BIOSTATISTICS 140.653

POPULATION ASSOCIATIONS

- SETTING: (X, Y) are variables associated with a single unit in a population.
- COVARIANCE: $\text{Cov}(X,Y) = \text{E}[(X - \text{E}[X])(Y - \text{E}[Y])]$,
  with $\text{E}[X]$ = population mean of X (etc.)
- CORRELATION:

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \, \sigma_Y}$$

  with σ_X^2 = population variance of X (etc.)

- PROPERTY: $-1 \le \rho_{XY} \le 1$

PROOF:

$$\begin{aligned} (1)\ \text{Var}\,(X/\sigma_X - Y/\sigma_Y) &= \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} - \frac{2\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \\ &= 2 - 2\rho_{XY} \\ &\ge 0 \quad \text{(variance property)} \end{aligned}$$

Therefore, $\rho_{XY} \le 1$

Similarly,

$$\begin{aligned} \text{Var}\,(X/\sigma_X + Y/\sigma_Y) &= 2 + 2\rho_{XY} \\ &\ge 0 \end{aligned}$$

Therefore, $\rho_{XY} \ge -1$.

- SHORTCUT COVARIANCE FORMULA:

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

POPULATION CALCULATIONS

- <u>$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$</u>

  <u>PROOF</u>: $\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$ (definition)

  $$E[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y]$$

  with $\mu_X = E[X]$ (etc.)

  $$E[XY] - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$$

  $$E[XY] - E[X]E[Y]$$

- <u>EQ(1)</u>: $\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)$

  $$E\left[\left\{\left(\frac{X}{\sigma_X} \quad \frac{Y}{\sigma_Y}\right)^2 - E\left[\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right]\right\}^2\right]$$

  $$E\left[\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)^2 + \left\{E\left[\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right]\right\}^2 - 2\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)E\left[\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right]\right]$$

  $$E\left[\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)^2 + \left(\frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\sigma_Y}\right)^2 - 2\left(\frac{X}{\ } - \frac{Y}{\ }\right)\left(\frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\sigma_Y}\right)\right]$$

  with $\mu_X = E[X]$ (etc.)

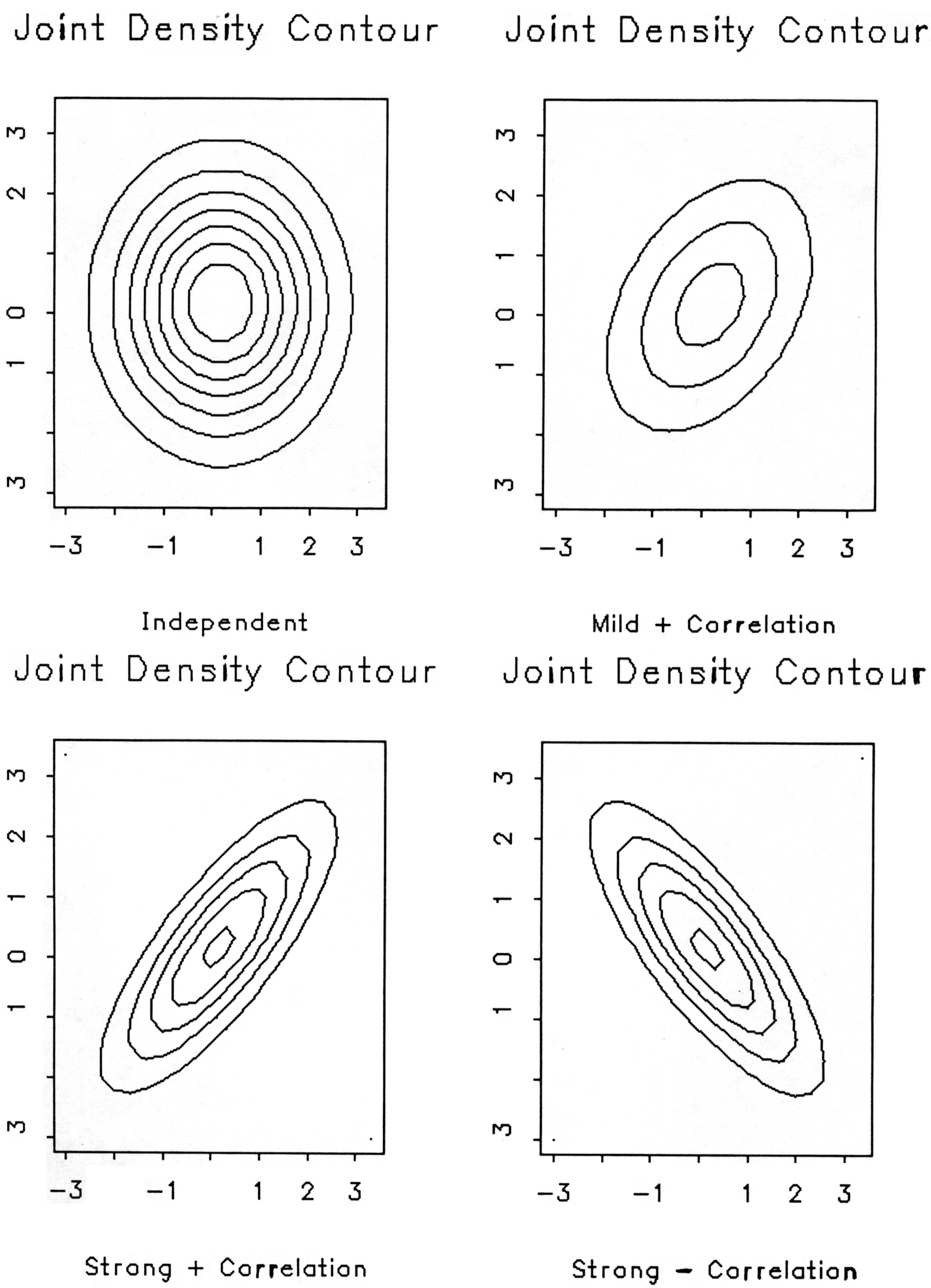
  $$= E\left[\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)^2\right] + \left(\frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\sigma_Y}\right)^2 - 2E\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)\left(\frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\sigma_Y}\right)$$

  $$= E\left[\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)^2\right] - \left(\frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\ }\right)^2$$

  $$= E\left[\frac{X^2}{\sigma_X^2} + \frac{Y^2}{\sigma_Y^2} - \frac{2XY}{\sigma_X\sigma_Y}\right] - \left(\frac{\mu_X^2}{\sigma_X^2} + \frac{\mu_Y^2}{\sigma_Y^2} - \frac{2\mu_X\mu_Y}{\sigma_X\sigma_Y}\right)$$

  $$= E\left[\frac{X^2}{\sigma_X^2}\right] - \frac{\mu_X^2}{\sigma_X^2} + E\left[\frac{Y^2}{\sigma_Y^2}\right] - \frac{\mu_Y^2}{\sigma_Y^2} - \frac{2E[XY]}{\sigma_X\sigma_Y} + \frac{2\mu_X\mu_Y}{\sigma_X\sigma_Y}$$

  $$= \frac{\mathbf{Var[X]}}{\sigma_{\mathbf{X}}^2} + \frac{\mathbf{Var[Y]}}{\sigma_{\mathbf{Y}}^2} - \frac{\mathbf{2Cov(X,Y)}}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}$$


Joint Density Contour
3
2
0
1
3
−3
−1
1
2
3
Independent
Joint Density Contour
3
2
1
0
1
3
−3
−1
1
2
3
Mild + Correlation
Joint Density Contour
3
2
1
0
1
3
−3
−1
1
2
3
Strong + Correlation
Joint Density Contour
3
2
1
0
1
3
−3
−1
1
2
3
Strong − Correlation

BIOSTATISTICS 140.653

SAMPLE ASSOCIATIONS

- SETTING: Two variables X, Y with measurements

$$X_1, \ldots, X_i, \ldots, X_n$$

$$Y_1, \ldots, Y_i, \ldots, Y_n$$

(X_i, Y_i) measured on same sampling unit "i"

- COVARIANCE:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}),$$

$$\text{with } \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \ ; \ \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

- CORRELATION: (Pearson, Product-moment)

$$r_{XY} = \frac{s_{XY}}{s_X s_Y}$$

$$\text{with } s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \ ; \ s_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

- SQUARED-AND CROSS-PRODUCT NOTATION

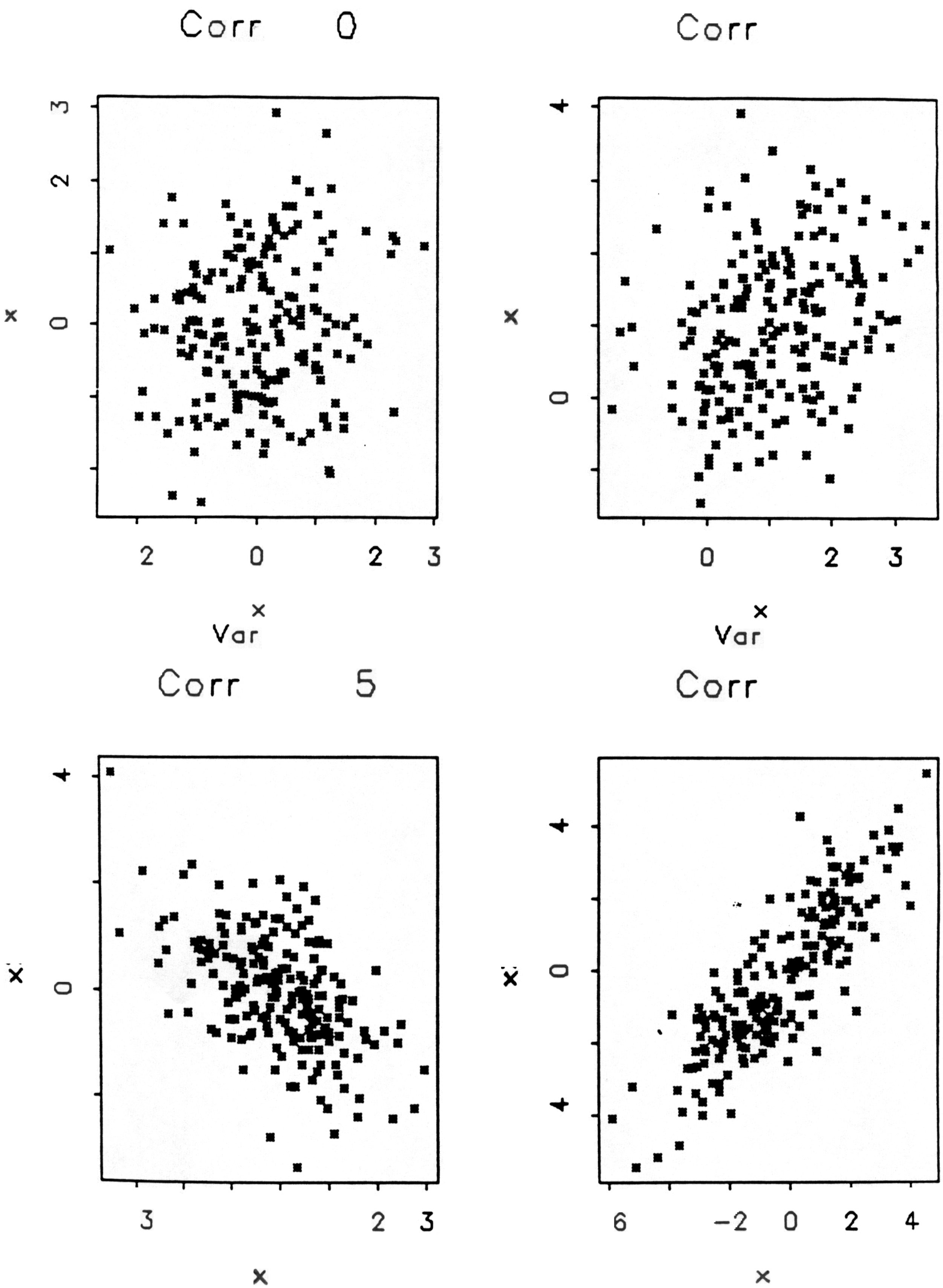
$$S_{XX} = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

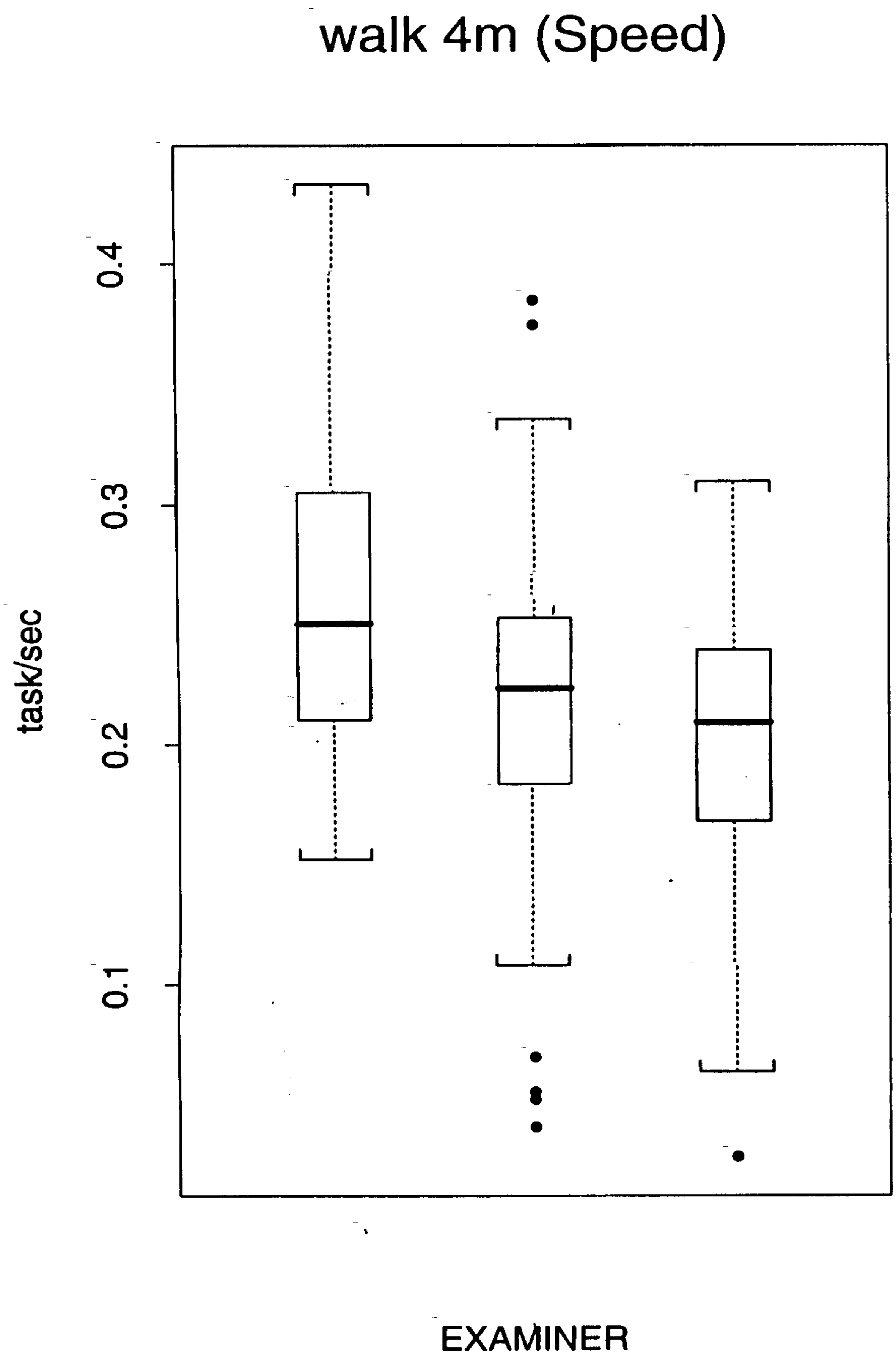
$$S_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

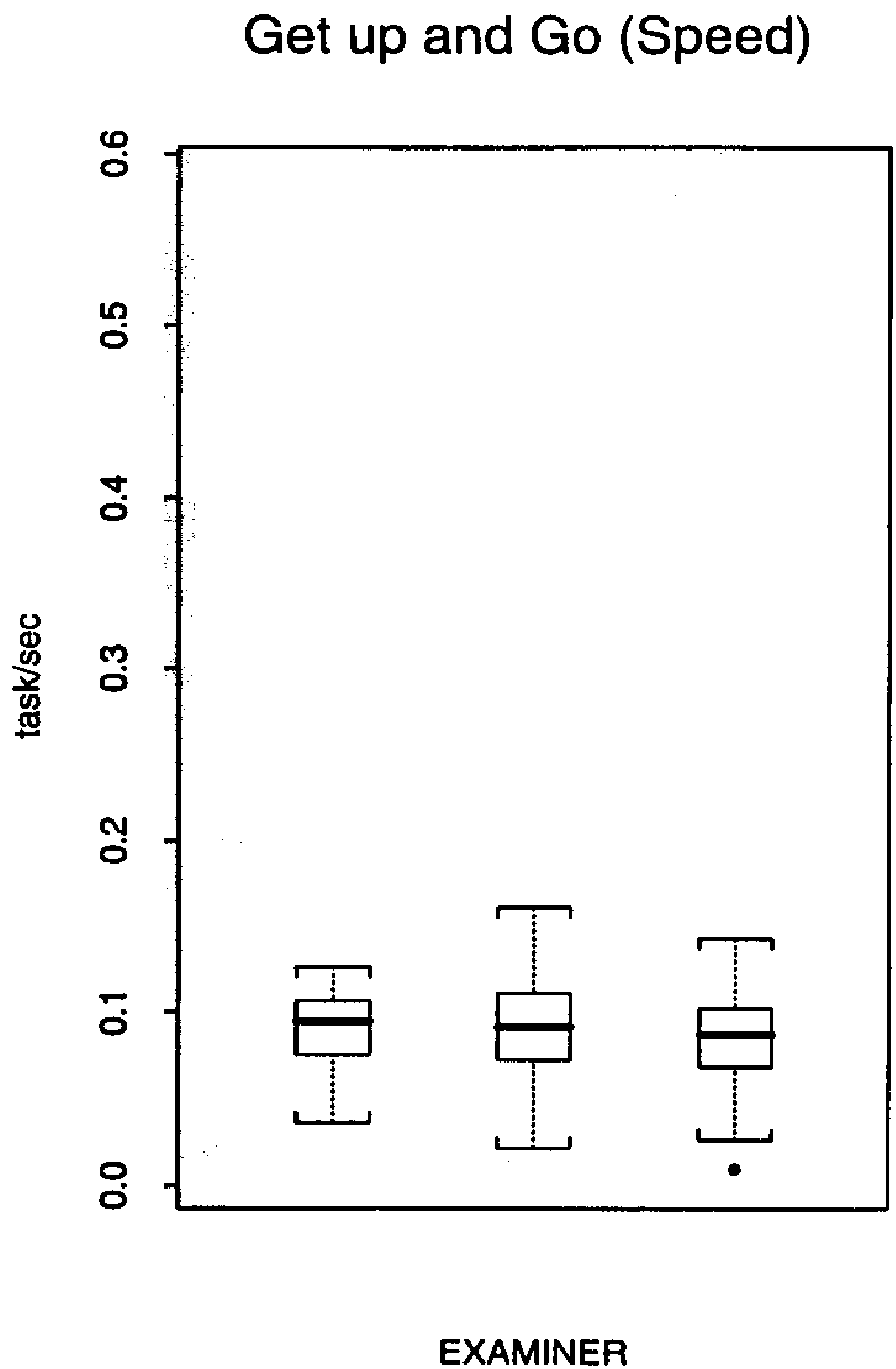
$$S_{XY} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$\Longrightarrow$ e.g.,

$$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$$


Corr 0
Corr
x
Var
x
Var
Corr 5
Corr
x
x
x
x


walk 4m (Speed)
task/sec
0.1
0.2
0.3
0.4
EXAMINER



Get up and Go (Speed)
task/sec
0.0
0.1
0.2
0.3
0.4
0.5
0.6
EXAMINER