

Module 2 Task

Name: Obubelebara George

Deliverables:

- Module 2 assignment word file (write your explanation on this word file);
 - Module_2_assignment_Watershed.xlsx (include your answers on this excel file)
 - Master data_yourname.xlsx (using the community survey, create your own master data. On the separate tab, include your descriptive statistics for all your variables)
-

PART 1.

1.

A local government operates a small park near to its city hall. The department uses the attendance to measure its workload in determining the budget for the park. The city manager has long believed that the park has too few visitors and that the department should reach out to more customers or its budget will need to be reduced. You are an analyst for the city. The city manager has asked you to prepare an analysis for the park to determine whether the budget request for the park is justified in the city's budget proposal. You randomly selected 25 days in the past year and calculated the park attendance data: 5, 3, 10, 1, 2, 3, 4, 3, 5, 100, 4, 3, 2, 4, 25, 150, 3, 3, 5, 4, 8, 7, 10, 15, and 30.

- a. Use computer software (Excel) to calculate the mean, the median, and the mode for the data. Write a paragraph to explain the meaning of these statistics. Do you recommend the use of the mean in your presentation? Why or why not? Do you recommend the use of the median or the mode? Why or why not?**

The data on park attendance can be better understood by examining the mean, median, and mode statistics. Hence, it is needed as they are all descriptive statistics. An average depiction of attendance over the sampled days is provided by the mean attendance of 16.36 visitors each day, but this figure may be impacted by extreme values, such as days with noticeably greater attendance. Thus, even if the mean offers a gauge of central tendency, its use alone might not adequately depict the usual attendance pattern, particularly in cases when extreme values are present. However, the dataset's median attendance of 4 visitors per day, less impacted by extreme numbers, shows the dataset's middle value and provides a more accurate picture of average attendance. Furthermore, the mean of three visits per day indicates the most typical amount of attendance found in the dataset available indicating the average regular days with this attendance count. I recommend using the median, mode, and mean in the presentation to provide a comprehensive understanding of park attendance, as they offer insights into typical and frequent attendance patterns while mitigating the influence of outliers.

2. This dataset (Watershed dataset) was developed from rating variables, explanations, descriptions, and specific data incorporated within the Environmental Protection Agency's (EPA) Index of Watershed Indicators. See the data description in the attached "Watershed Data Info.docx" file.

- a. Calculate the mean and median of the percentage of samples that exceed conventional pollution standards (variable name: Conpolut) and the number of fish and wildlife advisories (variable name: Advisory). Provide the descriptive data analysis tables for both variables (fill in the table below).

	Conpolut	Advisory
Mean	20.79148148	4.796296
Standard Error	1.826701774	1.029739
Median	19.315	3
Mode	14.55	0
Standard Deviation	13.42346178	7.567008
Sample Variance	180.1893261	57.25961
Kurtosis	0.700120215	24.15405
Skewness	0.949759267	4.309038
Range	59.17	50
Minimum	0	0
Maximum	59.17	50
Sum	1122.74	259
Count	54	54

- b. Provide the descriptive statistics table for Conpolut and Advisory **if the region is coded as '3'**.

	Conpolut	Advisory
Mean	20.79148148	4.796296
Standard Error	1.826701774	1.029739
Median	19.315	3
Mode	14.55	0
Standard Deviation	13.42346178	7.567008
Sample Variance	180.1893261	57.25961
Kurtosis	0.700120215	24.15405
Skewness	0.949759267	4.309038
Range	59.17	50
Minimum	0	0
Maximum	59.17	50
Sum	1122.74	259
Count	54	54

- c. Provide the descriptive statistics table for the number of samples that exceed **both toxic pollution and conventional pollution standards by the region.**

* Note that you need to create a new variable by adding the actual percentage of samples taken exceeding allowed toxic pollutant level (Toxicon) and the actual percentage of samples taken exceeding allowed conventional pollutant level (Conpolut).

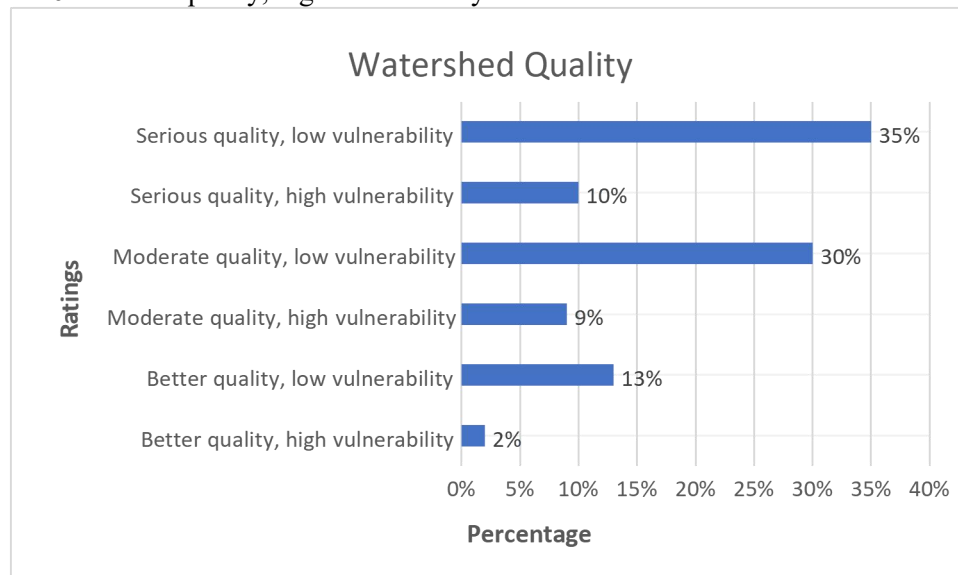
* Note that “0” value means 0% but blank means no information available. Therefore, you need to figure out how to deal with blanks in cells.

	New Variable (Toxicon + Conpolut)					
	Region 1	Region 2	Region 3	Region 4	Region 5	Total
Mean	18.69593	27.46824	23.18111	25.2375	25.69375	120.27653
Standard Error	1.698992	2.606981	1.975635	10.34557	4.564696	21.191874
Median	16.1	24.68	20.135	14.495	22.95	98.36
Mode	15.12	#N/A	15.12	15.12	15.12	#N/A
Standard Deviation	8.828219	10.74886	14.51789	35.83809	12.91091	82.843969
Sample Variance	77.93746	115.5379	210.7691	1284.369	166.6916	1855.3051
Kurtosis	2.211733	-0.66332	1.219383	9.695823	4.36248	16.826099
Skewness	1.286981	0.498313	1.102081	3.017974	1.971851	7.8772
Range	40.2	37.6	68.06	132.84	39.73	318.43
Minimum	3.8	12.2	1.97	1.76	15.12	34.85
Maximum	44	49.8	70.03	134.6	54.85	353.28
Sum	504.79	466.96	1251.78	302.85	205.55	2731.93
Count	27	17	54	12	8	118

d. **Make a bar chart** of the **quality of watersheds** (variable name=Quality).

Variable: Watersheds Quality

- 1= better quality, low vulnerability
- 2= better quality, high vulnerability
- 3= moderate quality, low vulnerability
- 4= moderate quality, high vulnerability
- 5= serious quality, low vulnerability
- 6= serious quality, high vulnerability



PART 2. Performance Report Project

Use the Community Satisfaction Data

1. Create a Master dataset.

- Based on the instructors' comments on your Module 1 Assignment Part 2 – Research Proposal, you are required to create a Master Dataset by selecting relevant variables (columns) from the dataset..
- This Master Dataset will be used for M2 through M7 assignments.
- In your Master data, you need to create a tab that includes the following items
 - Column A – ID copied and pasted from the dataset
 - From Columns B and after, copy your selection of DVs and IVs and paste them into the new Master Dataset.

2. Create descriptive statistics table for your OWN dependent variable and independent variables.

Please carefully read the instructor's comment on Module 1 Assignment and then complete this assignment.

- Based on your Master Dataset, create descriptive statistics table using Data ToolPak *Tips: Don't make changes in your master data. Instead create new tabs for new analyses.*

(Suggested table *example*)

Variable	Mean	Standard Deviation	Minimum	Maximum	N (Count)
DV (Resident Satisfaction with Snow Removal Services)	11.37562724	2.879299578	3	15	1395
IV1 (Snow Removal on Major City Streets)	4.136917563	0.810108635	1	5	1395
IV2 (Snow Removal on Neighborhood Streets)	3.698207885	1.026878906	1	5	1395
IV3 (Snow Removal from Cul De Sacs)	3.540501792	1.042312037	1	5	1395
IV5 (Race)	1.627957	1.138825	1	6	1395
IV4 (Years Lived in Naperville)	20.1957	13.10343	0	85	1395
IV6 (Gender)	1.488172	0.500039	1	2	1395

Step 3. Create a bar chart using your variables.

(Example)

