

# Case Study

1

Data Set			
12	a)	mean	15.91667
15	b)	median	15.5
14	c)	mode	15
16	d)	range	12
18		variance	11.07639
20			
22			
19			
17			
15			
13			
10			

According to the given data set, its variance is 11.076 (which is used to understand the spread of data), and the skewness value is 0.072873.

That indicates the given data set is almost have a normal distribution. For calculating range, the maximum value = 22, Minimum value = 10.

2

Data Set			
150	a)	range	20
160	b)	variance	38.1875
165	c)	SD	6.179604
170		mean	161.25
155			
160			
162			
168			

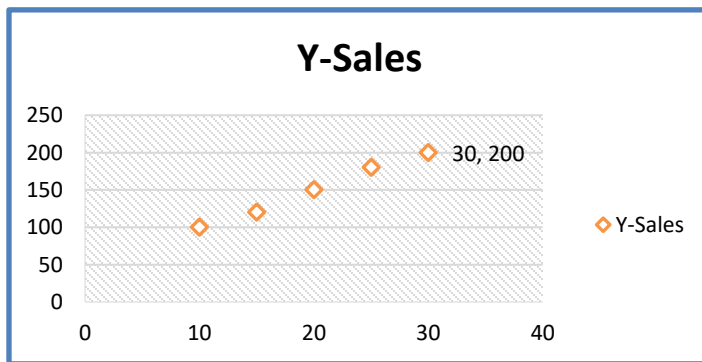
Here, the maximum value of the data set is 170, and minimum value-150. Both the variance and standard deviation (SD) are used to describe the spread of data. Variance is expressed in squared units and SD in same unit as the original value.

3

Data Set		
200	a) SD	38.43893
250	b) medium level	270
300		
220	c) variance	1477.551
270	mean	262.8571
290		
310		

4

X-Advertising Spending	Y-Sales	covariance	260		
10	100	correlation		X	Y
20	150		X	1	
15	120		Y	0.997054	1
25	180		very strong positive correlation		
30	200				



Covariance and correlation are both used to measure the relationship between two variables.

5

b) The correlation coefficient value lies in between -1 to +1, here in the case the coefficient is 1 so it has a perfect positive correlation.

c) IQR is inter quartile range, it indicate the mid spread of the value. It is calculates by finding diffence between upperquartile value and lower quartile value in a data set.

X- Study Hours	Y- Exam Score	a)	Correlation		
2	55			X	Y
4	60		X	1	
6	65		Y	1	1
8	70				
10	75		perfect positive correlation		
		c)	Quartile	Q1	4
				Q3	70
			IQR	-66	

6

Data Set	Answer a)		Answer b)		
200	First Quartile=	425	IQR	450	
300	Third Quartile=	875			
400					
500			Answer c)		
600			Whisker Points	Upper Fence	1550
700				Lower Fence	-250
800					
900					
1000					
1100					

Based on the above given results, there is no outliers in the data set. Because the data included in the upper and lower fences.

7

7			
	Data Set		
	22	Q1	28.5
	25	Q3	48.75
	28	IQR	20.25
	30	Upper Fence	58.875
	35	Lower Fence	-1.875
	40	Outliers =	100
	45		
	50	Hard Level	?
	55		
	100	Weighted Mean	?
		Mean	43

- a) in this data outlier value is 100
- b) Hard level
- c) There is no weights are provided. If we take a particular weight for each one, multiply the each data with corresponding weight and sum the weight and find weighted mean.

8

13											
	P.A	P.B	P.C	Count	Observed Fre.	Expected fre.	(o-E)^2	(o-E)^2/E	chi.squ:	0.8775	
	50	30	20	100	50	51.5625	2.44140625	0.047348485	alfa:	0.05	
	55	25	20	100	30	28.125	3.515625	0.125	n:	9	
	60	35	25	120	20	20.3125	0.09765625	0.004807692			
	Count	165	90	65	320	55	51.5625	11.81640625	0.229166667	H0: there is relation	
					25	28.125	9.765625	0.347222222	H1: There is no relation		
					20	20.3125	0.09765625	0.004807692			
					60	61.875	3.515625	0.056818182	P value	15.5073131	
					35	33.75	1.5625	0.046296296			
					25	24.375	0.390625	0.016025641			
							sum: chi.squ.value	0.877492877			

b) Here P value is greater than chi square value. So rejecting the H0. In this data there is f there is a significant difference in customer preferences among the three products.

9

14	Weight										
	70	Mean.P	70					Z score			
	65	S.L	5%	ttable value	2.26215716			-0.295540232			
	72	n	10	tcalculated	-0.93458027			-2.40654174			
	71	SD	2.368544					0.548860397			
	69	S.Mean	70.7					0.126660092			
	73	d.f(n-1):	9					0.126660092			
	72							-0.717740519			
	74							0.971060702			
	70							0.548860397			
	71							1.393261008			
								-0.295540214			
								0.126660092			

here t.cal. Is less than t.table .so, accept H0

H0:mean weight after diet program is 70

H1:mean weight after diet program is not 70

Z- Scores (X-mean)/SD

10



<i>bin</i>	<i>Frequency</i>
85	1
90	2
95	2
100	2
105	2
More	1

b)

Mean	96										
SD	7.141428429										
Normal Distri.	0.061741889										
cumulative probability up to a certain point on the normal distribution curve.											
The value 0.061741889 means that approximately 6.17% of the data lies below the given point on the curve. If this is a probability, it indicates the likelihood that a randomly selected value from the dataset will be less than the specified value you used in the function.											

12

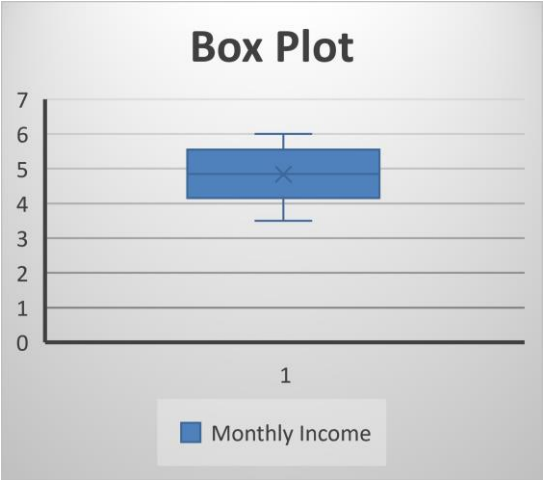
17	t Test			H0: Drug has no effect on blood pressuer( mean reduction in blood pressure is 0)			
	DataSet(xi)	mean	16	H1:Drug has effect on blood pressuer( not equal to zero)			
	12	SD	2.738613				
	14	n	10			16	
	15	n-1	9			0.866025404	
	13		a)	test statistic	18.47520861		
	16		b)	P value	2.262157163		
	18						
	17		c)	CI	95%(0.05)		
	19						
	20						

Here, p value (table value) less than the calculated value (test statistic).....so reject H0

13

a)

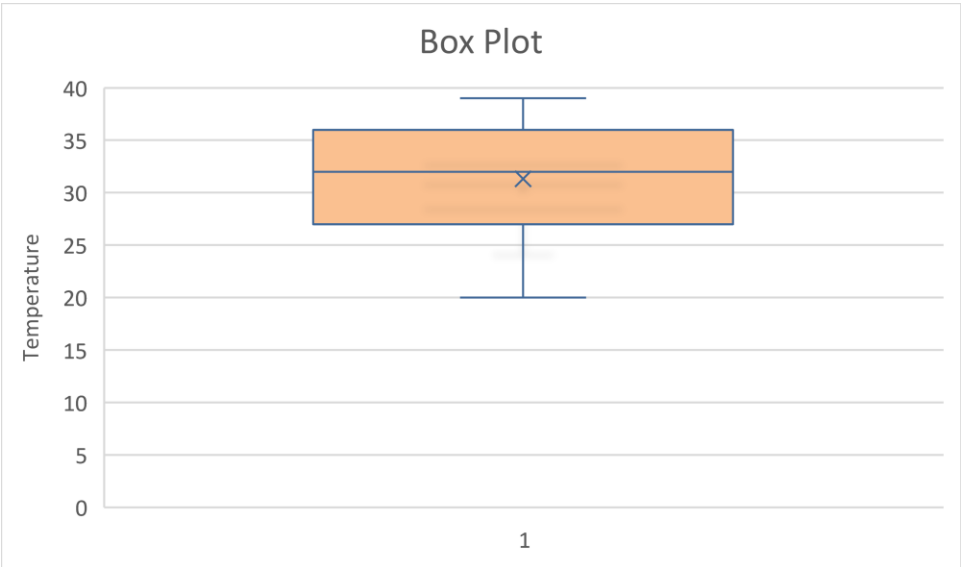
Monthly Income				
3.5		Mean	4.84	
4		Std.Dv	0.761839878	
4.2		confidence Interval	0.472184572	95%
4.5				
4.7				
5				
5.3				
5.5				
5.7				
6				



b) Box Plot

14

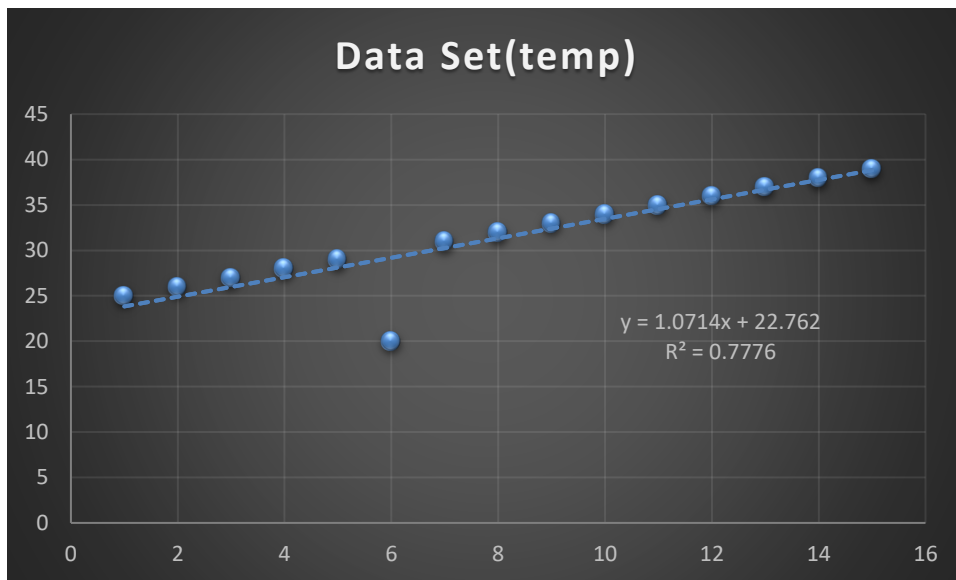
a)



b)

Q1	27.5
Q3	35.5
IQR	8
Upper Fence	47.5
Lower Fence	15.5
Outliers =	No Outliers

c)



15

