secular trend in height: ± about 1 in taller than ± (nutrition)

Case Study 17 (genetics): Human Inheritance of Height

How is your height related to that of your parents? Two eminent British scientists, Francis Galton and Karl Pearson, investigated this question in the late 1800s. They gathered data on the adult heights of 1,078 fathers and sons (one son per family). The picture below graphically summarizes this dataset; here are some numerical summaries:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father's height</td>
<td>68 inches</td>
<td>2.7 inches</td>
<td>1078</td>
</tr>
<tr>
<td>Son's height</td>
<td>69 inches</td>
<td>2.7 inches</td>
<td></td>
</tr>
</tbody>
</table>

If you were going to draw a smooth curve through the data that captured the basic trend of how sons' heights depend on the heights of their fathers, a straight line would probably do pretty well. The plot below gives two natural lines you might try to fit to the data, one called the SD line, the other the regression line. We will talk next time about which is which, and which one is better. The question for today is: How strong is the linear relationship between these two variables?
1. How strongly related (is a linear way) $\rho$.
2. $\rho$ = slope / positively associated / negatively associated
   - Correlation coefficients
   - No linear association
   - Non-linear association
   - Slope is zero
   - $Y = 0$
   - $x = 0$
   - $Y = x$
4 quadrants

\( r = \text{average of products of } x \& y \text{ variables in } Su \)

The official formula:

\[ r = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s_x^*} \right) \left( \frac{y_i - \bar{y}}{s_y^*} \right) \]

\( s_x^* = \sigma_x \) but dividing by \( n \) not \( (n-1) \)

\[ s_x^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \]

What's the corr. between \( F \) of \( F \& S \)?

\( r = +.5 \) (moderate)

Facts about \( r \):

- 4 quadrants (roughly equal)
- \( r > 0 \) (correlation)
- \( r = 0 \) (no correlation)
- \( r < 0 \) (inverse correlation)

(mentally convert \( x \& y \) to \( Su \))
1. \( r \) always converges between \([-1, 1]\).

- **Perfect negative linear association**
- **No linear association**
- **Perfect positive linear association**

2. \( r \) can be fooled by 2 things:
   - (a) nonlinearity
   - (b) outliers

If \( r = 1 \) or \(-1\), scatterplot can have only 1 possible appearance; but if \( r = 0 \) it could look like any of above 3 plots.
Training your eye to read correlation values.

All plots have \( \bar{x} = \bar{y} = 3 \), \( s_x = s_y = 1 \).
(a) units of $r^2$

ex. $X = \text{income}$

$Y = \text{height}$

(b) switching $X$ and $Y$ leaves $r$ unchanged.

(c) Adding a (pos. or neg.) constant to $X$ also doesn't change $r$.

(d) Adding a (pos. or neg.) constant to $Y$ also leaves $r$ alone.

(e) Multiplying $X$ or $Y$ by a positive constant (surprisingly) also leaves $r$ alone.

$r = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$
Sir Francis (Galton)

reasonable line for purpose:

SD line: slope \( \frac{5}{5} \)

not the best line for predicting \( y \) from \( x \)

multi. + 6