ATTACHMENT 2

REPUBLIC OF THE PHILIPPINES

PHILIPPINE STATISTICS AUTHORITY

Results from the 2010 Census of Population and Housing

2010 CPH

Kabilang Ako!

How well do we know our people?

2010 Census of Population and Housing

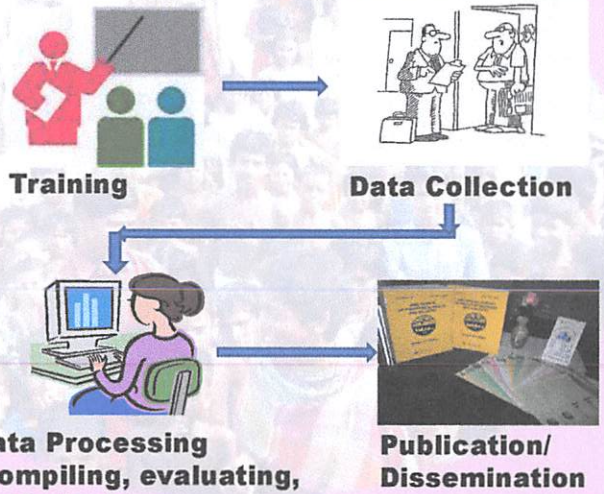
The Census tells the story of our people.

Census of Population and Housing

- inventory of the total population and housing units in the Philippines covering all barangays in the country.
- 2010 Census: 13th census of population,

  6th census of housing

Phases in the Census



Reference Date and Time of 2010 CPH

May 1, 2010 as of 12:01 a.m.

Legal Authority

Basic Law:

- ✓ Commonwealth Act No. 591, Section 2

Other directives:

- ✓ Executive Order No. 352
- ✓ P.D. No 418
- ✓ E.O. No. 121
- ✓ Batas Pambansa Blg 72
- ✓ E.O. No. 5

Uses of Census Data

In government:

- Redistricting and apportionment of congressional seats
- Allocation of resources and revenues
- Creation of political and administrative units
- Formulation of policies concerning population and housing
- Formulation of programs relative to delivery of basic services for health, education, housing, social welfare, and others.

Uses of Census Data (conc)

In business and industry:

- Determination of sites for establishing businesses
- Determination of consumer demands for various goods and services
- Determination of supply of labor for the production of goods and services.

In research and academic institutions

- Conduct of researches on population and other disciplines
- Study of population growth and distribution as basis in preparing projections.

Population census items

P1 – Name of household members
P2 – Relationship to head
P3 – Sex
P4 – Date of Birth
P5 – Age
P6 – Birth Registration
P7 – Marital Status
P8 – Religious Affiliation
P9 and 10 – Citizenship
P11 – Ethnicity
P12 – Disability
P13 – Functional Difficulty
P14 – Residence 5 years ago
P16 – Highest Grade/Year Completed
P19 – Overseas worker

Housing Census items

B1 – Type of Building
1 – Single House
2 – Duplex
3 – Multi-unit residential (three units or more)
4 – Commercial/Industrial/Agricultural (office, factory and others)
5 – Institutional living quarters (Hotels, Hospital and others)
6 – Other housing units (Boat, Cave and others)

B2 – Construction materials of the roof by type

Housing Census items (conc.)

B3 – Construction materials of the walls

B4 – State of repair of the building/house

B5 – Year building/house was built

H8 – Tenure status of the lot
1 – Owned/being amortized
2 – Rented
3 – Rent-free with consent of owner
4 – Rent-free without consent of owner
5 – Not Applicable

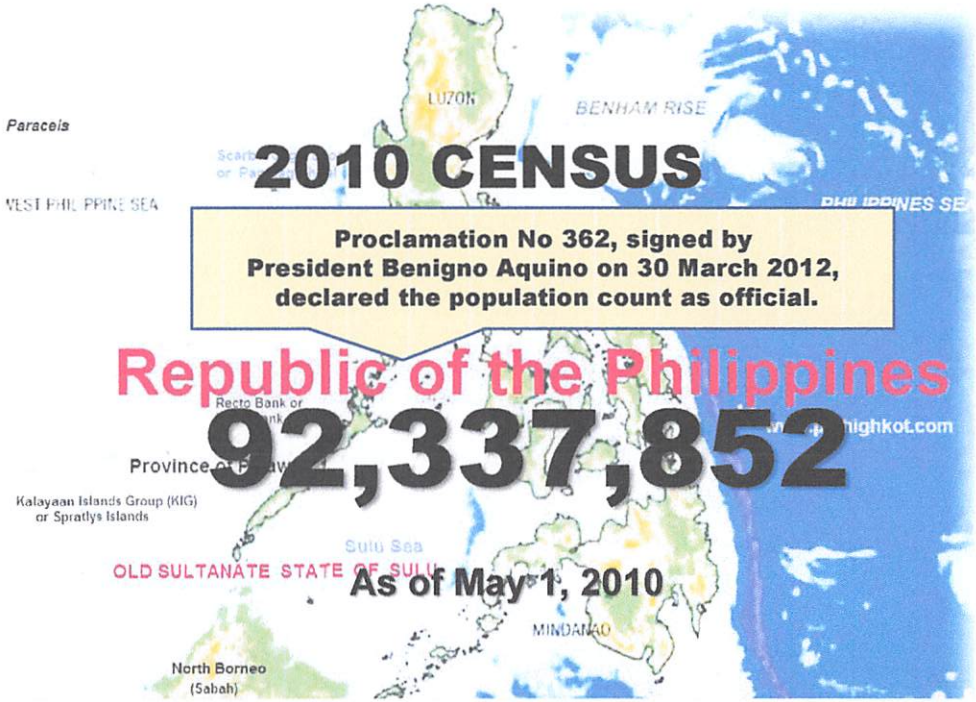
D1 – Floor area of the housing unit

Definition

Building – a building is defined as any structure built, designed or intended for enclosure, shelter or protection of any person, animal or property.

Housing unit - structurally separate and independent place of abode which, by the way it has been constructed, converted or arranged, is intended for habitation by one or more households

Household - a social unit consisting of a person living alone or a group of persons who sleep in the same housing unit and have a common arrangement in the preparation and consumption of food.



2010 CENSUS

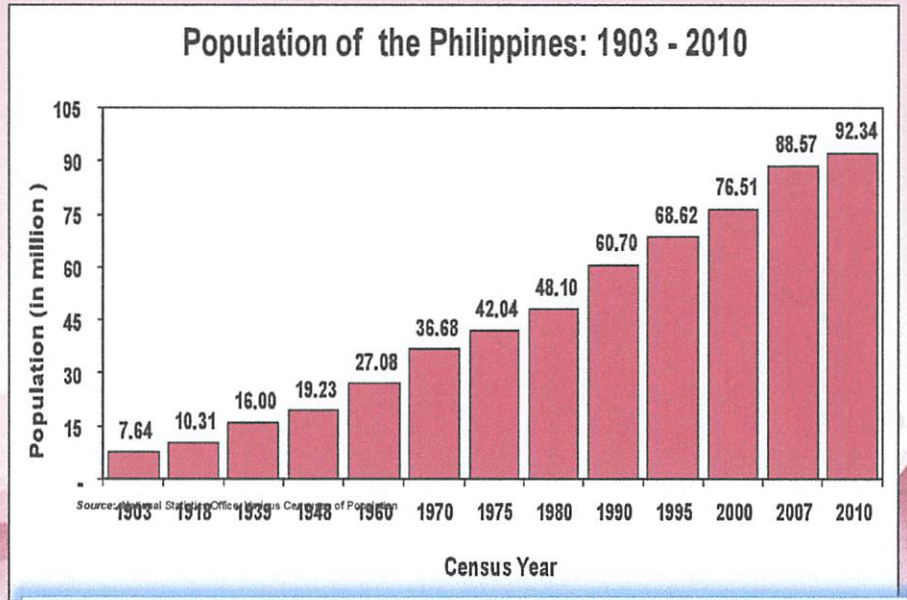
Proclamation No 362, signed by President Benigno Aquino on 30 March 2012, declared the population count as official.

Republic of the Philippines

92,337,852

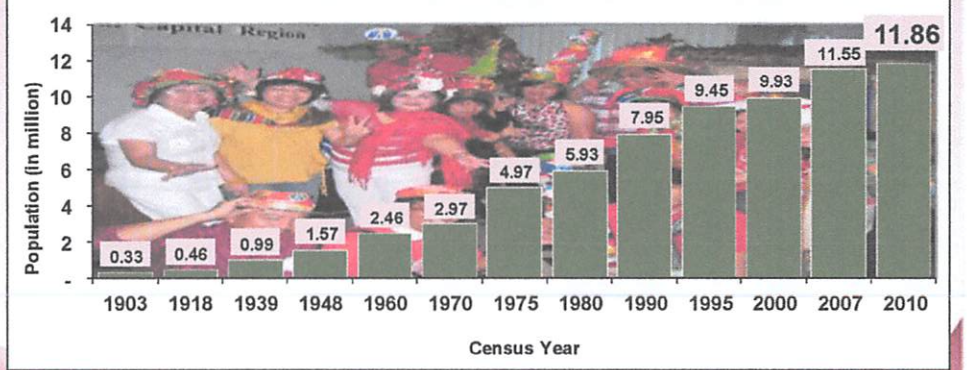
As of May 1, 2010

Population of the Philippines: 1903 - 2010



The Philippine population in 2010 is 12 times higher than the population in 1903; more than three times higher the population 50 years ago.

Total Population in NCR*: 1903 - 2010



NCR population in 2010 is nearly 5 times higher than the population 50 years ago.

* Population figures are based on current composition of NCR (16 cities and one municipality).

Provinces with more than Two Million Population, 2010

Province	Total Population
Cavite	**3.09 million**
Bulacan	**2.92 million**
Pangasinan	**2.78 million**
Laguna	**2.67 million**
Cebu (exc cities of Cebu, Lapu-Lapu & Mandaue)	**2.62 million**
Rizal	**2.48 million**
Negros Occidental (exc Bacolod City)	**2.40 million**
Batangas	**2.38 million**
Pampanga (exc Angeles City)	**2.01 million**

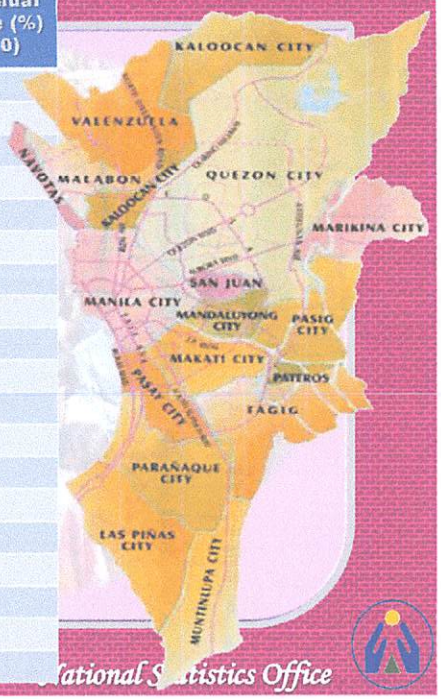
HUCs with more than One Million Population, 2010

City	2010 Total Population
Quezon City	**2.76 million**
City of Manila	1.65 million
Caloocan City	1.49 million
Davao City	1.45 million

Most Populous Barangays, 2010

Barangays	2010 Total Population
Barangay 176 (Caloocan City)	**243,890**
Commonwealth (Quezon City)	**186,543**
Batasan Hills (Quezon City)	**150,764**

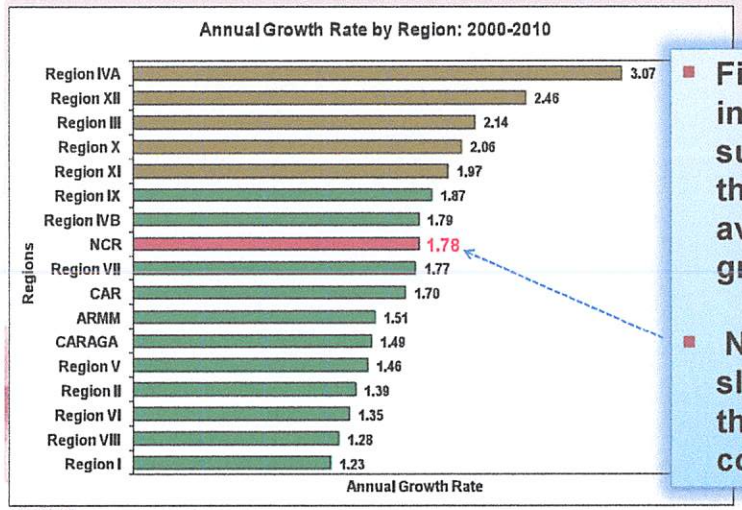
City	Population May 1, 2010	Average Annual Growth Rate (%) (2000-2010)
Philippines	92,337,852	1.90
NCR	11,855,975	1.78
Las Piñas	552,573	1.57
Makati	529,039	1.16
Malabon	353,337	0.42
Mandaluyong	328,699	1.67
Manila	1,652,171	0.44
Marikina	424,150	0.81
Muntinlupa	459,941	1.95
Navotas	249,131	0.78
Parañaque	588,126	2.72
Pasig	669,773	2.86
San Juan	121,430	0.31
Valenzuela	575,356	1.71
Caloocan	1,489,040	2.37
Pasay	392,869	1.02
Pateros	64,147	1.12
Quezon City	2,761,720	2.42
Taguig	644,473	3.26



National Statistics Office

Average Annual Growth Rate

Philippines: 1.90 % (2000-2010)



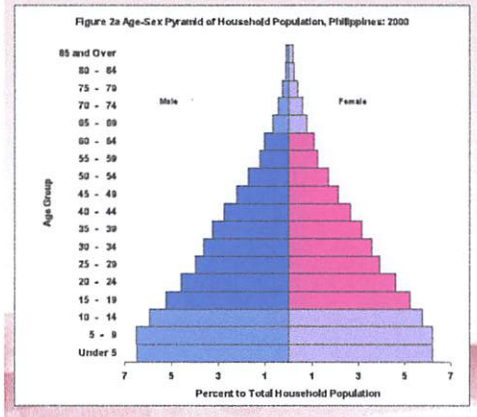
- Five regions in the country surpassed the national average growth rate
- NCR grew slower than the entire country.

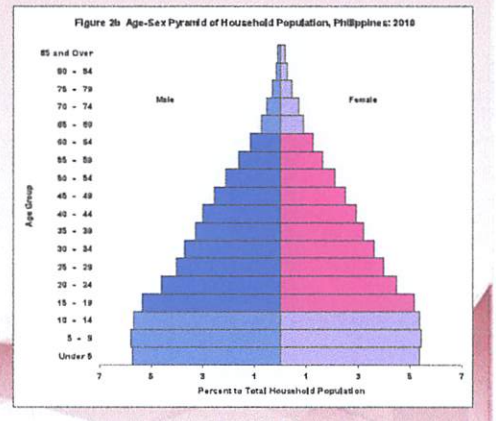
Average Annual Growth Rate

	Philippines	NCR
1990 - 2000	2.34%	2.25%
2000 - 2010	1.90%	1.78%

Age-Sex Structure of Household Population

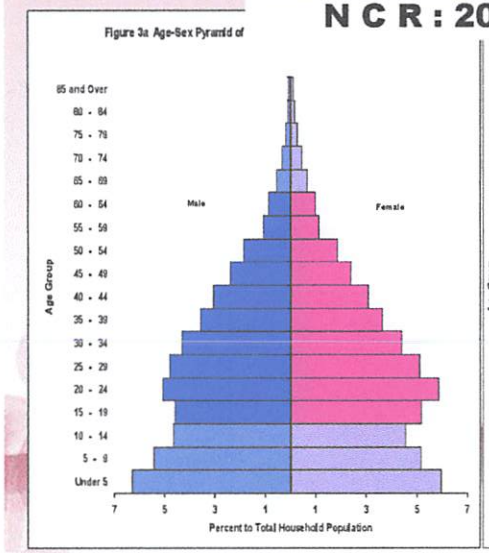
Philippines: 2000 & 2010

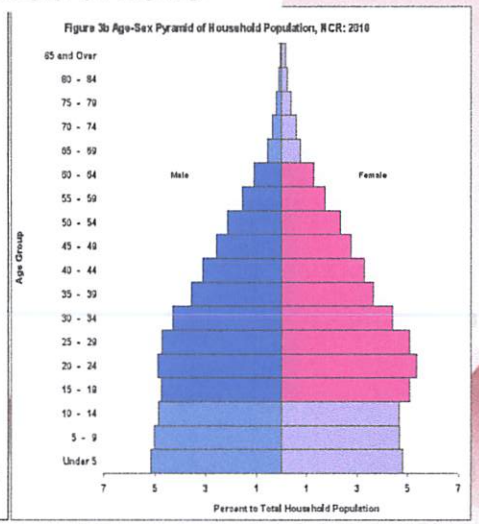




Age-Sex Structure of Household Population

N C R : 2000 & 2010



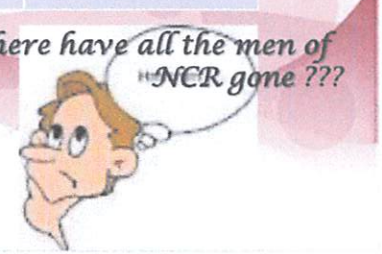


Source: 2000 and 2010 Census of Population and Housing, NSO

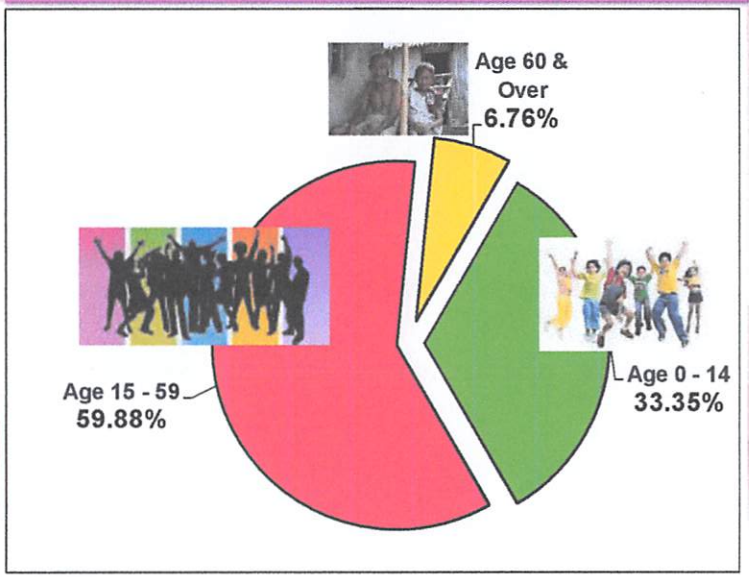
Sex Ratio

(Number of Males to 100 Females)

	Philippines	N C R
2010	102:100	96:100
2000	101:100	96:100
1990	101:100	94:100

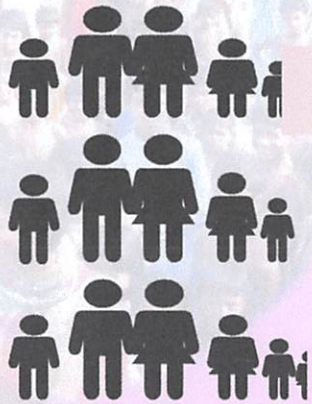


Age Composition, Philippines: 2010



Average Household Size, Philippines

Year	Number of Persons
2010	4.6
2000	5.0
1990	5.3

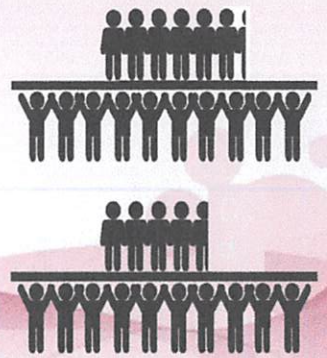


Median Age, Philippines 2010

	Median Age
Philippines	23.4
NCR	25.5

Dependency Ratio 2010

	Dependency Ratio
Philippines	61
NCR	48



Dependency ratio refers to the number of young (0-14 years) and old (65 & over) dependents for every 100 persons in the

The Five Most Populous Countries

The World: 6.97 billion *(est 2010)*

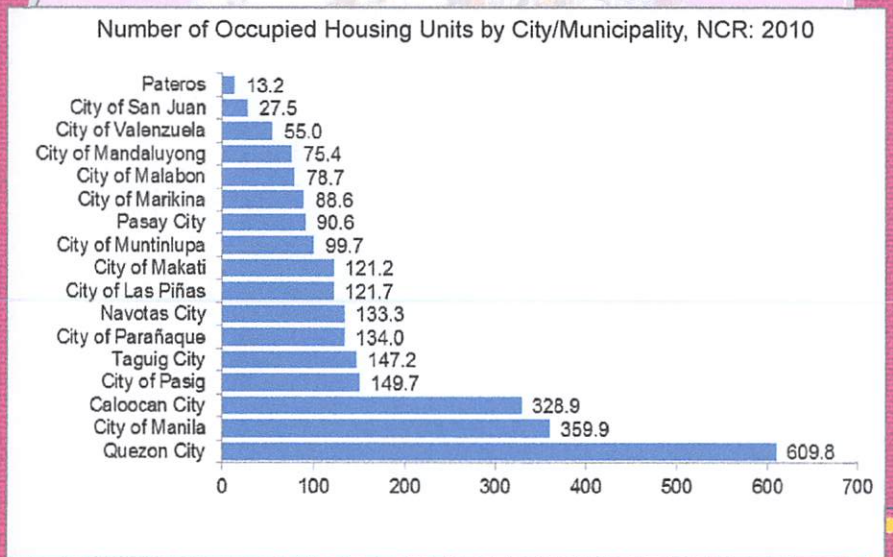
Country	Population	Reference Date
1. People's Republic of China	1.34 billion	Nov 1, 2010
2. India	1.21 billion	March 1, 2010
3. U S A	308.74 million	2010 Census
4. Indonesia	237.56 million	May 2010
5. Brazil	190.73 million	Aug 1, 2010
6. Pakistan	185 million	2010 Census
7. Nigeria	164 million	2010 Census
8. Bangladesh	158 million	2010 Census
9. Russia	142 million	2010 Census
10. Japan	127 million	2010 Census
11. Mexico	118 million	2010 Census
12. Philippines	92.3 million	2010 Census

Housing Characteristics

Total occupied housing units in the Philippines reached 19.7million

NCR occupied housing units reached 2.6 million

Housing Characteristics



Of the total occupied housing units in the country, ...

57.5%	**Single-type houses**
60.0%	With concrete/brick/stone for outer walls
86.0%	With galvanized iron/aluminum for roofs
84.5%	Needs minor repair or none at all
20.2%	Built from 2001 - 2010
20.3%	With floor area of 30-49 sq m
Ratio of households to occupied housing units	1.02

REPUBLIC OF THE PHILIPPINES

PHILIPPINE STATISTICS AUTHORITY

2010 Census of Population and Housing

2010 CPH

Kabilang Ako!

THANK YOU

Types of Data

Primary data

Sources

National Statistics Office (NSO)
Bureau of Agricultural Statistics (BAS)
Department of Agriculture (DA)
Bureau of Labor and Employment Statistics (BLES)
Department of Labor and Employment (DOLE)
Bangko Sentral ng Pilipinas (BSP)

Sources (from Administrative Records)

Bureau of Internal Revenue (BIR)
Department of Education (DepEd)

Types of Data

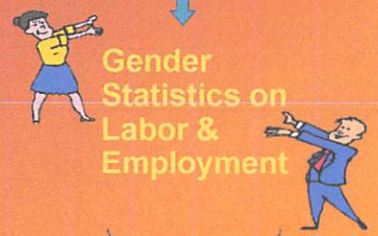
SECONDARY DATA

- **Data are provided directly by an organization or government agency in convenient form such as written report or census data in tapes or floppy diskettes.**
- **Data that are processed and re-processed by individuals or entities from sources other than the primary source of information.**

Types of Data

NSO → **Labor Force Survey** → **PRIMARY DATA**

BLES → **Gender Statistics on Labor & Employment** → **SECONDARY DATA**



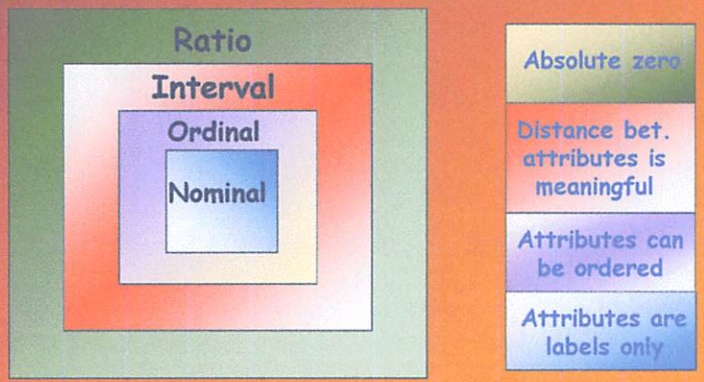
Calculating Machines and Use of Excel Software

- **Use of scientific calculator to compute parameters or statistics.**

- **MS Excel**

Levels of Measurement

COMPARATIVE SUMMARY



Levels of Measurement

IMPORTANCE OF UNDERSTANDING THE LEVELS OF MEASUREMENT

1. Helps you decide how to interpret the data.
2. Helps you decide what statistical analysis is appropriate on the values that were assigned.

Types of Data

2 Types

- PRIMARY DATA
- SECONDARY DATA

Types of Data

PRIMARY DATA

- Any set of data or information that are directly collected from the source (informants or respondents or records).
- Government statistical agencies are given the responsibility to collect, publish and disseminate statistical series.

Interval scale

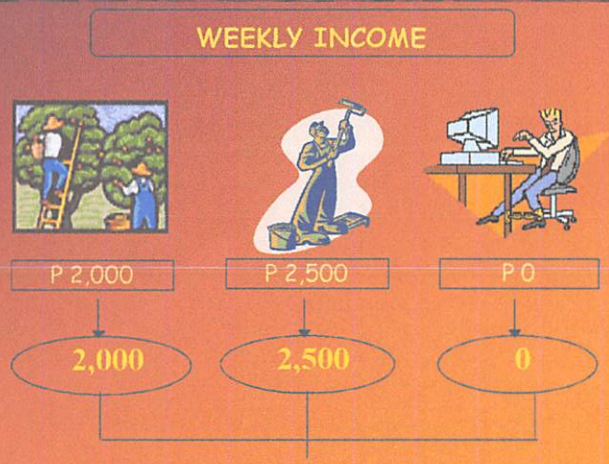
- zero point has no meaning

Example:

Celsius	-18	0	10	30	100
Fahrenheit	0	32	50	86	212

Ratio Measurement

- Has all the features of an interval scale.
- Requires on absolute, fixed and non-arbitrary zero point.
- Ratio of two numbers is meaningful



Ratio Measurement

HEIGHT TIME VOLUME

Years of school completed
Number of children born
Years in present job

Per capita GNP
Weeks of unemployment
Travel time to work (minutes)

Levels of Measurement

Nominal

↓

qualitative

↓

Exhaustive/Mutually exclusive

↓

equal in value

↓

TeamA=TeamB=TeamC

Ordinal

↓

qualitative

↓

Exhaustive/Mutually exclusive

↓

ranked

↓

1st place >2nd place >3RD

Levels of Measurement

Social Class

Upper

Middle

Lower

Academic Grades

A, B, C, D, E, F

Intensity of attitude

↓

Strongly agree, agree, neutral, disagree, strongly disagree

Quality of Service

↓

(5) Excellent
(4) Very Satisfactory
(3) Satisfactory
(2) Needs Improvement
(1) Poor

Levels of Measurement

Interval scale

- Assigning of numbers to observations is based not only on the order to which they possess a certain attribute but also indicates exactly how much they posses the attributes.
- In this measurement we can determine how many units' difference there are from one rank to the next.

Levels of Measurement

GRADES IN STATISTICS

80 85 90

80 85 90

Rule: INTERVAL

Level: INTERVAL

Levels of Measurement

The psychiatric system of a diagnostic groups

- Schizophrenic
- Paranoid
- Manic-depressive
- Psychoneurotic

Automobile license plates

NJB020401, NUU112900

jersey numbers

Employment classification

1 - Educator
2 – Construction worker
3 – Manufacturing worker
4 – Lawyer
5 – Doctor
6 - Others

Levels of Measurement

Nominal scale

Conditions:

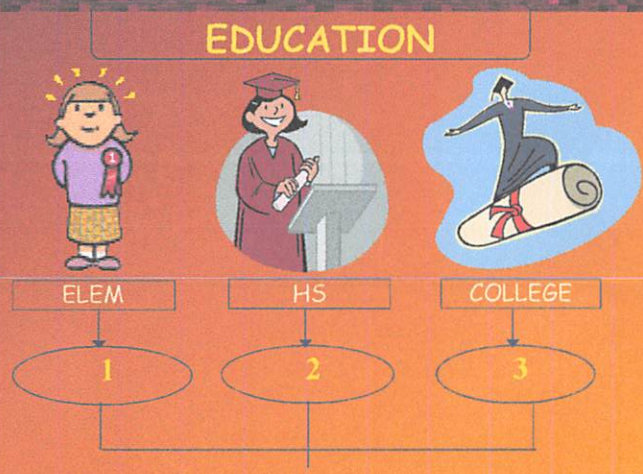
1. Exhaustive – every value or unit of data can be assigned to a category.
2. Mutually exclusive – it is not possible to assign a value to more than one category because the categories do not overlap.

Levels of Measurement

Ordinal scale

- It involves placement of values or codes in some rank order to create an ordinal scale variable.
- The relationship between observations takes on the form of "greater than" and "less than" or "higher than" and "lower than"

Levels of Measurement



Rule: MAGNITUDE
Level: ORDINAL

Levels of Measurement

MEASUREMENT

is a set of rules for assigning numbers to attributes of observations.

It is structured in such way that the existing relationship between the observations is preserved in the numbers assigned to them.

About Measurement

Levels of Measurement

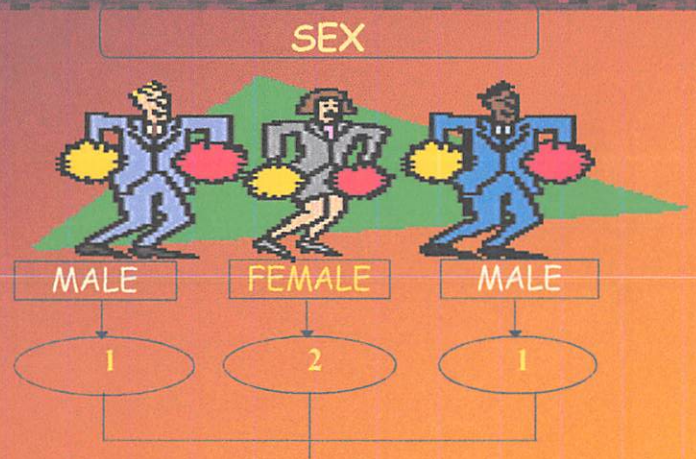
- Nominal
- Ordinal
- Interval
- Ratio

Levels of Measurement

Nominal scale

- Is the simplest scale of measurement where a value or unit of data is assigned to one of at least two qualitative classes or categories.

Levels of Measurement



RULE: Identification

LEVEL: Nominal

Kinds of Variables

Qualitative Variable

A qualitative variable takes on non-numerical values.

It simply describes which class or category the observations fall, thus also known as
categorical data

Kinds of Variables

Qualitative Variables

Hair color: Black, Blonde, Brown

Religion: Catholic, Protestants, INC

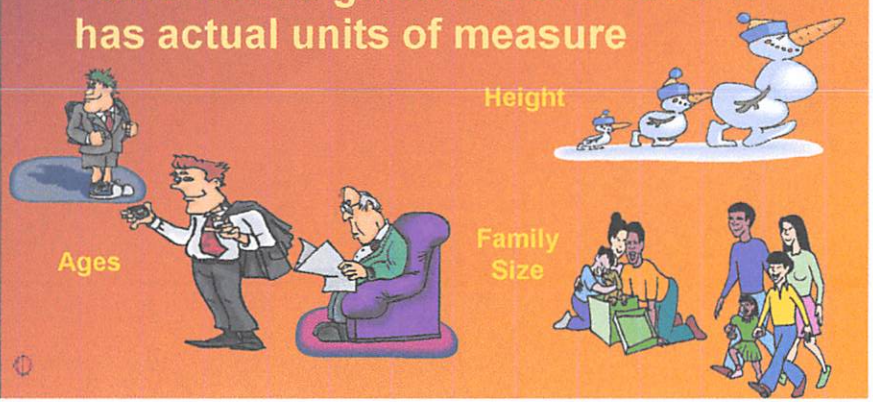
Occupation: Teacher, Doctor, Engineer

Nationality: Filipino, American, Hispanic

Kinds of Variables

Quantitative Variable

A quantitative variable may take any value from a given set of values. It has actual units of measure



Kinds of Variables

Quantitative Variables

- **Discrete**

  Number of overweight persons

  0, 1, 2, 3 ….

- **Continous**

  Weight in kilograms

  65.6 kg, 55.34 kg, 100 kg, ¾ kg . . .

Scope of Statistics

Methodology

(2) INFERENTIAL STATISTICS

Method or technique using small portion of the total set of data in order to draw conclusions or judgments regarding the entire set.

Scope of Statistics

Statistical Inference

- Predict life span of bulbs
- Compare effectiveness of two reducing diets

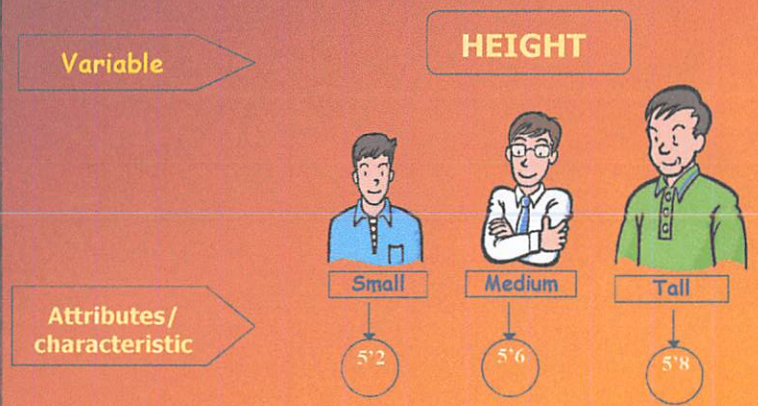
✓ Probability theory

✓ Risks/odds methods

VARIABLES

A variable is a characteristic of a unit of observation or subject that can take on different values for different units/subjects or for the same unit/subject at different periods.

About Variable



Definition of Statistics

STATISTICS is the art and science of

- Collection
- Organization
- Presentation
- Analysis and Interpretation

of numeric or quantitative data.

✗ Houses are built of hollow blocks, wood, etc

✓ How many houses were built of hollow blocks

✓ Proportion of houses built of wood.

Scope of Statistics

ORIGIN

- Government and public administration
  - for taxation purposes
  - for raising an army

  ✓ Statistical methods
- Games of chance
  - for betting on card games
  - for selecting the number which is most likely to appear

  ✓ Probability theory

Scope of Statistics

Methodology

(1) DESCRIPTIVE STATISTICS

is concerned with the gathering, classification, presentation of data and the collection of summarizing values to describe group characteristics of the data.

Scope of Statistics

DESCRIPTIVE STATISTICS

- Gathering
  - Survey
  - Observation
  - Use of Existing records
  - Experimental
- Classification (Data Organization)
  - Raw Data
  - Array
  - Frequency Distribution Table
  - Single Value Grouping
- Presentation
  - Textual
  - Tabular
  - Graphical
- Collection of Summarizing values (computing measures)
  - Central Tendency
  - Variability
  - Percentages/Ratio/Proportions
  - Others (Quantiles / Fractiles)

INTRODUCTION TO STATISTICS

INTRODUCTION

- Definition of Statistics
- Scope of Statistics
- Kinds of Variables
- Levels of Measurement
- Types of Data
- Calculating Machines and Use of MS Excel Software

Definition of Statistics

STATISTICS (singular)

- refers to the statistical principle and method that have been developed for handling numerical data.

- Mean, Median, Mode
- Z-test, T-test
- Tuckeys w, Kolmogorovs-Smirnovs test
- Statistical Power

Definition of Statistics

STATISTICS (plural)

- refers to the body of numerical fact of any kind.
- a set of quantitative data.

Table1: Legislated Wage Rate in Non-Agricultural Sector by Region: 1999 (in Pesos)

Region	Lowest	Highest
NCR	198.00	198.00
CAR	144.00	155.00
I	155.00	168.00
II	152.00	160.00
III	164.00	176.50
IV	159.00	200.00
V	123.00	163.00
VI	117.40	146.85
VII	111.00	165.00
VIII	149.00	153.00
IX	136.00	142.00
X	142.00	149.00
XI	136.00	145.00
XII	130.00	141.00
ARMM	129.20	140.00
CARAGA	128.00	140.00

Source of Data: National Wages and Productivity Commission

Statistics on

- Prices, GDP, GDI, GNP
- Employed, Accessions, Separations
- Violations, Compliance, Strikes


Thank
You!

TRAINING COURSE ON STATISTICAL CONCEPTS BEHIND RENTAL CONTROL STUDY AND ESTIMATION OF HOUSING NEEDS FOR SELECTED HUDCC STAFF

10-12 November 2014

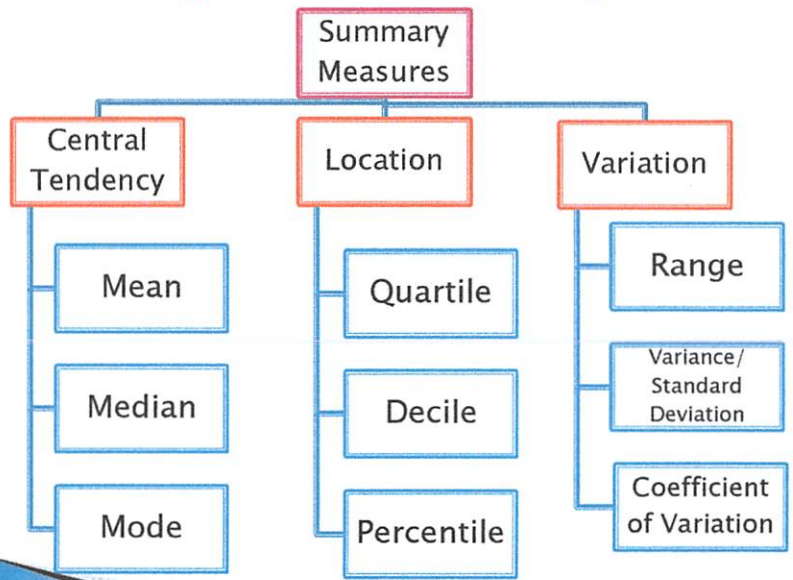
Measures of Central Tendency

REPUBLIC OF THE PHILIPPINES
PHILIPPINE STATISTICAL RESEARCH AND TRAINING INSTITUTE
J & S Building, 104 Kalayaan Avenue, Diliman, Quezon City 1101, Philippines
Telefax : (632) 433-1745/929-7543/436-1426/426-0620/920-9649
www.srtc.gov.ph

Objectives

- Learn when to use the different measures of central tendency
- Determine the different properties of the measures of central tendency

Describing Data with Summary Measures



Measures of Central Tendency

- Measure of central tendency is an index of the central location of a distribution. It is a single value that is used to identify the "center" of the data or the typical value.
- Precise yet simple
- Most representative value of the data

The Arithmetic Mean:

- the most frequently used measure of central tendency
- the sum of the observations divided by the total number of observations

Notations:

μ - used to denote population mean (parameter)

$\bar{x}$ - used to denote sample mean (statistic)

It can be computed in two ways:

- for ungrouped data
- for grouped data

THE MEAN

Example: Suppose the following represent the wages of ten employees taken as samples. Find the mean.

Employee	Wages
1	6,550
2	7,333
3	14,075
4	9,700
5	4,500
6	5,705
7	4,900
8	10,258
9	3,809
10	5,500

Apply the formula,

$\sum_{i=1}^{n} x_i$ = 6,550 + 7,333 +... + 3,809 + 5,500

= 72,330

n = 10

Thus,

$\bar{x}$ = 72,330 / 10 = Php7,233

THE MEAN

Mean of Grouped data

- associated with the frequency distribution
- also known as the weighted mean.

Population Mean : $$\bar{X}_w = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i} = \frac{W_1X_1 + W_2X_2 + \ldots + W_nX_n}{W_1 + W_2 + \ldots + W_n}$$

Sample Mean: $$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

where: W_i and w_i = the class frequencies

K and k = the number of classes.

The Weighted Mean

* if the individual values do not have equal importance, then we compute for the weighted mean.

* We assign weights to the observed values of the data set before we can get the weighted mean.

Example of Weighted Mean

Suppose a government agency gives scholarship grants to employees taking graduate studies. Courses in graduate studies earn credits of 1, 2, 3, 4, or 5 units. They can get a partial scholarship for the next semester if they get a weighted average of 1.5 to 1.75 and a full scholarship if the average is better than 1.5, which means an average of 1.0 to 1.49. What kind of scholarship will the 2 employees get given their grades for the previous semester?

Consider the grades of the two employees in the previous semester:

	Employee A			Employee B	
Subjects	Units	Grade	Subjects	Units	Grade
A	1	1.0	A	1	2.0
B	2	1.25	B	2	1.75
C	3	1.5	C	3	1.5
D	4	1.75	D	4	1.25
E	5	2.0	E	5	1.0

Solution:

We let the units be the weights Wi and the grade is the Xi.

Weighted average of employee A:

$$\overline{X}_w = \frac{1(1)+2(1.25)+3(1.5)+4(1.75)+5(2)}{1+2+3+4+5} = \frac{25}{15} = 1.67$$

Weighted average of employee B:

$$\overline{X}_w = \frac{1(2)+2(1.75)+3(1.5)+4(1.25)+5(1.0)}{1+2+3+4+5} = \frac{20}{15} = 1.33$$

Thus, employee A will get a partial scholarship while employee B will get a full scholarship.

The Combined Population Mean

We can obtain the mean of several data sets given the means and number of observations of each data set. This is what we call the combined mean. Suppose that k finite populations having measurements, $N_1, N_2, \ldots, N_k$ respectively, have means $\mu_1, \mu_2, \ldots, \mu_k$

The ***combined population mean,*** μ_c of all the populations is

$$\mu_c = \frac{\sum_{i=1}^{k} N_i \mu_i}{\sum_{i=1}^{k} N_i} = \frac{N_1\mu_1 + N_2\mu_2 + \ldots + N_k\mu_k}{N_1 + N_2 + \ldots + N_k}$$

If random samples of size $n_1, n_2, \ldots, n_k$, selected from these k populations, have the means respectively, the **combined sample mean** $\overline{X}_c$ of all the sample data is

$$\overline{X}_c = \frac{\sum_{i=1}^{k} n_i \overline{X}_i}{\sum_{i=1}^{k} n_i} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + \ldots + n_k\overline{X}_k}{n_1 + n_2 + \ldots + n_k}$$

Example of the Combined Mean

The Philippines have 6,028 male children deaths and 4,948 female children deaths for the age group 1-4 in 2002. The average number of deaths for male and female children is 376.8 and 309.2. What is the combined population mean for both sexes?

Solution: We let $N_1 = 6028$ and $N_2 = 4{,}948$.

$\mu_{males} = 376.8$ $\mu_{females} = 309.2$

Thus, $\mu_{both\ sexes} = \frac{6028(376.8) + 4948(309.2)}{6028 + 4948} = 346$

The average number of deaths for children 1-4 years old for both sexes is 346.

THE MEAN

Advantages of the MEAN:

- Takes into account all observations.
- Can be used for further statistical calculations and mathematical manipulation.
- The value of the mean always exists and unique.
- It is a widely understood measure of central tendency.

THE MEAN

Disadvantages of the MEAN:

- It may or may not be an actual observed value in the data set.
- Mean is easily affected by extreme values, especially if the number of observations is small.
- Mean cannot be computed if there are missing values due to omission or non-response.
- In grouped data with open-ended class intervals, the mean cannot be computed. It is dependent on all observed values.

THE MEAN

When to use the MEAN:

- When data is of interval and ratio scale
- The value of each score is desired
- Further statistical computation is needed
- When the distribution of the data is normal

THE MEDIAN

- the central value of a distribution
- the value that divides the distribution into two equal parts.

Median for Ungrouped Data:

* The first step in finding the median, denoted by Md, is to arrange the observations in an array.

If N is even: average of the two middle observed values

$$M_d = \frac{X_{(N/2)} + X_{(N+2)+1}}{2}$$

If N is odd: middle observed value

$$M_d = X_{(N+1)/2}$$

Examples of the Median

The annual per capita poverty threshold in pesos of the different regions of the Philippines are as follows:

15,693, 13,066, 12,685, 11,128 13,760, 13,657,
11,995, 11,372, 11,313, 9,656, 9,518, 9,116,
10,503, 10,264, 10,466, 10,896, 12,192.

Find the median.

Solution: We arrange the 17 annual per capita poverty threshold in pesos of the 17 regions of the Philippines from lowest to highest.

Array: 9116, 9518, 9656, 10264, 10466,
10503, 10896, 11128, 11313, 11372,
11995, 12192, 12,685,13066, 13657,
13760, 15693

THE MEDIAN

Advantages of the MEDIAN:

- Not affected by extreme values.
- If you want the exact middle value of the distribution.
- It can be computed even for grouped data with open-ended class intervals.

THE MEDIAN

Disadvantages of the MEDIAN:

- The median cannot be combined with other distributions with similar variates to obtain an overall median
- The median value does not have direct relation to the total number of observations and their total value. It merely indicates the value that divides the population into two parts.

THE MEDIAN

When to use:

- When data is of ordinal scale
- When middle value is desired
- When measure of central tendency that is not affected by extreme values is needed
- When data distribution is skewed
- If the distribution has open-ended intervals

THE MODE

- Mode is that value of a variable that occurs most frequently in a distribution.
- It is also referred to as the nominal average.

THE MODE

Determine the mode by counting the frequency of each observed value and finding the observed value with the highest frequency of occurrence.

Examples of Mode

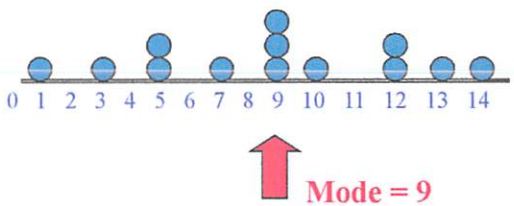
1. Given the data on number of children of 12 currently married women: 2, 2, 1, 1, 1, 3, 3, 4, 4, 2, 2, 2. Find the mode.

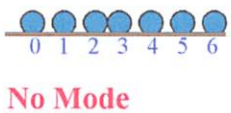
2. Given the data on number of cases resolved by a 10 lawyers: 5, 4, 1, 1, 3, 3, 2, 1, 3, 0. Find the mode.

3. Given the data on number of cases handled by 14 PAO lawyers : 629, 645, 356, 656, 231, 455, 412, 289, 444, 452, 642, 225, 335, 411. Find the mode.

Mode

- occurs most frequently
- may or may not exist





THE MODE

Advantage of the MODE:

- It can be easily identified through ocular inspection.
- Extreme values do not easily affect the mode.
- Its value is always one of the observed values in the data set.
- It can be obtained both for quantitative and qualitative types of data.

THE MODE

Disadvantages of the MODE:

- The mode is sometimes not unique and does not exist.
- We can have several modes for one data set. If there is one mode, it is unimodal. If there are two modes, we call it bimodal. If there are more than two modes, then we call it multimodal.
- It does not possess the desired algebraic property of the mean that allows further manipulation.
- To obtain a new mode of different distributions, all the raw data of the different distributions have to be merged to obtain a new mode.

Example of Mode for Qualitative Data

Number of Families by Tenure Status of House and Lot

Tenure of House and/or Lot	Number of Families
Total	
Own/owner-like possession of house & lot	21,476,446
Rent house/room including lot	14,839,335
Own house rent lot	1,545,227
Own house rent-free lot w/ consent of owner	252,316
Own house rent-free lot w/out consent of owner	3,090,604
Rent-free house & lot with consent of owner	648,550
Rent-free house & lot w/out consent of owner	1,017,995
Not applicable*	57,128

It can be observed that most families in our country own or are owner-like of house & lot

Thank you.

TRAINING COURSE ON STATISTICAL CONCEPTS BEHIND RENTAL CONTROL STUDY AND ESTIMATION OF HOUSING NEEDS FOR SELECTED HUDCC STAFF

10-12 November 2014

Measures of Location

REPUBLIC OF THE PHILIPPINES
PHILIPPINE STATISTICAL RESEARCH AND TRAINING INSTITUTE
J & S Building, 104 Kalayaan Avenue, Diliman, Quezon City 1101, Philippines
Telefax : (632) 433-1745/929-7543/436-1426/426-0620/920-9649
www.srtc.gov.ph

Objectives

- To list and define the most common measures of location
- To demonstrate and apply the use of measures of location;
- Interpret results obtained from each measure.

Measures of location

- not measures of central tendency but assist in measuring the variation of the distribution and its skewness
- also known as quantiles or fractiles

Definition Quantiles or Fractiles

- help describe or locate position of certain non-pieces of data relative to the entire set of data
- specific fraction or percentage of the observations that fall below a certain value
- allied to the median and are based on their positions in a distribution

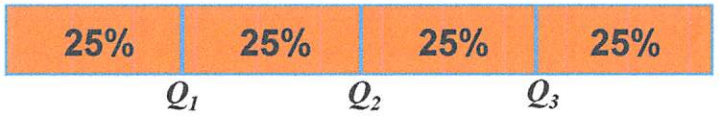
Kinds of Quantiles or Fractiles

A. Quartiles
B. Deciles
C. Percentiles

A. Quartiles

- divides the ordered observations into 4 equal parts
- any of the 3 values which divide the distribution into four equal parts

3 Quartiles



First Quartile

- denoted by Q_1
- distribution is divided into lower ¼ and upper ¾
- the value of x for which $<CF=n/4$

Second Quartile

- denoted by Q_2
- distribution is divided into 2 equal parts
- the value of x for which
- equal to the median

Third Quartile

- denoted by Q_3
- distribution is divided into lower ¾ and upper ¼
- the value of x for which $<CF = 3n/4$

B. Deciles

- Divides the ordered observations into 10 equal parts
- any of the 9 values that divide the distribution into 10 equal parts

The Deciles

We read and interpret the deciles as follows:

D_1, read as first decile, is the value below which 10% of the ordered values fall.

D_2, read as second decile, is the value below which 20% of the ordered values fall.

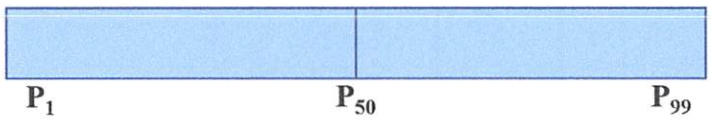
:

:

D_9, read as ninth decile, is the value below which 90% of the ordered values fall.

C. Percentiles

- Divides divide the ordered observations into 100 equal parts
- any of the 99 values that divide the distribution into 100 equal parts



Percentiles

We read and interpret the individual percentiles as follows:

P_1, read as first percentile, is the value below which 1% of the ordered values fall.

P_2, read as second percentile, is the value below which 2% of the ordered values fall.

:

P_{99}, read as ninety-ninth percentile, is the value below which 99% of the ordered values fall.

Procedure in calculating quantiles:

1. Arrange the data in ascending order.
2. Calculate the quantile location ***(i)***.

for decile, $i = \frac{nk}{4}$

for decile, $i = \frac{nk}{10}$

for percentile, $i = \frac{nk}{100}$

where:

k quantile number (e.g. for Q_1, ***k=1***)
n number of observations in the data set

Cont'd...

3. To determine the location of the quantile Z_i^{th} (Z_i can be Q_1, D_3, P_{10} etc.).

   a. If ***i*** is a whole number,

$$Z_i = \frac{(x_i + x_{i+1})}{2}$$ Equation 4.4

   b. If ***i*** is not a whole number,

$$Z_i = x_{i+1}$$ Equation 4.5

Relationship of Quartiles and Percentiles

- first quartile or lower quartile is the 25th percentile;
- second quartile or the median is the 50th percentile; and
- third quartile or the upper quartile is the 75th percentile.
- Quartiles are special cases of percentiles. Thus, the formulas we have for the percentiles are applicable for the quartiles.

Table of Equivalents for Quantiles or Fractiles

Quartile	Decile	Percentile
	D_1	P_{10}
	D_2	P_{20}
Q_1		P_{25}
	D_3	P_{30}
	D_4	P_{40}
Q_2	D_5	P_{50}
	D_6	P_{60}
	D_7	P_{70}
Q_3		P_{75}
	D_8	P_{80}
	D_9	P_{90}

Thus, P_k is a value such that at least k% of the ordered data are smaller than it and at least (100-k)% are larger than it, where k = 1, 2, 3, ..., 99.

For example, the 80th percentile of a distribution is a value such that at least 80 percent of the ordered observations are less than its value and at least 20 percent of the ordered observations are larger than its value.

Examples of Getting the Percentile Using the Empirical Distribution Number with Averaging

1. The annual per capita poverty threshold in pesos of the different regions of the Philippines are as follows: 15,693, 13,066, 12,685, 11,128 13,760, 13,657, 11,995, 11,372, 11,313, 9,656, 9,518, 9,116, 10,503, 10,264, 10,466, 10,896, 12,192. Find the 75th percentile.

Solution: Arrange the 17 annual per capita poverty threshold in pesos of the 17 regions of the Philippines from lowest to highest.

Array: 9116, 9518, 9656, 10264, 10466, 10503, 10,896, 11128, 11313, 11,372, 11995, 12192, **12,685**, 13066, 13657, 13760, 15,693

Compute for nk/100 where n = 17 and k = 75.
nk/100 = 17(75)/100 = 12.75 (not an integer)

Since nk/100 is not an integer, we use the second formula in the empirical number distribution with averaging.

The 75th percentile is 12,685. This implies that 75% of the 17 annual per capita poverty threshold falls below P12,685.

Thank you.

TRAINING COURSE ON STATISTICAL CONCEPTS BEHIND RENTAL CONTROL STUDY AND ESTIMATION OF HOUSING NEEDS FOR SELECTED HUDCC STAFF

10-12 November 2014

Measures of Variation

REPUBLIC OF THE PHILIPPINES
PHILIPPINE STATISTICAL RESEARCH AND TRAINING INSTITUTE
J & S Building, 104 Kalayaan Avenue, Diliman, Quezon City 1101, Philippines
Telefax : (632) 433-1745/929-7543/436-1426/426-0620/920-9649
www.srtc.gov.ph

Objectives

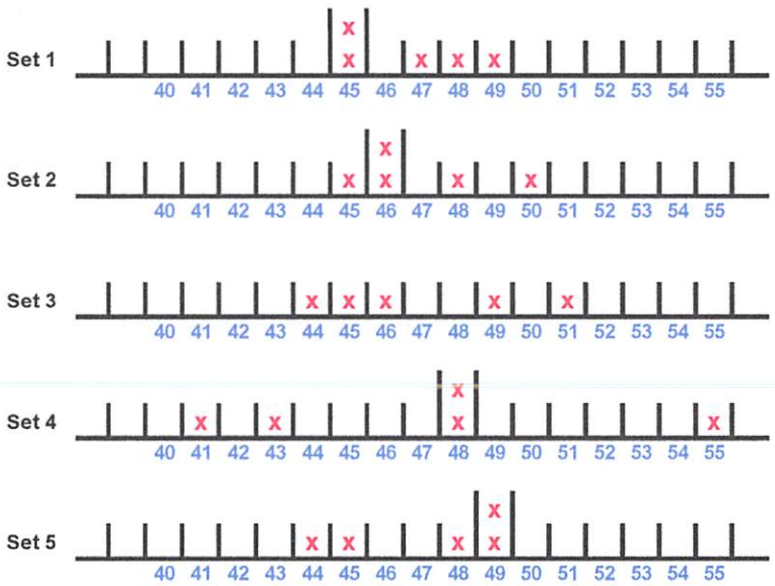
- Gain skills in the computation of the different quantitative measures of dispersion;
- Describe and compare groups and individuals within groups using the measures of dispersion;
- Interpret results obtained from each measure

Let us take 5 sets of observations

Set 1:	45	45	47	48	50	$\bar{x} = 47$
Set 2:	45	46	46	48	50	
Set 3:	44	45	46	49	51	
Set 4:	41	43	48	48	55	
Set 5:	44	45	48	49	49	

Questions remain unanswered even after getting the mean:

- how variable the data sets are?
- how the values in each data set differ from each other?
- how are the values in each data set clustered or dispersed from each other?



Measures of dispersion

- group of analytical tools that describes the spread or variability of a data set.
- indicate the extent to which individual items in a series are scattered about an average.

Importance of the measures of dispersion

- supplements an average or a measure of central tendency
- compares one group of data with another.
- indication on how representative the average is.

A measure of dispersion can be expressed in several ways:

Measures of Dispersion

- Range
- Quartile Deviation — *Based on the position of certain items in a distribution*
- Mean Absolute Deviation
- Variance/ Standard Deviation — *Measure the dispersion around an average*
- Coefficient of variation — *Expressed in a relative value*

Measures of Dispersion

- indicate the extent to which individual items in a series are scattered about an average.

1. **Measures of Absolute Dispersion**
   - use to compare two or more data sets with the same means and the same units of measurement.
2. **Measures of Relative Dispersion**
   - used to compare two or more data sets with different means and different units of measurement.

1. Range (R)

- **difference between the highest and lowest values in a given set of data.**

Formulas of the range:

For Ungrouped Data:

$$R = HV - LV$$

For Grouped Data:

$$R = X_{last\ class} - X_{first\ class}$$

$$R = UCB_{last\ class} - LCB_{first\ class}$$

Range

Characteristic:

- **most crude measure of dispersion.**

Uses:

- **when quickest measure of dispersion is needed.**
- **if information concerning extreme values is desired.**

Range

Advantages:

- **simplest measure of dispersion.**
- **includes the limits within which all of the items occurred.**

Disadvantages:

- **does not consider every observation in the data set.**
- **fails to measure the variability of the majority of the values.**
- **very sensitive to extreme values.**
- **cannot be computed for open-ended distributions.**
- **not amenable to algebraic manipulations.**

Range

Disadvantages:

- **does not consider every observation in the data set.**
- **fails to measure the variability of the majority of the values.**
- **very sensitive to extreme values.**
- **cannot be computed for open-ended distributions.**
- **not amenable to algebraic manipulations.**
- **is unreliable when computed from a frequency distribution table with gaps or zero frequencies.**

Example: Given below are the weights in pounds of five babies below 1 yr. old from Health Center 1, get the range.

10 pounds **lightest** | 12 pounds | 14 pounds | 16 pounds | 20 pounds **heaviest**

Health Center 1

Solution: The maximum or heaviest baby is 20 pounds and the minimum or lightest baby is 10 pounds. Thus, the weight range of babies is

heaviest – lightest = 20-10 = 10 pounds

We can say that the weights of babies range from 10 to 20 pounds.

Example: Given below are the weights of 5 babies from health center 2. Compare the weight range of the babies from health center 1 given in the previous example and health center 2.

12 pounds | 12 pounds | 14 pounds | 12 pounds | 16 pounds

Health Center 2

Solution: Weight range of babies in Health Center 1:

heaviest – lightest = 20 – 10 pounds

Weight range of babies in Health Center 2:

heaviest - lightest = 16 - 12 = 4 pounds

Variance and Standard Deviation

- **are the measures of dispersion that is preferred in most circumstances and by far the most important measure of variation.**

Variance

- **is the average of the squared deviations of each observation in the set from the mean of the dataset. (Shows variation about the mean)**
- **Population variance is denoted by σ^2 (sigma-squared) while the sample variance is denoted by s^2 (s-squared)**

Variance

- Sample variance:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

- Population variance:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Variance and Standard Deviation

Standard Deviation

- square root of the average squared deviations.
- also known as root-mean-square of the deviations from the mean.
- Standard Deviation of the population is represented by the Greek letter σ (sigma) while the sample standard deviation is denoted by s (small s)

Variance/Standard Deviation

Characteristic:

- can only be computed where n is at least 2.
- Variance is always greater than 0.
- Variance is not expressed in the same units as the observations.
- Standard deviation can be seriously affected if the mean is a poor measure of location.

Variance/Standard Deviation

Characteristic:

- if all values of a data set are the same, the standard deviation is zero.
- small standard deviation means a high degree of uniformity and homogeneity of the observed values.
- if the distribution has a few very extreme cases, the standard deviation can give misleading results.

Standard Deviation

- Sample standard deviation: $$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$$
- Population standard deviation: $$\sigma=\sqrt{\frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}}$$

Variance/Standard Deviation

Uses:

- when a dependable measure of dispersion is needed.
- if further statistical analysis is needed.
- when interpretation related to the normal distribution is required.
- when the mean is used as a measure of central tendency.
- Extremely useful in estimating the 'representativeness' of the mean
- when further mathematical computations are needed.
- most widely used measure of dispersion and the easiest to handle algebraically.

Variance/Standard Deviation

Advantages:

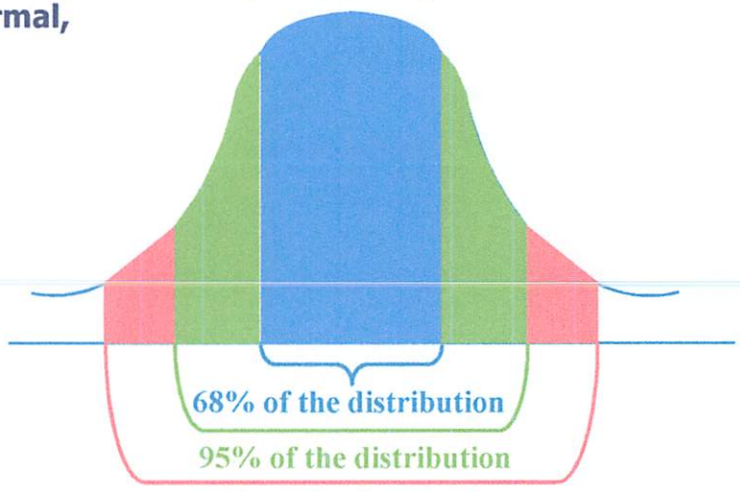
- takes into account every value in the data set.
- most reliable measure of dispersion.
- mathematically logical.
- amenable to further mathematical manipulations.
- can be used for in-depth analysis.

Disadvantages:

- harder to compute and more difficult to understand.
- generally affected by extreme values that may be due to skewness of data.

Variance/Standard Deviation

According to the Empirical Rule, if the distribution is normal,



Standard Deviation

Remarks:

1. If there is a large amount of variation in the data set, then on the average, the data values will be far from the mean. Hence, the standard deviation will be large.

2. If there is only a small amount of variation in the data set, then on the average, the data values will be close to the mean. Hence, the standard deviation will be small.

Variance/Standard Deviation

Steps in computing the Variance (steps 1-5) and Standard deviation (steps 1-6):

1. Calculate the mean.
2. Subtract the mean from each observation.
3. Square each result.
4. Add these squares.
5. Divide this sum by the number of observations.
6. Take the positive square root.

Example for ungrouped data

Data Set 1: Total employment from small establishments

x_i	x_i^2	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
45	2,025	-2	4
45	2,025	-2	4
47	2,209	0	0
48	2,304	1	1
50	2,500	3	9
$\sum x_i = 235$	$\sum x_i^2 = 11,063$		$\sum(x_i-\bar{x})^2 = 18$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{18}{4} = 4.5$$

$$s = \sqrt{4.5} = 2.12$$

Example for ungrouped data

Data Set 1: Total employment from small establishments

x_i	x_i^2
45	2,025
45	2,025
47	2,209
48	2,304
50	2,500
$\sum x_i = 235$	$\sum x_i^2 = 11,063$

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

$$= \frac{5(11,063) - (235)^2}{5(5-1)}$$

$$s^2 = \frac{55,315 - 55,225}{5(4)} = \frac{90}{20} = 4.5$$

$$s = \sqrt{4.5} = 2.12$$

Example for ungrouped data

Interpretation:

- The variability of **total employment from 5 small establishments** is 2.12.

Characteristics	Range	Quartile Deviation	M.A.D.	Standard Deviation
Computation based on	Lowest and highest values	Q_1 and Q_3	Every value	Every value
Affected by extreme values	Greatest	Not by values smaller than Q_1 or larger than Q_3	Affected by every value	Affected by every value
Degree of precision as a measure of dispersion	Rough estimate	Better than the range	Good, but it only measures absolute deviations from the mean or median	Excellent, measures squared deviations from the mean
Mathematical advantages	Easy to compute	Can be used to measure asymmetrical distribution	Easier to compute than standard deviation	Hard to compute, but suitable for further mathematical computations

Coefficient of Variation (CV)

- relative measure of dispersion.
- ratio of the standard deviation to the mean.

Formula of the *CV*:

$$CV = \frac{s}{\overline{x}} \times 100$$

Coefficient of Variation

Characteristics:

- an abstract number expressed in percent.
- demonstrates the relationship between standard deviation and mean, by expressing the risk as a percentage of the mean.

Uses:

- compares distributions where units are different.
- when measure of relative dispersion is needed.

Coefficient of Variation (CV)

Advantages:

- independent of any unit of measurement.
- easy to interpret.

Disadvantage:

- not useful when the mean is close to 0.

Examples

1. **To get the coefficient of variation (CV) using the distribution of the number of vacancies from 43 selected enterprises, we have the following:**

Given: $s = 42.16$

$$\bar{x} = 104.92$$

$$CV = \frac{s}{\bar{x}} \times 100$$

$$= \frac{42.16}{104.92} \times 100$$

$$= 0.4018 \times 100 = 40.18\%$$

- **The variability of the number of vacancies from 43 selected enterprises in relation to its mean is 40.18%.**

Examples

2. **Suppose that there are two sets of data, one for the weights of the employees and the other data set is their income. These two data sets have equal standard deviation. How do we compare these two data sets?**

Given:

weights of the employees:	**income of the employees:**
$s = 100\ lbs.$	$s = Php100$
$\bar{x} = 150\ lbs.$	$\bar{x} = Php30,000$

Examples

Computing the CV for each of the data set, we have:

For the weights of the employees:

$$CV = \frac{s}{\bar{x}} \times 100$$

$$= \frac{100\ lbs.}{150\ lbs.} \times 100$$

$$= 0.6667 \times 100$$

$$= 66.67\%$$

For the income of the employees:

$$CV = \frac{s}{\bar{x}} \times 100$$

$$= \frac{Php100}{Php30,000} \times 100$$

$$= 0.0333 \times 100$$

$$= 3.33\%$$

- **The CV of the weights of the employees is greater than the CV of the income of the employees. This means that the weights of the employees are more variable than their income despite that their standard deviations are equal.**

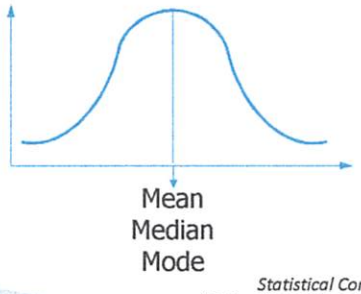
SKEWNESS

- **describes the degree to which the data deviates from symmetry.**
- **when the distribution of the data is not symmetrical, it is said to be asymmetrical or skewed.**
- **the more the mean departs from the mode, the greater the skewness**
- **frequency curve maybe either skewed to the left or skewed to the right**

SKEWNESS

Symmetric Distribution

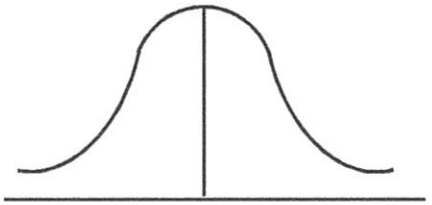
- distribution of data in which the right half is a mirror image of the left half.
- when the distribution has no skewness.
- when the mean, median and mode are equal and all are at the center of the distribution.



KURTOSIS

Mesokurtic distribution

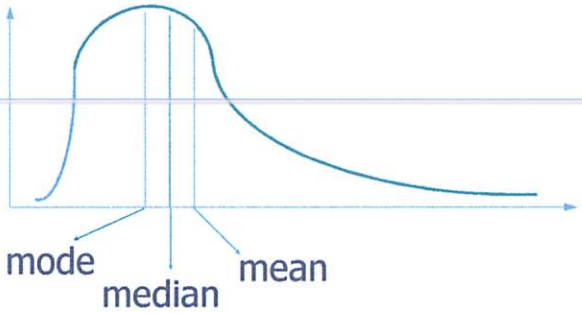
- between the two types of distributions discussed earlier, another type of distribution that is more "normal" in shape.



Mesokurtic distribution

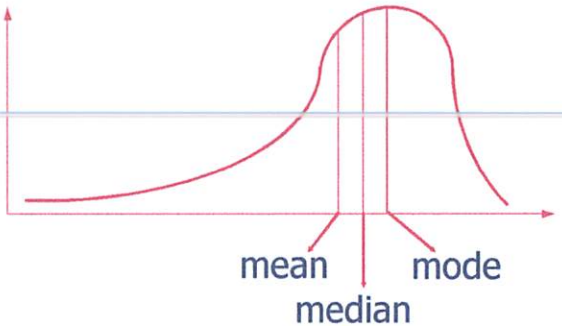
SKEWNESS

- when the mean is greater than the mode, the distribution is said to be positively skewed or skewed to the right.



SKEWNESS

- when the mode is greater than the mean, the distribution is said to be negatively skewed or skewed to the left.



SKEWNESS

Characteristic:

- when the distribution is symmetrical, the mean, median and the mode are all equal.
- a positive value - observations are clustered more to the left of the mean with most of the extreme values to the right of the mean.
- a negative skewness - clustering to the right.

SKEWNESS

Measuring Skewness:

1. Pearsonian Measure of Skewness
2. Measures of Skewness based on quartiles and percentiles

SKEWNESS

1. Pearsonian Measure of Skewness:

$$Skewness = \frac{(\bar{x} - Mode)}{s}$$

$$Skewness = \frac{3(\bar{x} - Median)}{s}$$

Example

Given:

$\bar{x} = 104.92\ vacancies$

$Md = 106.75\ vacancies$

$s = 42.16\ vacancies$

$$Skewness = \frac{3(\bar{x} - median)}{s}$$

$$Skewness = \frac{3(104.92 - 106.75)}{42.16}$$

$$= \frac{-5.49}{42.16}$$

$$= -0.13$$

- Since the value is negative, the skewness of the distribution is to the left.
- This is considered as slightly skewed to the left since the measure does not depart substantially form 0.

Coefficient of Variation

Characteristics:

- an abstract number expressed in percent.
- demonstrates the relationship between standard deviation and mean, by expressing the risk as a percentage of the mean.

Uses:

- compares distributions where units are different.
- when measure of relative dispersion is needed.

SKEWNESS

2. Measure of Skewness based on quartiles and percentiles:

$$\frac{(Q_3 - Median) - (Median - Q_1)}{Q_3 - Q_1} = \frac{(Q_1 + Q_3) - 2Median}{Q_3 - Q_1}$$

$$\frac{(P_{90} - Median) - (Median - P_{10})}{P_{90} - P_{10}} = \frac{(P_{10} + P_{90}) - 2Median}{P_{90} - P_{10}}$$

KURTOSIS

- tool that describes the peakedness of the distribution whether humpbacked, slender and narrow, or broad.
- measure of peakedness or flatness of a distribution.
- measured by making use of the fourth moment around the mean expressed in terms of the fourth power of the standard deviation

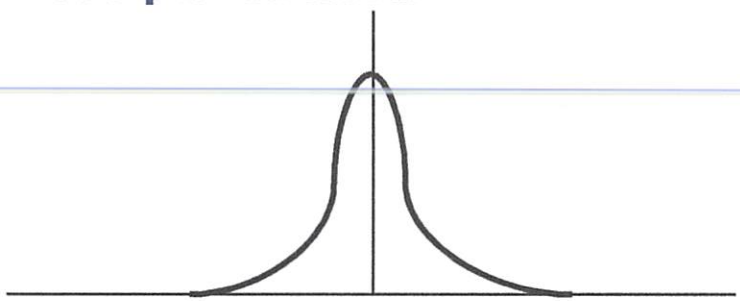
Formula of the kurtosis:

$$\pi_4 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{(n-1)s^4} - 3$$

KURTOSIS

Leptokurtic distribution

- when the values are concentrated in the area around the mode, the distribution has a peaked curve.

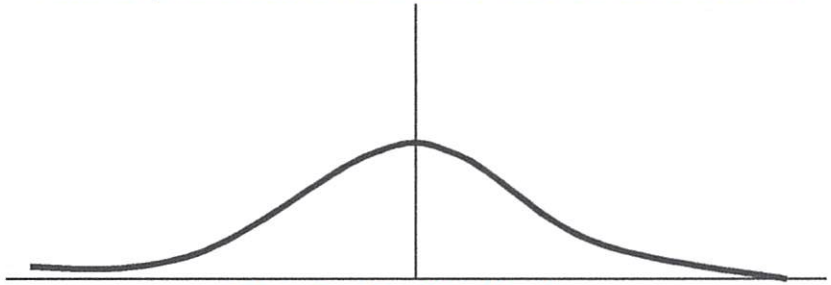


Leptokurtic distribution

KURTOSIS

Platykurtic distribution

- when the values are decentralized from the mode to both tails of the frequency curve, the distribution has a flat curve

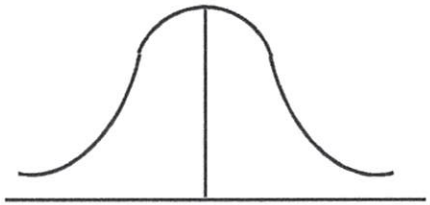


Platykurtic distribution

KURTOSIS

Mesokurtic distribution

- between the two types of distributions discussed earlier, another type of distribution that is more "normal" in shape.



Mesokurtic distribution

KURTOSIS

Characteristic:

- +3 - mesokurtic distribution or standard normal.
- >3 - leptokurtic distribution.
- <3 - platykurtic distribution.
- Kurtosis is less than or equal to the sample size, that is, $\pi_4 \leq n$.
- Kurtosis minus the squared of skewness is greater than or equal to 1, that is, $\pi_4 - Skewness^2 \leq 1$.

Thank you.