2017 Fall QMS102 Tip Sheet 2

(Covering Chapters 3.7 to 4.4)

Ogive graph: Gives readers cumulative relative frequencies of the data directly.

Always reaches to 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%

To draw an Ogive graph:

Sort data in ascending order (recall calculator lesson on sorting function).

Construct a frequency table.

Calculate cumulative frequency for each class.

Draw dots assigned with class boundary as X-value and crf as Y-value on the blank Ogive table. Then link all the dot to finish the line.

To read an Ogive graph: 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

To apply 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

To calculate 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 $P_k = (X_{i-0.5} + X_{i+0.5})/2$

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

Measure of Central Tendency

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: measures the average of a set of data.

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

Population Mean

$$\mu = \frac{\sum x}{N}$$

Sample Mean

$$\bar{X} = \frac{\sum x}{n}$$

Σx = sum of all data values

N = # of data items in population

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.)

$$\bar{x}_w = \frac{\sum w \cdot x}{\sum w}$$

X = the individual value

w = the individual value's weight (frequency)

(Reminder: assign weights list as the 1 Var Freq List in the Calculator)

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?

(10% Rule Application)

Step1: Calculate the difference between the mean and median.

Difference = mean – median

Step2: Calculate 10% of the smaller value

10% x Mean or Median, whichever is smaller

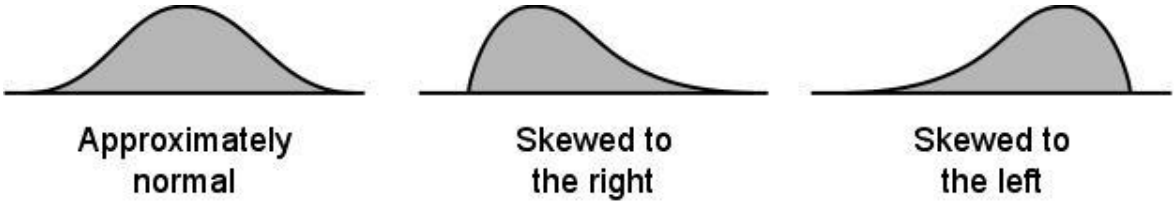
Step3: Compare the difference with the 10% Rule and conclude which is the preferred measure.

Decision**:**

Difference < 10% - The difference is lower than the 10% of the smaller value.	Mean is approximately equal to median	Mean is a better measure.

Difference > 10% - The difference is greater than the 10% of the small value.	Mean in not equal to median	Median is a better measure

Shape of Distribution



Mean is a better measure	Median is a better measure: mean is not equal to median	
Mean = median	Mean > Median: Right Skewed	Mean < Median: Left Skewed
symmetrical/normal distribution	most values in the lower portion	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 75^{th} and 25^{th} percentile of a variable. $IQR = Q_3 - Q_1$

Quartiles
- First Quartile = Q1 = 25th percentile
- Second Quartile = Q2 = 50th percentile = median
- Third Quartile = Q3 = 75th percentile

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

Standard Deviation (σ for the population, s for the sample): Square root of variance.

(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:

$$CV = \frac{\sigma}{\mu} * 100\%$$

CV for a sample:

$$CV = \frac{s}{\bar{x}} * 100\%$$

- To compare the variability of two or more sets of data
- Expressed in percentage (%)
- Higher the value, the more variable
- Lower CV means less risky

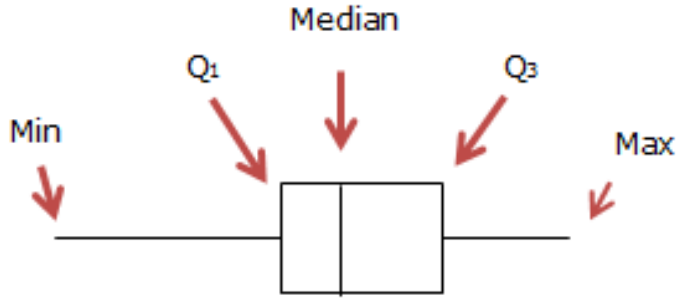
Five-Number Summary

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

The Five Numbers Are:

minX, $Q_1=P_{25}$, Median = $Q_2= P_{50}$, $Q_3= P_{75}$, maxX

Constructing a Box-Whisker Plot



Indicate the mean with a "+"

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

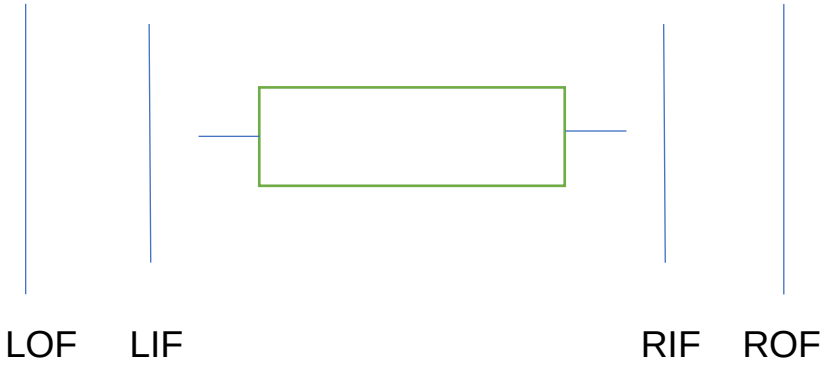
Determining Suspect Outliers and Outliers

Suspect outliers – lie between inner and outer fences

Outliers – lie outside the outer fences

Fences:

1. Right Inner Fences (RIF) = Q_3 + (1.5 x IQR)
2. Right Outer Fences (ROF) = RIF + (1.5 x IQR)
3. Left Inner Fences (LIF) = Q_1 – (1.5 x IQR)
4. Left Outer Fences (LOF) = LIF – (1.5 x IQR)



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

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