

### Question for Statistical Functions:

#### Data Source

Q. Download data of different aspects like agriculture, labour etc. from sites like indiastat.com, rbi.org, finmin.nic.in, nssso, cso, <http://www.mospi.nic.in/> etc. websites or databanks like provis etc. Import the downloaded data into spreadsheet.

Q. Generate 100 random numbers between 20 and 100 using formulas that use RAND and RANDBETWEEN functions.

Q. Generate some random integer numbers between 30 and 70. Find out how many numbers have been generated, also compute the sum of all the generated numbers.

*Hint: TRUNC*

Q. Generate 50 random integer numbers between -200 and 200. Fix the generated numbers so that they do not change each time a change is made in the spreadsheet.

*Hint: Paste Special + Values, TRUNC*

#### MIN, MAX, LARGE

Q. Generate 50 random integer numbers between 150 and 300. Fix the generated numbers so that they do not change each time a change is made in the spreadsheet. Find out the first, second and third maximum numbers from the generated numbers. Also find the minimum number generated.

*Hint: MAX finds only the 1<sup>st</sup> maximum number, however LARGE finds the maximum number as per the ranking.*

Q. You are given with the following information about the date of births of some students. Find the dates of birth of eldest and youngest student.

01/01/2054	Sajal
22/09/2052	Ramesh
25/09/2055	Raj
02/02/2053	Akshay
11/12/2052	Sanjay
31/03/2055	Rajesh

Q. You are given with the following information about the date of births of some students belonging to two sections. Find the dates of birth of eldest and youngest student.

Section A		Section B	
01/01/2054	Sajal	01/01/2051	Hiten
22/09/2052	Ramesh	18/09/2053	Umesh
25/09/2055	Raj	28/09/2054	Priya
02/02/2053	Akshay	02/11/2050	Lina
11/12/2052	Sanjay	11/02/2053	Lindsey
31/03/2055	Rajesh	31/09/2054	Valarie

*Hint: Use max with two ranges*

Q. You are given with the following information about the marks obtained by the students of some class. Find the marks attained by the topper, second topper and third topper.

Sajal	98%	Hiten	97.50%
Ramesh	92.75%	Umesh	95.25%
Raj	98.25%	Priya	96.50%

Akshay	90%	Lina	98%
Sanjay	NA	Lindsey	91.75%
Rajesh	89.75%	Valarie	92.75%

*Hint: LARGE*

Q. Find the minimum and maximum temperature reached in the data of 10 years

Data – 36.5, 18.2, 19.5, 45, 52, 7, 9, 0.5, 3, 0.7

Q. Using the share price data sheet (Data\_SharePrice.xls), find out the highest and lowest price attained by a share.

Q. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find the worst attendance achieved by any student (after the attendances have been computed)

Q4. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find the number of students enrolled in my class by referring to the serial number column of the sheet provided that serial numbers are not given to any vacant row.

Q6. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find out how the best attendance achieved by any student (After the attendances have been computed)

#### COUNTA, COUNT, COUNTBLANK

Q7. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find out the number of students enrolled in my class by referring to the column in which the names of the student are written.

Q8. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find out the attendance of all the students for each day provided the attendance is maintained only in the form of P for present and nothing for absent.

Q9. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find the student attendance for each day if the attendance is stored as 1 for presence and nothing for absence or roll number for presence and nothing for absence.

Q10. In some survey, a particular question was to be answered by writing one of the values 1, 2, 3 or 4. However some people did not answer that question. Suppose you have the following survey data of 10 people, find out how many of them did not answer the question.

Sajal	1
Ramesh	3
Raj	5
Akshay	
Sanjay	4
Rajesh	2
Hiten	
Umesh	3
Priya	2
Lina	4
Lindsey	
Valarie	1

*Hint: COUNTBLANK*

## COUNTIF

Q. Generate and fix 150 random integer numbers between 30 and 70. Compute the frequency of each integer generated. Compute how much percentage of the entire population numbers is more than 50. Compute how many of the population numbers are between 45 and 55 inclusive.

Q11. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find the student attendance for each day if the attendance is recorded as P for presence and A for absence or 1 for presence and 0 for absence.

Q12. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find out the number of days on which the attendance has been more than 20.

*Hint: Use COUNTIF with ">20" as condition*

Q13. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find out the number of days on which the presents was more than absents.

*Hint: Find the difference between the presents and absents for each day and then use that for countif*

Q14. Using the attendance sheet (Data\_Attendance\_Sheet.xls), find out the number of days on which the presents were between 15 and 20.

*Hint: Using COUNTIF, find counts for  $\geq 15$  and find counts for  $\leq 20$  separately*

Q15. The outcomes of a player's 10 matches is recorded in MS Excel from A1 to A10 in the form of YES or NO data. Write an Excel formula that finds the loosing rate of the player.

*Hint: Use COUNTIF, COUNTA*

## Class Intervals

Q. Generate and fix 150 random integer numbers between 30 and 70. Using the formula given by Struges, create classes for grouping the population of numbers. Also represent the classes in the form of inclusive, exclusive and open ended form. Also find the mid points of each class interval.

*Hint: Struges suggested that the approximate number of classes for N data items can be calculated as:*

$$\text{no of classes intervals} = 1 + 3.322 \log_{10} N$$

The magnitude of the class interval can be computed as:

$$\begin{aligned} &\text{magnitude of a class interval} \\ &= \text{approximate integer of } \frac{(\text{maximum data item} - \text{least data item})}{\text{no of classes intervals}} \end{aligned}$$

The first class interval can be created by adding magnitude of the classes to the minimum data item. Subsequent class intervals can be computed by adding the magnitude of the class intervals to the upper limit of the previous class interval.

Q. Generate 150 random integer numbers between 30 and 70. Using the formula given by Struges about number of class intervals, create classes for grouping the population of numbers. Also represent the classes in the form of inclusive, exclusive and open ended form.

*Hint: Inclusive class intervals include both the lower and upper limits. Exclusive class intervals do not include upper class limits. Open class intervals have either the lower limit of the first class interval not defined or upper class limit of the last class interval not defined.*

## FREQUENCY

Q. Convert the following class intervals into 1) inclusive class intervals, 2) exclusive class intervals and 3) open ended class intervals. Assume that a sample data includes a list of integer numbers ranging from 1 to 100.

Class Intervals	As close ended inclusive class Interval	As open ended inclusive class interval for first interval	As open ended inclusive class interval for last interval	As open ended inclusive class interval for first and last intervals
0 - 20				
20 - 40				
40 - 60				
60 - 80				
80 - 100				

Class Intervals	As close ended exclusive class Interval	As open ended exclusive class interval for first interval	As open ended exclusive class interval for last interval	As open ended exclusive class interval for first and last intervals
0 - 20				
20 - 40				
40 - 60				
60 - 80				
80 - 100				

*Hint: The inclusive class intervals include both the lower and upper limits. The exclusive class intervals do not include the upper limits. The open ended class intervals do not mention the lower limit of the first class interval and upper limit of the last class intervals.*

Q. Generate different types of class intervals as mentioned in the previous problems, if the values represent a continuous variable instead of a discrete one.

*Hint: Use the precision value of the data values.*

Q. Generate the bin ranges that can be used the FREQUENCY function for each type of class intervals as mentioned in the previous question. Assume the data values represent a discrete variable. Use the following format

Class Intervals	Bin range

*Hint: Separate the lower and upper limits separately*

**Q18.** Prepare a spread sheet in MS EXCEL to classify 50 given numbers (varying between 1 and 100, generated at random) according to the following class intervals. Before creating frequency distribution, fix the random numbers. Solve the problem using count method and frequency functions.

Class Intervals	Frequency using COUNT	Frequency using FREQUENCY function	Cumulative Frequency

0 - 20			
20 - 40			
40 - 60			
60 - 80			
80 - 100			
<b>Total</b>			

**Q19.** Using the attendance sheet (Data\_Attendance\_Sheet.xls), in which you have already calculated the total attendance of each student in the entire month, find the number of students under each of the following attendance ranges (class intervals)

Attendance	Number of students
0% - 20%	
20% - 40%	
40% - 60%	
60% - 80%	
80% - 100%	
<b>Total</b>	

*Hint: FREQUENCY function*

Q. Prepare a spread sheet in MS EXCEL to classify 50 given numbers (varying between 1 to 150, Generated at Random) according to the following class intervals:

Class Intervals	Frequency
Less than 20	
20-40	
40-60	
60-80	
80-100	
More than 100	
<b>Total</b>	

*Hint: FREQUENCY function*

## AVERAGE, AVERAGEIF, MEDIAN, MODE

### AVERAGE

**Q20.** Using the attendance sheet, compute the average attendance of the class by using the explicit method as well as the AVERAGE function.

*Hint: The explicit method uses the formula  $\frac{\sum_{i=1}^N x_i}{N}$ , where the numerator can be computed by summing all the raw values.*

**Q21.** A person has been saving whatever he could on each day of a month. In MS Excel, he recorded savings information for each day in a column. Against each date, he recorded the amount of his savings if could save some money or else he recorded “could not save today”. At the end of the month, he wanted to know his rate of savings (i.e. average money saved per day), over a period of 31 days. Can we use AVERAGE function to solve this problem?

**Q22.** The outcomes of a player’s 10 match is recorded in MS Excel from A1 to A10 in the form of TRUE or FALSE data. Write an Excel formula that finds the losing rate of the player.

*Hint: Use AVERAGEA*

**Q24.** From the following data compute the arithmetic mean from frequency of the grouped data.

Class Intervals	Frequency
0-10	5
10-20	10
20-30	25
30-40	30
40-50	20
50-60	10

Hint: The arithmetic mean of from the frequency distribution of grouped data is obtained using the following formula  $\frac{\sum_{i=1}^k f_i m_i}{N}$ , where  $m_i$  is the middle point of  $i^{\text{th}}$  class interval.

**Q25.** From the following frequency distribution, find out the arithmetic mean for the height of trees in a garden.

Class Intervals	Frequency
Less than 7	25
Less than 14	57
Less than 21	92
Less than 28	134
Less than 36	216
Less than 42	287
Less than 49	341
Less than 56	360

Hint: Convert the cumulative class interval into successive class intervals and their successive frequencies. The arithmetic mean of from the frequency distribution of grouped data is obtained using the following formula  $\frac{\sum_{i=1}^k f_i m_i}{N}$ , where  $m_i$  is the middle point of  $i^{\text{th}}$  class interval.

**Q26.** A contractor pays Rs 100 for a male worker, Rs 80 for a female worker and Rs 30 per day for a child. If the number of employees hired by him are as follows, compute the average wage paid to a person.

Type of Employee	Wage Rate	Number of Employees
Male	100	20
Female	80	15
Children	30	5

Hint: Mean of weighted data is computed by using the formula  $\frac{\sum_{i=1}^k f_i x_i}{N}$ , where  $f_i$  is the frequency of  $x_i$  value and  $k$  is the number of values.

## AVERAGEIF

**Q23.** The outcomes of 5 players in 3 matches is recorded in MS Excel in the form of a table that contains the name of the players in the first column and outcome in the form of 1 or 0 in the second cell. Find the winning rate of each player.

Hint: Use AVERAGEIF

## MEDIAN

**Q28.** Generate and fix 150 random integer numbers between 30 and 70. Compute the median of the individual observations (ungrouped) population of generated numbers using explicit method and MEDAIN function.

*Hint: The computation of median will require all the numbers to be rearranged in the ascending order. Once arranged, the median is the  $(\frac{n+1}{2})^{\text{th}}$  when  $n$  is odd else is mean of  $(\frac{n}{2})^{\text{th}}$  +  $(\frac{n}{2} + 1)^{\text{th}}$  data.*

**Q29.** From the following data find the value of median

Income	Number of persons
100	24
150	26
80	16
200	20
250	6
180	30

*Hint: Arrange the data in ascending or descending order of income. Find cumulative frequency. Compute median as = income of  $\frac{(n+1)}{2}$  th item. Ans: item no is 61.5 and it lies in the income of 150.*

**Q.** The frequency distribution of marks obtained by 60 students of a class in the college is given below. Compute the median marks using the given procedure.

Marks	Number of students
30-34	3
35-39	5
40-44	12
45-49	18
50-54	14
55-59	6
60-64	2

Procedure:

- Convert the inclusive class intervals to exclusive class intervals by adding 0.5 in the upper limit of previous class intervals and adding 0.5 in the lower limit of the next class intervals.
- Compute the cumulative frequency of the class intervals.
- Identify the class interval in which the  $N/2^{\text{th}}$  element will lie.
- Find the lower and upper limits of the class interval in which  $N/2^{\text{th}}$  element lies.
- Compute the formula by using the following formula  $L + \frac{(\frac{N}{2} - cf_i)}{f_i} \times i$ , where
  - $L$  is the lower limit of the class interval,
  - $N$  is the total of frequency i.e.  $\sum_{i=1}^k f_i$ ,  $k$  is the number of class intervals
  - $cf_i$  is the cumulative frequency of the selected class interval,
  - $f_i$  is the frequency of the selected class interval and
  - $i$  is interval size i.e. (upper limit – lower limit) of the selected class interval

## MODE

**Q28.** Generate and fix 150 random integer numbers between 30 and 70. Compute the mode of the individual observations (ungrouped) population of generated numbers using explicit method and MODE.SING and MODE.MULT functions.

*Hint: The computation of mode requires finding out frequency of unique data values generated in the data generation process. The data value with the highest frequency is considered as the mode of the population.*

**Q.** From the following data (discrete series) of height of 100 persons in a commercial concern, determine the modal height.

Heights in inches	Number of persons
58	4

60	6
61	5
62	10
63	20
64	22
65	24
66	6
68	2
70	1

Q. Compute the value of mode from the following distribution (continuous series). Use the procedure given below.

Daily Wages	Number of workers
20-25	40
25-30	60
30-35	90
35-40	170
40-45	140
45-50	110
50-55	60
55-60	50

Procedure:

1. Make the class intervals equal if required by splitting the larger intervals into smaller ones.
2. If given class intervals are inclusive, convert the class intervals to exclusive class intervals by adding 0.5 in the upper limit of previous class intervals and adding 0.5 in the lower limit of the next class intervals.
3. Identify the class interval which has the maximum frequency.
4. Compute the mode by using the formula  $L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$ , where
  - a. L is the lower limit of the class interval,
  - b.  $\Delta_1$  is the difference between  $f_i$  and  $f_{i-1}$ , i.e. the frequency of the identified class interval and the previous class interval
  - c.  $\Delta_2$  is the difference between  $f_i$  and  $f_{i+1}$ , i.e. the frequency of the identified class interval and the next class interval
  - d.  $i$  is interval size i.e. (upper limit – lower limit) of the selected class interval

### STDEV, VAR

Q. Generate and fix 150 random integer numbers between 100 and 200. Compute the standard deviation and variance of population (individual observations) of generated numbers using explicit method and Excel function.

*Hint: The computation of standard deviation is done using the formula  $\sqrt{\frac{\sum(x_i - \bar{X})^2}{N}}$  where  $\bar{X}$  represents the mean of the observations, i.e.,  $\frac{\sum X_i}{N}$  and N is the total number of observations. The excel function that finds the standard deviation is STDEV. Variance is computed using the formula (standard deviation)<sup>2</sup>*

Q. From the following data (discrete series) of height of 100 persons in a commercial concern, determine the standard deviation and variance of height.

Heights in inches (X)	Number of persons (f)
58	4
60	6
61	5



62	10
63	20
64	22
65	24
66	6
68	2
70	1

Hint: The computation of standard deviation is done using the formula  $\sqrt{\frac{\sum f(X_i - \bar{X})^2}{N}}$  where  $f$  represents the frequency of each unique data item,  $\bar{X}$  represents the mean of the observations, i.e.,  $\frac{\sum f_i X_i}{N}$  and  $N$  is the total number of observations, i.e.,  $\sum f_i$ . Variance is computed using the formula (standard deviation)<sup>2</sup>

Q. Compute the value of standard deviation and variance from the following distribution (continuous series). Use the procedure given below.

Class	Frequency
4-8	11
8-12	13
12-16	16
16-20	14
20-24	14
24-28	9
28-32	17
32-36	6
36-40	4

Procedure:

1. Make the class intervals equal if required by splitting the larger intervals into smaller ones.
2. If given class intervals are inclusive, convert the class intervals to exclusive class intervals by adding 0.5 in the upper limit of previous class intervals and adding 0.5 in the lower limit of the next class intervals.

3. Compute the mode by using the formula  $\sqrt{\frac{\sum f_i(x_i - \bar{X})^2}{N} - \left(\frac{\sum f_i(x_i - \bar{X})}{N}\right)^2}$  where
  - a.  $x_i$  and  $f_i$  represents the midpoint and frequency of  $i^{\text{th}}$  class intervals respectively
  - b.  $\bar{X}$  represents the mean of grouped data  $\frac{\sum f_i x_i}{N}$
  - c.  $N$  represents the total population i.e.  $\sum f_i$

Q. Prepare a spread sheet in MS EXCEL to classify 50 given numbers (varying between 1 to 150, Generated at Random) according to the following class intervals. Compute the statistical parameters such as mean and standard deviation both on the basis of discrete data and above frequency distribution.

Class Intervals	Frequency
1-20	
20-40	
40-60	
60-80	
80-100	
>100	
Total	

**MetaInfo:** QNo- L1\_Q1; Type- Payroll (Level1); Time-; Keywords- Statistical Parameters (Level1); Reference-Past Practical Papers; Resource-;

### SKEW, SKEW.P, KURT

Q. Generate and fix 50 random integer numbers between 200 and 300. Compute the Persons Coefficient of skewness of sample (individual observations) of generated numbers using explicit method and Excel function.

*Hint: Excel uses SKEW function for computing Pearson's skewness of a sample. The explicit computation can be done by using the adjusted value of  $\frac{\mu_3}{\sqrt{\mu_2^3}}$  which is computed as  $\frac{n\sqrt{(n-1)} \sum(x_i-\bar{x})^2}{(n-2) (\sum(x_i-\bar{x})^3)^{3/2}}$ .*

Q. Generate and fix 50 random integer numbers between 200 and 300. Compute the Persons Coefficient of skewness of population (individual observations) of generated numbers using explicit method and Excel function.

*Hint: Excel uses SKEW.P function while the computation of Pearson's skewness is done by using the formula  $\frac{\mu_3}{\sqrt{\mu_2^3}}$  where  $\mu_3$  is called the third moment and is computed by using formula  $\frac{\sum(x_i-\bar{x})^3}{N}$  and  $\mu_2$  is called the second moment and is computed by using formula  $\frac{\sum(x_i-\bar{x})^2}{N}$ . The value of  $\frac{\mu_3}{\sqrt{\mu_2^3}}$  can be directly computed by using the formula  $\sqrt{n} \frac{\sum(x_i-\bar{x})^2}{(\sum(x_i-\bar{x})^3)^{3/2}}$*

Q. Generate and fix 50 random integer numbers between 200 and 300. Compute the sample excess kurtosis of the sample (individual observations) of generated numbers by using Excel function.

*Hint: Excel uses KURT function while the computation of kurtosis is computed by using the formula  $\frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum(x_i-\bar{x})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$  where  $n$  is the count of the observations,  $x_i$  is the  $i^{\text{th}}$  observation,  $\bar{x}$  is the mean of individual observations and  $s$  is not known.*

### CORREL

Q. Compute the correlation coefficient for the following heights (in inches) of fathers(X) and their sons (Y) using the algebraic expression  $r = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$  as well as inbuilt function of MS Excel:

X:	65	66	67	67	68	69	70	72
Y:	67	68	65	68	72	72	69	71

State whether the correlation is perfectly correlated, strongly correlated, moderately correlated, weakly correlated or uncorrelated. Also represent the correlation between X and Y variables with the help of a scatter diagram.

*Hint: The parts of expression  $r = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$  are computed as:  $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$*

$$\sigma(X) = \sqrt{E[(X - E(X))^2]} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}$$

*If  $|r| > 0.8$  then the variables are strongly correlated; if  $|r|$  is between 0.5 and 0.8, the variables are moderately correlated and when  $|r|$  is less than 0.5, the variables are weakly correlated.*

Q. A prognostic test in Mathematics was given to 10 students who were about to begin a course in Statistics. The scores (X) in their test were examined in relations to scores (Y) in the final examinations in Statistics. The following results were obtained –

$$\Sigma X = 71, \quad \Sigma Y = 70, \quad \Sigma X^2 = 555, \quad \Sigma Y^2 = 526, \quad \Sigma XY = 527$$

Find the coefficient of correlation between X and Y using explicit method as well as in-built function of MS Excel.

Q. Following table gives the number of blind per lakh of population in different age-groups. Find out the correlation between age and blindness using expanded method as well as in built function of MS Excel.

Age in years	Number of blind per lakh population
0-10	55
10-20	67
20-30	100
30-40	111
40-50	150
50-60	200
60-70	300
70-80	500

*Hint: The values for age will be taken as the midpoint of the class intervals before computing the correlation coefficient. Ans – 0.89*

Q. Following table gives the distribution of items of production and also the relatively defective items among them, according to the size-groups. Is there any correlation between size and defect in quality? Use expanded method as well as in built function of MS Excel.

Size-group	15-16	16-17	17-18	18-19	19-20	20-21
No of items	200	270	340	360	400	300
No of defective items	150	162	170	180	180	120

*Hint: Here we cannot compute the correlation between class groups and no of defectives straight way as the other things are not same in each case. Hence, we need to get X by convert the class groups as their mid points and get Y by finding the percentage of defectives. Ans – 0.94*

### SLOPE, INTERCEPT (Regression)

Q. Ten recruits were subject to a selection test to ascertain their suitability for a certain course of training. At the end of training they were given a proficiency test. The marks secured by recruits in the selection test (X) and in the proficiency test (Y) are given below:

Serial No	1	2	3	4	5	6	7	8	9	10
X	10	15	12	17	13	16	24	14	22	20
Y	30	42	45	46	33	34	40	35	39	38

Compute the regression of X on Y using the discreet data and in built functions. Show the data and the line of regression through an appropriate chart.

Q. Find the parameters of lines of regression for the following data of heights (in inches) of fathers(X) and their sons (Y) using the discreet data and in built functions. Also Show the data and the regression lines through an appropriate chart. Also compute the estimate of X for Y=70.

X:	65	66	67	67	68	69	70	72
Y:	67	68	65	68	72	72	69	71

*Ans:  $X = 0.54Y + 30.74$ ,  $68.54$*

Q. Compute the coefficient of correlation from the following data. Also obtain the equations of the lines of regression and obtain an estimate of Y which should correspond on the average to X = 6.2.

X:	1	2	3	4	5	6	7	8	9
Y:	9	8	10	12	11	13	14	16	15

Ans:  $r=0.95$ ,  $Y-12 = 0.95(X-5)$ ,  $X - 5 = 0.95(Y-12)$ , 13.14

Q. Obtain the regression of Y on X, and X on Y from the following table and estimate the blood pressure when the age is 45 years.

Age in years	Blood Pressure
56	147
42	125
72	160
36	118
63	149
47	128
55	150
49	145
38	115
42	140
68	152
60	155

Hint: Ans.  $Y = 1.138X + 80.778$ ,  $Y = 131.988$  for  $X = 45.4$

Q. Develop a spreadsheet in MS Excel to compute the standard regression estimates for the set of fifteen observations. Assume any value between 20 and 100.

Y(dependent)                      X(Independent)                      Estimated Y

Your required to find the estimated values of Y series, given that  $Y = a + b * X$ . What shall be the value of Y when the value of Y when the value of X = 70. Given an appropriate graphical representation of the regression line.

Q. You are given the output of a product for the following four quarters. I-Jan to Mar; II Apr to Jun; III Jul to Sep; IV Oct to Dec.

Year	I	II	III	IV	Total
1997	34	54	98	224	
1998	37	58	93	228	
1999	39	56	95	236	
2000	36	59	99	250	
2001	33	53	87	256	
2002	37	58	96	256	
2003	38	50	92	254	
2004	37	55	94	290	

Develop a spreadsheet to conduct to conduct the trend series analysis for each year, by utilizing the standard technique of least square regression. Compute the estimated output for the year 2005 and also depict the actual and estimated output using a suitable graph.

Q. Estimate the age of husband whose wife's age is 31 years. Estimate the age of wife whose husband is 40 years old.

Age of Husband	Age of Wife				
	15-25	25-35	35-45	45-55	55-65

15-30	30	6	3	-	-
30-45	18	32	15	12	8
45-60	2	28	40	16	9
60-75	-	4	9	10	8

### Combined

Q. Generate and fix 150 random integer numbers between 30 and 70. Compute the mean, median and mode of the ungrouped population of generated numbers using explicit method. Also compute if the given population data is single modal or multi modal.

*Hint: The computation of median will require all the numbers to be rearranged in the ascending order. Once arranged, the median is the  $(\frac{n+1}{2})^{th}$  when  $n$  is odd else is  $(\frac{n}{2})^{th} + (\frac{n}{2} + 1)^{th}$  data. The computation of mode will require finding out frequency of unique data values generated in the data generation process. The data value with the highest frequency is considered as the mode of the population.*

Q. Generate 150 random integer numbers between 30 and 70. Compute the first four statistical moments of the given population of numbers.

*Hint: the first, second, third and fourth moments are mean, variance, skewness, and kurtosis.*

**Q1.** From attendance sheet find the information about the following questions:

1. How many students are there in this class
2. How many classes were conducted in the entire month
3. How many days classes were conducted in the entire month
4. Find out the total attendance scored by each student
5. Find out the total students present in the class each day
6. Find out the total number of students absent from class each day
7. What was the minimum attendances by any student
8. What was the maximum attendance achieved in the class over the month
9. How many students usually normally come to class each day
10. Find out how many students have got their attendance more than 85%
11. Find out if how many cells are filled with invalid data

Q. You did some survey and got survey data in Excel by some third party. Before doing some data analysis, you want to ensure, if the data is entered properly or not. Suggest how you would ensure that data entry operator has not done the following kinds of mistakes.

- a. If data has been missed in some cells
- b. If some names have been entered in numeric cells
- c. If the value of the cells is not in the required range (e.g. permitted values are 1, 2, 3 and 4 only)

*Hint: COUNTBLANK for blank cells, COUNTA for non-numeric data in places where numeric data, COUNTIF for values outside the limits*

Q. Generate 150 random integer numbers between 30 and 70. Using the Struges law, create classes for grouping the population of numbers. Also represent the classes in the form of inclusive, exclusive and open ended form. Compute the statistical parameters of mean, median, mode, standard deviation from the grouped data.

Q. Prepare a spread sheet in MS EXCEL to classify 50 given numbers (varying between 1 to 100, Generated at Random) according to the following class intervals. Before creating frequency distribution, fix the random numbers.

Class Intervals	Frequency
0 - 20	
20 - 40	
40 - 60	
60 - 80	
80 and Above	
	-----
Total	
	-----

Required:

- Prepare a pie chart for the above frequency distribution
- Compute the statistical parameters such as mean and standard deviation both on the basis of discrete data and above frequency distribution.

Q. Prepare a spread sheet in MS EXCEL to classify 50 numbers (varying between 100 to 500) that have been generated at random according to the following class intervals:

Class Intervals	Frequency
100 - 200	
200 - 400	
400 - 600	
600 - 800	
800 - 900	
800 and Above	

Required:

- Prepare a pie chart for the above frequency distribution
- Compute the statistical parameters such as mean and standard deviation both on the basis of discrete data and above frequency distribution.

Q. 14 Prepare a spread sheet in MS EXCEL to classify 80 given numbers (varying between 1 to 150, Generated at Random) according to the following class intervals:

Class Intervals	Frequency
0 - 20	
20 - 40	
40 - 60	
60 - 80	
80 - 100	
100 - 120	
120 and Above	
	-----
total	
	-----

Required:

- Prepare a pie chart for the above frequency distribution

- Compute the statistical parameters such as mean and standard deviation both on the basis of discrete data and above frequency distribution..