Statistics

Lecture 6 (follow up)

Descriptive statistics

By Dr.Chelli

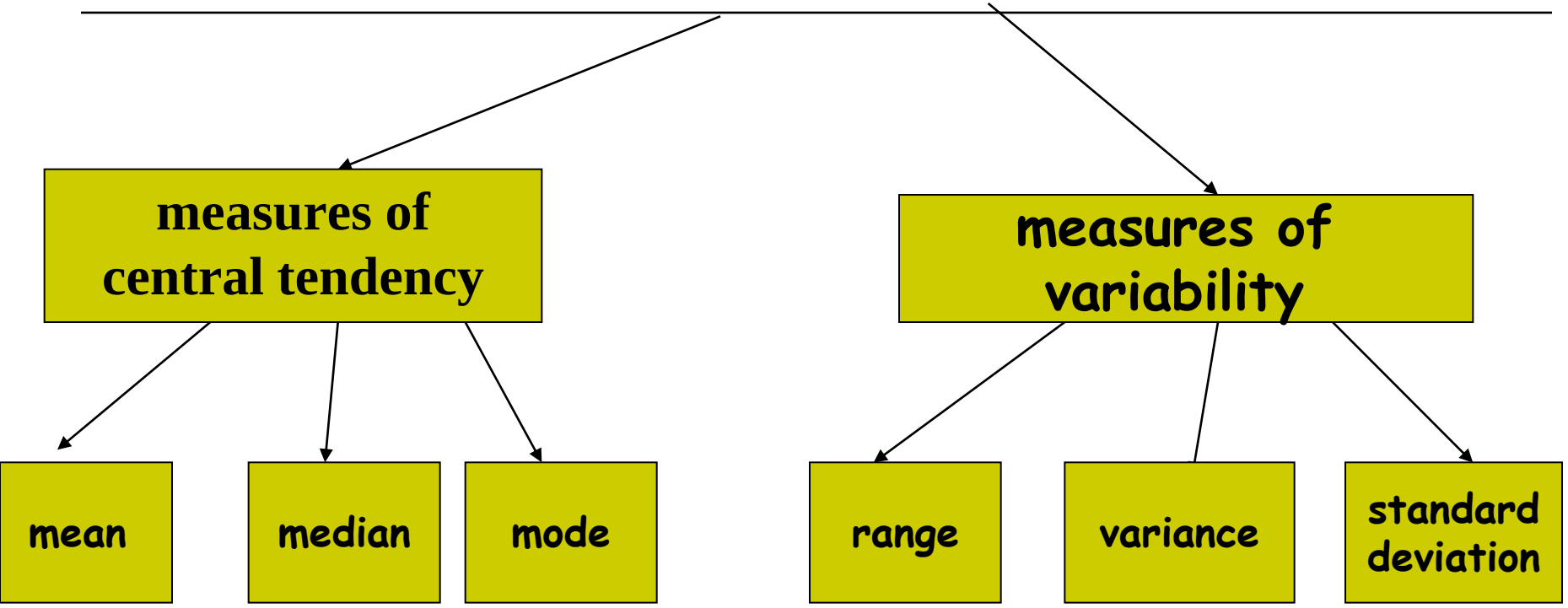
Lecture objectives

- **Training students in calculating and using measures of central tendency and measures of variability**

Introduction

To describe and summarize data, both measures of central tendency and measures of dispersion can be used. A measure of central tendency is a value that attempts to describe a set of data by identifying the central position within that set of data. Measures of variability provide indices of how dispersed or varied the scores are in a dataset. They include: This lecture introduces measure of central tendency classified as summary statistics and also measures of variability which measure how well an individual score represents the distribution.

Descriptive Statistics



Measures of central tendency

1. The mean, the most used measure can be used with both discrete and continous data though its use is most often with continous data.
- The mean generally makes sense for interval and ration scale measures.

It is calculated by summing all the values in the data set divided by the number of the values in that data set

The following formula is used:

$$\bar{x} = \frac{\sum X}{N}$$

continued

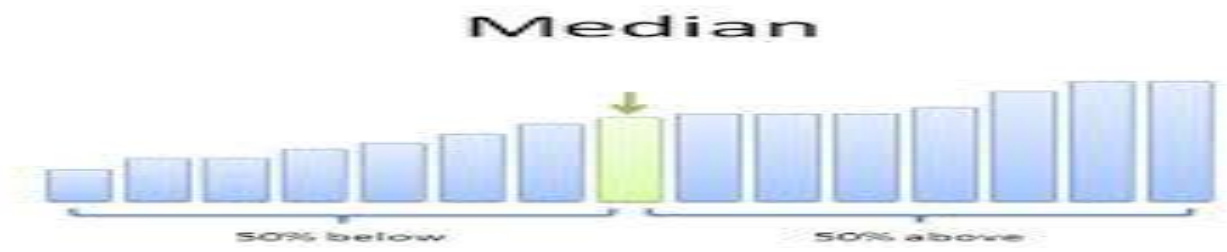
You notice that n refers to the sample an N to the population. Look at the following

Sample Mean	Population Mean
$\bar{x} = \frac{\Sigma x}{n}$	$\mu = \frac{\Sigma x}{N}$

where Σx is sum of all data values

N is number of data items in population

n is number of data items in sample



2. The median is the middle point for a set of data that has been arranged in order of magnitude as shown below.

95	27	18	65	14	46	9	73	24	52	20

9	14	18	20	24	**27**	46	52	65	73	95

How can you calculate the median in this example?

9	14	18	20	24	46	52	65	73	95

continued

We sum the two scores in the middle and divide by 2

24+ 46= 70/ 2= **35**

continued

3. The mode is the most frequent score in a data set as shown below (18 is the frequent score).

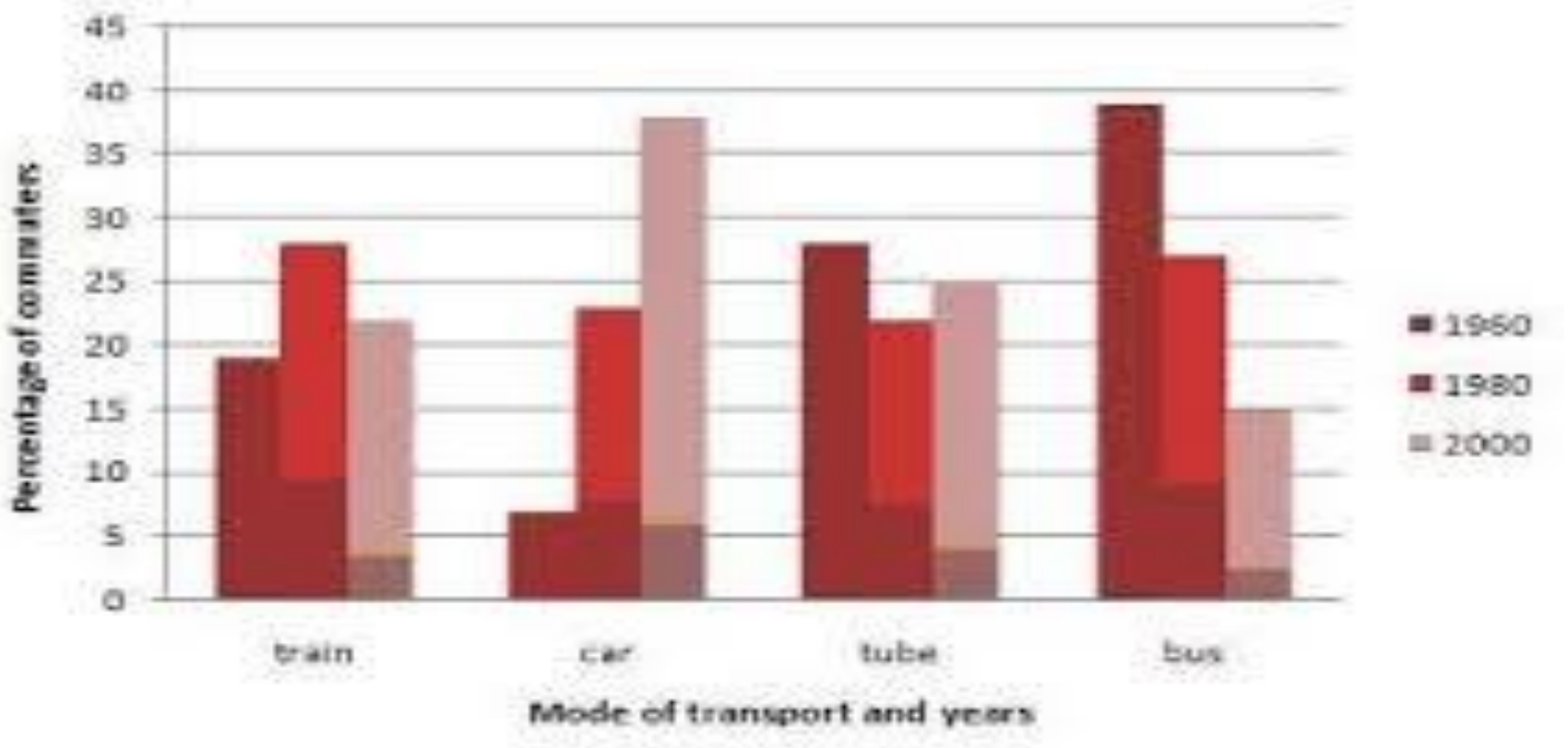
95	27	**18**	65	14	46	**18**	73	24	**18**	20

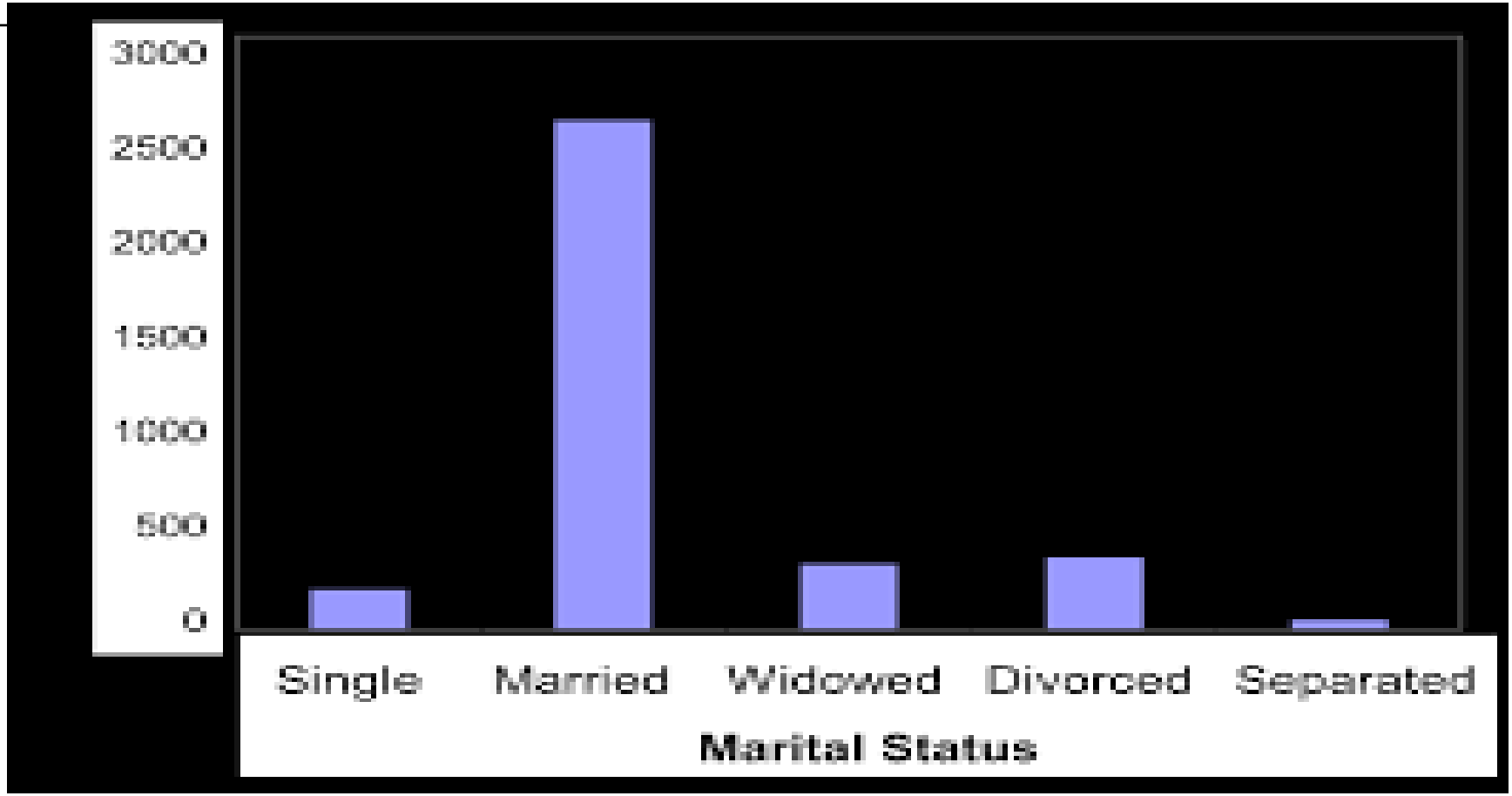
It is not used very often and is the only measure of centre appropriate for nominal data.

For example, if we are talking about the most frequently sold kind of TV, it does not make sense to talk about the mean or the median.

continued

The mode is generally used for categorical data where we want to know which is the most common category as shown in the graph.




3000
2500
2000
1500
1000
500
0
Single
Married
Widowed
Divorced
Separated
Marital Status

Measures of variability

The extent to which the mean and median are representatives of the values in the original data set depends upon the variability or dispersion in the original data set. This can be measured by the **range**, the **variance** and **the standard deviation.**

* Variability is a quantitative distance measure based on the differences between scores

The purposes of measures of variability are:

- To describe the distribution
- Measure how well an individual score represents the distribution

1. The range (continued)

1. **The range** is the distance covered by the scores in a distribution from the smallest to the highest value.

   It is the difference between the highest and the lowest score.

79 84 99 99 132 148 198

The range of the scores above is:

198 – 79= **119**

2. The variance (continued)

The variance is a necessary companion to standard deviation.

The variance measures the average squared distance from the mean.

Calculation of the variance

Score	Mean	Difference	Difference squared
49	63.5	- 14.5	210.25
99	63.5	35.5	1260.25
57	63.5	- 6.5	42.25
17	63.5	- 46.5	2162.25
63	63.5	- 0.5	0.25
100	63.5	36.5	1332.25 $\sum$= 5007.5 Variance= 5007.5: 6= 834.58

Calculation of the variance (continued)

1. Find the deviation fro each score from the mean.
2. Square the deviations.
3. Find the sum of squared deviations
4. Find the average of squared devations (divide the sum by the number of observations.

Formula

$$s^2 = \frac{\sum \left(x - \bar{x}\right)^2}{n-1} \qquad \sigma^2 = \frac{\sum \left(x - \mu\right)^2}{N}$$

Another formula

$$s^2 = \frac{\sum x^2}{N} - \bar{x}^2$$

3. The standard deviation

Most common and most important measure of variability is the standard deviation.

A measure of standard deviation, or average, distance from the mean describes whether the scores are clustered closely to the mean or are widely scattered.

continued

The calculation of the variance is necessary for that of the standard deviation because

The standard deviation is the square root of the variance.

$$S = \sqrt{\frac{\Sigma (X - \overline{X})^2}{N}}$$

where S = the standard deviation of a sample,
Σ means "sum of,"
X = each value in the data set,
$\overline{X}$ = mean of all values in the data set,
N = number of values in the data set.

continued

The Equation

Undoing the effect of earlier squaring

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

continued

$$\text{Standard Deviation (population)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \text{mean})^2}{n}}$$

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \text{mean})^2}{n-1}}$$

Exercise: Calculate the mean, variance and standard deviations for the following tests.

Participants scores in an experiment about the effects of brainstorming on developing learners writing are :

- Pre-test scores: 1, 5, 16, 2, 13, 6, 5, 10, 1, 6, 8, 7, 2, 4, 15, 9, 3, 8, 10, 7
- Post-test scores: 6, 9, 17, 12, 15, 10, 13, 11, 9, 12, 16, 14, 14, 10, 11, 10, 9, 12, 14, 11

You can watch these you tubes

- https://www.youtube.com/watch?v=E4HAYd0QnRc
- https://www.youtube.com/watch?v=ztfaPxCqGiE