Chapter– 15

ANALYSIS OF PETROL PRICES USING STATISTICS

Ch.Id:-ASU/GRF/EB/RAPOET/2022/Ch-15
DOI: https://doi.org/10.52458/9789391842765.2022.eb.grf.asu.ch-15

[1]NITIN SINGH RAWAT
[1]Apeejay Stya University, Sohna, Gurugram

ABSTRACT

This paper aims at analyzing the petrol prices in Delhi over a period of 5 years. Using statistical tools we will try to understand how petrol prices tend to change. This will help us uncover trends and make future assumptions. Firstly, the paper includes a brief introduction of statistics and various statistical tools. Then we will look at various applications of statistics. Then these tools are used to analyze petrol prices and obtained results are interpreted. In the end, we conclude our interpretations from the analysis.

INTRODUCTION

In modern world, we have large amounts of data. Data shapes most of the modern technologies. Many a times we are given data which we need to analyze and draw conclusions from. We might want to know about how the data is distributed, what are the extremes of the data, etc. Data itself can be very hard to understand and comprehend. If we want to understand the data, we need to draw out the essential information from the data. The field of mathematics which deals with this is called statistics. Statistics allows us to analyze quantitative data using mathematical tools and extract useful conclusions from it. Hence, it helps in understanding the data clearly.

STATISTICS

Statistics is a field of mathematics which involves collection, description, analysis, and inference of conclusions from data. Statistics deals with data using mathematical tools. Generally, the goal of statistics is to make inferences based on data.

Statistical analysis comprises of three processes: collecting data, describing data and analyzing data. We first collect raw facts and figures. Then, we describe in the data what is represented and how. Then, we analyze the data using statistical tools.

Some of the statistical tools are:

- Mean
- Median
- Mode
- Standard deviation
- Variance
- Skewness
- Kurtosis

MEAN

Mean refers to average of a given data.

- Arithmetic mean of an ungrouped data is given by dividing the sum of all observations by the total number of observations.

For example: Given heights of students in class,

167, 156, 168, 159, 162, 169, 171, 155, 172, 166

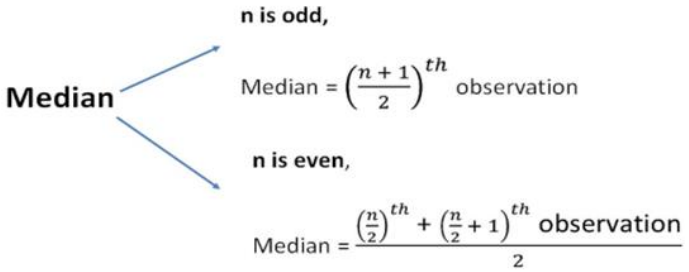
Mean Height = (167 + 156 + 168 + 159 + 162 + 169 + 171 + 155 + 172 + 166) /10= 164.5

- Arithmetic Mean of Grouped Data is given by taking sum of products of observations with their frequencies and dividing that sum by total number of observations (or sum of frequencies). If classes are given, we take mid-point of each class and use that value for calculation of mean.

MEDIAN

Median is a measure of central tendency that determines the observation in the middle after data is arranged increasingly.

- For ungrouped data, median is calculated by first arranging the given data in ascending order and then finding the middle value. If n is the number of observations, then after arranging in ascending order, median is given by:



- For grouped data, we first need to calculate cumulative frequencies for the data. Then, we need to find the median class. If n is the sum of all frequencies, then the class whose cumulative frequency is just greater than or equal to (n/2) is the median class. Then the following formula is applied to calculate the median.

$$Median = L + \left(\frac{N}{2} - C.F.\right)\frac{h}{f}$$

Where;

$L = Lower\ limit\ of\ median\ class$

$N = total\ frequency$

$C.F. = cumulative\ frequency (less\ than\ type)\ of\ the\ class\ preceding\ the\ median\ class$

$f = frequency\ of\ median\ class$

$h = width\ of\ median\ class$

MODE

Mode refers to the value that occurs most frequently in the data.

- For ungrouped data, we find mode by first finding frequency of each data value. The value with the largest frequency is the mode. There may be more than one mode.
- To find mode of grouped data, we first find the modal class. The class with the largest frequency is the modal class. The mode is calculated by using the following formula:

$$Mode = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$

Where,

l = Lower Boundary of modal class

$h = size\ of\ model\ class$

$f_m = Frequency\ corresponding\ to\ modal\ class$

$f_1 = Frequency\ preceding\ to\ modal\ class$

$f_2 = Frequency\ proceeding\ to\ modal\ class$

STANDARD DEVIATION

Standard deviation is a measure of dispersion that tells us how much the data is spread. The more dispersed the data, the higher the standard deviation or vice versa. It is calculated by taking square root of mean of squared deviations from mean of the data. Standard deviation can also be calculated directly from assumed mean method.

1. **Actual mean method:**

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum d^2}{n}}, \quad d = x - \bar{x}$$

2. **Assumed mean method:**

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}, \quad d = x - A$$

For grouped data, the assumed mean formula is modified to:

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times i, \quad \text{where } d = \frac{x - A}{i}$$

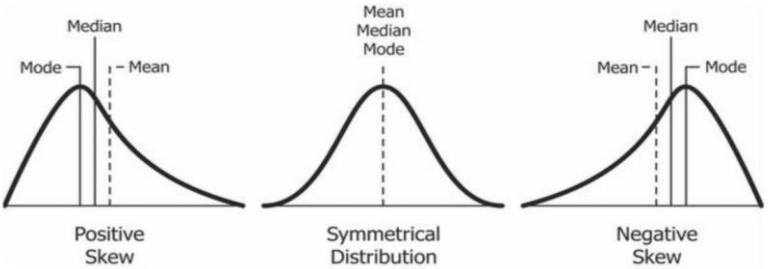
VARIANCE

Variance is the mean of squared deviations from mean. It is also a measure of dispersion and is square of standard deviation.

$$Variance(\sigma^2) = (Standard\ Deviation)^2 = \frac{\sum(x - \bar{x})^2}{N}$$

SKEWNESS

Skewness is a measure of the symmetry of the data. If we represent the data graphically, mode forms the peak of the data. If all 3 measures of central tendencies are equal, then the distribution is said to be perfectly symmetrical. If Mode < Median < Mean, then the data is said to be positively skewed. If Mean < Median < Mode, then data is said to be negatively skewed.



We can determine whether a given data is positively skewed or negatively skewed using graphical method easily. We can also measure the extent of skewness.

- Absolute Measure of skewness

$$S_k^* = \overline{X} - Mode$$

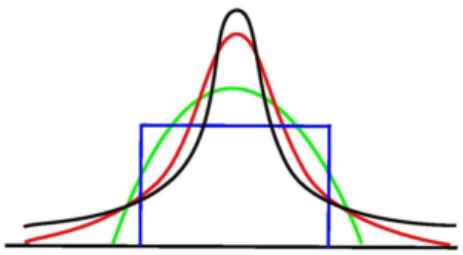
- Relative Measures of Skewness

There are four relative measures of skewness:

1. Karl-Pearson's Coefficient of Skewness
2. Bowley's Coefficient of Skewness
3. Kelly's Coefficient of Skewness
4. Measure of Skewness based on the Moments

KURTOSIS

Kurtosis is a statistical measure of peakness of distribution. It describes the degree of peakness of a distribution.



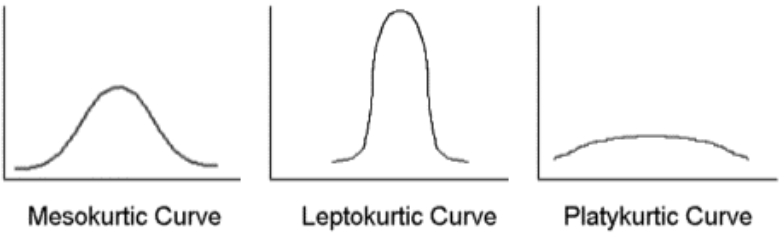
It is calculated using moments:

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2}$$

A distribution is said to be leptokurtic when kurtosis is greater than 3.

A distribution is said to be mesokurtic when kurtosis is equal to 3.

A distribution is said to be platykurtic when kurtosis is less than 3.



APPLICATIONS OF STATISTICS

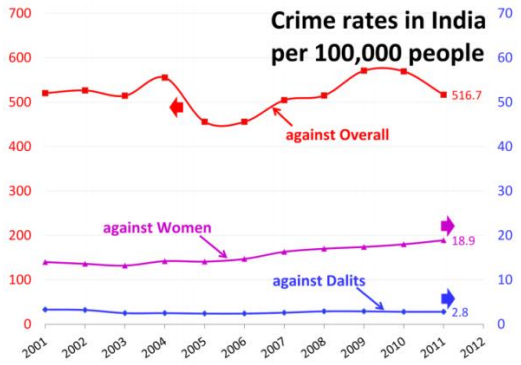
As we know, data is everywhere in today's modern world. Almost all fields have to deal with data at some point. Hence, when we have to analyze and extract information from such data, we need to resort to statistics. Hence, statistics plays a very important role in today's world and finds its application in almost all fields. Some of the fields where statistics is employed are:

- **Economics**

Economics generally deals with data like employment rates, incomes of a city/region/country, market fluctuations, stock prices, etc. To understand such economic data properly, statistics is employed. For example, we might want to analyze how wealth is distributed in a country or how employment changes with time. Statistics also helps economists to estimate future data. For example, looking at previous trends, we might want to predict future trends for a stock price.

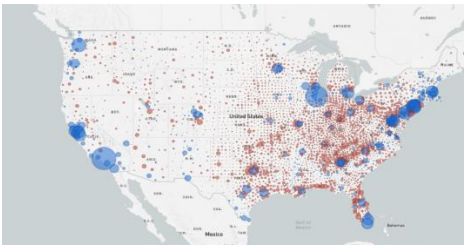
- **Sociology**

Usage of data and statistics is also very important in fields like sociology. In sociology, studies related to human population, crime rates, etc. are conducted. These researches often deal with data collected from real world. To better understand what each data represents and to get better interpretations from the data, statistics is used.



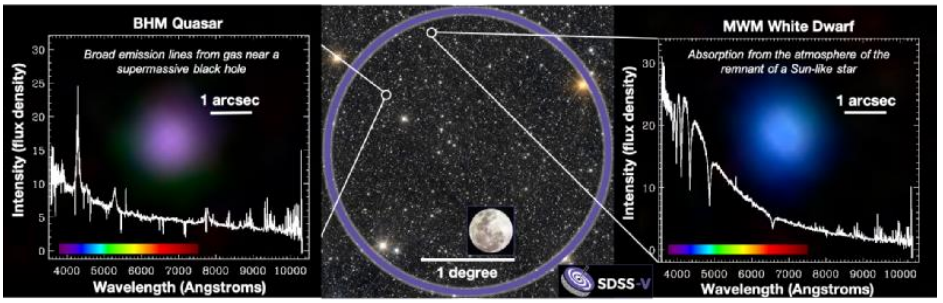
❖ **Elections**

Elections generate a massive amount of data. Such data usually tells us about number of votes for each party, number of votes from each region, number of votes from each gender/nationality, etc. To analyze and better understand the elections, statistical tools are employed. We might want to know about average number of voters from each state, which party has most number of votes, etc.



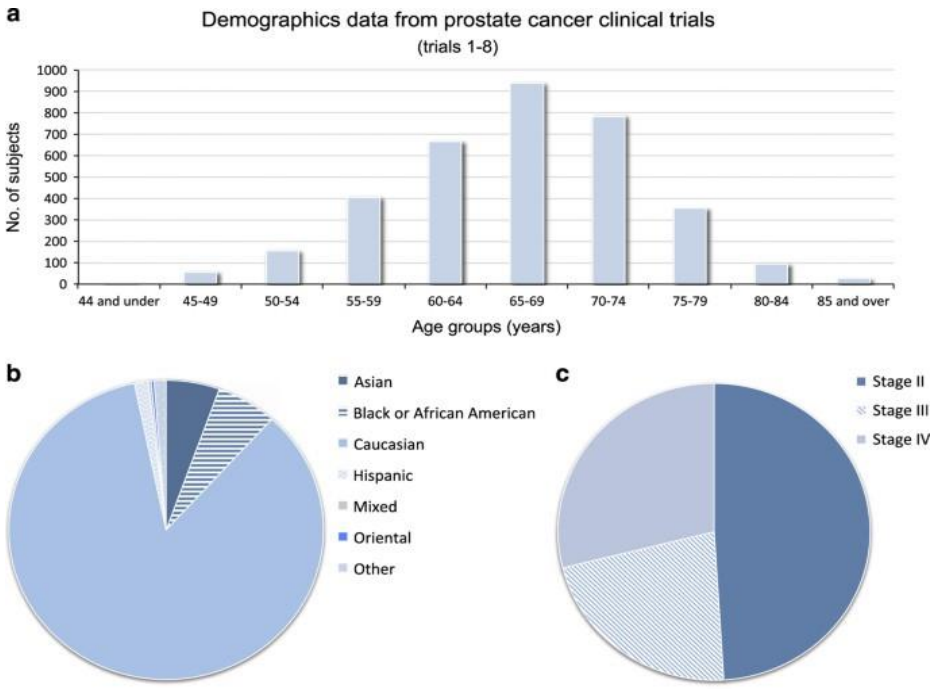
❖ **Astronomy**

The field of astronomy mostly comprises of researches. Data related to different stars, planets, their density, their distance, their luminosity, their gravity, etc. is collected on a daily basis. This data cannot be analyzed very easily. Hence, a number of statistical tools are employed to make this data easier to comprehend.



❖ **Clinical Trials**

Statistical analysis is also used in clinical trials. When a new drug or treatment is discovered, it is first tested on a group of people to check whether it is efficient and safe. The results are then observed and analyzed using different statistical tools to determine whether the drug performed well on the test subjects or not.



❖ **Designing and analyzing Surveys**

Surveys result in a collection of large amounts of data. Later, we need to analyze the data collected during the survey to find out the results. This is done by employing a number of statistical tools which makes it easy to report the findings of the survey. Statistics not only helps us in analyzing the data but it also helps in collection and organization of data.

Survey Statistics Report

Day	Offered	Accepted	Not Accepted	No Input Error	No Match Error	% Accepted
2017-09-11	23	23	0	58	17	100.00%
2017-09-12	23	23	0	62	21	100.00%
2017-09-13	28	28	0	21	9	100.00%
2017-09-14	47	44	3	17	23	93.62%
2017-09-15	42	42	0	26	62	100.00%
2017-09-18	0	0	0	3	0	0.00%
Total	163	160	3	187	132	98.16%

PETROL PRICE DATA

I have selected "Petrol Price in Delhi" as data for statistical analysis. I have taken into consideration petrol prices from 1st May 2017 to 1st June 2022. Petrol prices were considered at a period of 1 month.

This data was obtained from the site https://www.petroldieselprice.com.

DATE	PETROL PRICE
1-Jun-22	₹ 89.62
1-May-22	₹ 105.41
1-Apr-22	₹ 101.81
1-Mar-22	₹ 95.41
1-Feb-22	₹ 95.41
1-Jan-22	₹ 95.41
1-Dec-21	₹ 95.41
1-Nov-21	₹ 109.69
1-Oct-21	₹ 101.89
1-Sep-21	₹ 101.34
1-Aug-21	₹ 101.84
1-Jul-21	₹ 98.81
1-Jun-21	₹ 94.49
1-May-21	₹ 90.40

1-Apr-21	₹ 90.56
1-Mar-21	₹ 91.17
1-Feb-21	₹ 86.30
1-Jan-21	₹ 83.71
1-Dec-20	₹ 82.34
1-Nov-20	₹ 81.06
1-Oct-20	₹ 81.06
1-Sep-20	₹ 82.08
1-Aug-20	₹ 80.43
1-Jul-20	₹ 80.43
1-Jun-20	₹ 71.26
1-May-20	₹ 69.59
1-Apr-20	₹ 69.59
1-Mar-20	₹ 71.49
1-Feb-20	₹ 73.19
1-Jan-20	₹ 75.14
1-Dec-19	₹ 74.91
1-Nov-19	₹ 72.86
1-Oct-19	₹ 74.61
1-Sep-19	₹ 71.77
1-Aug-19	₹ 72.80
1-Jul-19	₹ 70.44
1-Jun-19	₹ 71.62
1-May-19	₹ 73.13
1-Apr-19	₹ 72.86
1-Mar-19	₹ 71.81
1-Feb-19	₹ 70.94
1-Jan-19	₹ 68.65
1-Dec-18	₹ 72.53
1-Nov-18	₹ 79.39
1-Oct-18	₹ 83.73
1-Sep-18	₹ 78.68

1-Aug-18	₹ 76.31
1-Jul-18	₹ 75.55
1-Jun-18	₹ 78.29
1-May-18	₹ 74.63
1-Apr-18	₹ 73.73
1-Mar-18	₹ 71.57
1-Feb-18	₹ 73.05
1-Jan-18	₹ 69.97
1-Dec-17	₹ 69.22
1-Nov-17	₹ 69.14
1-Oct-17	₹ 70.76
1-Sep-17	₹ 69.26
1-Aug-17	₹ 65.40
1-Jul-17	₹ 63.09
1-Jun-17	₹ 66.91
1-May-17	₹ 68.09

STATISTICAL ANALYSIS OF PETROL PRICES

The data was first exported to an excel spreadsheet for statistical analysis. Then, it was formatted so that it can be used for calculations. Then, various statistical measures were calculated for the given data.

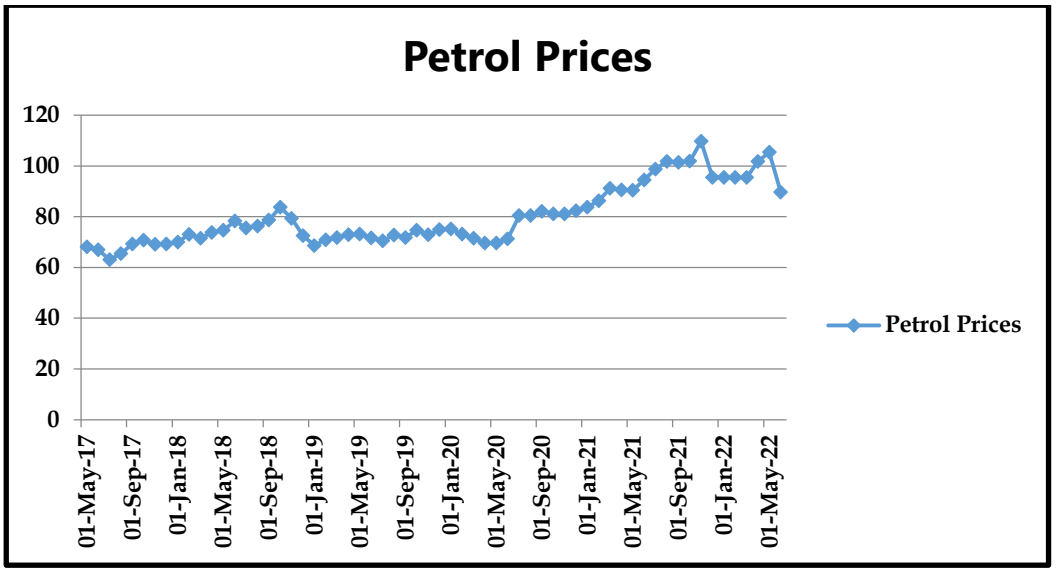
The following tables in excel was used for statistical analysis:

DATE	PETROL PRICE (RS./LIT)(x_i)	D = (x_i - MEAN)	D^2	D^4
1-Jul-17	63.09	-16.94290323	287.0619697	82404.57446
1-Aug-17	65.4	-14.63290323	214.1218568	45848.16957
1-Jun-17	66.91	-13.12290323	172.2105891	29656.48699
1-May-17	68.09	-11.94290323	142.6329375	20344.15485
1-Jan-19	68.65	-11.38290323	129.5704858	16788.5108
1-Nov-17	69.14	-10.89290323	118.6553407	14079.08987
1-Dec-17	69.22	-10.81290323	116.9188762	13670.02361
1-Sep-17	69.26	-10.77290323	116.0554439	13468.86606
1-May-20	69.59	-10.44290323	109.0542278	11892.8246

1-Apr-20	69.59	-10.44290323	109.0542278	11892.8246
1-Jan-18	69.97	-10.06290323	101.2620213	10253.99696
1-Jul-19	70.44	-9.592903226	92.0237923	8468.378349
1-Oct-17	70.76	-9.272903226	85.98673424	7393.718464
1-Feb-19	70.94	-9.092903226	82.68088907	6836.129418
1-Jun-20	71.26	-8.772903226	76.96383101	5923.431284
1-Mar-20	71.49	-8.542903226	72.98119553	5326.2549
1-Mar-18	71.57	-8.462903226	71.62073101	5129.52911
1-Jun-19	71.62	-8.412903226	70.77694069	5009.375333
1-Sep-19	71.77	-8.262903226	68.27556972	4661.55342
1-Mar-19	71.81	-8.222903226	67.61613746	4571.942045
1-Dec-18	72.53	-7.502903226	56.29355682	3168.964539
1-Aug-19	72.8	-7.232903226	52.31488907	2736.847619
1-Nov-19	72.86	-7.172903226	51.45054069	2647.158137
1-Apr-19	72.86	-7.172903226	51.45054069	2647.158137
1-Feb-18	73.05	-6.982903226	48.76093746	2377.629022
1-May-19	73.13	-6.902903226	47.65007294	2270.529452
1-Feb-20	73.19	-6.842903226	46.82532456	2192.61102
1-Apr-18	73.73	-6.302903226	39.72658907	1578.201879
1-Oct-19	74.61	-5.422903226	29.4078794	864.8233706
1-May-18	74.63	-5.402903226	29.19136327	852.1356894
1-Dec-19	74.91	-5.122903226	26.24413746	688.7547511
1-Jan-20	75.14	-4.892903226	23.94050198	573.1476349
1-Jul-18	75.55	-4.482903226	20.09642133	403.8661504
1-Aug-18	76.31	-3.722903226	13.86000843	192.0998336
1-Jun-18	78.29	-1.742903226	3.037711655	9.227692096
1-Sep-18	78.68	-1.352903226	1.830347138	3.350170647
1-Nov-18	79.39	-0.642903226	0.413324558	0.17083719
1-Aug-20	80.43	0.397096774	0.157685848	0.024864827
1-Jul-20	80.43	0.397096774	0.157685848	0.024864827
1-Nov-20	81.06	1.027096774	1.054927784	1.112872629
1-Oct-20	81.06	1.027096774	1.054927784	1.112872629

1-Sep-20	82.08	2.047096774	4.190605203	17.56117197
1-Dec-20	82.34	2.307096774	5.322695525	28.33108766
1-Jan-21	83.71	3.677096774	13.52104069	182.8185413
1-Oct-18	83.73	3.697096774	13.66852456	186.8285636
1-Feb-21	86.3	6.267096774	39.27650198	1542.643608
1-Jun-22	89.62	9.587096774	91.91242456	8447.893788
1-May-21	90.4	10.36709677	107.4766955	11551.24008
1-Apr-21	90.56	10.52709677	110.8197665	12281.02065
1-Mar-21	91.17	11.13709677	124.0349246	15384.66251
1-Jun-21	94.49	14.45709677	209.0076471	43684.19656
1-Mar-22	95.41	15.37709677	236.4551052	55911.01678
1-Feb-22	95.41	15.37709677	236.4551052	55911.01678
1-Jan-22	95.41	15.37709677	236.4551052	55911.01678
1-Dec-21	95.41	15.37709677	236.4551052	55911.01678
1-Jul-21	98.81	18.77709677	352.5793633	124312.2074
1-Sep-21	101.34	21.30709677	453.9923729	206109.0747
1-Apr-22	101.81	21.77709677	474.2419439	224905.4214
1-Aug-21	101.84	21.80709677	475.5494697	226147.2982
1-Oct-21	101.89	21.85709677	477.7326794	228228.513
1-May-22	105.41	25.37709677	643.9970407	414732.1884
1-Nov-21	109.69	29.65709677	879.5433891	773596.5733
	4962.04	**0.00**	**8273.130677**	**2861811.326**

The data was also represented graphically:



Calculation of Mean

$$MEAN = \sum \frac{x_i}{N} = \frac{4962.04}{62} = 80.03290323$$

Calculation of Median

First the petrol prices were arranged in increasing order. The total number of observations was 62 (N = 62). Since N is even, the median was the average of the middle two observations.

$$\boldsymbol{MEDIAN} = \frac{\frac{N^{th}}{2}\ observation + (\frac{N}{2}+1)^{th} observation}{2} = \frac{74.91 + 75.14}{2} = 75.025$$

Calculation of Mode

The most frequent occurring petrol price was Rs. 95.41 which occurred 4 times.

Hence, MODE = 95.41.

Calculation of Variance

To calculate variance, deviations from mean were calculated. Then, the squares of deviation were calculated and variance was calculated using formula:

$$\boldsymbol{\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N} = \frac{8273.130677}{62} = 133.4375916}$$

Calculation of Standard Deviation

Standard deviation was calculated as square root of variance.

$$\boldsymbol{\sigma = \sqrt{\sigma^2} = \sqrt{133.4375916} = 11.55151902}$$

Calculation of Skewness

- Absolute Skewness

$$S_k^* = MEAN - MODE = 80.03290323 - 95.41 = -15.37709677$$

- Karl-Pearson's coefficient of Skewness

$$S_{KP} = \frac{MEAN - MODE}{\sigma} = \frac{-15.37709677}{11.55151902} = -1.331175299$$

Calculation of Kurtosis

Kurtosis was calculated using the 4th moment.

$$\boldsymbol{\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = \frac{\frac{\sum(x_i - \overline{x})^4}{N}}{\sigma^4} = 2.592345719}$$

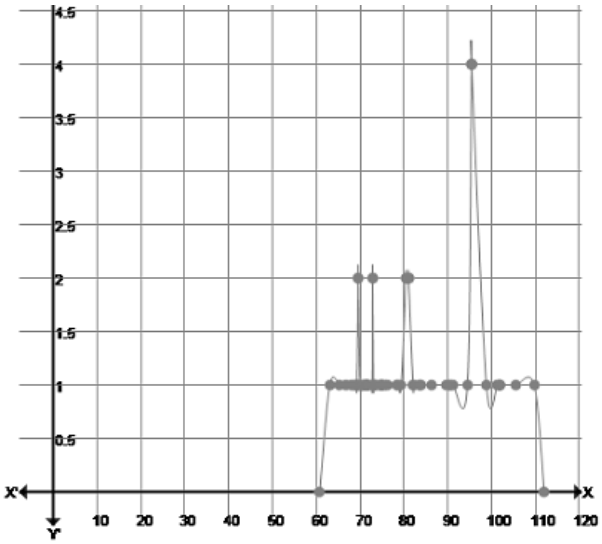
RESULT

MEAN	80.03290323
MEDIAN	75.025
MODE	95.41
VARIANCE	133.4375916
STANDARD DEVIATION	11.55151902
ABSOLUTE SKEWNESS	-15.37709677
KARL PEARSON'S COEF. OF SKEWNESS	-1.331175299
KURTOSIS	2.592345719

CONCLUSION

Statistical analysis of a data allows us to understand the whole dataset easily and also uncover trends. Petrol price is one of the major trend topics of our country. Statistical Interpretation of petrol gives us a better insight on the trends. In this report, I analyzed the petrol price data using various statistical tools. The conclusions drawn from the analysis are:

- The mean of the data was found to be around Rs. 80 per litre. It means that over the last 5 years petrol price in Delhi had an average of or revolved around Rs 80. Per litre.
- The median of data was found to be around Rs. 75.02 per litre. Hence, we can say that during this period of 5 years, half of the prices lied above Rs 75.02 per litre while half lied below.
- The mode of data was calculated as Rs. 95.41 per litre. Hence, petrol prices touched the mark of Rs. 95.41 per litre the most number of times during this time period.
- The standard deviation of the data was found to be 11.55. This value is not very high. Hence, we conclude that the data is not very dispersed and is fairly close. It does not make very large deviations. We can also check after looking at the graph that our conclusion is correct.
- Skewness of the data was found to be -1.33. Hence, this data is negatively skewed. The data has medium degree of skewness. This can also be inferred from the frequency graph below.



- Kurtosis of the data was calculated as 2.59. As 2.59 is less than 3, the given distribution is platykurtic. This means that the given distribution is less peaked than a normal distribution.

Hence from the above statistical analysis, I concluded that petrol prices over the last 5 years in Delhi were not very much varied. They typically lied between the Rs. 63 – Rs. 109 range. We can also observe that petrol prices are increasing with time. Thus, we

can infer that in future petrol prices are bound to increase further. As we already concluded that changes are not very extreme (less dispersed), we can estimate that future price hikes will also be gradual and not sudden. Hence, the data that I have chosen has been statistically analyzed completely and conclusions have been briefed above.

BIBLOGRAPHY

- *https://www.petroldieselprice.com/petrol-price-previous-historical-trend-chart-in-New-Delhi/Delhi*
- *https://atozmath.com/StatsGraph.aspx for frequency graph plotting.*
- *https://homeworkdoer.org/statistics/application-of-statistics.html*
- *https://en.wikipedia.org/wiki/List_of_fields_of_application_of_statistics*
- *https://www.kolabtree.com/blog/6-essential-applications-of-statistical-analysis*