Least Squares Linear Regression

Important Formulas

① $\bar{X} = \hat{\mu}$; $S_X^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ Most famous statistics

② Sample Covariance (S_{XY}) and correlation coefficient (r)

$$r = \hat{\rho} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_X}\right)\left(\frac{Y_i - \bar{Y}}{S_Y}\right) = \left(\frac{1}{S_X S_Y}\right)\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{S_{XY}}{S_X S_Y}$$ [1]

Coefficient of determination r^2 = fraction of variation in y explained by regression y on x= $\frac{\mathrm{Var}(\hat{y})}{\mathrm{Var}(y)}$

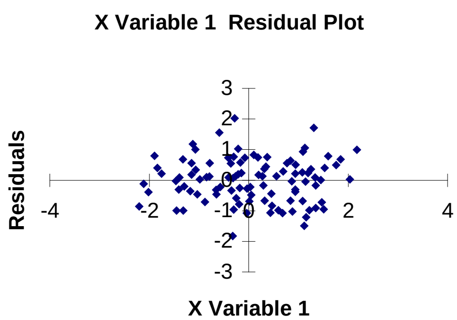
There are alternative formulas.[2]

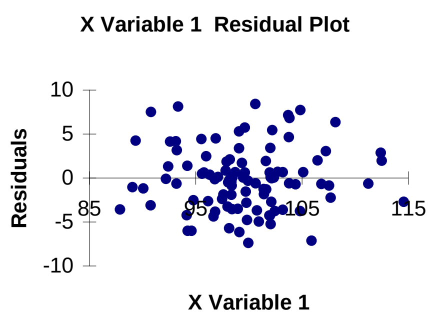
③ Regression Line: $\hat{y} = a + bx$, $\begin{cases} b = r\frac{s_y}{s_x} \\ a = \bar{y} - b\bar{x} \end{cases}$

NOTE: Slope of regression line $m \propto r$

④ Residuals (errors): $e_i = y_i - \hat{y}_i$, or $e = y - \hat{y} = y - a - bx$

$$\sum e^2 = \sum (y - a - bx)^2 = \sum (y_i - \hat{y}_i)^2$$





---

[1] Alternate: $S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n} X_i Y_i - \frac{n}{n-1}\bar{X}\bar{Y}$

[2] Alternate $b = \frac{\Sigma XY - \Sigma X \cdot \bar{Y}}{\Sigma X^2 - \bar{X}\Sigma X}$, and all the other formulas in your book.