

ANNEXE À LA TROISIÈME PARTIE :

NOTE DE TERMINOLOGIE :

ÉCART TYPE, ERREUR TYPE, ETC.

1. Définitions générales

- **Écart type** de la distribution d'une variable aléatoire v dans la **population** :

$$\sigma_v = \sqrt{\sigma_v^2}, \text{ avec}$$

$$\sigma_v^2 = \frac{1}{n} \sum_i (v_i - \mu_v)^2 \text{ (population finie)}$$

ou, de manière théoriquement rigoureuse, avec

$$\sigma_v^2 = E[(v_i - \mu_v)^2] \text{ (population finie ou infinie)}$$

- **Écart type d'un échantillon**

autrement dit, écart type d'une variable v dans un échantillon,

ou encore écart type de la distribution d'une variable v dans un échantillon

$$s_v = \sqrt{s_v^2}, \text{ avec } s_v^2 = \frac{1}{n-1} \sum_i (v_i - m_v)^2$$

s_v^2 est un estimateur non biaisé de σ_v^2

2. Application à une statistique et à sa distribution d'échantillonnage

Nous considérons désormais que

v est une **statistique** calculée sur les données d'un **échantillon**

Alors,

μ_v est la **moyenne** (l'espérance mathématique) de la **distribution d'échantillonnage** de v .

σ_v^2 est la **variance de la distribution d'échantillonnage** de v , autrement dit la variance d'échantillonnage de v .

σ_v est l'**écart type de la distribution d'échantillonnage** de v ,

autrement dit l'**écart type d'échantillonnage** de v

ou l'**erreur type** de v

ou l'**erreur d'échantillonnage** (*sampling error*) de v .

Toutes ces expressions sont synonymes.

3. Estimateurs de l'erreur d'échantillonnage

On aura remarqué que l'erreur d'échantillonnage est un paramètre inconnu de la distribution d'échantillonnage de la statistique v .

Mais on peut estimer ce paramètre à partir de données d'un échantillon. Cet estimateur est généralement dénoté s_v .

Il arrive que l'on désigne la valeur estimée s_v du paramètre inconnu σ_v comme l'« erreur d'échantillonnage ». Pour être parfaitement explicite, il faudrait dire « l'estimateur de l'erreur d'échantillonnage » ou « la valeur estimée de l'erreur d'échantillonnage ».

4. Exemples

4.1 ERREUR D'ÉCHANTILLONNAGE DE LA MOYENNE

On estime généralement μ_v , la moyenne inconnue de la population, au moyen de m_v la moyenne de l'échantillon.

La statistique mathématique montre que m_v a une distribution d'échantillonnage normale, avec

une moyenne égale à la vraie moyenne μ_v et un écart type égal à $\sqrt{\frac{\sigma_v^2}{n}} = \frac{\sigma_v}{\sqrt{n}}$: c'est l'erreur d'échantillonnage de la moyenne.

L'estimateur de cette erreur d'échantillonnage est donné par $\sqrt{\frac{s_v^2}{n}} = \frac{s_v}{\sqrt{n}}$.

On note en passant que le t de Student a la structure suivante :

$$t \text{ de Student} = \frac{\text{Différence entre la moyenne de l'échantillon et la moyenne hypothétique de sa distribution d'échantillonnage}}{\text{Erreur d'échantillonnage estimée de la moyenne}}$$

4.2 ERREUR D'ÉCHANTILLONNAGE DES COEFFICIENTS ESTIMÉS DANS LE MODÈLE CLASSIQUE DE LA RÉGRESSION LINÉAIRE NORMALE

La méthode des moindres carrés ordinaires produit

- des estimateurs b_j pour les coefficients β_j ,¹
- un estimateur s^2 pour la variance des termes d'erreur σ^2 ; cet estimateur est calculé au moyen de la formule

$$s^2 = \frac{1}{n-k} \sum_i (y_i - \hat{y}_i)^2 = \frac{SSR}{n-k} ;$$

- des estimateurs $s_{b_j}^2$ de la variance d'échantillonnage $\sigma_{b_j}^2$ des b_j (voir la note 2).

Le théorème de Gauss-Markov montre que, sous les hypothèses du modèle classique de la régression linéaire normale, les estimateurs b_j ont une distribution d'échantillonnage normale, avec une moyenne égale à β_j et une variance égale à $\sigma_{b_j}^2$. Par conséquent,

σ_{b_j} est l'**erreur d'échantillonnage** de b_j et s_{b_j} est l'estimateur de σ_{b_j} .

La statistique mathématique montre que la variable aléatoire

$$\frac{b_j - \beta_j}{s_{b_j}}$$

a une distribution de Student avec $n - k$ degrés de liberté. On note que cette variable a exactement la même structure que le t qui sert à tester une hypothèse simple sur une moyenne :

$$t \text{ de Student} = \frac{\text{Différence entre la valeur estimée du coefficient et la moyenne hypothétique de sa distribution d'échantillonnage}}{\text{Erreur d'échantillonnage estimée du coefficient}}$$

¹ En écriture matricielle, on a $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

² La matrice estimée des variances-covariances est donnée par $(\mathbf{X}'\mathbf{X})^{-1} s^2$.