

Descriptive Statistics

1

Variables

- **Categorical** variables (variables that take on a category rather than a number)
 - **Non-ordinal** (no rank order)
 - Hair Colour of a person: Black, Brown, Golden
 - Sex of dog: male, female
 - **Ordinal** (with rank order)
 - T-shirt size: small medium, large
 - Response of a patient to drug: None, Partial, Complete

2

Variables

- **Quantitative** variables (variables that can be assigned values)
 - **Discrete** (countable)
 - Number of pimples on the face
 - Number of young produced by a cow
 - **Continuous** (measurable)
 - Weight of an adult
 - Glucose concentration in a blood specimen

3

Population and Sample

- The totality of observations is called **population**
 - Finite population : Observe the birth weight of 800 puppies
 - Infinite population : Observe the height of wheat plants in 10 acres of land
- A **sample** is a subset of the population
 - Random sampling
 - Unbiased

4

Graphical Presentation of Frequency Distributions

- A frequency distribution is a tabulation of the number of occurrences of each unique value in a data set
- Graph
 - [Bar Chart](#)
 - [Grouped Frequency Distributions](#)

5

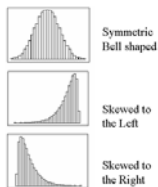
How do we describe data?

- Measures of Central Tendency
 - Measure the centre of a frequency distribution
 - Mode
 - Median
 - Mean
- Measures of Variability/Dispersion
 - Measure the amount of scatter in a dataset
 - Range
 - Variance
 - Standard Deviation

6

How do we describe data?

- Measure amount of skewness in a dataset
 - No skewness (0 value, symmetry)
 - Negatively skewed (negative values, skewed to the left)
 - Positively skewed (positive values, skewed to the right)



7

How to calculate the statistics?

- Using Formulae
- Using Calculator
- Using Statistical Software (Microsoft Excel, SPSS, etc.)

8

Using Formulae

9

Measures of Central Tendency

Mode

– Value with the highest frequency

Example A

Consider the set of data :

2,5,6,4,7,4,7,2,8,9,4,11,9,1,3

$n=15$

1, 2, 3, 4, 4, 4, 5, 6, 7, 7, 8, 9, 9, 11

Mode = 4

Median = 5

10

Measures of Central Tendency

Median

- Arrange in increasing order, the median is the
 - Middle value if the number of observations is odd
 - The arithmetic mean of the middle two values if the number of observations is even

Example B

Consider the set of data :

2,5,6,4,7,4,7,2,8,9,4,11,9,1,3

11

Measures of Central Tendency

Median

- Arrange in increasing order, the median is the
 - Middle value if the number of observations is odd
 - The arithmetic mean of the middle two values if the number of observations is even

Example C

Consider the set of data :

3,1,4,7,9,5,6,8,3,1,2,9,12,3,3,15

$n=16$

1, 1, 2, 3, 3, 3, 3, 4, 5, 6, 7, 8, 9, 9, 12, 15

Median = $\frac{4+5}{2} = 4.5$

12

Measures of Central Tendency

- Mean

- Population Mean $\mu = \frac{\sum_{i=1}^N x_i}{N}$

- Sample Mean $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

- Example D

Consider a sample of data, 2,5,7,10,11,13

$$\bar{x} = \frac{2+5+7+10+11+13}{6} = 8$$

13

Measures of Variability/Dispersion

- Range

- Range = Maximum - Minimum

- Variance and Standard Deviation

- Population

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} = \frac{\sum_{i=1}^N x_i^2}{N} - \bar{x}^2 \quad \sigma = \sqrt{\sigma^2}$$

← Sample (n ≥ 30)

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

↔ Sample (n < 30)

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

$$\bar{x}$$

$$\sigma_n \quad \sigma_{n-1}$$

$$\sigma \quad s$$

14

Example 6.24

Find the sample median and range for the data in

(a) Example 6.22: 4, 6, 2, 4, 2, 1, 5, 1, 3, 1, 3, 4, 5, 4, 3, 1, 6, 5, 6, 3, 1, 2, 4, 6

$$\text{Range} = 6 - 1 = 5$$

Example 6.24

Find the sample median and range for the data in

(b) Example 6.23: 106, 98, 95, 109, 99, 102, 101, 108, 94, 99, 96, 102

$$\text{Range} = 109 - 94 = 15$$

Using Calculator

17

Descriptive Statistics

- Example E

Consider a sample of data,

$y_i : 2, 5, 7, 10, 11, 13$

Using Calculator CASIO fx-992s

1. Set calculator mode to SD by pressing [MODE] 2
2. 2 [DATA], 5 [DATA], 7 [DATA], ... $M+$
3. [RCL] 1 gives $\sum x^2 = 468$
4. [RCL] 2 gives $\sum x = 48$
5. [RCL] 3 gives $n = 6$
6. [SHIFT] 1 gives $\bar{x} = 8$
7. [SHIFT] 3 gives $s = 4.09878\dots$

18

Example 6.22 (pg 412)

A die was tossed 24 times, producing the following results: 4, 6, 2, 4, 2, 1, 5, 1, 3, 1, 3, 4, 5, 4, 3, 1, 6, 5, 6, 3, 1, 2, 4, 6

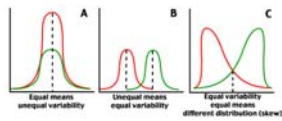
Find the sample average and standard deviation and compare results to Example 6.16 and 6.17.

Example 6.23 (pg 413)

Measured values of resistance (in Ω) for 12 nominally 100Ω resistors were as follows: 106, 98, 95, 109, 99, 102, 101, 108, 94, 99, 96, 102. Find the sample average and both forms of sample variance and standard deviation.

Use of Mean and Standard Deviation

- Consider 2 sets of data
 - Set A : 5, 3, 4, 4, 4
 - Set B : 5, 3, 4, 6, 2
- What is their means?
- What is their standard deviations?



21

Using Software ([Excel](#))

22