# Mathematics and Biostatistics

Variables and Measurements

1st Semester 2022

Lecture 2

# Review

- **Statistics** is the branch of mathematical science that is focused on the collection, organization, and interpretation of numerical information.

- **data = numerical information = measurements = observations**

- **Biostatistics** is a specialized discipline of statistics that deals with statistical applications in the biological and health sciences.

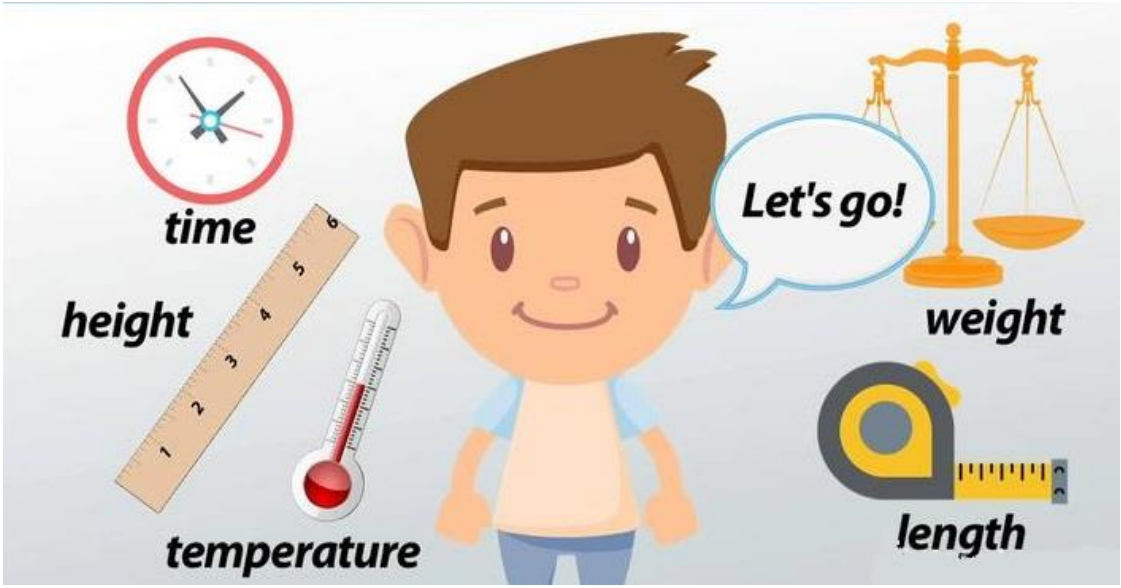
- The design of health surveys, clinical trials, vital statistics, cancer survivorship studies and biological field studies are some specific biostatistical applications.

- Inference is making a generalization from a few specific measurements (sample) to a larger set of measurements (population).

# Measurement

- Measurement is the assignment of a **numerical value** to some property or characteristic of a person or object.

- This simple definition **covers** the assignment of numeric codes for properties like gender, blood type, spoken languages or other non-quantitative properties.




Note: (These **codes do not** support arithmetic or numerical computation)

# Properties of Measurements

- **identity,** is the ability to identify differences between items regarding the property of the item being measured.

- **order** is the ability to rank the items based on the measurement. The ordering depends on whether one item has more or less of the property than another item.

- A **scale** of measurement allows one to not only say one item has more of a characteristic, but how much more or less in terms of units on a scale.

# Quantitative Types Of Variables

- **Discrete:** There are no fractional values between any two values, measurements or codes.

- **Examples:**

1. number of patients,
2. number of siblings,
3. rankings on a scale of 1 to 5, categories of spoken languages coded with numbers. Say 1=Arabic, 2=English, 3=French, 4=Chinese, etc.

# Quantitative Types Of Variables

- **Continuous:** There are no ‘breaks’ in the set of possible measurements. Between any two possible values there always can be found another value,
- Examples: Distance, mass, height.

## Notes:

The issue is to what degree of accuracy do require to carry out the work.

# Qualitative Types Of Variables

- **Nominal:** A variable with value assignments for a characteristic which only has the property of identity.
- A ‘measurement that has this property allows items to be identified as different
- Example: Blood types, female vs. male, Green vs. Red.

Note:

1. There is no quantitative aspect to such a ‘measurement’ or coding assignment.
2. Such variables are always discrete categories.

# Qualitative Types Of Variables

- **Ordinal:** A variable with a coding which has the properties of identity, and order indicating more or less of an attribute.
- An example is patient prognosis*: Poor, Fair, Good, Excellent. University Graduation degrees : Bachelor, Master, Doctorate.
- Note:

1. all of these reduce to a ranking.
2. also, that such rankings are discrete.

*prognosis : an opinion, based on medical experience, of the likely course of a medical condition.

# Other Types of Variables

- **Interval:** A variable with properties of identity, order and a scale of measurement. The scale quantitatively indicates how much more or less of a characteristic one item has relative to another.
- An interval measurement or variable does not have a true zero point. A value of zero does not mean a complete absence of the characteristic.
- Example: temperature in Celsius or Fahrenheit scale,
- where 0° does not correspond to an absence of heat (no true zero).

# Other Types of Variables

- Ratio: A variable with measurements which holds identity, order, scale of measurement, and has an absolute or true 'zero' ( distance, time, mass).
- Ratio is the most advanced level of measurement, which can handle most types of mathematical operations.

Ratio examples

1. Range of motion, no movement corresponds to zero distance.
2. The interval between 10 and 20 degrees is the same as between 40 and 50 degrees
3. Weight-lifting capacity, a person who is unable to lift scores zero. A person who lifts 30 kg can lift twice as much as one who lifts 15 kg.

# Rounding of Numbers

- How we round is related to the level of precision in the measurement. A question that has to be answered is: to what nearest digit have these numbers been measured? Is it the nearest whole number, nearest, $10^{th}$, nearest $100^{th}$ etc.

- When we report data in tables, or as statistics, then we only report to the level of precision in the measurements. So if you had values measured to the nearest $100^{th}$ you could only report values to two decimal places: ex: 123.45

# Rounding of Numbers

- When rounding for this course follow this rule:

- If the ‘extra’ digit is a 0,1,2,3 or 4 do not round up. While if the extra digit is a 5, 6, 7, 8 or 9 round up.

Ex: Values to nearest $10^{th}$: 12.45, 16.12 are rounded to 12.5 and 16.1.

# Rounding of Numbers

- A further annoyance is that if you report a weight like 80.0 kg, that value lies in an interval of 79.5 up to 80.4 inclusive.
- You may recall intervals in mathematics given as: [139.5, 140.4] or [139.5, 140.5).
- The square bracket implies the endpoint is in the interval, while the rounded bracket indicates the value is not, but that every value up to but not including 140.5 is included.
- Every now and then this type of issue pops up and can be annoying.

# Distribution of Values

- A primary concept in statistics is that of the distribution of the values for a variable.
- The ‘distribution’ is the frequency or relative frequency with which each value occurs.
- The relative frequency is the proportion of times a given value occurs. In line with the frequentist idea of the probability of a value occurring.
- The distribution can be viewed as a graph where the ‘**X**’ axis lists the possible values of the variable, and the ‘**Y**’ axis gives the frequency or relative frequency with which each value occurs.

# Distribution of Values

Here are the scores on the first exam in an introductory statistics course for 15 students in one section of the course. 80 73 92 85 75 98 93 55 80 90 92 80 65 58.

# Distribution of Values

- One of the first things to consider is the  type of variable: Continuous variable or Discrete variable.
- If we have a variable which is discrete then the distribution for that variable will be discrete: with gaps between the values.
- A continuous variable will have a continuous distribution of values: no gaps (at least theoretically).
- These considerations impact the choice of graph used for displaying the distribution.

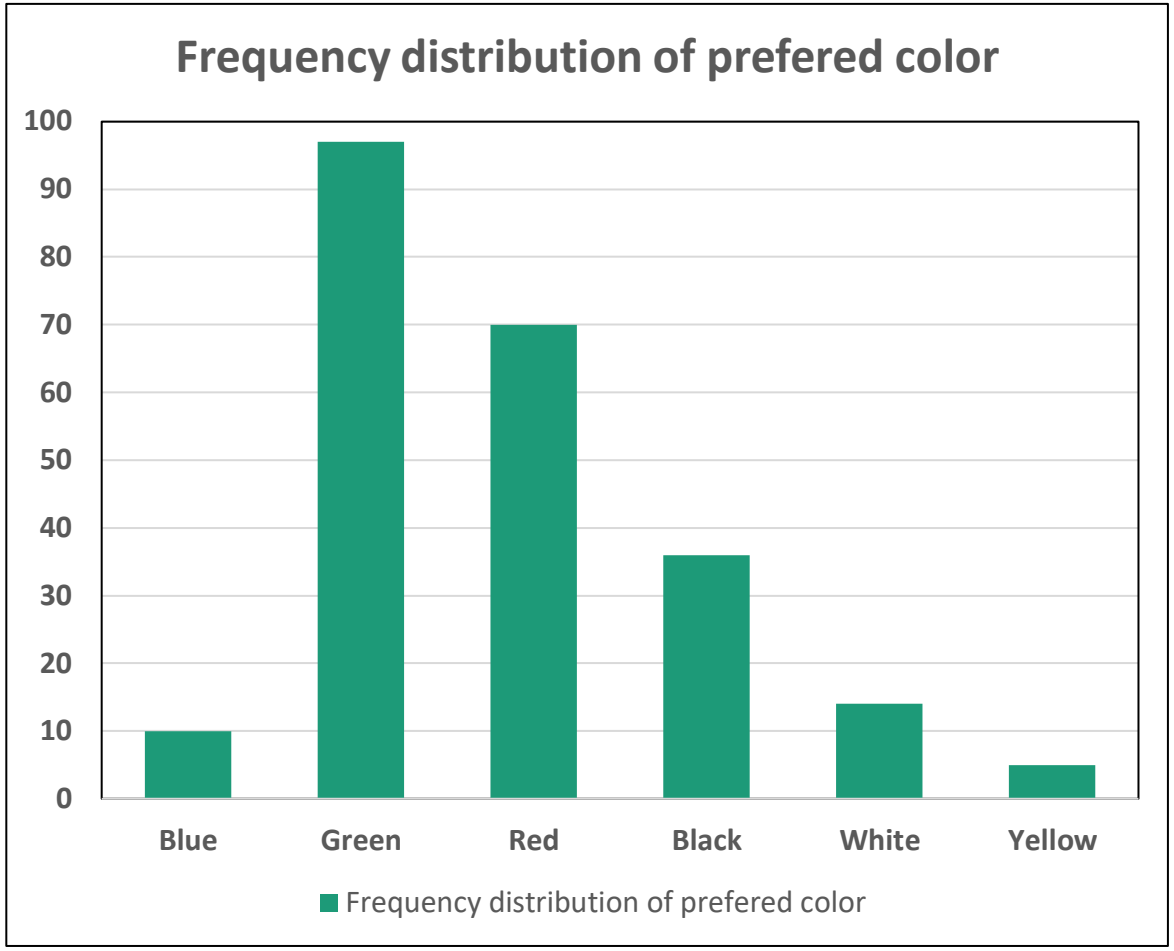
# Frequency Table

What is your favorite color? The survey about color preferences reported the distribution of the responses to that question:

| | Frequency | Relative Frequency |
|---|---|---|
| Blue | 10 | 0.04 |
| Green | 97 | 0.42 |
| Red | 70 | 0.30 |
| Black | 36 | 0.16 |
| White | 14 | 0.06 |
| Yellow | 5 | 0.02 |

# Frequency Distribution

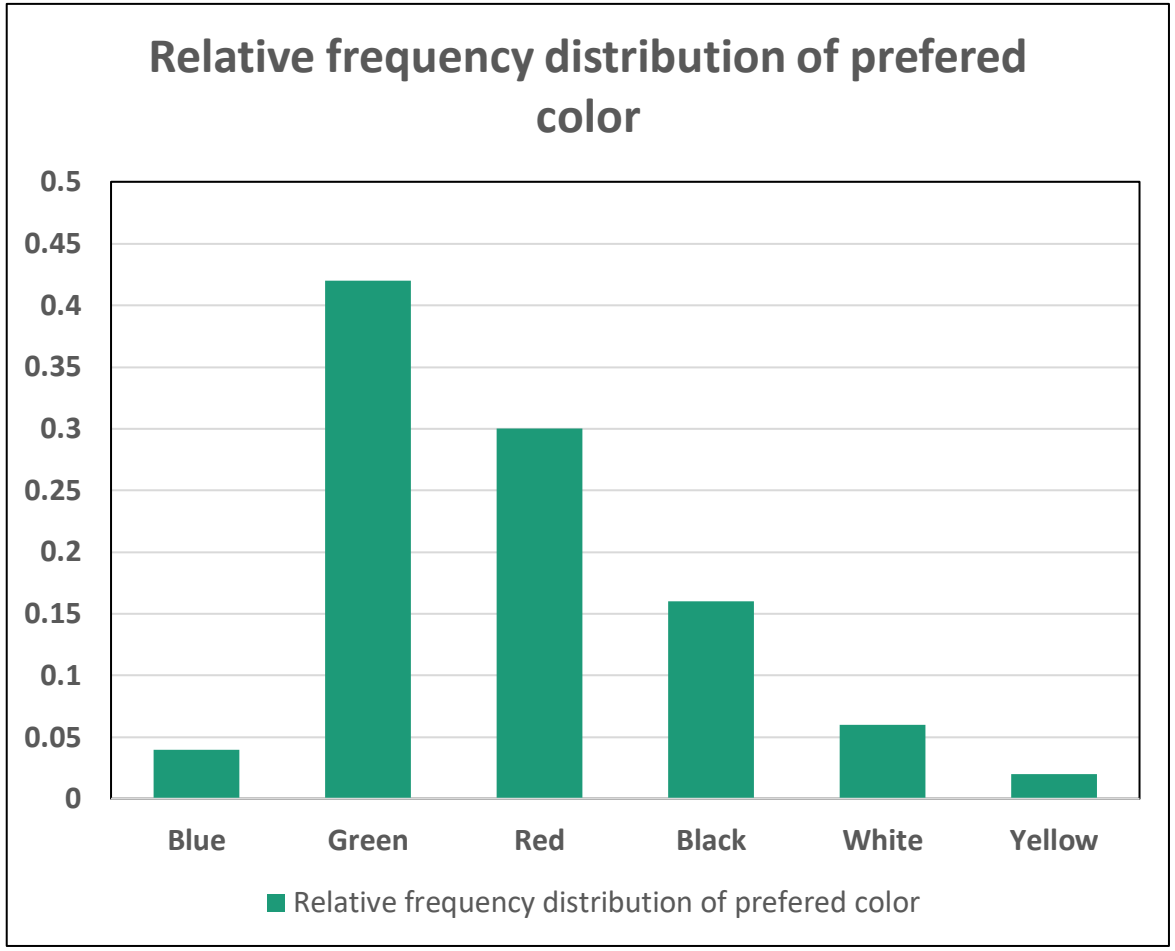
- Notice that the first column gives the color as: Blue, Green, Red, Black, White, and Yellow.
- The frequency of each color is given in the second column. The color Blue occurred 10 times while the Green occurred 97.
- There were a total of 232 persons examined (sum of the frequencies).



# Frequency Distribution

- The relative frequency is the proportional distribution and is calculated by dividing each of the frequency values by the total: 232.

- Note that the same shape is obtained by plotting either the frequencies or the relative frequencies.



# Things we look for in a distribution

1. Central Tendency
2. Variability of values and their spread
3. Shape of the distribution.
4. Gaps and clumping of values

# Central Tendency

- Measures of Central Tendency (location):

  MEAN -- average

  MEDIAN -- middle value

  MODE -- most frequently observed value.

- Arithmetic Mean: The mean of the sample is the arithmetic average of the sample values. It represents the center of data according to the size of the values.
- To calculate the mean, add all values of a series of numbers and then divided by the total number of elements.

# Arithmetic Mean

- Formula to calculate the mean:
- Mean of a sample

$$\overline{X} = \frac{\Sigma X}{n}$$

- Mean of population

$$\mu = \frac{\Sigma X}{N}$$

Where:
X is a command that adds all of the X values, n is the total number of values in the series of a sample and N is the same for a population.

# Arithmetic Mean

Example: Suppose the degree of student in exams was:

77,50,65,70,83,55,90

$$\bar{x} = \frac{77 + 50 + 65 + 70 + 83 + 55 + 90}{7} = 70$$

# Median

- The value that divides a series of values in half when they are all listed in order .*OR,* the middle value from a set of observation that has been ranked.
- When there are an odd number of values. The median is the middle value.
- When there are an even number of values. Count from each end of the series toward the middle and then average the 2 middle values.
- Example: Find the median from the set of values

a) 10, 9, 20, 23, 15, 4, 6 (odd No. & ranked)

4,6,9,10,15,20,23 → median = ?.

b) 12,20,25,27,5,8, 33,44 (even No.& ranked)

5,8,12,20,25,27,33,44 → median = ?

# Mode

- The most frequently occurring value in a series
- Example: find mode for the set of data:

- mode = ?

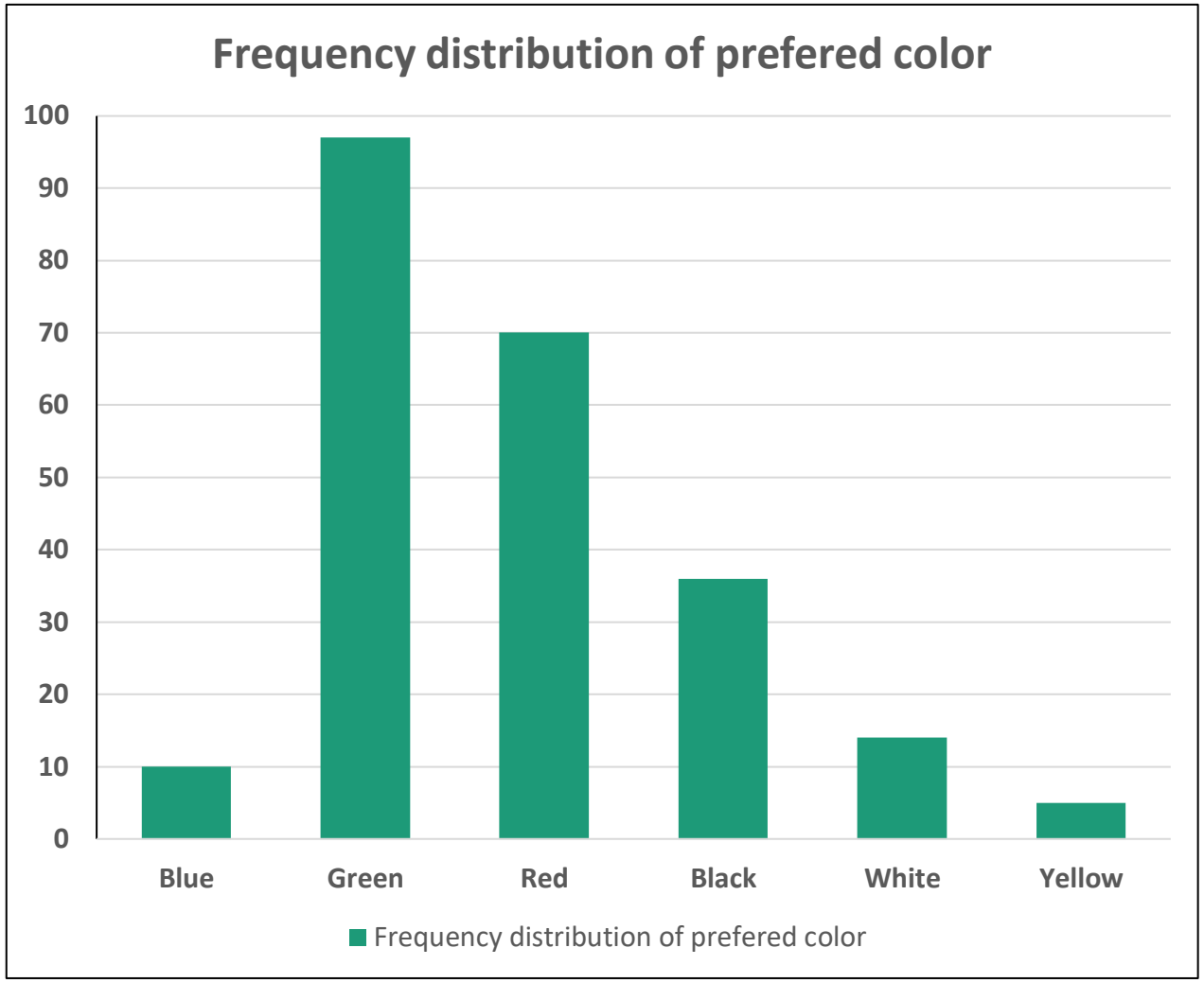
| | |
|---|---|
| 1 | 205 |
| 2 | 200 |
| 3 | 205 |
| 4 | 205 |
| 5 | 201 |
| 6 | 199 |
| 7 | 195 |
| 8 | 202 |
| 9 | 205 |
| 10 | 207 |

# Mode

The most frequently occurring value in a series

Example: find mode for the set of data shown in the plot:

mode = ?



# End of Lecture 2