Ann Arbor Section
ASQ The Global Voice of Quality™

Module 5:
Some Advanced Techniques:
Scatter Diagrams & Linear Regression

Ann Arbor Section
ASQ The Global Voice of Quality™

Scatter Diagrams

- What is it?
  - A plot of two variables to show the relationship between them
- Why is it important?
  - Helps the visual analysis of the relationship between two variables
  - Test the strength of the relationship
  - Basis for further statistical testing

Ann Arbor Section
ASQ The Global Voice of Quality™

Scatter Diagrams

- Creating the Scatter Plot or Diagram
  - Identify which variable is the dependent variable and which is the independent variable
  - Construct a table with the independent (x) variable in the first column and the dependent variable (y) in the second column
  - In Excel, highlight the columns and select Charts and Scatter buttons
  - Title the Plot

Scatter Diagrams

- Analyzing the Scatter Diagram
  - Look at the plot for any strong relationship between the variables
  - Fit a straight line to the plot such that there are an equal number of points on either side of the line
  - Visually decide if the line can predict the dependent variable given the independent variable
  - Conduct additional tests (following slides)

Ann Arbor Section
ASQ The Global Voice of Quality™

- Plot of two variables showing linear relationship

Data:

Intercept = -4

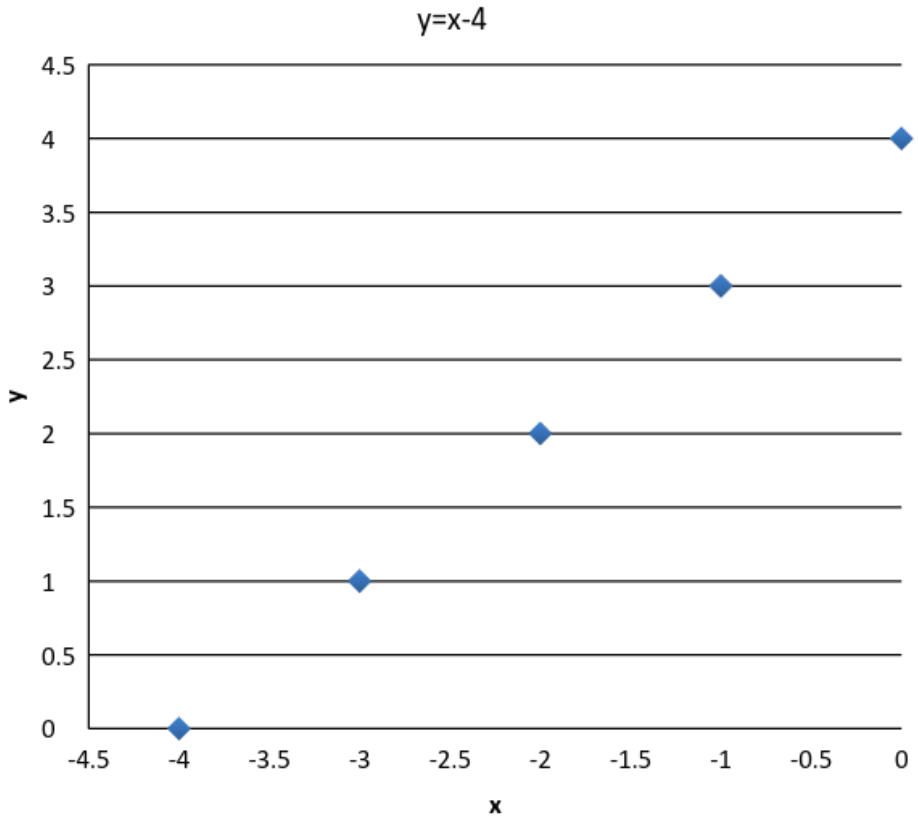
Slope = 1

Pearson Correlation Coefficient r = 1

p value = 0.00001 at alpha = 0.05

y=x-4

x	y
0	4
-1	3
-2	2
-3	1
-4	0



Correlation Coefficient

- Correlation is a technique for investigating the relationship between two quantitative, continuous variables, for example, age and blood pressure.
- Pearson's correlation coefficient (**r**) is a measure of the strength of the association between the two variables.
  - Source: http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1442

Correlation Coefficient

- Using Excel to calculate Pearson's Correlation Coefficient, **r**
  - Establish the data table and plot the Scatter Diagram
  - Select a cell near the table for the result
  - Enter and equal sign into the cell
  - From the menu select Pearson
  - Enter the start and stop cells for the x variable and then for the y variable in the specified format
  - Press Enter and **r** will appear in the cell

Correlation Coefficient

- Alternative Calculation Method
  - Via the Internet, go to the following website:
  - http://www.socscistatistics.com/Default.aspx
  - From the tabs at the top select “Statistical Test Calculators” and then select “Pearson Correlation Coefficient”
  - Follow the website directions remembering that the values you enter for x must be the independent variable and y must be the dependent variable
  - Record the value of **r** then, in the Note, click on the calculator for **p**

Correlation Coefficient

- Interpreting the result:
  - The calculated **r** will lie between -1 and +1
    - If **r**=0, there is no correlation between the variables
    - If **r** equals either a -1 or +1, there is very high correlation between the variables
    - The sign of **r** (- or +) shows a negative (-) or positive (+) correlation; i.e., as the x variable increases the y variable decreases (-) or increases (+)
    - Unless R**r**=-1 or **r**=+1 the significance of the **r** value needs to be tested

Correlation Coefficient

- Testing the Significance of the Relationship
  - You can see from **r** that if the value is near either -1 or +1 that the correlation is quite good.
  - Similarly, if **r** is close to 0 then the correlation is weak or poor
  - For other values of **r** we need to determine whether the value is significant or not
    - For this we must determine the value of **p**
      - This is found on the website http://www.socscistatistics.com/Default.aspx

Correlation Coefficient

- Determining the **p** value for the **r** statistic
  - This is found on the website http://www.socscistatistics.com/Default.aspx
  - After calculating **r** using the website calculator click on the Note text to calculate **r**
    - Enter the **r** value just calculated
    - Enter the sample size (**n**)
    - Select the Significance Level (usually 0.05)
    - Click “calculate” to obtain the **p** value and the statement of significance

Ann Arbor Section
The Global Voice of Quality™

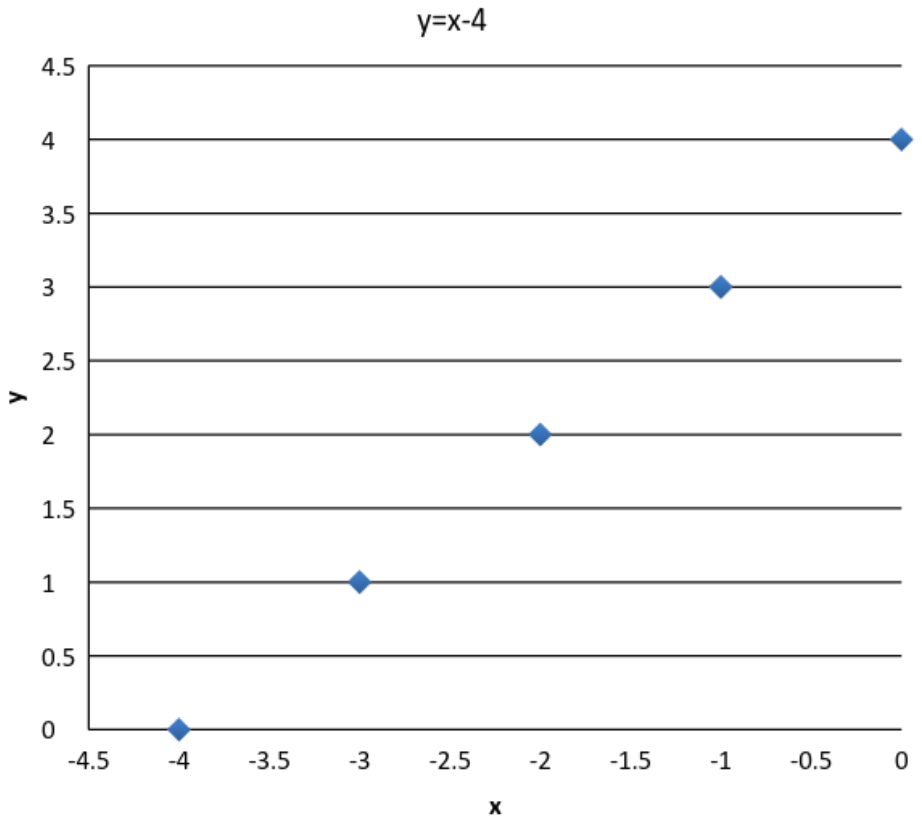
• Example of highest positive correlation

Data:

Intercept (b) = -4
Slope (m) = 1
y = 1x-4 = x-4

Pearson Correlation Coefficient r = 1
p = 0.00001 at alpha = 0.05
Highly significant correlation

X	Y
0	4
-1	3
-2	2
-3	1
-4	0



Ann Arbor Section
ASQ The Global Voice of Quality™

Data:

Intercept (b) = -114.75

Slope (m) = 1.06

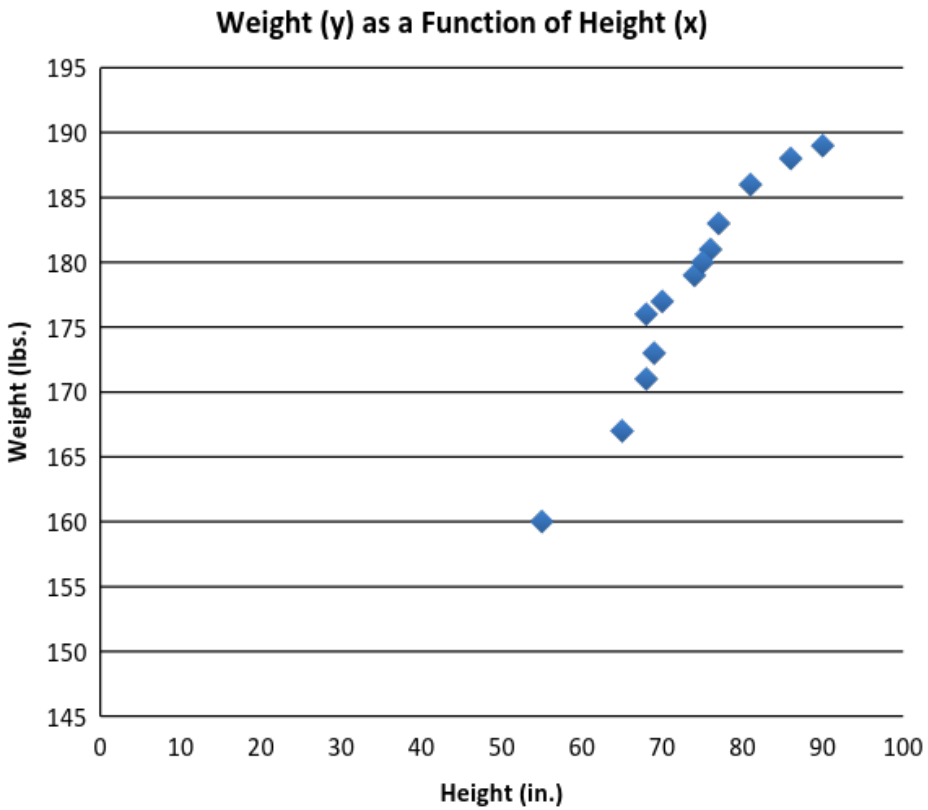
Pearson Correlation Coefficient

r = 0.97

p = 0.00001 at alpha = 0.05

Height	Weight
69	173
68	171
90	189
65	167
77	183
76	181
74	179
55	160
70	177
75	180
86	188
81	186
68	176

- Positive Linear Correlation High



Ann Arbor Section
ASQ The Global Voice of Quality™

- Negative Linear Correlation High

Data:

Intercept (b) = 109.69

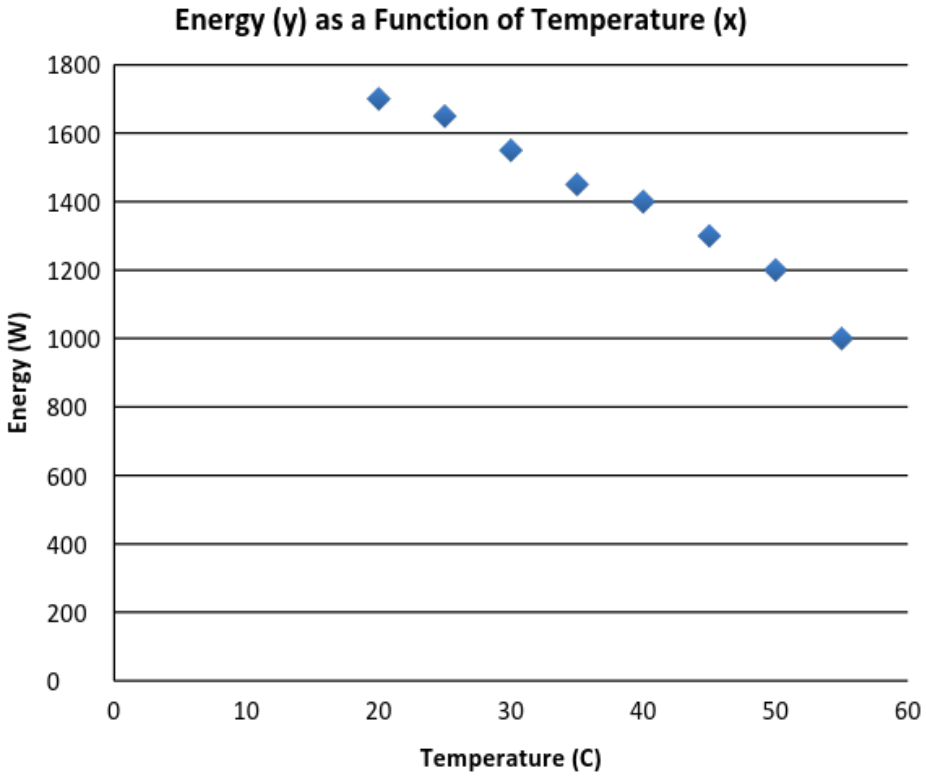
Slope (m) = -0.051

Pearson Correlation Coefficient

rR = -0.986

p = 0.00001 at alpha = 0.05

Temp. (F.)	Energy (W)
55	1000
50	1200
45	1300
40	1400
35	1450
30	1550
25	1650
20	1700



Ann Arbor Section
ASQ The Global Voice of Quality™

• Poorly Correlated Data

Data:

Intercept = 16.79

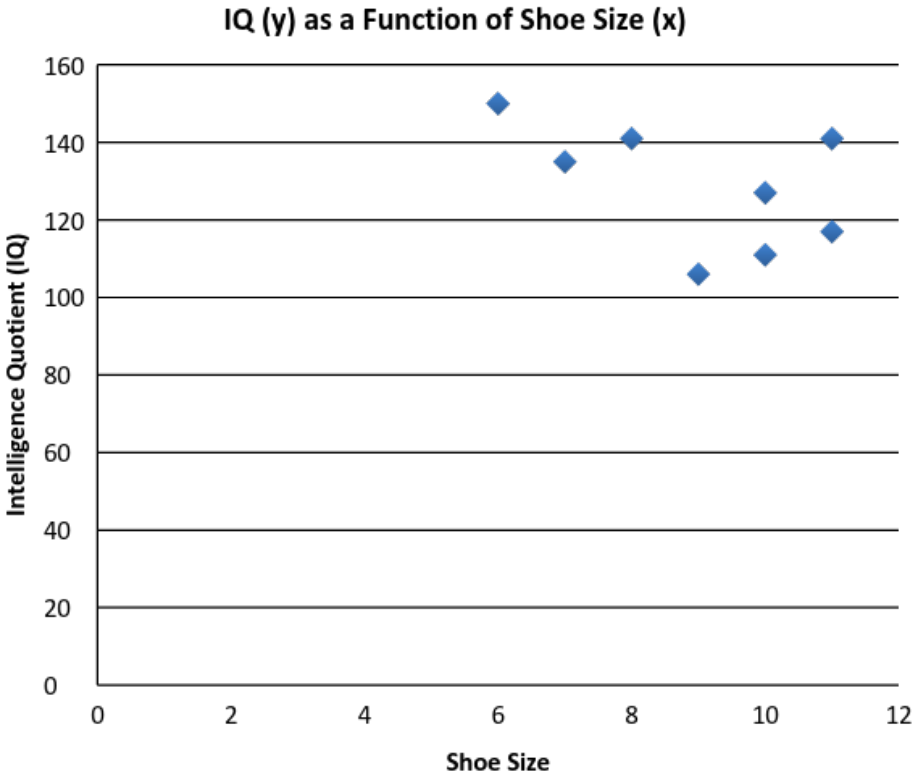
Slope (m) = -0.060

Pearson Correlation Coefficient

r = -0.520

p = 0.187 at alpha 0.05

Shoe Size	IQ
8	141
10	127
11	117
6	150
11	141
10	111
9	106
7	135



Correlation Coefficient

- Online references for this module:
  - www.socscistatistics.com (has statistical calculators for most common statistics)
  - www.mathisfun.com
  - www.learntech.uwe.as.uk (definitions of **r** and **p** in correlation calculations)
  - www.statisticshowto.com
  - www.onlinestatbook.com

Correlation Coefficient

- What to watch for:
  - The calculations and graphs can be made with any data but care must be taken to ensure that the correlation is logical.
  - Do not try to fit a linear regression to data whose plotted shape is not obviously linear; e.g., data in the shape of a parabola or other high order equation.
  - Note that correlation does not imply the cause