

STATISTICS

Statistics is the science of data. This involves collecting, summarizing, organizing, and analyzing data in order to draw meaningful conclusions about the universe from which the data is obtained.

Statistics

Descriptive Statistics

Inferential statistics

Descriptive Statistics

Descriptive Statistics is the science of collecting, organizing, and summarizing data, in addition to utilizing numerical and graphical methods to present and reveal information about the data.

Inferential Statistics

A statistical inference is an estimate or prediction or some other generalization about a population based on information contained in a sample.

Measure of reliability

A measure of reliability is a statement (usually quantified) about the degree of uncertainty associated with a statistical inference.

Example

How many vehicles does a typical family own? Thirty families were polled and the following results are obtained

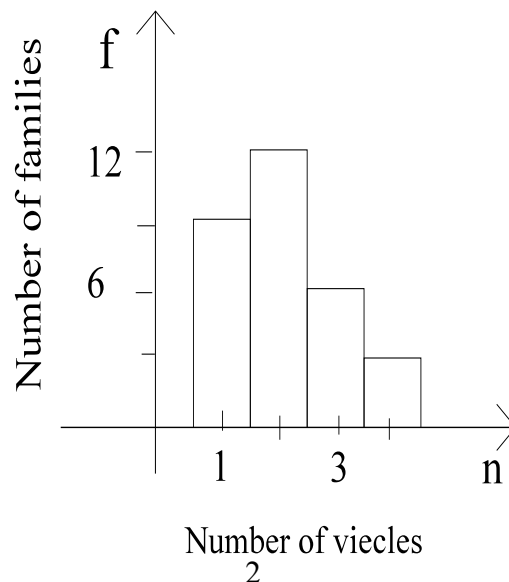
1 3 4 2 1 1 3 2 2 2
2 3 1 1 4 3 4 2 2 1
2 2 1 1 2 2 3 2 1 3

Before

Vehicles	Frequency	Rel. Freq.
1	9	0.3
2	12	0.4
3	6	0.2
4	3	0.1

Frequency distribution

After



The Fundamental Elements of Statistics

(Descriptive)

population

A population is a set of units (equally people, objects, transactions or events) that we are interested in studying.

Variable

A variable is a characteristic or property of an individual population unit.

Sample

A sample is a subset of the population.

Parameter

A statement or measurable characteristic pertaining to the Population.

Statistic

A statement or measurable characteristic pertaining to the sample.

The Fundamental Elements of Statistics

(Inferential)

Statistical inference

A statistical inference is an estimate or prediction or some other generalization about a population based on information contained in a sample.

Measure of reliability

A measure of reliability is a statement (usually quantified) about the degree of uncertainty associated with a statistical inference.

“The Uncertainty when cast in statistical terms, turns into variations.”

The methods of statistics applied to data reveal variations inherent in the data which can be used to summarize and predict outcomes form representative subsets of the population from which the data is obtained.

Example:

To be able to predict the amount of rain fall during rain fall season in the South West USA, a sample of 500 days is randomly selected from the last 100 years of rain fall season.

Population: The days of the year during rain fall season for the last century.

Variable: The number of inches rain fall each day in the Southwest USA during rain fall season.

Sample: Simple random sample of 500 measurements of rain fall in inches, one per day.

Descriptive Statistics:

- A Frequency Distribution

- A bar chart

- The average daily rain fall in inches during rain fall season

Inferential Statistics

- The average rain fall in inches per dy during a rain fall season in the South West USA next year.

- The error in the prediction of the average. rain fall per day next year.

1.2 The Need to Model Uncertainty

Deterministic outcomes

The outcome is determined by methods that produce certain outcomes, up to error in measurements.

Uncertain outcomes

The outcomes can not be determined with any degree of certainty.

Statistical outcomes

Outcomes that depend on the analysis of data.

Probabilistic Outcomes

Outcomes that relate to chance occurrence.

1.3 Random Variables And Distributions

Variables whose values can not be predicted with certainty are identified as random variables.

Examples:

- The weight of a new born baby.
- The closing price of an index in a financial market.
- The number of minutes a client waits in a teller line in a bank.
- The outcome of coin flip. (Head or Tail)

1.4

DATA TYPES

Qualitative

A qualitative data are measurements that can not be measured on a natural numerical scale; they can only be classified into one of a group of categories.

Examples

1. The inventory of a set of rain coats at a department store are classified by size as Small (S), Medium (M), Large(L) or Extra Large(XL)

2. The classification of group of 10 people as male or female.

M M F F M F F F M M

1.4 DATA TYPES

Quantitative or numerical:

Discrete

The data consists of counts

Continuous

The data consists of measurements that are recorded on a naturally occurring numerical scale.

Example Discrete

1. A sample of data that gives the number of school age children per family in santa Clara County.

A sample of twenty families yielded the following:

1 1 2 3 2 1 4 3 2 1 1 3 2 2 2 4 1 3
1 3

Example Continuous

2. The high daily temperature in Fahrenheit in Down Town Los Angeles during the month of July in 1998 .

101 98 99 94 90 85 84 90 91 92 88 89 92
....

1.5

Producing Data

1. Sampling

2. Experiments

3. Surveys

4. Census

5. Simulations

Random samples

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Proportionally stratified sampling
- Cluster sampling
- Convenience sampling

1.6 **Statistical Experiments**

Components of Comparative Experiments:

Experimental units

Explanatory Variable (Factor)

Levels or Categories

Response Variable

Responses to the different levels or categories

Randomization

Replication

Treatment

Placebo

Treatment Group

Control Group

Confounding

Blocking

Randomizations

Using chance to assign experimental units to experimental group.

Confounding

The affect of the lurking, unaccounted for variables on response of the experimental units to the treatment.

Replication

The idea of reducing chance variation in the response to treatment by the administration of the treatment to a large number of experimental units.

Blocking

A group of experimental units known to be similar in how they affect the response to the treatment.

