1. WHAT IS STATISTICS?

"Statistics is data analysis, together with everything you need to do data analysis."

(John Tukey)

"What is real, and what is an illusion?"

(Moody Blues)

"Understand Variation." (W.E. Deming)

"A recent poll indicates that 49 percent of the people are in favor of President Obama as president. Margin of error is ±3 percent.

(Gallup)

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write".

(H.G. Wells)

"You've got to know when to hold 'em, know when to fold 'em."

(The Gambler)

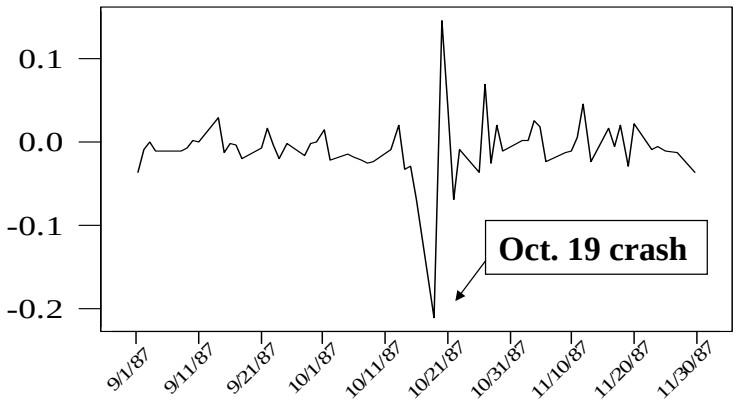
Statistics

A body of principles and methods for extracting useful information from data, for assessing the reliability of that information, for measuring and managing risk, and for making decisions in the face of uncertainty.

• Helps you to make better management decisions, instead of drowning in a flood of numbers. Statistics gives you a competitive advantage, since it is simply not enough to rely on intuition, experience and hunches.

• Statistics looks at the big picture (providing an overview and summary of the data), but it does not ignore the individual (the exceptions to the rule).

GM Returns



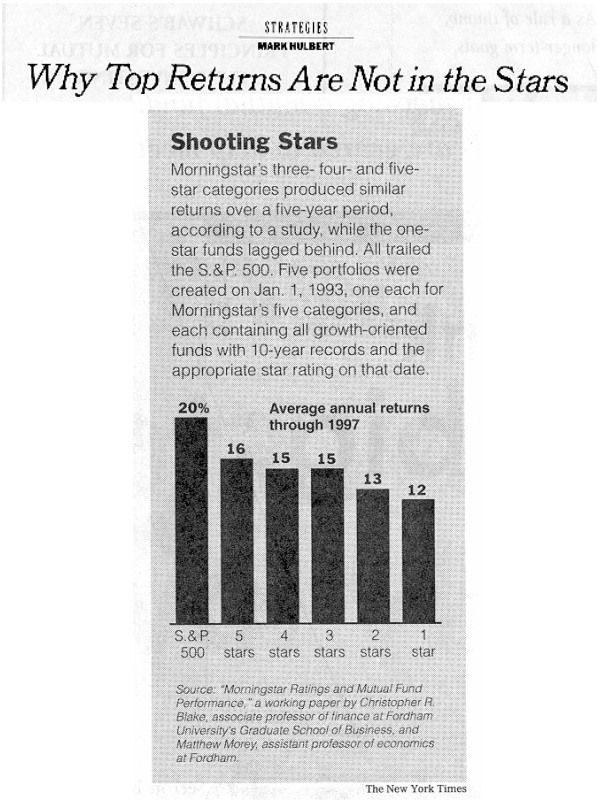
• You need Statistics to understand variability, a ubiquitous phenomenon that blurs the message of the data.

• Since variability is a measure of risk, statistical methods yield tools for risk management.

Problems that Statistics can help us solve

- Domino's Pizza wants to estimate the average time between orders and deliveries. (Goal: Improve quality of service). Why not use average of waiting times as provided by complaining customers?

- Is there a difference in performance between mutual funds with different Morningstar ratings?



STRATEGIES

MARK HULBERT

Why Top Returns Are Not in the Stars

Shooting Stars

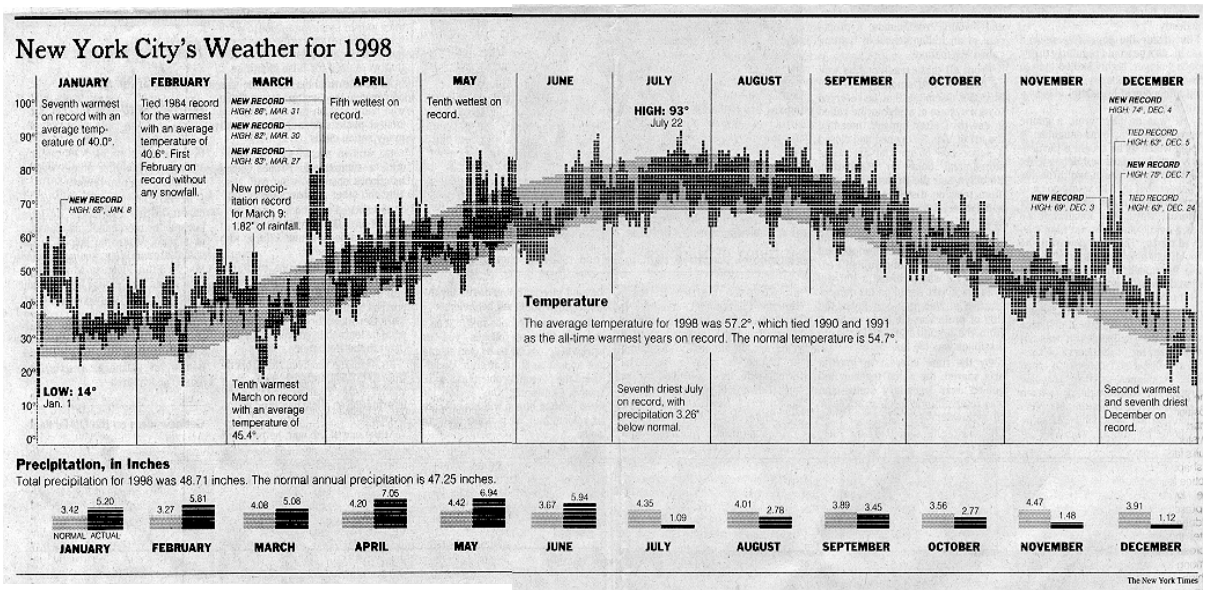
Morningstar's three- four- and five-star categories produced similar returns over a five-year period, according to a study, while the one-star funds lagged behind. All trailed the S.&P. 500. Five portfolios were created on Jan. 1, 1993, one each for Morningstar's five categories, and each containing all growth-oriented funds with 10-year records and the appropriate star rating on that date.

Source: "Morningstar Ratings and Mutual Fund Performance," a working paper by Christopher R. Blake, associate professor of finance at Fordham University's Graduate School of Business, and Matthew Morey, assistant professor of economics at Fordham.

The New York Times

Descriptive Statistics

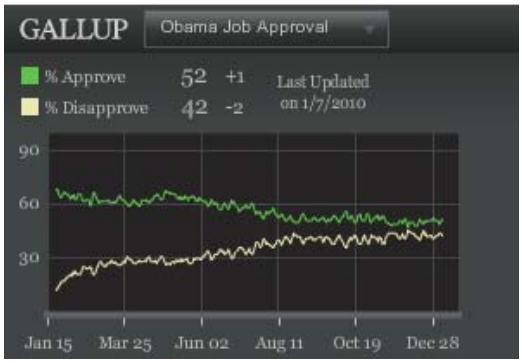
Methods of organizing, summarizing and presenting numerical data in a convenient form.



Inferential Statistics

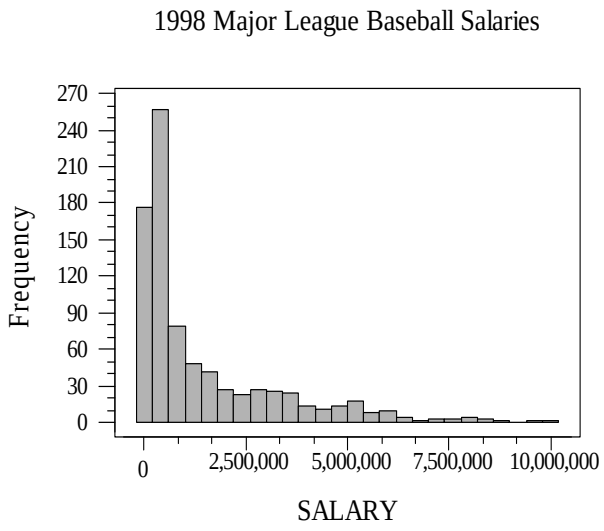
Methods for drawing conclusions (making inferences) about a population based on a small sample.

Eg: In a poll conducted the day before the 2008 presidential election, 54% of the 2,500 registered voters contacted favored Obama. Based on this data, how confident can we be that Obama will win the election, i.e., that at least 50% of the **population** of all voters are for Obama?



Some Key Statistical Concepts

Population: The set of all items of interest in a statistical problem. The size of a population can be finite or infinite.



Eg: The salaries $x_1,\cdots, x_N$ of the N = 837 Major League baseball players for the 1998 season.

Belle	Albert	10,000,000	CWS	AL
Sheffield	Gary	10,000,000	Fla	NL
Maddux	Greg	9,600,000	Atl	NL
Bonds	Barry	8,916,667	SF	NL
McGwire	Mark	8,333,333	StL	NL
Clemens	Roger	8,250,000	TOR	AL
Williams	Bernie	8,250,000	NYY	AL
Galarraga	Andres	8,000,000	Atl	NL
Piazza	Mike	8,000,000	LA	NL
Sosa	Sammy	8,000,000	ChC	NL

- N represents the population size.
- A descriptive measure of a population is called a **parameter**.

  Eg: The average salary of the 837 baseball players, μ = 1,447,690

Sample: A set of data drawn from the population.

Eg: The salaries $x_1,\cdots, x_n$ of the n = 100 top-earning baseball players.

- n represents the sample size.
- A descriptive measure of a sample is called a **statistic**.

  Eg: The average salary of the 100 players in our sample, $\bar{x}$ = 5,652,915.

The value of a statistic will fluctuate from sample to sample. We must take account of this variability in order to draw valid conclusions from the given data.

Statistical Inference: The process of making an estimate, forecast, or decision about a population, based on the sample information.

We usually don't have the time or money to examine an entire population (i.e., conduct a **census**). It is cheaper and much more efficient to base our conclusions about the population on a small sample.

Currently the decennial US "census" involves a very nonrandom sampling scheme: The people who respond are those who have addresses and wish to comply. This causes a systematic undercounting of certain groups. Statistical sampling would greatly improve the accuracy of the count.

Of course, it is not possible to decide or estimate with certainty without seeing the entire population. (This is one reason why some object to the use of statistical sampling in the US census). Fortunately, as long as we can control the sampling methodology, we will be able to provide a measure of the reliability of the inference.

Eg: An example of such a reliability measure is the following statement:

"The true proportion of voters for the Democrat will be within 0.5% of the estimate on 95% of the occasions when such a prediction is made."

Our Goals

Become informed and critical readers of quantitative statements and arguments.

Apply statistical techniques to analyze data, manage risk, and make decisions.

Understand and accept randomness and variation.

- To make good business decisions, you must take calculated risks.
- Every business decision is to some extent a gamble.
- Most people need to sharpen their intuition about randomness.

Labs and Demos

[Let's Make a Deal Demo].

Let's Make a Deal Website (Play the game interactively):
http://www.stat.sc.edu/~west/javahtml/LetsMakeaDeal.html

[Forecasting Lab].