What is Statistic?

OPRE 6301

In today's world...

... we are constantly being bombarded with statistics and statistical information. For example:

Customer Surveys
Medical News
Demographics
Political Polls
Economic Predictions
Marketing Information
Sales Forecasts
Stock Market Projections
Consumer Price Index
Sports Statistics

How can we make sense out of all this data?
How do we differentiate valid from flawed claims?

What is Statistics?!

"Statistics is a way to get information from data."

Data: Facts, especially numerical facts, collected together for reference or information.

Information: Knowledge communicated concerning some particular fact.

Statistics is a *tool* for creating an *understanding* from a set of numbers.

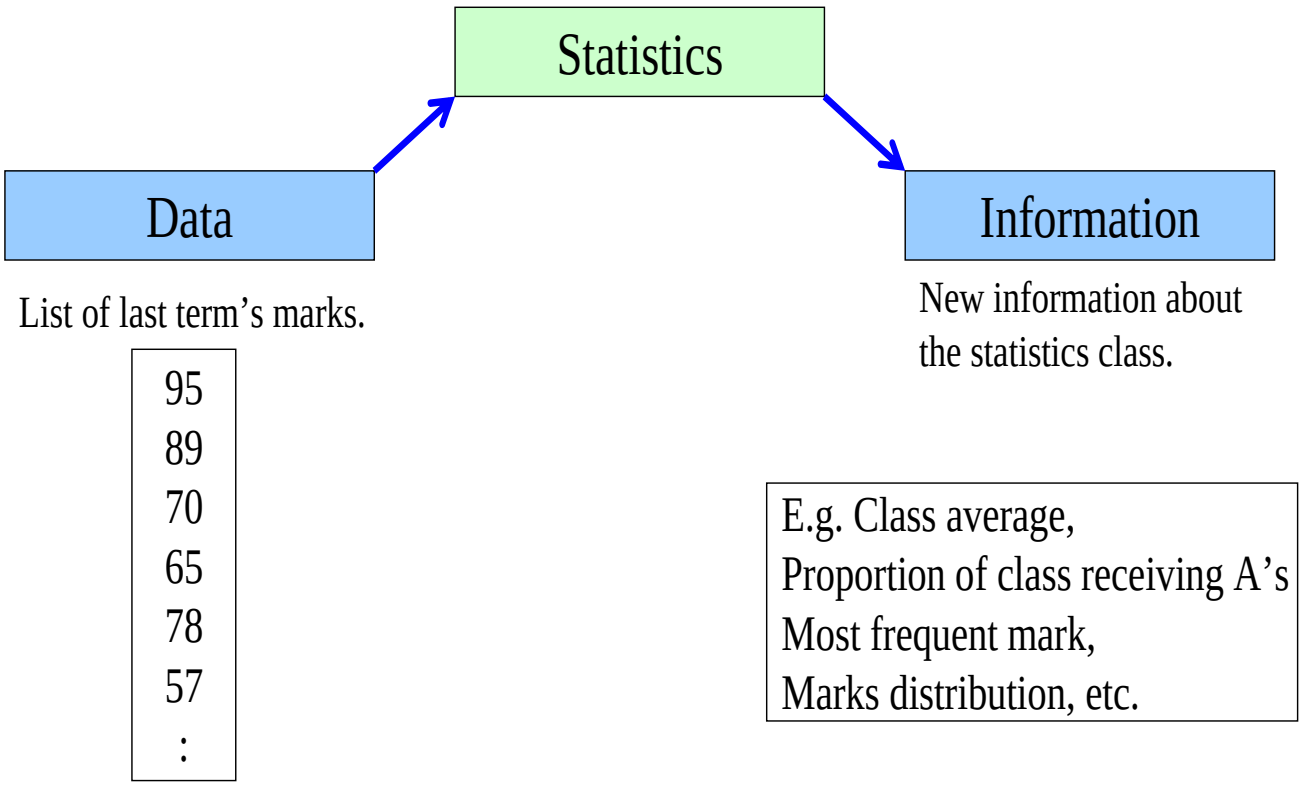
Humorous Definitions:

The Science of drawing a precise line between an unwarranted assumption and a forgone conclusion.

The Science of stating precisely what you don't know.

An Example: Stats Anxiety...

A business school student is anxious about their statistics course, since they've heard the course is difficult. The professor provides last term's final exam marks to the student. What can be discerned from this list of numbers?



Key Statistical Concepts...

Population

— a population is the group of *all* items of interest to a statistics practitioner.

— frequently very large; sometimes infinite.

E.g. All 5 million Florida voters (per Example 12.5).

Sample

— A sample is a set of data drawn from the population.

— Potentially very large, but less than the population.

E.g. a sample of 765 voters exit polled on election day.

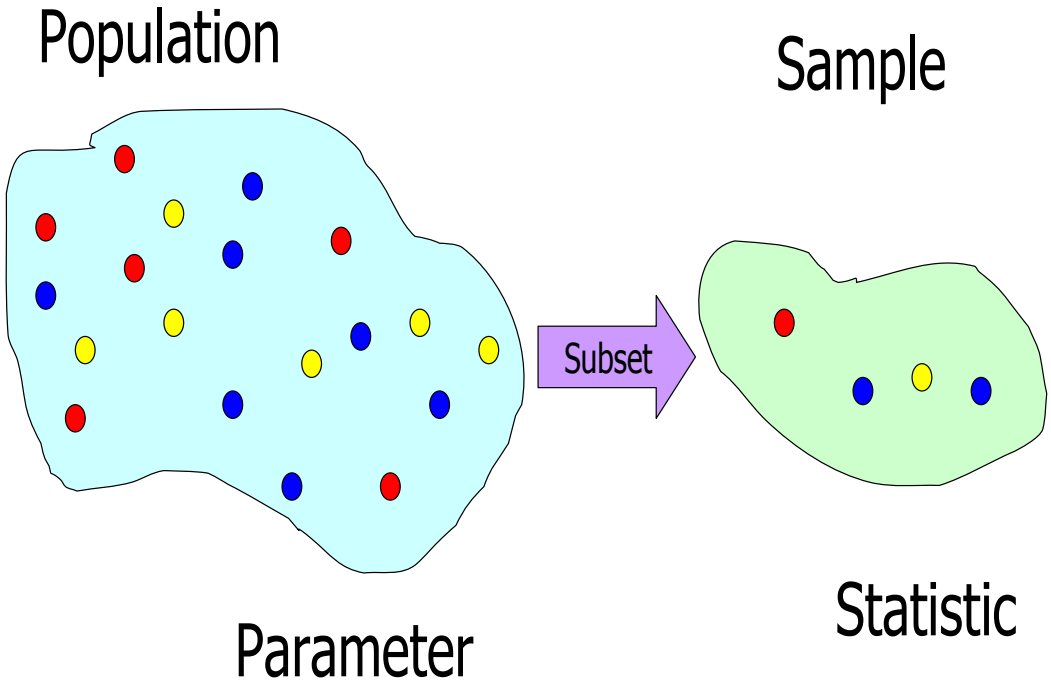
Parameter

— A descriptive measure of a *population*.

Statistic

— A descriptive measure of a *sample*.

Pictorially, we have. . .



Populations have Parameters,

Samples have Statistics.

Descriptive Statistics. . .

. . . are *methods* of organizing, summarizing, and presenting data in a convenient and informative way. These methods include:

Graphical Techniques (Chapter 2), and

Numerical Techniques (Chapter 4).

The actual method used depends on what *information* we would like to extract. Are we interested in. . .

measure(s) of central location? and/or

measure(s) of variability (dispersion)?

Descriptive Statistics helps to answer these questions. . .

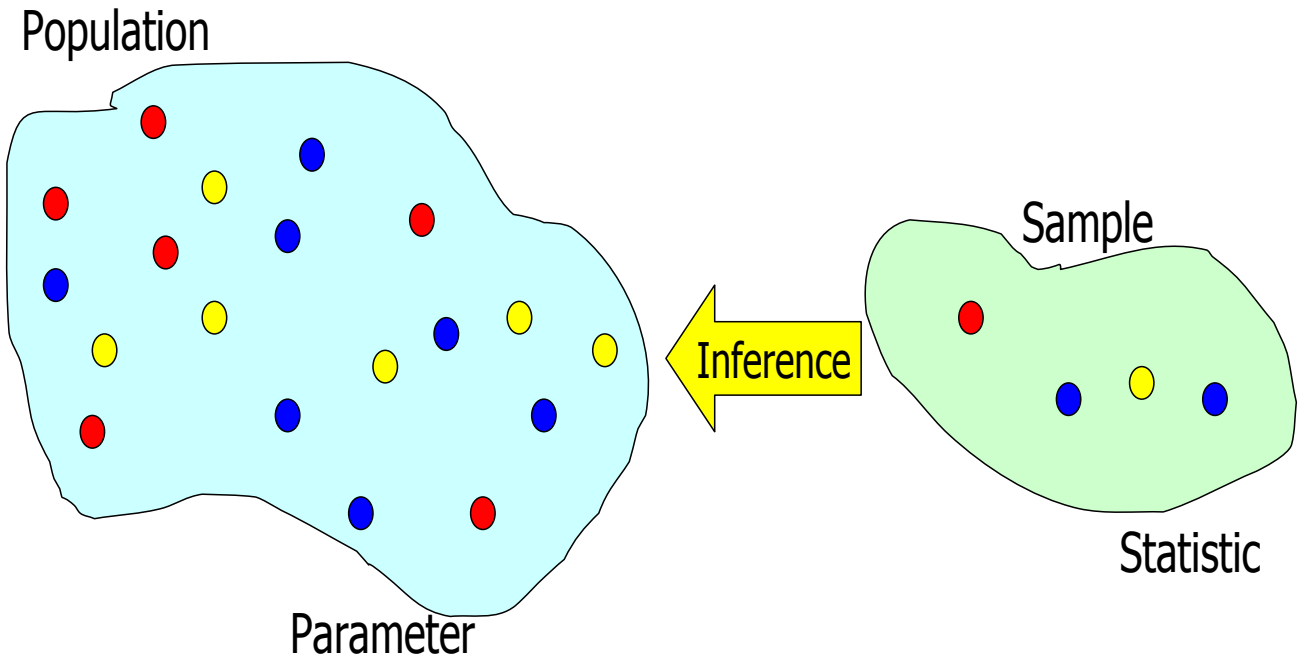
Inferential Statistics...

Descriptive Statistics describe the data set that's being analyzed, but doesn't allow us to draw any conclusions or make any interferences about the data. Hence we need another branch of statistics: *inferential statistics*.

Inferential statistics is also a set of methods, but it is used to draw conclusions or inferences about characteristics of *populations* based on data from a *sample*.

Statistical Inference...

Statistical inference is the process of making an estimate, prediction, or decision about a population based on a sample.



What can we *infer* about a Population's Parameters based on a Sample's Statistics?

We use *statistics* to make inferences about *parameters*.

Therefore, we can make an estimate, prediction, or decision about a *population* based on *sample* data.

Thus, we can apply what we know about a sample to the larger population from which it was drawn!

Rationale:

Large populations make investigating each member impractical and expensive.

Easier and cheaper to take a sample and make estimates about the population from the sample.

However:

Such conclusions and estimates are *not* always going to be correct. For this reason, we build into the statistical inference "measures of reliability," namely *confidence level* and *significance level.*

Confidence and Significance Levels...

The *confidence level* is the proportion of times that an estimating procedure will be correct.

E.g. a confidence level of 95% means that, estimates based on this form of statistical inference will be correct 95% of the time.

When the purpose of the statistical inference is to draw a conclusion about a population, the *significance level* measures how frequently the conclusion will be wrong in the long run.

E.g. a 5% significance level means that, in the long run, this type of conclusion will be wrong 5% of the time.

If we use α (Greek letter "alpha") to represent significance, then our confidence level is $1 - \alpha$.

This relationship can also be stated as:

$$\text{Confidence Level} + \text{Significance Level} = 1$$

Consider a statement from polling data you may hear about in the news:

"This poll is considered accurate within 3.4 percentage points, 19 times out of 20."

In this case, our confidence level is 95% (19/20 = 0.95), while our significance level is 5%.

Examples of Business Problems...

Accounting: An auditor is interested in the costs of business travel

Organization Behavior: A manager is interested in why turnover of employees seems to have increased

Marketing: Your firm wishes to expand its product line and determine which products are of interest to consumers

Finance: The CEO wishes to understand what factors are affecting the firm's stock price

Economics: Your firm is interested in the status of the economy over the next year

Operations Research/Management: Store managers have been reporting increased waiting lines at checkout counters

Management Information Systems: You are contemplating replacing your financial reporting system

International Management: You are interested in expanding your firm's services to a non-US market