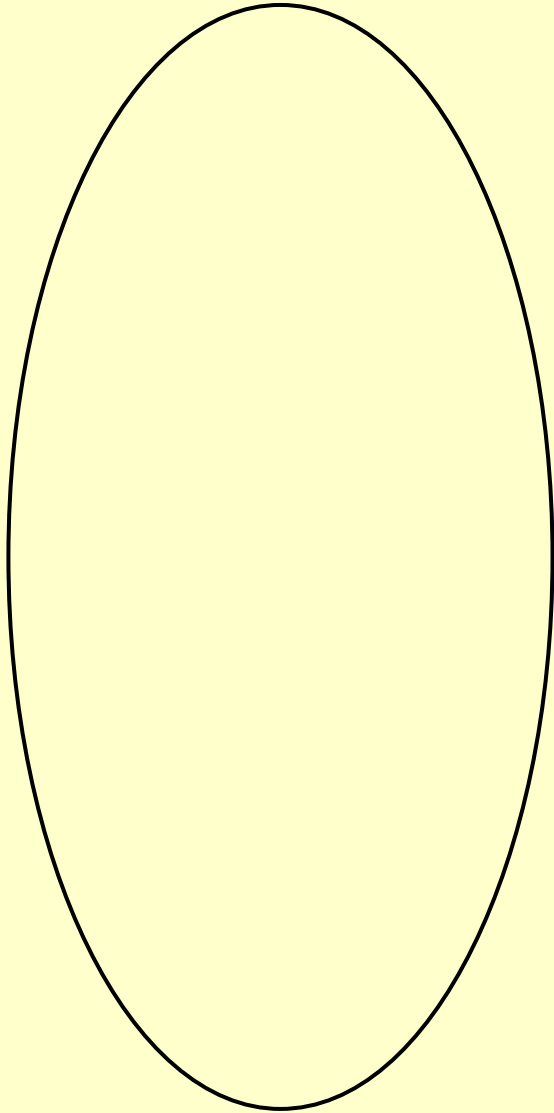


FLUCTUATION D'ÉCHANTILLONNAGE

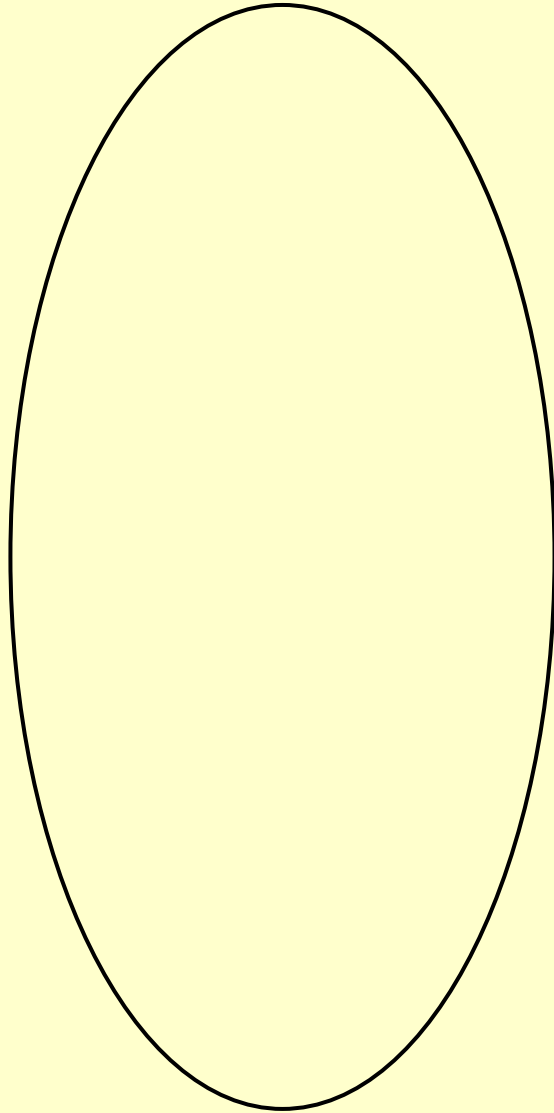
Équipe académique
Mathématiques
Bordeaux



Comment sont déterminés les
résultats d'un sondage ?



Population

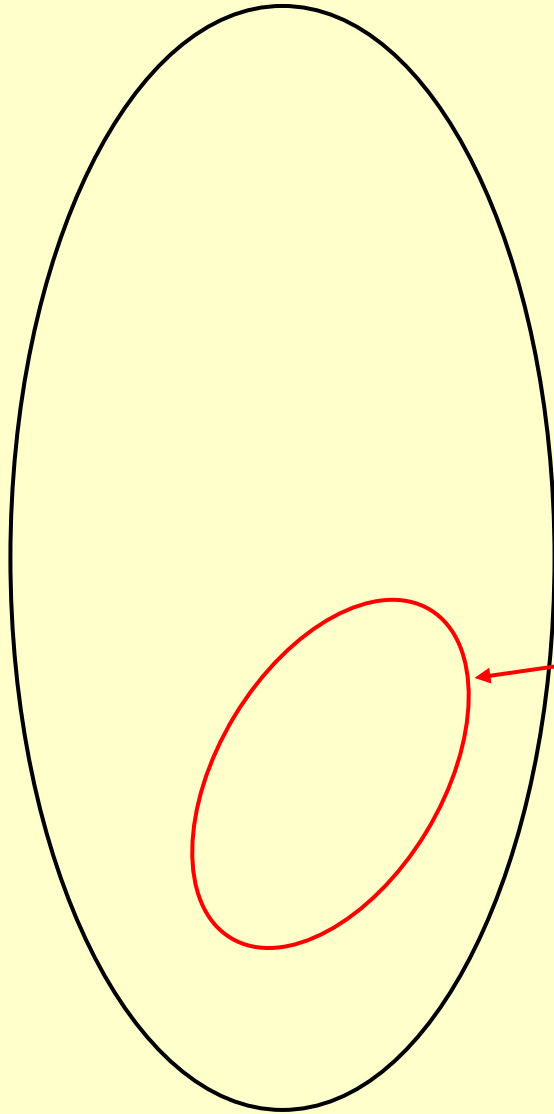


Population

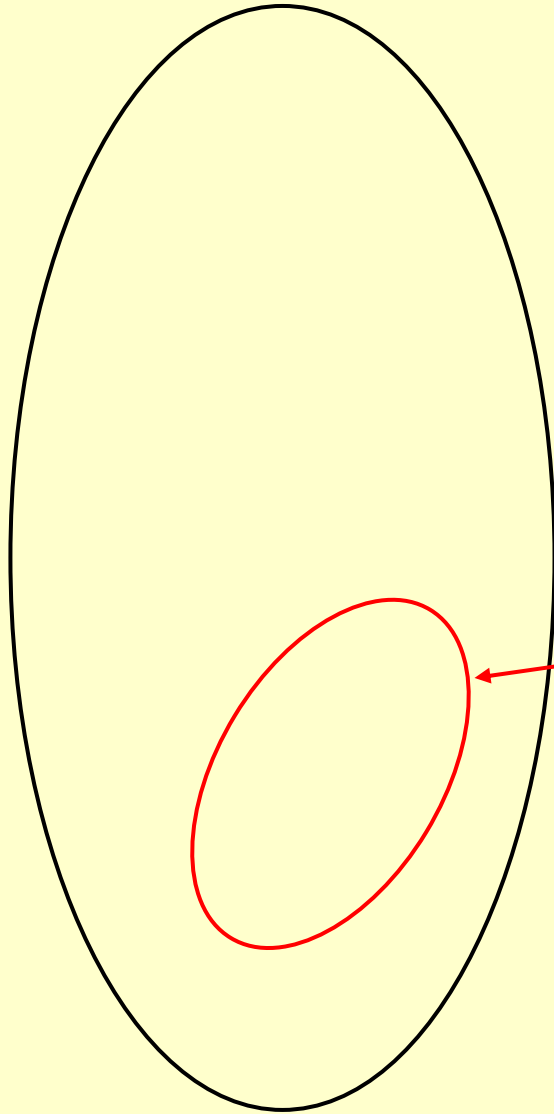
dans laquelle on cherche
un pourcentage

Population

dans laquelle on cherche
un pourcentage



Échantillon de taille n

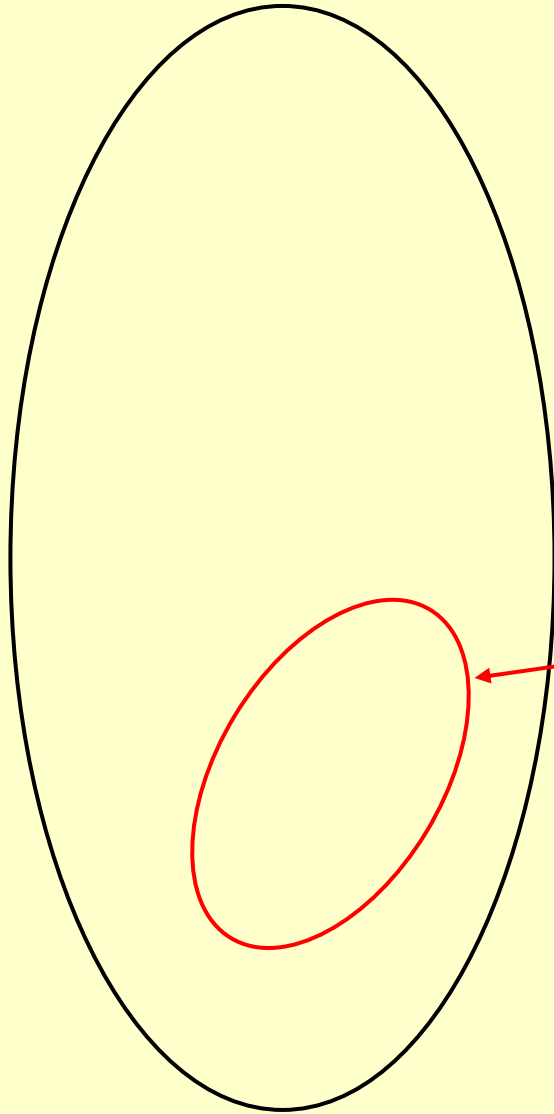


Population

dans laquelle on cherche
un pourcentage

Échantillon de taille n

On calcule le
pourcentage dans
cet échantillon,
par exemple 52 %



Population

~~dans laquelle on cherche
un pourcentage~~

**On attribue à la population
totale le pourcentage trouvé
dans l'échantillon : 52 %**

Échantillon de taille n

On calcule le
pourcentage dans
cet échantillon,
par exemple 52 %

On a fait une
ESTIMATION PONCTUELLE
d'un pourcentage
à partir du pourcentage obtenu
dans un échantillon.

Dans quelle mesure peut-on faire confiance au résultat obtenu dans un échantillon ?

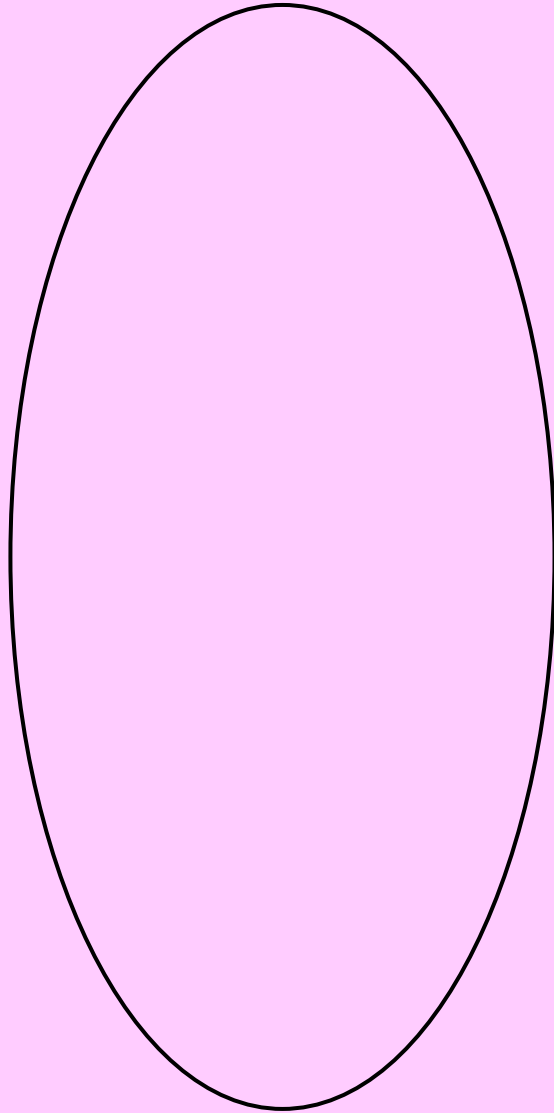
Le résultat aurait-il été le même si on avait pris, au hasard, un autre échantillon ?

On va passer de la théorie de
L'ESTIMATION

à celle de
L'ÉCHANTILLONNAGE

C'est-à-dire qu'on va se poser
la question suivante :

Comment sont constitués des
échantillons extraits d'une
population dont on connaît la
composition ?

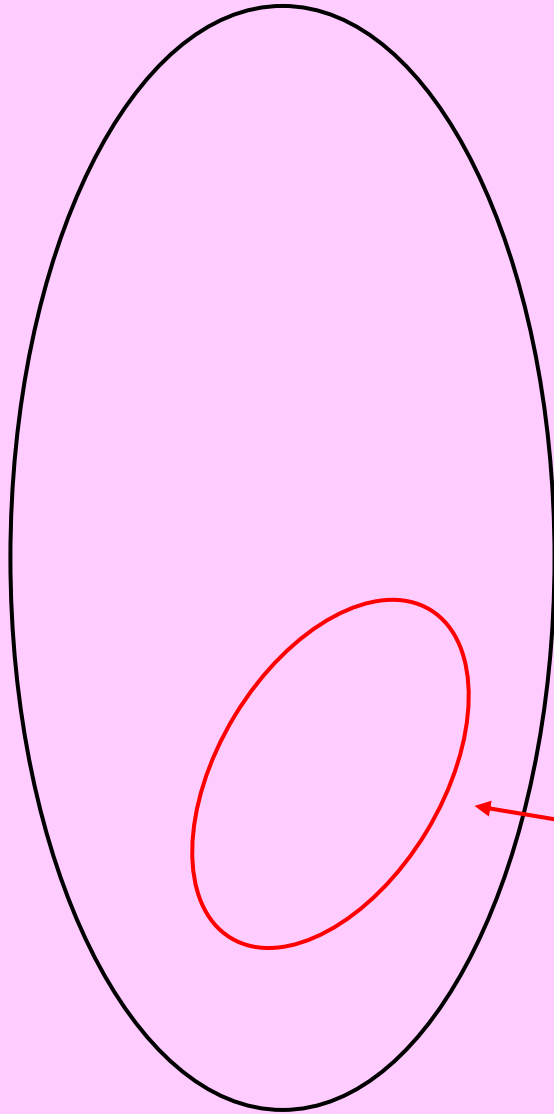


Population

pourcentage connu : 52 %

Population

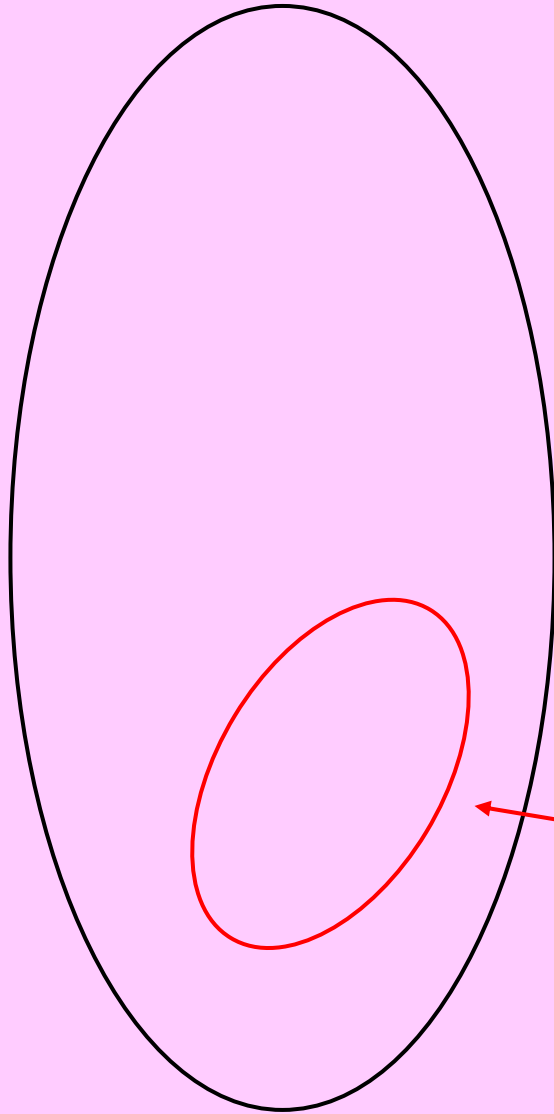
pourcentage connu : 52 %



Échantillon de taille n

Population

pourcentage connu : 52 %



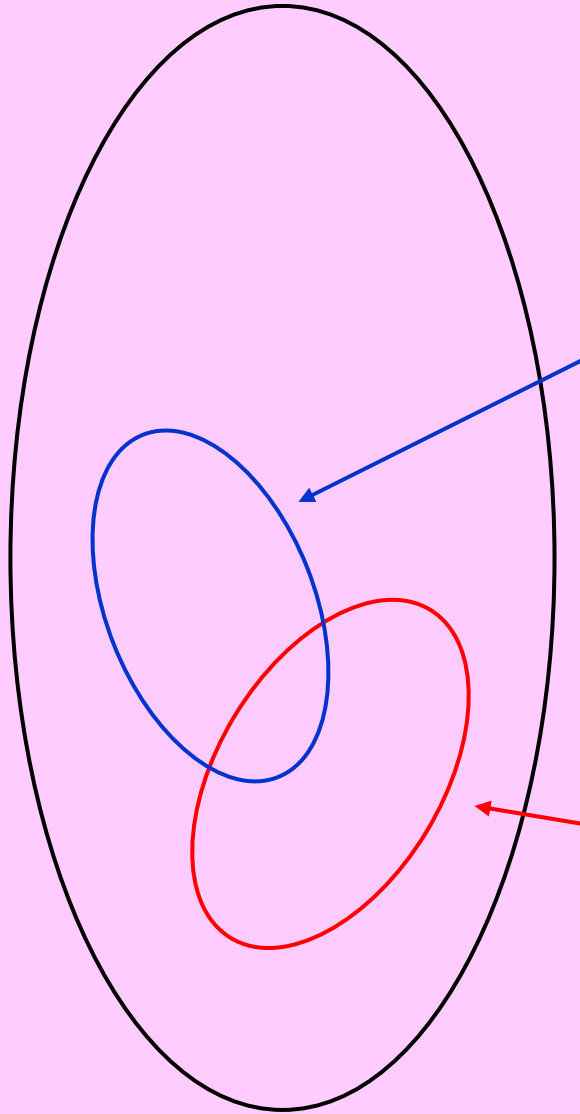
Échantillon de taille n

Quel est le pourcentage
dans cet échantillon ?

Population

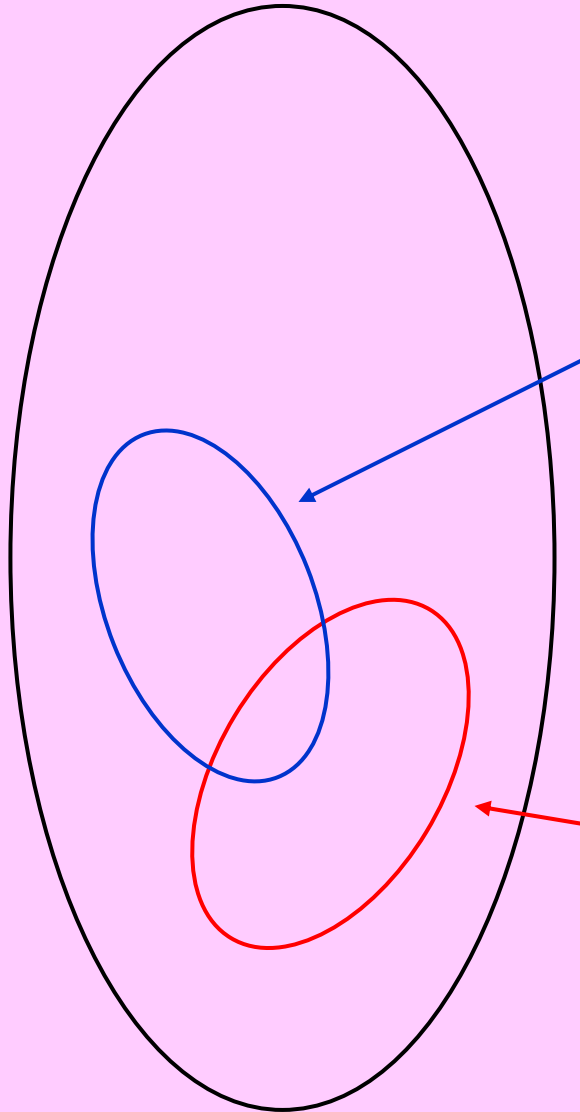
pourcentage connu : 52 %

Autre échantillon de taille n



Échantillon de taille n

Quel est le pourcentage
dans cet échantillon ?



Population

pourcentage connu : 52 %

Autre échantillon de taille n

Le pourcentage est-il le même dans cet échantillon que dans le précédent ?

Échantillon de taille n

Quel est le pourcentage dans cet échantillon ?

Grâce à un tableur,
on va simuler les résultats de calculs
de pourcentages dans des échantillons
extraits d'une population
dont on connaît la composition,
par exemple :
52 % de boules rouges et
48 % de boules noires.

Fonctions utilisées dans le tableur :

ALEA.ENTRE.BORNES(1;100)

qui tire un nombre au hasard entre 1 et 100

test logique : $a \leq b$

donne 1 si vrai, 0 si faux

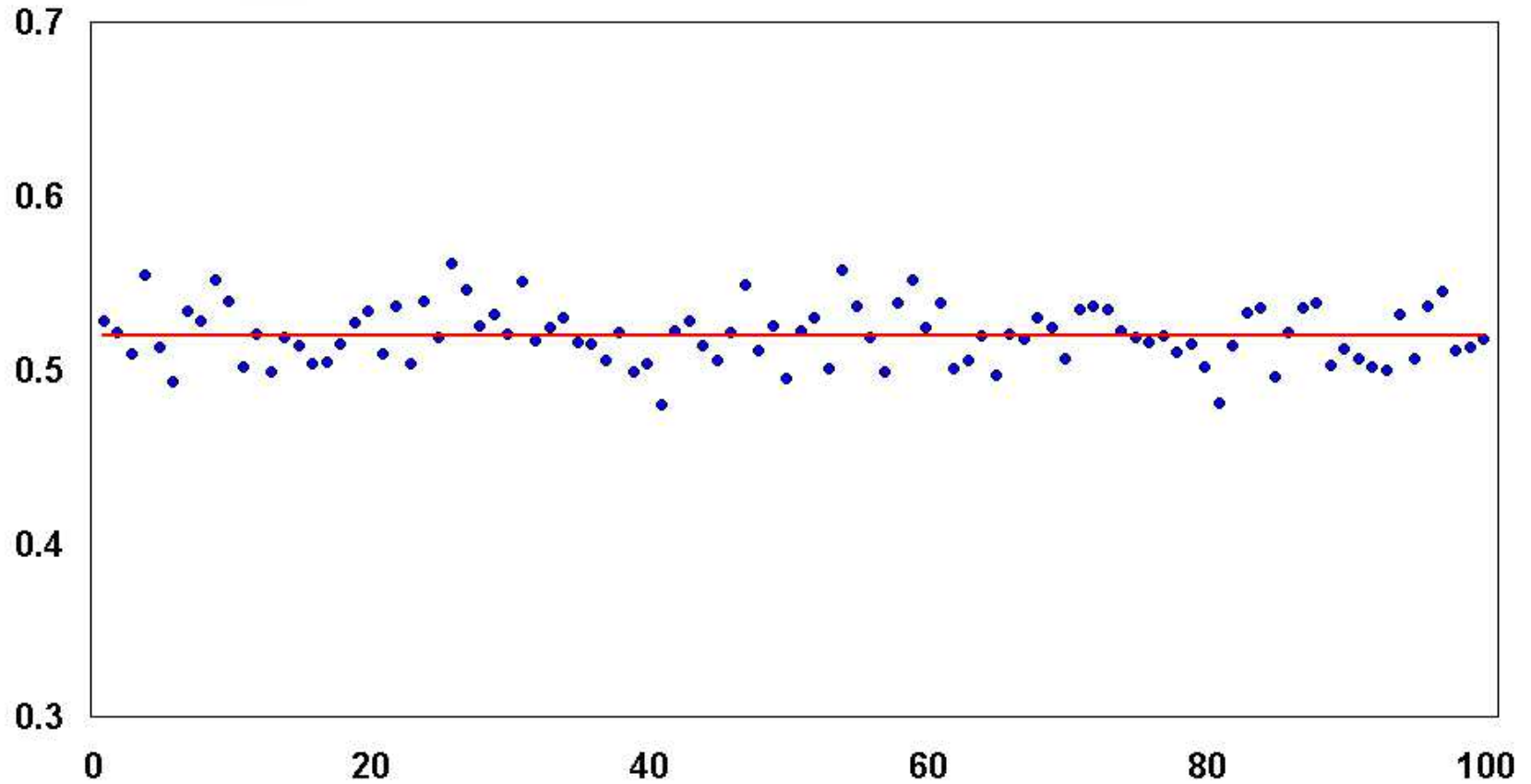
SOMME(...)

qui calcule la somme d'une série

Simulation sur tableur

Voici les résultats du tirage d'une série de 100 échantillons de 1000 nombres 1 et 0 répartis dans une proportion 52 % - 48 %

Échantillonnage



Après avoir effectué les simulations sur tableur, on se rend compte que les résultats obtenus varient en fonction des échantillons ;

c'est ce que l'on appelle :

LA FLUCTUATION D'ÉCHANTILLONNAGE

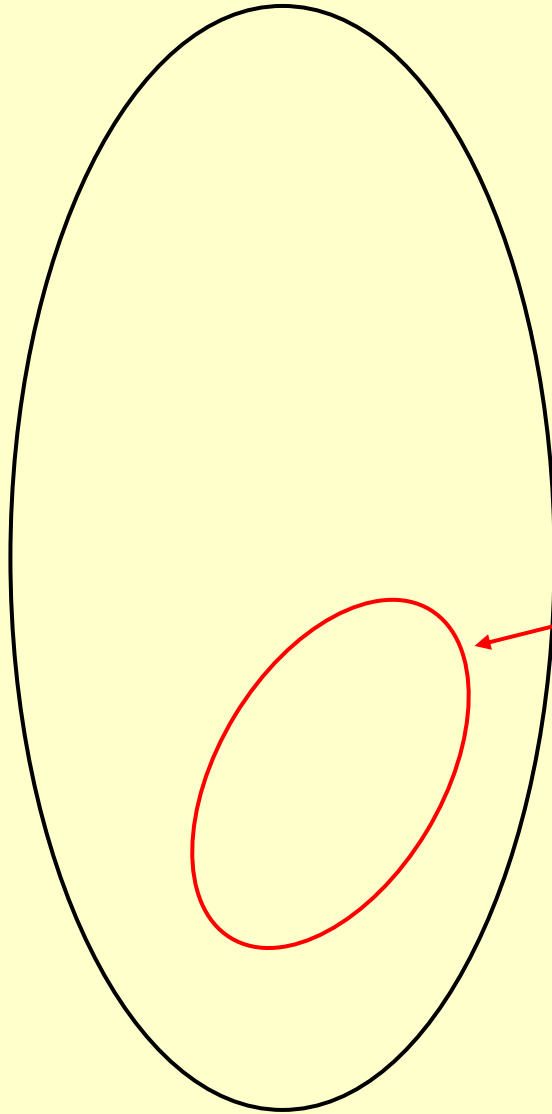
Comment minimiser les effets de la fluctuation d'échantillonnage ?

Une réponse :
l'estimation par
INTERVALLE DE CONFIANCE

On recommence l'expérience :
on extrait un échantillon de taille n
dans la population
et on détermine dans cet échantillon le
pourcentage de ce que l'on cherche.

Population

pourcentage inconnu



Échantillon de taille n

On calcule le
pourcentage dans cet
échantillon,
par exemple $p = 52 \%$

On choisit maintenant
un taux de risque a
(et donc un taux de confiance $1 - a$).
Puis on détermine un intervalle dans
lequel il y aura une probabilité de
 $1 - a$ que se trouve la vraie valeur du
pourcentage cherché dans la
population totale.

La détermination de cet intervalle dit :
INTERVALLE DE CONFIANCE
repose sur des résultats
de la loi normale dont
le théorème central limite.

Pour un risque de 5 %
(valeur couramment employée)
on a une très bonne approximation de
l'intervalle de confiance par l'intervalle

$$\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$$

où p est le pourcentage trouvé dans
l'échantillon et n est la taille de
l'échantillon.

Pour un sondage auprès
de 1000 personnes,
l'intervalle est donc
déterminé à
plus ou moins 3 points.

En guise de conclusion...

Quand on entend comme résultat d'un sondage :

*Il y a actuellement 52 % de gens
qui voteraient pour M. X
au deuxième tour d'une élection
(sondage effectué auprès de 948 personnes),*

il faut comprendre :

*Il y a 95 % de chance pour que l'intervalle
[49% ; 55%]
contienne le pourcentage de gens
prêts à voter pour M. X
au deuxième tour de cette élection.*

Ce qui n'est pas tout à fait la
même chose que de dire que
52 % sont prêts à voter pour lui !

Fin



Équipe académique Mathématiques
Bordeaux