average height of fathers ≈ 68 inches, SD ≈ 2.7 inches
average height of sons ≈ 69 inches, SD ≈ 2.7 inches, r ≈ 0.5

ALL VERTICAL STRIPE AVES.

x = 75
Possible father height
y = avg height of sons whose fathers
    HTS were ≈ 75

Freedman Pisani Pursero, 1980

- Statistics -
  Regrline = Plot of vertical stripe averse is particular
to elliptical plots.
\[ \bar{x} = 19.56 \]
\[ \bar{y} = 10.39 \]
\[ r = 0.775 \]
\begin{align*}
\{\text{Mean[hw1a], Mean[e1a]}\} & \quad 1.0 \\
\{\sigma[hw1a], \sigma[e1a]\} & \quad 1.0 \\
r[hw1a, e1a] & \quad 1.0 \\
r[hw1a, e1a] \sigma[e1a]/\sigma[hw1a] & \quad 1.0
\end{align*}

\begin{align*}
\{19.5652, 10.3913\} \\
\{2.96099, 1.83538\} \\
0.775338 \\
0.480595
\end{align*}
Lines of regression with and without the point (10, 9).
For elliptical plot-

If father's HT is $x_1$ above $\bar{x}$: $(x+\delta x) + 1$ if $\rho$ (correlation) is $\rho = .8$

Then, the avg HT of sons born to fathers of HT $\bar{x} + \delta x$ is $\bar{x}_2 = \bar{y} + \delta y$

$\bar{y} + .8 \delta y$

Another: If we look at all students whose exam 1 score is $\bar{x} + 2 \delta x$, then their avg score on exam 2 is $\bar{y} + 2 \delta y = \bar{y} + 1.6 \delta y$
WHERE CORRELATION CAME FROM

IDENTIFY PLOT THAT IS LIKE A LINE

\[
\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})/(n-1)
\]

\[
x \cdot y = n
\]

INSENSITIVE TO LOC + (POS) SCALE CHANGES.

\[|z| \leq 1 \quad \iff \quad |z_1 - 1| \leq 1 \quad \iff \quad \text{COINCIDENT w/ NAIVE LINE in this } \alpha = 1 \text{ CASE}
\]

\[|z| = 1 \iff \text{ALL THE POINTS UNE Perfectly on (THE REG) LINE}
\]

NOTE: \[
(x \bar{y} - \bar{x} \bar{y})/(\sqrt{x^2 - \bar{x}^2} \sqrt{y^2 - \bar{y}^2})
\]
STT 200 (K7 (cont.))

Exam 1 grade = 2 + 0.5(score - 8)

log(score) = grade

8 → grade = 2 + 0.5(8 - 8) = 2

10 → grade = 2 + 0.5(10 - 8) = 3

12 → grade > 12

Will send a post now due next Thu.

Keep up!
SON Y

\[
\sum \frac{(x - \bar{x})(y - \bar{y})}{n-1}
\]

SON'S HEIGHT IN INCHES

FATHER'S HEIGHT IN INCHES

GAITON'S DATA

\[
x = \frac{\sum (x - \bar{x})^2}{n-1}
\]

REGRESSION LINE

\[
y = \bar{X} + \beta_y (x - \bar{x})
\]

INFLUENCES INCL DIET, GENES, 

FACTOR X
average height of fathers ≈ 68 inches, SD ≈ 2.7 inches
average height of sons ≈ 69 inches, SD ≈ 2.7 inches, $r = 0.5$

**Claim:** For elliptical plots the vertical strip avoids PLOT as regression line.

For every PLOT LS LINE = REGRESSION LINE.
EX 1

OUTLIER?

\( \bar{x}_2 \)  \( \bar{y}_2 \)  

\{19.5652, 10.3913\}
\{2.96099, 1.83538\}
0.775338
0.480595

\( n = 23 \)
\( a_x = 2.96 \)  \( b_y = 1.83 \)
\( r = 0.775 \)

IF YOU WERE 2 \( a_x \) ABOVE \( \bar{x} \) (ON HW1) REG LINE PREDICTS YOU WILL BE \( \lambda (2b_y) = 2 (0.775) b_y \) ABOVE \( \bar{y} \).
\{\text{Mean[hw1a], Mean[e1a]} \} 1.0
\{\sigma[hw1a], \sigma[e1a]\} 1.0
r[hw1a, e1a] 1.0
r[hw1a, e1a] \sigma[e1a]/\sigma[hw1a] 1.0
\{19.5652, 10.3913\}
\{2.96099, 1.83538\}
0.775338
0.480595
Lines of regression with and without the point (10, 9).
Define the coefficient of determination (r):

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

\[ = \frac{\text{CAUTION} \ x \ y}{\text{ } \ x \ y} \]

\[ = \frac{\overline{xy} - \overline{x} \overline{y}}{(\text{AVE of products}) - (\text{PROD of AVGS})} \]

\[ \text{AVE} = \frac{x^2}{n} \quad \text{variance} = \frac{\sum (x-x)^2}{n-1} \]

\[ \text{AVE} = \frac{\overline{x}^2}{n} \quad \text{AVE} = \frac{\overline{x} \overline{y}}{n} \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x^2</th>
<th>y^2</th>
<th>xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>36</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \overline{x} = \frac{\sum x}{n} = \frac{9}{3} = 3 \]

\[ \overline{y} = \frac{\sum y}{n} = \frac{9}{3} = 3 \]

\[ s^2 = \frac{\sum (x-x)^2}{n-1} = \frac{9}{2} = 4.5 \]

\[ s = \sqrt{4.5} = \sqrt{2.25} = 1.5 \]

\[ r = \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}} \]

\[ \text{AVG of squares} = \frac{\sum x^2}{n} = \frac{9}{3} = 3 \]

\[ \text{AVG of products} = \frac{\sum xy}{n} = \frac{9}{3} = 3 \]

\[ \text{r} = \frac{3}{3} = 1 \]

\[ \text{r} = 1 \quad \text{perfect positive correlation} \]
Plot (x, y) pairs -