Lecture 3: Regression Analysis by Excel

Tasks for Simple Regression

Use ***Dataset 3***, which indicates a relationship between the mean annual temperature in the region and the dehydration index, to answer the following questions:

1. Is there a relationship between the temperature and the dehydration index?

2. Find out the possible regression line for the two variables mentioned in question 1.

3. Calculate the correlation matrix between the two variables.

4. Display the regression equation and coefficient of determination (R^2) on the chart.

5. Estimate the regression statistics between the dependent variables –dehydration index – and the explanatory variable – temperature – based on the following hypothesis: (i.e., regress temp on dehydration index).

H_0: There is no linear relationship between the dehydration index and the mean annual temperature.

H_I: There is a linear relationship.

Tasks for Multiple Regression

Use Data Set 4, which presents hypothetical information about a relationship between academic ability and its determinants, to answer the following questions:

1. Calculate the correlation matrix between the relevant variables under study.

2. How well can we predict academic ability if we know something about parents' education?

3. Estimate the regression statistics (i.e., the regression equation) between the dependent variables academic ability and its determinants based on the following hypothesis: (i.e., regress pe, r, c, g on aa).

H_0: The coefficients of all variables = 0

H_I: At least one of the coefficient of ≠ 0.

Tips for Simple Regression

1. *Plot your data: A scattered plot can quickly point out obvious problems in assuming that a linear model fits your data. From this information, click chart wizard and choose XY scatter from the list of chart types (see fig 1).*

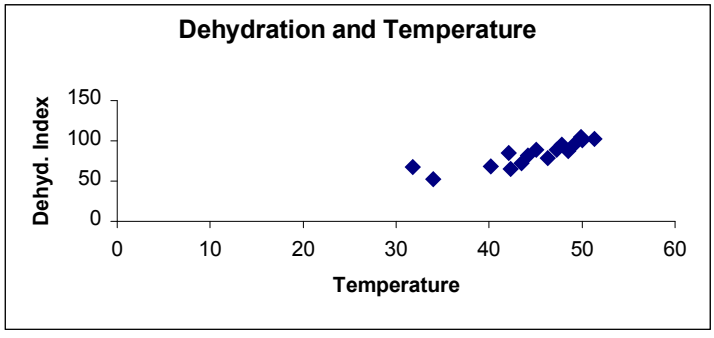

FIG 1

*Should you want to see more clearly, resize the horizontal scale (i.e., say the lower boundary is **30**) and then resize the vertical scale (i.e., the lower boundary is **50**) (see fig 2).*

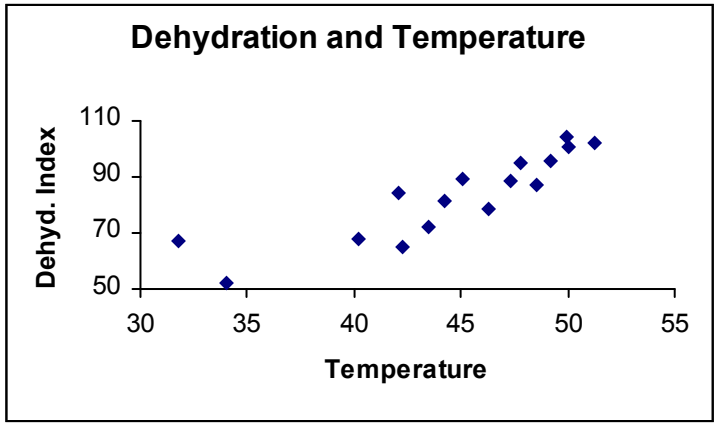

FIG 2

Hint: *To resize both the horizontal scale and vertical scale, follow the steps below: having plotted the original graph, first make the mouse pointer to show horizontal axis numbers then **right click** to choose **format axis > scale**. When you select the **scale** button, type first **30** into **minimum box** and the other information shown in the dialog box below, then press **ok** as in fig 3. Apply the same procedure for vertical axis numbers as in fig* 4.

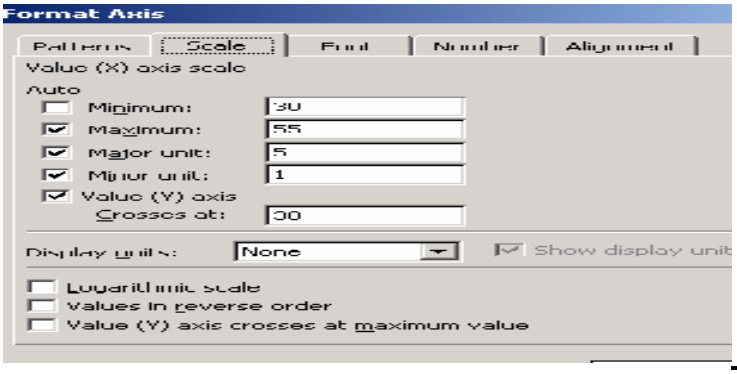

FIG 3

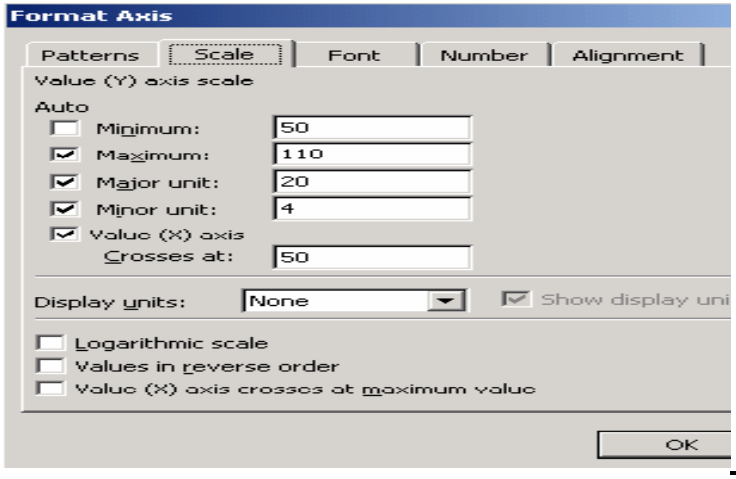

FIG 4

2. *To find out the possible regression line,* ***right click*** *any of the data points in the graph and* ***add trendline*** *from the menu (see fig 5).*

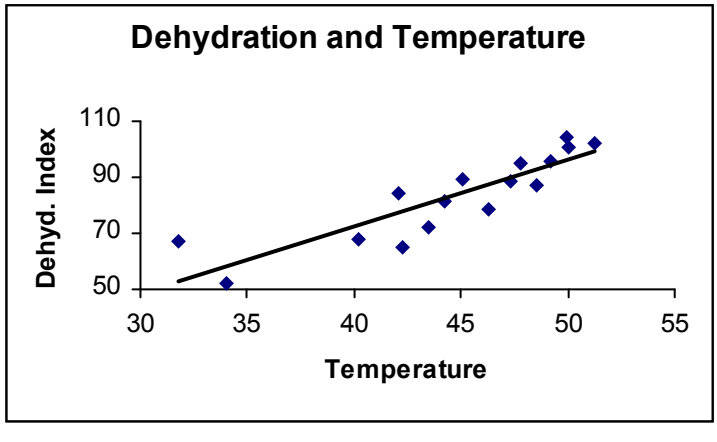

FIG 5

Regression line or trend on the plot gives a pretty good idea whether a straight line fits your data or not.

3. *To create the correlation matrix, choose the option;* ***tools> data analysis> correlation.*** *Make sure you select the two variables at the same time in the* ***input range*** *box as in fig 6 so that this gives the final outcome as appears in fig 7.*

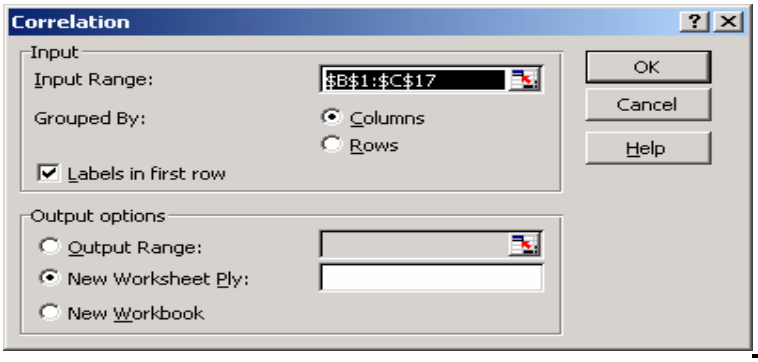

FIG 6

	A	B	C
1		*Temperature*	*Dehydration*
2	Temperatu	1	
3	Dehydratio	0.874854	1
4			

FIG 7

4. *To display the equation and coefficient of determination on the relevant graph, first make the mouse pointer to show any of the data points on the graph then* ***right click*** *to choose* ***format trendline*** *and select* ***option*** *button as appears in fig 8. Finally, you can* ***drag*** *the text containing the* ***regression equation*** *and* ***R^2- value*** *to a point above the plot as in fig 9.*

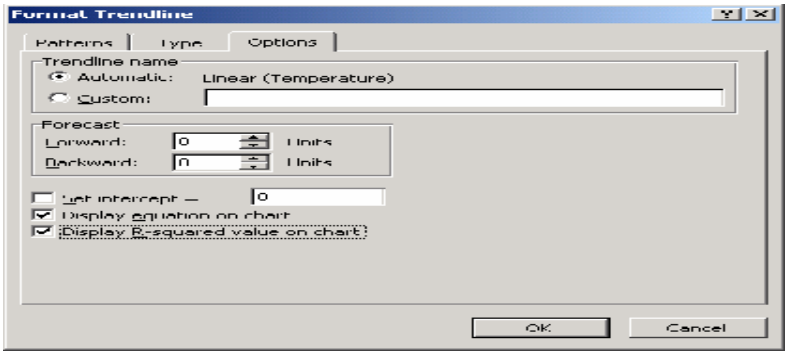

FIG 8

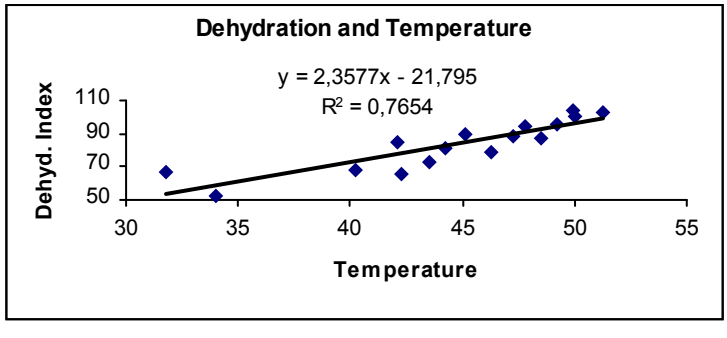

FIG 9

5. To create a table of regression statistics; click **data analysis** and scroll down the analysis tools list box, then select **regression** section and press **ok**. Further, enter the cell range **C1:**

C17 in the input Y range box and **B1: B17** in the input X range box and make the necessary changes appear as shown in fig 10. This gives the final outcome as in fig 11.

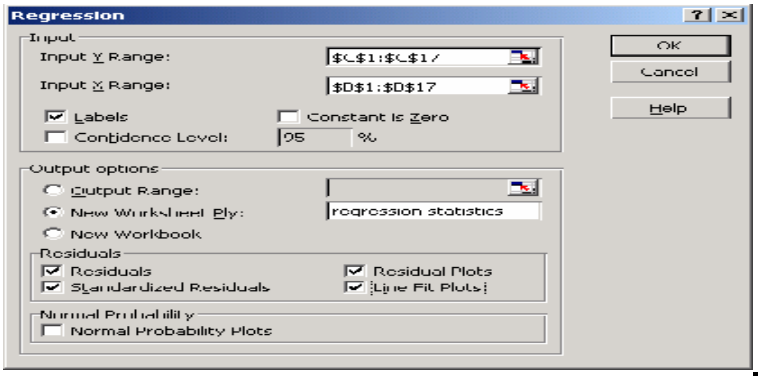

FIG 10

1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.874854				
5	R Square	0.76537				
6	Adjusted R	0.748611				
7	Standard E	7.544657				
8	Observatio	16				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*ignificance F*
12	Regressior	1	2599.534	2599.534	45.66847	9.2E-06
13	Residual	14	796.9058	56.92184		
14	Total	15	3396.439			
15						
16		*Coefficients*	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*
17	Intercept	-21.7947	15.6719	-1.39069	0.186032	-55.4076
18	Temperatu	2.357695	0.348883	6.757845	9.2E-06	1.609415
19						
20						
21						

FIG 11

Conclusion: Reject H_o (null hypothesis) and Accept H_1 based on the hypothesis on page 2

Tips for Multiple Regression

1. *To create the correlation matrix, choose the option;* ***tools> data analysis> correlation.*** *Make sure you select all variables (except* ***ns****) at the same time in the* ***input range*** *box as in fig 12 so that this gives the final outcome as appears in fig 13.*

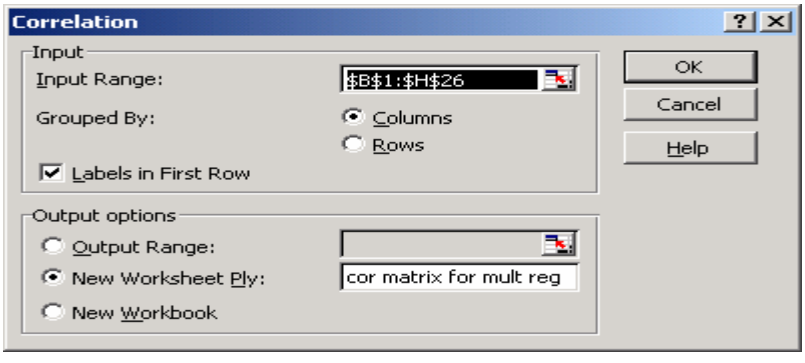

FIG 12

	A	B	C	D	E	F	G	H
1		*aa*	*pe*	*sm*	*ae*	*r*	*g*	*c*
2	aa	1						
3	pe	0.774039	1					
4	sm	0.696307	0.724498	1				
5	ae	0.67323	0.689738	0.737968	1			
6	r	0.167937	-0.03229	0.055414	0.076524	1		
7	g	0.100496	0.007435	0.150005	-0.04583	0.283742	1	
8	c	0.666832	0.530525	0.577415	0.392125	-0.05954	0.174595	1

FIG 13

2. *To determine the relationship between academic ability and parents' education click* ***tools>data analysis*** *and select* ***regression*** *section then press* ***ok****. Further, enter the cell range* ***B1: B26*** *in the input Y range box and* ***C1: C26*** *in the input X range box and make the necessary changes appear as shown in fig 14. This gives the final outcome as in fig 15.*

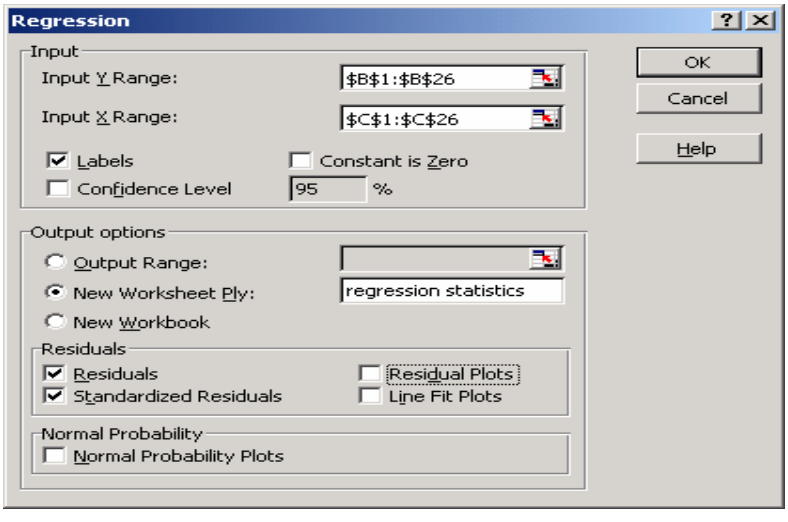

FIG 14

	A	B	C	D	E	F	G	H
1	SUMMARY OUTPUT							
2								
3	*Regression Statistics*							
4	Multiple R	0.7740389						
5	R Square	0.5991362						
6	Adjusted R	0.5817074						
7	Standard E	12.660723						
8	Observation	25						
9								
10	ANOVA							
11		*df*	*SS*	*MS*	*F*	*Significance F*		
12	Regression	1	5510.2799	5510.28	34.3761014	5.63734E-06		
13	Residual	23	3686.7601	160.2939				
14	Total	24	9197.04					
15								
16		*Coefficients*	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*
17	Intercept	2.1386379	11.561621	0.184977	0.85486826	-21.77836572	26.05564	-21.77837
18	pe	4.8921126	0.8343882	5.863114	5.63734E-06	3.166051515	6.618174	3.166052

FIG 15

3. *Apply the same procedure as you did in the previous step and find out the relationship between academic ability and its determinants. This time, select the cell range* ***B1: B26*** *in the input Y range box and* ***C1: H26*** *in the input X range box and make the necessary*

changes appear as shown in fig 16. This gives the final outcome as in fig 17 (This is called ***Multiple Regression****).*

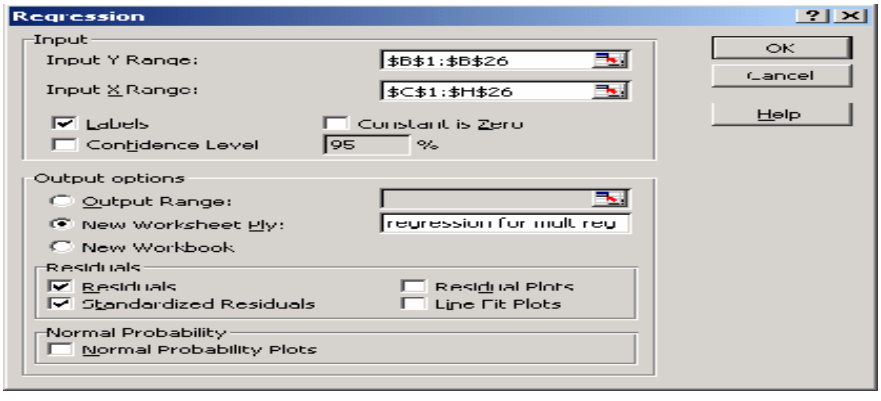

FIG 16

	A	B	C	D	E	F	G	H
1	SUMMARY OUTPUT							
2								
3	*Regression Statistics*							
4	Multiple R	0.870168						
5	R Square	0.757193						
6	Adjusted R	0.676258						
7	Standard E	11.13828						
8	Observatio	25						
9								
10	ANOVA							
11		*df*	*SS*	*MS*	*F*	*ignificance F*		
12	Regressior	6	6963.936	1160.656	9.3555	9.86E-05		
13	Residual	18	2233.104	124.0613				
14	Total	24	9197.04					
15								
16		*Coefficients*	*tandard Err*	*t Stat*	*P value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*
17	Intercept	16.1828	13.24885	1.221449	0.237673	-11.652	44.01762	-11.652
18	pe	2.706120	1.171709	2.377666	0.020711	0.224207	5.247962	0.224207
19	sm	0.147308	5.215291	0.028245	0.977777	-10.8096	11.10424	-10.8096
20	ac	5.157621	4.696086	1.098281	0.286556	4.7085	15.02374	4.7085
21	r	4.664954	3.030462	1.539354	0.141113	-1.70182	11.03172	-1.70182
22	g	-0.43708	4.955235	-0.0882	0.930688	-10.8476	9.973496	-10.8476
23	c	14.43676	5.944122	2.428623	0.025869	1.947982	26.92554	1.947982
24								
25								

regression for mult reg | **mult reg** | school data | cola data | hotel d

FIG 17