Probability and Statistics

Kinds of variables

Statistical correlation

Lecture 5.
Class 1.
Time: 8:30- 10:30
Department: Bit

Types/Classifications of Variables

A. Qualitative
B. Quantitative
    1. Discrete
    2. Continuous

Qualitative Data

- Describes the quality
- Non-numerical format
  Counts العد او الاعداد
  Cannot order or measure
- Examples
  - gender
  - marital status
  - geographical region
  - job title….

Quantitative Data

- Frequencies
- Measurements

Discrete

- Measurements القياسات are integers
- Examples:
  - number of employees of a company
  - number of incorrect answers on a test
  - number of participants in a program…

Continuous

- Measurements can take on any value - usually within some range
- Examples:
  - Age
  - Income
- Arithmetic operations such as differences and averages make sense.

Examples

Qualitatiave or Quantitative?

Discrete or Continuous?

- Score on a placement exam
- Preferred restaurant
- Dollar amount of a loan مبلغ الدولار من القرض
- Height
- Salary
- Length of time to complete a task
- Ethnic origin اصل عرقي

Analysis
Qualitative Data

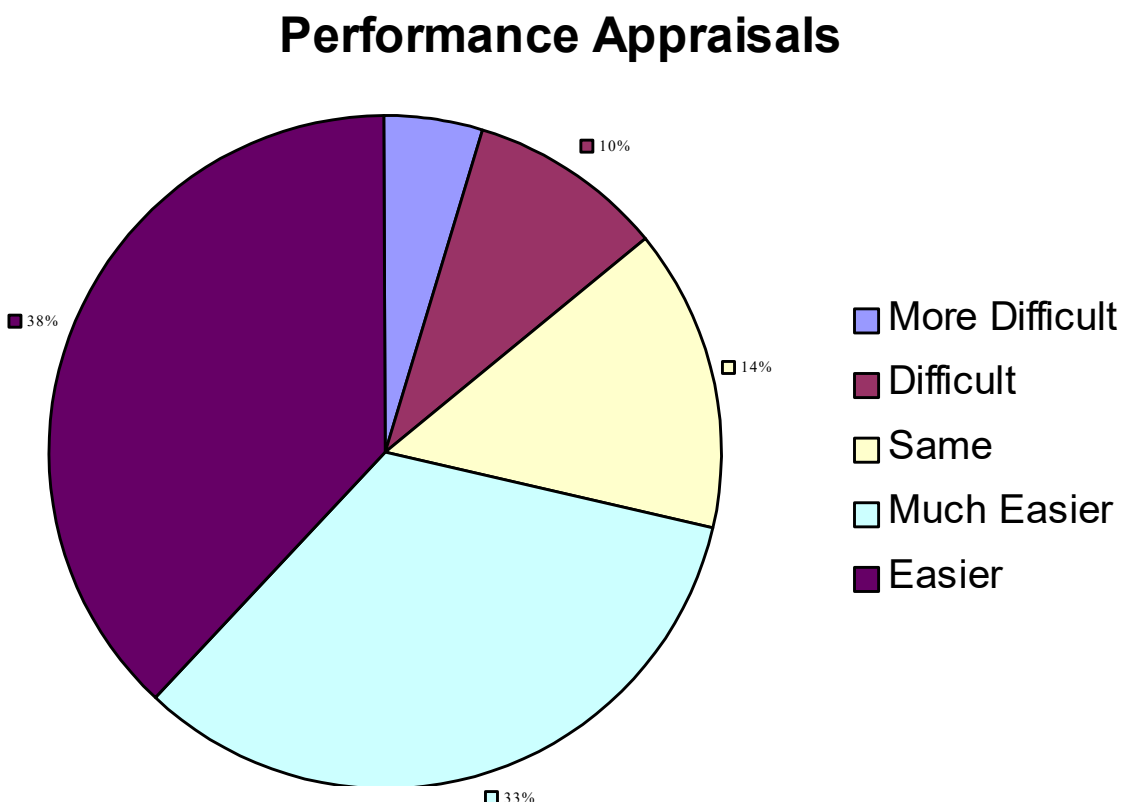
- Frequency tables
- Modes - most frequently occurring
- Graphs: Bar Charts and Pie Charts

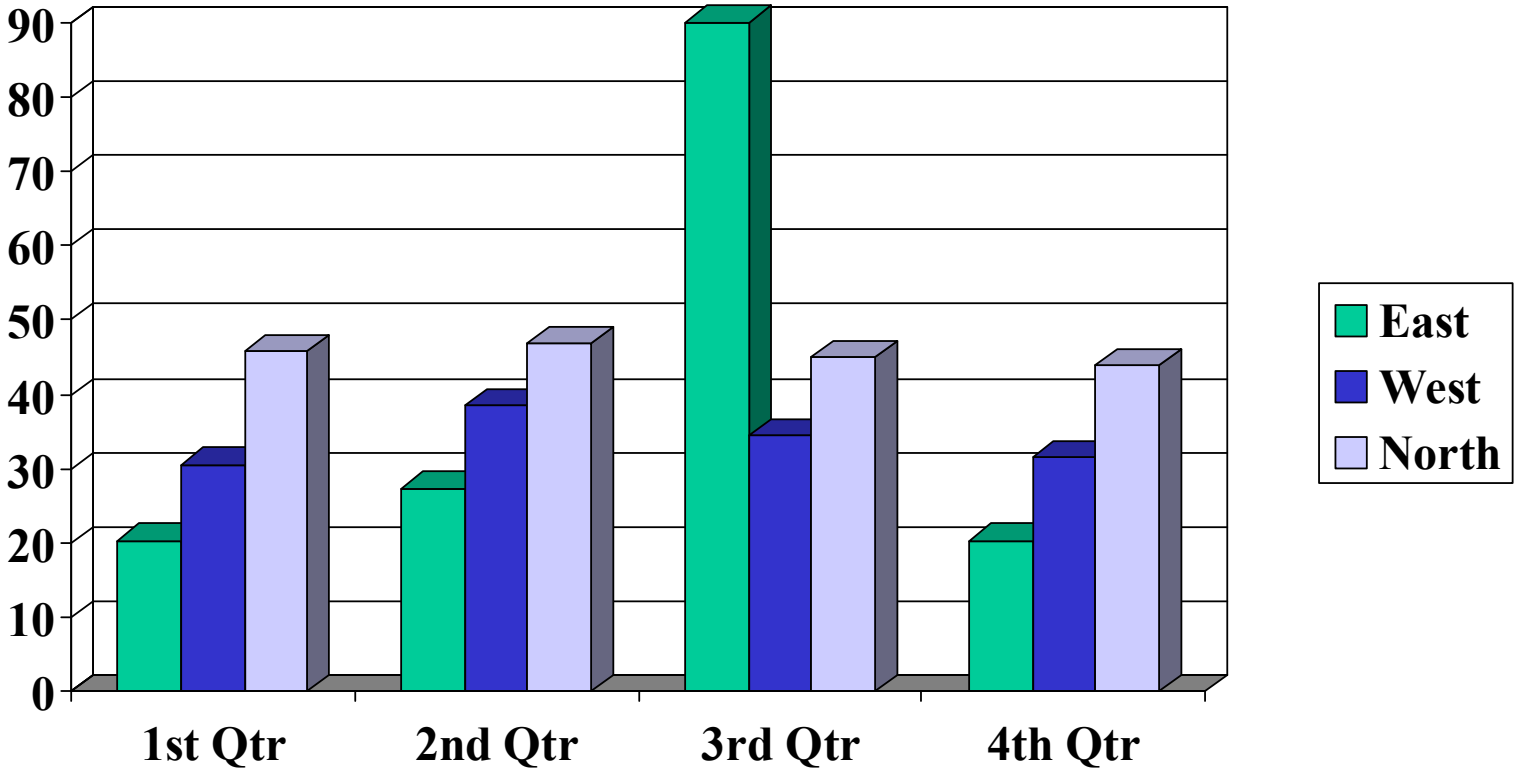
Analysis
Quantitative Data

- Any form
- Create groups or categories and generate frequency tables
- All descriptive statistics

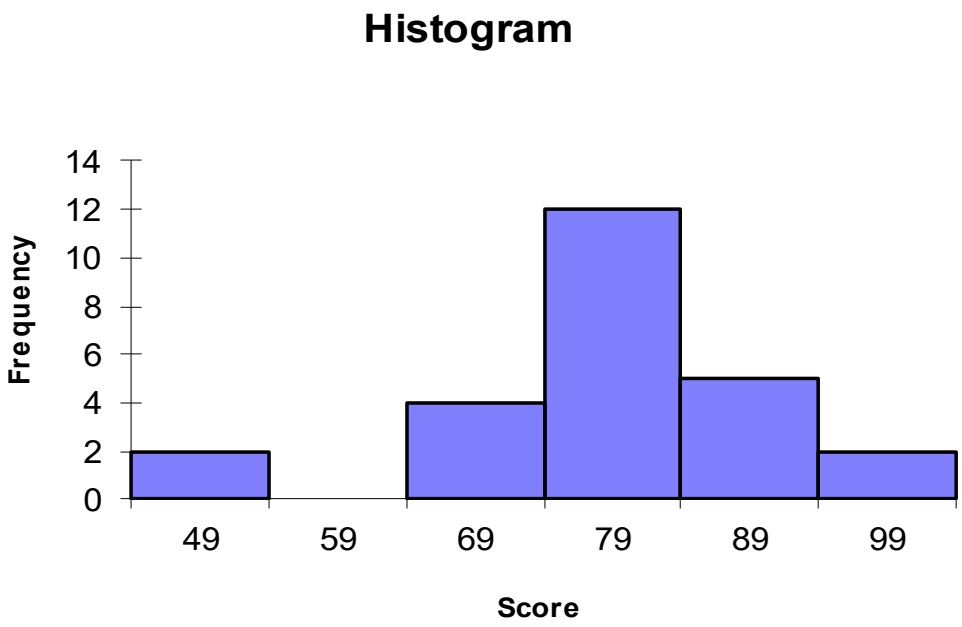
Examples of Graphs

Pie Chart

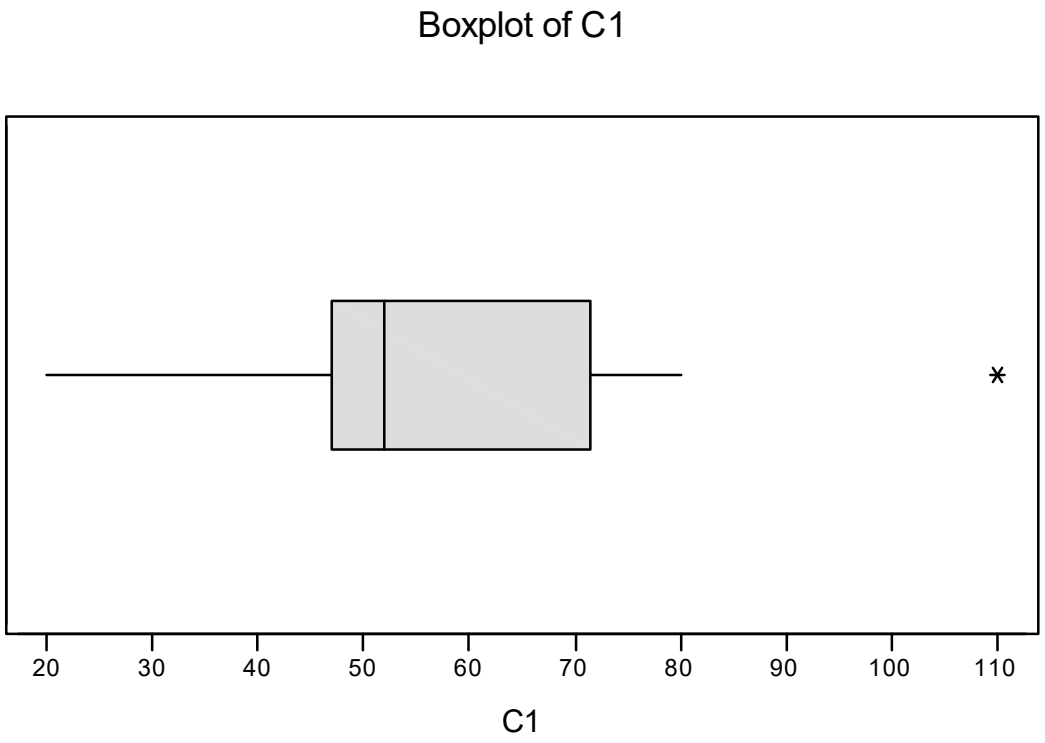

Performance Appraisals
10%
14%
38%
33%
More Difficult
Difficult
Same
Much Easier
Easier


90
80
70
60
50
40
30
20
10
0
1st Qtr
2nd Qtr
3rd Qtr
4th Qtr
East
West
North

Histogram



Boxplot



Bivariate Correlation

Introduction

Bivariate correlations measure the degree of association between two variables. If the two variables are continuous, the Pearson product moment correlation is an appropriate measure. If they are not continuous (that is, if they are discrete or categorical), it would be more appropriate to use Spearman's.

Bivariate Correlation

The correlation coefficient, which ranges from -1 to +1 is both a measure of the strength of the relationship and the direction of the relationship. A correlation coefficient of 1 describes a perfect relationship in which every change of +1 in one variable is associated with a change of +1 in the other variable. A correlation of -1 describes a perfect relationship in which every change of +1 in one variable is associated with a change of -1 in the other variable. A correlation of 0 describes a situation in which a change in one variable is not associated with any particular change in the other variable. In other words, knowing the value of one of the variables gives you no information about the value of the other.

قياس الارتباط

والجدول التالي يوضح أنواع الارتباط **واتجاه العلاقة وشكل الانتشار** لكل نوع :

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
من 0.70 إلى 0.99	ارتباط طردي قوي
من 0.50 إلى 0.69	ارتباط طردي متوسط
من 0.01 إلى 0.49	ارتباط طردي ضعيف
0	لا يوجد ارتباط

وما قيل عن الارتباط الطردي ينطبق على الارتباط العكسي (مع وضع إشارة سالبة)

معامل بيرسون للارتباط الخطي
Pearson linear correlation coefficient

- معامل بيرسون للارتباط الخطي من اكثر معاملات الارتباط استخداما خاصة في العلوم الإنسانية والاجتماعية.
- ومستوى القياس المطلوب عند تطبيق معامل بيرسون للارتباط هو ان يكون كلا المتغيرين مقياس فترة او نسبي او بمعنى اخر ان تكون بيانات كلا المتغيرين (الظاهرتين) بيانات كمية

يمكن حساب معامل بيرسون للارتباط الخطي بدلالة القراءات لبيانات المتغيرين X و Y باستخدام الصيغة التالية

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

حيث :

$\sum_{i=1}^{n} x_i y_i$: مجموع حاصل ضرب x في y

$\sum x$: مجموع قيم المتغير x

$\sum y$: مجموع قيم المتغير y

$\sum x^2$: مجموع مربعات قيم المتغير x

$\sum y^2$: مجموع مربعات قيم المتغير y

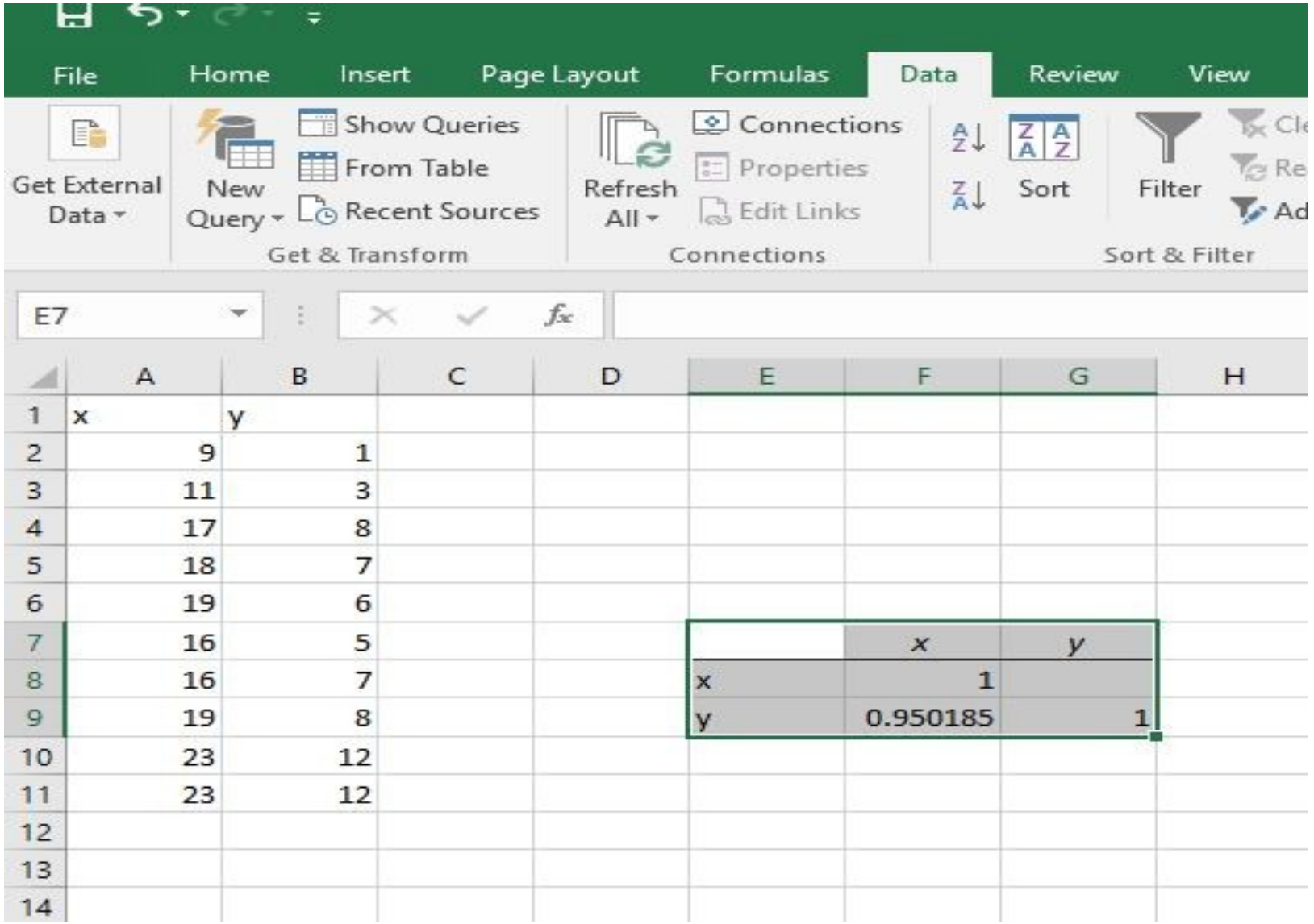
مثال

لدراسة علاقة الصادرات بالميزان التجاري خلال عدة سنوات اخذنا عشر قراءات تقريبية لقيمة صادرات احدى البلدان (X) وقيمة الميزان التجاري (Y) وكما يلي:

12	12	8	7	5	6	7	8	3	1	Y
23	23	19	16	16	19	18	17	11	9	X

احسب معامل الارتباط الخطي، ما مدى قوة العلاقة الخطية؟

X	Y	XY	X^2	Y^2
9	1	9	81	1
11	3	33	121	9
17	8	136	289	64
18	7	126	324	49
19	6	114	361	36
16	5	80	256	25
16	7	112	256	49
19	8	152	361	64
23	12	276	529	144
23	12	276	529	144
171	69	1314	3107	585



	A	B	C	D	E	F	G	H
1	x	y						
2	9	1						
3	11	3						
4	17	8						
5	18	7						
6	19	6						
7	16	5				*x*	*y*	
8	16	7			x	1		
9	19	8			y	0.950185	1	
10	23	12						
11	23	12						
12								
13								
14								

مـن الملاحـظ ان علاقـة الارتباط الخطـي بيـن قيمـة صـادرات البلـد وقيمـة الميزان التجاري موجودة وهـــي علاقـــة ارتباط طرديـة قويـة

What is a correlation

Scatter diagrams can be used to examine the relationship before we start more detailed calculations and judge what kind of relationship we are looking at.

Be aware that relationships may not be linear

- Correlation provides a measure of (linear or nonlinear) association between variables (how they move together)

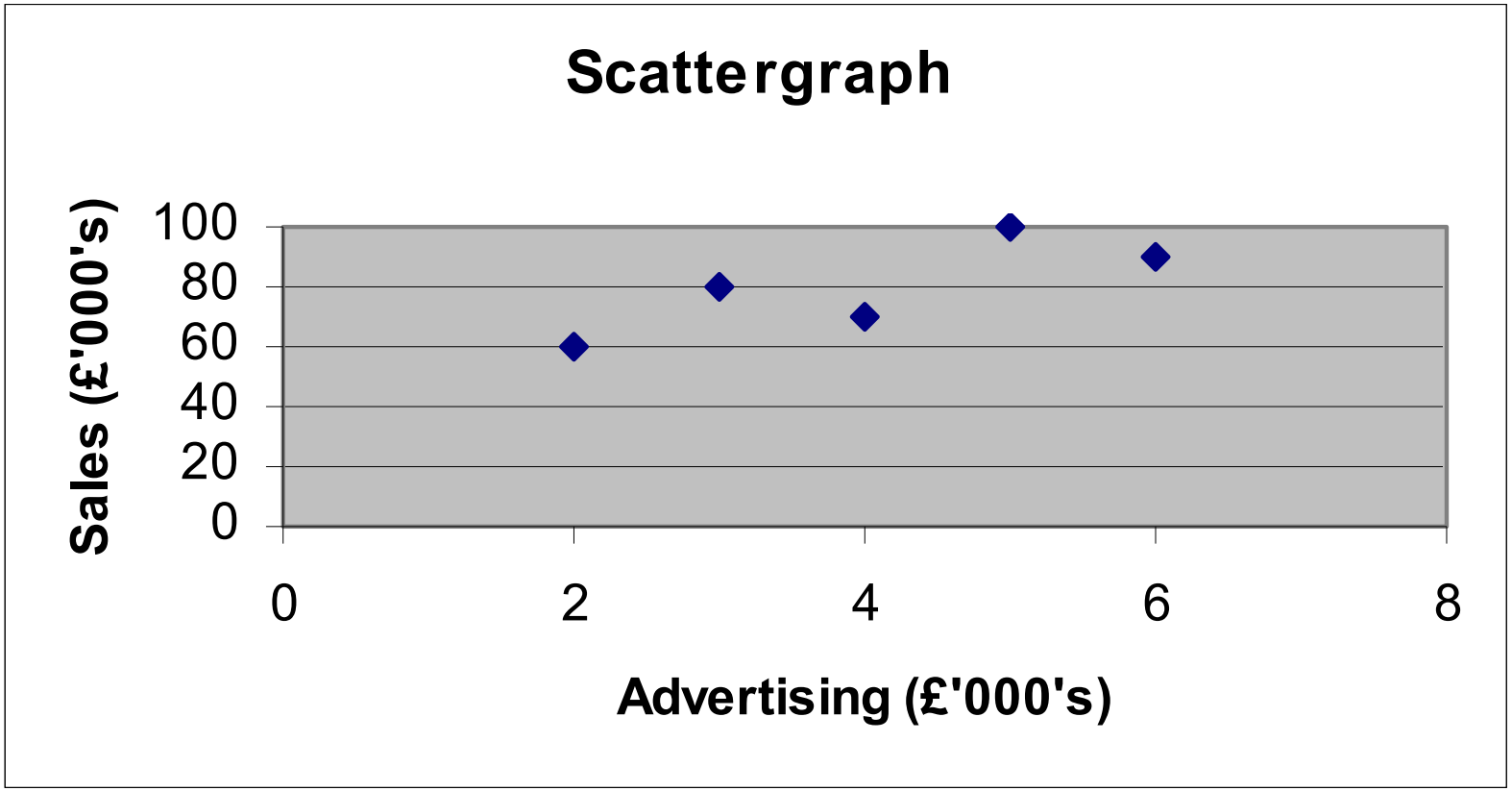
- Correlation will always be between -1 and +1 and can help decide whether there is a likely relationship or not

- Correlation is not proof of cause and effect

For example

X :2 : 3 : 4 : 5 :6

Y: 60 :80 :70 :100 :90



- We would expect a positive correlation as the points lie on an upward sloping line.
- We would expect a negative correlation as the points lie on an downward sloping line
- The formula:

$$-1 \leq r \leq +1$$


وقل ربي زدني علماً
UNIVERSITY OF INFORMATION TECHNOLOGY AND COMMUNICATIONS


جامعة تكنولوجيا المعلومات والاتصالات
University of Information Technology & Communications
2014
Businesses Informatics College
كلية معلوماتية الاعمال

THANK YOU