

# La « bonne » mesure de l'écart-type

Mesures et Incertitudes

Quelle formule pour le calcul de l'écart-type d'une série de données  $(n_i, x_i)$  ?

$$\sigma_{n-1} = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{n-1}} \quad \text{ou} \quad \sigma_n = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{n}} \quad ?$$

SPC

Maths

# 1. Écart-type $s$ et écart-type $\sigma$

Écart-type d'une série  $(n_i, x_i)$  de valeurs observées :

Par définition, l'écart-type est la moyenne quadratique des écarts à la moyenne  $\bar{x}$ . On le note habituellement  $s$  (de l'anglais *standard deviation*) :

$$\sqrt{\sum f_i (x_i - \bar{x})^2} = \sqrt{\sum \frac{n_i (x_i - \bar{x})^2}{n}}$$

Écart-type d'une série de valeurs possibles d'une variable aléatoire :

C'est la moyenne quadratique des écarts à l'espérance  $\mu$ . On le note  $\sigma$  :

$$\sqrt{\sum p_i (x_i - \mu)^2}$$

# Variance

- Série observée :  $s^2$
- Série des VA possibles :  $\sigma^2$

# Exemples :

- **Série de données observées**

Tailles de 7 personnes : 152 158 164 168 168 169 176

- Moyenne  $\bar{x} = \frac{1}{7} [152 + 158 + 164 + 2 \times 168 + 169 + 176] = 165,0$

- Écart-type  $s = \sqrt{\frac{1}{7} [(152 - 165)^2 + (158 - 165)^2 + (164 - 165)^2 + 2 \times (168 - 165)^2 + (169 - 165)^2 + (176 - 165)^2]}$   
 $= 7,3$

- **Série de valeurs possibles d'une VA**

Lancer de dé équilibré

- Espérance  $\mu = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \dots + \frac{1}{6} \times 6 = 3,5$

- Ecart-type  $\sigma = \sqrt{\frac{1}{6} \times (1 - 3,5)^2 + \frac{1}{6} \times (2 - 3,5)^2 + \dots + \frac{1}{6} \times (6 - 3,5)^2} = 1,71$

# Ecart-type d'une population

- Etude d'un caractère d'une population dans son entièreté

$$\circ p_i = f_i$$

$$\circ \mu = \bar{x}$$

$$\circ \sigma = s$$

$$\circ \sigma^2 = s^2$$

## 2. Estimation de $\sigma$ par l'écart-type $s$ d'un échantillon : le problème du biais d'estimation

- étude d'un **échantillon** constitué par des réalisations de VA
- $s$  est une estimation de  $\sigma$  : inférence statistique (extrapolation)
- répétition de l'échantillonnage : il existe un **écart de la moyenne des  $s$  avec  $\sigma$ .**



# Simulation lancer dé

$$\sigma \approx 1,71$$

$$\sigma^2 \approx 2,92$$

```
9 from numpy import *
10 from random import *
11
12 def Var(tailleEch):
13     L=[]
14     for i in range (tailleEch):
15         L.append(randint(1,6))
16         s=std(L)
17     return s**2
18
19 def moyVar(tailleEch,nbEch):
20     liste=[]
21     for k in range (nbEch):
22         liste.append(Var(tailleEch))
23     return average(liste)
```

$$\frac{n-1}{n} \sigma^2 \approx 2,34$$

```
In [38]: moyVar(5,100)
```

```
Out[38]: 2.4056
```

```
In [39]: moyVar(5,100)
```

```
Out[39]: 2.355200000000000004
```

```
In [40]: moyVar(5,100)
```

```
Out[40]: 2.21759999999999996
```

```
In [41]: moyVar(5,100)
```

```
Out[41]: 2.2312
```

```
In [42]: moyVar(5,100)
```

```
Out[42]: 2.3704
```

```
In [43]: moyVar(5,100)
```

```
Out[43]: 2.4232
```

```
In [44]: moyVar(5,100)
```

```
Out[44]: 2.30159999999999996
```



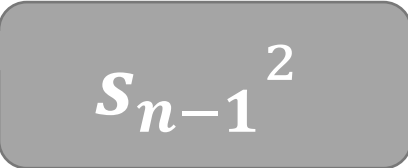
Correction de cet écart avec la moyenne des variances  $s^2$  des échantillons :

$$\frac{n-1}{n} \sigma^2$$

Donc on estime  $\sigma^2$  à partir de  $s^2$  par  $\frac{n}{n-1} s^2$

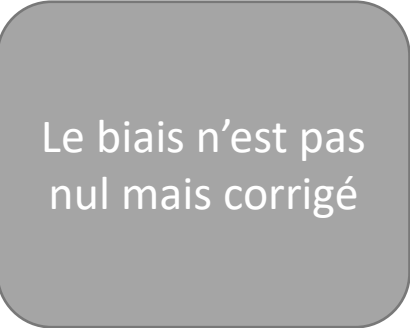
On a :

$$\sigma^2 \approx \frac{n}{n-1} \times \frac{\sum n_i (x_i - \bar{x})^2}{n}$$

$$\sigma^2 \approx \frac{\sum n_i (x_i - \bar{x})^2}{n-1}$$


$S_{n-1}^2$

$$\sigma^2 \approx S_{n-1}^2$$



Le biais n'est pas nul mais corrigé

# Intervalle de confiance

- Mesure le degré de précision que l'on a sur les estimations issues de l'échantillon.
- Deux sources principales de variations sur les données peuvent être la cause d'un manque de précision dans l'estimation d'une grandeur :
  - Un nombre insuffisant de données
  - Du bruit dans la mesure des données ce qui est pratiquement toujours le cas pour la mesure des grandeurs physiques.
- Estimation de l'espérance d'une loi à partir d'une moyenne empirique et d'une majoration de l'écart type

---

$$I_c = \left[ \bar{x} - t_\alpha \frac{s}{\sqrt{n}} ; \bar{x} + t_\alpha \frac{s}{\sqrt{n}} \right]$$

# Estimation d'une moyenne (loi normale)

- Moyenne  $\bar{X}$  d'une famille  $(X_1, X_2, \dots, X_n)$  ?
- On connaît la moyenne d'un échantillon  $\bar{x}$
- $I_{c95\%} = [\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}]$

Si  $\sigma$  inconnu, estimation par  $\sqrt{\frac{n}{n-1}} \times s = s_{n-1}$

$$I_c = [\bar{x} - 1,96 \frac{s_{n-1}}{\sqrt{n}}; \bar{x} + 1,96 \frac{s_{n-1}}{\sqrt{n}}]$$

ou encore  $I_c = [\bar{x} - 1,96 \frac{s}{\sqrt{n-1}}; \bar{x} + 1,96 \frac{s}{\sqrt{n-1}}]$

# Bibliographie

- *Quelle est la « bonne » formule de l'écart-type ?*, E. Grenier, Modulad, 2007