Démonstration de l'égalité qui apparait successivement sur les slides 41 et 42 du Cours 1

D'après le chapitre 1 du livre *Analyse de régression appliquée* de Yadolah Dodge, Dunod, 2004.

Vous trouverez ici la démonstration d'une propriété des moindres carrés (slides 41 et 42 du Cours 1) qui est la suivante :

Propriété 0.1.

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2. \tag{0.1}$$

Démonstration : On a :

$$y_i-\overline{y}=(y_i-\hat{y}_i)+(\hat{y}_i-\overline{y}). \tag{0.2}$$

On élève au carré l'égalité (**??**) et on somme sur i, qui est une variable muette. En utilisant la première identité remarquable

$$(a+b)^2=a^2+2ab+b^2,$$

où

$$a=(y_i-\hat{y}_i)$$

et

$$b=(\hat{y}_i-\overline{y}),$$

on obtient alors :

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\overline{y})+\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2. \tag{0.3}$$

Maintenant occupons nous du terme

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\overline{y}),$$

qui provient de l'égalité (**??**). Auparavant, établissons une égalité très importante :

Proposition 0.1.

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0. \tag{0.4}$$

Démonstration.

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \overline{x}) &= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} \\ &= n\overline{x} - n\overline{x} \\ &= 0.\end{aligned}$$

Développons le terme $\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y})$. On obtient alors :

$$\begin{aligned}\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) &= \sum_{i=1}^{n}(y_i\hat{y}_i - \overline{y}y_i - {\hat{y}_i}^2 + \overline{y}\hat{y}_i) \\ &= \sum_{i=1}^{n} y_i\hat{y}_i - \overline{y}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}{\hat{y}_i}^2 + \overline{y}\sum_{i=1}^{n}\hat{y}_i. \qquad (0.5)\end{aligned}$$

Établissons maintenant la proposition essentielle qui va nous permettre de démontrer la propriété **??** que l'on cherche à établir.

Propriété 0.2.

$$\sum_{i=1}^{n}\hat{y}_i = \sum_{i=1}^{n} y_i \tag{0.6}$$

et

$$\sum_{i=1}^{n}\hat{y}_i^2 = \sum_{i=1}^{n} y_i\hat{y}_i. \tag{0.7}$$

Démonstration de l'égalité (??) de la proposition ??:

Pour établir la première égalité de la proposition **??**, il faut se rappeler de deux définitions du cours de statistique bivariée. La première égalité qui va nous être utile pour commencer la démonstration de cette égalité est la suivante :

$$\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x_i$$

qui provient du slide 24. La seconde définition qui va servir dès la troisième ligne de la séquence d'égalités qui va suivre est la suivante :

$$\widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x}.$$

Cette égalité provient du slide 31. On obtient alors :

$$\begin{aligned}
\sum_{i=1}^{n} \widehat{y_i} &= \sum_{i=1}^{n} \left(\widehat{\beta_0} + \widehat{\beta_1} x_i \right) \\
&= \sum_{i=1}^{n} \widehat{\beta_0} + \sum_{i=1}^{n} \widehat{\beta_1} x_i \\
&= \sum_{i=1}^{n} (\overline{y} - \widehat{\beta_1} \overline{x}) + \sum_{i=1}^{n} \widehat{\beta_1} x_i \\
&= n\overline{y} - \widehat{\beta_1} \sum_{i=1}^{n} \overline{x} + \widehat{\beta_1} \sum_{i=1}^{n} x_i \\
&= n\overline{y} + \widehat{\beta_1} \left(\sum_{i=1}^{n} (x_i - \overline{x}) \right) \\
&= n\overline{y} + \widehat{\beta_1} \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} \right) \\
&= n\overline{y} + \widehat{\beta_1} \left(n\overline{x} - n\overline{x} \right) \\
&= n\overline{y} \\
&= \sum_{i=1}^{n} y_i,
\end{aligned}$$

ce qui achève la démonstration de l'égalité (**??**).

Démonstration de l'égalité (??) de la proposition ?? : .

$$\begin{aligned}
\sum_{i=1}^{n} \widehat{y_i}^2 &= \sum_{i=1}^{n} \left(\widehat{\beta_0} + \widehat{\beta_1} x_i \right)^2 \\
&= \sum_{i=1}^{n} \left(\overline{y} - \widehat{\beta_1} \overline{x} + \widehat{\beta_1} x_i \right)^2 \\
&= \sum_{i=1}^{n} \left(\overline{y} + \widehat{\beta_1} (x_i - \overline{x}) \right)^2 \\
&= n\overline{y}^2 + 2\overline{y}\beta_1 \sum_{i=1}^{n} (x_i - \overline{x}) + \widehat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \overline{x})^2 \\
&= \overline{y} n \overline{y} + 0 + \widehat{\beta_1}\widehat{\beta_1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\
&= \overline{y} n \overline{y} + \widehat{\beta_1} \sum_{i=1}^{n} \widehat{\beta_1} (x_i - \overline{x})^2
\end{aligned}$$

$$
\begin{aligned}
&= \overline{y}n\overline{y} + \widehat{\beta}_1 \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}), \\
&\quad \text{d'après le slide 33} \\
&= \overline{y}n\overline{y} + \widehat{\beta}_1 \sum_{i=1}^{n}((x_i - \overline{x})y_i) - \widehat{\beta}_1\overline{y}\sum_{i=1}^{n}(x_i - \overline{x}) \\
&= \overline{y}\sum_{i=1}^{n} y_i + \widehat{\beta}_1 \sum_{i=1}^{n}(x_i - \overline{x})y_i - 0 \\
&= \sum_{i=1}^{n}(\overline{y} + \widehat{\beta}_1(x_i - \overline{x}))y_i \\
&= \sum_{i=1}^{n}\widehat{y}_i y_i.
\end{aligned}
$$

Ces deux égalités démontrées achèvent la démonstration de la proposition **??**.

Ainsi l'égalité (**??**) devient nulle. Et par conséquent le double terme de l'égalité (**??**) devient nul et nous avons établi l'égalité (**??**).