AP Statistics Summer Institute 2016

E-mail: dave_ferris@nobl.k12.in.us
Web site: noblestatman.com

Introductory notes:

Perspectives on AP Statistics

By Chris Olsen
George Washington High School
Cedar Rapids, IA

Unlike the AP Statistics course description, which outlines the scope and nature of the course, this introduction focuses on the teaching and learning experience. The teachers who pioneered this program were deeply committed to it and excited about the benefits it could bring to their students. I believe their enthusiasm was rooted in (1) the discipline of statistics, (2) their experience with their students, and (3) the collective professional community they created.

Because the science, art, and practice of statistics differ significantly from other fields of mathematics, it is not surprising that this discipline is also taught differently. Among ecologists, there is a concept known as the "edge effect," the biologically active, interstitial region that forms a boundary --, for example, between a forest and a meadow. As the eminent statistician John Tukey noted, the field of statistics allows you to ***play in everyone else's backyard***. Statistics is positioned at the edge between the known and the unknown in all those backyards. Our classes are populated with students who possess a bewildering variety of interests, some of which are ***allegedly nonmathematical***. Statistics can encompass and expand those interests, and provide the methods and concepts for creatively extending knowledge in all of their backyards.

The AP Statistics classroom is nothing if not active. Students analyze data with calculators and computers, conduct classroom experiments, carry out individual and group projects, and perform simulations involving probabilistic concepts. AP Statistics students are active, engaged learners. ***Moreover, these students would not necessarily be enchanted by a traditional mathematics course***. The AP Statistics course not only accommodates students with a wide variety of interests, it also serves those with a relatively wide range of math proficiency. ***Discussion in an AP Statistics class is an activity for all students***. Group projects are less likely to be dominated by the most able student, and individuals can succeed by capitalizing on their individual interests. A more healthy learning and teaching environment is difficult to imagine.

It has long been a fact of life that AP Statistics teachers are ***lonely members*** of their math departments. The preservice preparation of most math teachers today does not include a statistics course, and high school statistics teachers have less opportunity to bounce ideas off their colleagues. (This phenomenon is also not unknown among statisticians teaching in some colleges.) Ironically, this isolation, together with the power of the Internet, has spawned what is possibly the most collegial resource available to high school teachers -- each other. From the very beginning of AP Statistics, an electronic discussion group (EDG) operating out of British Columbia attracted statistics teachers to ongoing discussions of content, philosophy, and pedagogy. This EDG, now operating under the aegis of the College Board®, created a synergistic network that has aided hundreds of high school teachers, as well as college and university professors who realize the importance of the precollege statistics curriculum. Sessions at the annual meetings of the National Council of Teachers of Mathematics, statistics institutes at the North Carolina School of Science and Mathematics, and a growing number of Web sites have been direct consequences of this long-range collegiality. Clearly, this is the way our profession ought to work -- and nobody has done it better than the AP Statistics teachers.

In the light of the first six years' experience, the AP Statistics phenomenon must be declared an incredible and enduring success. It is a success not merely by the numbers -- 173,944 students have taken the AP Statistics exam in the past six years* -- but because of the personal and professional experience of teachers like you, and the learning experience of students like yours.

1. In 2015, approximately 200,000 students took the AP Statistics Exam (830 readers)

Why Take Statistics?

The Philosophical:
Statistics deals with how we can learn about the world from observations when those observations are fallible. It teaches us to work with randomness and make it a tool for discovering the unknown rather than something to fear. The reasoning of statistics is the foundation of scientific reasoning. When examined closely, it is amazing that we actually can draw conclusions from a random sample or fallible experiment. If nature were just a bit more intransigent, we'd be unable to get anywhere. There is something quite amazing and beautiful in reasoning and methods that tread so close to the edge of the impossible and yet are so fundamental to scientific progress.

The Political:
Your parents may have told you that their statistics course was the worst experience of their academic lives, but things are different now. The AP course is a modern course that focuses on data rather than on probability theory and combinatorics. We use calculators and computers and deal with real-world problems. We use discovery learning methods and multimedia materials. I guarantee that your experience will be nothing like what your parents suffered through.

The Practical:
Statistics is used by charlatans to fool the masses. Statistical literacy is good self-defense in a world of advertising and political claims that only seem to be based on data. This course will arm you with the necessary understanding so you can smell a statistically- disguised lie a mile away. And think of the money you'll save not playing the lottery.

The Crass:
1) Statistics is one of those courses that will get you a job and/or a promotion. When we survey the companies that hire our graduates (I teach at Cornell) they tell us that they'd like our students to know more statistics and more computing. When we survey our recent graduates (5-years out), they tell us that statistics was one of the most important courses to them in getting their current positions.

2) Statistics is required by virtually every social science major, engineering, pre-med, and many others. It is now required for most law degrees. AP Statistics credit is therefore a useful commodity -- it will save you time and money in college.

--Paul Velleman

M&M's Activity:

My Guess:_______

<u>Guesses</u>: <u>Actual</u>:

Describe the distribution of the guesses:

Compare and contrast the distributions of guesses vs. actual:

FIRST DAY OF SCHOOL IDEAS:

- **Take Census At School survey; find "stories" in the data; introduce technology/calc.**
- **Discuss how to answer a "Big Question:" Do shorter people have more bodyfat? Can you find the right waist size in pants by wrapping the pants around your neck? Will women run faster than men in the Olympics? How often to people wash their hands in RR? Do cell phones cause brain cancer? Do females text more than males?**
- **Conduct a taste test (bottled vs. tap water, etc.)**
- **Show interesting/famous graphs, data stories, articles about stats**
- **Do Memory Game activity**
- **Look at some cool data sets: TuvaLabs.com, Census at School, gapminder.org, Bodyfat dataset, McDonald's menu, Arby's Menu, etc.**

Variation Activity: (shows how students perceive SAMPLING variation—a new idea)

"A committee of 10 members of a school band is to be chosen from a set of 30 males and 70 females. How many females might be on such a committee?"

- **A random draw of 10 names is completed 30 times. Graph your estimate of the number of females in those 30 samples on poster.**
- **Calculate the mean, median and mode of your graph.**
- **Discuss and justify your graph (shape, center, spread, etc.).**
- **Simulate a few draws using colored marbles or beads, etc.**
- **Simulate using a calculator or computer.**
- **Compare simulations to the graph of your estimates.**

From "The Known Mix: A Taste of Variation"
Nov. 2008, *Mathematics Teacher*

Essential resources from apcentral.collegeboard.com:

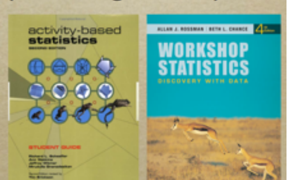
- **Course Overview**
- **Full Course Description (including AP StatsTopic Outline—later in this handout)**
  - **"Statistics is a discipline in which clear and complete communication is an essential skill."**
  - **"formulate cogent answers"**
- **AP Statistics Teacher's Guide (BIG pdf file)**
- **AP Teacher Community and e-mail discussion group (see green sidebar)**
- **Audit information**
- **Special Focus: Samping Dist'ns, Inference, Planning and Conducting a Study**

Textbooks:

- The Practice of Statistics 5e (Starnes, Yates, Moore)
- Stats: Modeling the World 4e (Bock, Velleman, Deveaux)
- Statistics: Learning From Data "1e" (Peck, Olsen)
- Statistics In Action 2e (Watkins, Schaeffer, Cobb)

Other Resources:

- www.apstatsmonkey.com
- Rossman/Chance applets
- TuvaLabs.com
- StatKey (online simulations and data analysis)
- AP Exams (FR) on AP Central (& scoring rubrics)
- Supplementals:
  - Activity-Based Statistics
  - Workshop Statistics
  - Exam Formulas (CB book)
- Fathom, StatCrunch, Google Forms



Data Collection Activity

First Day of School Name:____________________________________

Supply the data for the following categories/attributes/measurements.

1. First Name: ___________________________
2. Last Name: ___________________________
3. Gender: ___________________________
4. Blood Type: ___________________________
5. Age in Days: ___________________________
6. Height in cm: ___________________________
7. Belly button height above ground in cm.: ________________________
8. Foot length in cm.: ___________________________
9. Wrist circumference in cm.: ___________________________
10. Eye color: ___________________________
11. Month of birth: ___________________________
12. Day of birth (1–31): ___________________________
13. Circumference of head in cm.: ___________________________
14. Shoe size: ___________________________
15. Number of sisters: ___________________________
16. Number of brothers: ___________________________
17. Are you left-handed or right-handed? ___________
18. Measure your hand span in cm: ______________

2010 US Census at School Measurement Questions

Use Safari. Class ID: _______________ Password: ____________

The following questions require measurements. Please fill these out prior to taking the online survey.

4. How tall are you without your shoes on? Answer to the nearest centimeter. ________

5. What is the length of your right foot (without your shoe on)? Answer to the nearest centimeter.

____

6. What is your arm span? (Open arms wide and measure distance across your back from tip of right hand middle finger to tip of left hand middle finger.) Answer to the nearest centimeter.

____

9. How long does it usually take you to travel to school? Answer to the nearest minute.

____

14. What is the length of your left foot (without your shoe on)? Answer to the nearest centimeter.

____

16. What is the length of your index finger (finger next to your thumb) on your left hand? Answer to the nearest centimeter.

____

17. What is the length of your ring finger? (located between your middle finger and little finger) on your left hand? Answer to the nearest millimeter (there are 10 millimeters in one centimeter).

____

26. How many hours of sleep do you usually get when you have school the next day?

____

27. How many hours of sleep do you usually get when you don’t have school the next day?

____

The Kristen Gilbert Case

Kristen Gilbert worked as a nurse in the intensive care unit of the VA hospital in Northampton, Massachusetts in the 1990's. During her shifts, other nurses became suspicious that she was killing patients by injecting them with epinephrine, a heart stimulant. An analysis of 1641 eight-hour shifts was presented as evidence at her trial (Cobb and Gelbach, 2005). Is there statistical evidence from this table that Gilbert murdered patients?

	Death on shift?		
Gilbert Present?	Yes	No	**Total**
Yes	40	217	**257**
No	34	1350	**1384**
Total:	**74**	**1567**	**1641**

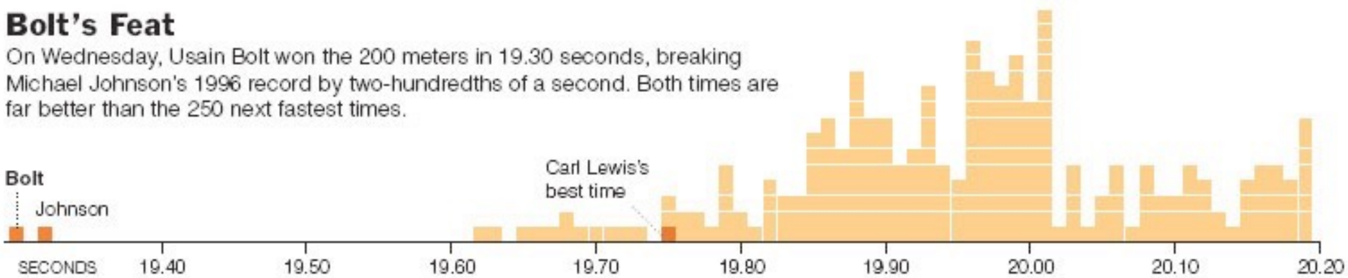
Using StatKey simulation

1. Open StatKey: http://lock5stat.com/statkey/
2. Choose Test for Difference in Proportions
3. Choose Edit Data, and fill in with 40, 257, 34, 1384. Click OK.
4. Generate 1 sample. What is the simulation doing?

   What do the numbers in the Randomization Sample table represent?

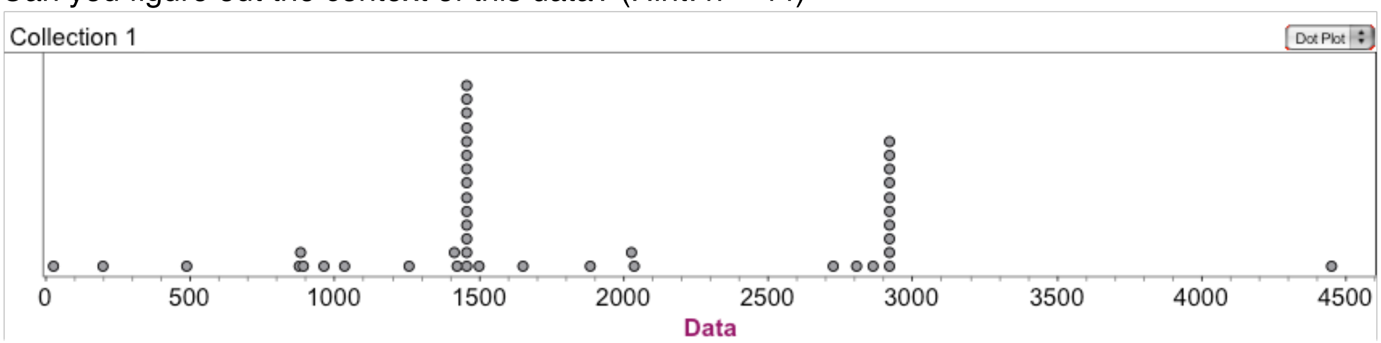
5. Generate 9 more samples. What do the 10 dots in the graph represent?
6. Generate 1000 Samples. Describe the distribution that is produced (shape, center, spread).
7. What does this distribution of 1010 dots tell us about the original Kristen Gilbert data?
8. Does this data suggest that Kristen murdered patients? Explain.

From the 2008 Summer Olympics:

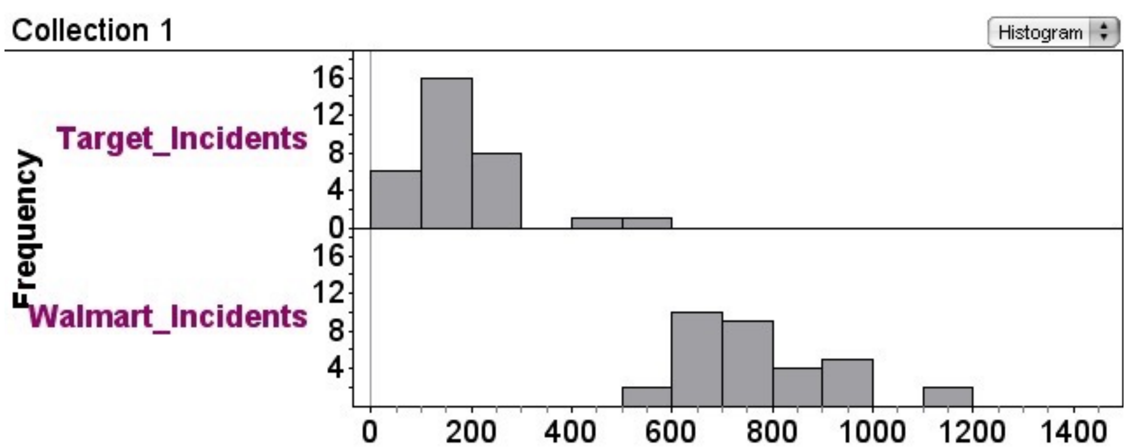


Outliers and Inliers

Can you figure out the context of this data? (Hint: n = 44)

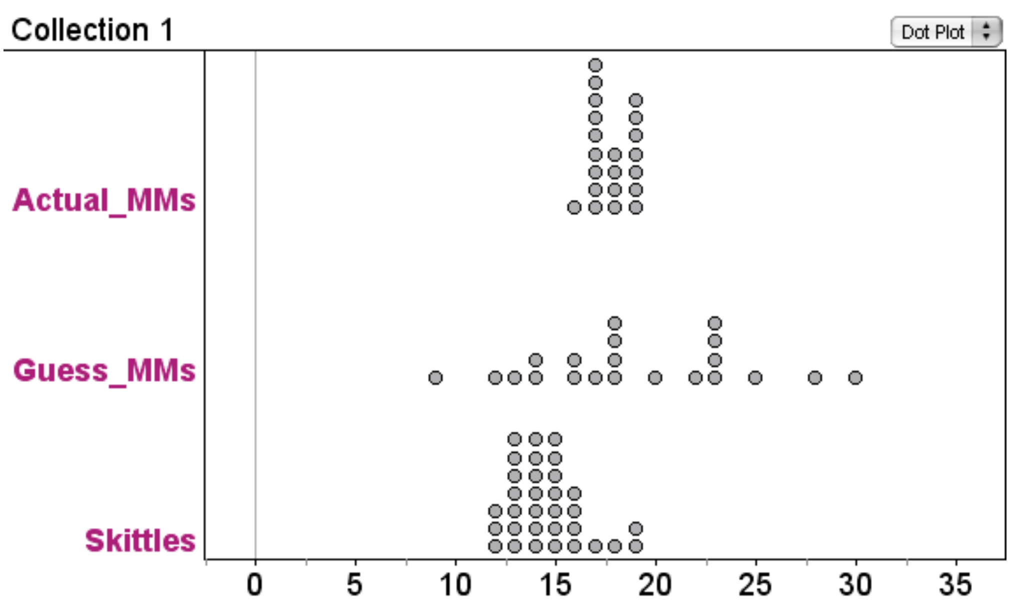


Is Wal-Mart Safe?[1]



[1] See notes at the end of this handout for details on how the data was collected.

Standard Deviation:



Actual: mean= __________

SD = __________

Guess: mean= __________

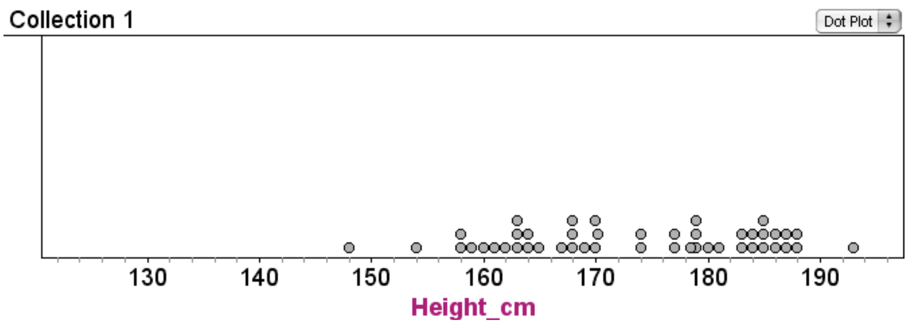
SD = __________

Skittles: mean= __________

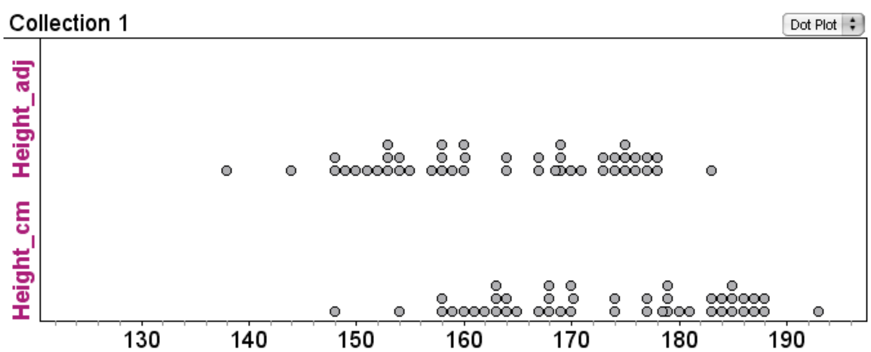
SD = __________

Transforming Data

Below is the distribution of student heights in cm.



1. What if we discovered that the tape measure used to measure height had 10cm cut off (i.e. it started at 10 instead of 0)? We need to subtract 10cm from each student's height. How would this transformation affect the mean and SD?



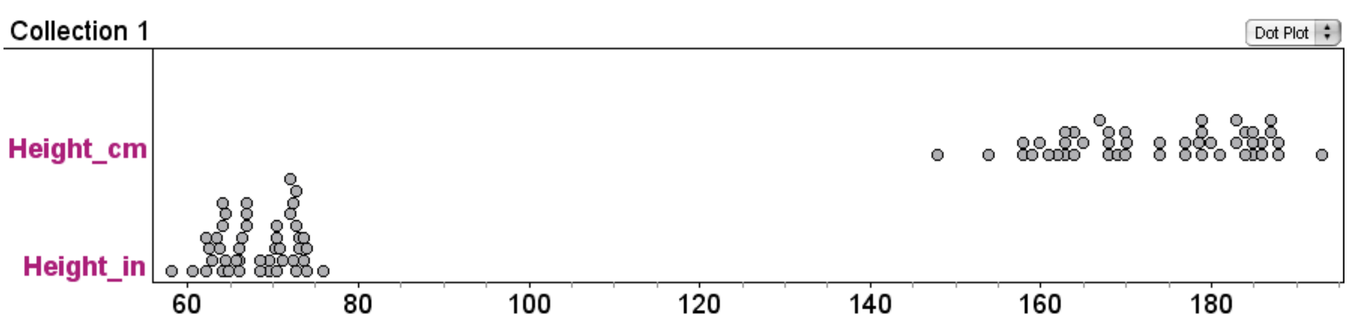
Mean = __________

SD = _________

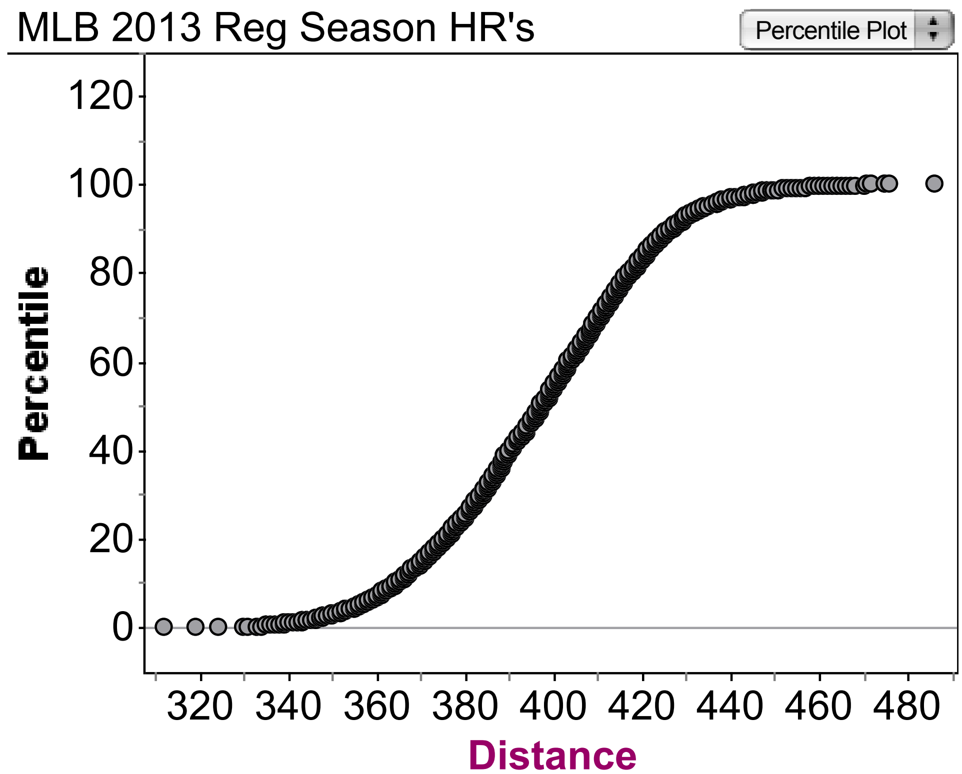
2. What if we had to convert each height to inches and recalculate the mean and SD? What is the new mean and SD?

Mean = __________

SD = _________



and and Ogive!



The cumulative frequency of the distances traveled by all 2013 MLB home runs is above.

1. Find the distance that represents the approximate value of the median. __________

2. Find the distance that represents the approximate value of the first (lower) quartile. _________

3. Approximate the interquartile range of these distances (IQR): __________

4. What is the approximate percentile of a home run that traveled 420 feet? __________

5. Find the distance that is closest to the 30th percentile. __________

Categorical Data

Below are some data and questions from the Titanic disaster.

http://lib.stat.cmu.edu/S/Harrell/data/descriptions/titanic.html
http://www.encyclopedia-titanica.org/

1. Give the marginal distribution of class (in %'s).

Titanic Passengers

		Survived		Row Summary
		No	Yes	
Class	1st	129	193	322
	2nd	161	119	280
	3rd	574	137	711
Column Summary		864	449	1313

S1 = count ( )

2. Give the conditional distribution of survival by class (in percent).

3. Of the first class passengers, what percent survived? P(survived I 1st class) = ________

4. Of the survivors, what percent were 1st class?   P(1st class I survived) = ________

5. What percent of passengers were 1st class? _________

6. What percent of passengers were either first class or survived? ________

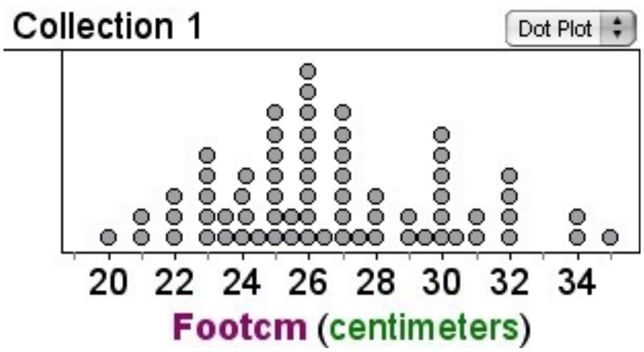
7. Do survivability and class appear to be independent? Explain.

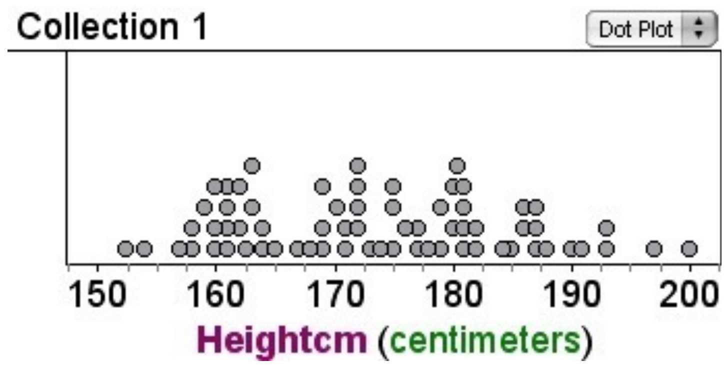
The Normal Model for Data:

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(Note difference between “math world” and “stats world.”)

Which is more extreme, a foot length of 32 cm or a height of 190 cm?

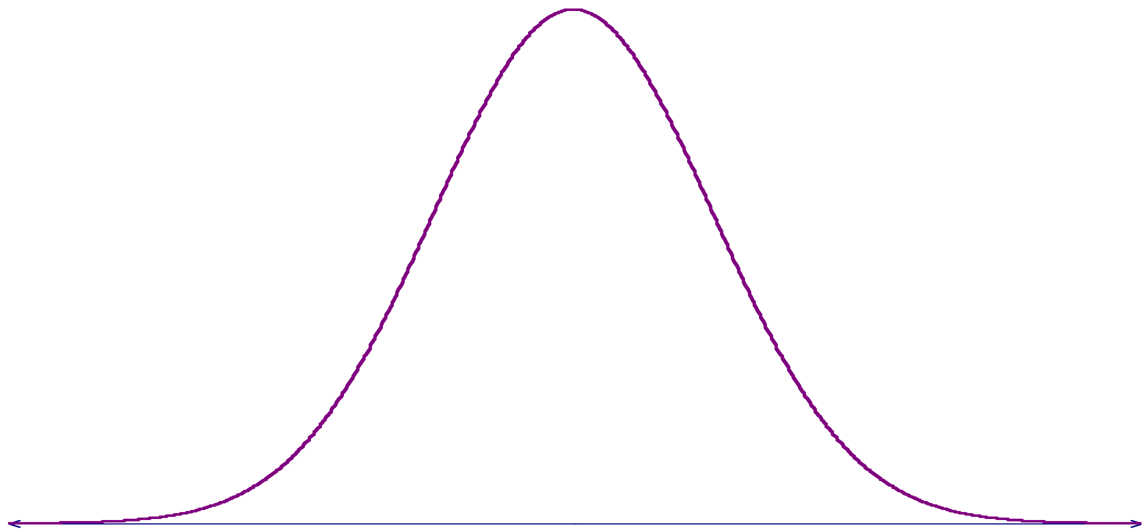




Heightcm	173.597 cm 11.36 cm
Footcm	26.6964 cm 3.37966 cm

S1 = mean ( )
S2 = stdDev ( )

Common exercises using the normal model:



1. Find the z-score for a student who scored 660 on the SAT Verbal, where the mean is 505 and the standard deviation is 110.

2. What percentile did this student score?

3. What SAT Verbal score represents the first (lower) quartile?

4. Approximately what percent of students scored between 450 and 560?

5. What score would a student have to earn to be in the top 5% of all SAT Verbal scorers?

6. Approximately what percent of students scored between 395 and 615?

7. Approximately what percent of student scored within two standard deviations of 505?

Age Guessing Activity

1. Guess the ages of the following people, and put your guess in this column:

Name:	Actual Age:	Your Guess:
Donald Trump	________	________
Nate Silver	________	________
Bill Gates	________	________
Johnny Depp	________	________
Adele	________	________
Alex Trebek	________	________
Daisy Ridley	________	________
Miley Cyrus	________	________
Tom Brady	________	________
J. K. Rowling	________	________
Mick Jagger	________	________
Mark Zuckerberg	________	________

2. Put the actual ages of each person in the first column.

3. Type both lists into your calculator. (Actual ages in L1, your guesses in L2.)

4. Make a scatterplot for these two lists. (x-axis is L1, y-axis is L2)

5. Calculate r: ________

6. Describe below what you discovered about your age guesses:

Bivariate Data

Will women run faster than men in the Olympics?

Winning times for Olympic 200-meter dash

Year	Female	Female _time	Country	Male	Country	Male_ time
1900				Walter Tewksbury	USA	22.2
1904				Archie Hahn	USA	21.6
1908				Robert Kerr	Canada	22.6
1912				Ralph Craig	USA	21.7
1920				Allan Woodring	USA	22.0
1924				Jackson Scholz	USA	21.6
1928				Percy Williams	Canada	21.6
1932				Eddie Tolan	USA	21.2
1936				Jesse Owens	USA	20.7
1948	F. Blankers-Koen	24.40	Netherlands	Mel Patton	USA	21.1
1952	Marjorie Jackson	23.70	Australia	Andrew Stanfield	USA	20.7
1956	Betty Cuthbert	23.40	Australia	Bobby Morrow	USA	20.6
1960	Wilma Rudolph	24.00	USA	Livio Berruti	Italy	20.5
1964	Edith McGuire	23.00	USA	Harry Car	USA	20.3
1968	Irena Szewinska	22.50	Poland	Tommie Smith	USA	19.8
1972	Renate Stecher	22.40	E. Germany	Valeri Borzov	USSR	20.0
1976	Barbel Eckert	22.37	E. Germany	Donald Quarrie	Jamaica	20.2
1980	Barbel Wockel	22.03	E. Germany	Pletro Mennes	Italy	20.2
1984	Valerie Brisco-Hooks	21.81	USA	Carl Lewis	USA	19.8
1988	Florence Griffith-Joyner	21.34	USA	Joe Deloach	USA	19.8
1992	Gwen Torrence	21.81	USA	Mike Marsh	USA	20.0
1996	Marie-Jose Perec	22.12	France	Michael Johnson	USA	19.3
2000	Marion Jones	21.84	USA	Konstantinos Kenteris	Greece	20.1
2004	Veronica Campbell	22.05	Jamaica	Shawn Crawford	USA	19.8
2008	Veronica Campbell-Brown	21.74	Jamaica	Usain Bolt	Jamaica	19.3

Shaughnessy, J. Michael, et. al: *Focus in High School Mathematics, Reasoning and Sense Making: Probability and Statistics,* Chapter 3; NCTM: 2009

Will women run faster than men in the Olympics?

Possible Discussion Questions

1. Discuss in your groups how you might use the information in the table to answer the question.

2. Create a scatterplot of the women's times vs. year. Describe what you see.

3. Graph the men's times <u>on the same scatterplot</u>, but use a different style or color for each dot. Describe what you see. Can you answer the question now? Why or why not?

4. Create a scatterplot of the DIFFERENCES in the times vs. year. Describe what you see.

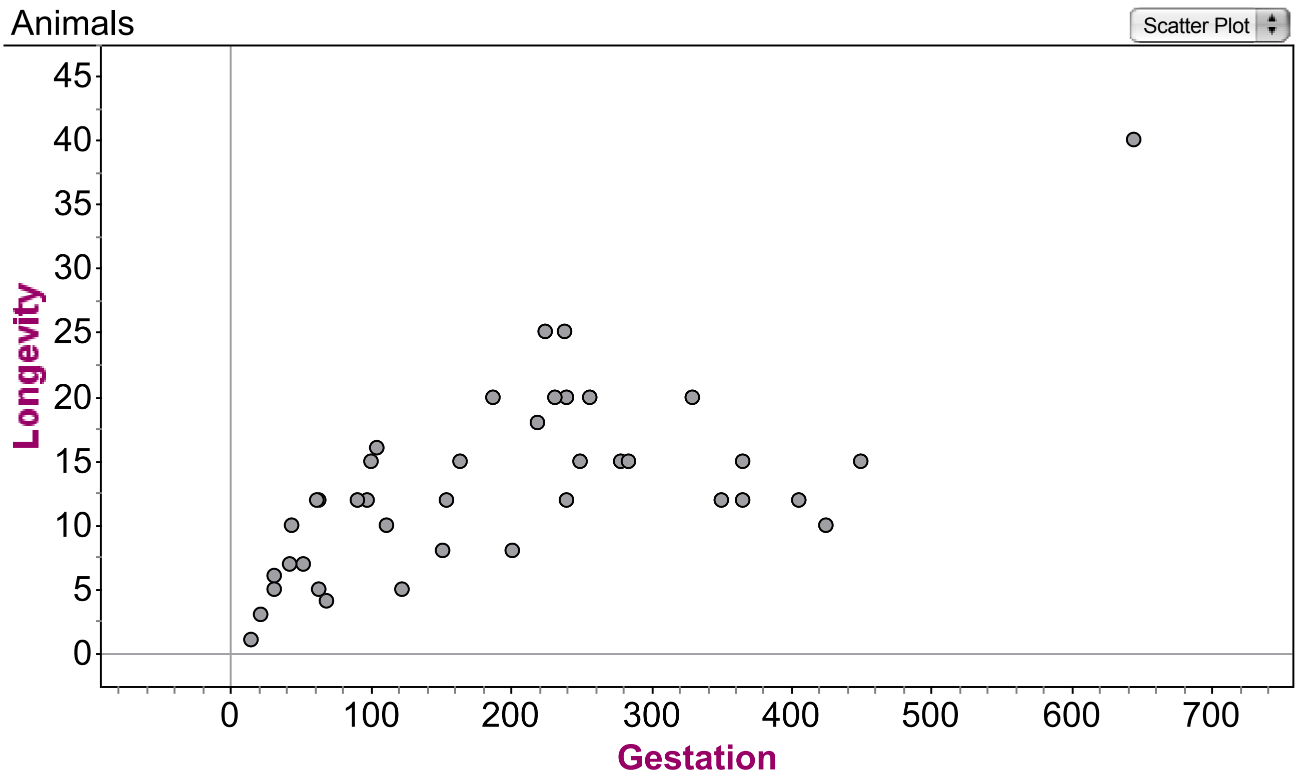
5. Would a linear model be appropriate for this relationship? Explain.

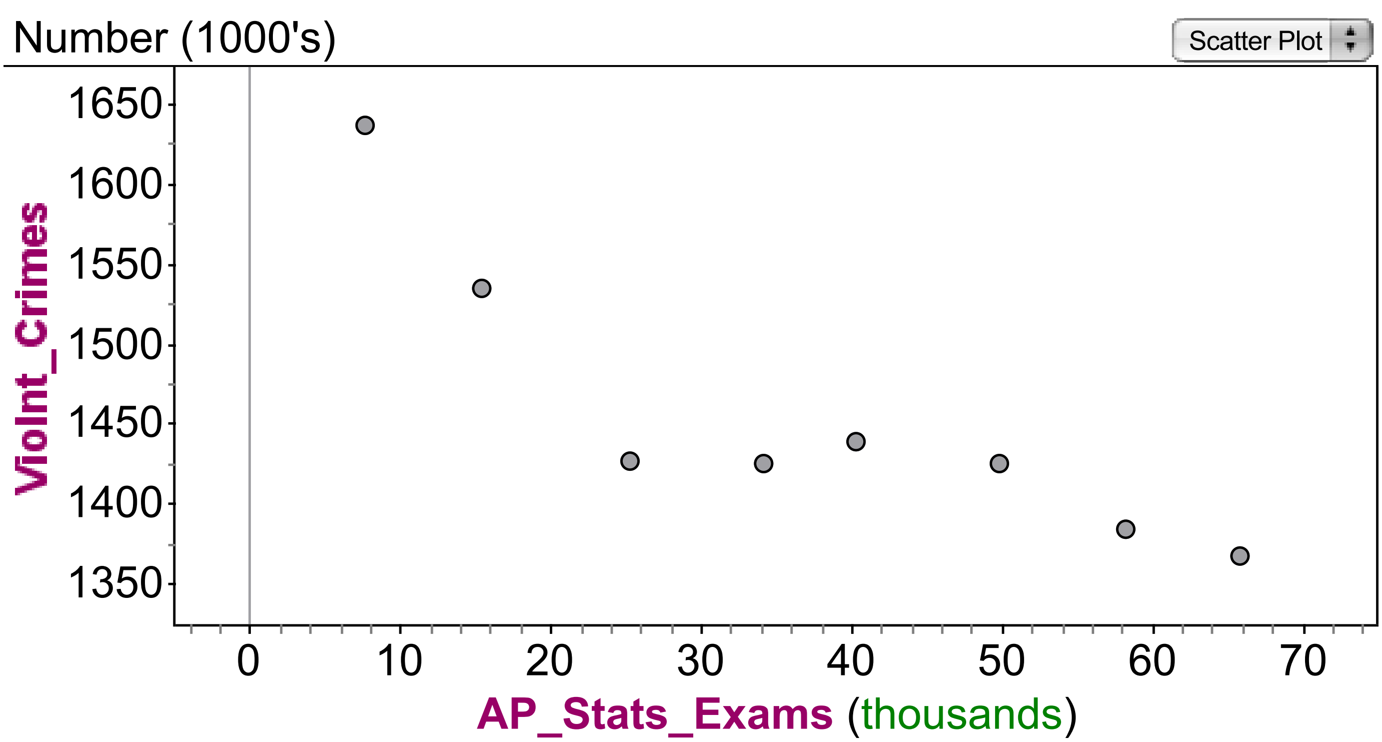
6. If the data DOES seem to have a linear relationship, is this relationship strong? Explain.

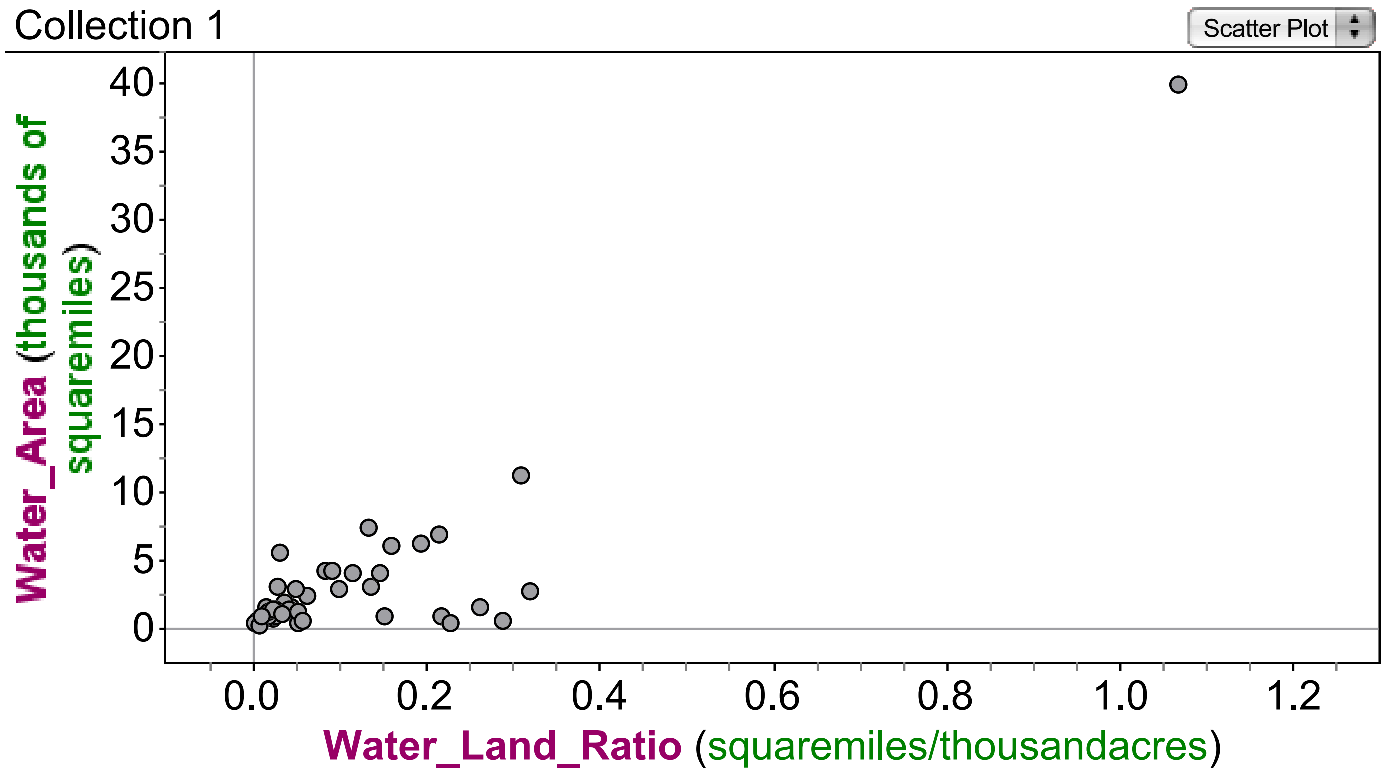
7. What are some possible sources of the variation in the olympic times?

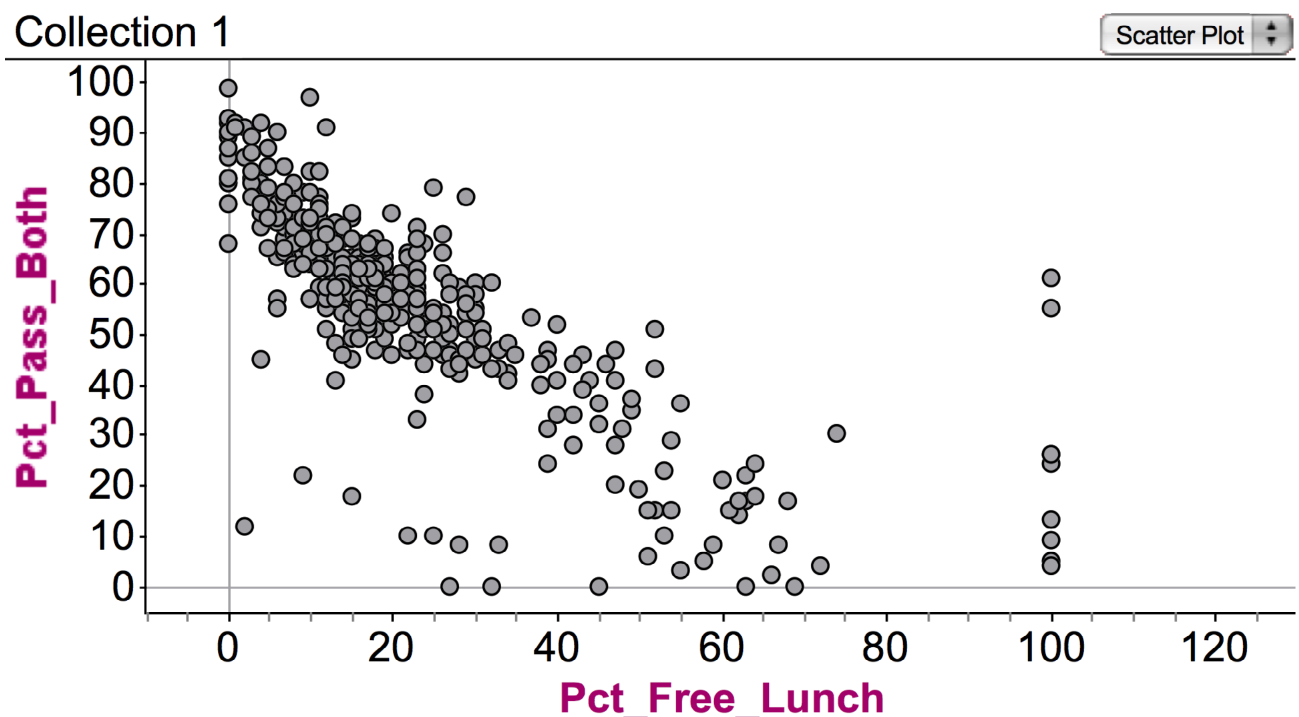
Some Scatterplot Examples:

The entire set is on my web site: web.me.com/noblestatman






Collection 1
Scatter Plot
Water_Area (thousands of squaremiles)
40
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Water_Land_Ratio (squaremiles/thousandacres)


Collection 1
Scatter Plot
Pct_Pass_Both
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
Pct_Free_Lunch

Commonly Asked Regression Questions

(as seen on previous AP Statistics exams)

1. Describe the association in context.

2. Is a linear model appropriate to describe this relationship? Explain.

3. Write the equation for the linear model on this data.

4. Explain the meaning of the slope in this linear model

5. Explain the meaning of the y-intercept in this linear model

6. Find the value and explain the meaning of the correlation coefficient.

7. Find the value of and interpret r-squared

8. a. Using the linear model, predict ______________ when _____________ is ____________.

   b. Is the residual for this data point positive or negative? Is the model over- or underestimating? Explain.

9. Comment on any outliers present. Fully describe their effect on the analysis, if any.

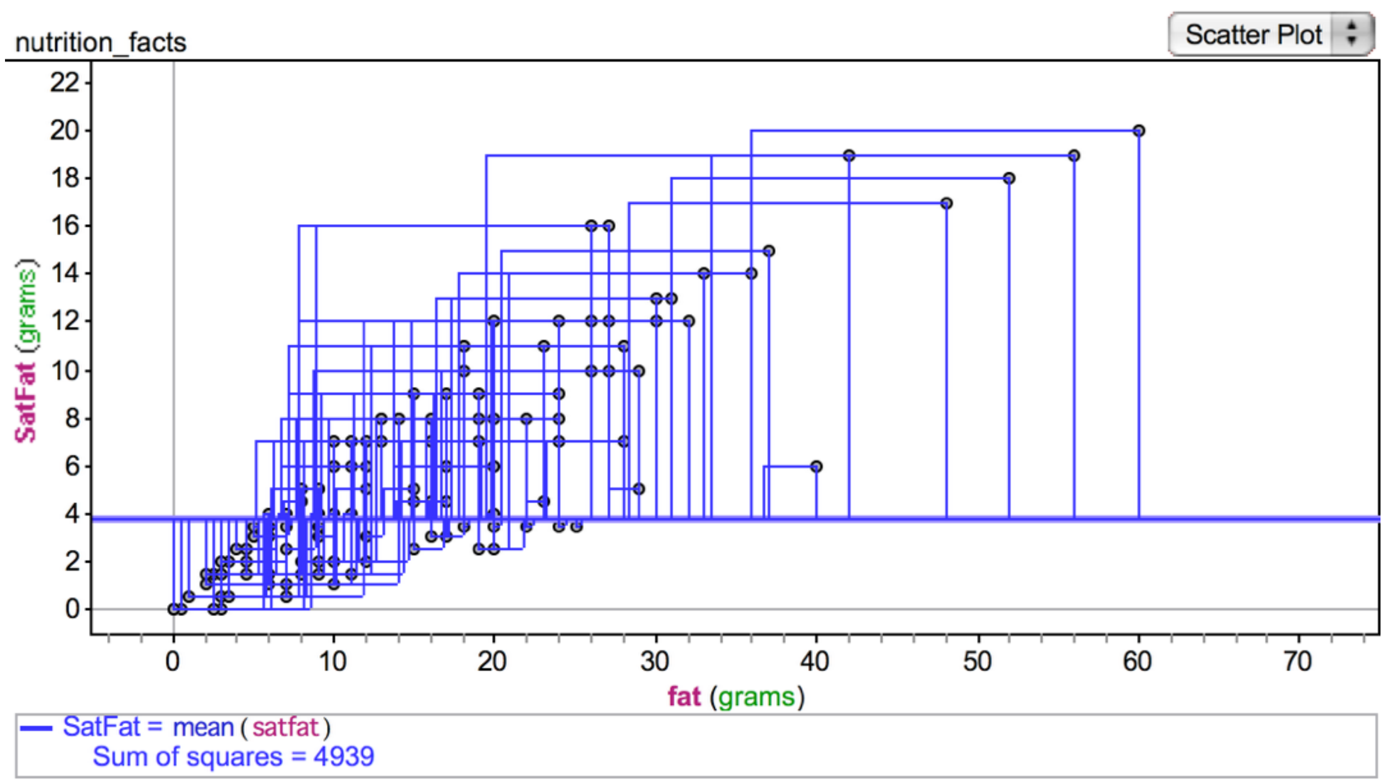
10. Interpret regression and model information from a computer printout.

What is R-Squared?

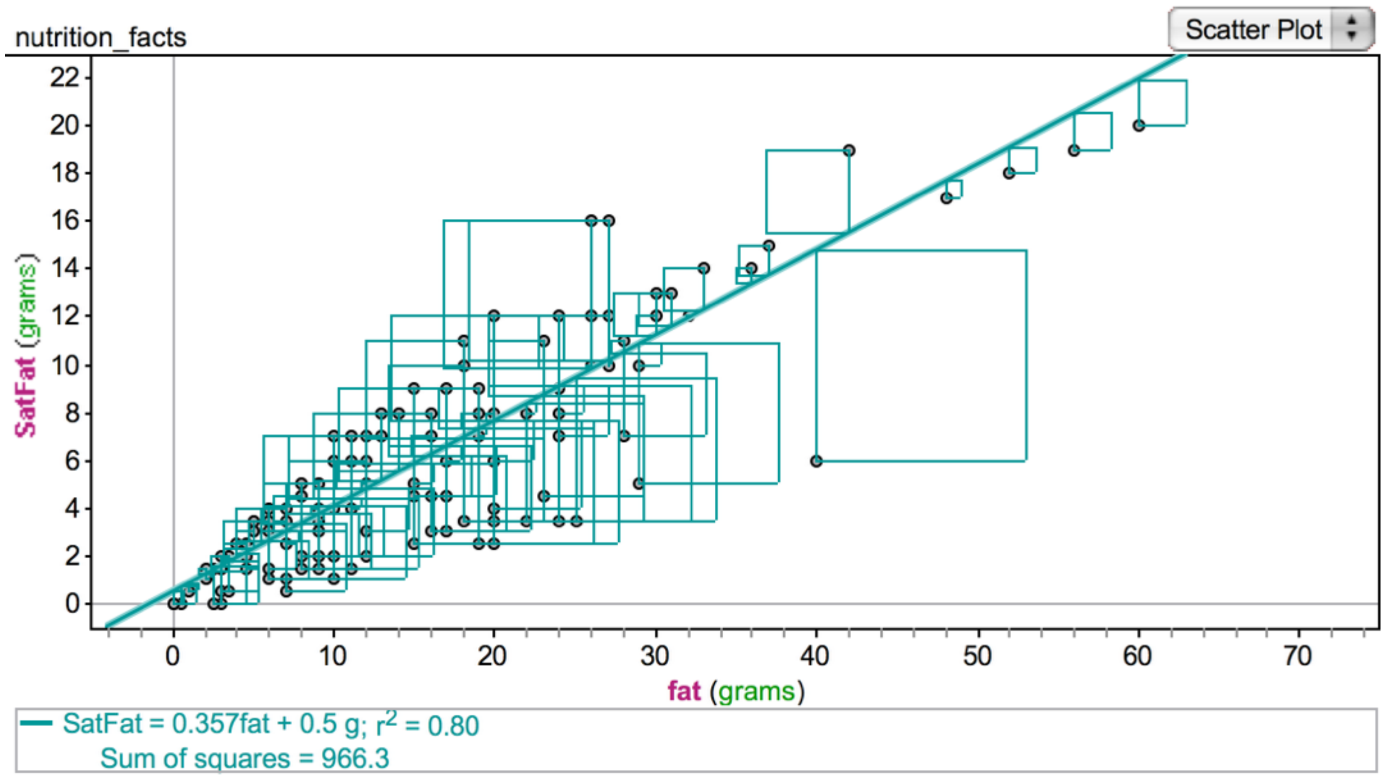
R-squared is frequently misunderstood, and its interpretation is often "parroted" without a clear, personal understanding. The following explanation will try to make r^2 clear.

Look again at the fat vs. saturated fat example. One of the common interpretations of r-squared is something like this: "80% of the variation in saturated fat in McDonald's menu items can be explained by the linear model on fat." To understand this statement, let's start with the question: "How much variation is in saturated fats *without* looking at the linear model on fat?"

A logical question to this question might be: "Variation from *what*?" Well, when we calculate standard deviation and variance, we usually measure variation from the mean. So let's measure the total amount of variation in saturated fats by finding the sum of the squares of the differences between each saturated fat and the mean of the saturated fats (called SST, or sum of squares total). Below is a graph showing all the squares constructed from each data point to the mean (represented by the horizontal line). The sum of the area of these squares (SST) is 4939.



But this horizontal line is clearly not the best linear model for the data. It looks like we could reduce the sum of the squares by "tilting" the line in an upward slope until we produced the least sum of the squares…making the least squares regression line!



(These squares are actually the squares of all the residuals.) Notice that the sum of the squares HAS been reduced…all the way down to 966.3 (this is called the SSE—the sum of the squares of the errors). How much of the variation was removed by the linear model?

SST – SSE = 4939 – 966.3 = 3972.7

What *percent* of the SST was this?

SSE ÷ SST = 3972.7 ÷ 4939 = .80435 (80%)

Notice this was the r^2 reported below the graph. The earlier interpretation should now make sense. The saturated fats for McDonald's menu items had some variation to begin with. But if we consider the variation in saturated fats with respect to a linear model on fat for the same menu items, the some of variation has been reduced/accounted for/explained by this linear model. The percent that was "reduced" is the value of r-squared.

AP Statistics Topic Outline

Following is an outline of the major topics covered by the AP Statistics Exam. The ordering here is intended to define the scope of the course but not necessarily the sequence. The percentages in parentheses for each content area indicate the coverage for that content area in the exam.

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

Exploratory analysis of data makes use of graphical and numerical techniques to study patterns and departures from patterns. Emphasis should be placed on interpreting information from graphical and numerical displays and summaries.

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

2. Center and spread
3. Clusters and gaps
4. Outliers and other unusual features
5. 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

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
5. Transformations to achieve linearity: logarithmic and power transformations

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

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

Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained. This plan includes clarifying the question and deciding upon a method of data collection and analysis.

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 and types of conclusions that can be drawn from observational studies, experiments, and surveys

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

Probability is the tool used for anticipating what the distribution of data should look like under a given model.

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
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

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

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
7. t-distribution
8. Chi-square distribution

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)
8. Confidence interval for the slope of a least-squares regression line

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 difference between two proportions
4. Test for a mean
5. Test for a difference between two means (unpaired and paired)
6. Chi-square test for goodness of fit, homogeneity of proportions, and independence (one- and two-way tables)
7. Test for the slope of a least-squares regression line

Age Guessing Key:

Guess the age of the following people:

Name:	Actual BD:	AGE:__
Donald Trump	6/14/46	___________
Nate Silver	1/13/78	___________
Bill Gates	10/28/55	___________
Johnny Depp	6/9/63	___________
Adele	5/5/88	___________
Alex Trebek	7/22/40	___________
Daisy Ridley	4/10/92	___________
Miley Cyrus	11/23/92	___________
Tom Brady	8/3/77	___________
J. K. Rowling	7/31/65	___________
Mick Jagger	7/26/43	___________
Mark Zuckerberg	5/14/84	___________

Wal-Mart vs. Target:
This phase of the study focused on the 50 Wal-Mart stores out of the 460 analyzed stores that experienced the "highest rate" of reported police incidents in 2004. Target stores chosen for the comparative analysis were within a 10-mile radius of the 50 "high incident" Wal-Mart stores. Of these 50 "high incident" Wal-Mart stores, three stores did not have a Target within 10 miles, leaving a sample of 47 Wal-Mart stores for further analysis. Because of further data restrictions, the sample for comparison was limited to 32 Wal-Mart stores and 30 nearby Target stores.