Lecture 7

HIGH YEILD

INFERENTIAL :
An area of statistics that is concerned about STATISTICS methods of drawing conclusions about *a population based on a sample*.

PARAMETER
is a piece of numerical information about a POPULATION”

STATISTIC
“ is a piece of numerical information about a SAMPLE”.

ESTIMATOR :
The random variable from which a statistic is calculated

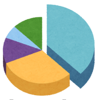
Descriptive Methods

1 Diagrams
- Pie Charts
- Bar Graphs
- Histograms

2 Measures of central tendency
- mean
- median
- mode

3 Measures of dispersion
- sample variance
- sample standard deviation

Pie Charts



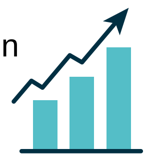
- Displays data in percentages.
- Should add to 100% adds to 99.9% due to round-off error
- Excellent in showing part vs. whole comparisons

Bar Graphs: Using frequencies

Excellent for showing Magnitude differences

Bar Graphs: Using Percentages

Allows easier comparisons between data sets of different sizes.

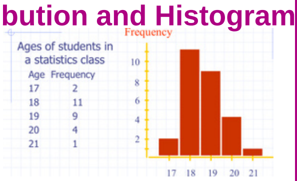


Dotplot

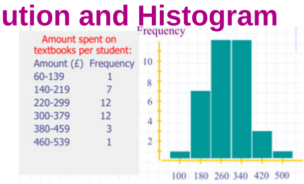
- Number line with dots representing data points
- Can visualize the "spread" of the data

Histogram

Ungrouped Frequency Distribution and Histogram

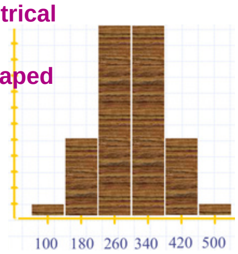


Grouped Frequency Distribution and Histogram

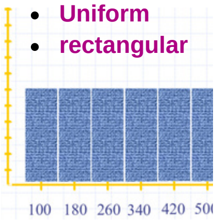


Shapes of Histograms

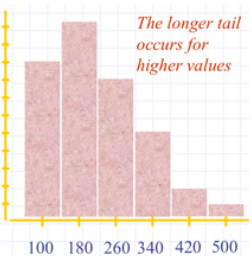
- Symmetrical
- normal
- bell-shaped



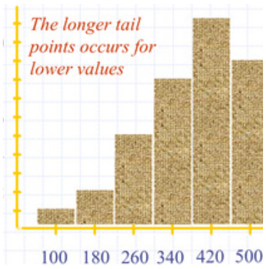
- Uniform
- rectangular

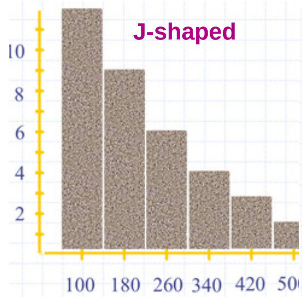


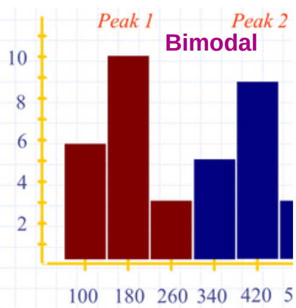
- Skewed right
- Positively skewed



- Skewed left
- negatively skewed







Mean

This is an ESTIMATOR $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

- The mean is sensitive to extreme values.

Median

- 50% of the observations are below the median and 50% are above the median
- The median is not as sensitive to extreme values as the mean.

Mode

- Most frequently occurring valueThis is the only measure of central tendency that can be used for qualitative data
- Does not always exist

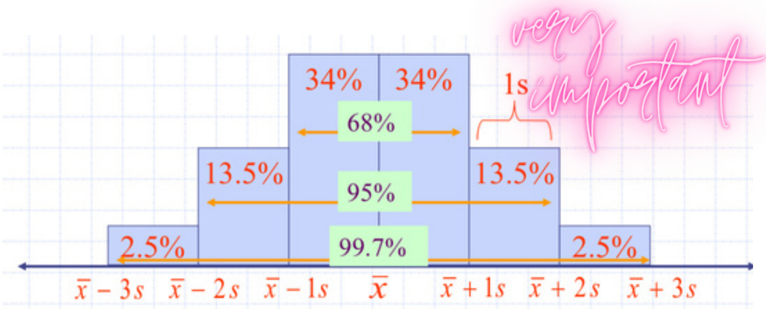
لجنة الطب والجراحة

HIGH YEILD

Variance

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

Sample standard deviation is the positive square root of the variance.



- **Approximately 68% of the data fall between** $\bar{x} - 1s$ and $\bar{x} + 1s$
- **Approximately 95% of the data fall between** $\bar{x} - 2s$ and $\bar{x} + 2s$
- **Approximately 99.7% of the data fall between** $\bar{x} - 3s$ and $\bar{x} + 3s$

HIGH YEILD

Frequency Distributions

- A frequency distribution is a model that indicates how the entire population is distributed based on sample data.
- Since the entire population is rarely considered, sample data and frequency distributions are used to estimate the shape of the actual distribution.
- This estimate allows inferences to be made about the population from which the sample data were obtained.
- It is a representation of how data points are distributed.
- It shows whether the data are located in a central location, scattered randomly or located, uniformly over the whole range.

The frequency distribution may be represented in the form of an
- **equation**
- **a graph**

- When using a frequency distribution, the interest is rarely in the particular set of data being investigated.
- In virtually all cases, the data are samples from a larger set or population.
- Sometimes, it is wrongfully assumed that data follow the pattern of a known distribution such as the normal.
- The data should be tested to determine if this is true. Goodness of Fit tests are used to compare sample data with known distributions.
- The inferences made from a frequency distribution apply to the entire population.

Central limit theory

statisticians deal with distributions formed from
1. individual measurement
2. sets of averages.

If the data are taken from the same population, there is a relationship between the distribution of individual measurements and the distribution of averages.

standard error = the standard deviation for a set of averages

Pattern of distribution of data

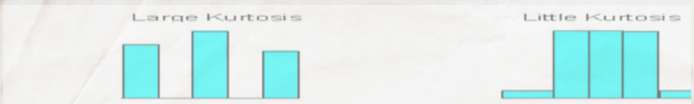
- **multimodal** "Some distributions have more than one point of concentration "
- it is likely that portions of the output were produced under **different conditions.**

unimodal "A distribution with a single point of concentration "

A distribution is **symmetrical** if the **mean**, **median** and **mode** are at the **same location**.

- **If a distribution is asymmetrical it is considered to be skewed.**
- **A symmetrical distribution has no skewness**

If a distribution has a relatively **high concentration** of data in the **middle** and **out on the tails**, but **little in between**, it has **large kurtosis**.
If it is relatively **flat** in the **middle** and has **thin tails**, it has **little kurtosis**.



THE NORMAL CURVE

The abscissa on the actual normal curve is denoted by x and the abscissa on the standard normal curve is denoted by Z.

transformation formula

$$Z = \frac{(X_i - \bar{X})}{s}$$