)
- perform a chi-squared test using Minitab
- perform 2-sample t-tests as follow up to the chi-test

How genetic information is stored: The information stored in DNA is remarkably similar to a low level computer language written with 4 letters: **A, C, G, T.** Each of these represents a different base arranged in a strand of DNA. A pair of complementary strands are joined with hydrogen bonds and form the famous double helix.



DNA Sequences: To sequence DNA, the bases are read from a single strand a few pieces at a time and then assembled into a lengthy string of As, T, G, and Cs. An entire set of these strings for a single organism is called a Genome.

The human genome contains 3,164.7 million of these letters as compare to 3.6 million letters in the King James Version of the Bible. The letters in the genome are arranged like the chapters in a book on 24 separate DNA molecules each contained in a different chromosome. Each chromosome contains numerous genes which are considered the basic functional units of heredity.

Genes vs. intergenetic regions: Although the human genome is thought to contain about 20,000 to 25,000 genes, the genes make up only about 2% of the human genome. The remaining 98% of DNA is called the intergenetic regions. It purpose is unknown, although, even a relatively simple analysis indicates that it contains some form of information--maybe useless, scrambled or abandoned information.

Genome punctuation: Quite simply there isn't any. There are no spaces, no punctuation marks, or capital letters to denote the beginnings and endings of words in a sequence. Try reading the paragraph at right for a comparison of how this would look in English.

genomepunctuationquitesimplythereisntanytherearenospacesnopunctuationmarksorcapitalletterstodenotethebeginningsandendingsofwordsinasequencetryreadingtheparagraphatrightforacomparisonofhowthiswouldlookinenglish

Statistical analysis of genomes: Statistics is probably the single most important mathematical tool for understanding genetics. The investigation below will help show some of the many ways it can be used.

Stats Investigation: Statistical Analysis of Genome Data - time approx 2 class periods

Purpose: Determine if a statistical method can determine whether a sequence is randomly generated or contains actual information..

Background

1. Begin by visually comparing the two sequences shown in the link. One is a real genetic sequence, the other a randomly generated sequence. Can you tell the difference just by looking?
2. Go the article about Genome mining, "Meaningful sequences" and read through it to the end. (You will have to click on the small arrow icon on the right side below the text.)
3. Now go to "Analysing DNA" and follow the exercises using Geneboy all the way to the end of the article.

Design

Using Analyze composition, Singles feature of Geneboy obtain the distribution of frequencies for A, C, G, T bases for Genetic 1, for Random 1 and for Intergenetic 1. What type of distribution would be predicted using the law of large numbers for a random distribution of As, Cs, Gs, and Ts?

Analysis

1. Plot 2 bar graphs using the distribution of frequencies for A, C, G, and T of Genetic 1. In the first scale the y-axis from 15 to 35%, the second from 0 to 300%. Repeat the process for Random 1.
2. What statistical tool could you use to confirm whether the 2 sets of data are or are not random ? Perform this analysis using Minitab. Note, the question is not asking if the two distributions are different from each other.

Conclusions

Answer the questions below provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions

1. Did the changing the scale of the above graphs alter the perception of their meaning and could it alter the meaning of a different set of graphs?
2. Which is more likely to be universally accepted, conclusions based on graphs or conclusions based on statistical analysis? Why?
3. If genes are only 2% of the sequence how would you find one and establish that it was one using statistics?

AP Statistics Standards

V. Statistical Inference: Estimating population parameters and testing hypotheses (continued)

B. Tests of significance

7. Test for the slope of a least-squares regression line

Objectives	Activities

Essential Question: How can you express the uncertainty in the slope of a line?

Ch. 14.1 Inference for Regression

1. State the 2 inferences drawn when using regression data. (slope & intercept)
2. State the assumptions for regression inference.

- for any x value, y-data is normally distributed
- for any x value, the y-data's standard deviation is the same
- the means of the y-data distribution at any value of x form a straight line relationship: $\mu_y = \alpha + \beta x$

3. Calculate the standard error of the least squares regression line.

$$s = [\Sigma(y - \hat{y})^2 / (n-2)]^{1/2}$$

4. Calculate the standard error of the slope.

$$SE_b = s/[(\Sigma(x - \bar{x})^{1/2}]$$

Homefun -- -- Read section 14.1, prob. 14.5,

Lesson 1

Key Concept: How to evaluate the slope obtained in a regression analysis

Interactive Discussion: Why is the slope of the line is a big deal. It often has physical meaning

Famous slopes X has known:

- electrical conductivity
- COF
- spring constant
- plank's constant
- density
- g
- the perfect gas law constant
- etc. etc. etc

2-person teams: Using data from an Einstein's photoelectric experiment find plank's constant using regression analysis in Minitab. Calculate a confidence interval on the slope of the line (plank's constant). Compare this technique to IB error analysis techniques. Would you draw different **conclusions** about

14.7, 14.9

Plank's constant from the 2 different forms of **analysis**? What steps could you take to **design** a better experiment for collecting the data, even if you knew no physics?

Essential Question: How can you spot a meaningless regression anaysis?

Evaluating Regression Results

5. Generate a confidence interval for the slope. (n-2 degrees of freedom)

$$b \pm t^* SE_b$$

6. Calculate the t-value for a hypothesis test of the slope H_o: $\beta = 0$. (n-2 degrees of freedom)

$$t = b / SE_b$$

7. Perform a significance test for the slope of a least squares regression line.
8. Correctly perform least squares regression using Minitab.
9. Correctly interpret least squares regression computer output (such as from Minitab).

Lesson 2

Key Concept: Spotting a meaningless regression analysis

Interactive Discussion: Objectives. While it's possible to spot a meaningless regression analysis, it's not possible to tell for certain if an analysis is meaningful.

1. Does a well **designed** experiment with two variables always yield a meaningful regression analysis?
2. Can any form of **analysis** compensate for a poorly designed experiment?
3. When drawing **conclusions** is it ever possible to be completely sure that a regression equation is meaningful?

2-person teams: (see the stats investigation below)

Stats Investigation: How to Spot a Meaningless Regression Analysis - time approx 2 class periods (individual work)

Purpose: Determine if a regression analysis using random numbers that has a high r-square value can be detected with hypothesis tests on the slope and intercept.

Design: Remember the stats investigation you did earlier in which you determined that even random data can produce a high r-square value.

Analysis: Redo the regression/correlation analysis in Minitab on the 4 sets of data you saved and interpret the results. Be sure to take all the recommended steps for producing a statistically significant regression analysis.

Conclusions: Answer the questions provided by the teacher and draw conclusions relative to the purpose. Remember, do not include any information in the conclusion that was not collected and analyzed in the previous section.

Questions:

1. Based on your data, could you spot randomness with the hypothesis tests on the slope and intercept of the regression equation. Explain
2. Outline all the steps which should be taken to produce a regression/correlation analysis with the best chance of being meaningful.
3. Can a thorough statistical analysis of bivariate data, by itself fully establish that a regression result is meaningful? Explain

Essential Question: How can we evaluate or include possible lurking variable in a regression anaysis?

Ch. 14

10. Perform multiple linear regression analysis using Minitab and correctly interpret the regression equation, R^2, and the hypothesis tests.
11. State would you plot and interpret residuals for a multiple regression analysis.

Lesson 3

Key Concept: Control of variables in a study

Interactive Discussion:

Review the following from chap 5:

16. Be as one with the three basic principles of experimental **design**.

- **Control** - effects of lurking variables.
- **Randomization** - prevents sampling bias.
- **Replication** - collect numerous data points.

2-person teams: Perform a multiple linear regression

analysis on the SAT data used earlier in the year. Compare the results with those obtained earlier using 2 variable regression analysis. We have changed the method of **analysis**. What weakness in the **design** would **conclusions** drawn from the data suffer from? Would it be possible to detect this weakness from the analysis alone?

Stats Investigation: How to Create a Confidence Interval for the Resistance Value of a Commercial Resistor.

Purpose: Create a confidence interval for the resistance value of a commercial resistor.

Design: Use the resistance data from the investigation in Chap. 3 (the last time we talked about regression).

Analysis: Derive an estimate of resistance from the slope of the regression line done earlier. Develop a confidence interval using this estimate.

Conclusion: Explain why a resistance value confidence interval obtained using regression is superior to merely measuring resistance a single time with an ohmmeter.

Mr. Rogers - AP Statistics Objectives

Review and Special Topics

AP Statistics Scores		
AP Grade	**1997 % Correct**	**2002 % Correct**
5	68-100	68-100
4	54-67	53-67
3	41-53	40-52
2	29-40	29-39
1	0-28	0-28

Our AP Test Hero: Sergeant Friday

Why: He writes everything down -- "just the facts..."

The AP Statistics test is tough but doable!

AP Test review and special projects. The 4th quarter will be devoted to test preparation interspersed with a few special investigations. While a full quarter for review may sound like a lot, keep in mind that the test will be in May, hence we will lose about a month of possible classroom time. Also we lose a week for spring break in the 4th quarter.

Special Investigations

Stats Investigation: How to Spot a Table full of Fake Data

Background: Benford's law has been used in investigation to detect tax fraud. Basically, the law says that the first digit in of the data numbers from a table of real data will not have a uniform distribution. First go to Benford's Law Part I at http://www.intuitor.com/statistics/Benford's%20Law.html and read the write up to learn the nature of the first digit distribution.

Purpose: Determine if Benford's law works for real data.

Design: Work with a partner and find a large table of data on the internet. This could be economic, sports or other types of statistics. Design a method of collecting data from the table to determine if Benford's law is valid for the table. Make sure that your table is different from all the others in the class.

Analysis: Choose one or more methods of analysis from those you've learned and analyze the data. Collect the conclusions about whether Benford's law works, yes or no, for all the groups in the class.

Conclusions: First answer whether Benford's Law worked for your data table. Next,

using the classes data draw a general conclusion about Benford's Law. Use hypothesis tests to answer the questions.

Stats Investigation: Are the colors in M&M Uniformly Distributed

Note: To compare results across the class, all groups must have exactly the same type of M&Ms. These should be half pound or larger plain style with traditional colors.

Purpose: Determine if the colors in M&M Uniformly Distributed.

Design: Work in groups of four with a large bag of plain M&Ms. Design a statistical study to answer the above purpose.

Analysis: Choose one or more methods of analysis from those you've learned and analyze the data. Compare you results to the rest of the class.

Conclusions: Answer the purpose using a hypothesis test in accordance with the analytical tools you selected.

Free response question preparation

Daily Assignment: work and grade 3 short free response problems or one long problem per day from old AP Statistics tests. After working the problem a student will trade papers with a partner who will grade it according to the official AP rubrics. Each paper will be signed by the person who grades it, have the grade from 0 to 4, and list the time spent on the problem.

Success Tips:

- Use the test question's available space as a guide for how much to write. Write too much you'll go too slowly.
- Know the rubrics--the AP Stats test is about documenting correct methods, not merely recording right answers.

- State the restrictions on any analytical tool you use and be sure to justify the tool's use. For example: if you use a t-distribution you should establish that the data is approximately normally distributed.

Grading: Students will turn in the completed problem sets for each year at a rate of every 2 to 3 days. Each completed set will receive a grade of 5 points (out of 5 pts possible). Each completed set beyond the 6th set will receive 3 points of extra credit. Problem sets can be found online at the statistics section of the American College Board site.

Using this system students will not only get feedback on their solutions but gain the benefit of evaluating another person's solution according to the AP Statistics philosophy. Students will typically work 4 to 8 sets of AP problems.

Multiple Choice Question Preparation

Weekly In-Class Tests: A multiple choice AP-type test will be given approximately once per week starting near the end of March. These will count 100 points each toward your grade. They will be curved to approximate an AP grading system. At worst 50% correct will be a "C". In addition, each student will receive an estimate from 1 to 5 of their future grade based on each test.

Take-Home Tests: A minimum of 2 multiple choice AP-type take-home tests will be given. These will count 50 points each and be curved but not as generously as the in-class tests. Your work is to be turned in on each question. You may collaborate with other students and may compare answers but only if each person has actually worked the problem and written down their work. Allowing a student to simply copy your answers is strictly forbidden and may result in a grade of zero for both of the students involved.

The Good News: The highest in-class test will be cloned to help compensate for having a bad day. Take-home tests will not be cloned.

Self Study

As mentioned on the first page, you will not reach your potential on the AP test without a lot of self study. This AP test study should start in December and be primarily based on the Barron's AP review book you have purchased. In addition a Princeton Review book will be provided in the 4th quarter for additional self study.