This time: correlation & regression
Next time: finish correlation & regression

Read: LN pp L-245 - 268  This time: LN p.p. L - 225

facts about r

\( r = 0 \) → 3 different scatterplot slope possibilities

- \( r \approx 0 \) no linear assoc.  non-linear association
- \( r \approx 0 \) outliers can distort, so not useful in predicting
- \( r \approx 0 \) strong non-linear relationship, especially with small n

adding a constant \((c > 0)\) to \(x\) \((c < 0)\)

leaves \(r\) unchanged!

multiply all \(x\) by 2: same tilt!

(mult. by a neg. number gives a reverse tilt)

\[ r = -0.87 \] (wing, tail) sparrows example

\( \hat{y} = \) predicted \(y\)

\( \hat{y} = \) predicted \(y\)

for population model: mean = \(\mu_y\), \(s_0 = 6\) \(y\), corr = \(\rho (\text{Roh})\)

math fact:

\[ SE_{\Sigma \rho} (r) = \frac{1 - \rho^2}{\sqrt{n-2}} \]

\[ SE_{\Sigma \rho} (r) = \frac{1 - r^2}{\sqrt{n-2}} \]
\( r = +0.87 \quad r \pm 1.96 \ \text{SE}(r) \quad \text{large-}n \text{ approx. 95\% CI} \)

\[ \text{SE}(r) = 0.16 \quad (12 \text{ sparrows is not large } n, \text{ do it anyway}) \]

\[ \text{95\% CI for } \rho \]

\[ (0.55 \pm 1) \quad 0.55 \quad 0.87 \quad \text{truncated at 1.00} \]

\textbf{Predicting y value from x.}

\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \]

\textit{regression line equation}

\[ \hat{\beta}_1 = r \frac{S_Y}{S_X} \]

\[ \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \overline{x} \]

\( \eta \approx 1000 \text{ families with at least 1 son} \)

\text{a random son + ht}

\text{fater + ht}

"regressing" toward the mean:

tall fathers have tall sons but not as tall as they were
short fathers had short sons, but not as short as them

\text{note: cm of tail length } \neq \text{ cm of wing length } \rightarrow \text{ units don't cancel!}
Another way to get the best line for pred. y from x

bad line

\[ y = \beta_0 + \beta_1 x \]

\[ \frac{1}{n} \sum_{i=1}^{n} \left[ y_i - (\beta_0 + \beta_1 x_i) \right]^2 \]

find \( \beta_0, \beta_1 \) to minimize

result: least squares line (Gauss 1800)

math fact: regression line = least squares line

reg. line for pred. x from y

5D line - capturing the trend

reg. line for predicting y from x

this quarter's focus