

3.2 Correlation

When a scatterplot takes a linear form, we use a numerical value called correlation (r) to discuss the direction and strength of the relationship between the two quantitative variables.

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

Not as scary as it looks - each value in parenthesis is the z-score for the value in the data. You calculate a z-score for an x , a z-score for the corresponding y , multiply the pair, do that for each pair and then add them up and divide by $n-1$. So the correlation is the average of the products of the standardized values.

We will use our calculator to calculate correlation (r).

Interpreting Correlation (r):

$r > 0$ means positive linear association

$r < 0$ means negative linear association

r close to 0 means very weak linear association

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

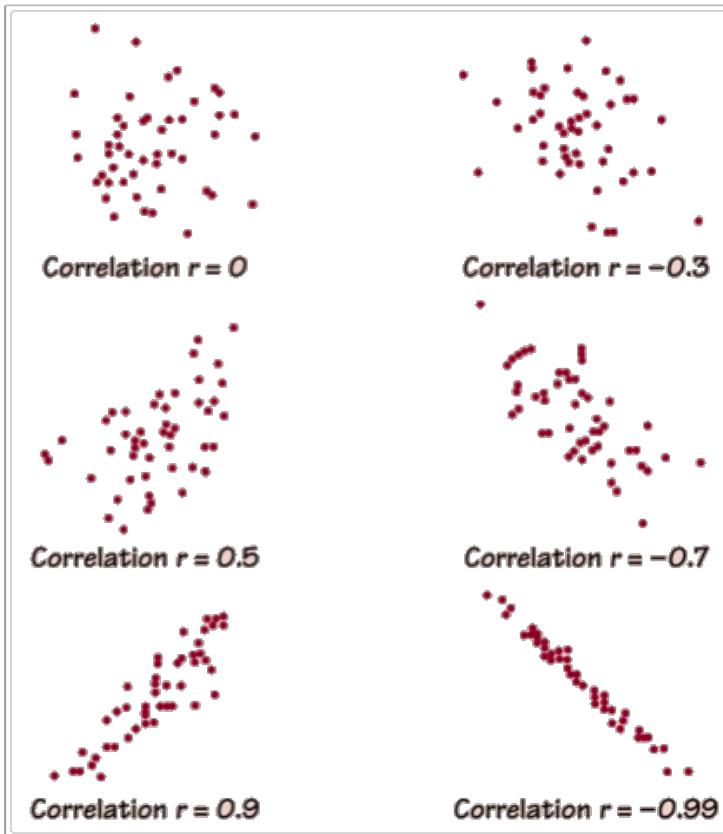
perfect correlation is when $r = 1$ or $r = -1$

*all points lie exactly on a straight line.

Good to know:

- r has no units (uses standardized values, doesn't change with conversions)
- r doesn't care which variable is x and which is y
- both variables must be quantitative to calculate r
- r is only used for measuring strength of linear relationships
- r is not resistant (if we take out an outlier, it strengthens the r score)
- r should be given with mean and std dev of x and y

Correlation Applet



Using the Calculator:

From Ex. 3.19 on P138 Archaeopteryx femur & humerus lengths

Enter into L_1 : 38 56 59 64 74

Enter into L_2 : 41 63 70 72 84

Calculate 2-var stats:

STAT | CALC | 2 | L_1, L_2 | Enter

Now go to your list editor and in the header of L_3 type:

$$\left((L_1 - \bar{x}) / s_x \right) (L_2 - \bar{y}) / s_y$$

VAR S | 5 for mean & std dev

Now from your main screen, type:

$$(1/(n-1) * \text{sum}(L_3))$$

sum is in the **2nd | STAT | MATH** menu