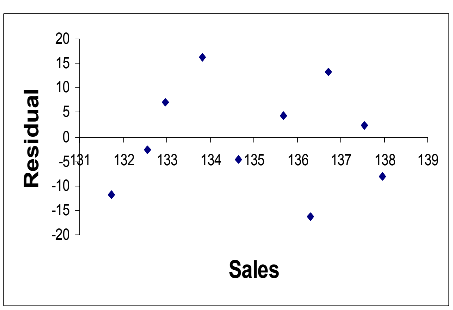
Prediction of new observations

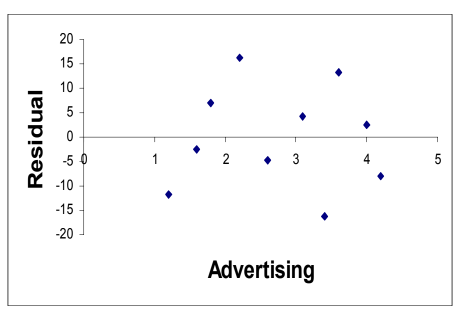
- Regression model: $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$
- A 100(1-α) percent interval on a future observation y_0 at the value x_0 is given by:

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

$$\leq y_0 \leq \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

Residual Analysis





Coefficient of determination (R^2)

- Definition

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \qquad 0 \le R^2 \le 1$$

- R^2 is the amount of variability in the data explained by the regression model

- Misconceptions about R^2

Lack-of-fit Test

- Hypotheses:
  - H_0: The simple linear regression model is correct
  - H_1: The simple linear regression model is not correct

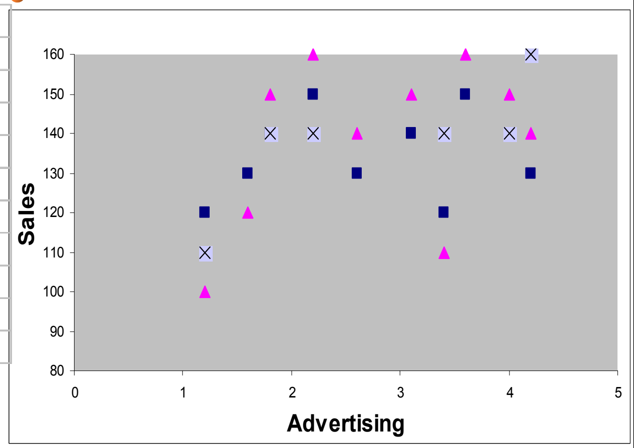
- Error or residual sum of square

$$SS_E = SS_{PE} + SS_{LOF}$$

  - SSPE: Sum of squares attributable to **pure error**
  - SSLOF: Sum of squares attributable to the **lack-of-fit** of the model

Lack-of-fit Test

x	y1	y2	y3
1.2	120	100	110
1.6	130	120	
1.8	140	150	140
2.2	150	160	140
2.6	130	140	
3.1	140	150	
3.4	120	110	140
3.6	150	160	
4	140	150	140
4.2	130	140	160



Lack-of-fit Test

$$SS_{PE} = \sum_{i=1}^{m} \sum_{u=1}^{n_i} (y_{iu} - \bar{y}_i)^2$$

$$\text{with } n_{pe} = \sum_{i=1}^{m} (n_i - 1) = n - m \text{ df}$$

$$SS_{LOF} = SS_E - SS_{PE}$$

$$\text{with } n - 2 - n_{pe} = m - 2 \text{ df}$$

- Test statistic $F_0 = \dfrac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)} = \dfrac{MS_{LOF}}{MS_{PE}}$

Reject H_0 if $f_0 > f_{\alpha,m-2,n-m}$