1 Échantillonnage et proportion

1.1 Rappels sur les échantillons

Définition 1. *Un **échantillon** de taille n est un ensemble de n individus prélevés au hasard dans une population.*

*La **fréquence** (ou proportion) d'un caractère dans un échantillon est le nombre d'individus possédant ce caractère divisé par la taille de l'échantillon.*

Exemple 1. *On prélève 100 personnes au hasard dans une population. Parmi elles, 38 sont favorables à une proposition.*

La fréquence de personnes favorables dans cet échantillon est $f = \dfrac{38}{100} = 0{,}38 = 38\%$

1.2 Loi binomiale et échantillonnage

Propriété 1. *Soit une population dans laquelle un caractère a une proportion p.*

On prélève un échantillon de taille n avec remise (ou dans une population très grande).

Si X est le nombre d'individus de l'échantillon ayant ce caractère, alors X suit une loi binomiale $\mathcal{B}(n, p)$.

La fréquence $F = \dfrac{X}{n}$ est une variable aléatoire telle que :

- $E(F) = p$
- $\sigma(F) = \sqrt{\dfrac{p(1-p)}{n}}$

Remarque 1.
- *L'espérance de la fréquence est égale à la proportion dans la population*
- *L'écart-type diminue quand n augmente : plus l'échantillon est grand, plus la fréquence est concentrée autour de p*

2 Intervalle de fluctuation

2.1 Définition

Définition 2. *Un **intervalle de fluctuation** au seuil de 95% de la fréquence d'un caractère de proportion p dans une population, pour des échantillons de taille n, est un intervalle I tel que :*

Pour environ 95% des échantillons de taille n, la fréquence observée appartient à I.

Remarque 2.
- *Le seuil de 95% signifie que dans 95% des cas, la fréquence observée sera dans l'intervalle*
- *Dans environ 5% des cas, la fréquence observée sera en dehors de l'intervalle*
- *C'est un outil pour juger si un échantillon est « normal » ou « surprenant »*

2.2 Intervalle de fluctuation (Programme de Seconde)

Propriété 2 (Intervalle simplifié - Seconde)**.** *Pour $n \geq 25$, $np \geq 5$ et $n(1-p) \geq 5$, un intervalle de fluctuation asymptotique au seuil de 95% est :*

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

Exemple 2. *Une pièce équilibrée a une probabilité $p = 0{,}5$ de tomber sur Pile.*
On lance 100 fois la pièce. Déterminer l'intervalle de fluctuation au seuil de 95%.
Solution :
Conditions : $n = 100 \geq 25$, $np = 50 \geq 5$, $n(1-p) = 50 \geq 5$ *: vérifiées.*
$I = \left[0{,}5 - \frac{1}{\sqrt{100}}; 0{,}5 + \frac{1}{\sqrt{100}}\right]$
$I = [0{,}5 - 0{,}1; 0{,}5 + 0{,}1] = [0{,}4; 0{,}6]$
Dans environ 95% des échantillons de 100 lancers, la fréquence de Pile sera entre 40% et 60%.

2.3 Intervalle de fluctuation (Programme de Première)

Propriété 3 (Intervalle de fluctuation - Première)**.** *Pour* $n \geq 30$, $np \geq 5$ *et* $n(1-p) \geq 5$, *l'intervalle de fluctuation asymptotique au seuil de 95% est :*

$$I = \left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}}; p + 1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$$

On utilise souvent l'approximation $1{,}96 \approx 2$.

Remarque 3. *Le coefficient 1,96 provient de la loi normale :* $P(-1{,}96 < Z < 1{,}96) \approx 0{,}95$ *pour* Z *suivant une loi normale centrée réduite* $\mathcal{N}(0; 1)$.

Exemple 3. *Dans une population, 30% des individus ont un certain caractère.*
On prélève un échantillon de 200 individus. Déterminer l'intervalle de fluctuation au seuil de 95%.
Solution :
$p = 0{,}3$, $n = 200$
Conditions : $n = 200 \geq 30$, $np = 60 \geq 5$, $n(1-p) = 140 \geq 5$ *: vérifiées.*
$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}3 \times 0{,}7}{200}} = \sqrt{\frac{0{,}21}{200}} \approx 0{,}0324$
$I = [0{,}3 - 1{,}96 \times 0{,}0324; 0{,}3 + 1{,}96 \times 0{,}0324]$
$I \approx [0{,}3 - 0{,}0635; 0{,}3 + 0{,}0635] = [0{,}2365; 0{,}3635]$
$I \approx [0{,}237; 0{,}364]$ *ou* $[23{,}7\%; 36{,}4\%]$

2.4 Prise de décision

Méthode 1 (Utiliser l'intervalle de fluctuation)**.** *Pour décider si un échantillon observé est compatible avec une proportion* p *connue :*

1. *Calculer l'intervalle de fluctuation* I *au seuil de 95%*
2. *Calculer la fréquence* f *observée dans l'échantillon*
3. *Si* $f \in I$ *: l'échantillon est compatible avec la proportion* p *(on accepte l'hypothèse)*
4. *Si* $f \notin I$ *: l'échantillon n'est pas compatible avec* p *au seuil de 95% (on rejette l'hypothèse)*

Exemple 4. *Un industriel affirme que 20% de ses produits sont défectueux.*
On prélève un échantillon de 100 produits et on trouve 32 défectueux.
Peut-on remettre en cause l'affirmation de l'industriel au seuil de 95% ?
Solution :
$p = 0{,}2$, $n = 100$, *fréquence observée :* $f = \frac{32}{100} = 0{,}32$
Conditions : $n = 100 \geq 30$, $np = 20 \geq 5$, $n(1-p) = 80 \geq 5$ *: vérifiées.*
Intervalle de fluctuation :

$\sqrt{\dfrac{0{,}2 \times 0{,}8}{100}} = \sqrt{0{,}0016} = 0{,}04$
$I = [0{,}2 - 1{,}96 \times 0{,}04; 0{,}2 + 1{,}96 \times 0{,}04]$
$I = [0{,}2 - 0{,}0784; 0{,}2 + 0{,}0784] = [0{,}1216; 0{,}2784]$
$I \approx [0{,}122; 0{,}278]$
Comme $f = 0{,}32 \notin I$, on peut remettre en cause l'affirmation de l'industriel au seuil de 95%.
Il semble y avoir plus de 20% de produits défectueux.

3 Intervalle de confiance

3.1 Principe

Définition 3. *Un **intervalle de confiance** est un intervalle qui, à partir d'une fréquence observée f dans un échantillon, permet d'estimer la proportion p inconnue dans la population.*
C'est la situation inverse de l'intervalle de fluctuation :

- *Intervalle de fluctuation : p connu, on prédit f*
- *Intervalle de confiance : f observé, on estime p*

3.2 Formule

Propriété 4 (Intervalle de confiance au niveau 95%)**.** *Pour $n \geq 30$, $nf \geq 5$ et $n(1-f) \geq 5$, un intervalle de confiance de la proportion p au niveau de confiance 95% est :*

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

Remarque 4.
- *On utilise la fréquence observée f (et non p qui est inconnu)*
- *L'intervalle est d'autant plus petit que n est grand*
- *Le niveau de confiance 95% signifie : si on répétait l'expérience de nombreuses fois, environ 95% des intervalles calculés contiendraient la vraie valeur de p*

Exemple 5. ***Sondage électoral :***
On interroge 1000 personnes. 520 déclarent vouloir voter pour le candidat A.
Estimer la proportion de votants pour A dans toute la population.
Solution :
$n = 1000$, *fréquence observée :* $f = \dfrac{520}{1000} = 0{,}52$
Conditions : $n = 1000 \geq 30$, $nf = 520 \geq 5$, $n(1-f) = 480 \geq 5$ *: vérifiées.*
$I = \left[0{,}52 - \dfrac{1}{\sqrt{1000}}; 0{,}52 + \dfrac{1}{\sqrt{1000}}\right]$
$\dfrac{1}{\sqrt{1000}} \approx 0{,}0316$
$I \approx [0{,}52 - 0{,}032; 0{,}52 + 0{,}032] = [0{,}488; 0{,}552]$
$I \approx [48{,}8\%; 55{,}2\%]$
Au niveau de confiance 95%, la proportion de votants pour A dans la population est comprise entre 48,8% et 55,2%.

3.3 Marge d'erreur

Définition 4. *La **marge d'erreur** d'un sondage est la demi-largeur de l'intervalle de confiance.*
Pour un sondage de taille n, la marge d'erreur au niveau 95% est :

$$e = \frac{1}{\sqrt{n}}$$

Exemple 6. ***Taille d'échantillon pour une marge donnée :***
On veut réaliser un sondage avec une marge d'erreur de 3%. Quelle doit être la taille de l'échantillon ?
Solution :
$\frac{1}{\sqrt{n}} = 0{,}03$
$\sqrt{n} = \frac{1}{0{,}03} = 33{,}33$
$n = 33{,}33^2 \approx 1111$
Il faut interroger environ 1111 personnes.

Taille d'échantillon	**Marge d'erreur**
100	$\pm 10\%$
400	$\pm 5\%$
1000	$\pm 3{,}2\%$
2500	$\pm 2\%$
10000	$\pm 1\%$

4 Tests d'hypothèse

4.1 Principe du test

Définition 5. *Un **test d'hypothèse** est une procédure qui permet de décider si une hypothèse sur une population est vraisemblable ou non, à partir d'un échantillon.*
On distingue :

— *L'**hypothèse nulle** H_0 : l'hypothèse que l'on teste (généralement « pas de changement »)*

— *L'**hypothèse alternative** H_1 : ce qu'on suspecte si H_0 est fausse*

Définition 6. *Le **seuil de risque** α (souvent 5% = 0,05) est la probabilité de rejeter H_0 alors qu'elle est vraie.*
C'est le risque d'erreur que l'on accepte de prendre.

4.2 Procédure du test

Méthode 2 (Test d'une proportion)**.** *Pour tester si une proportion dans une population vaut p_0 :*
Étape 1 : *Formuler les hypothèses*

— $H_0 : p = p_0$ *(hypothèse nulle)*

— H_1 : $p \neq p_0$ *(hypothèse alternative)*

Étape 2 : *Choisir le seuil de risque* α *(généralement 5%)*
Étape 3 : *Calculer l'intervalle de fluctuation de* p_0 *au seuil* $1-\alpha$
Étape 4 : *Observer la fréquence* f *dans l'échantillon*
Étape 5 : *Décision*

— *Si* $f \in I$ *: on accepte* H_0 *(on ne rejette pas)*

— *Si* $f \notin I$ *: on rejette* H_0 *au seuil* α

Exemple 7. ***Test d'une pièce truquée :***
On suspecte qu'une pièce est truquée. On la lance 200 fois et on obtient 120 fois Pile.
Peut-on affirmer au seuil de 5% que la pièce est truquée ?
Solution :
Hypothèses :

— H_0 *: la pièce est équilibrée,* $p = 0{,}5$

— H_1 *: la pièce est truquée,* $p \neq 0{,}5$

Seuil : $\alpha = 5\%$
Intervalle de fluctuation :
$n = 200$, $p_0 = 0{,}5$
$\sqrt{\dfrac{0{,}5 \times 0{,}5}{200}} = \sqrt{\dfrac{0{,}25}{200}} = \sqrt{0{,}00125} \approx 0{,}0354$
$I = [0{,}5 - 1{,}96 \times 0{,}0354; 0{,}5 + 1{,}96 \times 0{,}0354]$
$I \approx [0{,}5 - 0{,}069; 0{,}5 + 0{,}069] = [0{,}431; 0{,}569]$
Fréquence observée : $f = \dfrac{120}{200} = 0{,}6$
Décision :
$f = 0{,}6 \notin I$
On rejette H_0 *au seuil de 5%. On peut affirmer que la pièce est truquée.*

4.3 Tests unilatéraux

Définition 7. *Un* ***test unilatéral*** *ne s'intéresse qu'à un sens de variation.*

— *Test unilatéral à droite :* $H_0 : p \leq p_0$ *contre* $H_1 : p > p_0$

— *Test unilatéral à gauche :* $H_0 : p \geq p_0$ *contre* $H_1 : p < p_0$

Pour un test unilatéral au seuil 5%, on utilise le coefficient 1,64 au lieu de 1,96.

Exemple 8. ***Efficacité d'un médicament :***
Un médicament est efficace dans 60% des cas. Un laboratoire développe un nouveau médicament et veut prouver qu'il est plus efficace.
On teste le nouveau médicament sur 100 patients et il est efficace pour 72 d'entre eux.
Le nouveau médicament est-il significativement plus efficace au seuil de 5% ?
Solution :
Hypothèses :

— H_0 *: le nouveau médicament n'est pas plus efficace,* $p \leq 0{,}6$

— H_1 *: le nouveau médicament est plus efficace,* $p > 0{,}6$

Test unilatéral à droite.
Valeur critique :
$p_0 = 0{,}6$, $n = 100$

$\sqrt{\dfrac{0{,}6 \times 0{,}4}{100}} = \sqrt{0{,}0024} \approx 0{,}049$
Seuil de rejet : $0{,}6 + 1{,}64 \times 0{,}049 \approx 0{,}6 + 0{,}08 = 0{,}68$
On rejette H_0 si $f > 0{,}68$.
Fréquence observée : $f = \dfrac{72}{100} = 0{,}72$
Décision :
$f = 0{,}72 > 0{,}68$
On rejette H_0 au seuil de 5%. Le nouveau médicament est significativement plus efficace.

4.4 Erreurs de décision

Définition 8. *Dans un test d'hypothèse, deux types d'erreurs sont possibles :*
Erreur de type I (ou de première espèce) : *Rejeter H_0 alors qu'elle est vraie.*
Probabilité : α (seuil du test)
Erreur de type II (ou de seconde espèce) : *Accepter H_0 alors qu'elle est fausse.*
Probabilité : β (dépend de la vraie valeur de p et de n)

Décision / Réalité	H_0 **vraie**	H_0 **fausse**
Accepter H_0	Bonne décision	Erreur de type II (β)
Rejeter H_0	Erreur de type I (α)	Bonne décision

Remarque 5.
- *On ne peut pas diminuer simultanément α et β à taille d'échantillon fixée*
- *Pour diminuer les deux risques, il faut augmenter la taille de l'échantillon*
- *En pratique, on fixe α (souvent à 5%) et on essaie de minimiser β*

5 Applications

5.1 Contrôle qualité

Application 1. ***Exercice 1 :*** *Une entreprise produit des pièces avec un taux de défauts de 2%.*
On prélève un échantillon de 500 pièces et on trouve 15 pièces défectueuses.
Le processus de fabrication s'est-il dégradé ? (seuil 5%)
Solution :
$p = 0{,}02$, $n = 500$, $f = \dfrac{15}{500} = 0{,}03$
Conditions : $n = 500 \geq 30$, $np = 10 \geq 5$, $n(1-p) = 490 \geq 5$ *: vérifiées.*
$\sqrt{\dfrac{0{,}02 \times 0{,}98}{500}} = \sqrt{0{,}0000392} \approx 0{,}00626$
$I = [0{,}02 - 1{,}96 \times 0{,}00626; 0{,}02 + 1{,}96 \times 0{,}00626]$
$I \approx [0{,}02 - 0{,}0123; 0{,}02 + 0{,}0123] = [0{,}0077; 0{,}0323]$
$f = 0{,}03 \in I$
On ne peut pas affirmer que le processus s'est dégradé au seuil de 5%.

5.2 Sondage

Application 2. ***Exercice 2 :*** *Un sondage sur 800 personnes donne 456 intentions de vote pour un candidat.*

Donner un intervalle de confiance au niveau 95% pour le score du candidat.

Le candidat peut-il espérer être élu (majorité absolue = 50%) ?

Solution :

$n = 800,\ f = \dfrac{456}{800} = 0{,}57$

$I = \left[0{,}57 - \dfrac{1}{\sqrt{800}}; 0{,}57 + \dfrac{1}{\sqrt{800}}\right]$

$\dfrac{1}{\sqrt{800}} \approx 0{,}0354$

$I \approx [0{,}57 - 0{,}035; 0{,}57 + 0{,}035] = [0{,}535; 0{,}605]$

$I \approx [53{,}5\%; 60{,}5\%]$

Comme tout l'intervalle est au-dessus de 50%, le candidat peut raisonnablement espérer être élu.

5.3 Essai clinique

Application 3. ***Exercice 3 :*** *Un médicament classique guérit 70% des patients.*

On teste un nouveau médicament sur 150 patients et 115 guérissent.

Le nouveau médicament est-il plus efficace ? (seuil 5%)

Solution :

$H_0 : p = 0{,}7$ *(même efficacité)*

$H_1 : p > 0{,}7$ *(plus efficace) - test unilatéral*

$n = 150,\ f = \dfrac{115}{150} \approx 0{,}767$

$\sqrt{\dfrac{0{,}7 \times 0{,}3}{150}} = \sqrt{0{,}0014} \approx 0{,}0374$

Seuil critique : $0{,}7 + 1{,}64 \times 0{,}0374 \approx 0{,}7 + 0{,}061 = 0{,}761$

$f = 0{,}767 > 0{,}761$

On rejette H_0 au seuil de 5%. Le nouveau médicament est significativement plus efficace.

6 Tableaux récapitulatifs

6.1 Formules essentielles

Intervalle de fluctuation (Première)

Proportion connue	p
Conditions	$n \geq 30,\ np \geq 5,\ n(1-p) \geq 5$
Formule	$\left[p - 1{,}96\sqrt{\dfrac{p(1-p)}{n}}; p + 1{,}96\sqrt{\dfrac{p(1-p)}{n}}\right]$
Utilisation	Décider si un échantillon est compatible avec p

Intervalle de confiance

Fréquence observée	f
Conditions	$n \geq 30$, $nf \geq 5$, $n(1-f) \geq 5$
Formule	$\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$
Utilisation	Estimer la proportion p dans la population

6.2 Comparaison

	Intervalle de fluctuation	**Intervalle de confiance**
On connaît	p (proportion théorique)	f (fréquence observée)
On estime	f (fréquence attendue)	p (proportion réelle)
Question	L'échantillon est-il normal ?	Quelle est la vraie valeur de p ?
Formule	Utilise p	Utilise f

6.3 Démarche du test

Étape	**Action**
1	Formuler H_0 et H_1
2	Choisir le seuil α (souvent 5%)
3	Calculer l'intervalle de fluctuation de p_0
4	Observer f dans l'échantillon
5	Si $f \in I$: accepter H_0 ; si $f \notin I$: rejeter H_0
6	Conclure en contexte

6.4 Aide-mémoire

Coefficient	**Utilisation**
1,96	Test bilatéral, seuil 5%
1,64	Test unilatéral, seuil 5%
$\dfrac{1}{\sqrt{n}}$	Marge d'erreur simplifiée
$\sqrt{\dfrac{p(1-p)}{n}}$	Écart-type de la fréquence