PrivateTeacher
Cours Privés de Science

Variance Estimateur

Version 1.0

Au sujet de la variance

La variance est une grandeur qui mesure la dispersion des valeurs autours de leur moyenne.

Formellement, la variance est une somme des carrés(*) des écart entre chaque observation X_i et leur moyenne μ.

$$\boxed{Var(X) := \frac{1}{n} \sum_i (X_i - \mu)^2}$$

n : le nombre d'observation

μ : la moyenne des observation

X_i : la valeur de chaque observation

Cette définition est totalement générale

Elle s'utilise aussi bien pour calculer la variance d'une population que celle d'un échantillon.

(*)

On utilise le carré des écarts $(X_i - \mu)^2$ pour éviter que les différences positives ne compensent les différences négatives.

Si on l'utilise pour calculer la <u>variance d'une population</u>, la variance obtenue sera la <u>variance vraie</u>, notée σ^2.

$$\boxed{\sigma^2 = \frac{1}{n_{pop}} \sum_i (x_i - \mu)^2}$$

n_{pop} = nombre total d'individus au sein de la population

μ = moyenne de la population (= moyenne vraie)

Si on l'utilise pour calculer la <u>variance d'un échantillon</u>, la variance obtenue sera la <u>variance estimée</u>, notée s^2

$$\boxed{s^2 = \frac{1}{n_{spl}} \sum_i (x_i - \bar{x})^2}$$

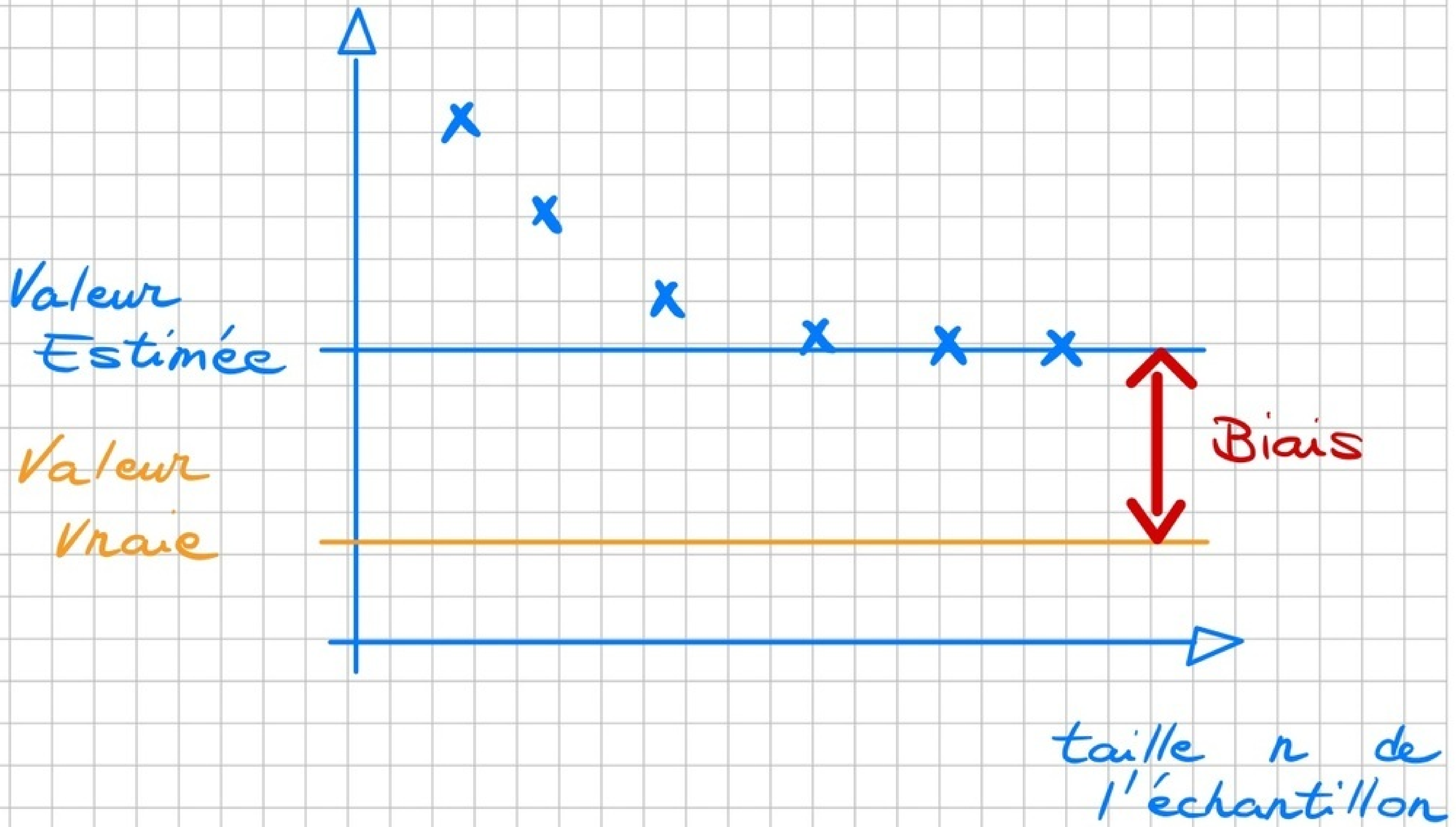
n_{spl} = nombre total d'individus au sein de l'échantillon

$\bar{x}$ = moyenne de l'échantillon (= moyenne estimée)

Corriger le biais

Un certain nombre d'estimateurs convergent vers un valeur différente de la valeur vrai. On appel cette différence <u>le biais</u> d'un estimateur.

Variance



Pour compenser ce biais, on utilise une correction différante pour chaque estimateur

Concernant l'estimateur s^2 de la variance vraie σ^2, on corrige son biais en remplacant $\frac{1}{n}$ par $\frac{1}{n-1}$ dans la formule.

Exercice

Cet exercice va nous permettre de réaliser la différence entre une valeur biaisée et une valeur non biaisée.

Voici l'énoncé :

We consider a sample $x_1, x_2, ..., x_{40}$ of weights from 40 people taken from a large population. Compute the variance of this sample

$$s^2 = \frac{1}{40-1}\sum_{i=1}^{40}(x_i - \bar{x})^2 \quad (*)$$

knowing that

$$\bar{x} = \frac{1}{40}\sum_{i=1}^{40} x_i = 77.825, \quad \sum_{i=1}^{40}(x_i)^2 = 243723.$$

(a) 37.27628 (100%)
(b) 36.34437
(c) 6.02863
(d) 6.105431
(e) None of the proposed answers

Solution:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{40-1}\left\{\sum_{i=1}^{40} x_i^2 - 40\,(\bar{x})^2\right\} = 37.27628$$

(*)

On nous demande donc ici d'utiliser la formule de l'estimateur non biaisé.

Corrigé Détaillé

On commence par la définition :

$$\boxed{s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2}$$

Puis on développe :

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

$$= \frac{1}{n-1} \sum_i \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2 \right)$$

$$= \frac{1}{n-1} \left[\sum_i x_i^2 - \sum_i 2x_i\bar{x} + \sum_i \bar{x}^2 \right]$$

Propriété d'additivité de la fonction addition

$$f(x+y) = f(x) + f(y)$$

$$\sum_i (x_i + y_i) = \sum_i x_i + \sum_i y_i$$

$$= \frac{1}{n-1} \left[\sum_i x_i^2 - 2\bar{x} \sum_i x_i + \bar{x}^2 \sum_i 1 \right]$$

On factorise le terme constant

Propriété d'homogénéité de la fonction addition

$$f(\alpha x) = \alpha f(x)$$

$$\sum_i (4x_i) = 4 \sum_i x_i$$

$$= \frac{1}{n-1}\left[\sum_i x_i^2 - 2\bar{x}\sum_i x_i + \bar{x}^2 \sum_i 1\right]$$

$$= \frac{1}{n-1}\left[\sum_i x_i^2 - 2\bar{x}\sum_i x_i + \bar{x}^2 \cdot n\right]$$

$$\sum_{i=1}^{n} 1 = \underbrace{1+1+\dots+1}_{n}$$

$$= \frac{1}{n-1}\left[\sum_i x_i^2 - 2\bar{x}\cdot n\bar{x} + n\bar{x}^2\right]$$

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

$$n\bar{x} = \sum_i x_i$$

$$= \frac{1}{n-1}\left[\sum_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_i x_i^2 - n\bar{x}^2\right]$$

On a : $n = 40$

$$\bar{x} = \frac{1}{n}\sum_i x_i = 77.825$$

$$\sum_i x_i^2 = 243723$$

$$= \frac{1}{40-1}\left[243723 - 40 \cdot 77.825^2\right]$$

$$= \frac{1}{39}\left[243723 - 242269.225\right]$$

$$= \underline{\underline{37.27628}}$$

Fausse bonne idée

Une erreur facile aurait été de partir de la mauvaise définition

Si l'on reprend le même développement, mais en partant de la définition de l'estimateur biaisé :

$$s^2 = \frac{1}{n} \sum_i \left(x_i - \bar{x} \right)^2$$

On obtient :

$$s^2 = \frac{1}{n} \left[\sum_i x_i^2 - n \bar{x}^2 \right]$$

En utilisant les même valeurs :

$$n = 40$$
$$\bar{x} = \frac{1}{n} \sum_i x_i = 77.825$$
$$\sum_i x_i^2 = 243723$$

On calcul

$$= \frac{1}{40} \cdot 243723 - \left(77.825 \right)^2$$

$$= \underline{\underline{36.344375}}$$