

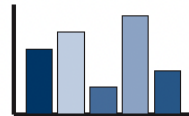
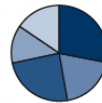
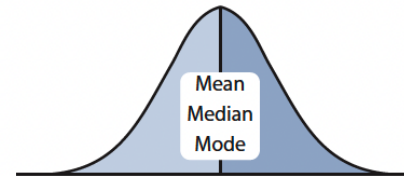
Project: Exploring Descriptive Statistics Using Microsoft Excel



Descriptive statistics

Descriptive statistics will include the following.

- Mean
- Mode
- Median
- Bar charts
- Pie charts
- Infographics
- Quartiles
- Standard deviation



Done By :
Jewel Alam

Introduction

In this project, I delved into the realm of descriptive statistics using Microsoft Excel as the primary tool. Descriptive statistics are essential in analyzing and summarizing datasets to gain insights into their characteristics. Through this hands-on exercise, I explored various statistical measures to understand the central tendency, dispersion, and distribution of data.

Project Objectives:

- The primary objectives of this project were:
- To gain practical experience in applying descriptive statistics techniques.
- To explore and utilize Microsoft Excel's functions for calculating statistical measures.
- To analyze a dataset comprehensively using a range of descriptive statistics methods.

Methods:

- Data Preparation:
 - Imported a dataset into Microsoft Excel.
 - Ensured data cleanliness by handling missing values, outliers, and inconsistencies.
- Calculating Measures of Central Tendency:
 - Utilized Excel to calculate mean, median, and mode.

- Assessing Variability and Dispersion:
 - to compute standard deviation and variance.
 - Calculated range, interquartile range (IQR), and percentiles to understand data spread.
- Exploring Data Distribution:
 - Utilized histogram tool in Excel to visualize data distribution.
 - Calculated skewness and kurtosis to understand the shape and peakedness of the distribution.
- Analyzing Relationships:
 - Calculated correlation coefficient to explore relationships between variables.
 - Constructed scatter plots to visualize correlations.
- Interpretation and Conclusion:
 - Interpreted the results obtained from descriptive statistics.
 - Drew conclusions about the characteristics and patterns present in the dataset.

Results

Result 1 : Calculation of frequency distribution table from categorical dataset

1

Categorical variables. Visualization techniques

2

Ice cream shop

3

4

Note: You may solve these problems both on paper or in a software of your choice. The medium is not crucial.

5

6

Background

There is an ice cream shop, that is operating in New York, LA and San Francisco.

7

Data

You have sold 12,327 ice creams in New York; 17,129 in LA and 19,923 in San Francisco.

8

Task

Order the data in a frequency distribution table.

9

10

Task By: Jewel Alam

11

12

13

Ice cream shop

14

15

S.No

City

Frequency of sales

16

1

New York

12,327

17

2

LA

17129.00

18

3

San Francisco

19923

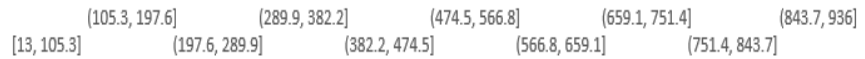
19

20

Result 2 : Formation of frequency distribution table from numerical dataset

Task :1 calculation of intervals		Task:2 Frequency distribution table			
initial number	8	Desired interval	6		
final number	282	Interval Wirdth	46		
dividing it to 6 intervals: wirth	$(282-8)/6$	Interval Starts	Interval Ends	Frequency	Relative Frequency
	45.6666667	1	46	3	0.15
	46	47	93	2	0
		94	146	3	0.15
		147	193	3	0.15
		194	240	2	0.10
		241	282	7	0.35
				20	
Task 3: calculationof intervals of exact wirtht and the frequency distribution table					
initial number	8	Interval starts	interval ends	frequency	relative frequency
final number	282	8	53.67	3	0.15
		53.68	99.35	3	0.15
dividing it into 6	45.6666667	99.36	145.02	2	0.10
		145.03	190.69	3	0.15
Interval	45.67	190.7	236.36	1	0.05
		236.37	282.03	8	0.40
				20	

7 Note: creating a histogram in Excel has some peculiarities.

[illegible]

Result 4: Calculation of Mean ,Mode , Median from the given dataset

1

Mean, median and mode

2

Average income in the United States

3

4

Background

You have a sample of 11 people and their personal annual income.

5

Task 1

Calculate the mean, median and mode

6

Task 2

Try to interpret on the numbers you got

7

8

9

10

	Annual income
11	\$ 48,000.00
12	\$ 49,000.00
13	\$ 51,000.00
14	\$ 53,000.00
15	\$ 54,330.00
16	\$ 55,000.00
17	\$ 62,000.00
18	\$ 64,000.00
19	\$ 64,000.00
20	\$ 324,000.00
21	\$ 1,264,000.00
22	

TASK BY : JEWEL ALAM

21

TASK:1

TO CALCULATE THE MEAN MEDIAN AND MODE

MEAN:	189848.18				
Median:	55,000.00				
Mode:	64.000.00				

Result 5: Obtained skewness from the given dataset

Skewness

- Background** You are given two datasets
- Task 1** Identify the skewness of dataset 1. You may use the formula from the lesson, the skewness formula in excel (=SKEW) or you can plot it on a graph
- Task 2** Identify the skewness of dataset 2. You may use the formula from the lesson, the skewness formula in excel (=SKEW) or you can plot it on a graph

TASK BY :JEWEL ALAM

Dataset 1

212
869
220
654
511
624
420
121
428
865
799
405
230
670
870
366
99
55
489
312
493
163
221
84
144
48

Task 1:

skewness 0.63

Task:1 :positive skewness

Interval	Frequency
[48, 130.2]	7
(130.2, 212.4]	4
(212.4, 294.6]	3
(294.6, 376.8]	3
(376.8, 459]	3
(459, 541.2]	3
(541.2, 623.4]	0
(623.4, 705.6]	3
(705.6, 787.8]	0
(787.8, 870]	4

Dataset 2

586
760
495
678
559
415
370
659
119
288
241
787
522
207
160
526
656
848
720
676
581
929
653
661
770
800
588

Task 2:

skewness -0.37

Task:2:Negative skewness

Interval	Frequency
[119, 206.6]	2
(206.6, 294.2]	3
(294.2, 381.8]	1
(381.8, 469.4]	1
(469.4, 557]	4
(557, 644.6]	3
(644.6, 732.2]	7
(732.2, 819.8]	4
(819.8, 907.4]	1
(907.4, 995]	4

Result 6: Obtained variance from the given dataset

A	B	C	D	E	F	G	H	I	J	K	L
1	Variance										
2	Average income in the United States										
3											
4	Background	You have the annual personal income of 11 people from the USA. You have the mean income from the exercise on mean, median and mode									
5	Task 1	Decide whether you have to use sample or population formula for the variance									
6	Task 2	Calculate the variance of their income									
7	Task 3	Generally, what does this number tell you?									
8											
9											
10											
11	Annual income	Mean	\$189,848.18								
12	\$ 62,000.00										
13	\$ 64,000.00										
14	\$ 49,000.00										
15	\$ 324,000.00										
16	\$ 1,264,000.00										
17	\$ 54,330.00										
18	\$ 64,000.00										
19	\$ 51,000.00										
20	\$ 55,000.00										
21	\$ 48,000.00										
22	\$ 53,000.00										
23	MEAN	189,848.18									
24			Sum								
25			1334334095363.66								
26											
27											

				Sample variance
	Annual income	x-mean	square of (x-mean)	sum of square of mean /(n-1)
	\$ 62,000.00	\$ (127,848.14)	16345146901	133433409536.37
	\$ 64,000.00	\$ (125,848.14)	15837754341	
	\$ 49,000.00	\$ (140,848.14)	19838198541	
	\$ 324,000.00	\$ 134,151.86	17996721541	
	\$ 1,264,000.00	\$ 1,074,151.86	1153802218341.46	
	\$ 54,330.00	\$ (135,518.14)	18365166269	
	\$ 64,000.00	\$ (125,848.14)	15837754341	
	\$ 51,000.00	\$ (138,848.14)	19278805981	
	\$ 55,000.00	\$ (134,848.14)	18184020861	
	\$ 48,000.00	\$ (141,848.14)	20120894821	
	\$ 53,000.00	\$ (136,848.14)	18727413421	
	MEAN	189,848.18		
			Sum	
			1334334095363.66	

Result 7: Obtained standard deviation and coefficient of variation from given two datasets

Standard deviation and coefficient of variation

Average income in the United States and Denmark

Background

You have the annual personal income of 11 people from the USA and 11 from Denmark. You have the mean income for USA from previous exercises

Task 1

Decide whether you have to use sample or population formula for the standard deviation and the coefficient of variation

Task 2

Calculate the standard deviation of income in the USA and in Denmark

Hint: You may start by calculating the mean and the variance

Task 3

Calculate the coefficient of variation of income in the USA and in Denmark

Task 4

Try to interpret the numbers you got

Annual income USA		Annual income Denmark		Mean US	\$	189,848.18	TASK BY :JEWEL ALAM	
4	\$ 62,000.00	462,852.37 kr.		Variance US	\$ ²	133,433,409,536.36		
5	\$ 64,000.00	470,317.73 kr.		standard deviation of		365285.381		
6	\$ 49,000.00	567,367.42 kr.		coefficient of variation	\$	1.92		
7	\$ 324,000.00	589,763.50 kr.						
8	\$ 1,264,000.00	500,179.17 kr.						
9	\$ 54,330.00	492,713.81 kr.						
10	\$ 64,000.00	515,109.89 kr.		Annual income Denma	MEAN -Xi	Square of (mean-xi)	sample variance=sum/(n-1)	standard deviation of Denmark
11	\$ 51,000.00	507,644.53 kr.	462,852.37 kr.	-42,077.48 kr.	1770514741.71	2098548471	45809.91673	
12	\$ 55,000.00	425,525.56 kr.	470,317.73 kr.	-34,612.12 kr.	1197999144.92			
13	\$ 48,000.00	522,575.25 kr.	567,367.42 kr.	62,437.57 kr.	3898449547.29		Coefficient of variation	
14	\$ 53,000.00	500,179.17 kr.	589,763.50 kr.	84,833.65 kr.	7196747726.47		0.09 kr.	
15			500,179.17 kr.	-4,750.68 kr.	22568973.22			
16			492,713.81 kr.	-12,216.04 kr.	149231683.83			
17			515,109.89 kr.	10,180.04 kr.	103633216.64			
18			507,644.53 kr.	2,714.68 kr.	7369484.16			
19			425,525.56 kr.	-79,404.29 kr.	6305041048.85			
20			522,575.25 kr.	17,645.40 kr.	311360170.67			

Result 8: Cross table and scattered plot

Cross table and scatter plot

Cross table

Background

You have employment data about country X. You have been asked to prepare a cross-table showing that data.

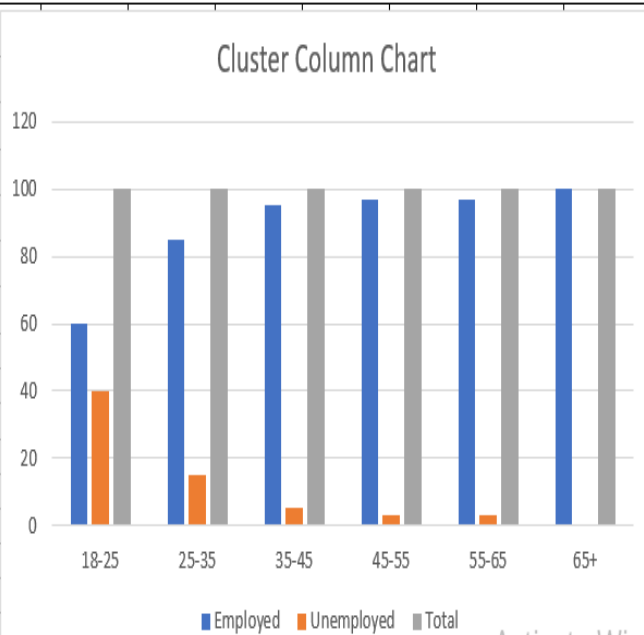
60% of 18 to 25-year-olds are employed
85% of 25 to 35-year-olds are employed
5% of 35 to 45-year-olds are unemployed
3% of 45 to 55-year-olds are unemployed
3% of 55 to 65-year-olds are unemployed
All 65+ are employed. Note: the definition of unemployed is: without a job, but actively searching for one. That's probably why all 65+s are employed.

Task 1 Create a cross table summarizing the data you have been given.

Task 2 Create a side-by-side bar chart (it is called clustered column chart in Excel), in order to visually enhance your summary.

TASK BY : JEWEL ALAM

Age group/ Employment %	Employed	Unemployed	Total
18-25	60	40	100
25-35	85	15	100
35-45	95	5	100
45-55	97	3	100
55-65	97	3	100
65+	100	0	100



Result 9: Covariance

Covariance

SAT scores

Background You are given data on the SAT reading and writing scores of several students from our lesson on cross tables and scatter plots

Task 1 Determine if this is sample or population

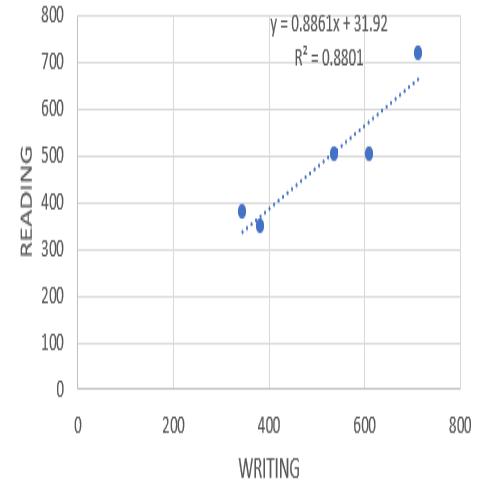
Task 2 Calculate the covariance of the two datasets

Task 3 Plot the data on scatter plot and using your previous knowledge comment on whether there is a noticeable relationship between the two variables.

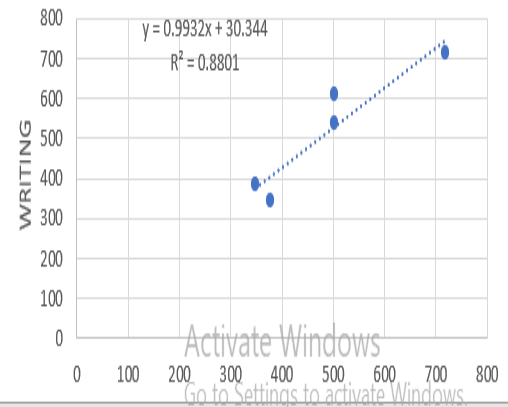
TASK BY : JEWEL ALAM

Writing	Reading								
344	378								
383	349	TASK :1 The following dataset is a sample of a population							
611	503								
713	719	Task: 2 Covariance of the two datasets:							
536	503								
		Writing	Reading	X-MEAN	Y-MEAN	(X-MEAN)(Y-MEAN)	SUM/(N-1)		
		344	378	-173.4	-112.4	19490.16	21155.55		
		383	349	-134.4	-141.4	19004.16			
		611	503	93.6	12.6	1179.36			
		713	719	195.6	228.6	44714.16			
		536	503	18.6	12.6	234.36			
		mean	mean			SUM			
		517.4	490.4			84622.2			
		Reading	Writing						
		378	344						
		349	383						

SCATTER PLOT 1



SCATTER PLOT2



Result 10: Correlation between two dataset

Correlation

SAT scores

Background You are given data on the SAT scores from the correlation exercise.

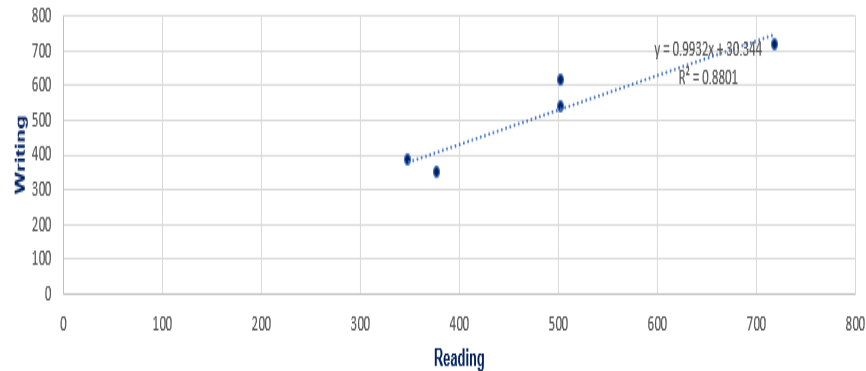
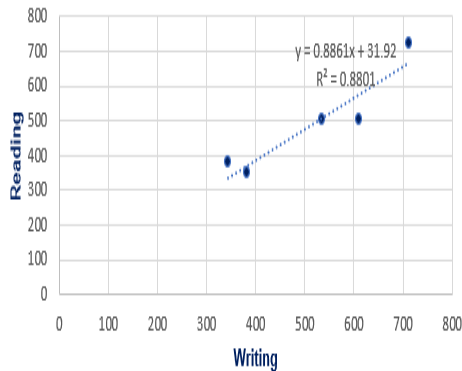
Task 1 Calculate the correlation coefficient of the two datasets.

Task 2 Comment on the strength of the correlation between the two datasets

Writing	Reading
344	378
383	349
611	503
713	719
536	503

Mean	517	490
------	-----	-----

Sum
Sample size
Cov. Sample
Correlation coefficient

$$\begin{array}{r} 19,490.16 \\ 19,004.16 \\ 1,179.36 \\ 44,714.16 \\ 234.36 \end{array}$$
[illegible]

Results:

- Through this project, I successfully completed exercises covering all aspects of descriptive statistics:
- Calculated various measures of central tendency and dispersion.
- Visualized data distributions using histograms and interpreted skewness.
- Analyzed relationships between variables through correlation analysis.
- Provided meaningful interpretations and insights into the dataset.

Conclusion:

- This project has provided invaluable hands-on experience in applying descriptive statistics techniques using Microsoft Excel. By completing exercises on every facet of descriptive statistics, I have enhanced my understanding of statistical analysis fundamentals and developed proficiency in utilizing Excel for data analysis. These skills are pivotal in various fields ranging from business analytics to scientific research, empowering me to make informed decisions based on data-driven insights.