# Mathematics and Biostatistics

Distribution of Values

1st Semester 2021

Lecture 3

# Review

- Measurement is the assignment of a numerical value to some property or characteristic of a person or object.
- Properties of measurements (identity, order, and scale)
- Quantitative types of variables (discrete and continuous)
- Qualitative types of variables (nominal, and ordinal)
- Other types of variables (interval, and ratio)
- Rounding of numbers
- Endpoint is in the interval

# 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.
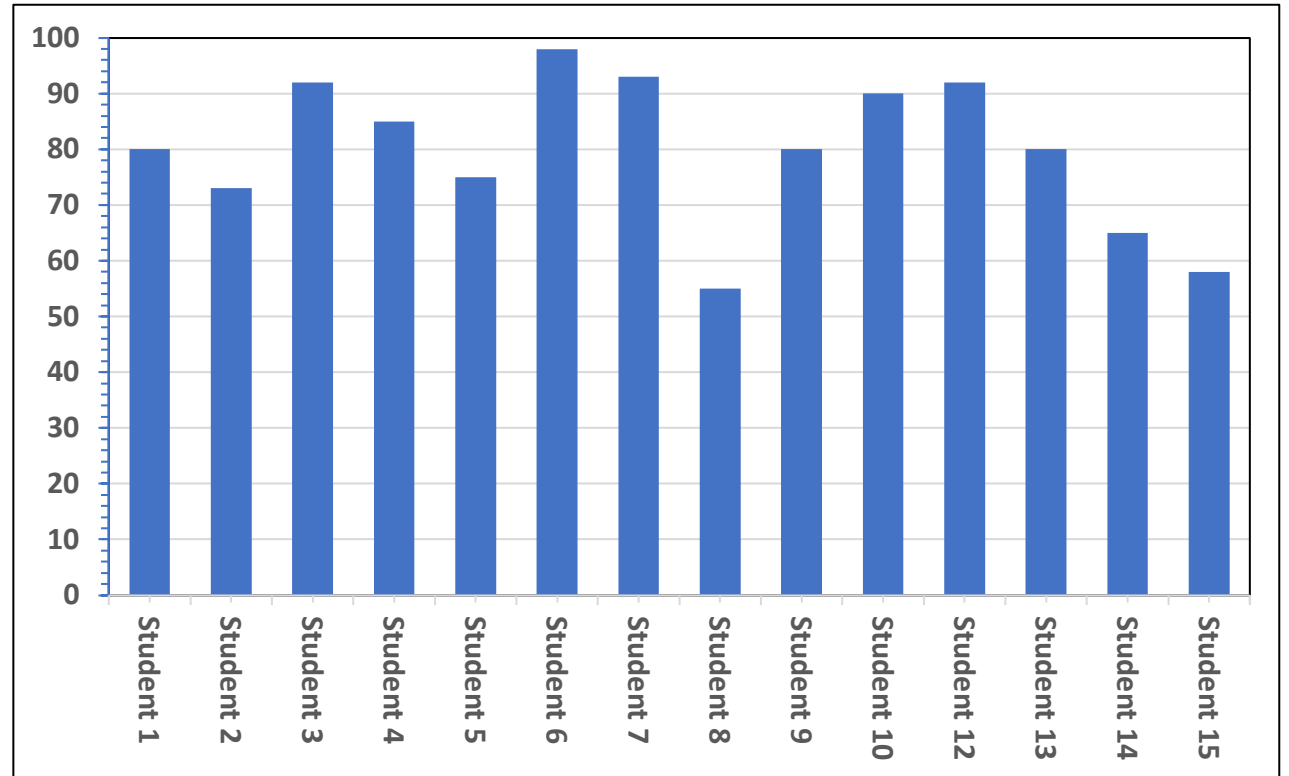
# Distribution of Values

- 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 of prefered color**
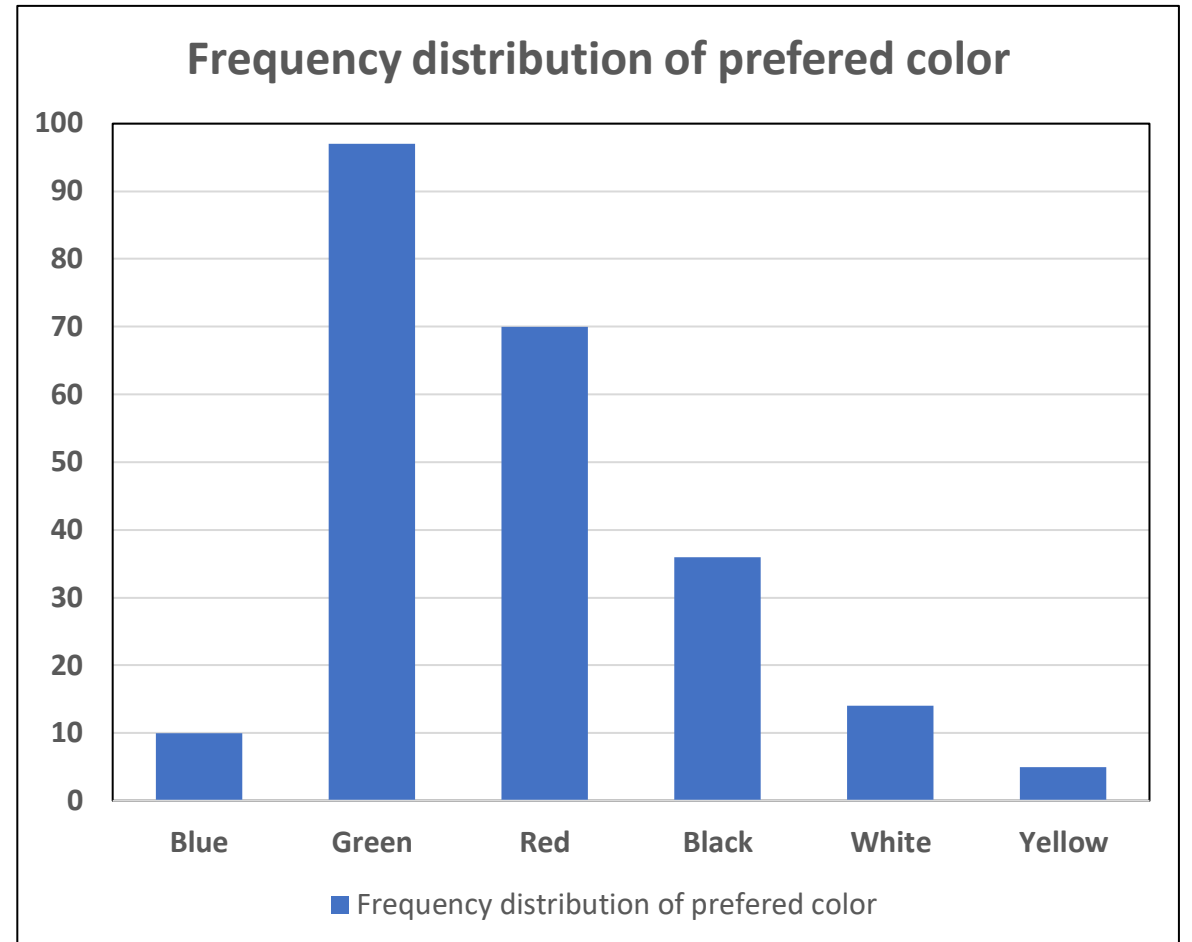
# Frequency Distribution

- 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.



Relative frequency distribution of prefered color
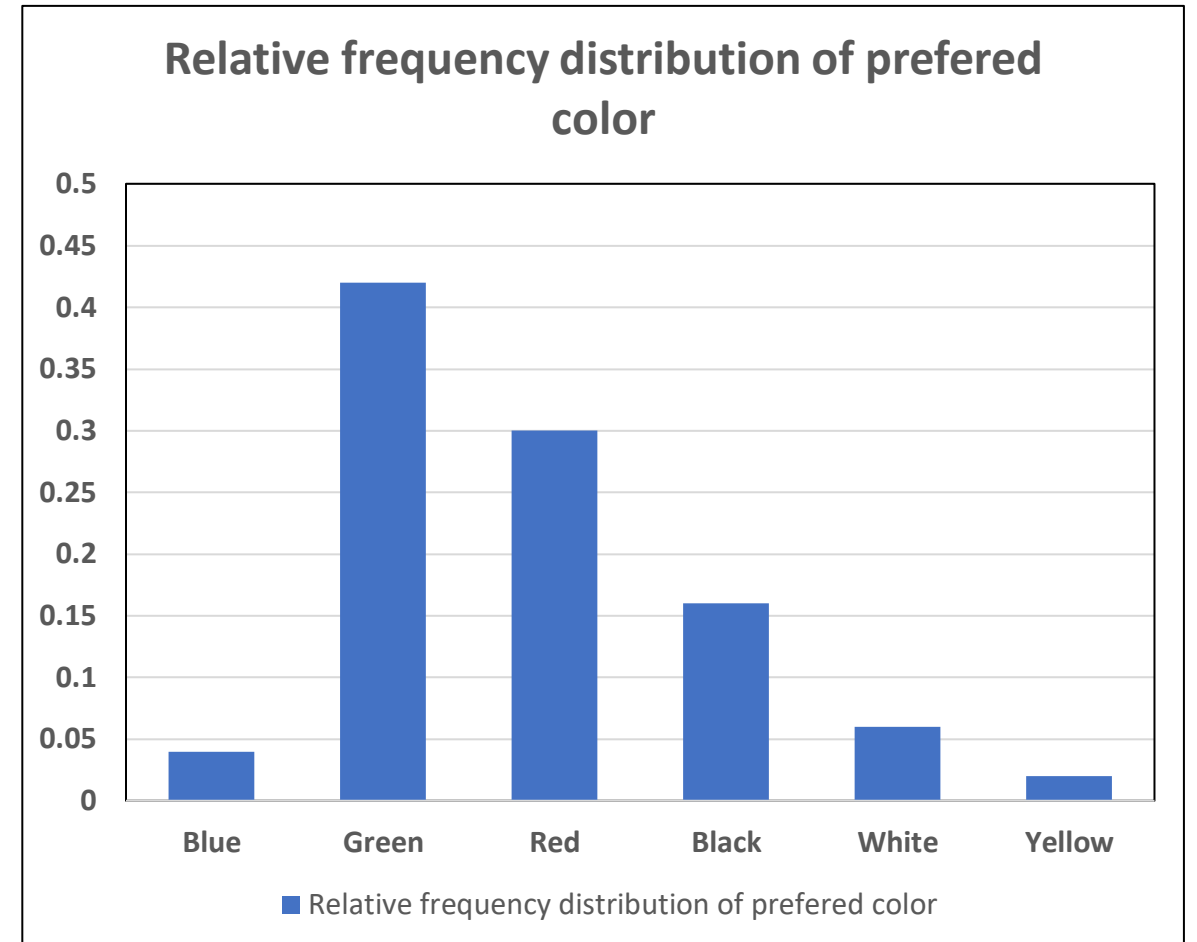
# Things we look for in a distribution

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 <u>average </u>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
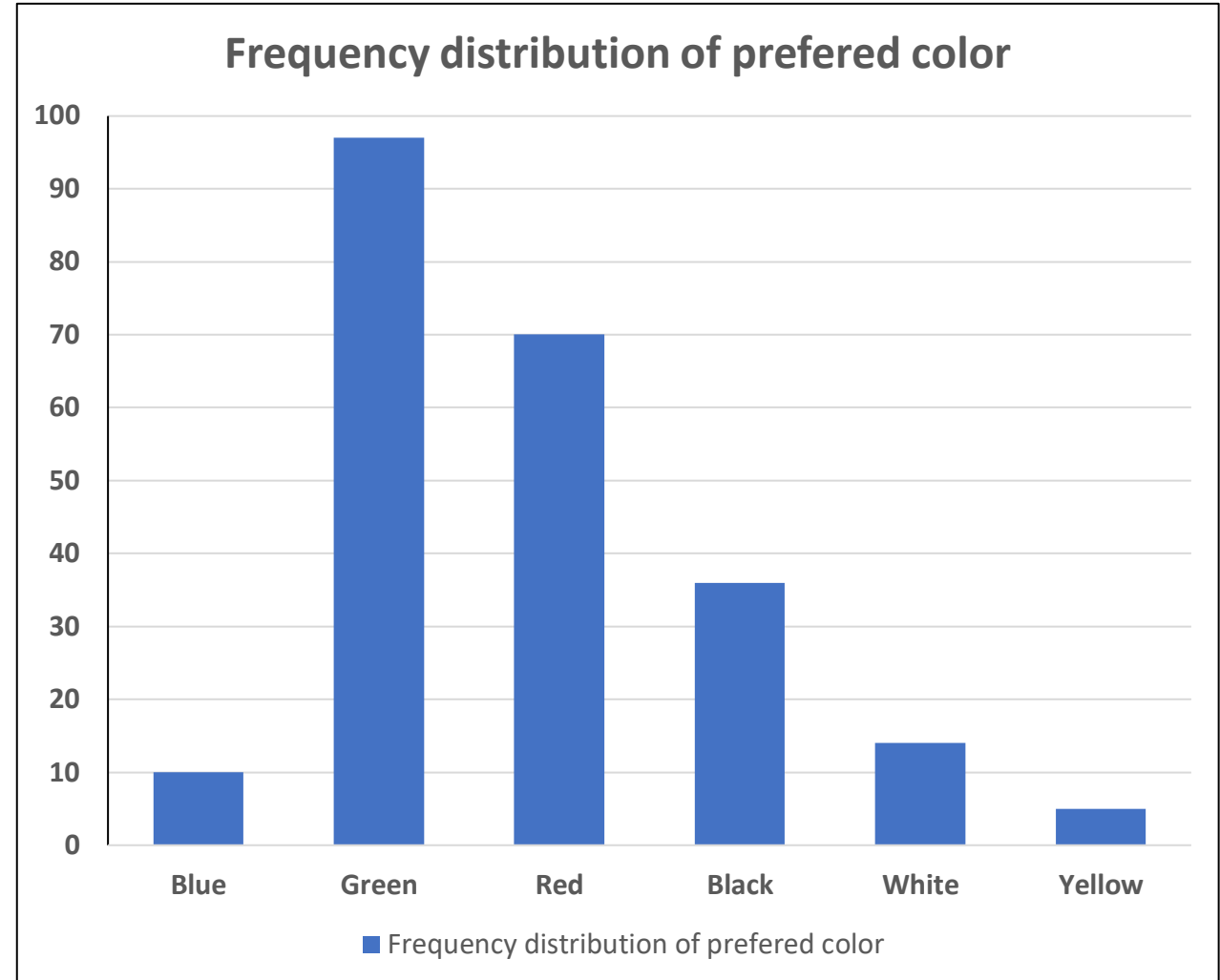- Example: find mode for the set of data:

- 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 = ?



Frequency distribution of prefered color

# End of Lecture 3