This:

**Correlation**

**Time:**

**Regression**

**Read:** FDP Ch. 10, 11

**Ass:**

**HWkS**

Next:

**No:**

**Wed:**

**Jun:**

Due Wed 7 Jun;

**Final Exam Review**

Mon 12 Jun 4-6 PM Here media theater

**HWk 6** Due Mon 12 Jun 4 PM at Final Exam Review

**Office Hrs the 6 Jun Rescheduled**

**Note Change**

**CS 12:**

**Correlation** (VIII A. p. 8)

\[ y \]

\[ x \]

\[ X \]

\[ Y \]

\[ X_1 \]

\[ Y_1 \]

\[ X_2 \]

\[ Y_2 \]

\[ \vdots \]

\[ X_n \]

\[ Y_n \]

\[ n \]

\[ \bar{X} = 68 \, \text{IN} \]

\[ S_x = 2.7 \, \text{IN} \]

\[ n = 1078 \]

\[ \bar{Y} = 69 \, \text{IN} \]

\[ S_y = 2.7 \]

Here \( X \) = HT. of Father

Here \( Y \) = HT. of Son

**Independent (Predictor) Variable**

**Outcome (Response) Variable**

**Dependent (Dependent) Variable**
Positive association ($r > 0$)

Point of average ($\bar{x}, \bar{y}$)

Pearson negative association

$r = \text{strength of straight-line association between } x \text{ and } y$

① ② $r$ can be fooled by:
- **Nonlinearity**
- **Outliers**

$v = 0$
\[
\sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s_x} \right) \cdot \left( \frac{y_i - \bar{y}}{s_y} \right) = \bar{r}
\]

(AVE OF PRODUCTS OF VARIABLES IN STANDARD UNITS)

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

\[s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2}\]

CORRELATION COEFFICIENT

Basic Facts

\[\bar{r}\] ALWAYS COMES OUT BETWEEN -1 AND +1

-1
0
+1

PERFECT NEGATIVE LINEAR ASSOCIATION

\[\text{NON-LINEAR}\]

PERFECT POSITIVE LINEAR ASSOCIATION

OUTLIER

HEALTHY, NO ASSOCIATION
Outlier (can confuse $r$, especially when $n$ is small)

3. Units or $r$? (Ex. $x = \text{income (s)}$)

4. Switching $x$ & $y$ leaves unchanged

5. Adding a constant to $x$ or $y$ (Ex. $x \rightarrow x + 5$)

Also leaves $r$ unchanged
1) Multiplying $x$ or $y$ by a positive constant seems to change $\checkmark$ visually but in fact this also leaves $\checkmark$ unchanged.

\[ \frac{\text{mult} x}{y} \]

\[-5 \quad +5 \quad +10 \quad 15 \]