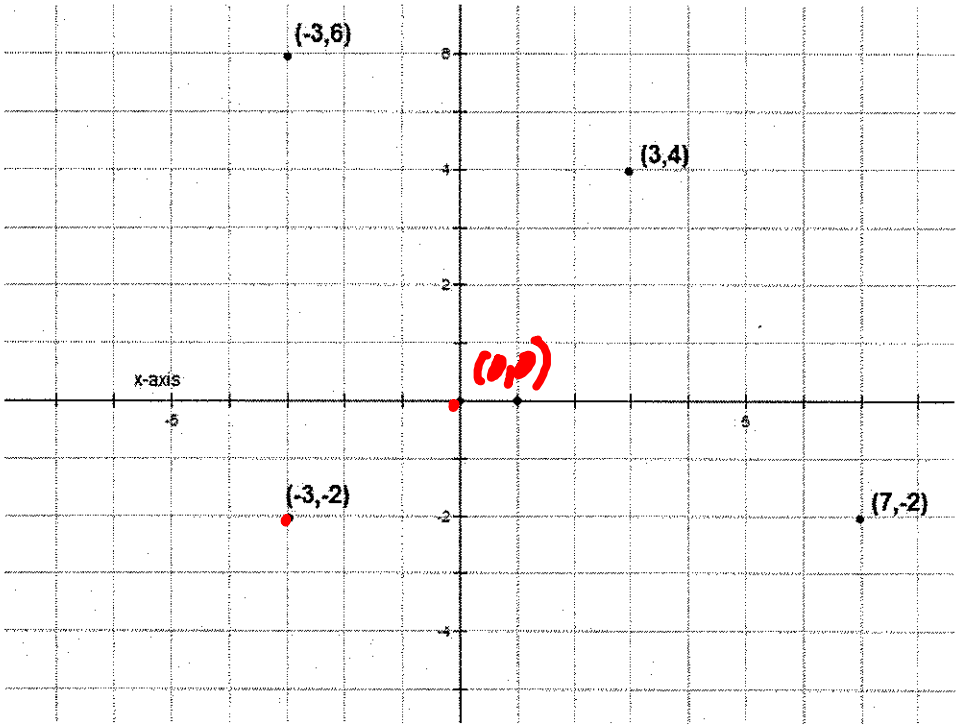
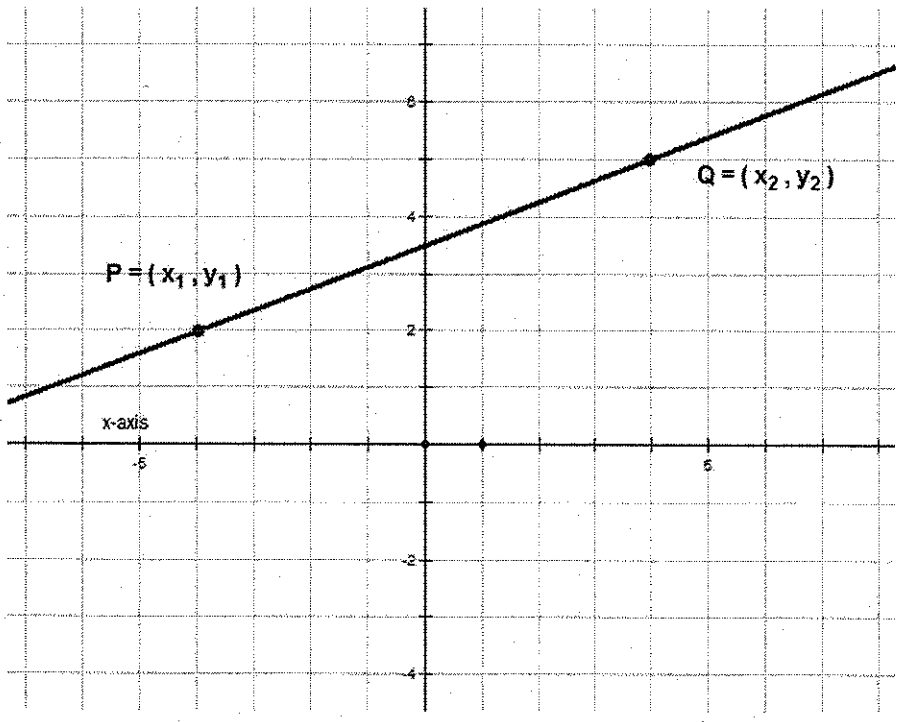
chapter 7

Review of Linear Equations:

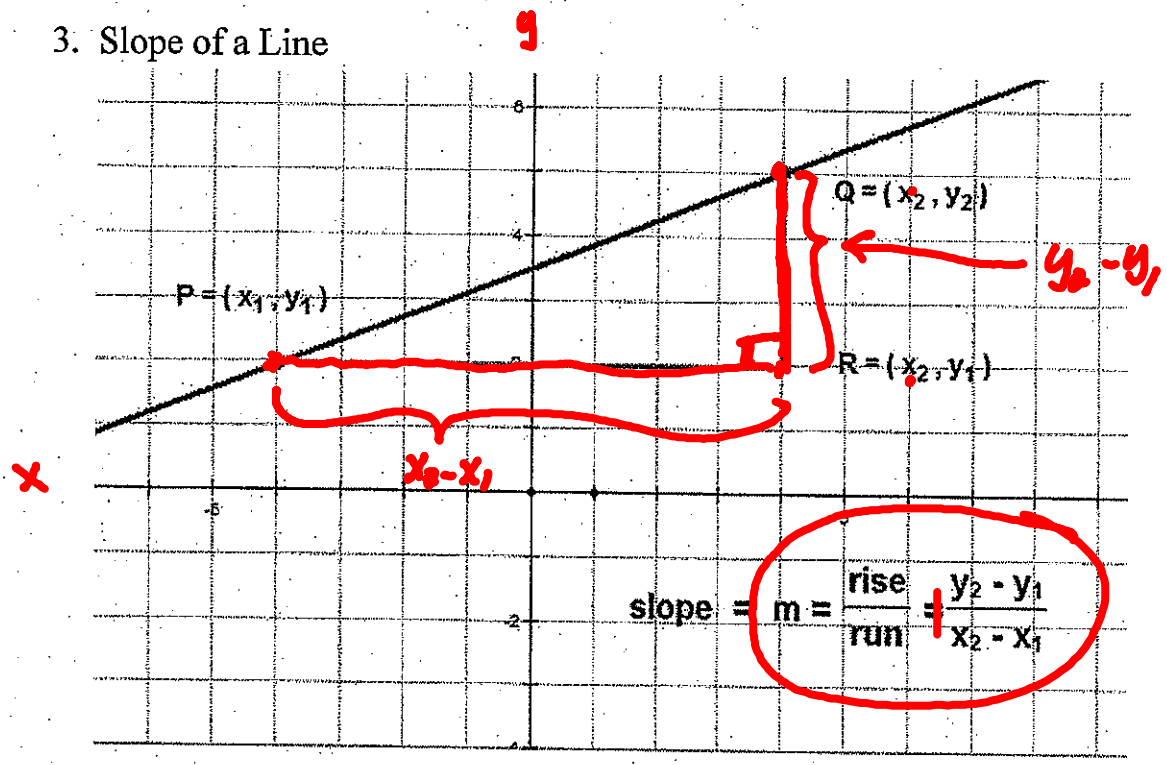
1. Cartesian Coordinate System

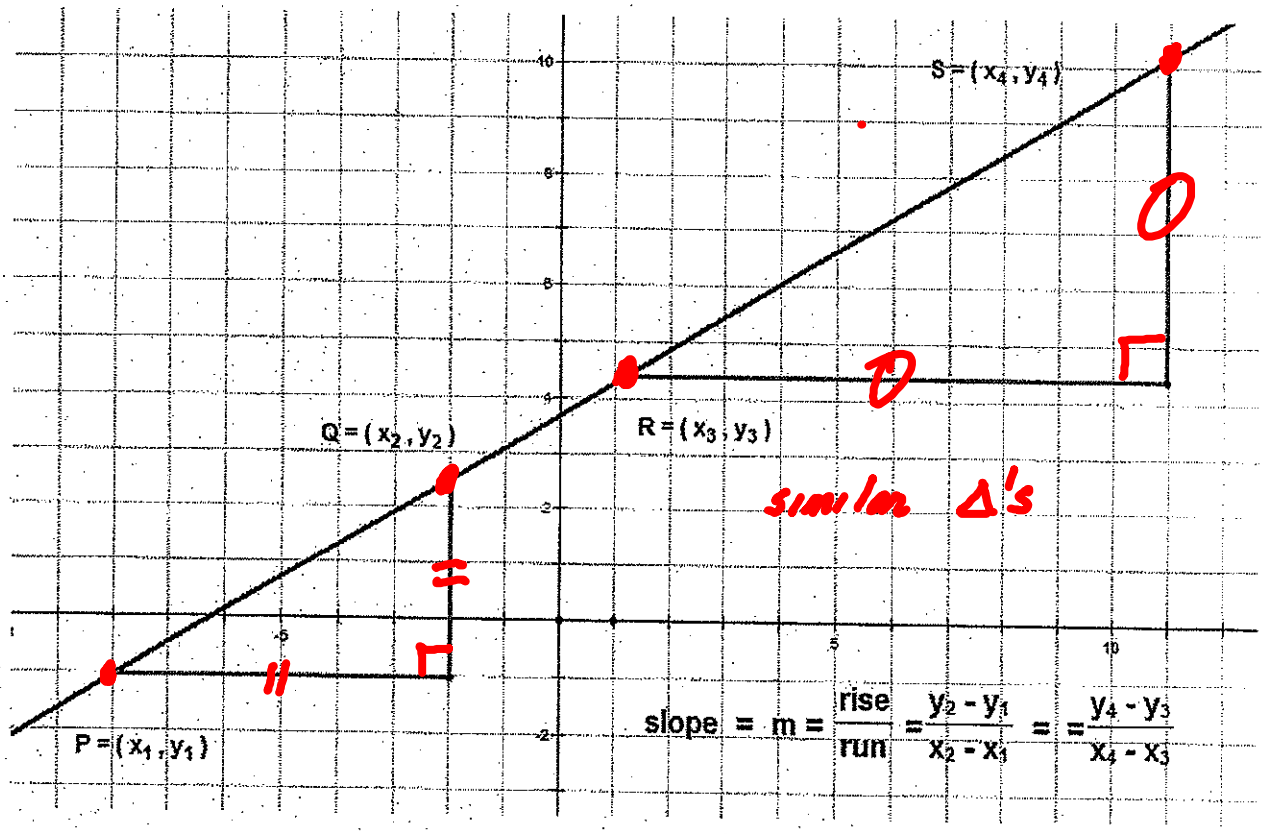


2. Two Points Determine a Line

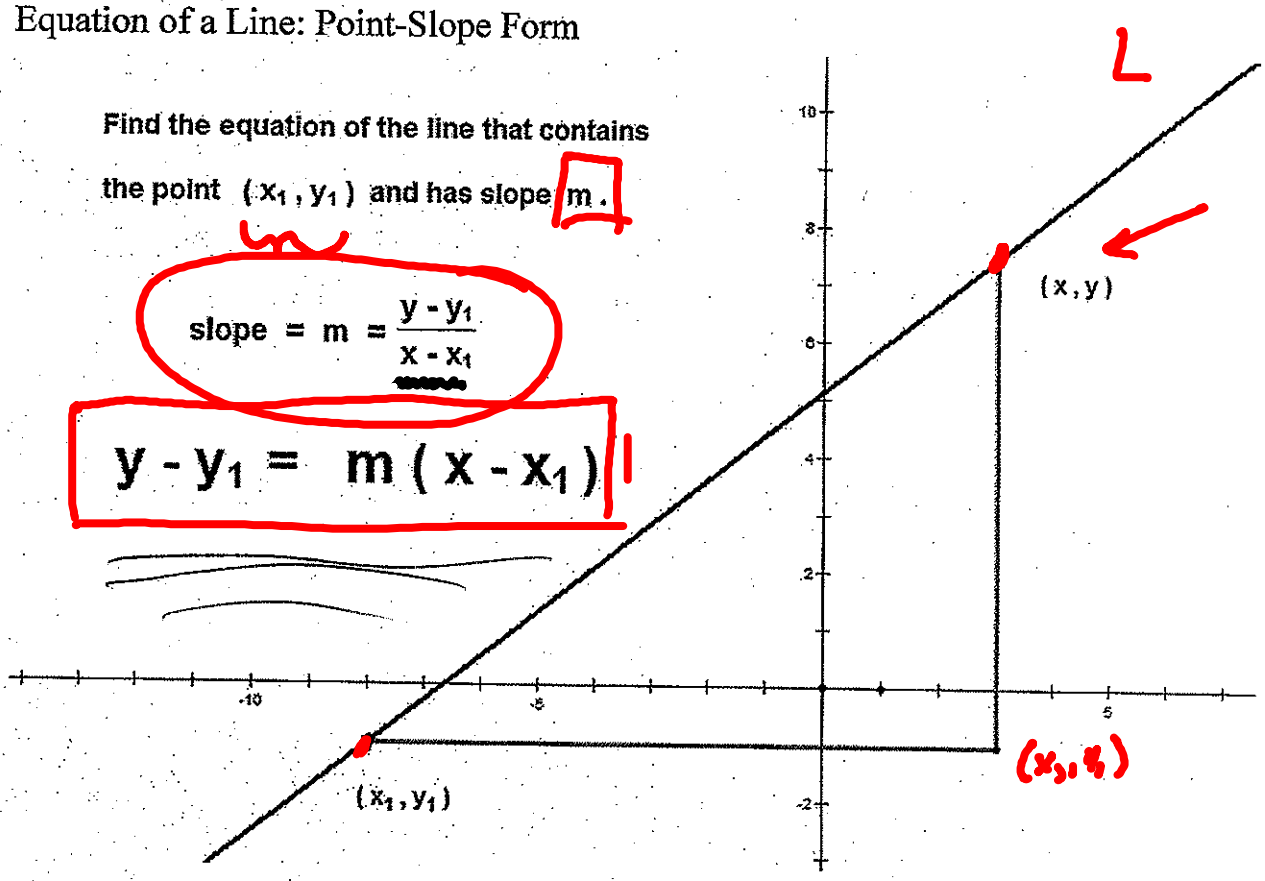


3. Slope of a Line





4. Equation of a Line: Point-Slope Form



5. Find the equation of the line containing the points (-3 , 5) and (1 , -7). Find the y-intercept and the x-intercept.

$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{-7-5}{1-(-3)} = \frac{-12}{4} = -3$

or $\frac{5-(-7)}{-3-1} = \frac{12}{-4} = -3$

$y - y_1 = m(x - x_1)$

$y - 5 = -3(x+3)$, $y - 5 = -3x - 9$

$y = -3x - 4$,

(0, -4)

when $x = 0$, $y = -4$

when $y = 0$, $0 = -3x - 4$

$-3x = 4$

$(-\frac{1}{3}, 0)$ $x = -\frac{1}{3}$

Scatter Diagram

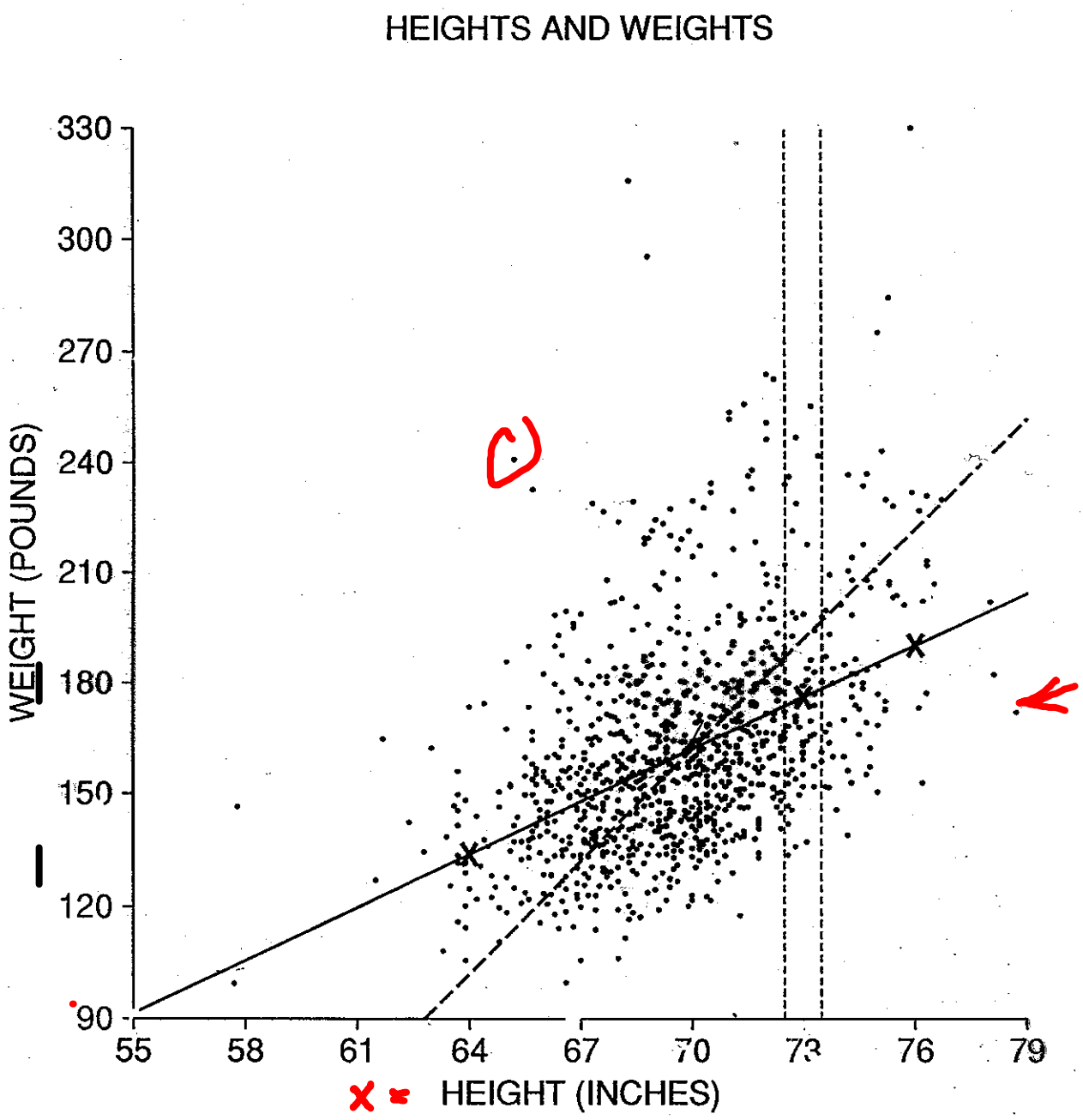


Table 3.5 Indices of cigarette smoking and lung cancer mortality for English occupational groups

Occupational group	Smoking	Mortality
Farmers, foresters, and fisherman	77	84
Miners and quarrymen	137	116
Gas, coke, and chemical makers	117	123
Glass and ceramics makers	94	128
Furnace, forge, foundry, and rolling mill workers	116	155
Electrical and electronics workers	102	101
Engineering and allied trades	111	118
Woodworkers	93	113
Leather workers	88	104
Textile workers	102	88
Clothing workers	91	104
Food, drink, and tobacco workers	104	129
Paper and printing workers	107	86
Makers of other products	112	96
Construction workers	113	144
Painters and decorators	110	139
Drivers of stationary engines, cranes, etc.	125	113
Laborers not included elsewhere	133	146
Transport and communications workers	115	128
Warehousemen, storekeepers, packers, and bottlers	105	115
Clerical workers	87	79
Sales workers	91	85
Service, sport, and recreation workers	100	120
Administrators and managers	76	60
Professionals, technical workers, and artists	66	51

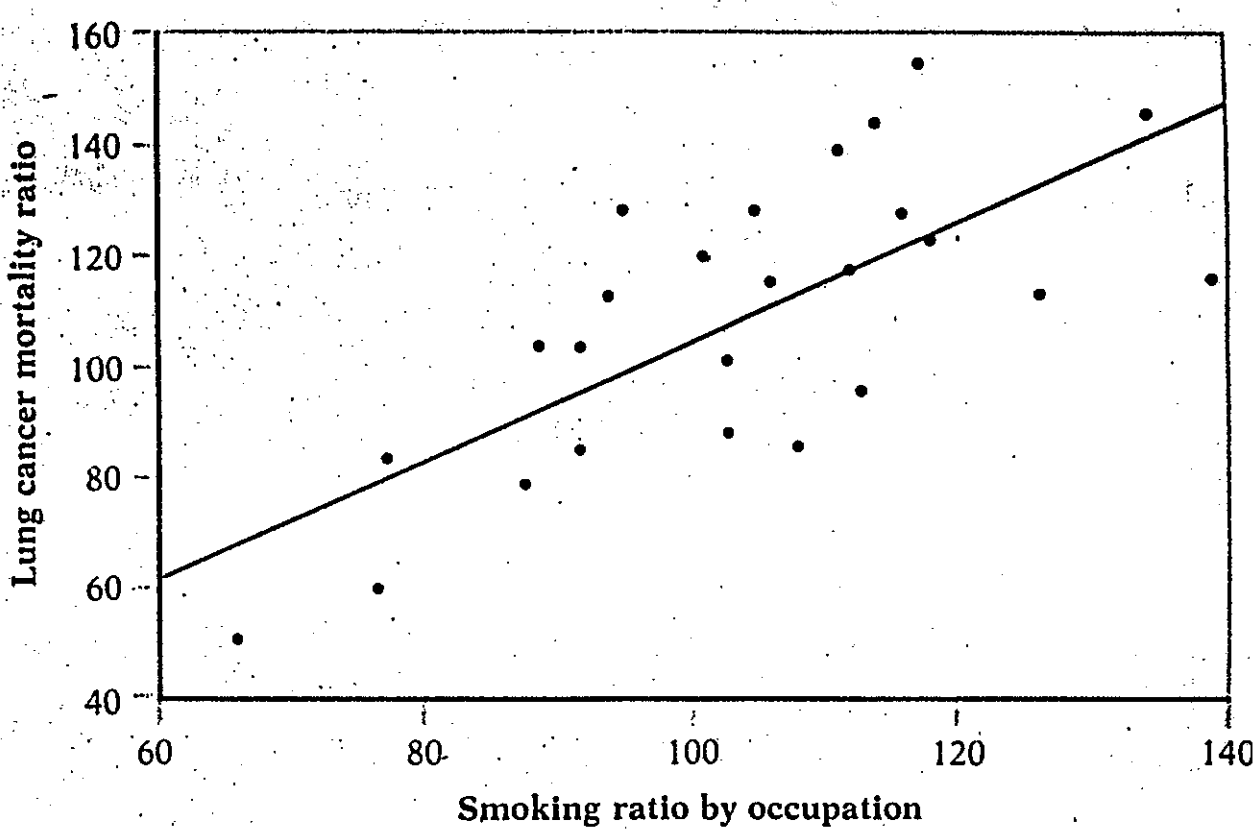


Figure 3.34 The regression of a measure of deaths from lung cancer on a measure of smoking for British occupational groups. See Example 3.21.

Francis Galton (1822 - 1911)

Galton had earned a Cambridge mathematics degree and completed two years of medical school when his father died, leaving him a substantial inheritance. Free to travel, he became an explorer of some note, but when *Origin of the Species* was published in 1859, his interests began to shift from geography to statistics and anthropology (Charles Darwin was his cousin). It was Galton's work on fingerprints that made possible their use in human identification. He was knighted in 1909.

Sir Francis Galton developed the statistical concept of correlation and regression. On the day before his fifth birthday, he wrote to his sister:

My dear Adele,

I am four years old and I can read any English book. I can say all the Latin substantives and adjectives and active verbs besides 52 lines of Latin poetry. I can cast up any sum in addition and can multiply by 2 through 10. I can also say the pence table. I read French a little and know the clock.

February 15, 1827

Galton went on to write over 200 papers and 15 books.

Regression Toward Mediocrity in Heredity Stature,
Journal of the Anthropological Institute, 15 (1885), pages 246-263.

TABLE 10.3.1 NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES (ALL FEMALE HEIGHTS HAVE BEEN MULTIPLIED BY 1.08)

Height of the Mid-Parents (inches)	Heights of the Adult Children														Total Number of:		Medians or Values of *M*
	Below	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	Above	Adult Children	Mid-Parents	
Above 72.5	—	—	—	—	—	—	—	—	—	—	—	1	3	—	4	5	
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4	19	6	72.2
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2	43	11	69.9
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3	68	22	69.5
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68.9
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—	219	49	68.2
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—	211	33	67.6
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—	78	20	67.2
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—	66	12	66.7
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—	23	5	65.8
Below 64.5	1	—	2	4	1	2	2	1	1	—	—	—	—	—	14	1	
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0					

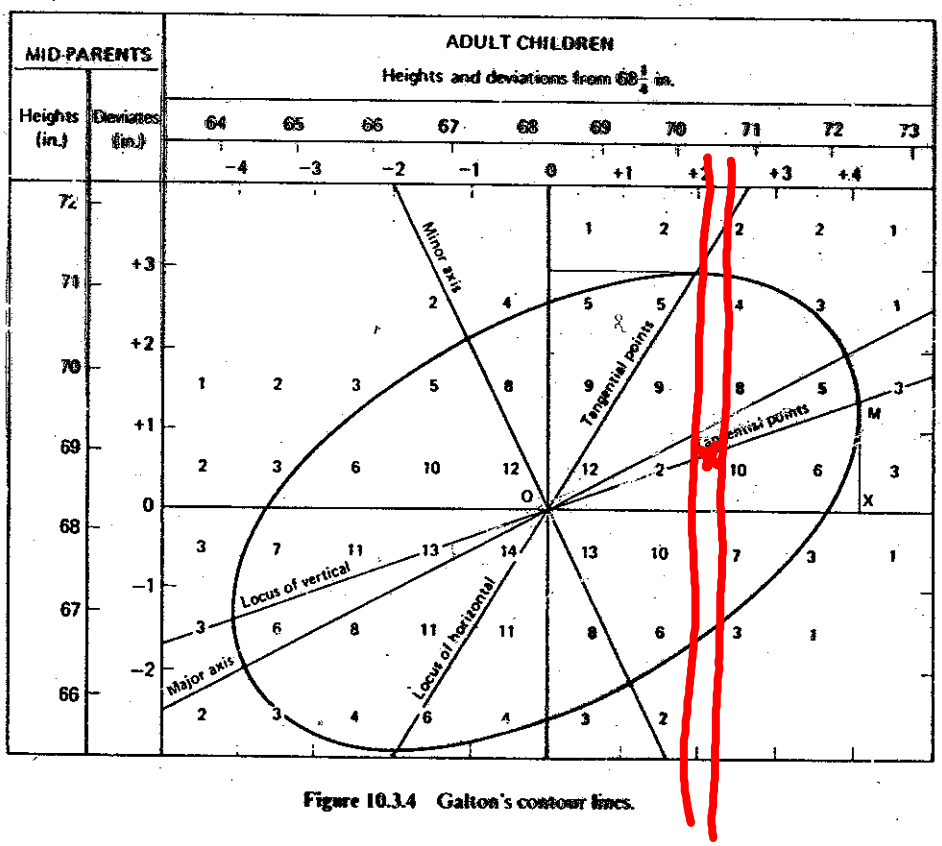


Figure 10.3.4 Galton's contour lines.

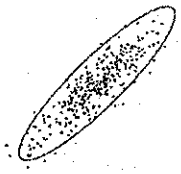
Galton (51) gave the following recollection of the birth of property 4:

> At length, one morning, while waiting at a roadside station near Ramsgate for a train, and poring over the diagram in my notebook, it struck me that the lines of equal frequency ran in concentric ellipses. The cases were too few for my certainty, but my eye, being accustomed to such things, satisfied me that I was approaching the solution. More careful drawings strongly corroborated the first impression.

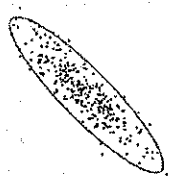
[1] Of this discovery, Pearson said, "That Galton should have evolved all this from his observations is to my mind one of the most noteworthy scientific discoveries arising from pure analysis of observations" (122).

Association

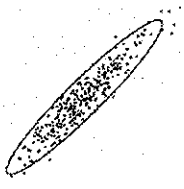
Positive association



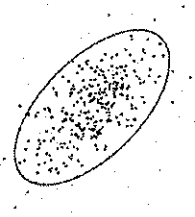
Negative association



Strong association



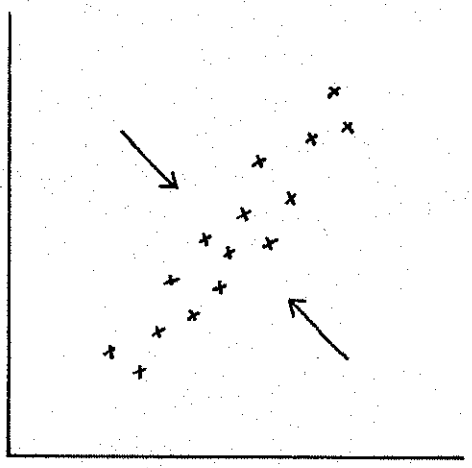
Weak association



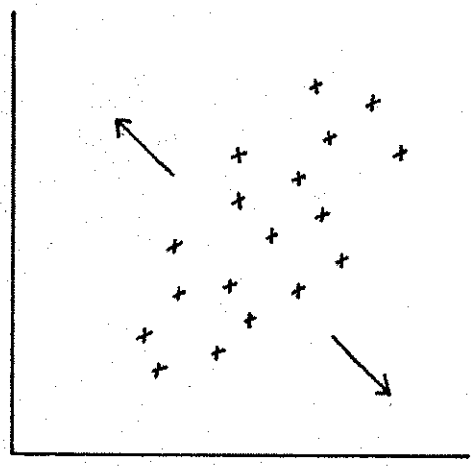
Strong: knowing one helps a lot in predicting the other
Weak: knowing one doesn't help much in predicting the other

Figure 5. Summarizing a scatter diagram. The correlation coefficient measures clustering around a line.

(a) Correlation near 1 means tight clustering.

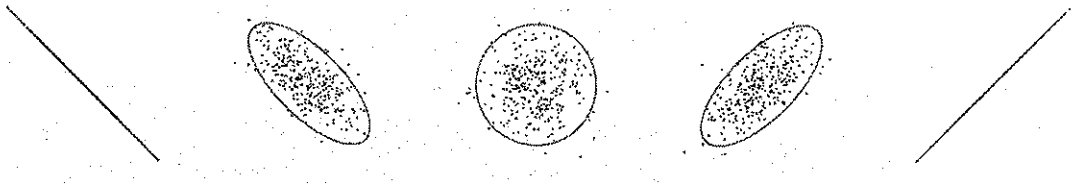


(b) Correlation near 0 means loose clustering.



The Correlation Coefficient

The correlation coefficient, r, measures the strength of association.



r = -1 **r = -.7** **r = 0** **r = .7** **r = 1**

- **r is always between -1 and 1.**
- **r is -1 if the points are on a line with a negative slope.**
- **r is 1 if the points are on a line with a positive slope.**

Goal: Analyze the association or correlation between two variables
Why? Prediction, Understanding, Control

The correlation coefficient is a measure of linear association between two variables.

Computing the correlation coefficient:

1. Convert the x - values to standard units.

2. Convert the y - values to standard units.

3. Multiply each x - value (in standard units) by each y - value (in standard units) .

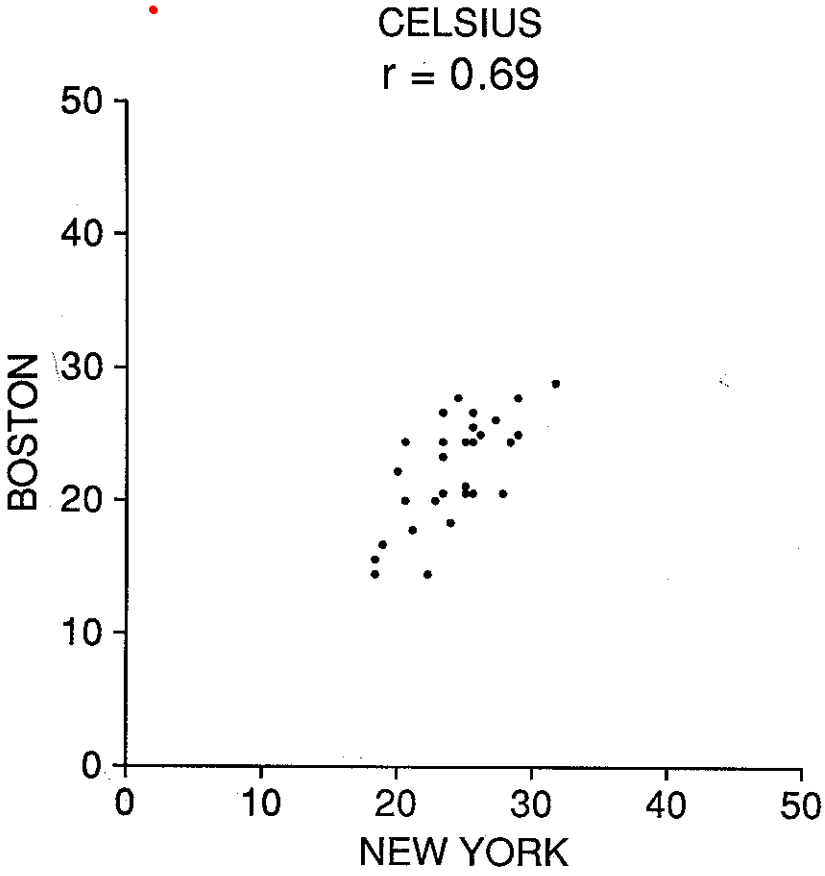
4. The average of these products is equal to r.

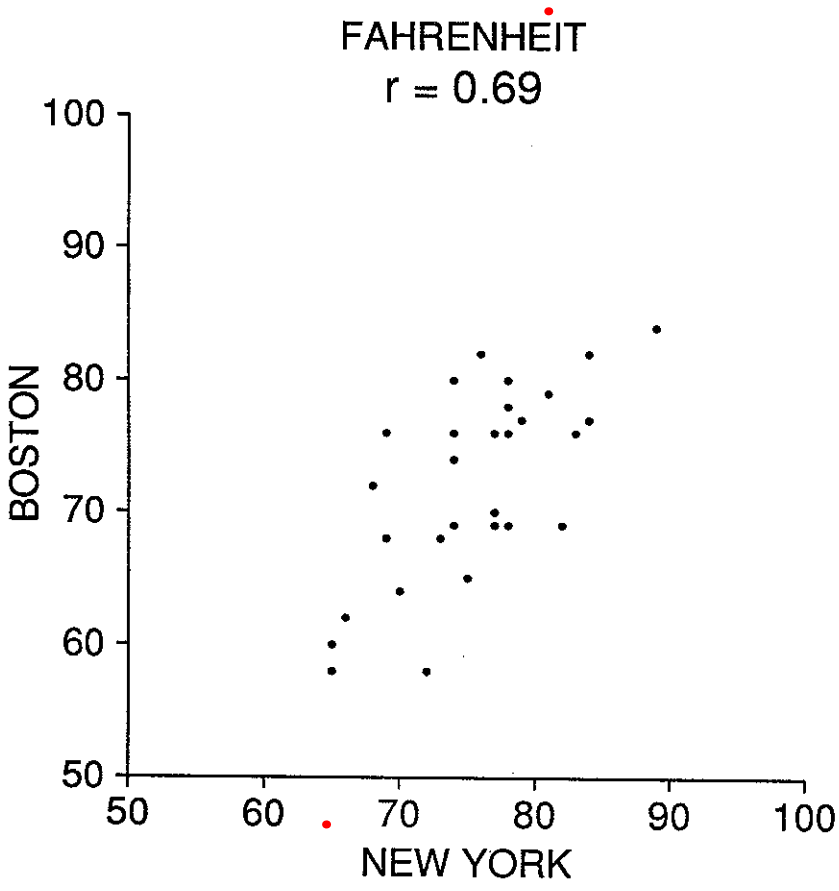
$$r = \text{average of the products } \left(\frac{x-AV_x}{SD_x}\right) \cdot \left(\frac{y-AV_y}{SD_y}\right)$$

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

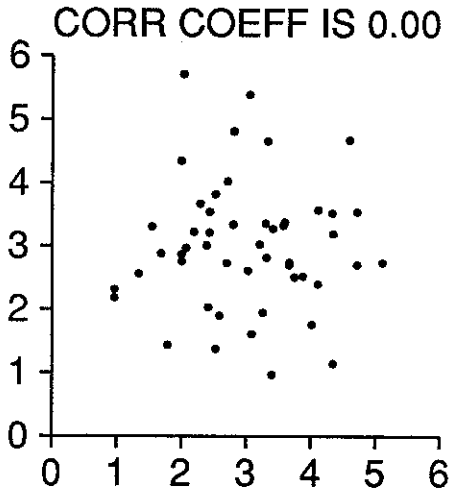
TEMPERATURES IN BOSTON AND NEW YORK

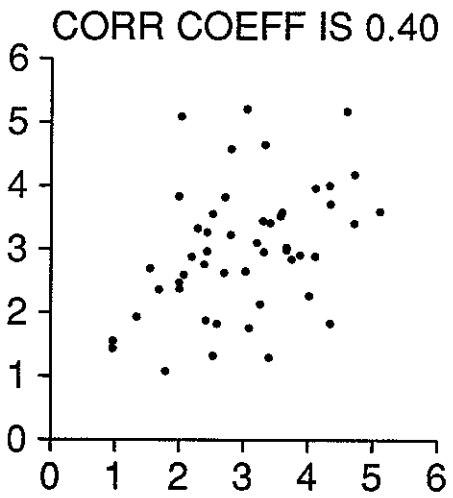
June Daily Maximum

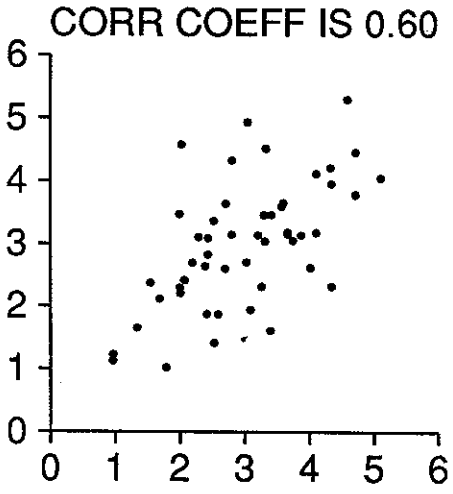


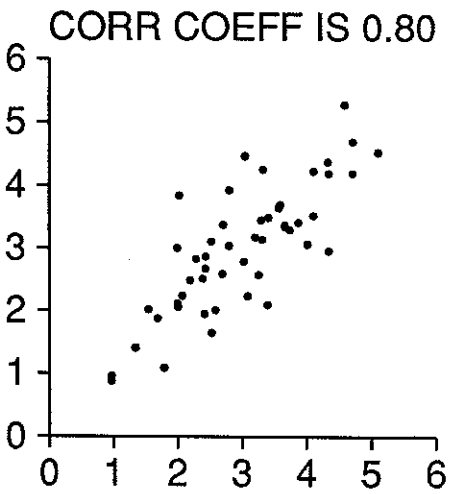


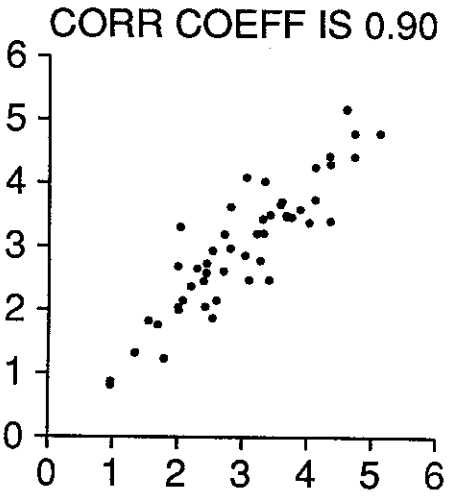
SIX POSITIVE CORRELATIONS

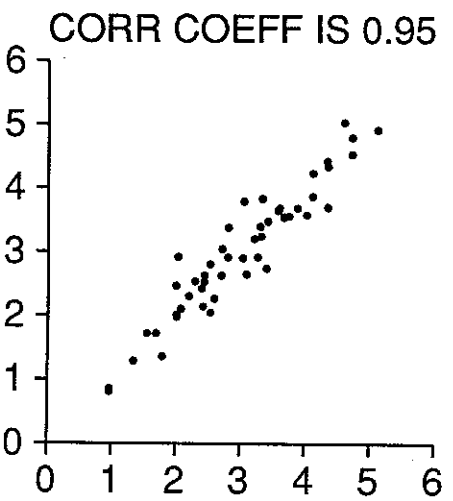




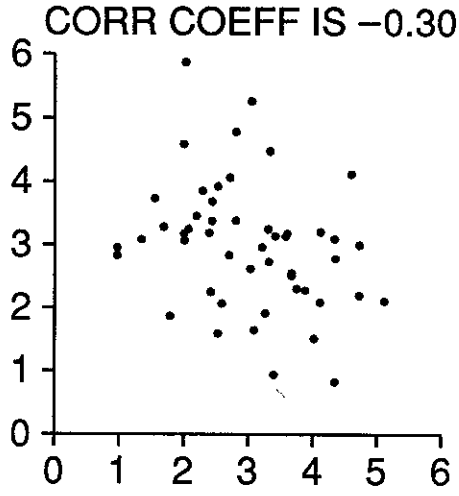


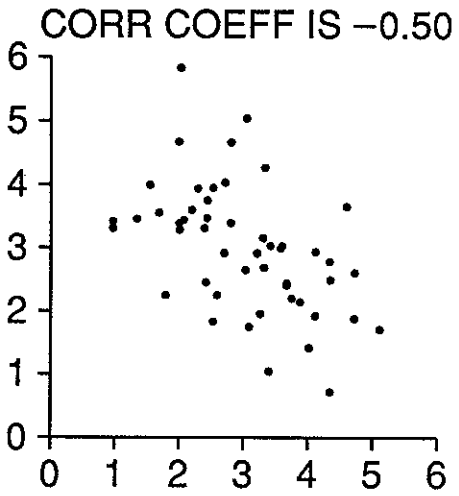


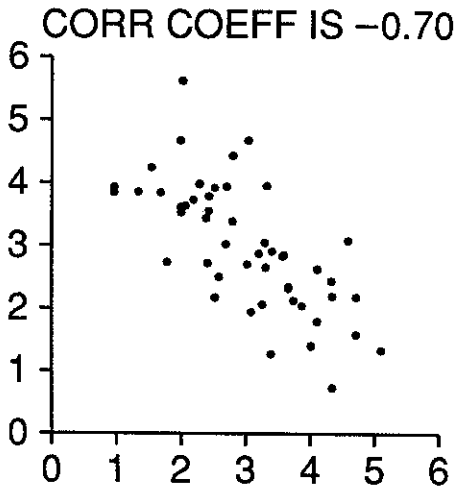


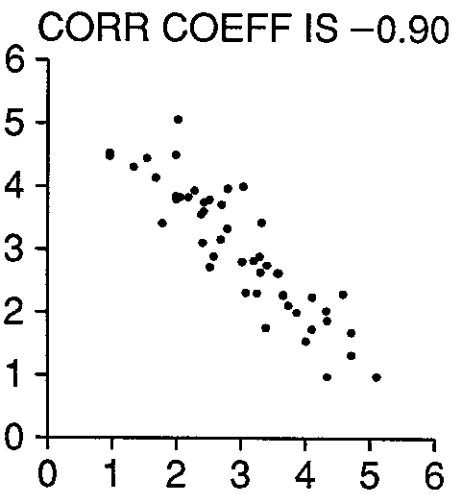


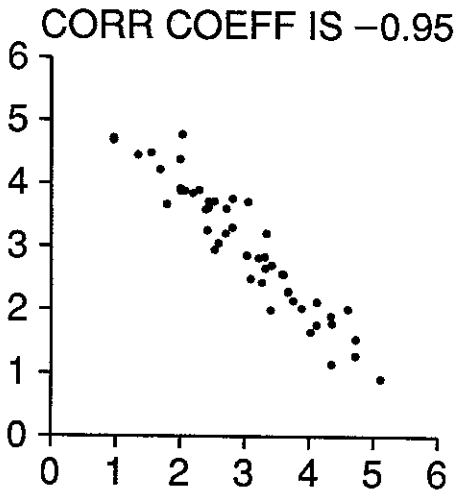
SIX NEGATIVE CORRELATIONS

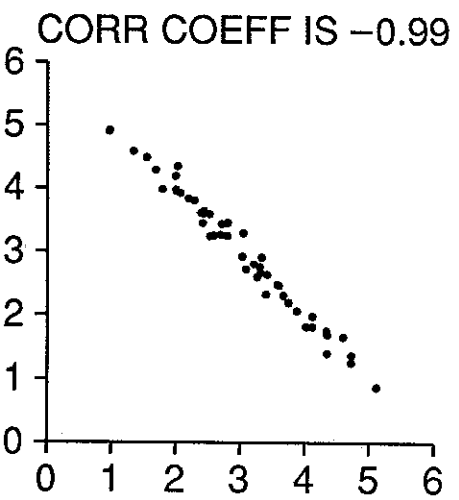












Calculating the correlation coefficient, r

For x, average = 4, SD = 2

For y, average = 7, SD = 4

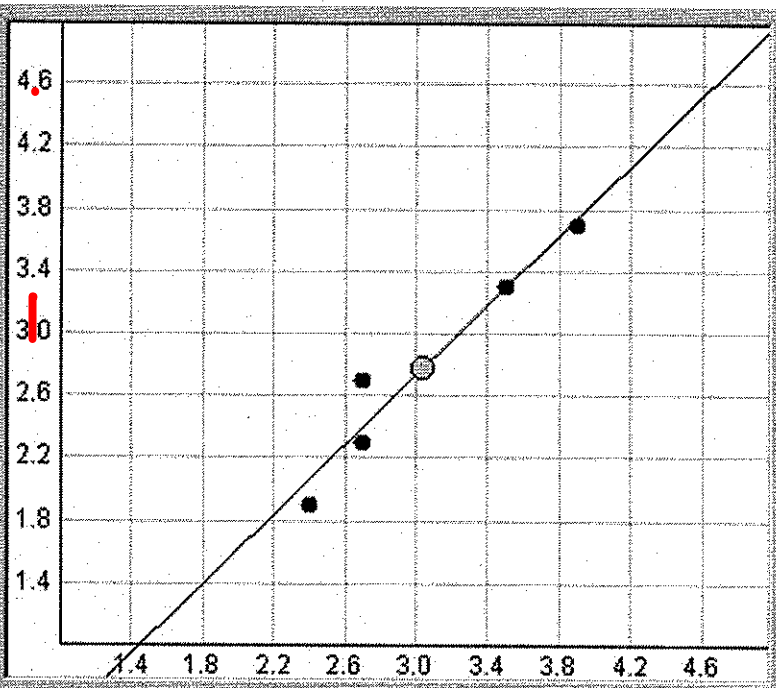
x	y	*x in standard units*	*y in standard units*	*Product*
1	5	−1.5	−0.5	0.75
3	9	−0.5	0.5	−0.25
4	7	0.0	0.0	0.00
5	1	0.5	−1.5	−0.75
7	13	1.5	1.5	2.25

$$r = \frac{0.75 - 0.25 + 0.0 - 0.75 + 2.25}{5} = 0.4$$

r = 0.4

		X	Y	H.S. GPA	USU GPA	X-Y
		H. S. GPA	USU GPA	Standard Units	Standard Units	Products
		2.7	2.3	-0.603	-0.736	0.44
		3.5	3.3	0.816	0.798	0.65
		2.4	1.9	-1.135	-1.349	1.53
		3.9	3.7	1.53	1.411	2.16
		2.7	2.7	-0.603	-0.122	0.07
	AV	3.04	2.78		r =	0.97
	SD	0.564	0.652			

	x	y
1	2.7	2.3
2	3.5	3.3
3	2.4	1.9
4	3.9	3.7
5	2.7	2.7
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		



Scale

x min 1 x max 5

y min 1 y max 5

Apply Clear

n = 5

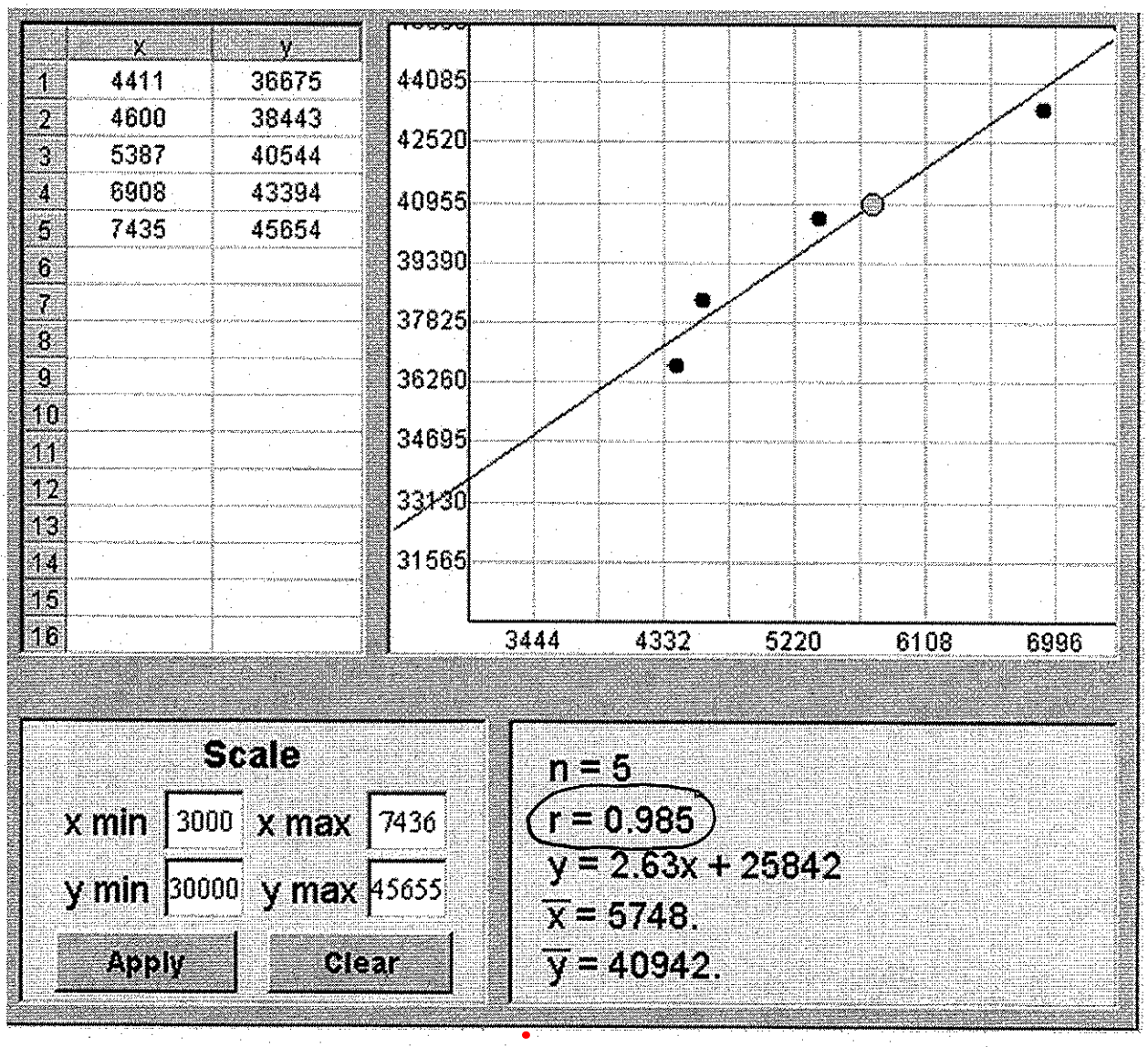
r = 0.969

y = 1.12x - 0.63

$\overline{x}$ = 3.0

$\overline{y}$ = 2.8

Year	AV Teacher Salary	Beer Consumption $
1994	$36,675	$4411
1996	$38,443	$4600
1998	$40,544	$5387
2000	$43,394	$6908
2002	$45,654	$7435



Examples of Correlation

1. Would the correlation between the age of a used car and its price be positive or negative?

2. Why are children shoe sizes positively correlated with reading ability?

3. Why is daily chocolate consumption positively correlated with daily auto accidents?

4. For a representative sample of cars, would the correlation between the age of the car and its gasoline economy (miles per gallon) be positive or negative?

5. The correlation between gasoline economy and income of owner turns out to be positive. How do you account for this positive association?

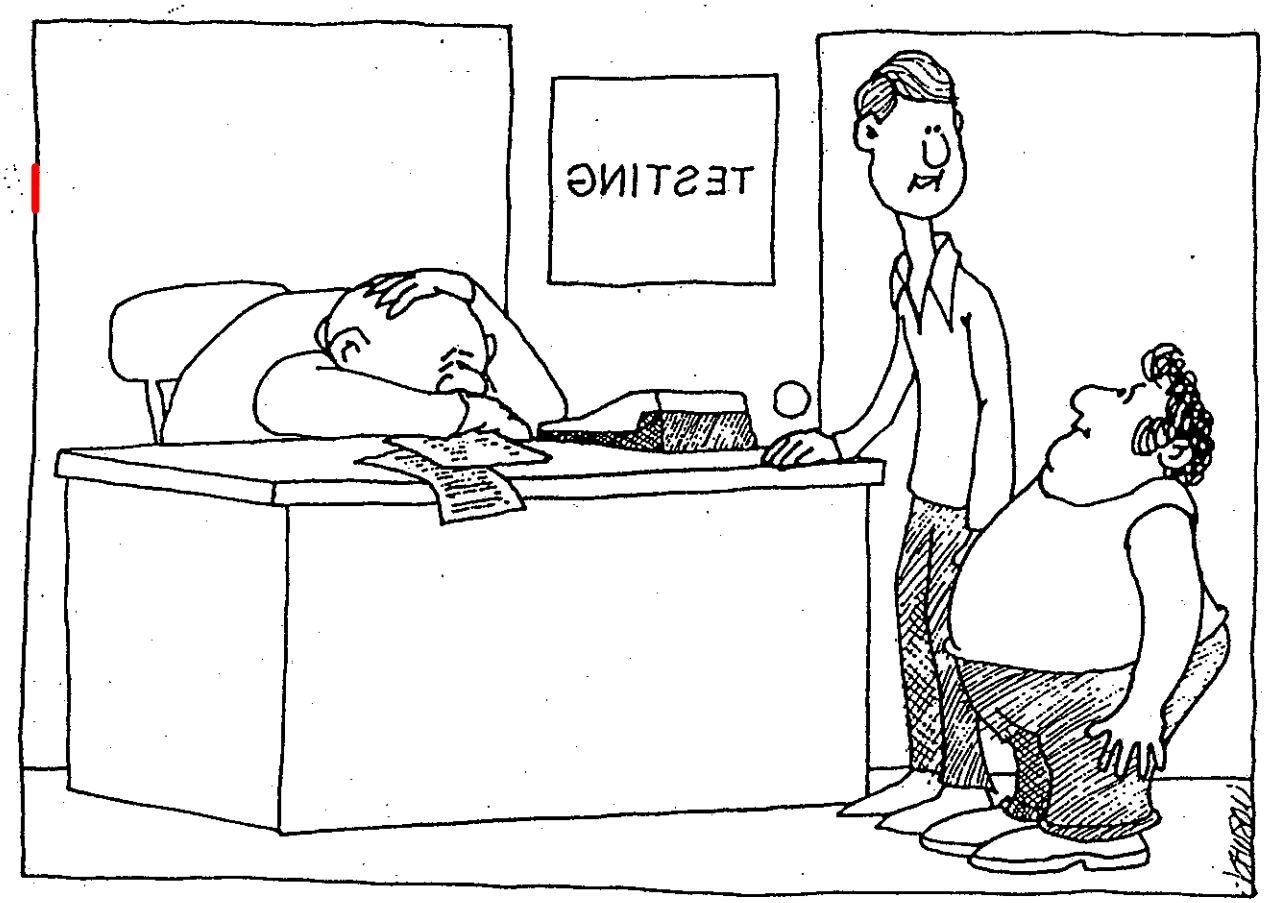
6. Is the correlation between the heights of husbands and wives in the U.S. around -0.9 , -0.3 , 0.3 , or 0.9 ?

7. If women always married men who were exactly 4 inches taller, what would the correlation between their heights be?

8. Three data sets are collected, and the correlation coefficient is computed in each case. The variables are

   (i) grade point average in freshman year and in sophomore year
   (ii) grade point average in freshman year and in senior year
   (iii) length and weight of two-by-four boards

   Possible values for the correlation coefficients are -.50 , 0.0 , 0.30, 0.60, and 0.95. Match the correlations with the data sets.



"He says we've ruined his positive association between height and weight."