1.6. Le coefficient de détermination R^2 : une mesure de la qualité de l'ajustement

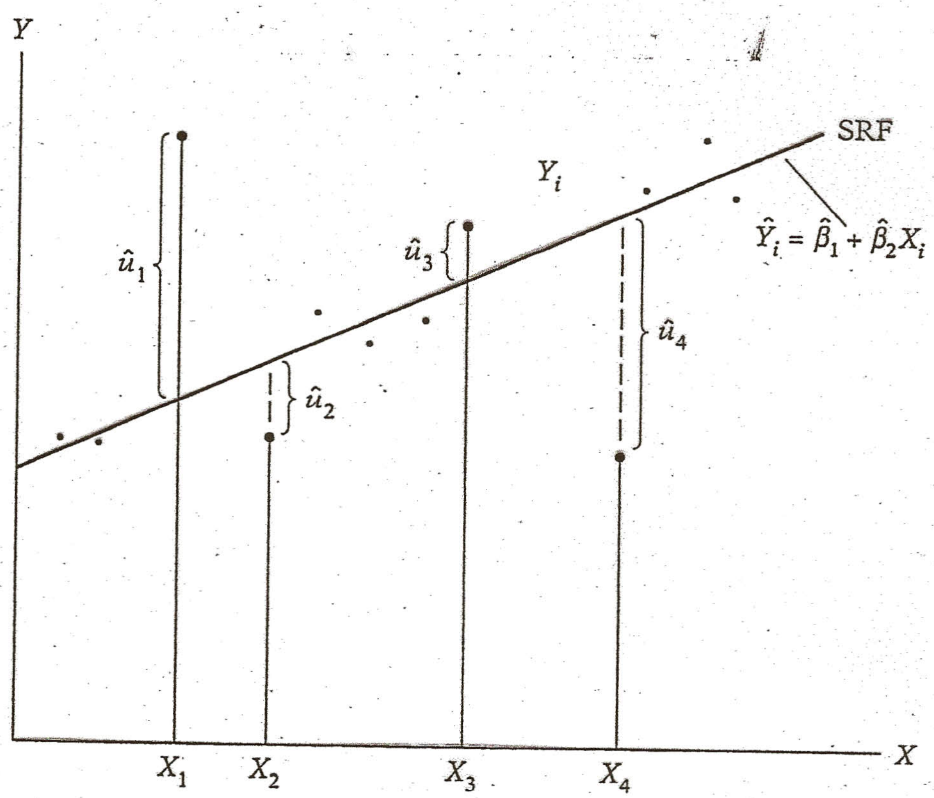

Figure 3.1 : Least-squares criterion.

La qualité de l'ajustement des observations par la droite de régression est mesurée par le coefficient de détermination R^2

Nous savons que : $Y_i = \hat{Y}_i + \hat{u}_i \implies y_i = \hat{y}_i + \hat{u}_i$

$$\boxed{\sum y_i^2} = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i$$

$$= \boxed{\sum \hat{y}_i^2 + \sum \hat{u}_i^2} \quad \text{(car } \sum \hat{y}_i \hat{u}_i = 0 \text{, propr. numérique des MCO)}$$

$$= \boxed{\hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2} \quad \text{(car } \hat{y}_i = \hat{\beta}_2 x_i \text{)}$$

- Que représentent ces ≠ Σ de carrés ?

① $\sum y_i^2 = \sum (Y_i - \bar{Y})^2$: variat° totale des valeurs observées autour de leur moyenne (échantillon)

$\equiv$ Somme des carrés totale (SCT)

② $\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{\hat{Y}})^2$

$= \sum (\hat{Y}_i - \bar{Y})^2$ (car $\bar{\hat{Y}} = \bar{Y}$, prop. num. de MCO)

$(= \hat{\beta}_2^2 \sum x_i^2)$

: variation des valeurs estimées autour de leur moyenne (échantillon)

$\equiv$ Somme des carrés expliquée (SCE)

③ $\sum \hat{u}_i^2$ $\equiv$ Somme des carrés résiduelle (SCR)

: variation résiduelle

$\Rightarrow$ $\boxed{SCT = SCE + SCR}$

La variation totale des valeurs observées (Y) autour de leur moyenne peut ê subdivisée en 2 parties : l'une qui attribuable à la droite de régression et l'autre à des forces aléatoires.

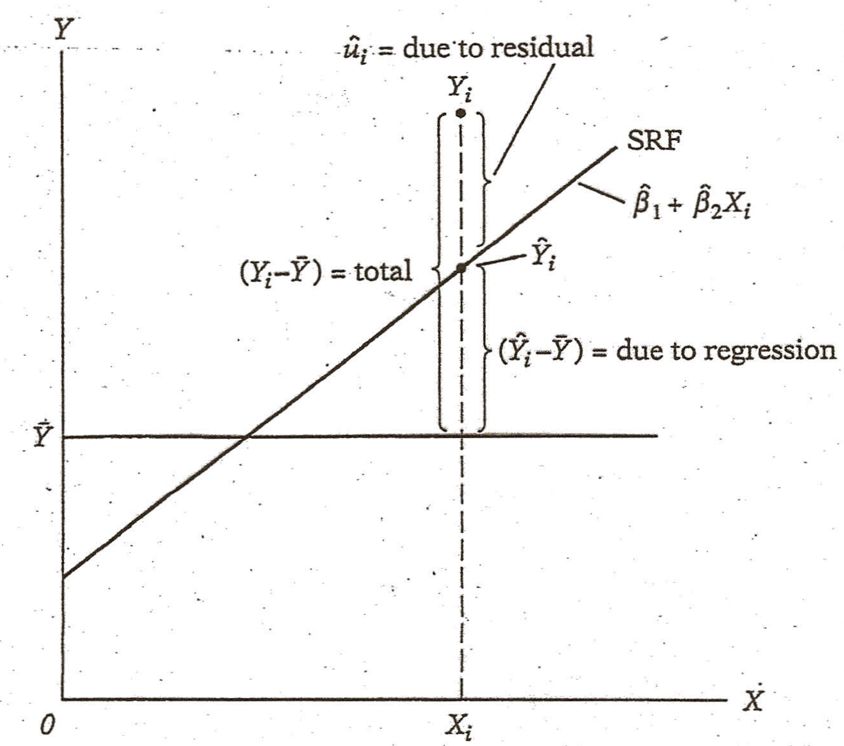

Fig. 3.10. Breakdown of the variation of Y_i into two components.

- $\boxed{SCT = SCE + SCR}$

$$\Rightarrow \quad 1 = \frac{SCE}{SCT} + \frac{SCR}{SCT}$$

$$= \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2} + \frac{\Sigma \hat{u}_i^2}{\Sigma(Y_i - \bar{Y})^2}$$

- $$\boxed{R^2 = \frac{SCE}{SCT}} = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2}$$

ou encore,

$$\boxed{R^2 = 1 - \frac{SCR}{SCT}} = 1 - \frac{\Sigma \hat{u}_i^2}{\Sigma(Y_i - \bar{Y})^2}$$

<u>Propriétés du R^2</u> :

a) Il ne prend jamais de valeur négative.

b) Il est compris entre 0 et 1 $(0 \leq R^2 \leq 1)$

si $R^2 = 1$, ajustmt parfait $(\hat{Y}_i = Y_i)$

si $R^2 = 0$, pas de relation tel X et Y $\rightarrow \hat{Y}_i = \hat{\beta}_1 = \bar{Y}$

- <u>En pratique</u> ?

Echantillon aléatoire de 10 ménages dont on connait le revenu et la consommation :

Table 3.2. Dép. et revenu (hebdomadaire) en \$ de 10 ménages

Revenu X	Dép. de cons. Y
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

- Estimation de la relation entre les revenus (X) et les dép. de cons. (Y) par les MCO :

$\hat{\beta}_1 = 24,4545$, $\hat{\beta}_2 = 0,5091$

$se(\hat{\beta}_1) = 6,4138$, $se(\hat{\beta}_2) = 0,0357$

$\hat{\sigma}^2 = 42,1591$, $cov(\hat{\beta}_1, \hat{\beta}_2) = -0,2172$

$R^2 = 0,9621$

$$\Rightarrow \boxed{\hat{Y}_i = 24,4545 + 0,5091\, X_i}$$

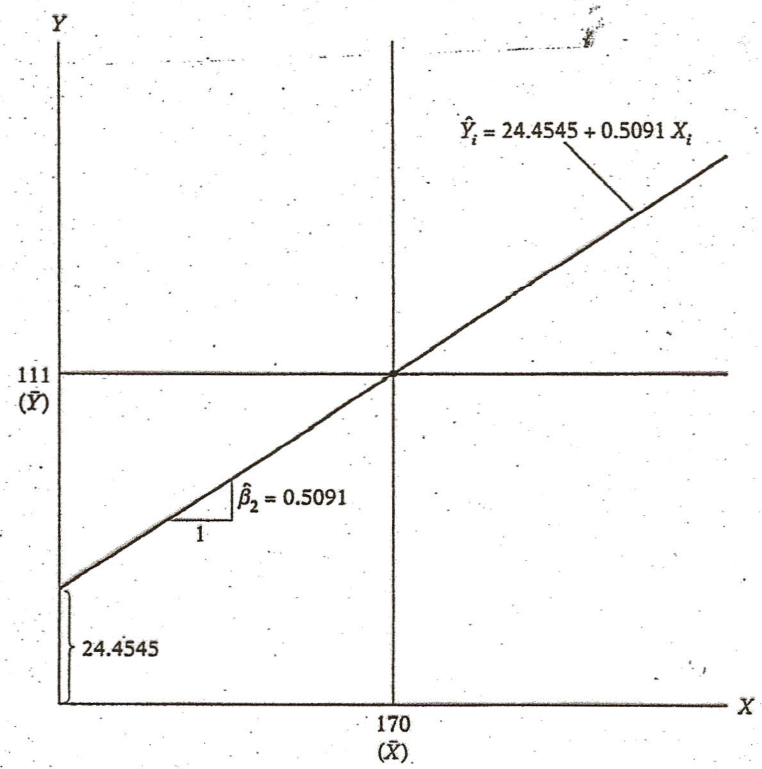

FIGURE 3.12 Sample regression line based on the data of Table 3.2.