QMS 210: Ogive Graph, Measures of Central Tendency and Variability

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS **CONTENT BY: ELIAS MA**

WHAT IS AN OGIVE GRAPH?

An ogive graph gives readers **cumulative relative frequencies** of the data directly. It always reaches 100% (CRF % values on the vertical axis) or sum of frequencies (CRF on the vertical axis) at the top end of the line.

Cumulative Relative Frequency (CRF)

CRF = relative frequency of the class interested + sum of frequencies from all previous classes
CRF % = (CRF ÷ total frequency) x 100%

HOW TO DRAW AN OGIVE GRAPH

1. **Sort data** in ascending order (recall calculator lesson on sorting function).
2. Construct a **frequency table**.
3. **Calculate cumulative frequency** for each class.
4. **Draw dots** assigned with **class boundary** as X-value and **CRF** as Y-value on the blank Ogive table.
5. **Link the dots** to finish the line
6. **Check**: make sure the data is in ascending data array

MEASURE OF LOCATION

P_k = k^{th} percentile
The rank of the data P_k = r = half round {[(n x k)100] + 0.5}, where n= total # of observations in the data set; k = % of observations less than or equal to P_k

Apply the Half Rounding Rule

When k < 50, round 0.25 down to 0 and 0.75 down to 0.5
When k > 50, round 0.25 up to 0.5 and 0.75 up to 1
When k = 50, the observation wanted is the median

Toronto Metropolitan University
Program Advising and Student Success
Ted Rogers School of Management, TRS 1-004
https://www.torontomu.ca/trsm-pass/
TED ROGERS SCHOOL OF MANAGEMENT | Program Advising and Student Success

QMS 210: Ogive Graph, Measures of Central Tendency and Variability

CONTENT BY: ELIAS MA

Calculating P_k

If the rank from above is an integer i, then $P_k = X_i$

If the rank from above is a fractional half i.5, then $P_k = \frac{(X_{i-0.5} + X_{i+0.5})}{2}$

NOTE: If the question is asking you to find a top % data from the graph, then k = 100% - the top% given

MEASURE OF CENTRAL TENDANCY

Calculator Symbols

$\bar{x}$ = sample mean
Σx = sum of values
σx = population standard deviation
sx = sample standard deviation

n = # of values
$minX$ = minimum value
$Q1$ = first quartile = P_{25}
Med = median/second quartile = P_{50}
$Q3$ = third quartile = P_{75}

$maxX$ = maximum value
Mod = mode(s)
Mod: n = # of mode(s)
Mod: F = frequency of mode(s)

The Mean: The mean measures the average of a set of data.

Arithmetic mean = average = $\bar{x}$

Population mean: $\mu = \frac{\Sigma x}{N}$, where Σx = sum of all data values, N = # of data items in population

Sample mean: $\bar{x} = \frac{\Sigma x}{n}$, where Σx = sum of all data values, n = # of data items in sample

Weighted Mean: used when grouped data is given and each category has a different weighted value.

NOTE: for a frequency distribution table, calculate the midpoint of each class to find the individual "x" value

$\bar{x}_w = \frac{\Sigma w \bullet x}{\Sigma w}$, x = the individual value, w = the individual value's weight (frequency)

REMEMBER: Assign weights list as the 1 Var Freq List in the Calculator

QMS 210: Ogive Graph, Measures of Central Tendency and Variability

CONTENT BY: ELIAS MA

The Median

Occupies the middle position of the data, after the set is sorted in ascending or descending order. Also known as the second quartile or 50th percentile.

The Mode

Value(s) corresponding to the highest frequency or the number that appears the most in a data set There can be one mode or a few modes.

WHICH IS A BETTER MEASURE OF CENTRAL TENDENCY: MEAN OR MEDIAN?

Steps for the 10% Rule Application

1. **Calculate the difference** between the mean and median
   Difference = mean – median

2. Calculate **10% of the smaller value**
   10% x Mean or Median, whichever is smaller

3. **Compare the difference with the 10% Rule** and conclude which is the preferred measure

4. Make decision
   - If **Difference < 10%** - the difference is lower than the 10% of the smaller value; Mean is approximately equal to median;
     Mean is a better measure
   - If **Difference > 10%** - the difference is greater than the 10% of the small value; Mean in not equal to median;
     Median is a better measure

QMS 210: Ogive Graph, Measures of Central Tendency and Variability

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS **CONTENT BY: ELIAS MA**

Shape of Distribution

Mean is a better measure | Median is a better measure: mean is not equal to median

Approximately Normal	**Skewed to the Right**	**Skewed to the Left**
Mean = median Symmetrical/Normal distribution	Mean > Median: Right Skewed Most values in the lower portion	Mean < Median: Left Skewed Most values in the upper portion

MEASURE OF VARIABILITY

Range: Measures the total spread in the data by calculating the difference between the highest and lowest value (ignoring the values in between)

Range (R) = Maximum – Minimum

Interquartile Range (IQR): The difference between the 75th and 25th percentile of a variable.

IQR = $Q3 - Q1$

Quartiles:

First Quartile = $Q1$ = 25th percentile

Second Quartile = $Q2$ = 50th percentile = median

Third Quartile = $Q3$ = 75th percentile

Variance = σ^2 (population variance) or s^2 (sample variance)

Toronto Metropolitan University

Program Advising and Student Success
Ted Rogers School of Management, TRS 1-004
https://www.torontomu.ca/trsm-pass/

TED ROGERS SCHOOL OF MANAGEMENT | Program Advising and Student Success

QMS 210: Ogive Graph, Measures of Central Tendency and Variability

CONTENT BY: ELIAS MA

Standard Deviation: The square root of variance (σ for the population, s for the sample)
NOTE: Variance and Standard deviation can be found in the calculator result

Characteristic of Measure of Variability

Dispersion	Range, IQR, Variance, Standard Dev
Spread out	Larger number
Concentrated	Smaller number

- Less variability means that it is more consistent
- 0 means no variability (cannot be lower than 0)

Coefficient of Variation: a relative measure of variability

CV for a population: - To compare the variability of two or more sets of data
$CV = \frac{\sigma}{\mu} \bullet 100\%$ - Expressed in percentage (%)
CV for a sample: - Higher the value, the more variable
$CV = \frac{s}{\bar{x}} \bullet 100\%$ - Lower CV means less risky

FIVE-NUMBER SUMMARY

Summarizes the various positions of a set of data. They are used to construct the box whisker plot

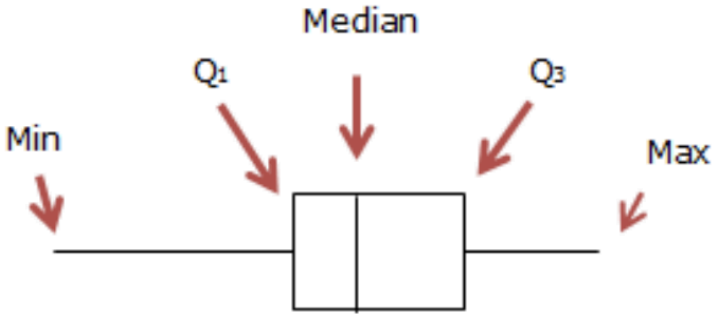
The Five Numbers:

$mixX$ | $Q1 = P_{25}$ | $Median = Q2 = P_{50}$ | $Q3 = P_{75}$ | $maxX$

QMS 210: Ogive Graph, Measures of Central Tendency and Variability

CONTENT BY: ELIAS MA

Constructing a Box-Whisker Plot



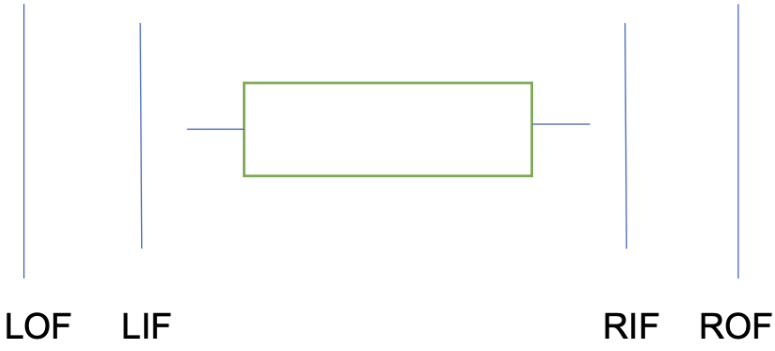
Indicate the mean with a "+"

Need to check whether maximum and minimum are outliers or suspect outliers

Determining Suspect Outliers and Outliers

Suspect outliers – lie between inner and outer fences
Outliers – lie outside the outer fences

Fences:
1. Right Inner Fences (RIF) = $Q3$ + (1.5 x IQR)
2. Right Outer Fences (ROF) = RIF + (1.5 x IQR)
3. Left Inner Fences (LIF) = $Q1$ – (1.5 x IQR)
4. Left Outer Fences (LOF) = LIF – (1.5 x IQR)



NOTE: Do not plot all the values → Only plot the suspect outliers and the outliers!

QMS 210: Ogive Graph, Measures of Central Tendency and Variability

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

CONTENT BY: ELIAS MA

Determining Whiskers

Left whisker: ends at a minimum value greater than the LIF
Right whisker: ends at a maximum value less than the RIF

Calculator Tips
1. Enter data into List 1
2. Press F1 (GRPH)

Graph Type: select Box
X List: List 1 (F1)
Frequency: 1
Outlier: ON