

Data Analysis using SPSS

IS2202: Engineering Mathematics IV

Department of Interdisciplinary Studies

Faculty of Engineering

University of Sri Jayewardenepura

Name : BRBSB Baruwaththa

IDX No : 16/ENG/016

Reg No : EN82645

Exercise 01

- Describe the center of the data, skewness of the data, intervals of most common values lies, and presence of outliers of Chest Injuries

Chest Stem-and-Leaf Plot

Frequency	Stem &	Leaf	
.00	3 .		
13.00	3 .	5567778999999	
31.00	4 .	00000111 222222222 3333344444444	(42 is the mode)
30.00	4 .	55556666666667777778888999999	
31.00	5 .	0000000111111222222233334444444	
31.00	5 .	55555666667777777888889999999	
17.00	6 .	00001112224444444	
7.00	6 .	5667799	
5.00	7 .	12344	

1.00 Extremes (>=97) (Outliears)

Most common values lies between 40-60

Stem width: 10

Each leaf: 1 case(s)

Descriptives

	Statistic	Std. Error
<u>Mean (Central tendency measures)</u>	<u>51.64</u>	.740
Lower Bound	50.18	
95% Confidence Interval for Mean		
Upper Bound	53.11	
5% Trimmed Mean	51.19	
<u>Median (Central tendency Measures)</u>	<u>51.00</u>	
Variance	90.982	
<u>Chest</u> Std. Deviation	9.538	
Minimum	35	
Maximum	97	
Range	62	
Interquartile Range	14	
Skewness	.891	.188
Kurtosis	2.085	.375

Skewness: (0.891/0.188) > 2 Skewed Left

2. Compute the mean and median for Chest variable. Which estimate best describe the central tendency of the data? Why?

Mean = 51.64

Median = 51

Best is Mean ∴ Mean is calculating using all data so, it is a measurement that represent whole sample

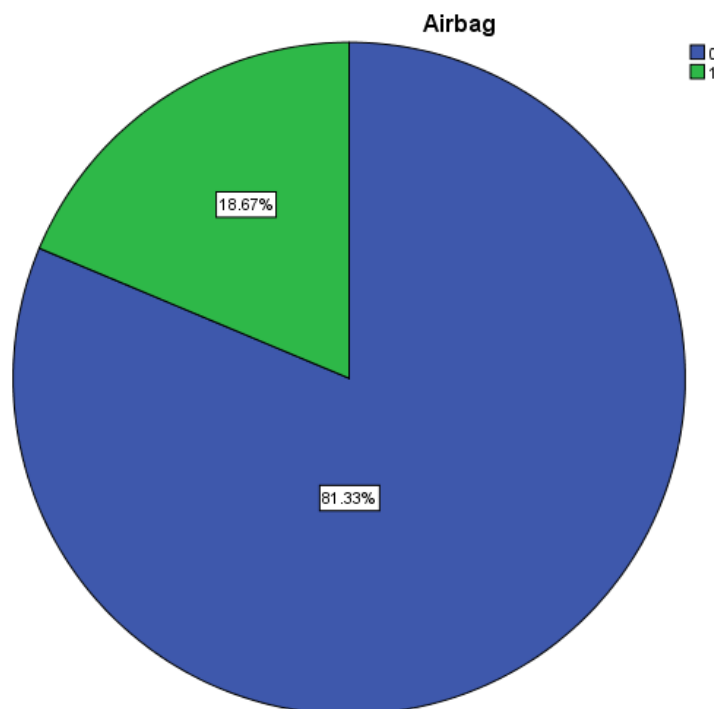
3. Select simple type of Histogram from the Charts. Then select Chest as the variable and create graph. In this histogram of the Chest variable, the program automatically selected the number of intervals. How many intervals are there?

13 intervals. 3 have empty frequencies

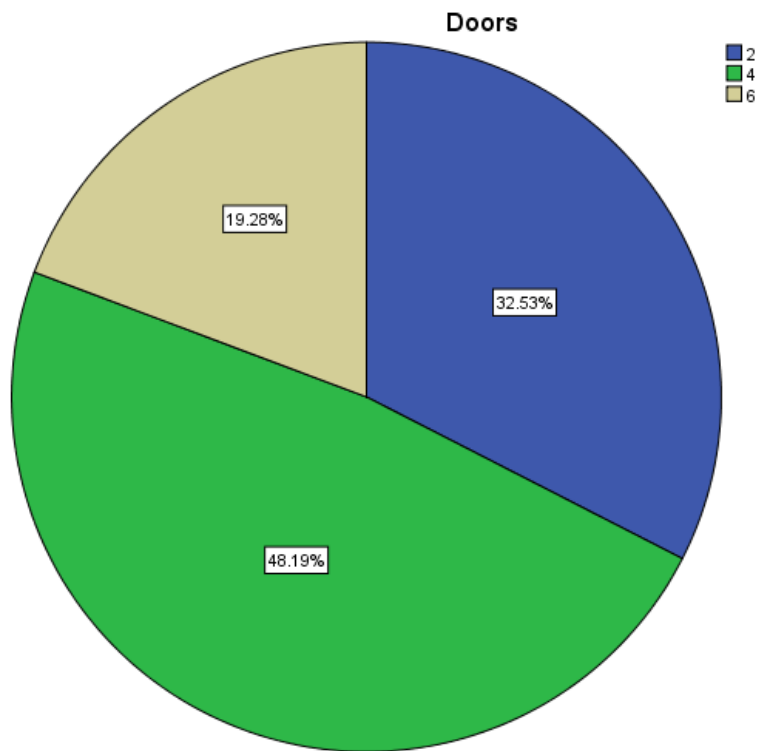
4. Experiment with changing the number of intervals. Change the number of intervals to “10” and “25” and generate Histogram for both instances. Which of the 3 Histograms describe the dataset?

25

5. Run frequencies for categorical variables: Airbag and Doors. Answer following question based on the frequency table. What percentage of vehicle has airbags? What percentage of vehicle has 2 doors?

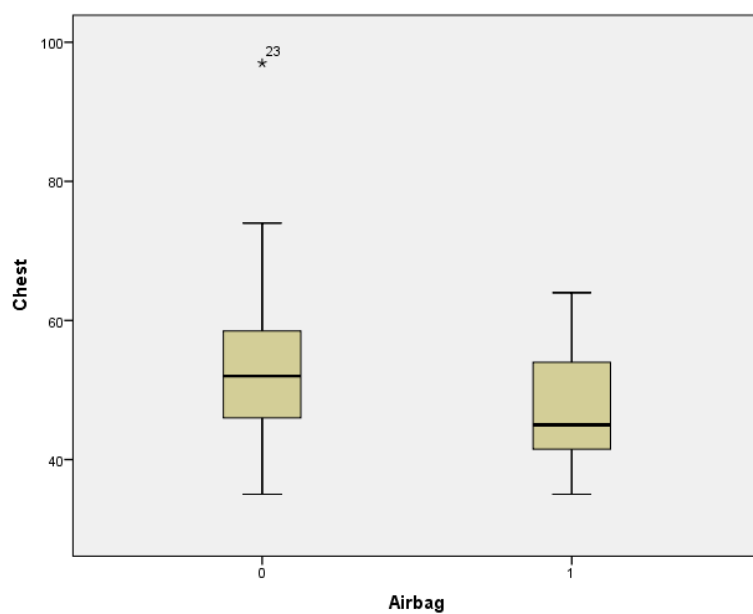


18.67% of vehicles have airbags



32.53 % have 2 doors

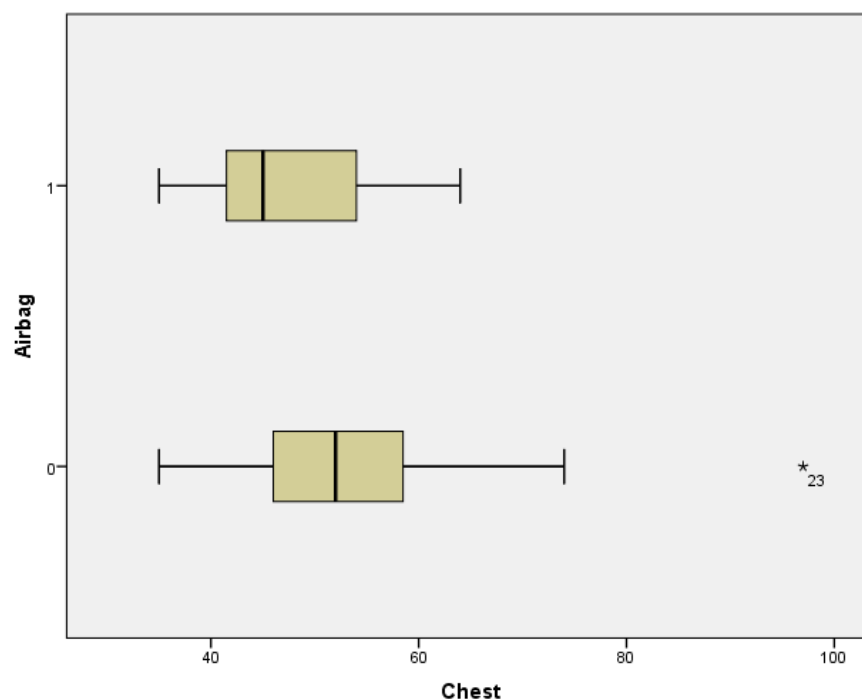
6. Suppose we wanted to see if Airbags prevented injuries. Since Airbag is a Categorical or Qualitative variable and a good form of analysis would be a side-by-side boxplot, compare the distribution of chest deceleration for Airbag (Airbag= 1) and non-Airbag (Airbag= 0) cars. Do Airbags appear to prevent injury?



The Interquartile range, mean and min/max (excluding the outlier 23rd case 97) of distribution of injuries with airbags, is lower than the same measurements of injuries without airbags

Measurement	Airbag 0	Airbag 1
Mean	52.72	46.97
Std	9.661	7.465

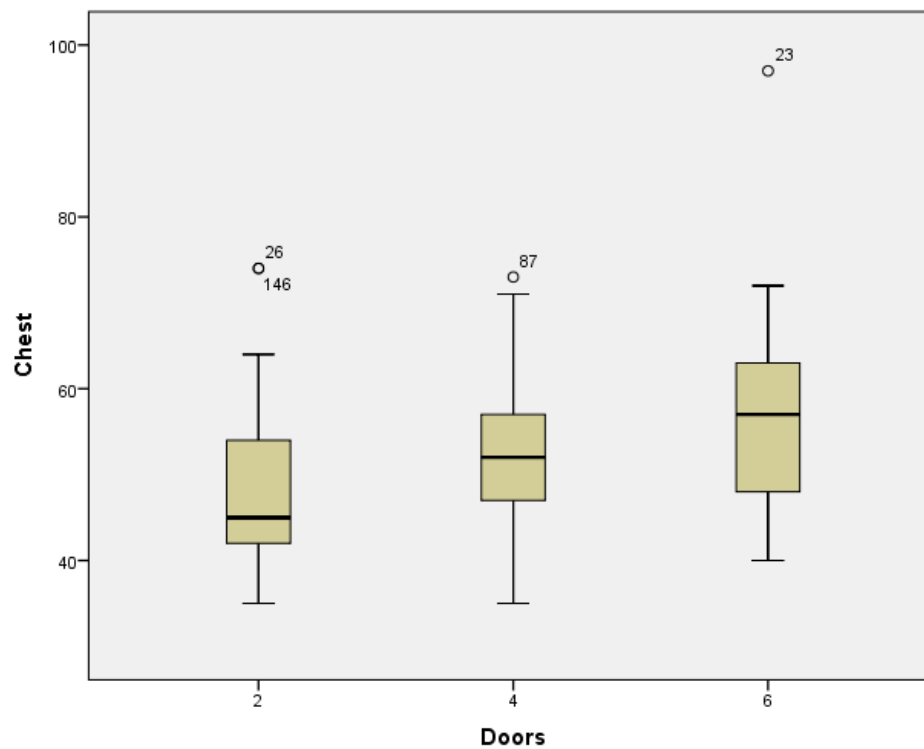
7. Is either distribution skewed? If so, describe the skewness. Are there any outliers in box plot? (A boxplot can give you information regarding the shape, variability, and center or median of a statistical data set. It is particularly useful for displaying skewed data)



Airbag 1 is Skewed Positively

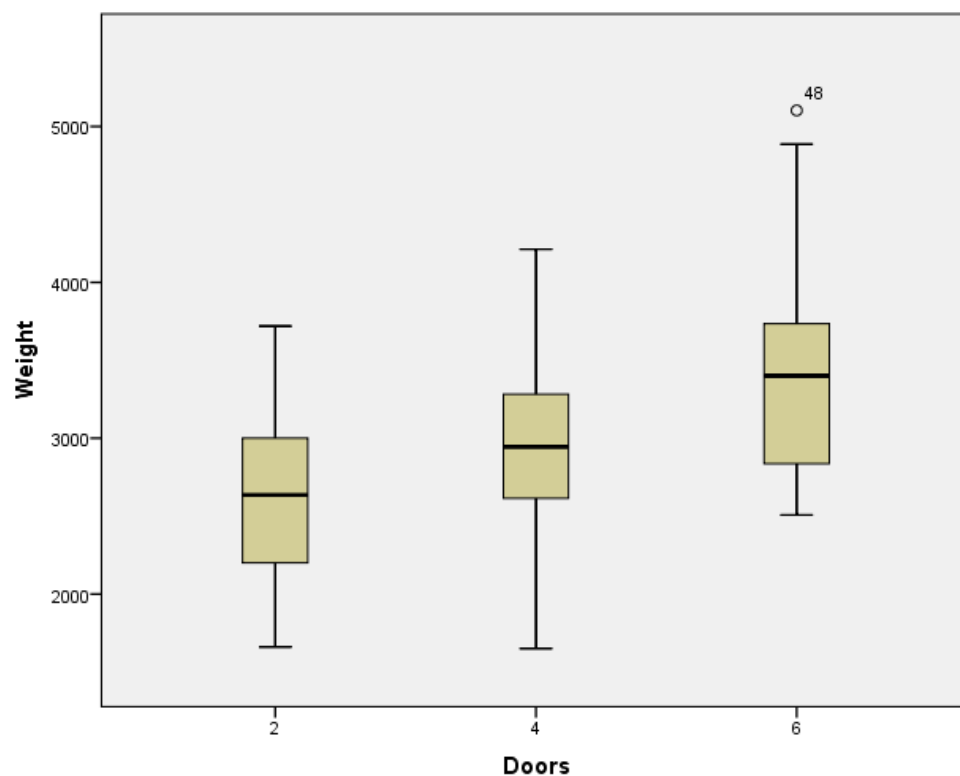
Airbag 0 also if the outliers considered

8. Compare the chest injuries of vehicles with different number of doors. Create side-byside boxplots. Does there appear to be a relationship between the number of doors and chest deceleration? Does the plot suggests a relationship between the above variables? • Which type of Vehicle (In terms of number of doors) tends to have the least severe injuries?



Mean of chest injuries increase with number of doors of the vehicle

9. Create boxplot of weight by doors. What do the results of this boxplot suggest about the results of the previous boxplot comparing Chest deceleration by number of doors?



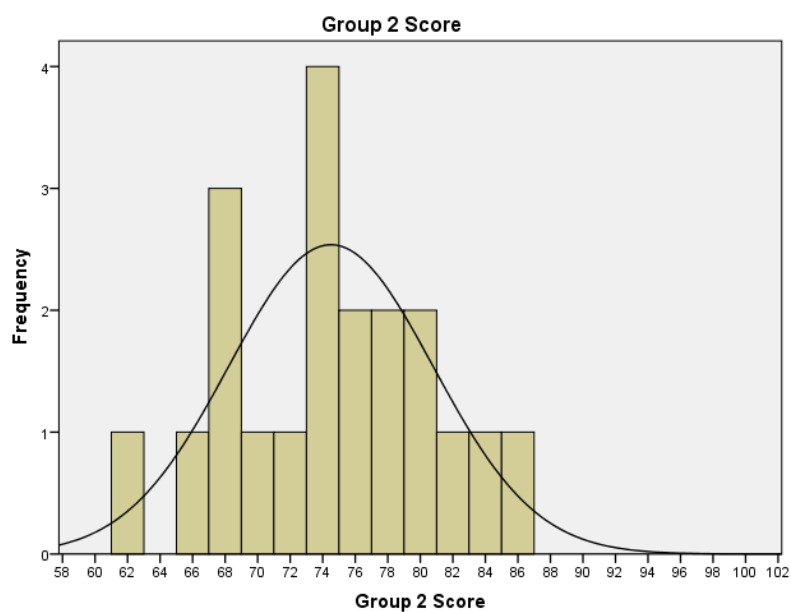
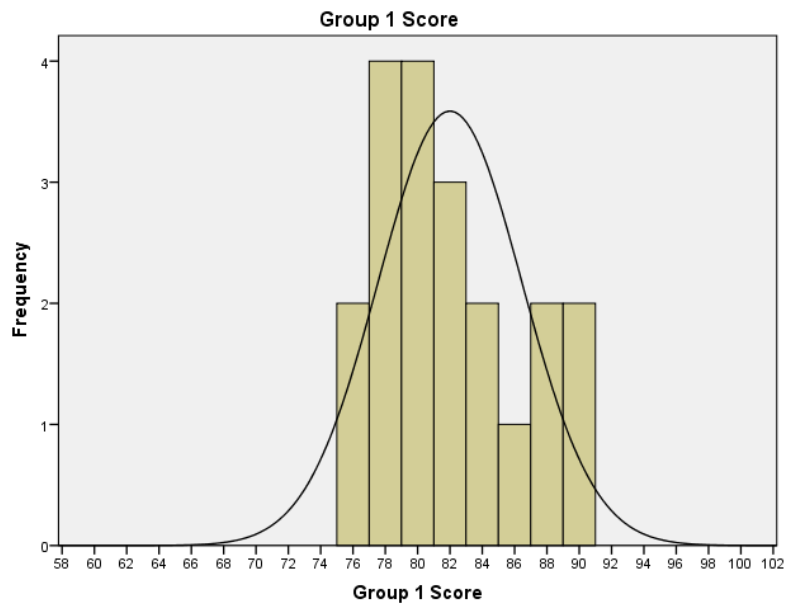
Mean of weight of the vehicle increase by the number of doors. The behavior is same as the chest injuries by doors.

Exercise 02

Q1

Statistics

		Group 1 Score	Group 2 Score
N	Valid	20	20
	Missing	0	0
Mean		82.00	74.50
Std. Deviation		4.449	6.287



Group 1	Group 2
Mean is high (82)	Mean is low 1 (74.5)
Std is low (4.449) \therefore student's marks are more packed and lower variance from the mean	Std is high (6.287) so, marks are more distributed away from mean. Overall students have different level of performances

Q2

Group 1 shows better performance and homogeneous of intelligence.

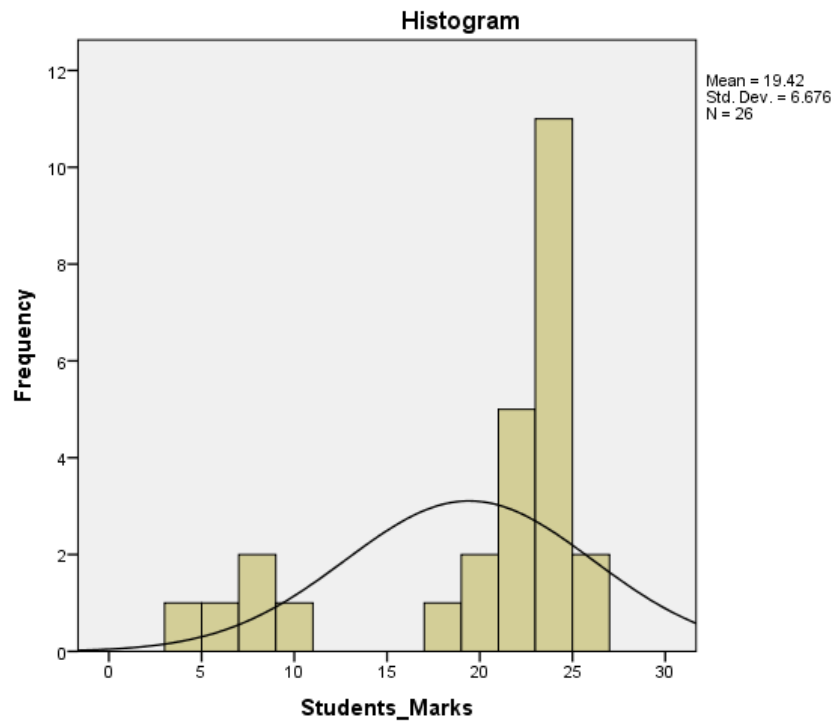
Q3

Group 1

Exercise 03

Distribution of Scores:

Shape



Descriptives

		Statistic	Std. Error
Students_Marks	Mean	19.42	1.309
	95% Confidence Interval for Lower Bound	16.73	
	Mean Upper Bound	22.12	
	5% Trimmed Mean	19.96	
	Median	22.50	
	Variance	44.574	
	Std. Deviation	6.676	
	Minimum	4	
	Maximum	25	
	Range	21	
	Interquartile Range	5	
	Skewness	-1.483	.456
	Kurtosis	.716	.887

Skewness –

$$(-1.483/0.456) < -2$$

$$\text{Mean (19.42)} < \text{median (22.5)}$$

Negatively Skewed

Center

Mean : 19.42

Median : 22.5

Mod : 23

Spread

Range : 21

Variance : 44.574

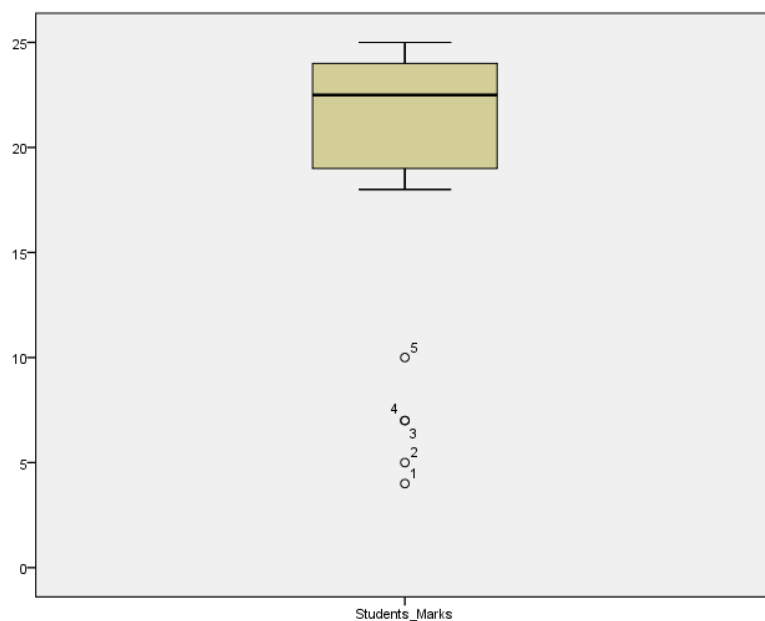
Std Deviation : 6.676

Quartiles 25 : 18.75

50 : 22.5

75 : 24

Outliers



Values <18 outliers according to quartiles range