SECOND YEAR OF SURVEYING DEPARTMENT

THEORY OF ERROR 2017

LECTURE 2

OBSERVATIONS AND THEIR ANALYSIS

DR. ENG. HASSAN MOHAMED

LECTURE ELEMENTS

1. " INTRODUCTION"
2. "SAMPLE VERSUS POPULATION"
3. "RANGE AND MEDIAN"
4. "GRAPHICAL REPRESENTATION OF DATA"
5. "NUMERICAL METHODS OF DESCRIBING DATA"
6. "MEASURES OF CENTRAL TENDENCY"
7. "ADDITIONAL DEFINITIONS"
8. " NUMERICAL EXAMPLES "

DR. ENG. HASSAN MOHAMED

INTRODUCTION

- SETS OF DATA CAN BE REPRESENTED AND ANALYZED USING EITHER GRAPHICAL OR NUMERICAL METHODS.
- A BAR CHART SHOWING DAILY HIGH TEMPERATURES OVER THE PAST MONTH IS AN EXAMPLE.
- OR, INSTEAD OF USING A BAR CHART, THE DAILY HIGH TEMPERATURES COULD BE TABULATED AND THE MEAN COMPUTED.

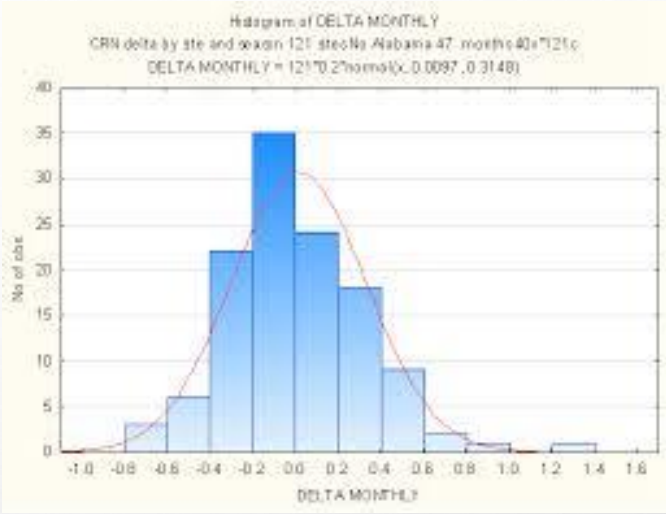
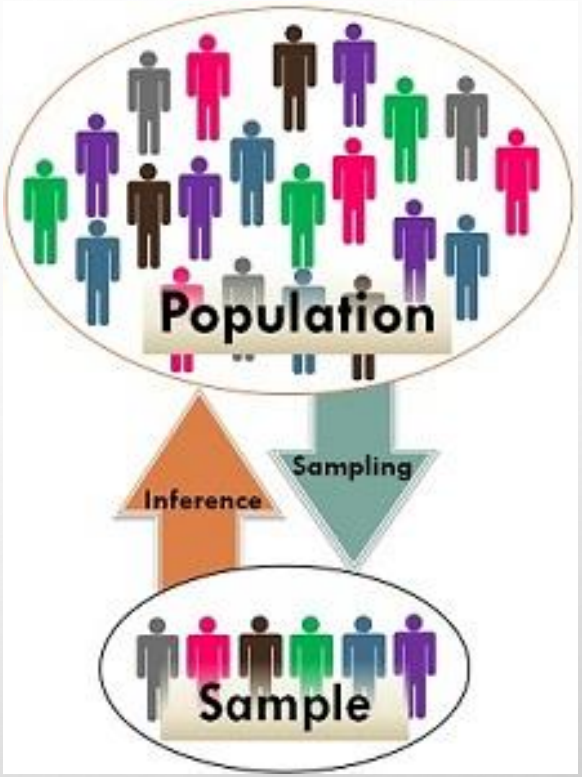


TABLE 3. Comparison of annual rainfall (mm)for Armidale from 1857 to 2015					
Statistic	all records	last 100	last 75	last 50	last 25
min	421	460	460	460	537
10%	558	558	606	597	608
20%	642	636	647	646	636
25%	664	649	673	659	647
30%	680	675	683	675	658
40%	740	737	746	749	678
50%	769	765	772	781	766
average	788	770	781	764	757
60%	810	789	817	818	816
70%	859	846	860	847	829
75%	894	866	881	866	838
80%	918	901	902	887	867
90%	1006	953	973	915	936
max	1508	1309	1309	1048	1048

DR. ENG. HASSAN MOHAMED

SAMPLE VERSUS POPULATION

- *POPULATION.* A POPULATION CONSISTS OF ALL POSSIBLE MEASUREMENTS THAT CAN BE MADE ON A PARTICULAR ITEM OR PROCEDURE. OFTEN, A POPULATION HAS AN INFINITE NUMBER OF DATA ELEMENTS.
- *SAMPLE.* A SAMPLE IS A SUBSET OF DATA SELECTED FROM THE POPULATION.



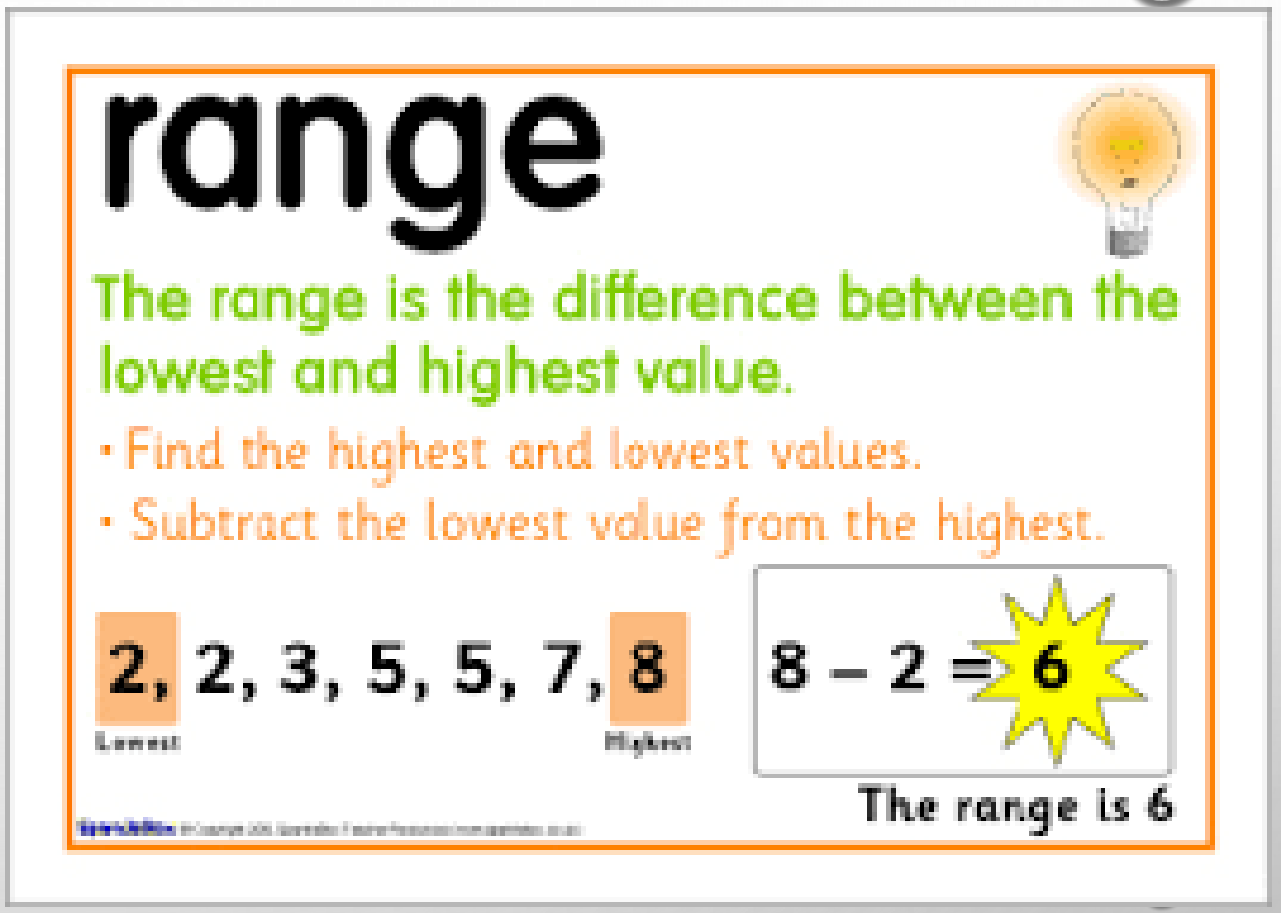
DR. ENG. HASSAN MOHAMED

RANGE AND MEDIAN

ONE QUICK NUMERICAL METHOD USED TO ANALYZE DATA IS TO COMPUTE ITS RANGE,

ALSO CALLED DISPERSION. A RANGE IS THE DIFFERENCE BETWEEN THE HIGHEST AND

LOWEST VALUES. IT PROVIDES AN INDICATION OF THE PRECISION OF THE DATA.

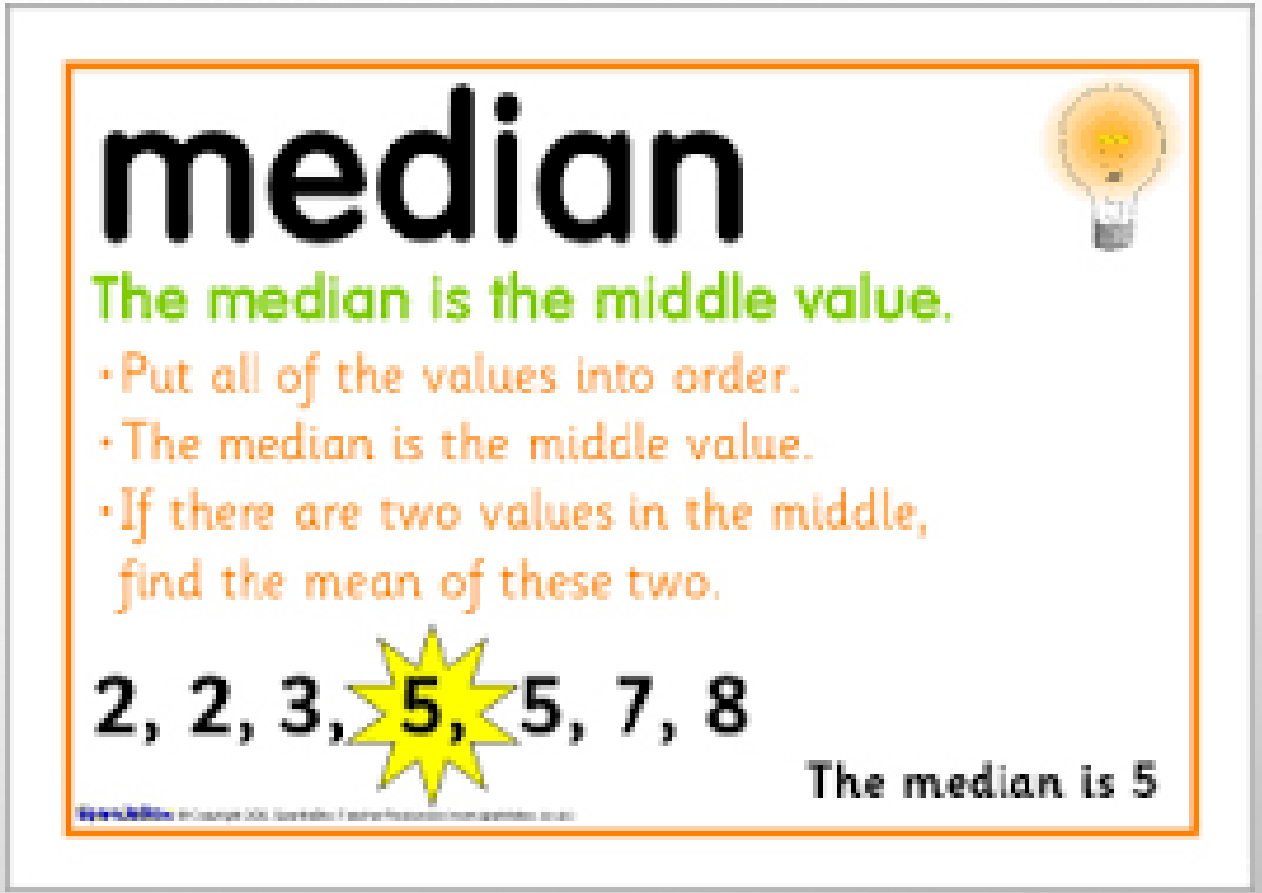


DR. ENG. HASSAN MOHAMED

RANGE AND MEDIAN

THE MIDPOINT VALUE IS ALSO KNOWN AS THE MEDIAN.

SINCE THERE ARE AN EVEN NUMBER OF VALUES IN THIS EXAMPLE, THE MEDIAN IS GIVEN BY THE AVERAGE OF THE TWO VALUES CLOSEST TO (WHICH STRADDLE) THE MIDPOINT.



DR. ENG. HASSAN MOHAMED

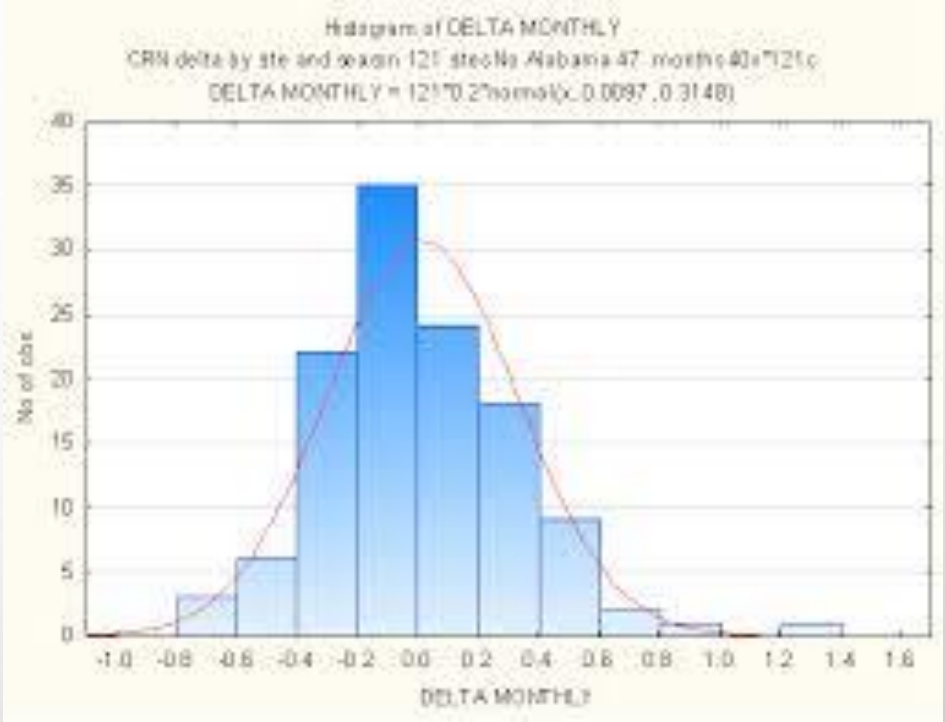
GRAPHICAL REPRESENTATION OF DATA

HISTOGRAMS ARE BAR GRAPHS THAT SHOW THE FREQUENCY DISTRIBUTIONS IN DATA. TO CREATE A HISTOGRAM, THE DATA ARE DIVIDED INTO CLASSES.

THESE ARE SUBREGIONS OF DATA THAT USUALLY HAVE A UNIFORM RANGE IN VALUES, OR CLASS WIDTH. ALTHOUGH THERE ARE NO UNIVERSALLY APPLICABLE RULES FOR THE SELECTION OF CLASS WIDTH, GENERALLY 5 TO 20 CLASSES ARE USED.

IN GENERAL, THE SMALLER THE DATA SET, THE LOWER THE NUMBER OF CLASSES USED.

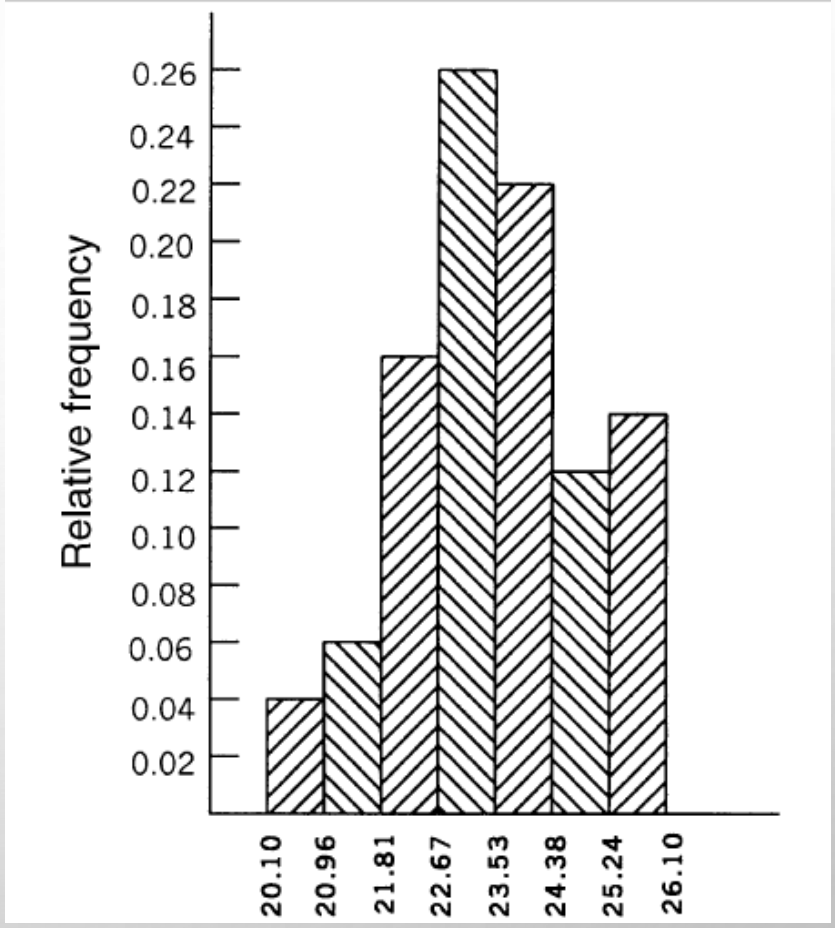
THE HISTOGRAM CLASS WIDTH (RANGE OF DATA REPRESENTED BY EACH HISTOGRAM BAR) IS DETERMINED BY DIVIDING THE TOTAL RANGE BY THE SELECTED NUMBER OF CLASSES.



DR. ENG. HASSAN MOHAMED

GRAPHICAL REPRESENTATION OF DATA

OFTEN, IT IS ALSO USEFUL TO CALCULATE THE CLASS RELATIVE FREQUENCY FOR EACH INTERVAL. THIS IS FOUND BY DIVIDING THE CLASS FREQUENCY BY THE TOTAL NUMBER OF OBSERVATIONS. FOR THE DATA IN TABLE 2.2, THE CLASS RELATIVE FREQUENCY FOR THE FIRST CLASS INTERVAL IS 2/50 = 0.04.

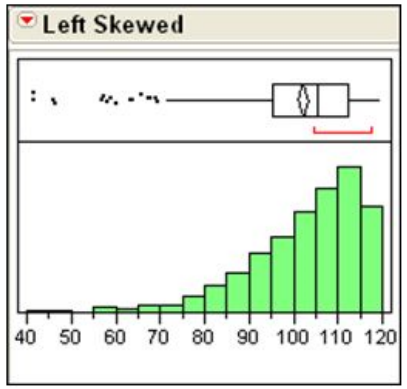


DR. ENG. HASSAN MOHAMED

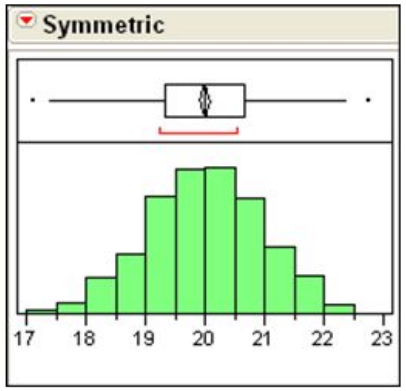
GRAPHICAL REPRESENTATION OF DATA

Histograms

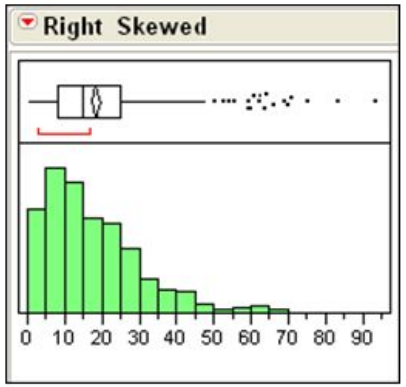
Common Histogram Shapes



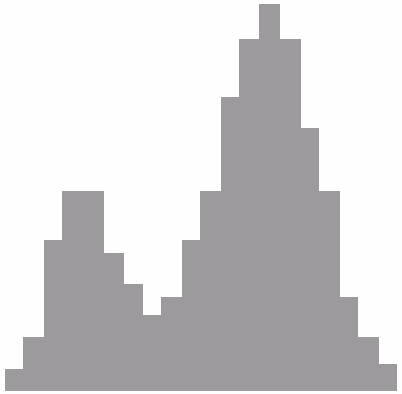
Left Skewed: Data trails off to the left.



Symmetric: Data has approximately the same distribution on either side of the center.



Right Skewed: Data trails off to the right.



Bimodal (double-peaked) distribution

DR. ENG. HASSAN MOHAMED

NUMERICAL METHODS OF DESCRIBING DATA

NUMERICAL DESCRIPTORS ARE VALUES COMPUTED FROM A DATA SET THAT ARE USED TO INTERPRET ITS PRECISION OR QUALITY. NUMERICAL DESCRIPTORS FALL INTO THREE CATEGORIES:

(1) MEASURES OF *CENTRAL TENDENCY*,

(2) MEASURES OF *DATA VARIATION*,

(3) MEASURES OF *RELATIVE STANDING*.

THESE CATEGORIES ARE ALL CALLED *STATISTICS*.

DR. ENG. HASSAN MOHAMED

MEASURES OF CENTRAL TENDENCY

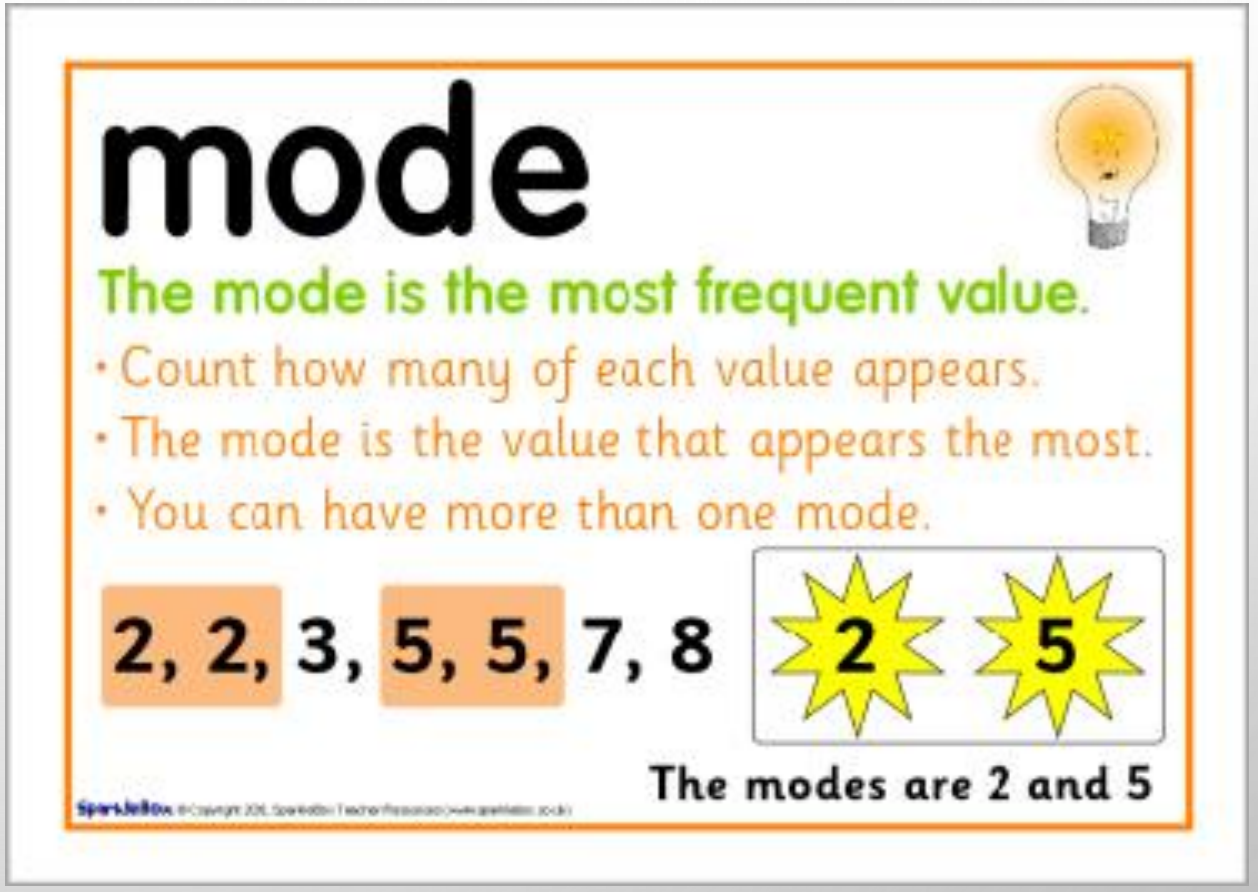
ARITHMETIC MEAN: FOR A SET OF N OBSERVATIONS, X1, X2, . . . , XN, THIS IS THE AVERAGE OF THE OBSERVATIONS. ITS VALUE, X, IS COMPUTED FROM THE EQUATION.

$$\text{arithmetic mean} = \frac{\sum_{n=1}^{k} x_n}{k}$$

DR. ENG. HASSAN MOHAMED

MEASURES OF CENTRAL TENDENCY

MODE. WITHIN A SAMPLE OF DATA, THE MODE IS THE MOST FREQUENTLY OCCURRING VALUE.



DR. ENG. HASSAN MOHAMED

ADDITIONAL DEFINITIONS

1. *TRUE VALUE, μ:* A QUANTITY'S THEORETICALLY CORRECT OR EXACT VALUE. THE TRUE VALUE CAN NEVER BE DETERMINED.
2. *ERROR, ε:* THE DIFFERENCE BETWEEN A MEASURED QUANTITY AND ITS TRUE VALUE.
3. *MOST PROBABLE VALUE, $\acute{Y}$:* THAT VALUE FOR A MEASURED QUANTITY WHICH, BASED ON THE OBSERVATIONS, HAS THE HIGHEST PROBABILITY OF OCCURRENCE.
4. *RESIDUAL, V:* THE DIFFERENCE BETWEEN ANY INDIVIDUAL MEASURED QUANTITY AND THE MOST PROBABLE VALUE FOR THAT QUANTITY.

$$v_i = \bar{y} - y_i$$

ADDITIONAL DEFINITIONS

5. VARIANCE, σ^2: A VALUE BY WHICH THE PRECISION FOR A SET OF DATA IS GIVEN. POPULATION VARIANCE APPLIES TO A DATA SET CONSISTING OF AN ENTIRE POPULATION. IT IS THE MEAN OF THE SQUARES OF THE ERRORS AND IS GIVEN BY

$$\sigma^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n}$$

6. SAMPLE VARIANCE: APPLIES TO A SAMPLE SET OF DATA.

$$S^2 = \frac{\sum_{i=1}^{n} v_i^2}{n - 1}$$

DR. ENG. HASSAN MOHAMED

ADDITIONAL DEFINITIONS

7. STANDARD DEVIATION, S: THE SQUARE ROOT OF THE SAMPLE VARIANCE. IT IS CALCULATED USING THE EXPRESSION

$$S = \sqrt{\frac{\sum_{i=1}^{n} v_i^2}{n - 1}}$$

8. STANDARD DEVIATION OF THE MEAN: THE ERROR IN THE MEAN COMPUTED FROM A SAMPLE SET OF MEASURED VALUES THAT RESULTS BECAUSE ALL MEASURED VALUES CONTAIN ERRORS.

$$S_{\bar{y}} = \pm \frac{S}{\sqrt{n}}$$

NUMERICAL EXAMPLES

THE DATA SET SHOWN BELOW ALSO REPRESENTS THE SECONDS' PORTION OF 50 THEODOLITE OBSERVATIONS OF A DIRECTION. COMPUTE THE MEAN, MEDIAN, MODE, AND THE STANDARD DEVIATION. ALSO CONSTRUCT A HISTOGRAM.

34.2 33.6 35.2 30.1 38.4 34.0 30.2 34.1 37.7 36.4

37.9 33.0 33.5 35.9 35.9 32.4 39.3 32.2 32.8 36.3

35.3 32.6 34.1 35.6 33.7 39.2 35.1 33.4 34.9 32.6

36.7 34.8 36.4 33.7 36.1 34.8 36.7 30.0 35.3 34.4

33.7 34.1 37.8 38.7 33.6 32.6 34.7 34.7 36.8 31.8

DR. ENG. HASSAN MOHAMED

NUMERICAL EXAMPLES

MEAN: $\acute{Y} = \Sigma Y_I / N = 1737.0/50 = 34.74$”

MEDIAN: THE MEDIAN IS BETWEEN THE 25TH AND 26TH VALUES, WHICH ARE BOTH 34.7”. THUS, THE MEDIAN IS 34.7”.

MODE: THE DATA HAVE THREE DIFFERENT VALUES THAT OCCUR WITH A FREQUENCY OF THREE. THUS, THE MODES FOR THE DATA SET ARE THE THREE VALUES 32.6”, 33.7”, AND 34.1”.

RANGE: THE RANGE OF THE DATA IS 39.3” - 30.0” = 9.3”.

DR. ENG. HASSAN MOHAMED

NUMERICAL EXAMPLES

- VARIANCE: BY EQUATION (2.10), USING THE SUM OF OBSERVATIONS SQUARED IN TABLE 2.5, THE SAMPLE VARIANCE IS

$$S^2 = \frac{\sum y_i^2 - n\bar{y}^2}{n-1} = \frac{60{,}584.48 - 50(34.74)^2}{50-1} = 4.92$$

AND THE SAMPLE STANDARD DEVIATION IS

$$S = \sqrt{4.92} = \pm 2.22''$$

NUMERICAL EXAMPLES

Class	Class Frequency	Relative Class Frequency
29.15–30.01	1	0.02
30.01–30.87	2	0.04
30.87–31.73	0	0.00
31.73–32.59	3	0.06
32.59–33.45	6	0.12
33.45–34.31	11	0.22
34.31–35.17	7	0.14
35.17–36.03	6	0.12
36.03–36.89	7	0.14
36.89–37.75	1	0.02
37.75–38.61	3	0.06
38.61–39.47	3	0.06
	50	1.00

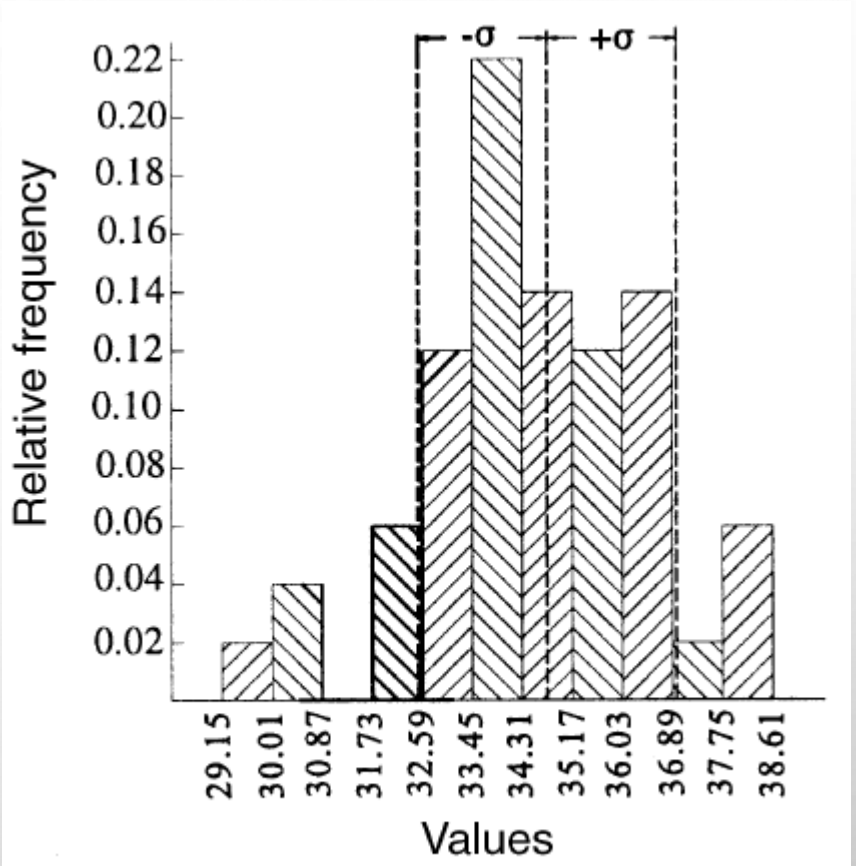


Figure Histogram for Example 2.2.

DR. ENG. HASSAN MOHAMED

SUPPLEMENTARY FILES

1. PLEASE VISIT THE LINK HTTPS://WWW.YOUTUBE.COM/WATCH?V=81ZCJULLH58
2. PLEASE VISIT THE LINK HTTPS://WWW.YOUTUBE.COM/WATCH?V=KCH_ZDYGRM4
3. PLEASE VISIT THE LINK HTTPS://WWW.YOUTUBE.COM/WATCH?V=DIYKISXPAB8
4. PLEASE VISIT THE LINK HTTPS://WWW.YOUTUBE.COM/WATCH?V=E4HAYD0QNRC

DR. ENG. HASSAN MOHAMED