Biostatistics

DR.ALAA MOHAMMED

3RD LEVEL, 2ND SEMESTER, BIOMEDICAL INSTRUMENTATION AND BIOMECHANICS BRANCHES, BIOMEDICAL ENG. DEPARTMENT

Review of Engineering Statistics

LECTURE 1

Biostatistics

Francis Galton is called as the 'Father of Biostatistics'. He created the statistical concepts 'correlation'. Also, he is the first researcher which used the statistical tools to study differences among human population.

What is Statistics?

Statistics is a branch of applied mathematics which deals with the collection, classification, analysis and interpretation of data.

What is Biostatistics?

Biostatistics is a branch of biological science which deals with the study and methods of collection, presentation, analysis and interpretation of data of biological research.

Biostatistics is also called as biometrics since it involves many measurements and calculations.

In biostatistics, the statistical methods are applied to solve biological problems.

A word of *Biostatistics* made from *biology* and *statistics*

Biostatistics

Biostatistics focuses on methods with applications in biomedical sciences, including

- public health : epidemiology, health services, environmental health
- design and analysis of clinical trials
- genetics and molecular biology

Review of Engineering Statistics

Types of variables in statistics:

Statistical variables can be classified based on two criterion:

- Nature of variables (Numerical and Categorical).
  - Numerical (discrete and continuous)
  - Categorical (ordinal and nominal).
- Source of variables (Primary data and Secondary data).

Review of Engineering Statistics (cont.)

Statistics deals with methods and techniques that can be used to draw conclusions about the characteristics of a large number of data points--commonly called a population by using a smaller subset of the entire data.

Statistics is sometimes described as the science of decision making under uncertainty and can be divided into two broad areas:

Descriptive Statistics and Inferential Statistics

- **Descriptive Statistics** Quantities and techniques used to describe a sample characteristic e.g. mean, standard deviation, box-plot.
- **Inferential Statistics** which covers those statistical procedures used to help draw conclusions or inferences about a population on the basis of a sample of data collected from the population.

Review of Engineering Statistics (cont.)

Variables are the quantities measured in a sample

Random Variable is the data may come from a survey, a questionnaire or from an experiment e.g. heights of university students.

They may be classified as:

- **Quantitative** i.e. numerical
  - **Continuous** (e.g. pH of a sample, patient cholesterol levels)
  - **Discrete** (e.g. number of bacteria colonies in a culture)
- **Categorical**
  - **Nominal** (e.g. gender, blood group, hair colour, marital status)
  - **Ordinal** (ranked e.g. mild, moderate or severe illness). Often ordinal variables are re-coded to be quantitative.

Review of Engineering Statistics (cont.)

Interval - Values of the variable are ordered as in Ordinal, and additionally, differences between values are meaningful, however, the scale is not absolutely anchored. Calendar dates and temperatures on the Fahrenheit scale are examples.

Population: the complete set of individuals, objects or scores of interest.

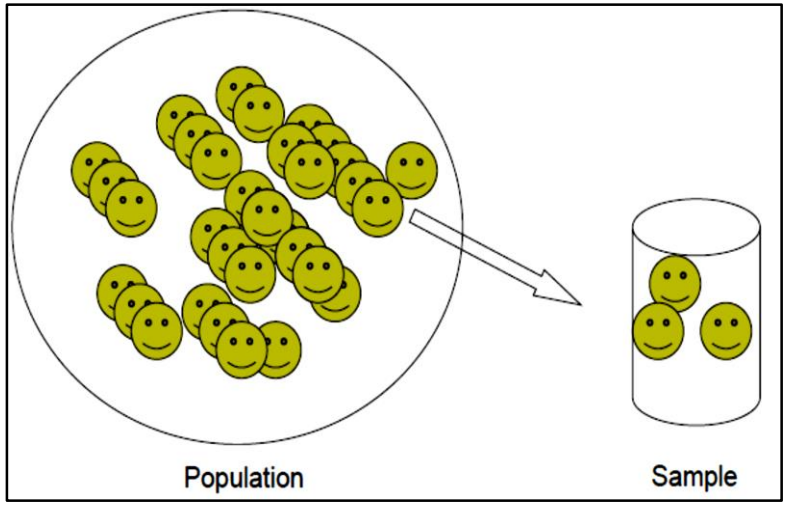
- Often too large to sample in its entirety
- It may be real or hypothetical (e.g. the results from an experiment repeated ad infinitum)

Review of Engineering Statistics (cont.)

Sample: A subset of the population.

- A sample may be classified as **random** (each member has equal chance of being selected from a population) or **convenience** (what's available).
- Random selection attempts to ensure the sample is representative of the population.

❑ **Sample vs. Population**



Descriptive Statistics

❖Descriptive Statistics

- Describes the important characteristics of a set of data.
- Organize, present, and summarize data:
  1. Graphically
  2. Numerically

Types of descriptive statistics of Data:

1. **Organize Data**
   - Tables
   - Graphs
     - Bar Chart or Histogram

Descriptive Statistics (cont.)

2. Summarize Data

- Central Tendency (or Groups' —Middle Values). A value that represents a typical, or central, entry of a data set. Most common measures of central tendency:
  - Mean
  - Median
  - Mode

$$\text{Mean} = \frac{\text{Sum of all entries}}{\text{Number of entries}}$$

$$\text{Median} = \frac{\text{Sum of middle two values}}{2}$$

Descriptive Statistics (cont.)

The *sample* arithmetic *mean* is defined as: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

For *population* use the parameter μ: $\mu = \frac{1}{N}\sum x$

- Variation (or Summary of Differences Within Groups)
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation
  - *coefficient of variation etc*.

Measures of Variability (Measures of Dispersion)

1. Range: is the difference between the largest and smallest values of a data distribution or observations in the sample

$$\boldsymbol{Range = Largest\,Value - Smallest\,Value}$$

2. Sample Variance:

The **sample variance**, s^2, is the arithmetic mean of the squared deviations from the sample mean:

The variance of a set of observations is the average of the squares of the deviations of the observations from their mean.

$$\boldsymbol{s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}}$$

Measures of Variability (cont.)

In symbols, the variance of the n observations $x_1, x_2, \ldots, x_n$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Variance of 5, 7, 3? Mean is (5+7+3)/3 = 5 and the variance is

$$\frac{(5-5)^2 + (3-5)^2 + (7-5)^2}{3-1} = 4$$

Measures of Variability (cont.)

3. Standard Deviation

The standard deviation of a data set is the positive square root of the variance.

Sample Standard Deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Population Standard Deviation $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

s has the advantage of being in the same units as the original variable **x**

The standard deviation for a grouped distribution is calculated from:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2 f}{n-1}}$$

Finding Sample Variance & Standard Deviation

1. Find the mean of the sample data set.

$$\bar{x} = \frac{\Sigma x}{n}$$

2. Find deviation of each entry.

$$x - \bar{x}$$

3. Square each deviation.

$$(x - \bar{x})^2$$

4. Add to get the sum of the deviations squared.

$$\Sigma (x - \bar{x})^2$$

5. Divide by $n - 1$ to get the **sample variance**.

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$

6. Find the square root to get the **sample standard deviation**.

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}}$$

Inferential Statistics

- Inferential Statistics

Inferential Statistics which covers those statistical procedures used to help draw conclusions or inferences about a population on the basis of a sample of data collected from the population.

Important areas inferential statistics include confidence intervals, hypothesis tests, regression analysis and experimental design.

Underlying inferential statistics is the idea of probability and probability distributions.

Methods of presentation of data

Numerical presentation

Graphical presentation

Mathematical presentation

Smoking	Lung cancer				Total	
	Cases		Control			
	No.	%	No.	%	No.	%
Smoker	15	75%	8	20%	23	38.33
Non smoker	5	25%	32	80%	37	61.67
Total	20	100	40	100	60	100

Types of variables

- *Numerical:* variables that take on numerical values.
  - Examples: weight, age, blood pressure, body temperature, annual expenditure on health product ...
  - computable and ordered (<,>,=)
- *Categorical:* variables that take values corresponding to categories
  - for example, gender, race, eye color, blood type(A,B,AB,O), final grade(A,B,C,F)
  - incomputable; unordered or ordered

Numerical variables

Types of numerical variables

1. Discrete: integer value
   - years of age
   - number of children
   - number of accidents
   - number of hospital visits

2. Continuous: real numbers, theoretically these can go to unlimited decimal points
   - age calculated from date of birth to study date
   - weight and height
   - alcohol consumption
   - price

Categorical variables

Types of Categorical variables

1. Nominal (unordered):

- – gender: male/female
- – smoking status: smoker/non-smoker
- – major: public health, biology, other

2. Ordinal (ordered)

- – smoking status: non-smoker/light/heavy
- – grades: A/B/C/F
- – Olympic medals: gold/silver/bronze
- – your steak: rare, medium, well-done
- – exercise frequency: often/rarely/never

H.W.

Find the Standard Deviation for the following data

Data	Deviation	Deviation^2
151	13.86	192.02
124	-13.14	172.73
132	-5.14	26.45
170	32.86	1079.59
146	8.86	78.45
124	-13.14	172.73
113	-24.14	582.88
Sum = 960.0	Sum = 0.00	Sum = 2304.86
$\bar{x} = 137.14$		

THANKS FOR LISTENING