

---

## 6.5. Model I: Simple linear regression

In this section we consider the calculations in two BOXES. The first (BOX 6.3.) describes how to calculate a regression line that allows you to model the association or make predictions. In BOX 6.4, the significance of this **association** is tested. Both calculations are illustrated in relation to Example 6.2.

---

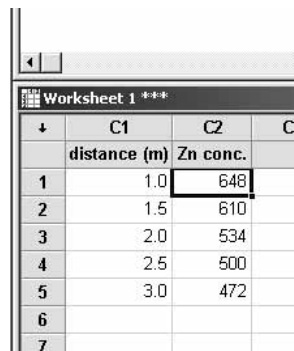
### EXAMPLE 6.2. Heavy metal contamination of soil under electricity pylons

---

**BOX 6.3.** How to carry out a Model I: Simple linear regression: drawing a regression line

**BOX 6.4.** How to carry out a Model I: Simple linear regression: testing the significance of the association

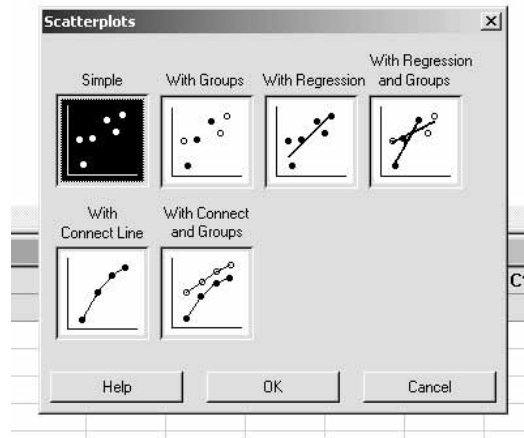
**Step 1.** Enter the data into the Minitab Worksheet window. Give your columns sensible names.



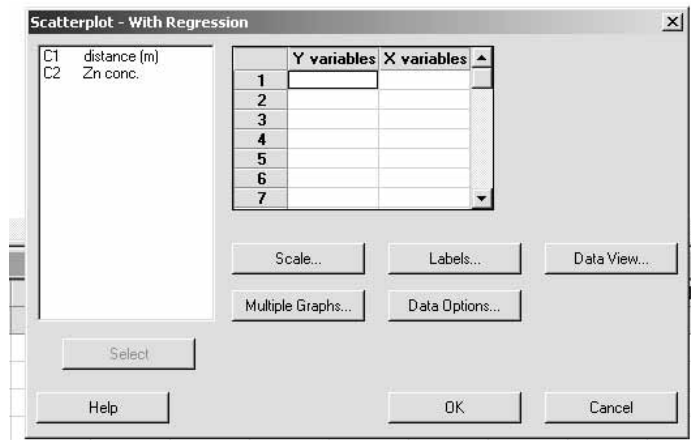
The image shows a screenshot of a Minitab Worksheet window titled "Worksheet 1 \*\*\*". The window contains a table with the following data:

	C1	C2	C3
	distance (m)	Zn conc.	
1	1.0	648	
2	1.5	610	
3	2.0	534	
4	2.5	500	
5	3.0	472	
6			
7			

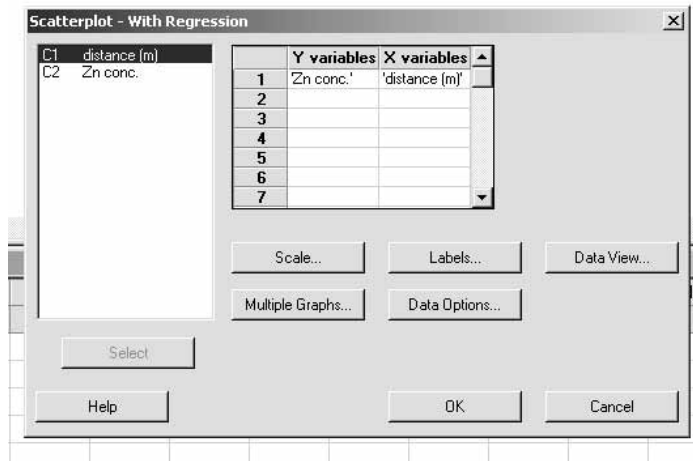
Step 2. Plot the graph. Go to 'Graph', 'Scatterplot'.



Click on 'With Regression', and click on 'OK'.

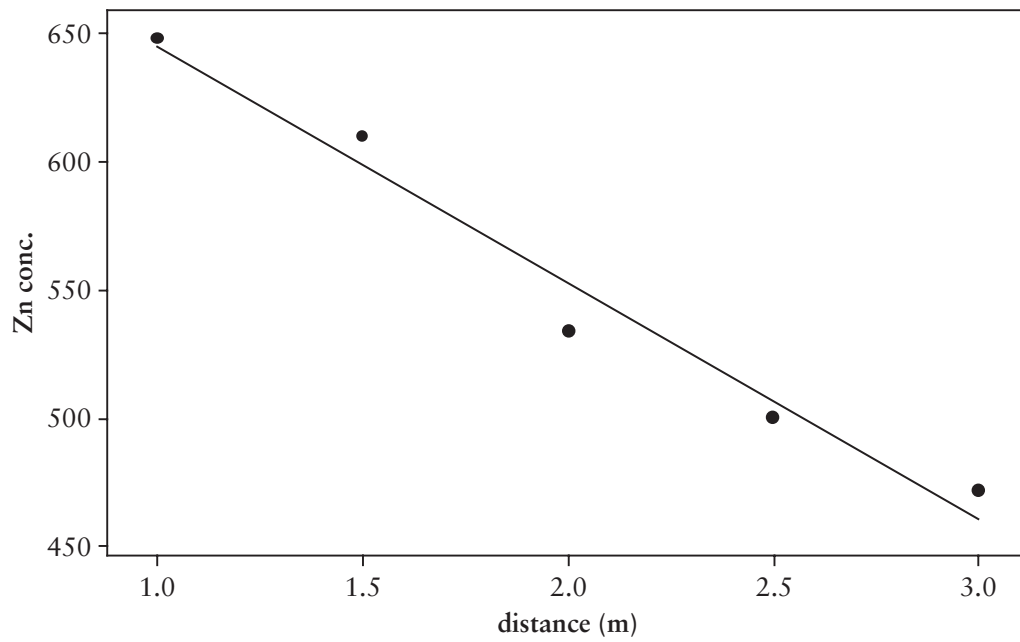


The  $x$  variable is the one you controlled (in this case, the distance from the pylon), and the  $y$  variable is the one you measured (in this case, the zinc concentration). Transfer them across into the 'Y-variables' and 'X-variables' boxes by clicking on them and clicking on 'Select'.



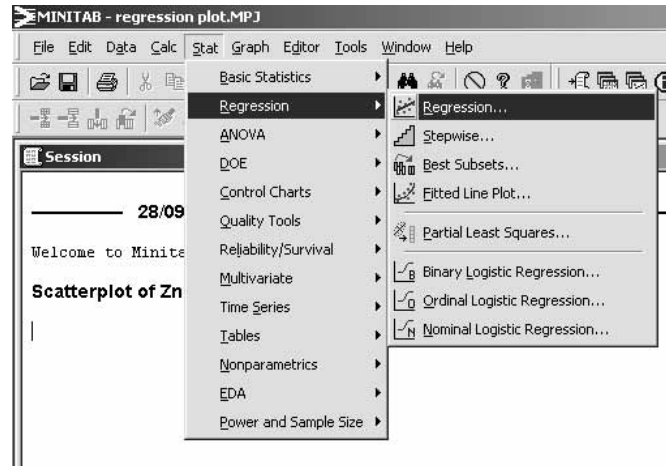
Click on 'OK'. The graph will appear in a separate window.

Scatterplot of Zn conc. vs distance (m)

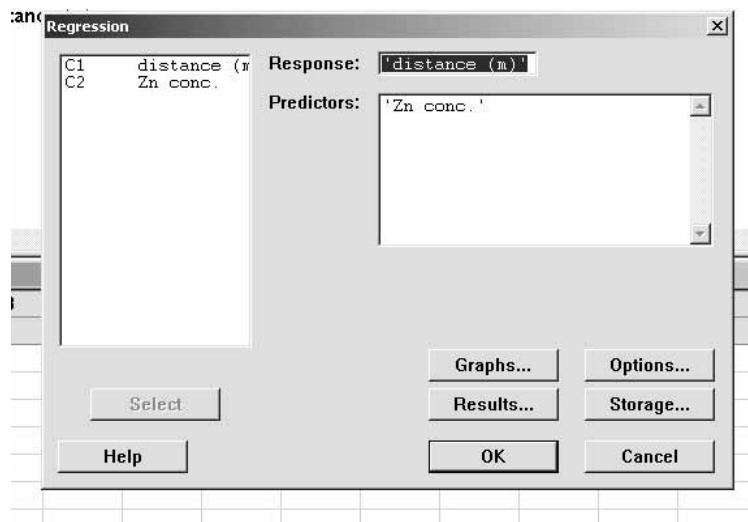


**Step 3.** Calculate the equation of the regression line.

Go to ‘Stat’, ‘Regression’, ‘Regression’.



The response is the quantity you measured, in this case the zinc concentration. Highlight it, and transfer it across to the ‘response’ window by clicking on ‘select’. The predictor is the quantity you controlled, in this case the distance from the pylon at which the measurements were made. Highlight it, and transfer it across to the ‘Predictor’ window by clicking on ‘Select’.



There are all sorts of other things you can do in the ‘Graphs’, ‘Options’, ‘Results’, and ‘Storage’ menus, but the default settings are OK for us now. Click on ‘OK’. The results will appear in the ‘Session’ window.

**Regression Analysis: Zn conc. versus distance (m)**

The regression equation is

$$\text{Zn conc.} = 738 - 92.4 \text{ distance (m)}$$

Predictor	Coef	SE Coef	T	P
Constant	737.60	19.91	37.05	0.000
distance (m)	-92.400	9.384	-9.85	0.002

S = 14.8369      R-Sq = 97.0%      R-Sq(adj) = 96.0%

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	21344	21344	96.96	0.002
Residual Error	3	660	220		
Total	4	22005			

This includes a lot of information that is probably surplus to requirements at present, but the important bit is right at the start: the equation of the regression line is  $\text{Zn conc.} = 738 - 92.4 \text{ distance (m)}$

**Step 4.** Decide the significance of any link between the two variables. See BOX 6.4.

Using the output from the regression calculations (Step 3), we look at the 'Analysis of Variance' part of the results.

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	21344	21344	96.96	0.002
Residual Error	3	660	220		
Total	4	22005			

The  $p$  value here is the significance level for the association. In this case, we have an association which is significant at better than  $p = 0.01$ , but not at  $p = 0.001$ .