Development Statistics

S11 Correlation

Fujikawa, Kiyoshi
Nagoya University, GSID

Sample Covariance

- Positive correlation
- Negative correlation

$$\mathrm{cov}(v, y) = s(x, y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

- Variance is a special case that x=y

Illustration 1

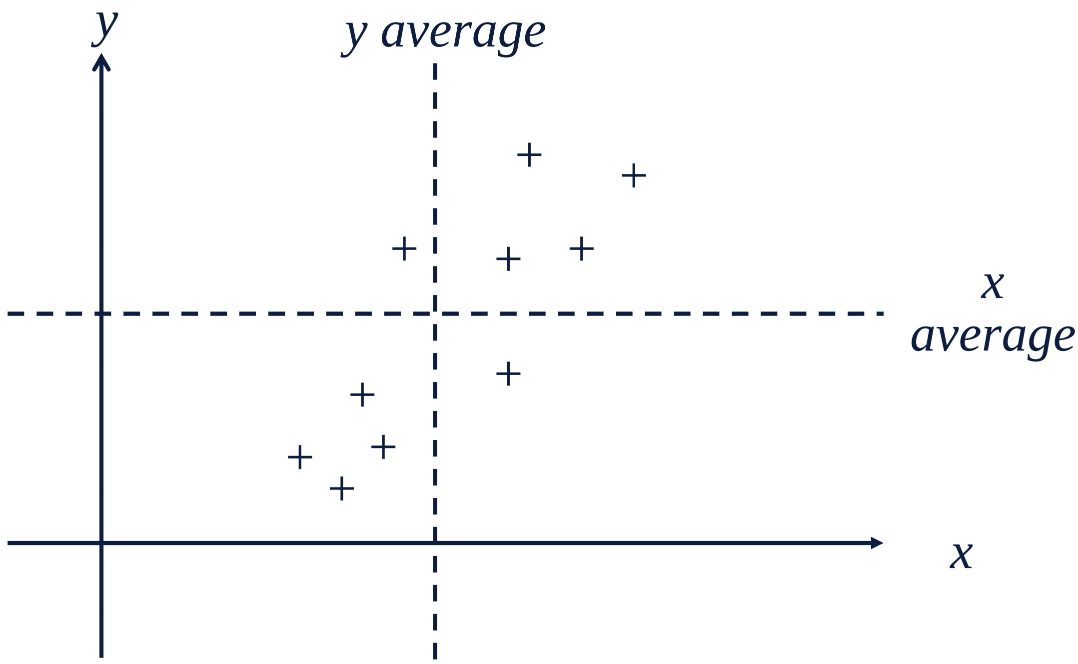
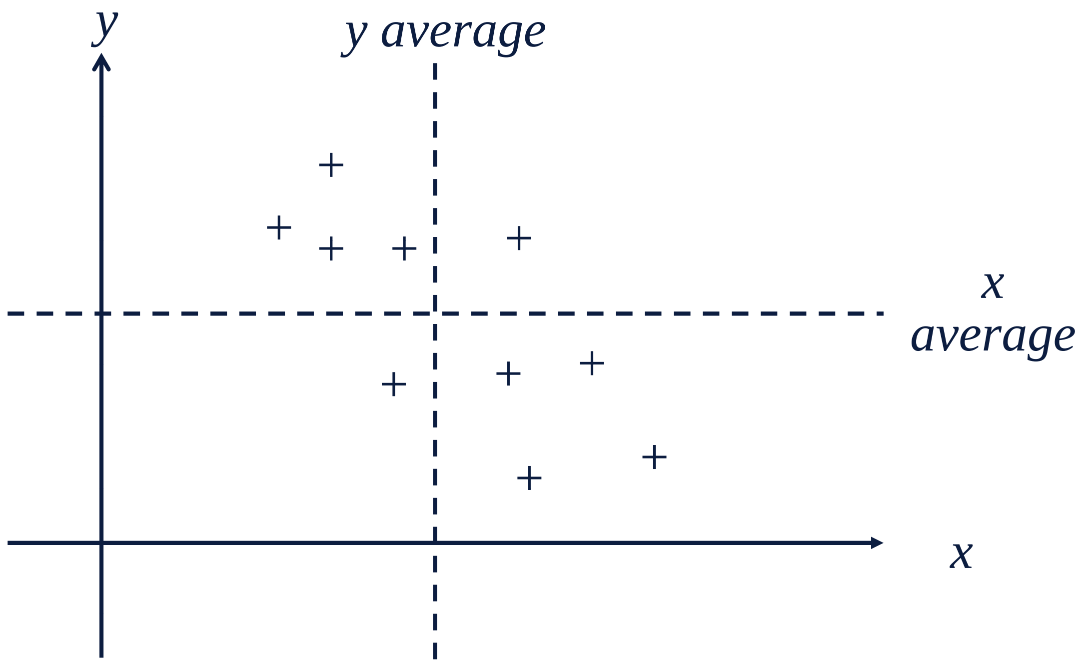


Illustration 2



Sample Correlation

- However, variance and covariance are unit-dependent
- Normalization of covariance

$$r(x,y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} = \frac{S(x,y)}{S(x)S(y)}$$

$$-1.0 \le r(x,y) \le 1.0$$

Test of the correlation of the population

- Suppose the population follows bivariate normal distribution
- If the correlation of the population is 0
- r : sample correlation, n : sample size

$$T(r) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$$

Correlation & t-value

- T-test of regression analysis

$$y_i = \alpha + \beta x_i + u_i$$

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{b-\beta}{SE(b)} \sim t(n-2)$$

Standard error of regression coefficient

- Regression line $\hat{y}_i = a + bx_i$
- Residual $y_i - \hat{y}_i = y_i - (a + bx_i) = e_i$
- Estimator of variance $s^2 = \sum_i e_i^2 \big/ (n-2)$
- Standard error $s = \sqrt{\sum_i e_i^2 / (n-2)}$