$y = mx + b$

$y =$ y-coordinate - dependent variabl

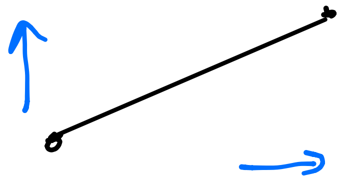
$m =$ slope $= \dfrac{y_2 - y_1}{x_2 - x_1}$

Given: (x_1, y_1) and (x_2, y_2)

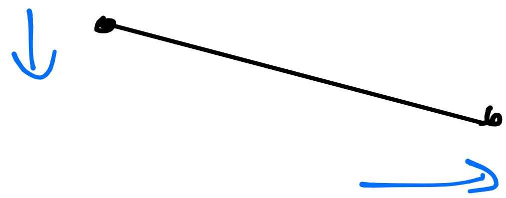
$x =$ x-coordinate - independent variable

$b \Rightarrow (0, b)$ is y-intercept

if m is positive



if m is negative



Example 12.4

Svetlana tutors to make extra money for college. For each tutoring session, she charges a one-time fee of $25 plus $15 per hour of tutoring. A linear equation that expresses the total amount of money Svetlana earns for each session she tutors is $y = 25 + 15x$.

* $25 starting price, when $x = 0$

* $15 per hour charge

$m = \dfrac{\$15}{1 \text{ hr}}$ slope

Once we've found comparative data, we can then plot a graph designating one set of data as the x variable or **independent variable,** and the other as the y variable or **dependent variable.** This graph is called a *scatter plot*.

A **scatter plot** is a graph of the ordered pairs (x, y) consisting of data from two data sets.

Example 1 Drawing and Analyzing a Scatter Plot

A medical researcher selects a sample of small hospitals in his state and hopes to discover if there is a relationship between the number of beds and the number of personnel employed by the hospital.

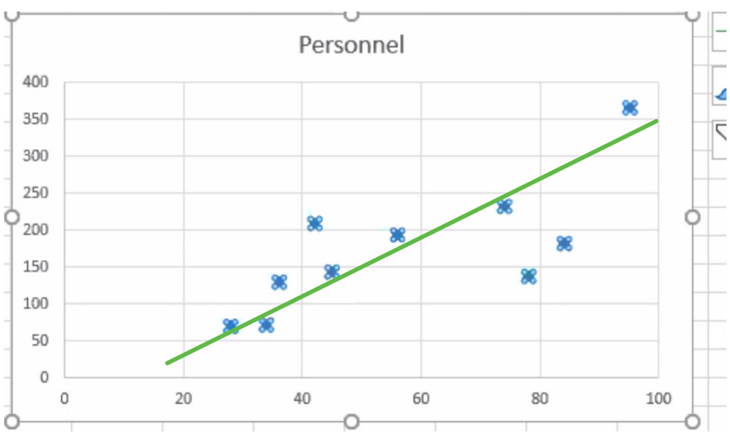
(a) Do you think there should be a relationship between these two data sets? Describe any relationship you'd expect.

(b) Draw a scatter plot for the data shown. Does it look like there's a relationship between the data sets?

No. of beds (x)	28	56	34	42	45	78	84	36	74	95
Personnel (y)	72	195	74	211	145	139	184	131	233	366

a.) Expect more beds → more personnel

b.)



Looks like it's growing, but not "perfect relationship".

Analyzing a Scatter Plot

There are several types of relationships that can exist between the *x* values and the *y* values in a scatter plot. These relationships can be identified by looking at the pattern of the points on the graphs. The types of patterns and corresponding relationships are:

1. *A positive linear relationship* exists when the points fall approximately in an ascending straight line from left to right, where the *x* and *y* values increase at the same time. See Figure 11-26a.

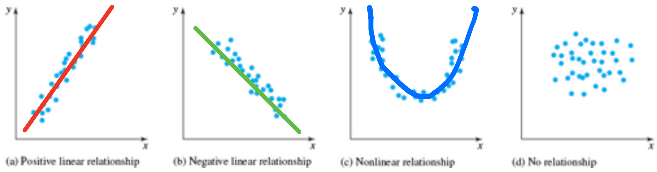


Figure 11-26

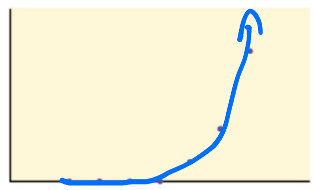
2. *A negative linear relationship* exists when the points fall approximately in a descending straight line from left to right. See Figure 11-26b. In this case as the *x* values are increasing, the *y* values are decreasing.
3. *A nonlinear relationship* exists when the points fall in a curved line. See Figure 11-26c. The relationship is described by the nature of the curve.
4. *No relationship* exists when there is no discernible pattern to the points. See Figure 11-26d.

(a) straight line, w/positive slope

(b.) straight line, w/negative slope

(c) nonlinear relationship "polynomial?"

(d) random looking



(a) Exponential growth pattern

Correlation Coefficient (r) – the number that describes how close to a linear relationship between two data sets

→ if r is positive → positive linear

→ if r is negative → negative linear

$-1 \le r \le 1$

Figure 11-27

Example 2 **Estimating Correlation Coefficients**

Based on our description of the correlation coefficient for two data sets, make an estimate of the correlation coefficient for each scatter plot in Figure 11-26.

a.) positive linear, closer to +1

b.) negative linear, closer to −1

c/d) no linear relationship.

→ if r = +1 perfect positive linear relationship

r = −1 perfect negative linear relationship

Calculating the Value of the Correlation Coefficient

Lecture: Signicance Levels and Regression

In order to find the value of the correlation coefficient, we will use the following formula:

Formula for Finding the Value of r:

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

Pearson
product
Coefficient

where

n = the number of data pairs

$\sum x$ = the sum of the x values

$\sum y$ = the sum of the y values

$\sum xy$ = the sum of the products of the x and y values for each pair

$\sum x^2$ = the sum of the squares of the x values

$\sum y^2$ = the sum of the squares of the y values

Example 3 **Calculating a Correlation Coefficient**

Find the correlation coefficient for the data in Example 1, and discuss what you think indicates.

I13 =(A11*(D12)-B12*C12)/SQRT((A11*E12-B12^2)*(A11*F12-C12^2))

	A	B	C	D	E	F	G	H	I	J	K	L
1		Number of beds (x)	Personnel (y)	xy	x^2	y^2						
2		28	72	2016	784	5184						sumx
3		56	195	10920	3136	38025						sum y
4		34	74	2516	1156	5476						sum xy
5		42	211	8862	1764	44521						sum x^2
6		45	145	6525	2025	21025						sum y^2
7		78	139	10842	6084	19321						
8		84	184	15456	7056	33856						
9		36	131	4716	1296	17161						
10	n=10	74	233	17242	5476	54289						
11	10	95	366	34770	9025	133956						
12	**sum**	**572**	**1750**	**113865**	**37802**	**372814**						
13								R=	0.748292	closer to 1	pretty good relationship	

$r = .748$

$n = 10$

Table 11-4 Significant Values for the Correlation Coefficient

Sample Size	5%	1%
4	.950	.990
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
21	.433	.549
22	.423	.537
23	.412	.526
24	.403	.515
25	.396	.505
30	.361	.463
40	.312	.402
60	.254	.330
120	.179	.234

confidence 95% 99%

significance 5% 1%

chances we're incorrect.

the minimum r value to have 1% chance of being wrong for sample size of 15.

Using Significant Values for the Correlation Coefficient

If $|r|$ is greater than or equal to the value given in Table 11-4 for either the 5% or 1% significance level, then we can reasonably conclude that the two data sets are linearly related.

5% value when $n=10$

$|r| = .748 > .632$?

Yes,

→ # beds and personnel is linearly related under 5% significance

Example 4 Deciding if a Correlation Coefficient Is Significant

Determine if the correlation coefficient $r = 0.748$ found in Example 3 is significant at the 5% level.

at 1%? $|r| = .748 > .765$?

1% value $n = 10$

No

→ # beds and personnel is not linearly related @ 1% significance

Formulas for Finding the Values of *a* and *b* for the Equation of the Regression Line

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \text{ slope}$$

$$a = \frac{\sum y - b(\sum x)}{n} \; y \text{ intercept}$$

Example 5 Finding a Regression Line

Find the equation of the regression line for the data in Example 1.

$y = bx + a$

Least Squares Regression

$y = 2.7077x + 20.118$

see Excel Sheet

The Relationship between *r* and the Regression Line

Two things should be noted about the relationship between the value of *r* and the regression line. First, the value of *r* and the value of the slope, *b*, always have the same sign. Second, the closer the value of *r* is to +1 or −1, the better the points will fit the line. In other words, the stronger the relationship, the better the fit. Figure 11-29 shows the relationship between the correlation coefficient and the regression line.

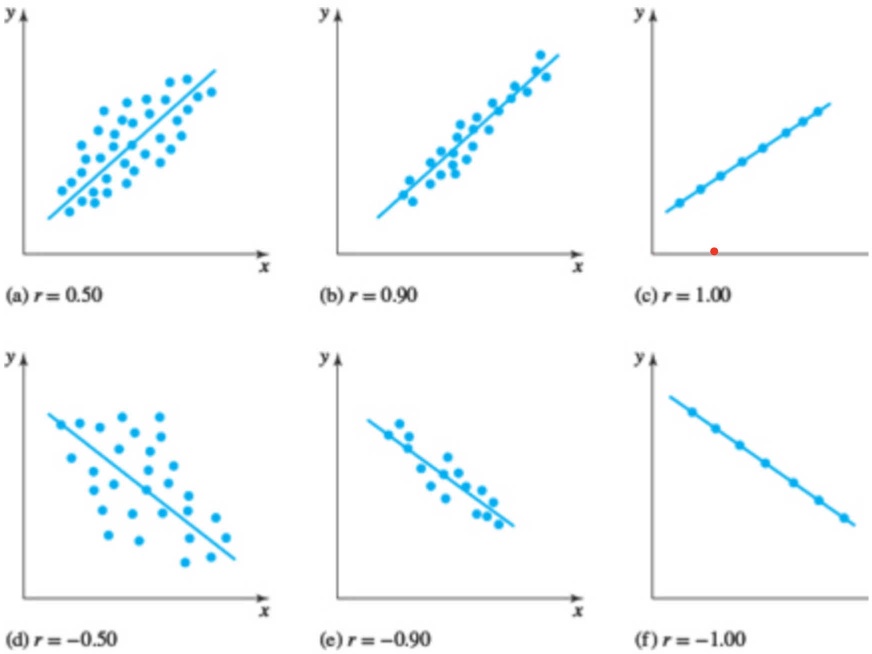


Figure 11-29

One of the most useful aspects of finding a regression line is that it can be used to make predictions for one of the variables given a value for the other. This is illustrated in Example 5.

Example 6 **Using a Regression Line to Make a Prediction**

Use the equation of the regression line found in Example 5 to predict the approximate number of personnel for a hospital with 65 beds.

personnel   beds

$$y = 2.7077x + 20.118$$

$$y = 2.7077(65) + 20.118$$

$y \approx 196.12$ personnel

$y \approx 196$ personnel.