

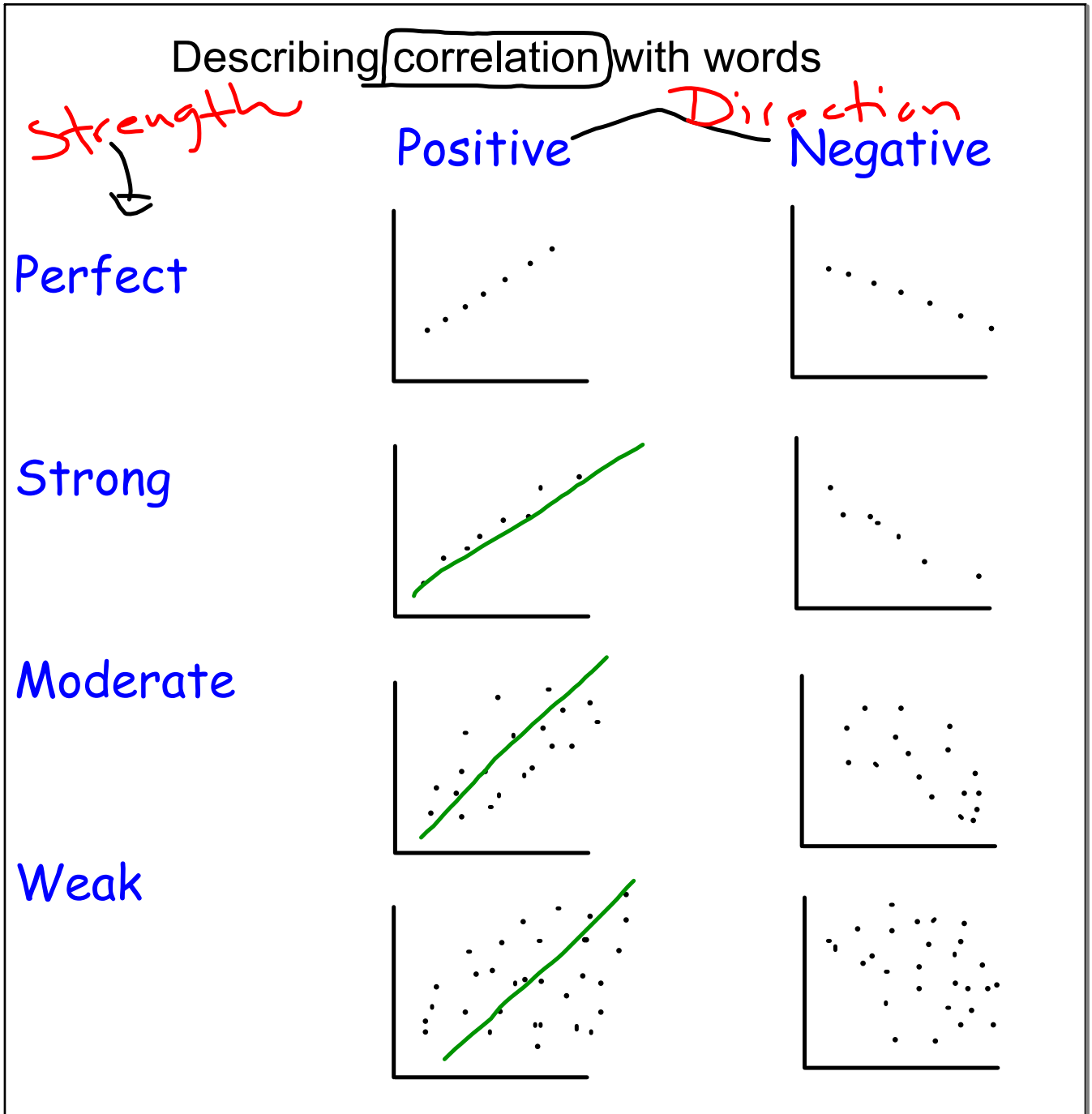
Correlation:

Describes the strength and direction of a linear relationship between two quantitative variables.

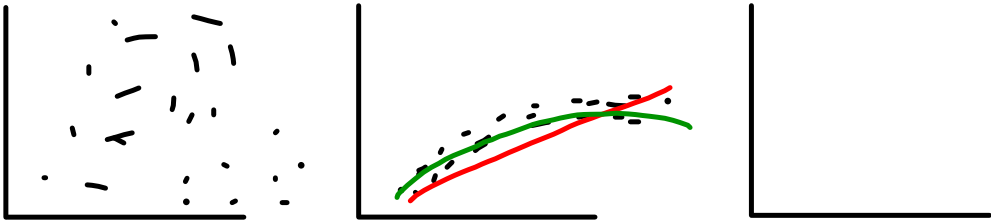
"strength" can be *perfect, strong, moderate, or weak*

"direction" can be *positive or negative*

NOTE: It is possible to have no correlation.

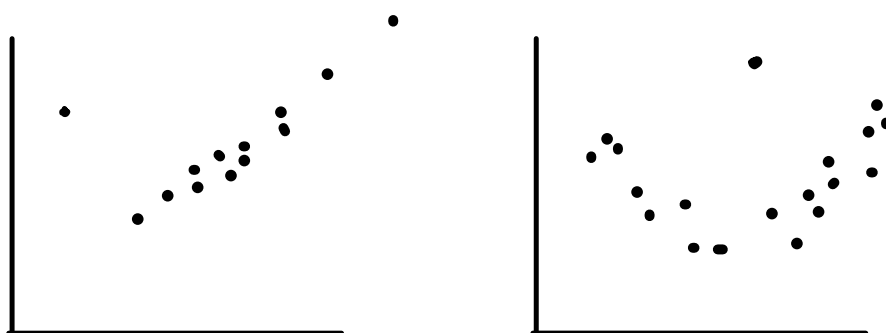


No correlation



Outliers in a Scatter Plot

Outliers fall outside of the overall pattern of points.



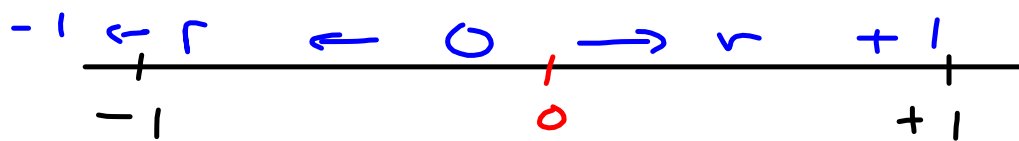
Correlation Formula

$$r = \frac{1}{n-1} \sum \left(\frac{x-\bar{x}}{s_x} \right) \left(\frac{y-\bar{y}}{s_y} \right)$$

Use calculator !!

Correlation

$$r = \pm 1$$



negative

down

Positive

positive

$-r$

0 to -1

↑

weak

↑
perfect

$+r$

0 to $+1$

↑

weak

↑
perfect

* Correlation of zero

* still can be a correlation just not linear.

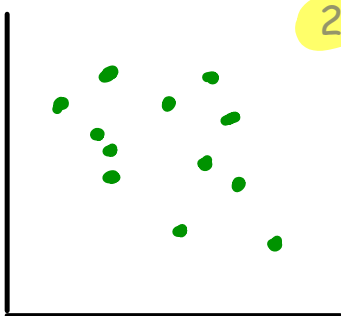
Correlation from -1 to 1

<http://www.duxbury.com/authors/mcclellandg/tiein/johnson/correlation.htm>



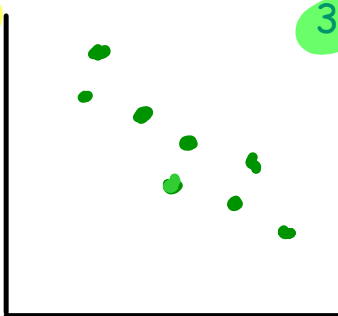
Matching Correlations

1.



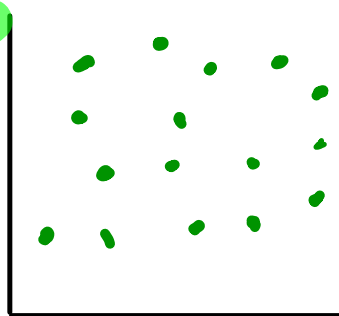
$r = -.56$

2.



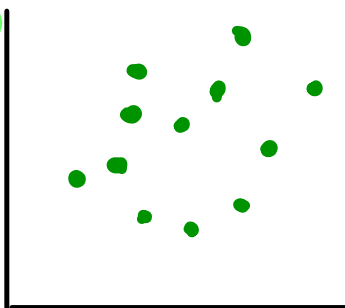
$r = -.87$

3.



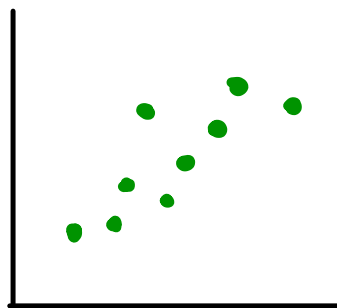
$r = 0.08$

4.



$r = .25$

5.



$r = .85$

A. $r = 0.25$

B. $r = -.87$

C. $r = -.56$

D. $r = .85$

E. $r = 0.08$

"r" = correlation coefficients

Guessing Correlations Applet

<http://istics.net/stat/Correlations/>


Facts About Correlation

1. The variable that represents correlation is r (the correlation coefficient).
2. $_____ \leq r \leq _____$
3. When $r = \pm 1$, there is $_____$ correlation.
4. When $r = 0$, there is $_____$ correlation.
5. r has no unit.
6. r is $_____$ to outliers
7. The value of r is not changed if you...
 - ...switch x & y ,
 - ...add/subtract a number to/from all of the x or y values, or
 - ...multiply/divide each x or y value by a positive number.

L_1 X	L_2 Y
5	9
3	8
1	6
4	5
6	7
8	2
2	7
7	1

$$\# 4 \quad y = ax + b$$

$$x: L_1$$

$$y: L_2$$

$$\# 8 \quad y = a + bx$$

$$\frac{d}{dx} ? \quad cn$$

$$r^2 =$$

$$r =$$

$$y = a + bx$$

$$a = 8.7$$

$$b = -.68$$

$$r^2 =$$

$$r = -.588$$

$$y = 8.7 - .68x$$

$$r = -.59$$

Finding correlation in the calculator

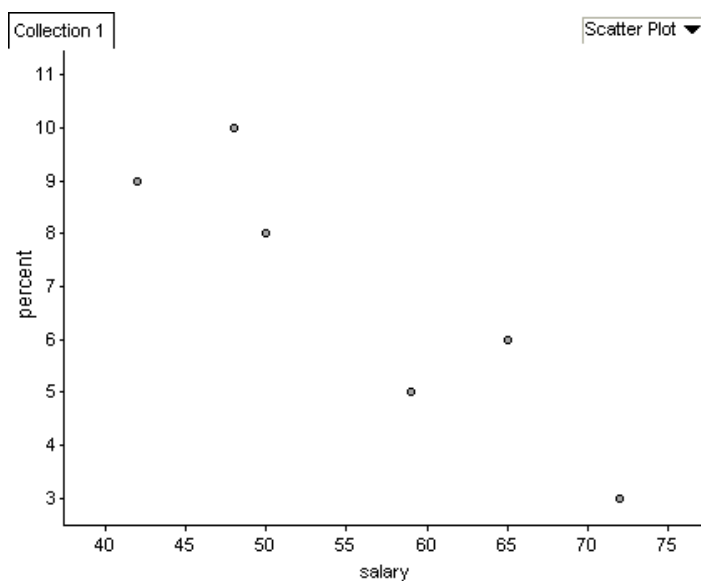
Ex:

Income Level (in 1000s)	Donating Percent
42	9
48	10
50	8
59	5
65	6
72	3

Make a scatter plot, estimate the correlation, and then find the exact correlation in the calculator.

Estimating Correlation: The Ellipse Method

$$r \approx \pm \left(1 - \frac{\text{minor}}{\text{major}} \right)$$



Homework:

Read p. 156 "Correlation Properties"
& p. 160 "What Can Go Wrong?"

Do Ch. 7 # 19-23, 28, 35

Barron's ... Page 123

Summary

- A scatterplot gives an immediate indication of the shape (linear or not), strength, and direction (positive or negative) of a possible relationship between two variables.
- If the relationship appears roughly linear, then the correlation coefficient, r , is a useful measurement.
- The value of r is always between -1 and $+1$, with positive values indicating positive association and negative values indicating negative association; and values close to -1 or $+1$ indicating a stronger linear association than values close to 0 , which indicate a weaker linear association.
- Evidence of an association is not evidence of a cause-and-effect relationship!
- Correlation is not affected by which variable is called x and which y or by changing units.
- Correlation can be strongly affected by extreme values.
- The differences between the observed and predicted values are called residuals.
- The best-fitting straight line, called the regression line, minimizes the sum of the squares of the residuals.
- For the linear regression model, the mean of the residuals is always 0 .
- A definite pattern in the residual plot indicates that a nonlinear model may fit the data better than the straight regression line.
- The coefficient of determination, r^2 , gives the percentage of variation in y that is accounted for by the variation in x .
- Influential scores are scores whose removal would sharply change the regression line.
- Nonlinear models can sometimes be studied by transforming one or both variables and then noting a linear relationship.
- It is very important to be able to interpret generic computer output.

Generated by CamScanner

