NOMINAL- No ordering of the items
ORDINAL- are **categorical data where there is a logical ordering** to the categories. **(The simplest ordinal scale is a ranking)**
INTERVAL-'0' is only an arbitrary reference point '0' does not mean "nothing" The **interval scale of measurement *only*** permits mathematical operations of **addition and subtraction**.
CHECK TO SEE IF '0' does mean the absence of the characteristic being measured, i.e., '0'='nothing'
CHECK TO SEE IF Addition/Subtract data values and get meaningful results.
RATIO-'0' does mean the absence of the characteristic being measured, '0'='nothing'. Ratio of (division) data values is meaningful. Addition/Subtract data values and get meaningful results. 0' does mean the absence of the characteristic being measured, '0'='nothing'

Example 1

Step 1: Arrange the data into an *ascending* data array:

18 18 20 20 20 20 21 22 22 24 25 28 29 29 38 40 45 52 63

Step 2: Calculate the rank of the kth percentile

$$r = HalfRound\left[\frac{nk}{100}+\frac{1}{2}\right]$$

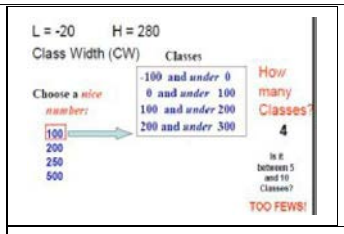
$$r = HalfRound\left[\frac{(20)(80)}{100}+\frac{1}{2}\right] = HalfRound[16.5] = 16.5$$

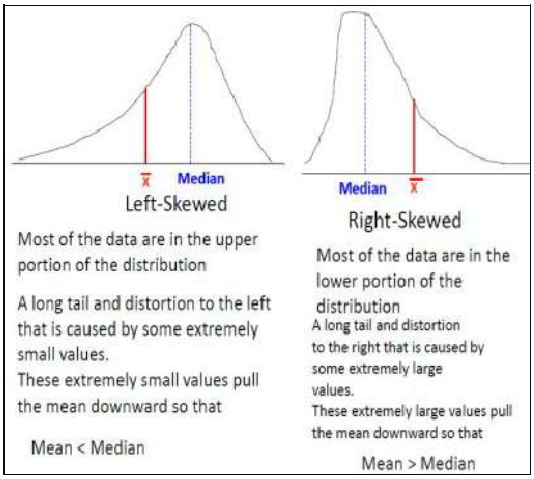
Step 3: Compute P_k

Data	18	18	18	20	20	20	20	21	22	22	24	25	28	29	29	38	40	45	52	63
Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Rank =16.5 is between rank 16 and rank 17

$$P_{80} = \frac{(16th\ value + 17th\ value)}{2} = \frac{38+40}{2} = 39 years$$





Relative Frequency Distribution

...shows the portion of observations in each class.

Time Studying (hrs)	f	rf
10.0 and under 15.0	7	7/30 = 0.233
15.0 and under 20.0	12	12/30 = 0.400
20.0 and under 25.0	7	7/30 = 0.233
25.0 and under 30.0	3	3/30 = 0.100
30.0 and under 35.0	1	1/30 = 0.033
Total	30	30/30 =1

OGIVE

Plot the crf for each class at the upper boundary
Join the points with straight lines. (Start the graph at the 1st boundary)

Amount in a 2 liter bottle (ml)	# of bottles f	rf	crf	c%
1999.00 and under 1999.50	5	5/255=0.020	0.020	0.020*100=2.0
1999.50 and under 2000.00	10	10/255=0.039	0.059	0.059*100=5.9
2000.00 and under 2000.50	20	20/255=0.078	0.137	13.7
2000.50 and under 2001.00	35	35/255=0.137	0.274	27.4
2001.00 and under 2001.50	55	55/255=0.216	0.490	49.0
2001.50 and under 2002.00	50	50/255=0.196	0.686	68.6
2002.00 and under 2002.50	40	40/255=0.157	0.843	84.3
2002.50 and under 2003.00	30	30/255=0.118	0.961	96.1
2003.00 and under 2003.50	10	10/255=0.039	1.000	100.00
TOTAL	255	1.00		

Horizontal axis

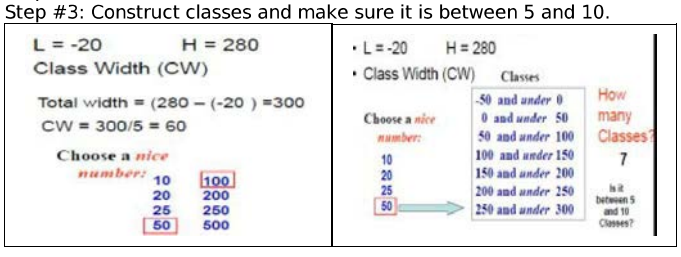
Vertical axis

Nominal	Ordinal	Interval	Ratio
1. Religious preference 2. Ethnicity 3. Gender 4. Marital status 5. zip codes 6. Area of country	Movie ratings (0, 1 or 2 thumbs up). U.S.D.A. quality of beef ratings (good, choice, prime) When a market researcher asks you to rank 5 types of beer from most flavourful to least flavourful, he/she is asking you to create an ordinal scale of preference	Temperature (Degrees F or Degrees C) Most personality measures. WAIS intelligence score (IQ scores) SAT scores	Annual income in dollars. Household size Length or distance in centimeters, inches, miles, etc. Height, weight, volume

Stems:
1. Should be from 6 to 13 stems
2. Should be consecutive numbers or repeated numbers. The numbers may each be repeated twice or 5 times.
3. Units must be indicated if stem not be taken at face value
4. There must be at least one leaf associated with the first and the last stem.

Leaves:
1. The leaf for each data value is the next single digit after the stem.
2. There is no rounding off.
3. They are written in ascending order.
4. They must be evenly spaced.
5. No commas or dashes between the numbers are allowed.

Frequency Distribution:/Relative Frequency
Step #1: CW = (H-L) divided by 5
CW = (280 – (-20)) divided by 5 = 60
Step #2: Find a nice number 1, 2, 2.5, 5, 10, 20, 25, 50 etc.
Step #3: Construct classes and make sure it is between 5 and 10.



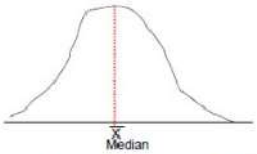
Measure of Position
Percentile
*The percentile symbol is Pk = kth percentile
The kth percentile in a data set is the value such that at most k% of the data is lower than the value and at most (100 – k)% of the data is higher than the value.
****For example, the 32th percentile is the value (or score) below which 32 percent of the observations may be found.****
P32=32th percentile

Example 1
The following data represents the ages, in years, of 20 employees working at a retail outlet. The data has already been arranged in an ascending data array.
18, 18, 18, 20, 20, 20, 20, 21, 22, 22, 24, 25, 28, 29, 29, 38, 40, 45, 52, 63
a) Determine the 80th percentile.

25th percentile = P_{25} = Q_1 = first quartile
✓ **50th percentile = P_{50} = Q_2 = Second quartile = the median**
✓ **75th percentile = P_{75} = Q_3 = Third quartile**
✓ MEAN sum of all data/number of data (average)
✓ MEDIAN middle of a data set when sorted in ascending order
✓ MODE the most frequent number that is occurring within the data set

- ✓ SAMPLE DATA the data obtained from a sample
- ✓ POPULATION DATA the data obtained from a population

Symmetrical Shape



Each half of the curve is a mirror image of the other half of the curve.
The low and high values on the scale balance,

Mean = Median

Sample data The data obtained from a sample	**Population data** The data obtained from a population
$\bar{x} = \frac{\sum x}{n}$	$\mu = \frac{\sum x}{N}$
Notations:	
$\bar{x}$ = sample mean	μ = population mean
x= data (from sample)	x=data (from population)
n-sample size	N=sample size
∑ = summation therefore ∑x = add up the data	

Type of Mean	Arithmetic Mean	Weighted Mean	"Total" Mean
Data Source	Sample	Sample	Sample
Nature of data	Raw (ungrouped) Data	Grouped data	Grouped data

Example 1: Raw Data

A researcher, engaged by a Retail Industry, took a random sample of 10 employees working in the retail outlets located in the GTA (Greater Toronto Area). These 10 employees earned **$37, 22, 22, 19, 19, 19, 28, 21, 20, 32** thousand annually. Calculate the average salary of these employees.

In this example you are given a set of raw data to compute the average: **$37, 22, 22, 19, 19, 19, 28, 21, 20, 32**

To calculate the mean you use the formula, $\bar{x} = \frac{\sum x}{n}$

$$\bar{x} = \frac{37+22+22+19+19+19+28+21+20+32}{10} = \frac{239}{10}$$

$\bar{x}$=$23.9 thousand per employee

CONTINUED .....

37 EXE
22 EXE
22 EXE
19 EXE
19 EXE
28 EXE
21 EXE
20 EXE
32 EXE

5. Press F6 ($ET) key

You will get the following display:

1Var	XList	List 1
1Var	Freq	1
2Var	Xlist	List 2
2Var	YList	List 2
2Var	Freq	List 1

$$\bar{X} = \frac{\Sigma X}{n} = \frac{37+22+22+19+19+28+21+20+32}{1+1+1+1+1+1+1+1+1} \quad \text{Number of employees}$$

$$= \frac{37+22+22+19+19+28+21+20+32}{7}$$

To calculate the MEAN.

You will see at the bottom of the screen:

GRPH	CALC	TEST	INTR	DIST	▷
F1	F2	F3	F4	F5	F6

For other menu Choices

4. Press F2 (CALC) key

You will see at the bottom of the screen:

1 VAR	2 VAR	REG			SET
F1	F2	F3	F4	F5	F6

5. Press F6 ($ET) key

5. Press EXIT key to return to the display of the data

You will see at the bottom of the screen:

1 VAR	2 VAR	REG			SET
F1	F2	F3	F4	F5	F6

6. Press F1 (1 VAR) key

You will see a full range of 1- variable statistics on the screen

7. Read the relevant results

Example 2: Grouped Data

X=Annual Salary (in $ 000)	W=Number of employee
37	1
32	1
28	1
22	2
21	1
20	1
19	3

Calculate the average salary of these employees.

To calculate you use formula, $\bar{x} = \frac{\sum (w)(x)}{\sum w}$

$\bar{x}$=$23.9 thousand per employee

Note that the mean salary in Example 2 is the same as in Example 1. Whether the data is ungrouped (raw) or grouped with the same data values, it should give you the same mean values.

Calculator

To obtain the mean from a set of grouped data, you should follow these calculator steps are;

DATA (taken from Example 2) :37, 32, 29, 22, 21, 20, 19

WEIGHTS: 1, 1, 1, 2, 1, 1, 3

Select STAT mode, enter the data values in List 1 and the weights values in List 2

Press F2 to select CALC
Press F6 to select SET
Set the values as follows:
1Var XList : Press F1 to select LIST 1
1Var Freq : Press F1 to select LIST 2
Press EXE
Press F1 to select 1Var for the results

Example 4: Grouped data

Using Example 1 data set, the researcher grouped the data as shown in Table 3:

Column A	Column B	Column A x Column B=Column C
X = Annual Salary (in $ 000)	W =Number of employee	WX= Total Annual Salary (in $000)
37	1	(37 x 1) = 37
32	1	(32 x 1) =32
28	1	(28 x 1) = 28
22	2	(22 x 2) = 44
21	1	(21 x 1) = 21
20	1	(20 x 1) = 20
19	3	(19 x 3) = 57

Arithmetic Mean	Weighted Mean	"Total" Mean
Example 1	Example 2	Example 4
Raw (ungrouped) Data X values: **37, 22, 22, 19, 19, 19, 28, 21, 20, 32**	Grouped data X Values: 37, 32, 28, 22, 21, 20, 19 Weights: 1,1,1,2,1,1,3	Grouped data
$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum wx}{\sum w}$	$\bar{x} = \frac{\sum Total}{\sum w}$
SET MODE: 1 Var XLIST : X values 1 Var FREQ : 1	SET MODE 1Var XList : X values 1Var FREQ : Weights Press EXE	You cannot use Stat mode for this kind of mean calculation. Just use the regular calculator to do the simple division.

10% Rule

1: Calculate the 'difference' between the mean and median.
Difference = |mean- median|
2: Calculate '10% of the smaller' of the mean or median.
10% smaller = Minimum (10%*mean, 10%*median)
3: Compare 'Difference' with '10% smaller'

The following decision rule is:

****If Difference **is less than** 10% of smaller, you conclude that mean is approximately equal to median, in which case the mean is the preferred measure.
****If Difference **is greater than** 10% of smaller, you conclude that mean is **NOT** equal to median, in which case the median is the preferred measure.
If the mean and median are **CLOSE, then the **mean** will give the correct impression and is the preferred measure.
If the mean and median are **NOT CLOSE, then the **median** will give the correct impression

Example 1

Consider the following:
Mean = 12 and Median = 14

Apply the 10% rule, you have

1. Difference = |Mean-Median|=|12 – 14|=2
2. 10% smaller = Minimum (10%*12, 10%*14) = Minimum (1.2, 1.4) = 1.2

3. Since Difference of 2 is **greater** than 10% smaller of 1.2, which indicate that mean is **NOT** equal to median, in which case the median is the preferred measure.

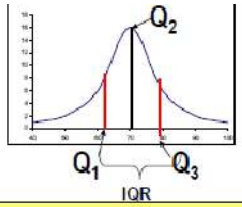
Measures of Variability

1. **Range (R)**
   R=High - Low
2. **Interquartile Range (IQR)**
   IQR = Q3 – Q1
   That is, it is the difference between the 75th and 25th percentiles of a variable.
3. **Variance**
4. **Standard Deviation**: standard deviation is simply the square root of the variance

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

Obtain the standard Deviation (s) from The CASIO calculator

Quartiles



Quartile	Notation	Location
First Quartile	Q_1	Is a value for which 25% of the observations are smaller and 75% are larger
Second Quartile	Q_2	Is a value for which 50% of the observations are smaller and 50% are larger
Third Quartile	Q_3	Is a value for which 75% of the observations are smaller and 25% are larger

Summary
Characteristics of the Range, IQR, Variance and Standard Deviation (Std Dev)

1. The more spread out, or dispersed, the data are, the larger the R, IQR, Var, Std Dev.
2. The more concentrated, or homogeneous the data are, the smaller the R, IQR, Var, Std Dev.
3. If the values are all the same (so that there is no variation in the data), the R, IQR, Var and Std Dev = 0
4. None of the measure of variation (R, IQR, Std Dev) can ever be negative.

Coefficient of Variability

Symbol : CV

Sample data: $CV = \frac{s}{\bar{x}} x100$

Population data: $CV = \frac{\sigma}{\mu} x100$

Stock	Price	Dividend
ABC Company	$25	$0.70
DEF Company	47	1.50
HIJ Company	78	2.00
XYZ Company	92	3.00

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

$s=\$30.23$, $\bar{x}=\$60.50$: $CV=\frac{30.23}{60.50} \times 100 = 50\%$

$s=\$0.963$, $\bar{x}=\$1.80$: $CV=\frac{0.963}{1.80} \times 100 = 54\%$

Conclusion: CV price < CV dividend. Dividend is more variable than Price.

*****USE SAMPLE DEVIATION

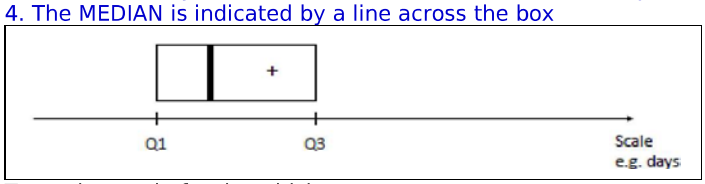
Box Whisker Plot	**Box Whisker Plot summarizes:**
This is referred as the "FIVE – NUMBER SUMMARY" Consists of: 1. Minimum 2. Q1 3. Median (Q2) 4. Q3 5. Minimum	•Measures of Central Tendency (Median) •Measure of Variation (Range, Interquartile Range) •Measure of Skewness (Shape of a data set)

Box Whisker Plot

Box Whisker Plot provides a graphical representation of the data based on the 5 number summary:
Minimum (the smallest observation)
Q_3 - The upper quartile
Q_2 - The median
Q_1 - The lower quartile
Maximum (the largest observation)

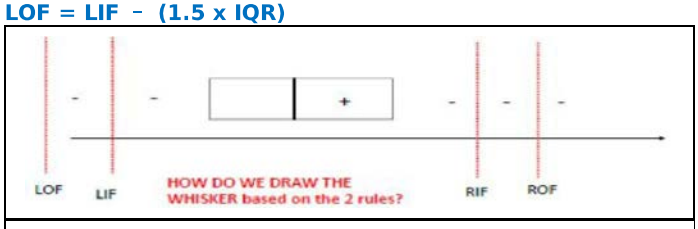
What is the purpose of a Box Whisker Plot?
1. How a data set look like?
2. Is there any outliers?

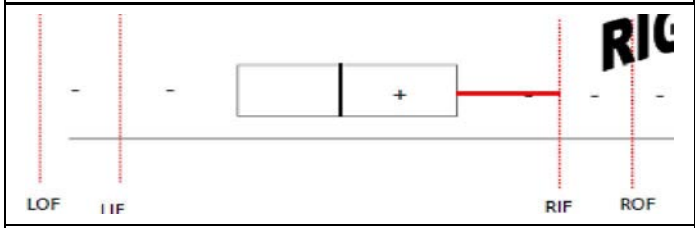
How to construct a box whisker plot?
Scale -- e.g. days
Q1 Q3
1. The first step in drawing the box whisker plot is to lay out an appropriate horizontal scale.
2. The 'box' is formed with Q1 and Q3.
3. The MEAN may be indicated inside the box with a "+" sign.
4. The MEDIAN is indicated by a line across the box

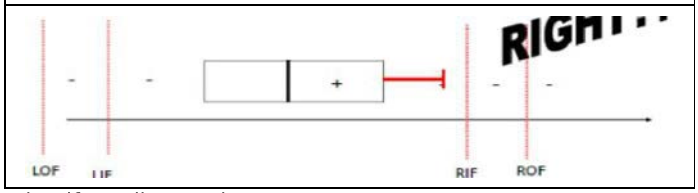


Two rules apply for the whiskers
Each whisker is as long as possible, but
• They cannot go past the inner fence
• They must end at a data point

Inner fences: RIF = Q3 + (1.5 x IQR)
LIF = Q1 – (1.5 x IQR)
Outer Fences: ROF = RIF + (1.5 x IQR)
LOF = LIF – (1.5 x IQR)



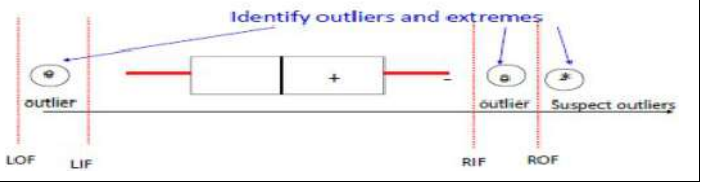




Identify outliers and extremes

OUTLIERS	SUSPECT OUTLIERS

All data values that fall between the FENCES Use 'o' symbol to indicate outliers	All data values that lie beyond the OUTER FENCES Use * to indicate extremes



1. The following sample data was obtained at 8:00 pm at a popular downtown restaurant. There were 15 tables occupied at that time.

Number of guests at the table	Food Bill for the table ($)	Liquor Bill for the table ($)
2	48.75	15.75
2	36.75	26.00
4	90.55	22.90
4	87.45	42.45
2	41.25	14.00
4	83.30	29.75
6	109.40	44.05
4	88.25	21.55
4	93.45	22.00
2	36.50	16.45
2	42.60	18.00
4	105.80	37.57
4	84.65	18.95
5	110.35	43.95
4	83.55	22.50

a) What is the 40th percentile table liquor bill?
b) What was the average food bill for the 15 tables?
c) How much did the average person spend on food?
d) Was there more relative variability in the food bill for a table or the liquor bill?
e) In order to be in the top 32% of the amount spent on food, a table would have to spend at least what amount?

2. A referendum was held on a particular issue affecting the GTA Megacity. The following table shows the results.

Municipality	Number of votes	% in Favour
City of Toronto	482,000	45
East York	152,000	63
North York	365,000	27
Etobicoke	298,000	48
Scarborough	456,000	33

a) What was the overall percent in favour of the issue?
b) Is the percent in favour discrete or continuous?

3. A company has invited its entire human resources staff from each office across the country to attend a conference at the head office in Toronto. The following information is available:

Office	Return Airfare	# of Offices	# of HR Staff per Office
Calgary	$400	2	4
Halifax	350	1	2
Montreal	330	2	3
Ottawa	300	2	3
Vancouver	500	3	5

a) What is the median return airfare per office?
b) What is the mean return airfare per person?

4. The following table shows some data regarding the top 4 chains of toy stores. For the chains and stores in this table:

Chain	# of Stores Average	Sales/Store ($ 000)
Toys 'R' Us	144	7236
Child World	79	3582
Kay bee	361	507
Lionel	56	3232

a) What is the average sales of a toy store? What is the standard deviation?
b) What is the mean sales of the 4 toy store chains?

Company/ Chain	Sales (000,000)	% Gain (Loss)	#of Stores	% Gain (Loss)
Radio Shack	1515	28	4398	7
Mervyn's	1336	26	92	15
Toys 'R' Us	1042	33	144	20
Marshall's	930	35	137	27
Saks Fifth Avenue	710	2	34	6
Lerners	682	-4	790	3
Nordstrum	613	17	36	6

c) Data regarding several retail chains is shown in the table below. What is the overall % gain in sales for the chains shown?

5. An extensive study was conducted to determine whether there are differences in the characteristics of holiday travellers that are less than 50 years old as compared to those that are more than 50 years old. One of the items of interest was the amount spent for a one-day trip. The results are shown below:

Cost of One-day Trip	Under 50 years old (n = 480)	Over 50 years old (n = 325)
$ 0 and under 50	**2**	**1**
50 " " 100	**7**	**5**
100 " " 200	**12**	**9**
200 " " 300	**20**	**113**
300 " " 400	**29**	**2**
400 " " 500	**18**	**17**
500 " " 750	**6**	**14**
750 " " 1000	**3**	**10**
1000 " " 2000	**2**	**7**
2000 " " 3000	**1**	**4**

a) For which group of travellers, if any, would the median be a better measure of central tendency than the mean?
b) What is the standard deviation of cost of a one-day trip for those under 50 yrs old?
c) For the under 50 group, 360 of those surveyed would have spent less than $.?

6.		Profit ($ billion)	Rev. per Employee ($)	Employees
1	Exxon Corp./ Mobil Corp.	11.8	96.170	122,700
2	Royal Dutch/ Shell Group	7.8	74.286	105,000
3	British Petroleum/ Amoco	4.0	70.859	56,450
4	Total SA/ Petrofina SA	2.9	67.391	69,066
5	Texaco Inc.	2.7	92.109	29,313
6	Elf Auitaine	0.96	11.469	83,700
7	ENI	3.0	37.417	80,178
8	Chevron Corp	3.3	84.615	39,362
9	PDVSA	4.8	84.818	56,592
10	SK	0.125	4.086	30,595

a) For the 10 companies shown, what is the mean profit per company? What is the standard deviation?
b) What is the overall mean revenue per employee for the ten companies shown?
c) What is the overall mean profit per employee?