NITESH
nitesh7@s.amity.edu

AMITY UNIVERSITY
UTTAR PRADESH

DATA SCIENCE LAB PRACTICAL FILE

Submitted to -- **Dr. Aanchal Anant Awasthi**

NAME: Nitesh

PROGRAMME TITLE: M.Sc. (Data Science)

SEMESTER: 1

SUBJECT CODE: STAT671

ENROLLMENT NUMBER: A044161824008

PRACTICAL 1

QUESTION: Using “Lung cancer_2108202” datasheet and do the following:

1. Define the data for each of the characteristics under study.
2. Compute appropriate measure of central tendency along with dispersion for each variable.
3. Also compute coefficient of skewness and kurtosis for numeric data types.

OBJECTIVE: To understand and describe the type of data and perform exploratory data analysis.

FORMULA USED:

1. AVERAGE (range)

2. MODE (range)

3. MIN (range)

4. MAX (range)

5. RANGE = MAX - MIN

6. QUARTILE (range, 1)

7. QUARTILE (range, 3)

8. INTER QUARTILE RANGE = Q3 – Q1

9. PERCENTAGE for categorical data.

10. VAR.S(range)

11. STDEV.S(range)

12. IF (logical_test, value_if_true, value_if_false)
13. COUNTIF (range, criteria)

PROCEDURE

1. Assign values to the categorical variables for calculations such as, in case of Gender: M=1 and F=2 and in SES: LOW=1, MEDIUM=2 and HIGH=3.

2. Compute the measures of central tendency and measures of dispersion using the above formulae.

3. For categorical data calculate mode and percentage as a measure of central tendency and dispersion.

RESULT:

The data type for each of the characteristics under study:

VARIABLE	Data Type
PATIENT_ID	Nominal
PINCODE	Nominal
AGE Years	Scale Ratio
GENDER	Nominal
Height in cm	Scale Ratio
Weight in kg	Scale Ratio
SES	Ordinal
SBP in mmHg	Scale Ratio
Body Temp	Scale Ratio
Diabetes	Nominal
Smoking	Nominal
Lung cancer	Nominal
Staging	Nominal`
Date of First visit	Scale
Date of Diagnosis	Scale

Measure of central tendency along with dispersion for each variable:

VARIABLE	DATA TYPE	MEASURE OF CENTRAL TENDENCY	CENTRAL TENDENCY	MEASURE OF DISPERSION	DISPERSION
PATIENT ID	NOMINAL	MEDIAN	ABC30 & ABC31	RANGE	ABC01 - ABC60
PINCODE	NOMINAL	MODE			58.33%
AGE	SCALE-RATIO	MEAN	29.1166	STANDARD DEVIATION	3.17
GENDER	NOMINAL	MODE	M		
HEIGHT	SCALE-RATIO	MEAN	156.483	STANDARD DEVIATION	4.11
WEIGHT	SCALE-RATIO	MEAN	74.466	STANDARD DEVIATION	10.29
SES	ORDINAL	MEDIAN		RANGE	LOW - HIGH

SBP	SCALE-INTERVAL	MEAN	131.866	STANDARD DEVIATION	15.75
BODY TEMP	SCALE-INTERVAL	MEAN	99.191	STANDARD DEVIATION	1.37
SMOKING	NOMINAL	MODE	NO		
DIABETES	NOMINAL	MODE	NO		
CANCER	NOMINAL	MODE	NO		
STAGE	ORDINAL	MODE	N/A	RANGE	NA - 4
DATE OF VISIT	NOMINAL	MEDIAN	3/1/2024	RANGE	1/6/2024 - 6/28/2024
DATE OF DIAGNOSIS	NOMINAL	MEDIAN	3/23/2024	RANGE	1/13/2024 - 7/4/2024

Coefficient of skewness and kurtosis for numeric data types:

VARIABLE	KURTOSIS	SKEWNESS
PATIENT ID		
PINCODE		
AGE	-1.28	0.26
GENDER		
HEIGHT	-1.34	-0.34
WEIGHT	-1.11	-0.12
SES		
SBP	-1.29	0.19
BODY TEMP	-0.67	0.26
SMOKING		
DIABETES		
CANCER		
STAGE		
DATE OF VISIT	-1.38	0.21
DATE OF DIAGNOSIS		0.22

PRACTICAL 2

QUESTION: Using “Lung cancer_2108202” datasheet and do the following:

A. Compute outliers for systolic blood pressure based on quartiles

B. What will be the average height of males and females separately

C. Generate a new variable called BMI (Body Mass Index) and perform descriptive

statistics on it.

OBJECTIVE: To find if there are any outliers based on quartile, average height based on genders and find how healthy a patient is by calculating their BMI.

FORMULA USED:

1. QUARTILE (range, 1)

2. QUARTILE (range, 3)

3. INTER QUARTILE RANGE = Q3 – Q1

4. Upper Bound = Q3 + (1.5*IQR)

5. Lower Bound = Q1 - (1.5*IQR)

6. OUTLIER = OR (value > upper bound, value < lower bound)

PROCEDURE:

1. Compute the upper and lower bounds for the SBP data.

2. Using the above formula find the outliers for each variable and value.

3. If the answer is “FALSE” means outlier is not present and if the answer is “TRUE” means outlier is present in the data.

RESULT:

PATIENT_ID	SBP_mmHg	Upper Bound	Lower Bound	Outliers
ABC01	130	187	75	FALSE
ABC02	135	187	75	FALSE
ABC03	110	187	75	FALSE
ABC04	120	187	75	FALSE
ABC05	160	187	75	FALSE
ABC06	140	187	75	FALSE
ABC07	115	187	75	FALSE
ABC08	110	187	75	FALSE
ABC09	112	187	75	FALSE
ABC10	115	187	75	FALSE

Average height of Males (in cm)	157.133
Average height of Females (in cm)	155.833

PATIENT_ID	HEIGHT_cm	WEIGHT_Kg	HEIGHT_m	BMI
ABC01	151	80	1.51	35.09
ABC02	156	81	1.56	33.28
ABC03	156	85	1.56	34.93

ABC04	154	90	1.54	37.95
ABC05	157	79	1.57	32.05
ABC06	150	80	1.5	35.56
ABC07	151	82	1.51	35.96
ABC08	152	65	1.52	28.13
ABC09	158	54	1.58	21.63
ABC10	162	60	1.62	22.86
ABC11	157	60	1.57	24.34
ABC12	158	65	1.58	26.04

	BMI
Average	30.501
Median	30.38
Mode	34.92767916
Max	39.11
Min	21.63
Range	17.48
First Quartile	26.79350258
Third Quartile	34.57717655
IQR	7.783673971
Variation	21.741
Standard Deviation	4.663
Skewness	-0.163
Kurtosis	-1.052

PRACTICAL 3

QUESTON: Visualize the given dataset as per your choice of graphs.

OBJECTIVE: To perform graphical representation on the given dataset.

FORMULA USED: INSERT > RECOMMENDED CHARTS.

PROCEDURE:

1. Select the cell of the data to obtain the required charts.

2. Click on the "Insert" tab in the Excel ribbon at the top of the screen.

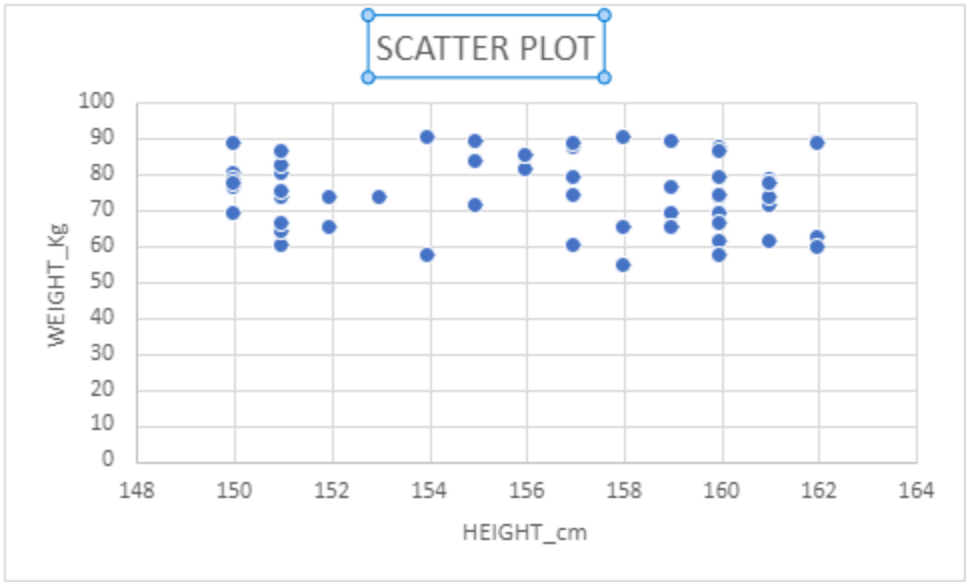
3. Click on the "Charts" group to see various chart types available in Excel. Select the chart type that best suits the data.

4. Customize Chart Elements by right clicking on the chart obtained.

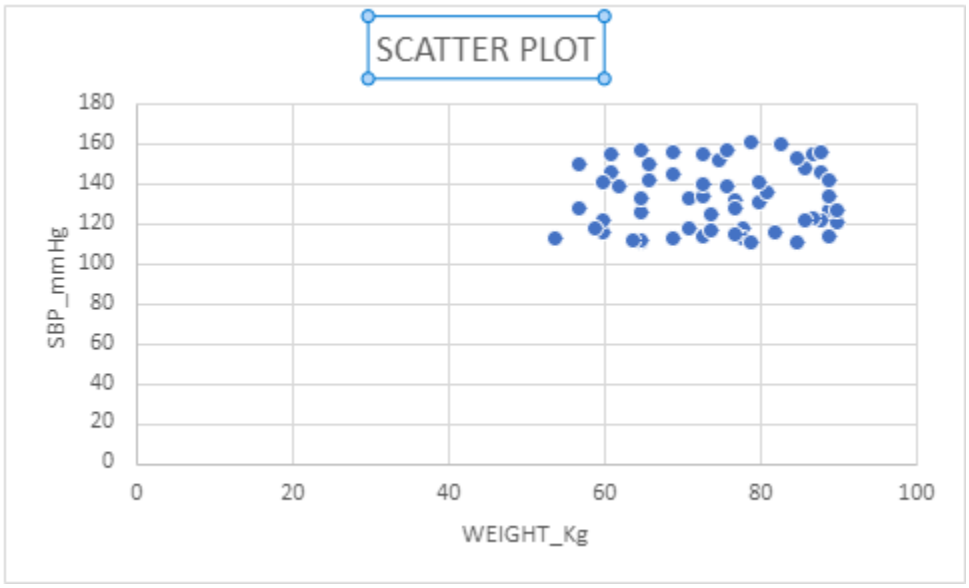
5. Go to format and edit information like “Chart Title”, “Axis Labels”, “Data Labels”, color, fonts, gridlines and more.

RESULT:

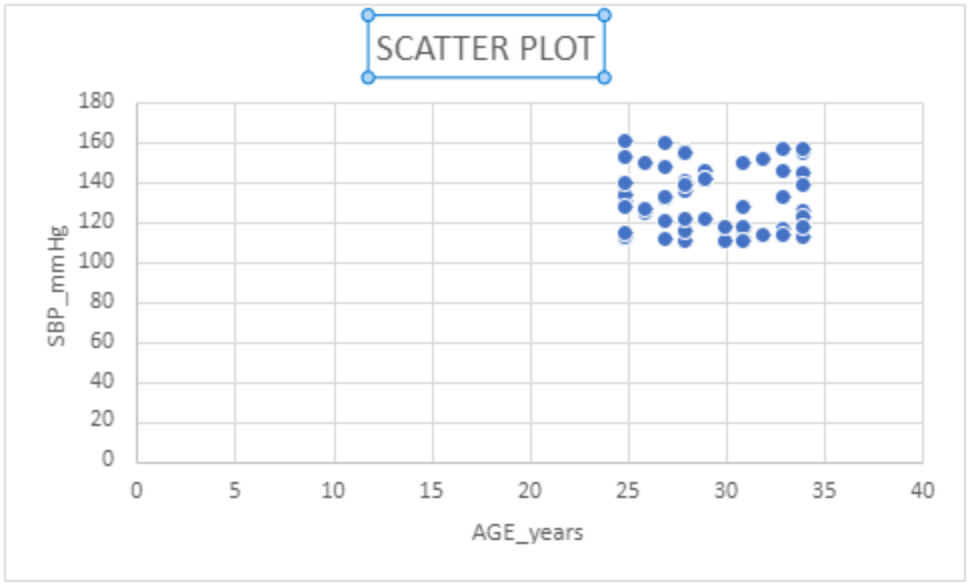
A. RELATIONSHIP BETWEEN HEIGHT (cm) AND WEIGHT (Kg)



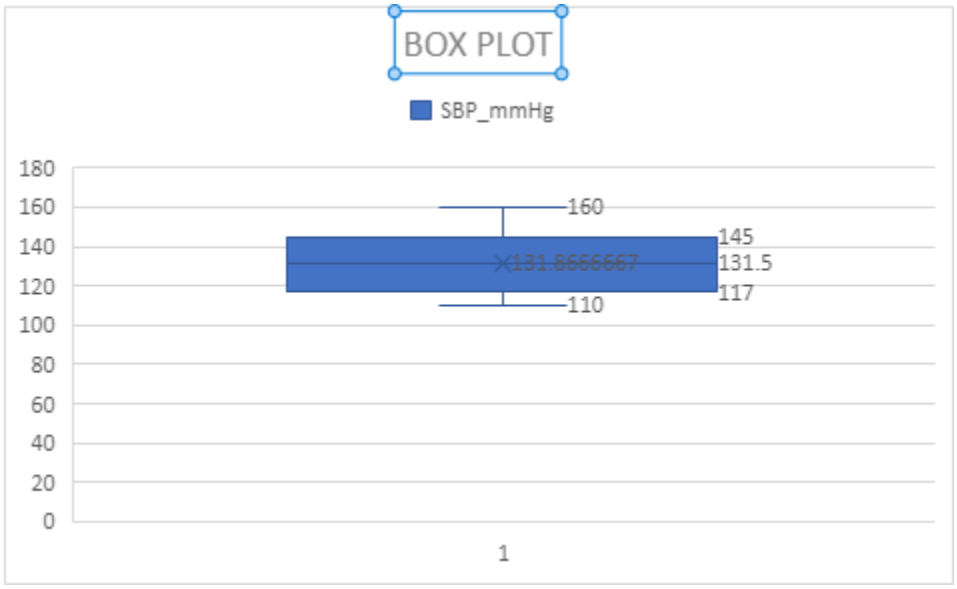
B. RELATIIONSHIP BETWEEN WEIGHT (Kg) AND SBP (mmHg)



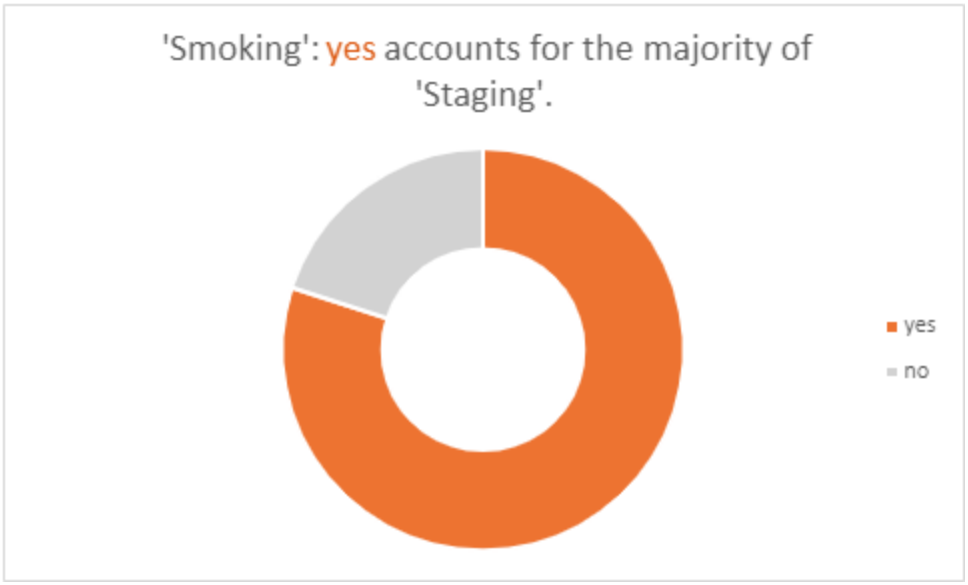
C. RELATIONSHIP BETWEEN AGE (Yrs.) AND SBP (mmHg)



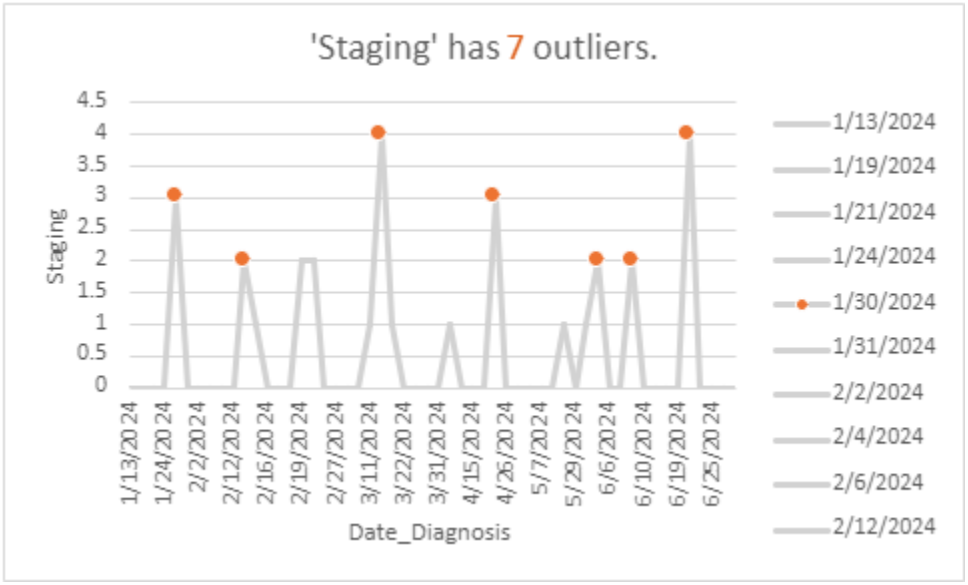
D. GRAPGICAL VISUALIZATION OF SBP (mmHg)



E. PIE CHART


'Smoking': yes accounts for the majority of 'Staging'.
yes
no

F.


'Staging' has 7 outliers.
Staging
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
1/13/2024
1/24/2024
2/2/2024
2/12/2024
2/16/2024
2/19/2024
2/27/2024
3/11/2024
3/22/2024
3/31/2024
4/15/2024
4/26/2024
5/7/2024
5/29/2024
6/6/2024
6/10/2024
6/19/2024
6/25/2024
Date_Diagnosis
1/13/2024
1/19/2024
1/21/2024
1/24/2024
1/30/2024
1/31/2024
2/2/2024
2/4/2024
2/6/2024
2/12/2024

PRACTICAL 4

QUESTION: Marks obtained by 10 students in Mathematics and Statistics are given in the table.

DATASET:

SCORES IN MATHEMATICS	SCORES IN STATISTICS
75	85
30	45
60	54
80	91
53	58
35	63
15	35
40	43
38	45
48	44

Is there any correlation between Mathematics and Statistics? If yes, how scores in Mathematics affect scores in Statistics.

OBJECTIVE: To find is there any correlation association between scores in Mathematics and scores in Statistics.

FORMULA USED:

1. CORREL ()
2. HOME > ADD INS > TOOL PAK > LINEAR REGRESSION.

PROCEDURE:

1. Compute correlation between scores in Mathematics and Statistics using above mentioned function,
2. If correlation exits between the two, we will calculate the regression equation.
3. For that, in Linear Regression tool Pak put scores in Statistics in Y input range and scores in Mathematics in X input range,
4. Click OK.

RESULT:

Pearson Correlation	
0.859349402	

The result states that there is a high positive correlation between the scores in Mathematics and scores in Statistics, i.e., increase in scores in Mathematics increase the scores in Statistics.

Therefore, we will calculate how scores in Mathematics effect scores in Statistics.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.812618569							
R Square	0.660348939							
Adjusted R Square	0.611827359							
Standard Error	10.33864835							
Observations	9							

ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	1454.675341	1454.675341	13.60938654	0.00776592			
Residual	7	748.2135478	106.8876497					
Total	8	2202.888889						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	21.25865468	9.296577207	2.286718456	0.056078117	-0.724257096	43.24156646	-0.724257096	43.24156646
75	0.718476461	0.194757087	3.689090205	0.00776592	0.257949134	1.179003788	0.257949134	1.179003788

CONCLUSION: Y (STATISTICS) = 18.726 + 0.793X (MATHEMATICS)

The regression equation states that unit change in scores in Mathematics gives on an average 0.8 unit change in scores in Statistics.

PRACTICAL 5

SERIAL NO.	WEIGHT (Kg)	HEIGHT (cm)	AGE (Months)
1	6.4	57	8
2	7.1	59	10
3	5.3	49	6
4	6.7	62	11
5	5.5	51	8
6	5.8	50	7
7	7.7	55	10
8	5.7	48	9
9	5.6	42	10
10	5.1	42	6
11	7.6	61	12
12	6.8	57	9

A. Is there any correlation between Weight and Height?

B. Is there any correlation between Weight and Age?

C. Perform Simple Linear Regression to predict weight of children.

D. How height and age effects on the weight of children?

A.

OBJECTIVE: To find is there any correlation between Weight and Height.

FORMULA USED: CORREL ()

PROCEDURE: Compute correlation between Weight and Height using above mentioned function,

RESULT:

Correlation b/w h &w
0.81425695

CONCLUSION:

The correlation between ‘Weight’ and ‘Height’ is +0.81 which is close to +1, so they are highly positively correlated, i.e., increase in height increases the weight.

B.

OBJECTIVE: To find is there any correlation between Weight and Age.

FORMULA USED: CORREL ()

PROCEDURE: Compute correlation between Weight and Height using above mentioned function,

RESULT:

Correlation b/w W & A	
0.769816802	

CONCLUSION:

Correlation between 'Weight' and 'Age' is +0.77 which is close to +1, so they are highly positively correlated (although not as high as between height and weight), i.e., increase in age increases the weight.

C.

OBJECTIVE: To find Simple Linear Regression to predict weight of children.

FORMULA USED: HOME > ADD INS > TOOL PAK > LINEAR REGRESSION.

PROCEDURE:

1. For that, in Linear Regression tool Pak put weight in Y input range and Age in X1 input range and Weight in X2 input range.
2. Click OK.

RESULT:

SLR (W&H)	SLR (W&A)
0.618984871	3.057142857
0.107223036	0.364285714

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.882938577							
R Square	0.779580531							
Adjusted R Square	0.724475664							
Standard Error	0.494231573							
Observations	11							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	2	6.911335761	3.455667881	14.14721729	0.002360477			

Residual	8	1.954118784	0.244264848					
Total	10	8.865454545						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	0.650160099	1.178640239	0.551618787	0.596276799	-2.067789164	3.368109362	-2.067789164	3.368109362
57	0.072472245	0.029641417	2.444965602	0.040251543	0.004119015	0.140825476	0.004119015	0.140825476
8	0.204124751	0.10546214	1.935526353	0.088958623	-0.03907138	0.447320882	-0.03907138	0.447320882

PRACTICAL 6

QUESTION-A: What is the probability of having 2 girl babies out of 3 babies?

QUESTION-B: What is the probability of having more than 2 girl babies out of 3 babies?

OBJECTIVE-A: To find the probability of having 2 girl babies out of 3 babies using Binomial Distribution.

FORMULA USED: BINOM.DIST(number_ s, trial, probability_ s, cumulative)

PROCEDURE:

1. Define:

total no of trials (n) = 3, no. of success (girl babies) (x) = 2 and probability of success (p) = 0.5.

2. Find the required probability i.e., P(X=x) using the above-mentioned Binomial function formula.

3. To compute probability mass function, select FALSE in cumulative.

4. Required result is obtained.

RESULT:

P(X=x)	0.375

CONCLUSION: The probability of having 2 girl babies out of 3 babies is 0.375.

OBJECTIVE-B: To find the probability of having more than 2 girl babies out of 3 babies using Binomial Distribution.

FORMULA USED: BINOM.DIST(number_ s, trial, probability_ s, cumulative)

PROCEDURE:

1. Define:

total no of trials (n) = 3, no. of success (girl babies) (x) > 2 and probability of success (p) = 0.5.

2. Find the required probability i.e., P (X > x) = 1 – P (X <= x).

3. To compute cumulative function, select TRUE in cumulative to find P (X <= x).

4. Required result is obtained.

RESULT:

P(X<=2)	0.875	
p(X>2)	1-0.875	0.125

CONCLUSION: The probability of having more than 2 girl babies out of 3 babies is 0.125.

PRACTICAL 7

QUESTION: In a class student grade for a single exam follows beta distribution with parameter with alpha = 8.28 and beta = 3.16. What is the probability that student is below the mean?

OBJECTIVE: To find the probability that student grade is below the mean using Beta Distribution.

FORMULA USED:

1. Mean: alpha/alpha + beta
2. Variance: alpha*beta/ (alpha + beta) ^2 * (alpha+beta+1)
3. BETA.DIST(x, alpha, beta, commutative)

PROCEDURE:

1. Compute the mean for the given information using above formula.
2. Define:

   x = mean value, alpha = 8.28 and beta = 3.16.

3. Find the required probability i.e., P (alpha, beta, X< 0.72).
4. To compute cumulative function, select TRUE in cumulative to find P (X > x).
5. Required result is obtained.

RESULT:

MEAN	0.723776224
ALPHA	8.28
BETA	3.16
P(X<0.723)	0.460204547

CONCLUSION: Average of student grade = 0.72 and the probability that the student is below mean is 0.462.

PRACTICAL 8

QUESTION: Based on the given data set test the hypothesis that sample mean is equal to population mean of 100.

DATASET:

DATA POINTS
70
120
110
101
88
83
95
98
107
100

OBJECTIVE:

H0: There is no significant difference between the sample mean and population mean, i.e., $\mu = 100$.

H1: There is a significant difference between the sample mean and population mean, i.e., $\mu \neq 100$.

FOEMULA USED:

1. Sample Mean = AVERAGE ( )

2. Standard Deviation = STDEV.S ( )

3. No. of Observation = COUNT ( )

4. Standard Error = Standard Deviation / Square root of no. Of observation

5. Degree of freedom = No. of Observation - 1

6. t statistics = | (Sample Mean – Hypothesised Mean) / Standard Error|

7. P value = TDIST(x, degree_of _freedom, tails)

PROCEDURE:

1. Set up the Null Hypothesis H0.

2. Set up the Alternative Hypothesis H1.

3. Choose the 5% level of significance (α) or p value.

4. Compute the test statistics using above formulae.

5. If p value > 0.05, accept H0 otherwise reject.

RESULT:

Sample Mean	97.2
Standard Deviation	14.2735186
Standard Error	4.513682901
Count	10
Degree of Freedom	9
Hypothesized Mean	100
t-statistic	0.62
P-value	0.55

As P value is greater than 0.05, we will accept the null hypothesis, i.e., There is no significant difference between the sample mean and population mean, i.e., μ= 100.

PRACTICAL 9

QUESTION: Weight of 10 participants is given as follows –

DATASET:

WEIGHT (Kg)
70
67
62
68
61
68
70
64
64
66

Is the average weight of participants being 64 at 5% level of significance.

OBJECTIVE:

H_0: There is no significant difference between the average weight of participants, i.e., $\mu = 64$.

H_1: There is a significant difference between the average weight of participants, i.e., $\mu \neq 64$.

FOEMULA USED:

1. Sample Mean = AVERAGE ( )
2. Standard Deviation = STDEV.S ( )
3. No. of Observation = COUNT ( )
4. t statistics =

(x−x bar)s/√n *(x−x bar)s/√n*

PROCEDURE:

1. Set up the Null Hypothesis H_0.
2. Set up the Alternative Hypothesis H_1.
3. Choose the 5% level of significance (α) or p value.
4. Compute the test statistics using above formulae.
5. If calculated t < tabulated t value, accept H_0 otherwise reject H_0.

   OR, if p value > 0.05, accept H_0 otherwise reject.

RESULT:

sample mean	66	
standard deviation	3.162278	
N =	10	
Mu =	64	
Standard error	1	
df	9	
	T statistic	2
	P value	0.076553

CONCLUSION: Since, p value (0.076553) > 0.05, it is significant at 5%. Therefore, we accept H_0 and conclude that the average weight of participants is 64 Kg.

PRACTICAL 10

QUESTION: Maths score of two groups is given as follows. Is average score of two groups are same.

DATASET:

GROUP 1	GROUP 2
25	44
32	34
30	22
34	10
24	47
14	31
32	40
24	30
30	32
31	35
35	18
25	21
	35
	29
	22

OBJECTIVE:

H_0: There is no significant difference between the average maths score of two groups,

i.e., $\mu_1 = \mu_2$.

H_1: There is a significant difference between the average maths score of two groups,

i.e., $\mu_1 \neq \mu_2$.

FORMULA USED:

1. HOME > ADD INS > TOOL PAK > F TEST TWO SAMPLE FOR VARIANCE
2. HOME > ADD INS > TOOL PAK > t-TEST: TWO SAMPLE.

PROCEDURE:

1. Set up the Null Hypothesis H_0.
2. Set up the Alternative Hypothesis H_1.
3. Choose the appropriate level of significance (α).
4. Compute F test to test whether the sample variance is equal or not.
5. If calculated F < tabulated F value, accept H0 (or p value > 0.05) and apply t-test two sample assuming equal variances.
6. If calculated F > tabulated F value, reject H0 and apply t-test two sample assuming unequal variances.
7. Conclude the results.

RESULT:

F test for testing equality of variances

H_0: There is no difference in variance of both the groups

H_1: There is a difference between the variance of both the groups

F-Test Two-Sample for Variances		
	25	44
Mean	28.27273	29
Variance	37.01818	92.30769
Observations	11	14
df	10	13
F	0.40103	
P(F<=f) one-tail	0.077024	
F Critical one-tail	0.346359	

Since, p value < 0.05 at 5% level of significance. Therefore, we reject H0. It means that we fail to assume equal variance of two samples.

t-TEST: TWO SAMPLE ASSUMING UNEQUAL VARIANCES.

H_0: There is no difference in means of both the groups

H_1: There is a difference between the means of both the groups

T Test: Two-Sample Assuming Unequal Variances		
	25	44
Mean	28.27273	29
Variance	37.01818	92.30769
Observations	11	14
Hypothesized Mean Difference	0	
df	22	
t Stat	-0.23046	
P(T<=t) one-tail	0.409932	
t Critical one-tail	1.717144	
P(T<=t) two-tail	0.819864	
t Critical two-tail	2.073873	

CONCLUSION: Since, p value (0.525) > 0.05 at 5% α. Therefore, we accept H_0 and conclude that there is no significant difference between the average maths score of two groups,

i.e., $\mu_1 = \mu_2$.

PRACTICAL 11

QUESTION: Weight gain in Kg due to food A and food B is given as –

DATASET:

FOOD A	FOOD B

49	52
53	55
51	52
52	53
47	50
50	54
52	54
53	53

A. Assuming that 2 samples are independent. Test the hypothesis that weight gain due to food A is better than food B.
B. When same set of 8 animals were used for both food A and B then, test the same hypothesis.

OBJECTIVE A:

H0: There is no significant difference in weight gain due to food A and food B.

H1: Weight gains due to food A is better than food B.

FORMULA USED:

1. HOME > ADD INS > TOOL PAK > F TEST TWO SAMPLE FOR VARIANCE
2. HOME > ADD INS > TOOL PAK > t-TEST: TWO SAMPLE.

PROCEDURE:

1. Set up the Null Hypothesis H0.
2. Set up the Alternative Hypothesis H1.
3. Choose the appropriate level of significance (α) or p value.
4. Compute F test to test whether the sample variance is equal or not.
5. If calculated F < tabulated F value, accept H0 or p value > 0.05 and apply t-test two sample assuming equal variances.
6. If calculated F > tabulated F value, reject H0 and apply t-test two sample assuming unequal variances.
7. Conclude the results.

RESULT:

F-Test Two-Sample for Variances		
	FOOD A	**FOOD B**
Mean	50.875	52.875
Variance	4.410714	2.410714
Observations	8	8
df	7	7
F	1.82963	
P(F<=f) one-tail	0.22193	
F Critical one-tail	3.78704354	

Since, p value (0.22) > 0.05 at 5% level of significance. Therefore, we accept H0 and will compute **t-TEST: TWO SAMPLE ASSUMING EQUAL VARIANCES.**

T Test: Two-Sample Assuming Equal Variances		
	FOOD A	**FOOD B**
Mean	50.875	52.875
Variance	4.410714	2.410714

Observations	8	8
Pooled Variance	3.410714	
Hypothesized Mean Difference	0	
df	14	
t Stat	-2.16589	
P(T<=t) one-tail	0.024036	
t Critical one-tail	1.76131	
P(T<=t) two-tail	0.048071	
t Critical two-tail	2.144787	

CONCLUSION: Since, p value (0.0240) < 0.05 at 5% α. Therefore, we reject H0 and conclude that weight gain due to food A is better than food B.

OBJECTIVE B:

H0: There is no significant difference in weight gain due to food A and food B.

H1: Weight gains due to food A is better than food B.

FORMULA USED: HOME > ADD INS > TOOL PAK > t-TEST: PAIRED TWO SAMPLE FOR MEAN.

PROCEDURE:

1. Set up the Null Hypothesis H0.
2. Set up the Alternative Hypothesis H1.
3. Choose the appropriate level of significance (α) or p value.
4. Compute t statistics for paired sample.
5. If calculated t < tabulated t value, accept H0 otherwise reject H0.

   OR, if p value > 0.05, accept H0 otherwise reject.

RESULT:

T Test: Paired Two Sample for Means		
	FOOD A	**FOOD B**
Mean	50.875	52.875
Variance	4.410714286	2.410714286
Observations	8	8
Pearson Correlation	0.783106226	
Hypothesized Mean Difference	0	
df	7	
t Stat	-4.320493799	
P(T<=t) one-tail	0.001739042	
t Critical one-tail	1.894578584	
P(T<=t) two-tail	0.003478084	
t Critical two-tail	2.364624236	

CONCLUSION: Since, p value (0.0240) < 0.05 at 5% α. Therefore, we reject H0 and conclude that weight gain due to food A is better than food B.

PRACTICAL 12

QUESTION: Birth weight in Kg of babies born to women in particular hospital, classify according to their health status in table –

DATASET:

HEALTH STATUS OF WOMEN		
GOOD	**FAIR**	**POOR**
3.1	3	2.4
2.9	2.6	2.2
3.5	2.5	2
3.6	2.7	2.3
3.4	2.6	2.2
	2.8	2.1
	2.7	2
		2.4
		2.3
		2.5

[Text Wrapping Break]

Test whether the mean birth weight of babies among the 3 groups of women are same or not. (Test for $\alpha = 5\%$ and 1%)

OBJECTIVE:

H0: All the 3 groups of women have equal mean, i.e., $\mu 1 = \mu 2 = \mu 3$.

H1: At least on group have different mean, i.e., $\mu 1 \neq \mu 2 \neq \mu 3$.

FORMULA USED: HOME > ADD INS > TOOL PAK > ANOVA: SINGLE FACTOR.

PROCEDURE:

1. Set up the Null Hypothesis H0.
2. Set up the Alternative Hypothesis H1.
3. Choose the level of significance (α) or p value at 5% and 1%.
4. Compute F statistics using ANOVA table.
5. If calculated F < tabulated F value, accept H0 otherwise reject H0.

   OR, if p value > 0.05, accept H0 otherwise reject.

6. In case H0 is rejected, perform STUDENT NEWMAN-KEULS TEST to find which group is significantly different.

RESULT:

Anova: Single Factor (alpha 5%)						
SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
GOOD	5	16.5	3.3	0.085		
FAIR	7	18.9	2.7	0.026667		
POOR	10	22.4	2.24	0.029333		
ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	3.799636	2	1.899818	47.24679	4.23E-08	3.521893
Within Groups	0.764	19	0.040211			
Total	4.563636	21				
ANOVA: Single Factor						

(alpha = 1 %)						
SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
GOOD	5	16.5	3.3	0.085		
FAIR	7	18.9	2.7	0.026667		
POOR	10	22.4	2.24	0.029333		
ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	3.799636	2	1.899818	47.24679	4.23E-08	5.925879
Within Groups	0.764	19	0.040211			
Total	4.563636	21				

CONCLUSION:

- Since, p value < 0.05 at 5% α. Therefore, we reject H0 and conclude that at least on group of women have different mean, i.e., $\mu 1 \neq \mu 2 \neq \mu 3$.
- Also, p value < 0.01 at 1% α. Therefore, we reject H0 and conclude that at least on group of women have different mean, i.e., $\mu 1 \neq \mu 2 \neq \mu 3$.

PRACTICAL 13

QUESTION: Given the dataset in the workbook Module3 in the sheet Level 1, use excel formula to mention the Company's name against the Employee ID.

OBJECTIVE: To mention the Company's name against each Employee ID from the dataset.

FUNCTION USED: VLOOKUP STATEMENT

PROCEDURE:

1. Apply the VLOOKUP function and select the cell containing Employee ID.
2. Select the whole dataset from the sheet Data.
3. Write the column number containing the Company name.
4. The required Company name is found for the given Employee ID.
5. Drag down the remaining cells.

RESULT:

Emp ID	Company Name
3721991182	Google
8839343394	Google
2472705883	Google
3378854170	Google
6906990252	Amazon
3388492123	Amazon
4738642548	EXL
5409103851	Google
3720565805	Amazon
3287299135	EXL
4061369981	American Express
2423281492	American Express
920205936	American Express
576284994	American Express
6689574105	American Express
6907591325	American Express
7306772207	American Express
7365692086	Gartner
521793959	Gartner

5671070342	American Express

PRACTICAL 14

QUESTION: Given the dataset in the workbook Module3 in sheet Level 2, populate column B with Total number of Employees for each Company from base data using excel functions.

OBJECTIVE: To fill column B with Total number of Employees from the base data.

FUNCTION USED: COUNTIF STATEMENT

PROCEDURE:

1. Use the COUNTIF function and select the whole dataset from the datasheet.
2. Select the cell containing the Company name.
3. The number of Employees is obtained for that Company.
4. Drag down the remaining cells.

RESULT:

Company Name	**Total Employees**
Google	2250
Gartner	1698
Amazon	477
CBRE	70
EXL	1843
American Express	489
Moody's	1018
Grand Total	**7845**

PRACTICAL 15

QUESTION: Given the dataset in the workbook Module3 in Level 3 excel sheet, populate column B to L with Total number of Employees for each Company and Location from base data using “COUNTIFS” function.

OBJECTIVE: To fill column B to L with the Total number of Employees for each Company and their respective Locations.

FUNCTION USED: COUNTIFS STATEMENT

PROCEDURE:

1. Use the COUNTIFS function and select the Company name data from the sheet Data.

2. Select the cell containing the company name.
3. Select the Location data from the sheet Data.
4. Select the cell containing the location with absolute reference i.e., using the $ sign.
5. Required result is obtained. Drag down the remaining cells.
6. Repeat the same procedure for the remaining columns having different Locations.

RESULT:

Company Name/Location	**Ahmedabad**	**Bangalore**	**Chandigarh**	**Chennai**	**Hyderabad**	**Kochi**	**Kolkata**	**Mumbai**	**New Delhi**	**Pune**	**Grand Total**
Google	39	298	6	134	100	12	116	747	666	132	2250
Gartner	29	159	3	175	74	13	53	547	617	28	1698
Amazon	2	42	5	7	13	0	7	221	178	2	477
American Express	4	31	3	11	14	4	17	113	285	7	489
Moody's	16	125	2	15	18	0	51	314	470	7	1018
EXL	18	292	2	117	69	6	59	689	477	114	1843
CBRE	0	2	0	1	2	0	2	24	39	0	70
Grand Total	**108**	**949**	**21**	**460**	**290**	**35**	**305**	**2655**	**2732**	**290**	**7845**

PRACTICAL 16

QUESTION: Given the dataset in the workbook Moduld3, do the following:

1. In Level 4 sheet of excel, calculate percentage of on contracts in column E.
2. Populate column F if percentage of on contracts exceed 5 % then “OK” else “NOT OK” using excel function.

OBJECTIVE:

1. To calculate the percentage of ‘On Contracts’ and fill column E.
2. Find the status whether the percentage is greater than 5% or not.

FUNCTION USED: IF STATEMENT

PROCEDURE:

1. Use the formula (on contacts/ grand total) *100% to calculate the percentage for corresponding cell.
2. Drag down the remaining cells.
3. Use the IF function to find the status.

4. Select the on contracts cell, write the first condition “OK” and the other condition as “NOT OK”.
5. Drag down the remaining cells.

RESULT:

Company Name	On Payroll	On Contracts	Grand Total	%age On Contracts	Status
Google	989	31	1020	3%	NOT OK
Moody's	460	15	475	3%	NOT OK
Amazon	1622	72	1694	4%	NOT OK
Gartner	380	26	406	6%	OK
EXL	2217	40	2257	2%	NOT OK
American Express	58	11	69	16%	OK
CBRE	1642	209	1851	11%	OK
Grand Total	**7368**	**404**	**7772**	5%	OK

PRACTICAL 17

QUESTION: Given the dataset in the workbook Module3_Part2 in excel sheet Level 5 using conditional formatting do the following:

1. Red highlight cells with average work experience of less than equal to 2 years.
2. Yellow highlight cells with average work experience between 2 to 3 years.

3. Green highlight cells with average work experience of greater than 3 years.

OBJECTIVE: To perform conditional formatting on the given dataset.

FUNCTION USED: CONDITIONAL FORMATTING

PROCEDURE:

1. Apply the first rule less than 2 and select the red colour to highlight.
2. Apply the second rule between 2 and 3 and select the yellow colour to highlight.
3. Apply the third rule greater than 3 and select the green colour to highlight the cell.

RESULT:

Company Name /Location	Ahmedabad	Bangalore	Chandigarh	Chennai	Hyderabad	Kochi	Kolkata	Mumbai	New Delhi	Pune	Grand Total
Google	3.293150685	1.789976082	1.852054795	2.058082192	3.327853881	5.554794521	1.38113804	2.410464713	1.994051377	1.536986301	2.107488139
Amazon	1.956447803	2.428706815	1.689497717	2.431076321	2.664124398	1.859220232	4.291289739	2.808139477	2.514931947	2.453522505	2.637529108
EXL	1.8	1.601289283	3.5	1.35	1.442781876	2.431076321	0.894063927	1.230365297	1.522693202	1.797260274	1.363789690
American Express	2.5	4.185554172	3.7	2.3	5.554794521	3.6	2.431076321	4.817808219	5.374775626	2.431076321	4.902896923
Gartner	6	2.693551106	1.175342466	2.522200478	3.015029354	2.242922374	4.582912761	2.957037217	2.815245167	2.587940891	2.915609054
Moody's	1	2.901369863	2.7	2.652054795	4.234246575	4.9	4.575342466	4.357648402	3.604425711	2.431076321	3.874716243
CBRE	3.421232877	7.187892179	3.602739726	4.856537983	2.921526419	2.13630137	11.38694601	6.025433386	4.121046933	2.519373777	4.932320465
Grand Total	2.767098153	2.901590531	2.006131768	2.515743069	2.925687293	2.679686888	4.062047258	3.054097248	2.863456392	2.652158003	2.948988613

PRACTICAL 18

QUESTION: Using Module3_Part2 workbook create a pivot table using "Data" sheet and provide the following data:

1. Total number of employees in each location by level also format reports to look professional.

OBJECTIVE: To create a pivot table to represent total number of employees in each location by level.

FUNCTION USED: INSERT > PIVOT TABLE

PROCEDURE:

1. Go to insert and click on pivot table.
2. Drag Location variable in Rows box.
3. Drag Levels variable in Columns box.
4. Drag Emp ID variable in Values box and change the value field as Count of Emp ID.

5. Click on ok.

RESULT:

Count of Emp ID	Location						
Level	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**	**Grand Total**
Ahmedabad	2	5	9	20	41	31	108
Bangalore	21	74	109	151	355	239	949
Chandigarh		1	1	2	12	5	21
Chennai	11	24	40	61	179	145	460
Hyderabad	6	22	26	56	108	72	290
Kochi		2	4	8	11	10	35
Kolkata	6	20	39	58	78	104	305
Mumbai	93	223	277	467	916	679	2655
New Delhi	71	253	270	462	923	753	2732
Pune	6	16	25	40	104	99	290
Grand Total	**216**	**640**	**800**	**1325**	**2727**	**2137**	**7845**

PRACTICAL 19

QUESTION: Use Module3_Part2 workbook to calculate experience in years as of 17 January, 2017.

OBJECTIVE: To calculate the experience in years of the employees as of 17 Jan, 2017.

FUNCTION USED: DATEDIF STATEMENT

PROCEDURE:

1. Use the DATEDIF function and select the date of joining against the respective employee id.
2. Using the DATE function enter the date 17 Jan, 2017.
3. Enter “Y” to get the result in years.
4. The required result is obtained.

RESULT:

Emp Id	Date of Joining	Experience	End Date
4139280866	5/1/1979	37 Year(s)	1/17/2017
9069716721	4/1/1996	20 Year(s)	1/17/2017
3721991182	11/15/1995	21 Year(s)	1/17/2017
8839343394	12/1/1995	21 Year(s)	1/17/2017
2472705883	2/5/1997	19 Year(s)	1/17/2017
3378854170	2/1/1997	19 Year(s)	1/17/2017
6906990252	2/1/2000	16 Year(s)	1/17/2017
3388492123	1/1/1998	19 Year(s)	1/17/2017
4738642548	9/5/1991	25 Year(s)	1/17/2017
4739720562	7/1/1996	20 Year(s)	1/17/2017
4335947116	5/18/1989	27 Year(s)	1/17/2017
4255831918	7/3/1996	20 Year(s)	1/17/2017
5482619147	9/1/1991	25 Year(s)	1/17/2017
1081318783	5/12/1997	19 Year(s)	1/17/2017
3156583728	12/21/1995	21 Year(s)	1/17/2017
2698088540	12/2/1996	20 Year(s)	1/17/2017
5409103851	5/1/1997	19 Year(s)	1/17/2017
3720565805	1/5/1996	21 Year(s)	1/17/2017
3287299135	7/1/1994	22 Year(s)	1/17/2017
2091669707	9/1/1999	17 Year(s)	1/17/2017
2842709710	1/17/2000	17 Year(s)	1/17/2017
1785456686	4/24/2000	16 Year(s)	1/17/2017
3770845043	11/21/1995	21 Year(s)	1/17/2017
9681739924	5/1/1998	18 Year(s)	1/17/2017
5749423456	7/7/1999	17 Year(s)	1/17/2017
4921870396	6/1/2012	4 Year(s)	1/17/2017
3324560655	6/1/2012	4 Year(s)	1/17/2017
8048302618	6/1/2012	4 Year(s)	1/17/2017
8308439340	6/1/2012	4 Year(s)	1/17/2017
1047610265	6/1/2012	4 Year(s)	1/17/2017

PRACTICAL 20

QUESTION: Based on the given constraints and objective function, obtain the value of P and Q such a way that it will maximise the objective function 45x + 60y – 6000 subjected to the constraints:

15x + 10y <= 2400

15x +15y +15y <= 2400

10x +5x +5y <= 2400

$15x + 15y <= 2400$

$x <= 100$

$y <= 50$

$x,y >= 0$

x,y should be integers

OBJECTIVE: To maximise the objective function based on given information.

FORMULA USED: DATA > SOLVER

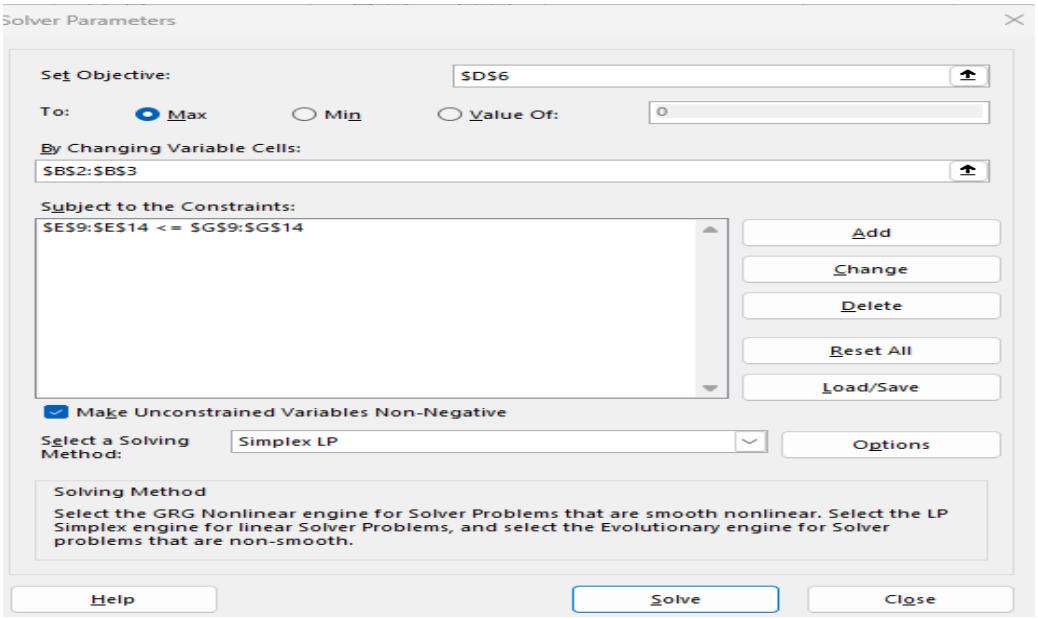
PROCEDURE:

1. Enter the values 0 for both variables whose value we need to compute (in yellow colour).
2. Write the maximising objective function (in purple colour).
3. Writ all the constraints as mentioned (in pink colour).
4. Go to Solver and:
5. Specify Objective Cell (objective function).
6. Specify Changing Cells (decision variables).
7. Specify Constraints with respective signs.
8. Specify Solver Settings (Simplex LP) and (GRE Nonlinear).

9. Solve Problem to find Optimal Solution.

RESULT:

BY SIMPLEX METHOD



BY NONLINEAR METHOD

VARIABLES						
P Qty	100	UNITS	SAY X			
Q Qty	30	UNITS	SAY Y			
OBJECTIVE FUNCTION						
REVENUE	45x + 60 y - 6000		300	MAXIMIZE		
CONSTRAINTS				L.H.S	SIGN	R.H.S
capacity of A		3x + 2y <= 480		360	<=	480

capacity of B		x + 2y <=160		160	<=	160
capacity of C		3x + y <= 480		330	<=	480
capacity of D		3x + y <= 480		330	<=	480
P order Qty		x		100	<=	100
Q Order Qty		y		30	<=	50

CONCLUSION:

P Qty (x) = **100 units**

Q Qty (y) = **30 units**

And Maximize Revenue = **Rs 300**

PRACTICAL 21

QUESTION: Consider the function $f(x) = x^4 + 8x^2 - 3$. Use MS -Excel to find the value of x that minimises the function f.

OBJECTIVE: To minimise the objective function of given f(x).

FORMULA USED: DATA > SOLVER

PROCEDURE:

1. Enter the values 0 for variable x whose value we need to compute (in orange colour).
2. Write the minimising objective function as f(x) (in purple colour).
3. Go to Solver and:
4. Specify Objective Cell (objective function).
5. Specify Changing Cells (decision variables).
6. Specify Solver Settings (GRG Nonlinear).
7. Solve Problem to find Optimal Solution.

RESULT:

CONCLUSION:

The function $f(x) = x^4 + 8x^2 - 3$ is thus minimised by the value $x = 0$.

PRACTICAL 22

QUESTION: With exactly 2700 cm^2 of cardboard, we wish to construct a box (width x, depth y and height z) that can contain a Volume V. We require the width to be double its depth. We would like to maximize the volume the box can hold. Which values of x, y, z fulfils our objective.

OBJECTIVE: To maximize the box volume expressed as V (x, y, z) = x*y*z.

FORMULA USED: DATA > SOLVER

PROCEDURE:

1. First, identify the variables, define the objectives and the constraints.
2. Enter the values 0 for all 3 variables whose value we need to compute (in yellow colour).
3. Write the maximising objective function (in blue colour).
4. Writ all the constraints as mentioned (in pink colour).
5. Go to Solver and:
6. Specify Objective Cell (objective function).
7. Specify Changing Cells (decision variables).
8. Specify Solver Settings (GRG Nonlinear).
9. Solve Problem to find Optimal Solution.

RESULT:

Let x, y, z be the width, depth and height of the cardboard box.

Objective Function:

Maximize V (x, y, z) = x*y*z

Subject To:

2xy+2yz+2xz = 2700

x = 2y

x, y, z > 0

This can occur if the variables are initially on a **fixed point** of the objective function that does not satisfy the requested criteria (maximum).

In this problem, if the cells of variables x, y, z are empty before questioning the Solver, they take on the value 0 by default. Therefore, we get the message of error. To fix the situation, we need to modify the initial values of the variables. For example, by giving the variables x, y, z as the values 1, 1, and 1, then the Solver will return the following solution:

VARIABLES						
WIDTH	SAY X	29.99	cm			
DEPTH	SAY Y	15.00	cm			
HEIGHT	SAY Z	20.01	cm			
OBJECTIVE FUNCTION						
V(x,y,z)	900	MAXIMIZE				
CONSTRAINTS				L.H.S	SIGN	R.H.S
1	2xy+2yz+2xz = 2700			2700.00	=	2700.00
2	x = 2y			29.99	=	29.99

CONCLUSION:

Width (x) = 30 cm

Depth (y) = 15 com

Height (z) = 20 cm

And Volume = 9000 cm^3

PRACTICAL 23

QUESTION: Create a database and label it as "Library DB". Also add "Borrow _Dataset" and "Inventory _Master" after creating this database

1. Count the number of rows in both the tables.
2. To check the variable names or field names and to get an idea of the data view top 10 rows of each table.

OBJECTIVE: To create a database named “LibraryDB” and do the following:

1. To count the number of rows in table, “Borrow _Dataset” and “Inventory _Master".
2. To check the variable names or field names and to get an idea of the data view top 10 rows of each table.

SOFTWARE USED: MICROSOFT SQL SERVER MANAGEMENT STUDIO

PROCEDURE:

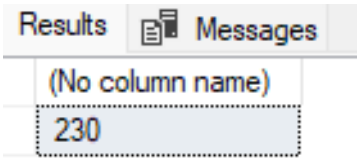
1. Open SQL Server Management Studio.
2. Right-click on Database’ > New Database > Enter Name “LibraryDB”.
3. Right-click on “LibraryDB” > Task > Import Flat File > ”Borrow_Dataset”.
4. Right-click on “LibraryDB” > Task >I mport flat file > ”Inventory_Master”.
5. Refresh the database.
6. Use ‘count’ function to know the number of rows for each table.
7. Use ‘top 10’ function to know field names and to get an idea of the data view top 10 rows of each table.

RESULT:

1. **CODE:**

```
select count(*) from Inventory_Master
```

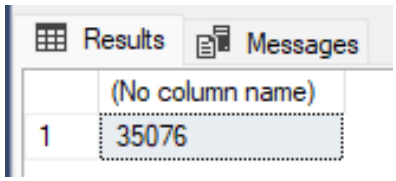
OUTPUT:



CODE:

```
select count(*) from Borrow_Dataset
```

OUTPUT:



CODE:

```
select top 10 * from Inventory_Master
```

OUTPUT:

Results | Messages

	Inv_ID	Inv_Type	Inv_Title	Inv_Entry
1	BK79599	Book	A Career in Statistics: Beyond the Numbers	2010-07-06
2	BK17632	Book	A clinician's guide to statistics and epidemiology i...	2016-06-18
3	BK97413	Book	A Course in Morphometrics for Biologists: Geomet...	2011-11-18
4	BK33683	Book	A course on statistics for finance	2018-02-26
5	BK12869	Book	A first course in probability and statistics	2003-06-19
6	BK14069	Book	A first course in statistics for signal analysis	2019-03-04
7	BK60979	Book	A first course in statistics for signal analysis (seco...	2017-03-03
8	BK44619	Book	A Guide to Teaching Statistics: Innovations and ...	2002-01-15
9	BK22360	Book	A total science: Statistics in liberal and Fascist Italy	2016-04-16
10	BK34505	Book	A Vision and Roadmap for Education Statistics	2008-05-12

CODE:

```
select top 10 * from Borrow_Dataset
```

OUTPUT:

Results | Messages

	Month	Year	Inv_ID	Days_Borrowed
1	1	2000	BK10968	9
2	1	2000	BK11516	8
3	1	2000	BK12656	5
4	1	2000	BK12869	5
5	1	2000	BK15517	5
6	1	2000	BK15703	5
7	1	2000	BK16307	2
8	1	2000	BK20605	8
9	1	2000	BK25996	8
10	1	2000	BK26042	6

PRACTICAL 24

QUESTION: Use LibraryDB database and do the following.

1. The Administration wants to know the type of inventory maintained by the library and its count. Summarize it at one table.
2. Filter the dataset for inventory type as Book and inventory entry date greater than

   2020-01-01.

OBJECTIVE: To use LibraryDB database and do the following.

1. To know the type of inventory maintained by the library and its count.
2. To filter the dataset for inventory type as Book and inventory entry date greater than 2020-01-01.

SOFTWARE USED: MICROSOFT SQL SERVER MANAGEMENT STUDIO

PROCEDURE:

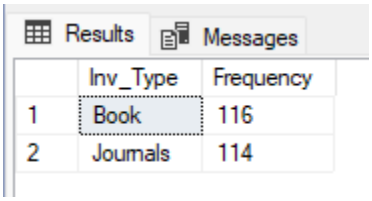
1. Open the SQL Server Management Studio.
2. Use the ‘count’ function to know the type of inventory maintained by the library and its count.
3. Use the ‘from’ and ‘where’ function to filter the dataset for inventory type as Book and inventory entry date greater than 2020-01-01.

RESULT:

1. CODE:

```
select Inv_Type, count(*) as Frequency
from Inventory_Master
group by Inv_Type
```

OUTPUT:



Results | Messages

	Inv_Type	Frequency
1	Book	116
2	Journals	114

2. CODE:

```
select *
from Inventory_Master
where Inv_Type = 'Book' and Inv_Entry > '2020-01-01'
```

OUTPUT:

Results | Messages

	Inv_ID	Inv_Type	Inv_Title	Inv_Entry
1	BK66874	Book	Advances in Business Statistics, Methods and Data Collection	2020-07-31
2	BK65535	Book	Algebraic statistics for computational biology	2022-06-20
3	BK66555	Book	Algorithmic Trading Methods: Applications Using Advanced St...	2022-12-10
4	BK61575	Book	Analytic Methods in Sports: Using Mathematics and Statistics t...	2020-11-07
5	BK26702	Book	Analyzing event statistics in corporate finance: Methodologies, ...	2020-04-14
6	BK99287	Book	Applied biostatistics for the health sciences: Second edition	2022-04-26
7	BK91016	Book	Applied medical statistics	2022-09-25
8	BK53733	Book	Applied statistics: Handbook of GENSTAT analysis	2022-07-19
9	BK79562	Book	Basic Biostatistics for Geneticists and Epidemiologists: A Practi...	2022-06-20
10	BK26042	Book	Basic Statistics in Criminology and Criminal Justice: Fifth Edition	2021-09-11
11	BK38707	Book	Bayesian Statistics and Marketing	2022-10-01
12	BK96692	Book	Best practices in teaching statistics and research methods in th...	2020-08-10
13	BK91795	Book	BIOEQUIVALENCE and STATISTICS in CLINICAL PHARMAC...	2020-08-30
14	BK74515	Book	Biostatistics decoded: Second edition	2023-04-28
15	BK14146	Book	Biostatistics for Medical and Biomedical Practitioners	2021-07-26
16	BK86088	Book	Bird strike in aviation: Statistics, analysis and management	2021-08-08
17	BK64016	Book	Complex and chaotic nonlinear dynamics: Advances in econo...	2022-08-19
18	BK25996	Book	Computational Statistics Handbook with MATLAB	2020-02-18
19	BK63184	Book	Core statistics	2021-05-29

PRACTICAL 25

QUESTION: Library administration wants to know the details of "Borrow_Dataset" along with inventory type, title and entry data as one table. Provide this table.

OBJECTIVE: To know the details of "Borrow_Dataset" along with inventory type, title and entry data as one table.

SOFTWARE USED: MICROSOFT SQL SERVER MANAGEMENT STUDIO.

PROCEDURE:

1. Open Microsoft SQL Server Management Studio.
2. Use the function ‘left join’ to know the details of “Borrow_Dataset” along with inventory type, title and entry data as one table.
3. Required table is obtained.

RESULT:

CODE:

```
select * into borrow_updated from
 (select a.*, b.Inv_Type, b.Inv_Title, Inv_Entry
 from Borrow_Dataset a left join  Inventory_Master b
 on a.Inv_ID = b.Inv_ID) as c
```

OUTPUT:

Results Messages

	Month	Year	Inv_ID	Days_Borrowed	Inv_Type	Inv_Title	Inv_Entry
1	1	2000	BK12869	5	Book	A first course in probability and statistics	2003-06-19
2	1	2000	BK15517	5	Book	Applied Statistics for Environmental Science with R	2017-03-30
3	1	2000	BK16307	2	Book	Advanced statistics with applications in R	2019-03-17
4	1	2000	BK20605	8	Book	Algebraic statistics: Computational commutative alge...	2013-12-04
5	1	2000	BK38983	7	Book	Advanced Mapping of Environmental Data: Geostati...	2011-08-21
6	1	2000	BK44885	7	Book	Applied Statistics For Economics And Business	2008-12-11
7	1	2000	BK45356	5	Book	Bayesian Statistics 9	2005-08-21
8	1	2000	BK48179	7	Book	Beyond Basic Statistics: Tips, Tricks, and Techniqu...	2015-02-20
9	1	2000	BK52178	4	Book	Biostatistics for radiologists: Planning, performing, an...	2007-11-22
10	1	2000	BK64016	7	Book	Complex and chaotic nonlinear dynamics: Advances...	2022-08-19
11	1	2000	BK83772	4	Book	Advanced medical statistics: 2nd edition	2007-08-30
12	1	2000	BK91882	7	Book	Contributions on Theory of Mathematical Statistics	2007-07-24
13	1	2000	JN146...	10	Journals	Swiss Journal of Economics and Statistics	2013-02-24
14	1	2000	JN155...	9	Journals	Journal of the Royal Statistical Society. Series A, (St...	2016-02-24
15	1	2000	JN176...	1	Journals	Communications in Statistics. Part C: Sequential Ana...	2009-04-01
16	1	2000	JN200...	2	Journals	Chilean Journal of Statistics	2009-03-12
17	1	2000	JN240...	4	Journals	Journal of the Indian Society for Probability and Stati...	2023-02-15

PRACTICAL 26

QUESTION: Use “LibraryDB” database and answer the following:

1. Report the year wise addition to library.
2. How many books are added each year?
3. How many journals are added each year?
4. How many journals were added after 2002?
5. Which inventory type borrowed most?

OBJECTIVE: To use “LibraryDB” database and answer the following:

1. Report the year wise addition to library.
2. To the number of books that are added each year.
3. To know the number of journals added each year.
4. To know the number of journals added after 2002.
5. To know the inventory type borrowed the most.

SOFTWARE USED: MICROSOFT SQL SERVER MANAGEMENT STUDIO.

PROCEDURE:

1. Open the Microsoft SQL Server Management Studio.
2. Use the function ‘count’ to report the year wise addition to library.
3. Use the functions ‘count’, ‘from’, ‘where’ to know the number of books added each year.
4. Use the functions ‘count’, ‘from’, ‘where’ to know the number of journals added each year.
5. Use the functions ‘from’, ‘where’, ‘group’ to know the number of journals added after year 2002.
6. Use function ‘sum’ to know the inventory type borrowed most.

RESULT:

1. CODE:

```
select year(Inv_Entry) as Entry_Year, count(*) as Entry_Count
from Inventory_Master
group by year(Inv_Entry);
```

OUTPUT:

.00 %

Results | Messages

	Entry_Year	Entry_Count
1	2000	12
2	2001	16
3	2002	6
4	2003	11
5	2004	9
6	2005	8
7	2006	9
8	2007	11
9	2008	6
10	2009	11
11	2010	10
12	2011	11
13	2012	6
14	2013	9
15	2014	7
16	2015	6
17	2016	9
18	2017	14
19	2018	12
20	2019	5
21	2020	10
22	2021	14
23	2022	13
24	2023	5

2. CODE:

```
select year(Inv_Entry) as Entry_Year, count(*) as Books_Count
from Inventory_Master
where Inv_Type = 'Book'
group by year(Inv_Entry);
```

OUTPUT:

100 %

Results | Messages

	Entry_Year	Books_Count
1	2000	2
2	2001	9
3	2002	4
4	2003	7
5	2004	4
6	2005	5
7	2006	1
8	2007	6
9	2008	4
10	2009	4
11	2010	6
12	2011	5
13	2012	1
14	2013	4
15	2014	5
16	2015	3
17	2016	5
18	2017	11
19	2018	7
20	2019	4
21	2020	6
22	2021	4
23	2022	8
24	2023	1

3. CODE:

```
select year(Inv_Entry) as Entry_Year, count(*) as Journals_Count
from Inventory_Master
where Inv_Type = 'Journals'
group by year(Inv_Entry);
```

OUTPUT:

Results | Messages

	Entry_Year	Journals_Count
1	2000	10
2	2001	7
3	2002	2
4	2003	4
5	2004	5
6	2005	3
7	2006	8
8	2007	5
9	2008	2
10	2009	7
11	2010	4
12	2011	6
13	2012	5
14	2013	5
15	2014	2
16	2015	3
17	2016	4
18	2017	3
19	2018	5
20	2019	1
21	2020	4
22	2021	10
23	2022	5
24	2023	4

4. CODE:

```
select year(Inv_Entry) as Entry_Year, count(*) as Journals_Count
from Inventory_Master
where Inv_Type = 'Journals' and year(Inv_Entry) > 2002
group by year(Inv_Entry);
```

OUTPUT:

Results Messages

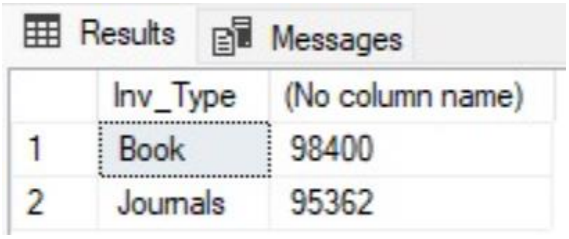
Entry_Year	Journals_Count
2003	4
2004	5
2005	3
2006	8
2007	5
2008	2
2009	7
2010	4
2011	6
2012	5
2013	5
2014	2
2015	3
2016	4
2017	3
2018	5
2019	1
2020	4
2021	10
2022	5
2023	4

uery executed successfully.

5. CODE:

```
select year(Inv_Entry) as Year,
sum(case when Inv_Type = 'Book' Then 1 Else 0 END) as Books_Count,
sum(case when Inv_Type = 'Journals' Then 1 Else 0 END) as Journals_Count
from Inventory_Master
group by year(Inv_Entry);
```

OUTPUT:



Results Messages

	Inv_Type	(No column name)
1	Book	98400
2	Journals	95362