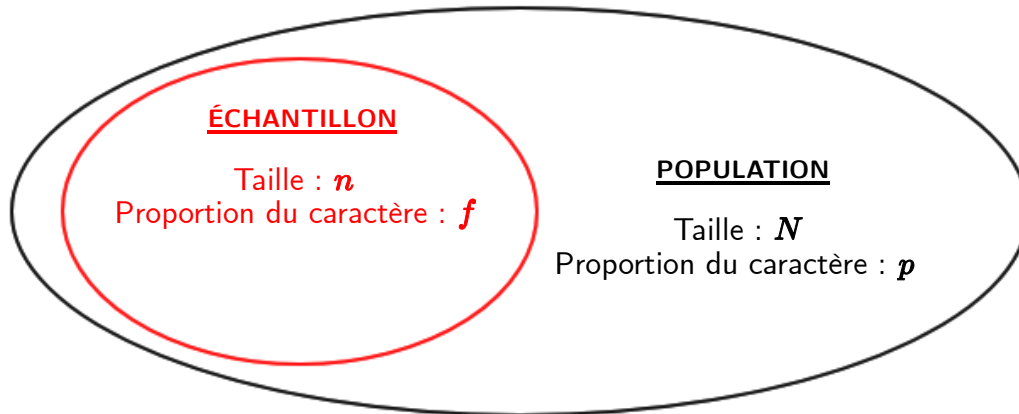


N₁ Intervalle de fluctuation

D Echantillon

Soit une **population** (ce n'est pas forcément des êtres humains) de $N \in \mathbb{N}$ individus. On choisit avec remise n individus, on obtient un **échantillon** de taille n . On note p la proportion du caractère étudié dans la population et on note f cette proportion dans l'échantillon.



P Intervalle de fluctuation

Quand on choisit un échantillon de taille n (tel que $n \geq 25$) dans une population qui contient une proportion p (telle que $p \in [0, 2; 0, 8]$) du caractère étudié alors la proportion f du caractère dans l'échantillon est telle que :

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

avec une probabilité de **95%**.

Cet intervalle est appelé **intervalle de fluctuation asymptotique à 95%** qui a pour amplitude $\frac{2}{\sqrt{n}}$.

- 1 On prélève un échantillon de taille **100** dans une population qui contient **23%** d'éléments du type **A**. Donner l'intervalle de fluctuation asymptotique à **95%** de la proportion f d'éléments du type **A** dans l'échantillon.
- 2 Dans l'article du site Wikipédia consacré au sex-ratio (c'est à dire au rapport du nombre d'hommes sur le nombre de femmes), on trouve l'affirmation suivante : "en France, il y a environ **105** garçons pour **100** filles à la naissance". Dans une clinique il y a environ **1 500** naissances par an.
 - a) Quel est l'intervalle de fluctuation asymptotique à **95%** de la proportion f de garçons dans cette clinique ?
 - b) Quel est l'intervalle de fluctuation asymptotique à **95%** de la proportion f de filles dans cette clinique ?
- 3 Lors du deuxième tour de l'élection présidentielle, il ne reste plus que deux candidats, nommés **A** et **B**. Un sondage est effectué la veille du second tour. Le candidat **A** est crédité de **48%** des intentions de vote.
 - a) Si on choisit un échantillon de taille $n = 1\,000$ au hasard dans la population française, quel est l'intervalle de fluctuation asymptotique à **95%** de la proportion f de personnes favorables au candidat **A** ?
 - b) Si $f = 48\%$ et si on veut que la limite supérieure de l'intervalle de fluctuation à **95%** soit **50%**, quelle valeur de la taille de l'échantillon n doit-on choisir ?

N₂ *Prise de décision***M** *Méthode*

On prélève un échantillon de taille n tel que $n \geq 25$. La proportion p (telle que $p \in [0, 2; 0, 8]$) dans la population est connue et la proportion f dans cet échantillon est connue.

L'échantillon provient-il de la population ? **On fait l'hypothèse que oui.**

$$\text{Si : } f \notin \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$$

Il y a **95%** de chance que l'hypothèse précédente soit fautive : on rejette l'hypothèse au seuil de **95%** ou avec un risque de **5%** (sinon on ne la rejette pas).

Une chaîne de production de tablettes numériques est contrôlée régulièrement. Ces contrôles ont fait apparaître que **20%** des tablettes présentent un léger défaut au niveau de la finition de l'écran qui nécessite un travail supplémentaire.

On décide de changer les méthodes de travail des équipes de cette chaîne de production. Quelques temps après cette modification, on prélève un échantillon de **300** tablettes, et on compte **50** tablettes ayant un défaut de finition de l'écran. Peut-on conclure que la réorganisation du travail a été efficace ?

N₃ *Intervalle de confiance***P** *Intervalle de confiance*

Quand un échantillon de taille n (tel que $n \geq 25$) contient une proportion f (telle que $f \in [0, 2; 0, 8]$) du caractère étudié alors la proportion p du caractère dans la population est telle que :

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

avec une probabilité de **95%**.

Cet intervalle est appelé **intervalle de confiance à 95%** qui a pour amplitude $\frac{2}{\sqrt{n}}$.

En Italie, un sondage est effectué sur la politique budgétaire du pays. La question posée est : "pensez-vous que la rigueur actuelle soit suffisante pour régler le problème du déficit de l'Etat".

Sur les **1 000** personnes interrogées, **30%** ont répondu non. Donner l'intervalle de confiance à **95%** du pourcentage d'Italiens qui répondraient non.

E₁ *Dé pipé*

Un dé semble sortir un nombre anormalement élevé de "6". Il est lancé **150** fois, on relève que la face portant le "6" est sortie **40** fois. Tester l'hypothèse suivante : ce dé n'est pas pipé.

E₂ *Urne*

Une urne contient **40%** de boules noires et le reste de boules rouges. Elle contient un très grand nombre de boules. On présente trois échantillons de boules noires et rouges mais on ignore si ces échantillons proviennent effectivement de l'urne.

- L'échantillon **A** contient **30** boules dont **12** noires.
- L'échantillon **B** contient **81** boules dont **40** noires.
- L'échantillon **C** contient **260** boules dont **80** noires.

Pour chacun des échantillons, examiner l'hypothèse suivante : cet échantillon a été prélevé dans l'urne.

N₄ Comparaison de 2 échantillons**M** Méthode

Soient deux échantillons de proportions respectives f_1 et f_2 d'un caractère. On détermine les deux intervalles de confiance de ces deux échantillons. On fait l'hypothèse qu'ils proviennent d'une même population.

Si les deux intervalles à 95% sont disjoints (aucun nombre en commun) alors on rejette l'hypothèse, sinon on ne la rejette pas.

Le premier janvier **2012** a été décidé une augmentation sur les prix du tabac. Pour étudier l'impact de cette augmentation sur la consommation de cigarettes, un institut de sondage a demandé à deux cents personnes choisies au hasard dans la population française s'ils ont fumés des cigarettes durant la dernière semaine.

Pour la semaine qui a précédé l'augmentation du prix du tabac, le pourcentage de personnes ayant répondu favorablement est de **35%**.

Pour la semaine qui a suivi l'augmentation du prix du tabac, le pourcentage de personnes ayant répondu favorablement est de **32%**.

Peut-on conclure qu'il y a une baisse significative de la consommation de cigarettes ?

E₃ Mouches drosophiles

On élève des mouches drosophiles dans un aquarium. Dans cette grande population de drosophiles de plusieurs milliers d'insectes, il y a **25%** de mouches qui ont les yeux rouges. On prélève un échantillon d'une centaine de mouches et on appelle f la proportion de mouches qui ont les yeux rouges. Donner l'intervalle de fluctuation à **95%** de la proportion f .

E₄ Précaution

Lors du second tour de l'élection présidentielle, on considère que l'on peut annoncer la victoire d'un candidat si l'intervalle de confiance à **95%** ne "mord" pas la barre des **50%**.

- 1** Dans un échantillon de **100** personnes, un candidat est crédité de **53%** des voix. Peut-on annoncer sa victoire ?
- 2** Dans un échantillon de **2 000** personnes, un candidat est crédité de **53%** des voix. Peut-on annoncer sa victoire ?

E₅ Lecteur MP3

Une unité de production réalisant des lectures MP3 a modifié récemment ses méthodes. Avant la modification, sur un échantillon de **130** lecteurs MP3, on a constaté que **42** lecteurs MP3 avaient un défaut. Après la modification, sur un échantillon de **90** lecteurs MP3, on constate que **25** lecteurs MP3 ont un défaut.

Peut-on dire que la modification des méthodes de production a eu un impact significatif sur la qualité de la production ?

E₆ Surpoids

Une nouvelle pilule est testée qui serait censée aider les personnes qui souffrent de surpoids à maigrir. Cette pilule est administrée à une centaine de personnes en surpoids et, après un traitement d'un mois, on constate que **30%** des personnes ont effectivement perdu du poids (on considère que la perte de poids est significative à partir de **2** kilos). Par ailleurs, un placebo est administré à une centaine de personnes en surpoids et on constate que, dans cet échantillon, **28%** des personnes ont perdu du poids.

Peut-on conclure que cette pilule est efficace au vu des résultats obtenus sur ces deux échantillons ?

E₇ **Epidémie**

Le département de la Réunion a fait face à une épidémie. Pour évaluer le budget nécessaire qui permettra d'affronter efficacement cette maladie, il faut connaître le nombre de patients qui seront atteints par cette maladie. L'épidémie est partie du Sud de la Réunion et on estime qu'elle est dans cette partie en phase de disparition. Dans un village de **1 800** habitants, on a comptabilisé **430** cas maintenant que l'épidémie est finie.

- 1** Donner par intervalle de confiance à **95%** la proportion de réunionnais qui seront atteints.
- 2** Sachant que sur l'île de la Réunion on compte environ **800 000** personnes, donner par intervalle de confiance à **95%** le nombre de réunionnais qui seront atteints par cette maladie.
- 3** Si le coût par personne de cette maladie est de **120 €**, quel est par l'intervalle de confiance à **95%**, le budget qu'il faut prévoir pour affronter cette maladie ?

E₈ **Répression des fraudes**

La répression des fraudes a enquêté au casino « Le Lion Vert ». Elle a estimé que la roulette française n'était pas équilibrée et que le casino avait fraudé avec une probabilité d'environ **95%**. Ils ont, pour cela, effectué une analyse informatique des tirages de l'année précédente et remarqué que la couleur rouge était sortie avec une fréquence de **49%** sur un certain nombre de lancers au jeu de la mise sur chance simple Noir-Rouge.

Quelle est la taille minimale de l'échantillon pour que la répression des fraudes puisse arriver à cet conclusion ? La veille, un client notant le déséquilibre en a profité pour l'utiliser. Quelle stratégie le client a-t-il utilisé ?

E₉ **Sur internet**

En décembre 2012, un sondage a été réalisé auprès de **1 003** personnes résidant en France, âgées de **18** ans et plus. L'échantillon a été constitué d'après la méthode des quotas (sexe, âge, catégorie socioprofessionnelle du répondant) par région et taille d'agglomération. **772** personnes interrogées ont déclaré avoir déjà été confrontées à une arnaque ou une tentative d'arnaque sur Internet. Dans le même temps, **211** personnes interrogées déclarent avoir déjà été piégées sur Internet par un mail ou un site Internet leur demandant leurs coordonnées personnelles.

Estimer le pourcentage de personnes en France confrontées à une arnaque sur Internet puis le pourcentage de personnes en France ayant déjà été piégées sur internet.

E₁₀ **Acteur préféré**

Un institut de sondage interroge un groupe de filles sur leur acteur préféré.

- 1** Sur un premier échantillon de **800** filles, **38 %** ont répondu : Léonard Ducape. Déterminer l'intervalle de confiance de ce premier échantillon.
- 2** Sur un deuxième échantillon de **650** filles, **42%** ont, elles, répondu Brad Flip. Déterminer l'intervalle de confiance de ce deuxième échantillon.
- 3** Ces deux intervalles sont-ils disjoints ? Peut-on en déduire que chez les filles Brad Flip a plus de succès que Léonard Ducape ?

E₁₁ **Photographe**

Un photographe vend des appareils photographiques. Il veut estimer par un intervalle de confiance le pourcentage **p** d'acheteurs d'appareils autofocus avec zoom dans sa clientèle

- 1** Dans un échantillon de **100** clients, **60** achètent un tel appareil. Donner une estimation de **p**.
- 2** On considère l'affirmation suivante : « la fréquence **p** est obligatoirement dans l'intervalle de confiance obtenu à la question précédente ». Est-elle vraie ?
- 3** Déterminer la taille **n**, **n ≥ 30**, d'un échantillon de clients pour qu'un intervalle de confiance de **p**, au seuil de **95%** soit **[0, 557; 0, 643]**.