Rappels et précisions préalables

Calcul des paramètres statistiques et de leurs estimations

▶ paramètres de position, paramètres de dispersion, facteurs de forme

Statistique descriptive **Décrire** *ab minima* **/ Résumer** **la distribution**

Paramètres

- **Position :** moyenne, mode, médiane, min, max…
  (*ordre de grandeur*); tendance centrale
- **Dispersion :** écart-type, variance, quantiles, étendue, …
  (*répartition autour de l'ordre de grandeur*)
- **Forme :** symétrie, aplatissement,…
  (tendance)

Tableaux synthétiques rassemblant **paramètres** caractéristiques et leurs **Intervalles de Confiance**

Graphes Histogrammes, boîtes à moustache, lissage normal, …

Test normalité de la distribution (K-S, Lilliefors, Shapiro-Wilk, Chi-deux,...)

⇨ **Comparer** **échantillon (s) - population (s)**

Informations (fichier) à décrire du mieux possible pour permettre leur analyse

Rappel : somme de n valeurs

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

i : indice de l'individu, de l'observation ou de la mesure

x_i : valeur observée ou mesurée (observation/mesure) pour l'individu i

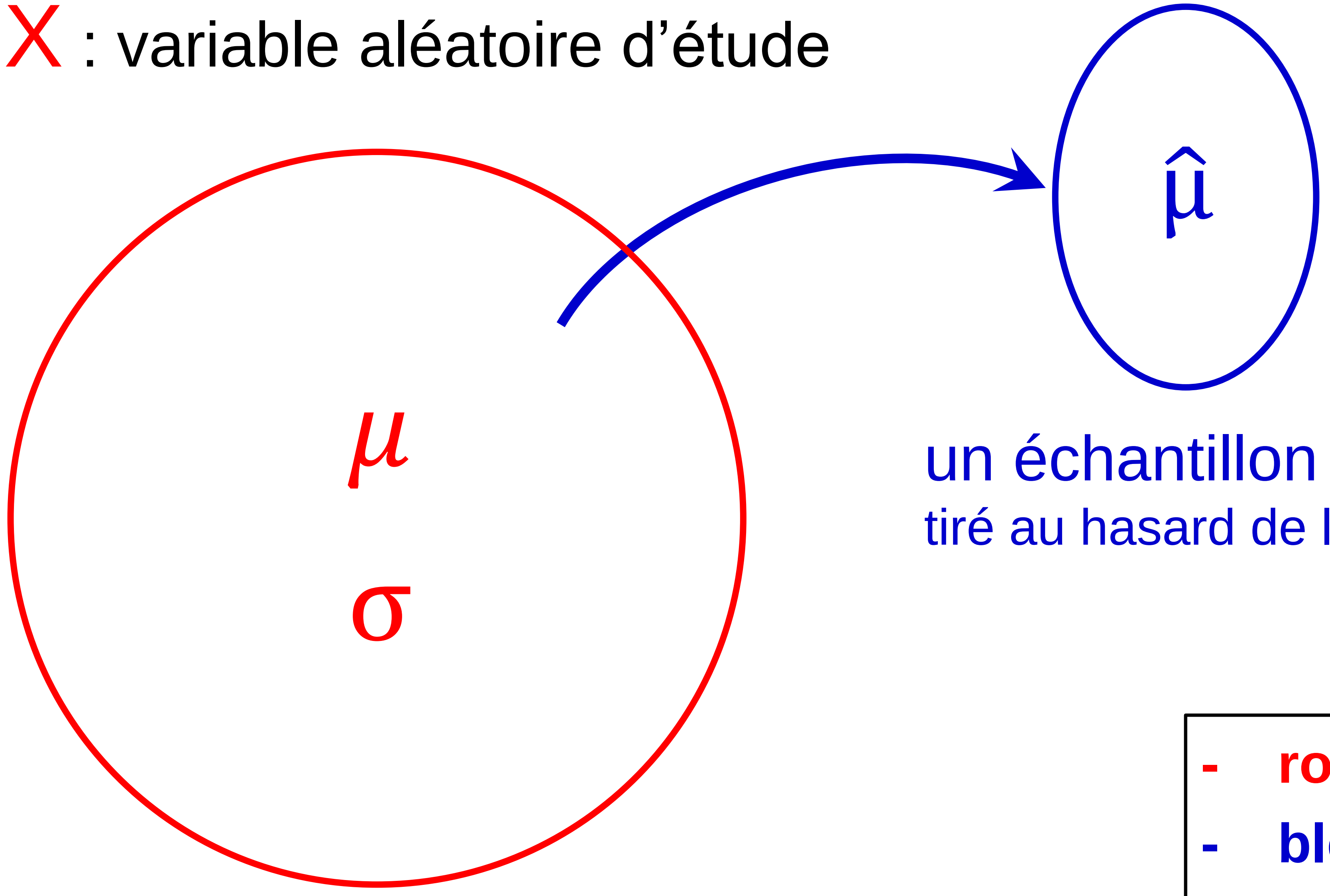
n : nombre total d'individus, d'observations ou de mesures

Une notation synthétique très pratique

Calcul des paramètres statistiques et de leurs estimations

▶ deux mots sur les codes couleurs adoptés

X : variable aléatoire d'étude



$(x_1, x_2 , \dots x_i , \dots x_n)$

un échantillon de taille n
tiré au hasard de la population

Population : caractérisée par
2 paramètres exacts : μ et σ

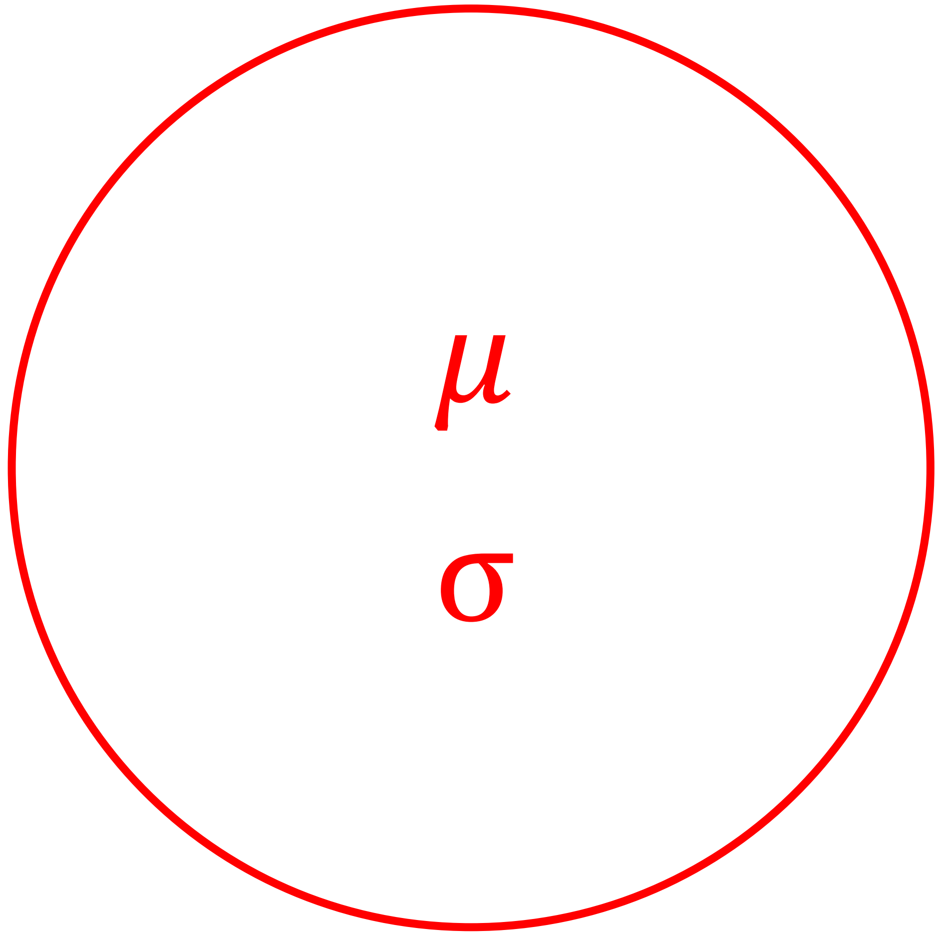
- **rouge** pour ce qui se rapporte **à la population**
- **bleu** pour ce qui se rapporte **à l'échantillon**
- lettres **minuscule** ▶ **observations**, **mesures**, **estimations**
- Lettres **Majuscules** ▶ **Variables Aléatoires**
- lettres **grecques** ▶ **paramètres exacts** d'une **population**
- **chapeaux** ▶ **estimations**

Calcul des paramètres exacts dans une population d’étude

Calcul des paramètres statistiques et de leurs estimations

▶ paramètres de position, espérance, moyenne exacte d'une population

X : variable aléatoire d'étude



Population : caractérisée par 2 paramètres exacts : **μ** et **σ**

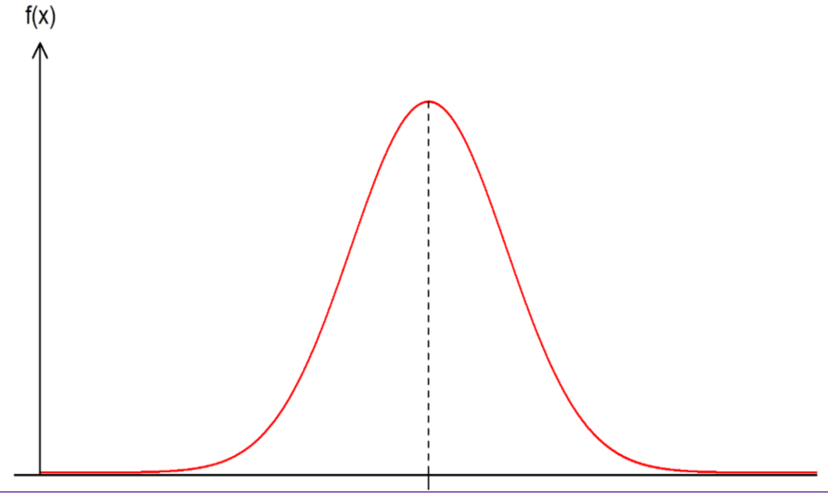
$$\mu = \mathrm{E}(X) = \frac{1}{N}\sum_{i=1}^{N} x_i \quad ; X \in \{x_1, x_2, \ldots, x_N\}$$

si X suit une loi de probabilité discrète :

$$\mu = \mathrm{E}(X) = \sum_{i=1}^{N} x_i \,.\, P(X = x_i)$$

si X suit une loi de probabilité continue; $f(x)$ densité de probabilité :

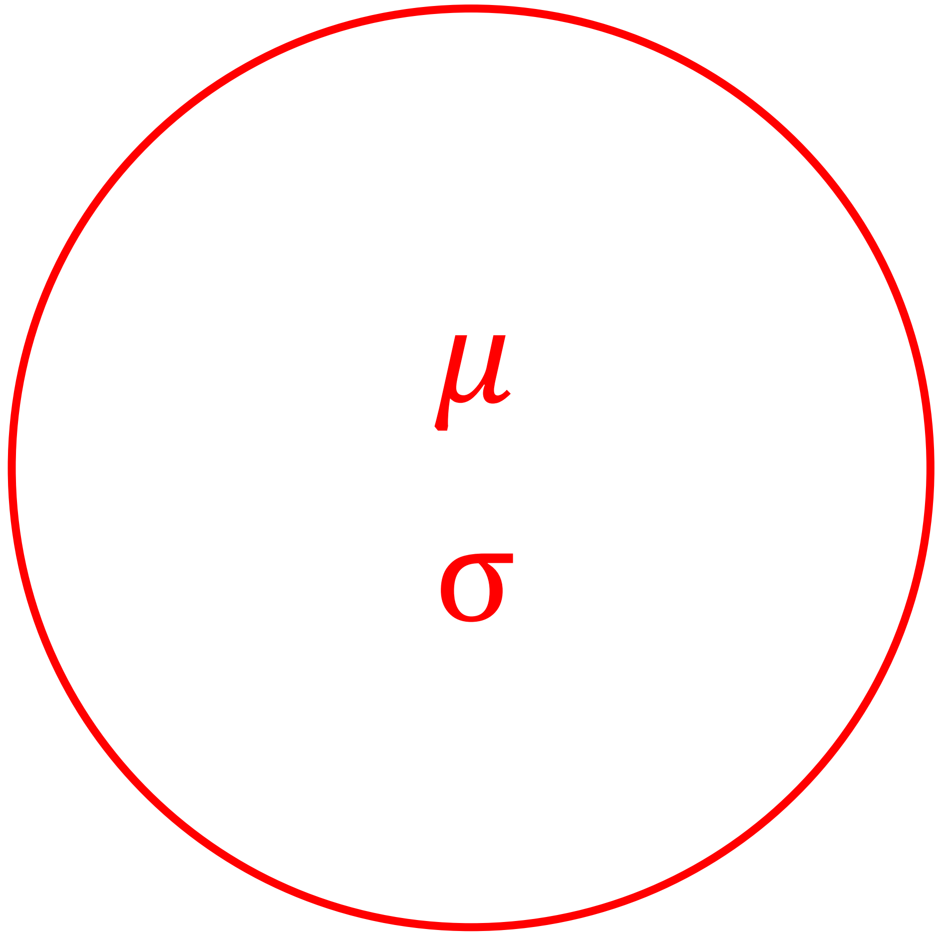
$$\mu = \mathrm{E}(X) = \int_{-\infty}^{+\infty} x.f(x)dx$$



Une définition et un mode de calcul dépendant du contexte

X : variable aléatoire d’étude

$$X \in \{x_1, x_2, \ldots, x_N\}$$



$$\sigma^2 = \mathrm{VAR}(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\sigma^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2$$ *(théorème de König-Huygens)*

Population : caractérisée par

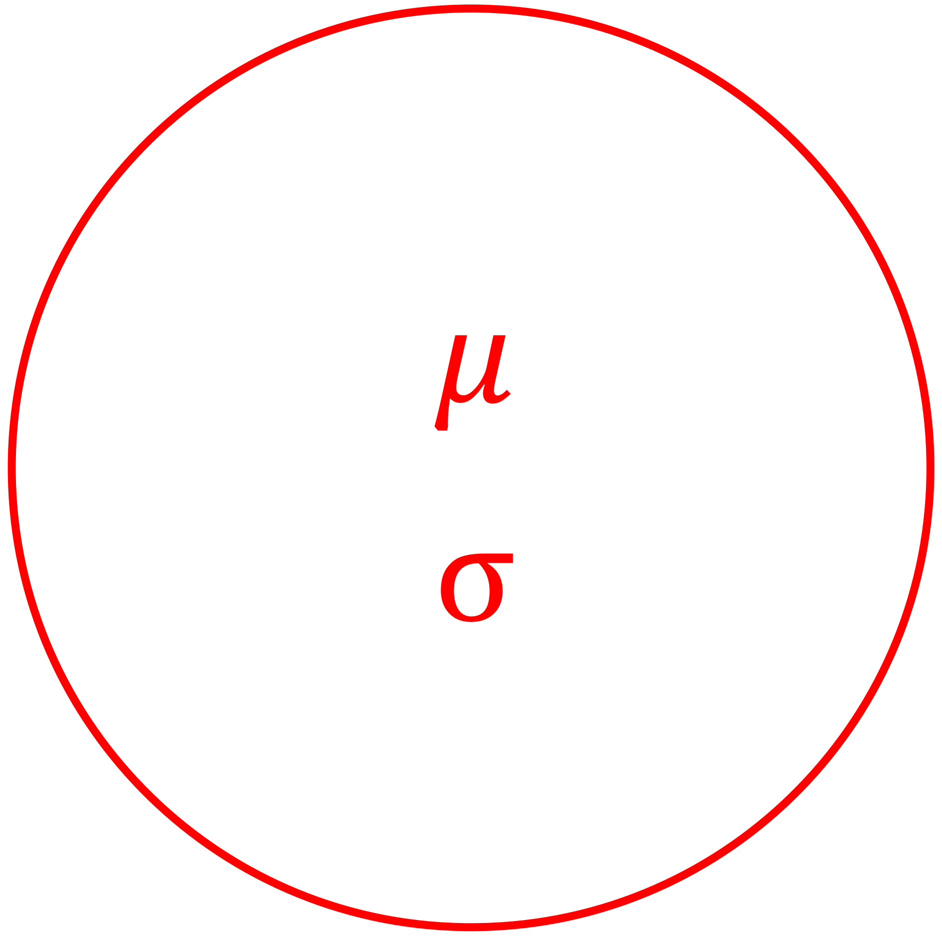
2 paramètres exacts : **μ** et **σ**

$$\sigma^2 = \mathrm{VAR}(X) = \frac{1}{N}\sum_{i=1}^{N}{x_i}^2 - \mu^2$$

Calcul des paramètres statistiques et de leurs estimations

▶ paramètres de dispersion, variance exacte d'une population

X : variable aléatoire d'étude



Population : caractérisée par 2 paramètres exacts : μ et **σ**

$$\mathrm{X} \in \{x_1, x_2, \ldots, x_N\}$$

$$\sigma^2 = \mathrm{VAR}(\mathrm{X}) = \frac{1}{N}\sum_{i=1}^{N} {x_i}^2 - \mu^2$$ *(Könnig- Huyguens)*

si X suit une loi de probabilité discrète :

$$\sigma^2 = \mathrm{VAR}(\mathrm{X}) = \sum_{i=1}^{N} {x_i}^2 . P(X = x_i) - \mu^2$$

si X suit une loi de probabilité continue; *f(x)* densité de probabilité :

$$\sigma^2 = \mathrm{VAR}(\mathrm{X}) = \int_{-\infty}^{+\infty} {x_i}^2 . f(x)dx - \mu^2$$

Une définition et un mode de calcul dépendant du contexte

Calcul des estimations des paramètres exacts dans un échantillon de taille n

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la moyenne exacte : moyenne calculée dans un échantillon

X : variable aléatoire d'étude

$\hat{\mu}$

$\hat{\sigma}$

μ

σ

un échantillon de taille n
tiré (avec remise) au hasard de la population
(échantillon représentatif de la population)

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

estimation ponctuelle de μ
(non biaisée)

Population : caractérisée par
2 paramètres exacts : μ et σ

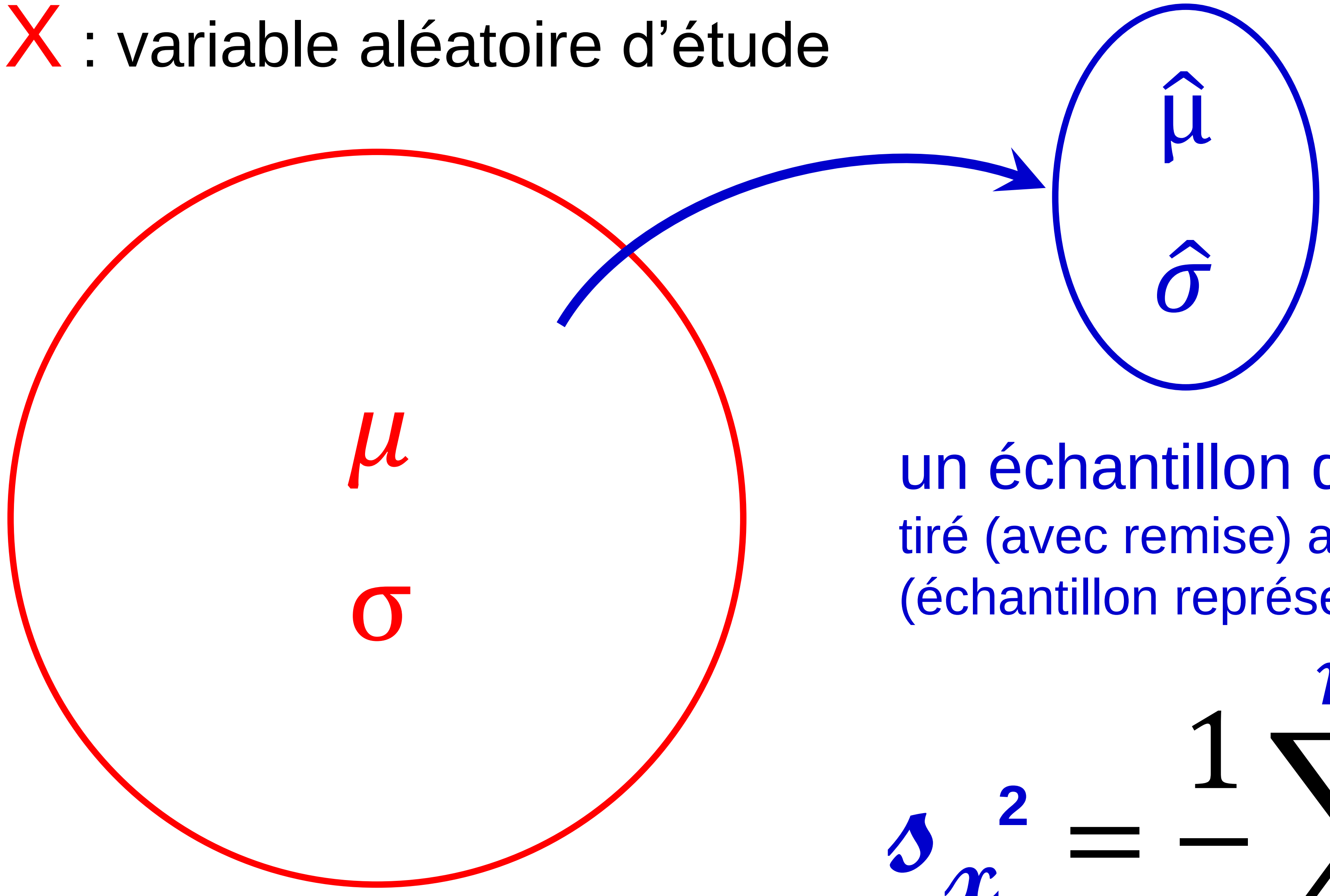
Si l'échantillon est tiré **au hasard** et de **taille suffisante**,
$\hat{\mu}$ est proche de la valeur de la moyenne exacte μ dans la population

La moyenne calculée dans un échantillon n'est qu'une estimation de la moyenne exacte dans la population

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la variance exacte : variance calculée dans un échantillon

X : variable aléatoire d'étude



$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

estimation ponctuelle de μ

(non biaisée)

un échantillon de taille n

tiré (avec remise) au hasard de la population

(échantillon représentatif de la population)

$$s_x^{\,2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

estimation ponctuelle de σ

(biaisée)

Population : caractérisée par

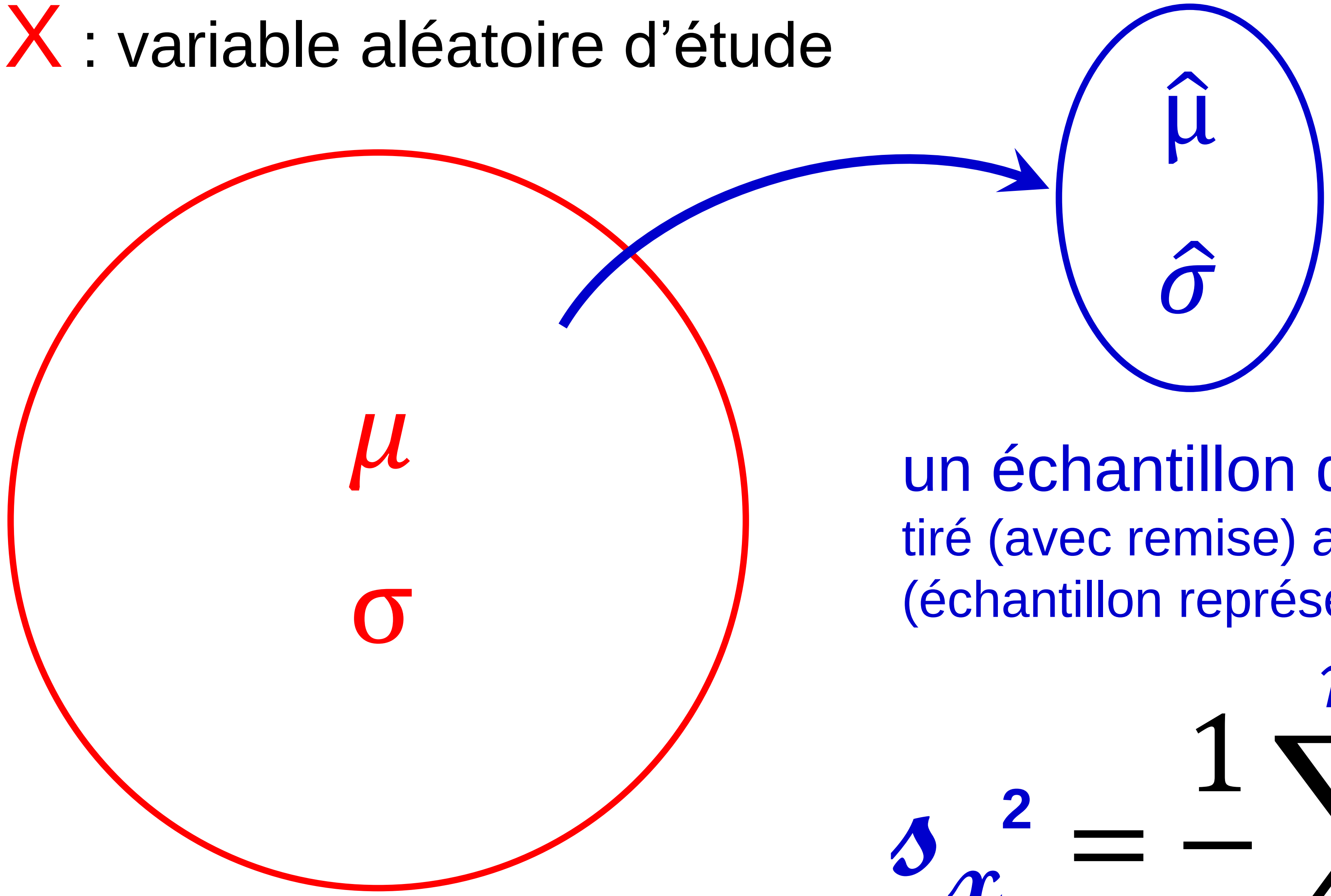
2 paramètres exacts : μ et σ

La variance calculée dans un échantillon est une estimation biaisée de la variance exacte de la population

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la variance exacte : variance calculée dans un échantillon

X : variable aléatoire d'étude



$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

estimation ponctuelle de μ

(non biaisée)

un échantillon de taille n

tiré (avec remise) au hasard de la population

(échantillon représentatif de la population)

$$s_x^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\ 2} - \hat{\mu}^2$$

formule de Könnig- Huyguens

Population : caractérisée par

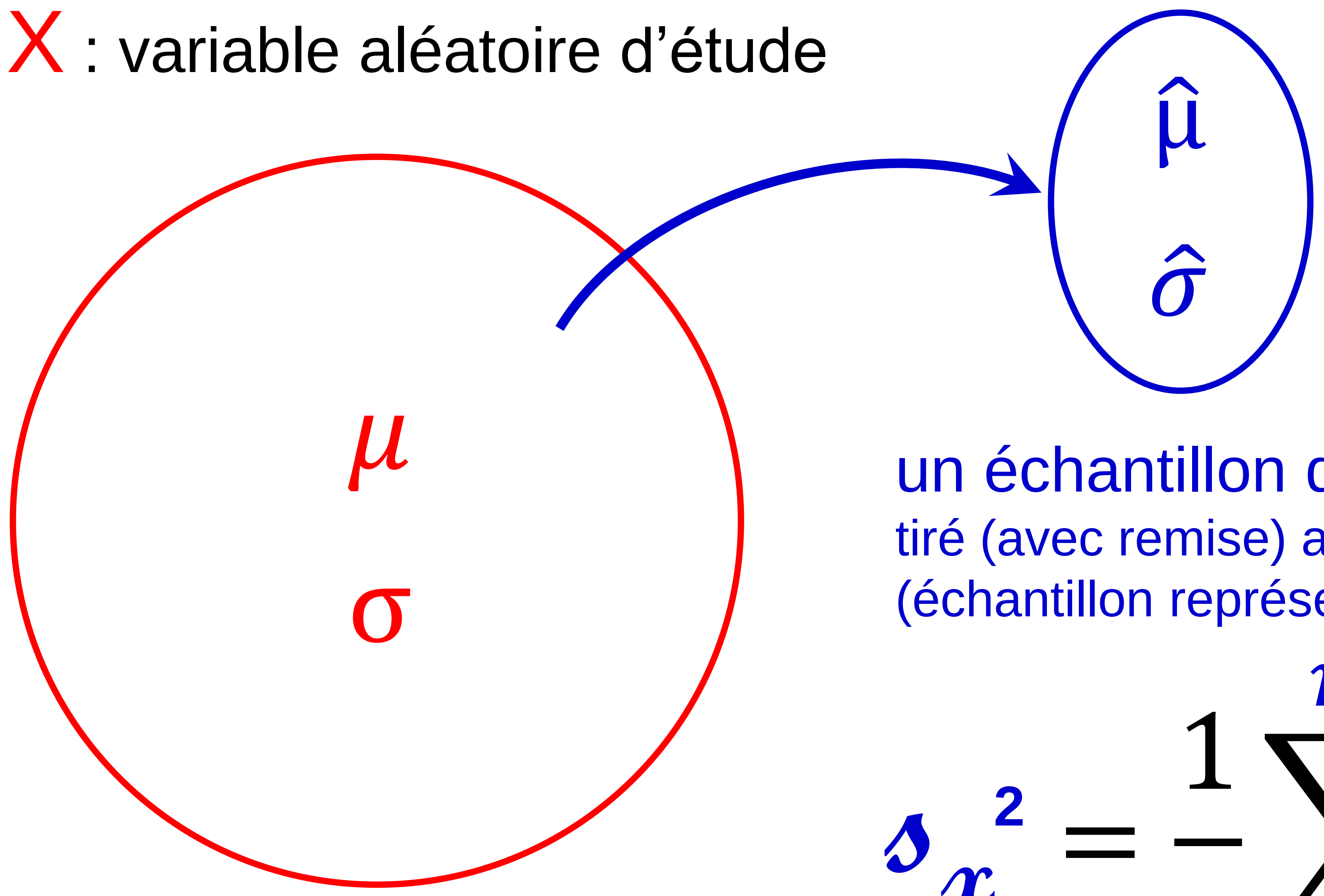
2 paramètres exacts : μ et σ

La variance calculée dans un échantillon est une estimation biaisée de la variance exacte de la population

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la variance exacte : estimation non biaisée de la variance

X : variable aléatoire d'étude



$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

estimation ponctuelle de μ (non biaisée)

un échantillon de taille n

tiré (avec remise) au hasard de la population
(échantillon représentatif de la population)

$$s_x^{2} = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \hat{\mu}^2$$

formule de Könnig- Huygens

Population : caractérisée par

2 paramètres exacts : μ et σ

$$\boxed{{s^*_x}^{2}} = \frac{n}{n-1}\, s_x^{2}$$

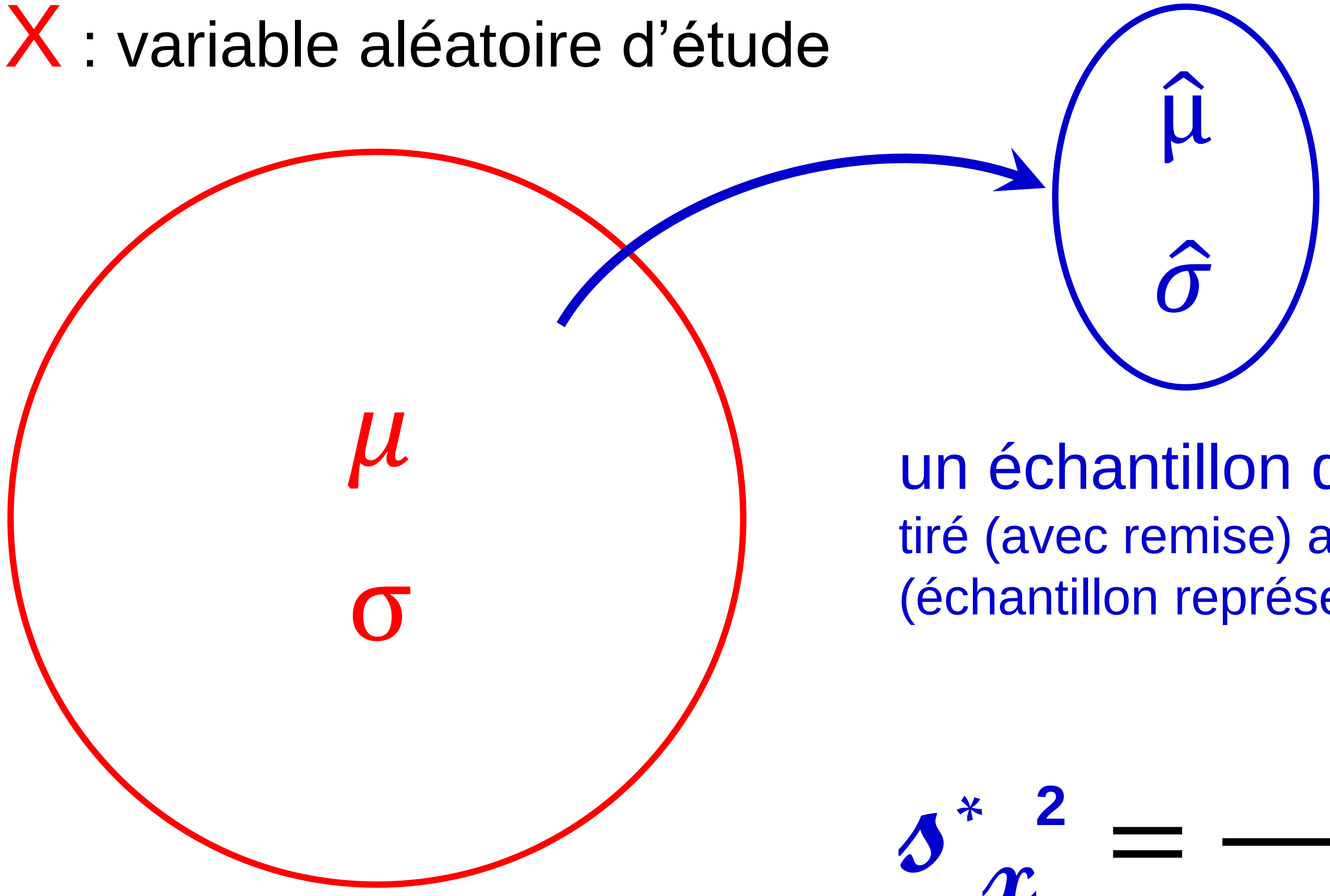
estimation ponctuelle de σ corrigée du biais

La variance calculée dans un échantillon est une estimation biaisée de la variance exacte de la population

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la variance exacte : estimation non biaisée de la variance

X : variable aléatoire d’étude



$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

estimation ponctuelle de μ (non biaisée)

un échantillon de taille n

tiré (avec remise) au hasard de la population

(échantillon représentatif de la population)

Population : caractérisée par

2 paramètres exacts : μ et σ

$$s^*{}_x{}^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i{}^2 - \frac{n}{n-1}\hat{\mu}^2$$

estimation ponctuelle de σ (corrigée du biais)

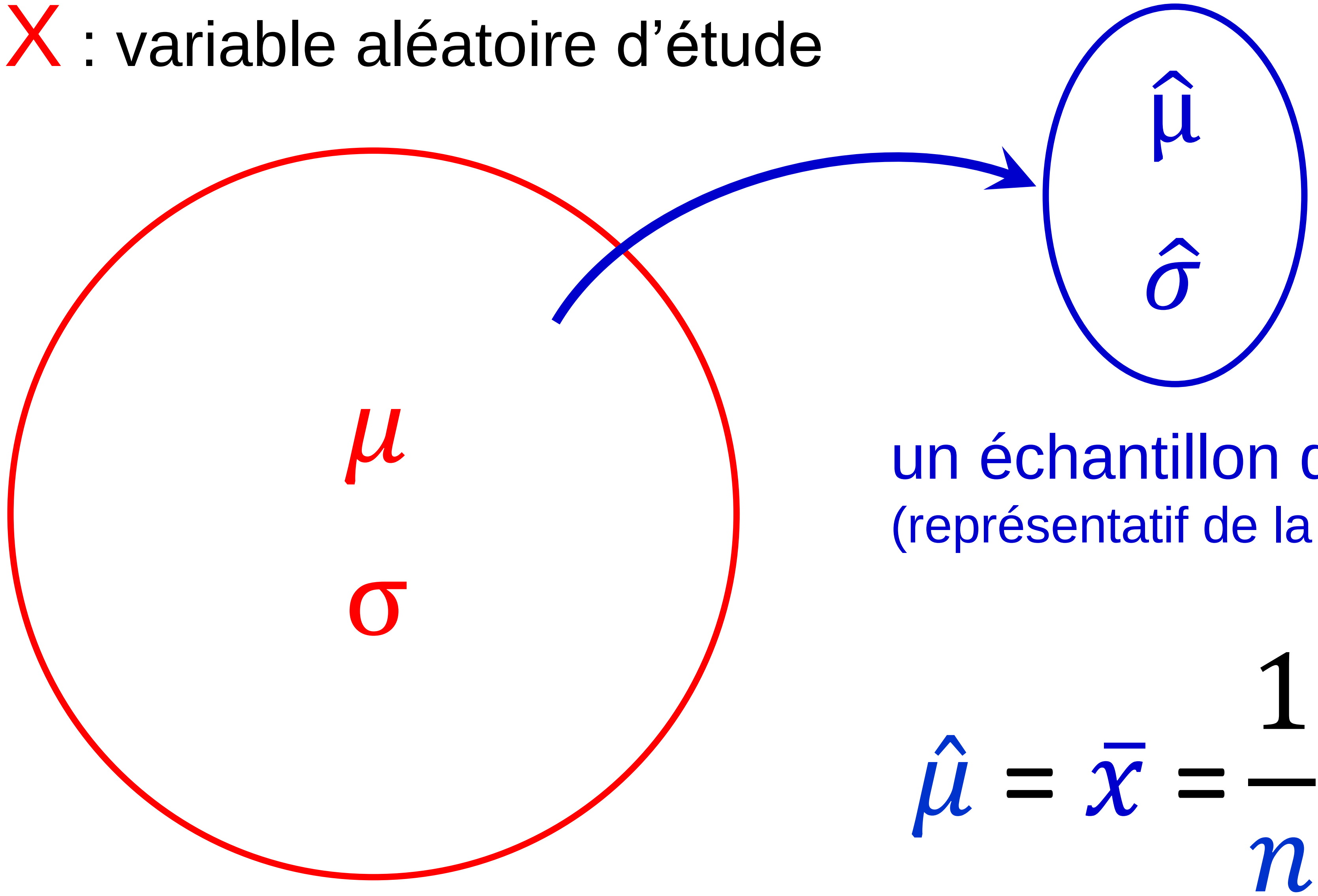
La variance calculée dans un échantillon est une estimation biaisée de la variance exacte de la population

Calcul des estimations des paramètres exacts lorsque les données sont ordonnées-classées (telles que représentées sur un histogramme)

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la moyenne exacte : moyenne calculée à partir de données ordonnées classées

X : variable aléatoire d'étude



un échantillon de taille n
(représentatif de la population)

Population : caractérisée par 2 paramètres exacts : μ et σ

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{k} n_j x_{jc}$$

classe (qt/ha)	ni	x_{ic} (qt/ha)	$n_i x_{ic}$ (qt/ha)
]80,85]	0	82,5	0,0
]85,90]	1	87,5	87,5
]90,95]	3	92,5	277,5
]95,100]	5	97,5	487,5
]100,105]	8	102,5	820,0
]105,110]	11	107,5	1182,5
]110,115]	13	112,5	1462,5
]115,120]	13	117,5	1527,5
]120,125]	8	122,5	980,0
]125,130]	2	127,5	255,0
]130,135]	1	132,5	132,5
Sommes	65		7212,5

$\hat{\mu}$ = 7212,5/65 = 111,0 qt/ha

estimation ponctuelle de μ
(moins précise)

On utilise les **effectifs de classes** n_j et les valeurs centrées x_{jc} de chacune des **k classes** pour calculer une **moyenne approchée**; j est **l'indice de classe**.

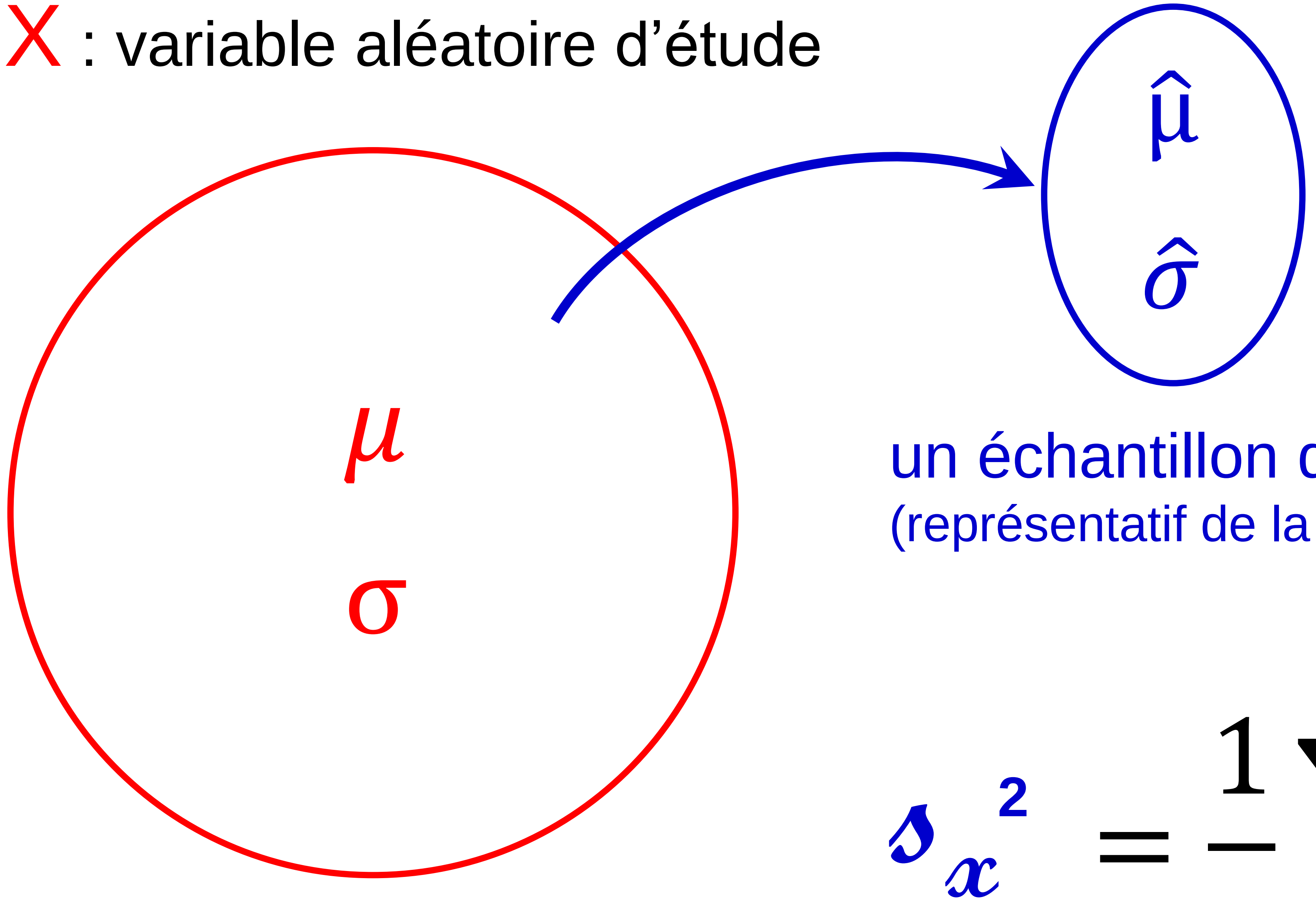
On **perd de la précision** sur la valeur ciblée. L'estimation est moins bonne.

La moyenne calculée dans un échantillon n'est qu'une estimation de la moyenne exacte dans la population

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la moyenne exacte : variance calculée à partir de données ordonnées classées

X : variable aléatoire d'étude



un échantillon de taille n
(représentatif de la population)

classe (qt/ha)	ni	x_{ic} (qt/ha)	$n_i x_{ic}$ (qt/ha)
]80,85]	0	82,5	0,0
]85,90]	1	87,5	87,5
]90,95]	3	92,5	277,5
]95,100]	5	97,5	487,5
]100,105]	8	102,5	820,0
]105,110]	11	107,5	1182,5
]110,115]	13	112,5	1462,5
]115,120]	13	117,5	1527,5
]120,125]	8	122,5	980,0
]125,130]	2	127,5	255,0
]130,135]	1	132,5	132,5
Sommes	65		7212,5

$\hat{\mu}$ = 7212,5/65 = 111,0 qt/ha

$$s_x^2 = \frac{1}{n}\sum_{j=1}^{k} n_j (x_{jc} - \hat{\mu})^2$$

Population : caractérisée par

2 paramètres exacts : μ et σ

estimation ponctuelle (moins précise) de σ (biaisée)

On utilise les **effectifs de classes** n_j et les valeurs centrées x_{jc} de chacune des k classes ; j est l'indice de classe.

La variance calculée dans un échantillon n'est qu'une estimation de la variance exacte dans la population

Calcul des paramètres statistiques et de leurs estimations

▶ estimation de la moyenne exacte : variance calculée à partir de données ordonnées classées

X : variable aléatoire d'étude

μ

σ

$\hat{\mu}$

$\hat{\sigma}$

un échantillon de taille n
(représentatif de la population)

classe (qt/ha)	ni	x_{ic} (qt/ha)	$n_i x_{ic}$ (qt/ha)
]80,85]	0	82,5	0,0
]85,90]	1	87,5	87,5
]90,95]	3	92,5	277,5
]95,100]	5	97,5	487,5
]100,105]	8	102,5	820,0
]105,110]	11	107,5	1182,5
]110,115]	13	112,5	1462,5
]115,120]	13	117,5	1527,5
]120,125]	8	122,5	980,0
]125,130]	2	127,5	255,0
]130,135]	1	132,5	132,5
Sommes	65		7212,5

$\hat{\mu}$ = 7212,5/65 = 111,0 qt/ha

$$s^{*}{}_{x}^{2} = \frac{1}{n-1}\sum_{j=1}^{k} n_j {x_{jc}}^2 - \frac{n}{n-1}\hat{\mu}^2$$

Population : caractérisée par

2 paramètres exacts : μ et σ

estimation ponctuelle (moins précise) de σ (corrigée du biais)

On utilise les **effectifs de classes** n_j et les valeurs centrées x_{jc} de chacune des k classes ; j est l'indice de classe.

La variance calculée dans un échantillon n'est qu'une estimation de la variance exacte dans la population