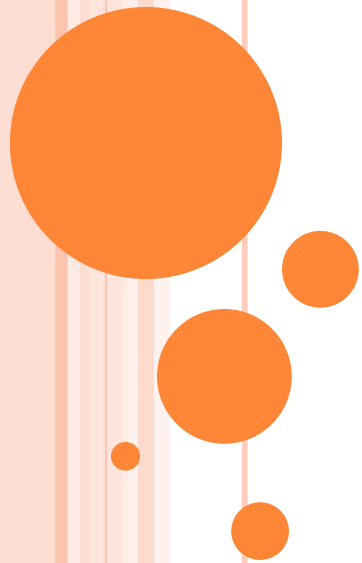


BASIC STATISTICS



THE MEANING OF A VARIABLE

- A **variable** refers to any quantity that may take on more than one value
 - Population is a variable because it is not fixed or constant – changes over time
 - The unemployment rate is a variable because it may take on any value from 0-100%
- A **random variable** can be thought of as an unknown value that may change every time it is inspected.
- A **random variable** either may be discrete or continuous
 - A variable is **discrete** if its possible values have jumps or breaks
 - Population - measured in integers or whole units: 1, 2, 3, ...
 - A variable is **continuous** if there are no jumps or breaks
 - Unemployment rate – needs not be measured in whole units: 1.77, ..., 8.99, ...



DESCRIPTIVE STATISTICS

- **Descriptive statistics** are used to describe the main features of a collection of data in quantitative terms.
 - Descriptive statistics aim to quantitatively summarize a data set
- Some statistical summaries are especially common in descriptive analyses. For example
 - Frequency Distribution
 - Central Tendency
 - Dispersion
 - Association



**Descriptive
Statistics:**

**Frequency
Distributions**



FREQUENCY DISTRIBUTION

- In statistics, a frequency distribution is a tabulation of the values that one or more variables take in a sample.
- Consider the hypothetical prices of something

Month	Price (cents/lb)
May	67.05
June	66.89
July	67.45
August	68.39
September	67.45
October	70.10
November	68.39



FREQUENCY DISTRIBUTION

- **Univariate frequency distributions** are often presented as lists ordered by quantity showing the number of times each value appears.
 - A frequency distribution may be **grouped** or **ungrouped**
 - For a small number of observations - ungrouped frequency distribution
 - For a large number of observations - grouped frequency distribution

Ungrouped	
Price (X)	Frequency
67.05	1
66.89	1
67.45	2
68.39	2
70.10	1

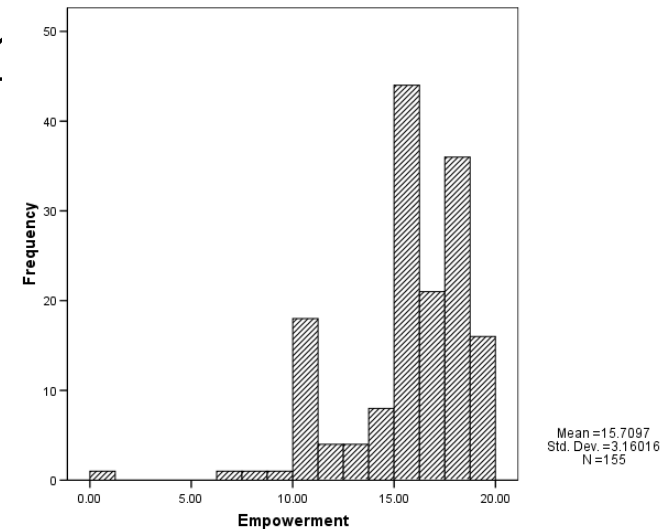
Grouped	
Price (X)	Frequency
65.00-66.99	1
67.00-68.99	4
69.00-70.99	1
71.00-72.99	0
73.00-74.99	0

- The three commonly used graphic forms are:
 - Histograms,
 - Bar charts,
 - Frequency polygons.

GRAPHIC PRESENTATION OF A
FREQUENCY DISTRIBUTION



- Histogram: A graph in which the classes are marked on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are represented by the heights of the bars and the bars are drawn adjacent to each other

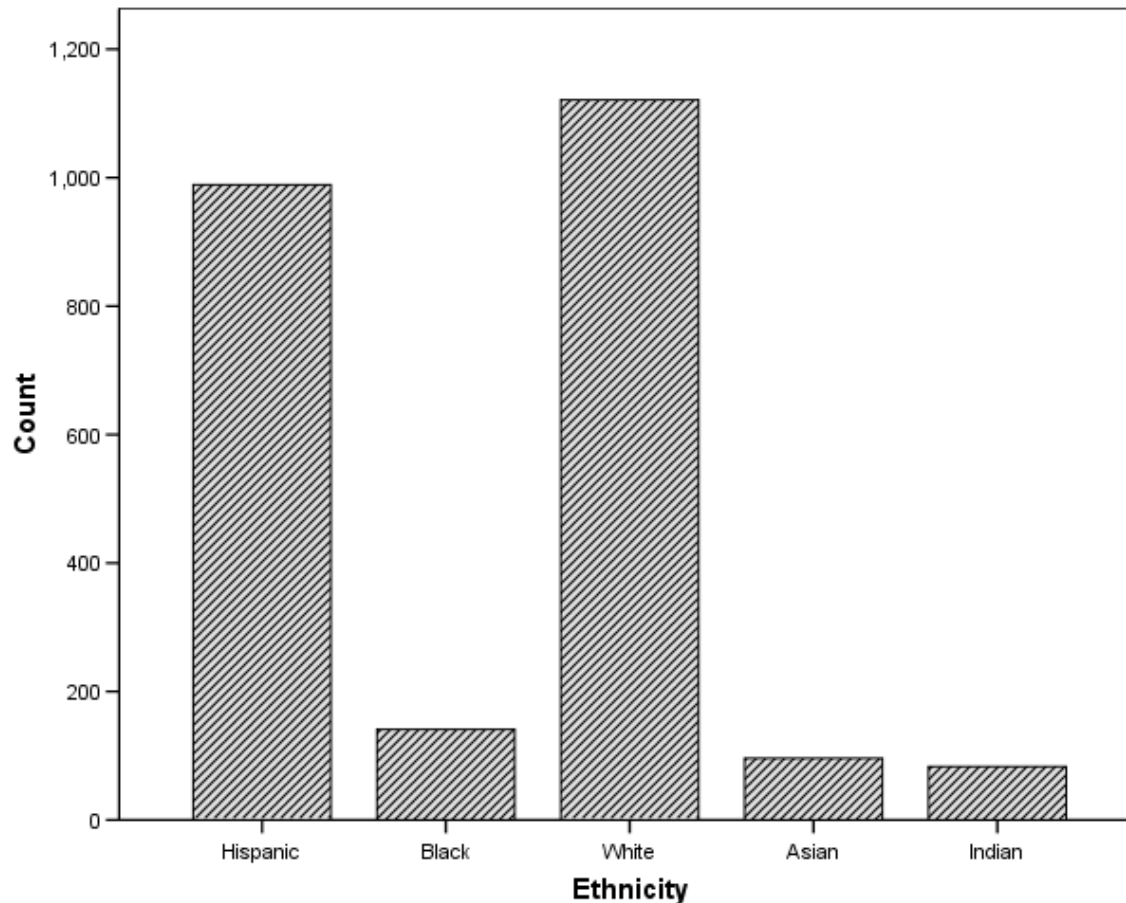


BAR GRAPHS

- A bar graph illustrates nominal data with the scores/categories along the x -axis and the frequencies on the y -axis.
- The scores are not ordered.
- The bars do not touch
(unlike a histogram).
- The heights correspond to the number of times the score occurs.

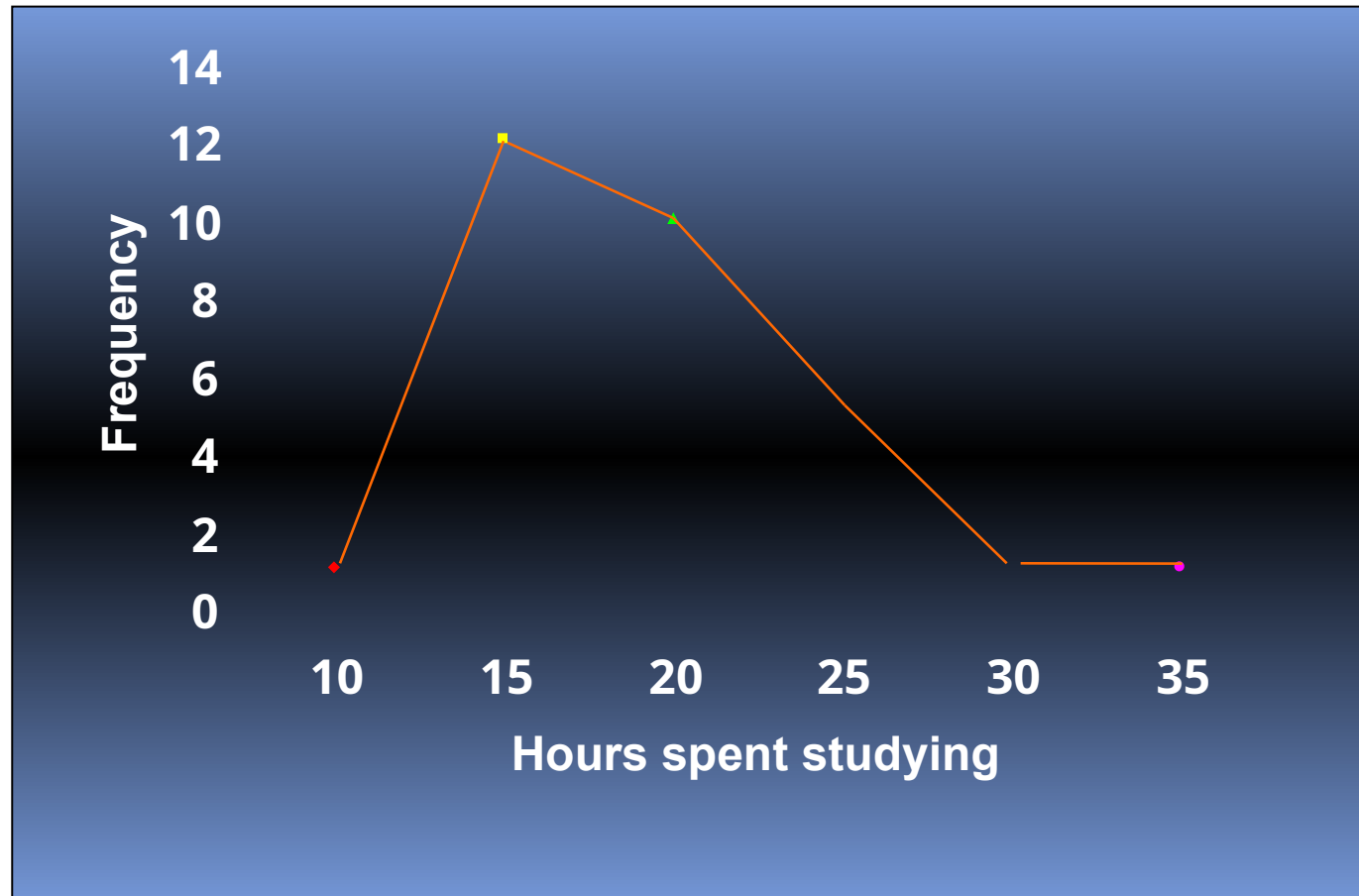


The majority of students in the sample were White. The next largest group was Hispanic. The smallest representative groups were Black, Asian, and Indian.



- A frequency polygon consists of line segments connecting the points formed by the class midpoint and the class frequency.





COMPARE THE FREQUENCY POLYGON TO THE HISTOGRAM

