GENERAL ASSEMBLY

DATA AGGREGATION IN EXCEL

Quickly create Analytics summaries using:

Sum, min, max, avg, median, mode, CountA, CountBlanks, Countif, And Countifs

Matthew Morris

Git: Morrisdata

MatthewMorris.DA@gmail.com

GENERAL ASSEMBLY

Previously in Data Analytics

Apply; Vlookup, Hlookup, INDEX, and Match to reference data dynamically

SYNTAX AGGREGATE FUNCTIONS

Date	Sales
1-Mar	1398
2-Mar	1017
3-Mar	1736
4-Mar	280
5-Mar	370
6-Mar	404
7-Mar	2347
8-Mar	281
9-Mar	1126
10-Mar	1192
11-Mar	1185
12-Mar	304
13-Mar	2167
14-Mar	2296
15-Mar	2206
16-Mar	1171
17-Mar	1588
18-Mar	1898
19-Mar	407
20-Mar	159
21-Mar	1966
22-Mar	1773

Date		*Sales*	
Mean	40253	Mean	1210.5161
Standard Error	1.632993162	Standard Error	132.94584
Median	40253	Median	1185
Mode	#N/A	Mode	#N/A
Standard Deviation	9.092121131	Standard Deviation	740.21114
Sample Variance	82.66666667	Sample Variance	547912.52
Kurtosis	-1.2	Kurtosis	-1.4607957
Skewness	-3.1648E-17	Skewness	-0.0444947
Range	30	Range	2188
Minimum	40238	Minimum	159
Maximum	40268	Maximum	2347
Sum	1247843	Sum	37526
Count	31	Count	31

OPENING

Review of MIN, MAX, SUM, AVERAGE, COUNT, COUNTIF, COUNTA, COUNTIFS, COUNTBLANKS to summarize data sets.

A fantastic quick deliverable!

SYNTAX AGGREGATE FUNCTIONS

=MIN(number1, [number2]…)

- Finds the minimum value of a range of numbers.

=MAX(number1,[number2]…)

- Finds the maximum value of a range of numbers.

=SUM()

- Adds a range of numbers.

=AVERAGE()

- Adds all numbers in a range and divides by the number of values.

SYNTAX AGGREGATE FUNCTIONS

=COUNT(value1,[value2]…)

- Counts the number of numeric values in a range

=COUNTA(value1, [value2]…)

- Counts the number of alpha values in a range

=COUNTBLANK(range)

- Counts the number of blanks in a range

SYNTAX AGGREGATE FUNCTIONS

=COUNTIF(range,criteria…)

- Counts the number of values in the range that meet the given criteria.

=COUNTIF((H7:H10,7) this will only count if the value in the given range is 1

=COUNTIFS(range1,criteria1, [range2],criteria2,…)

- Similar to COUNTIF except it can take multiple ranges and multiple criteria.

=COUNTIF((H7:H10,7, I7:I10,5) this will only count if the value in the in the first range is 1 and the value in the second range is 5 for each instance it will count once.

SYNTAX AGGREGATE FUNCTIONS

IMPUTATION
1
2
0
4
NULL
6
7

9

INTRO TO STATISTICS

Describe, Categorize, Interpret Data

Data

Personal budgets
Events in town
Likes Dislikes
Home or Office inventories
Diet & Exercise
Etc.

Data Set

Observational Units
Variables

Dimensions
Measures

Data Set

Population
Sample

STATISTICS

Collecting Analyzing Interpreting

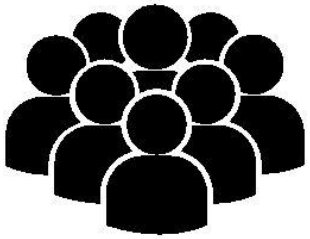
Descriptive

Know all the data

Inferential

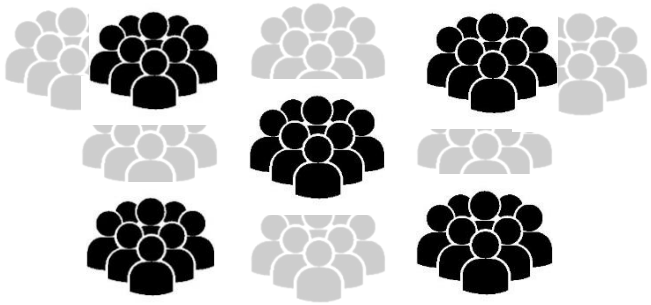
Use a Sample of Data to make guesses about all the data

Descriptive



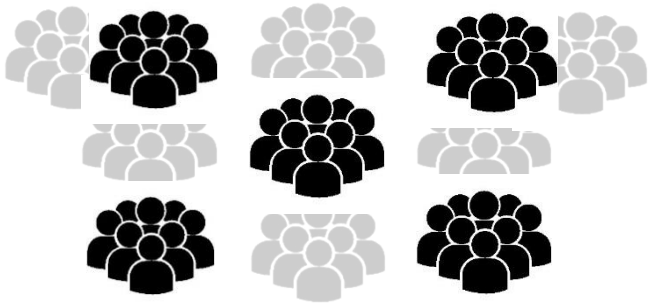
Count: 40
Karaoke: 30
Bowling: 10
Karaoke: 75%
Bowling: 25%

Inferential



Count: 210
Karaoke: 141
Bowling: 69
Karaoke: 67%
Bowling: 33%

Inferential



Count: 210
Karaoke: 141
Bowling: 69
Karaoke: 67%
Bowling: 33%

How close?

Karaoke: 67% + 3%

70% ⟷ 64%

How confident?

95%

"95% certainty that Karaoke has an approval rating among students of 67% plus or minus 3% "

Intro to Statistics

MEAN (average) add up all the numbers and divid by the number of numbers you added up. 1,3,3,5
(1+3+3+5)/4 = 12/4=3

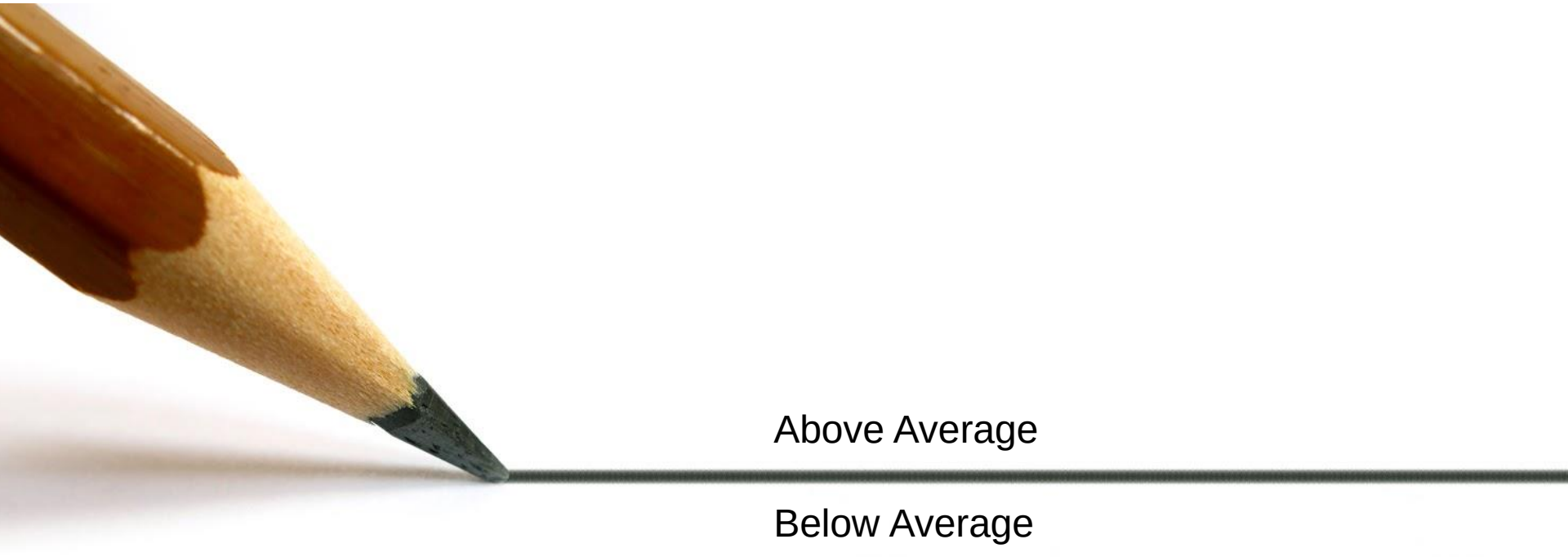
MEDIAN (middle) sort numbers highest to lowest. Then take the middle number as the Median

Odd Numbers 1,2,5,7,8
Even Numbers - Take the average of the two middle numbers 1,2,5,7,8,10
(5+7)/2= 12/2=6

MODE (most often) count the number that appears most often. If there are no repeats, there is no mode.

13, 18, 13, 14, 16, 13, 20, 15, 18, 13 The mode = 13

Intro to Statistics



Data Aggregation in Excel

Data Distribution

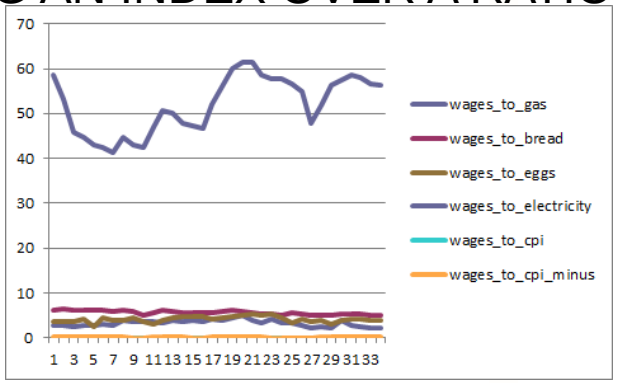
		FREQUENCY			
Score Bins	TALLY	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE FREQUENCY	RELATIVE AND CUMULATIVE FREQUENCY
90-99	IIII	4	4/20=20%	4	4/20 = 20%
80-89	~~IIII~~	5	5/20=25%	9	9/20 = 45%
70-79	~~IIII~~ I	6	6/20=30%	15	15/20 = 75%
60-69	IIII	4	4/20=20%	19	19/20 = 95%
50-59	I	1	1/20=5%	20	20/20 = 100%

Intro to Statistics

Data Aggregation in Excel

Intro to stats

CREATING AN INDEX OVER A RATIO



Data Aggregation in Excel

Q & A

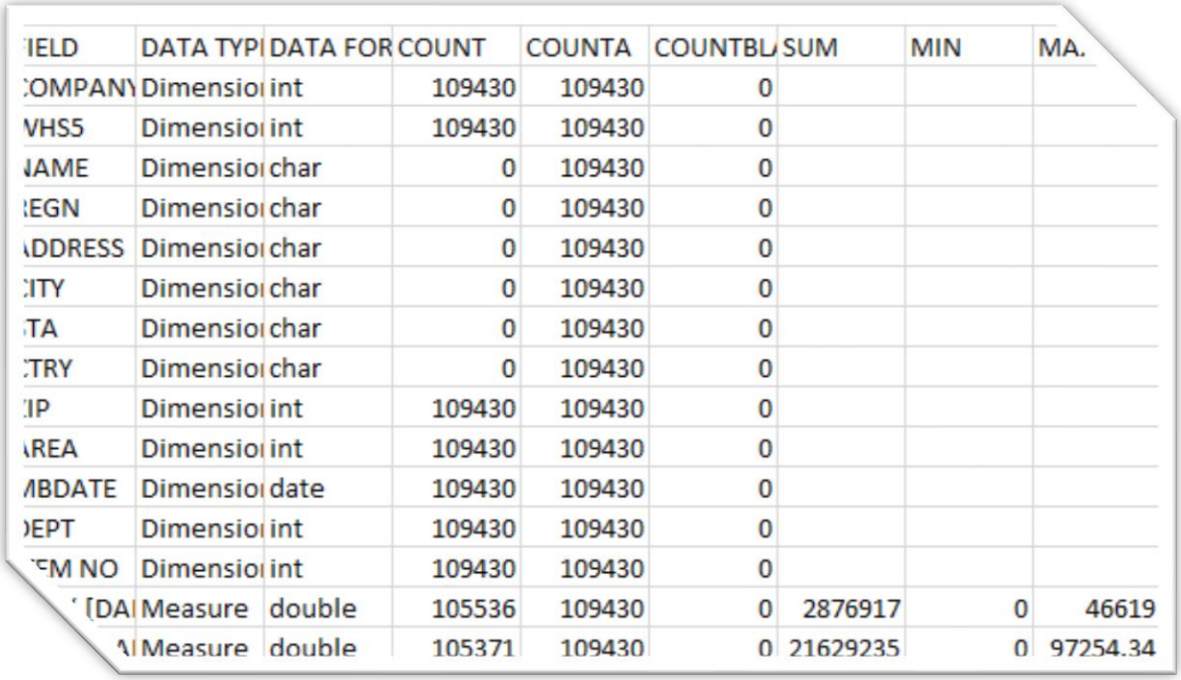
“The goal is to turn data into information, and information into insight.”

-Carly Fiorina, prior CEO of Hewlett-Packard


GA
Extra Practice

INDEPENDENT PRACTICE: USING AGGREGATE FUNCTIONS

SYNTAX AGGREGATE FUNCTIONS



IELD	DATA TYP	DATA FOR	COUNT	COUNTA	COUNTBL	SUM	MIN	MA.
OMPANY	Dimensio	int	109430	109430	0			
VHS5	Dimensio	int	109430	109430	0			
AME	Dimensio	char	0	109430	0			
EGN	Dimensio	char	0	109430	0			
DDRESS	Dimensio	char	0	109430	0			
ITY	Dimensio	char	0	109430	0			
TA	Dimensio	char	0	109430	0			
TRY	Dimensio	char	0	109430	0			
IP	Dimensio	int	109430	109430	0			
REA	Dimensio	int	109430	109430	0			
BDATE	Dimensio	date	109430	109430	0			
EPT	Dimensio	int	109430	109430	0			
M NO	Dimensio	int	109430	109430	0			
[DA	Measure	double	105536	109430	0	2876917	0	46619
A	Measure	double	105371	109430	0	21629235	0	97254.34

Before we begin the exercise; Use each of your Aggregates to build a data Analytics Summary over L3_independent_activity_p1**)**

INDEPENDENT PRACTICE

ACTIVITY: USING AGGREGATE FUNCTIONS

EXERCISE

DIRECTIONS

1. Open L3_independent_activity_p1.xlsx
2. Based on your experience, choose either the BASE or STRETCH tab to complete (20 min)

You may work with a partner, checking in with each other after answering each question.

DELIVERABLE

Complete BASE or STRETCH tab in L3_independent_activity_p1.xlsx