

## MATHEMATICS 201-510-LW

Business Statistics

Martin Huard

Fall 2008

# Regression and Correlation with Excel

## Example

Sherlock Holmes thought that he could determine the height of an individual based on the shoe size of the person. To observe this relationship, the shoe size of 10 randomly selected men was noted, along with their heights in centimeters.

Make the usual heading (cells A1:A4) and label sheet 1 appropriately.

In cell A6 write “ Shoe Size ” in bold and in cell B6 “Height (cm)”. Adjust the cell format and the width of the columns so that the words appear in full.

Open the worksheet “Data – Shoe size & Height” from my web site and copy the sample results in cells A7:B16.

## Regression

In cell D6:D9 write the following (adjusting the size of cells accordingly).

### Regression

slope

y - intercept

Least-squares line:

We will now get Excel to calculate this. The functions can be found directly in  $f_x$  -

STATISTICAL. Here is how to find these measures:

Slope

SLOPE(y range, x range)

y – intercept

INTERCEPT(y range, x range)

where the “y range” is the height (cells B7:B16) and “x range” the shoe size (cells A7:A16).

For the equation of line, make cell reference and space the equation in 5 columns (where the column width is adjusted) .

Suppose that on the scene of a crime, Sherlock Holmes discovers footprints for a shoe of size 8. What does the least-squares line forecast as a height?

In cell D11 write “Least squares forecast for shoe size 8”, adjusting the cell accordingly so that all the words appear. In cell E11, use the FORECAST(y range, x range) formula to determine the height. (The answer you should obtain is 171 cm).

## Correlation

Write the following in cells D13:D15.

### Correlation

Coefficient of correlation  $r$

Coefficient of determination  $r^2$

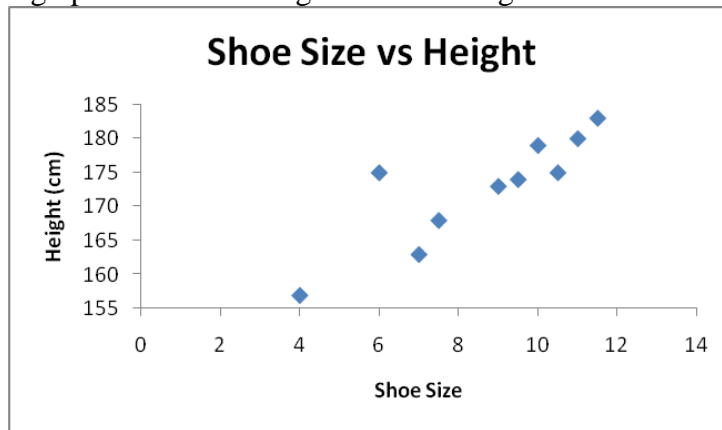
The coefficient of correlation  $r$  is found using CORREL(y range, x range) and the coefficient of determination  $r^2$  is found by squaring the result obtained previously (make Excel do this!).

You should obtain (with proper format of cells).

	A	B	C	D	E	F	G	H
1	<b>Business Statistics</b>							
2	Measures with Excel							
3	By Martin Huard							
4	18-Aug-08							
5								
6	<b>Shoe Size</b>	<b>Height (cm)</b>		<b>Regression</b>				
7	4	157		slope	2.84			
8	7	163		y - intercept	148.28			
9	7.5	168		equation of line	y = 148.28 + 2.84 x			
10	9	173						
11	9.5	174		Least squares forecast for shoe size 8:	171 cm			
12	10.5	175						
13	6	175		<b>Correlation</b>				
14	10	179		Coefficient of correlation $r$	0.858			
15	11	180		Coefficient of determination $r^2$	73.6%			
16	11.5	183						

### Scatter Diagrams

Let us draw a scatter diagram along with the regression line. Use the Chart Wizard in the same way that we made the other graphs. You should get the following:



The last thing we need to do is add the least-squares line. For this, click on the graph once and then go to CHART TOOLS – LAYOUT - TRENDLINE – MORE TRENDLINE OPTIONS. Click on LINEAR, on DISPLAY EQUATION ON CHART and on DISPLAY R-SQUARED VALUE ON CHART. Note: you may have to move the equation and  $r^2$  value a little bit so that the data points and the line are not obstructed. Here is what your graph should look like.



Note: Compare the equation of the line with the one you obtained previously. If all went right, they should be the same.