

## Statistiques inférentielles

# I. Échantillonnage

On connaît la loi de  $X$  sur une population. On s'intéresse à un échantillon.

## I. 1) Loi des moyennes

Dans une population de moyenne  $\mu$  et d'écart-type  $\sigma$ .

On prélève un échantillon de taille  $n$ .

On relève la moyenne  $\overline{X}_n$  de l'échantillon.

$\overline{X}_n$  suit approximativement une loi  $\mathcal{N}$  de moyenne  $m$  d'écart-type  $\frac{\sigma}{\sqrt{n}}$

Si la population suit une loi normale, ce n'est pas une approximation.

## I. 1) Loi des moyennes

Dans une population de moyenne  $\mu$  et d'écart-type  $\sigma$ .

On prélève un échantillon de taille  $n$ .

On relève la moyenne  $\overline{X}_n$  de l'échantillon.

$\overline{X}_n$  suit approximativement une loi  $\mathcal{N}$  de moyenne  $m$  d'écart-type  $\frac{\sigma}{\sqrt{n}}$

Si la population suit une loi normale, ce n'est pas une approximation.

**Exemple :** Une usine fabrique des tiges. En moyenne, le diamètre des tiges est de 20 mm, avec un écart type de 0,1 mm.

On prélève un échantillon de 100 tiges. Soit  $\overline{X}_{100}$  la moyenne des diamètres de cet échantillon. Quelle est la probabilité que  $19,99 \leq \overline{X}_{100} \leq 20,01$  ?

# I. 1) Loi des moyennes

Dans une population de moyenne  $\mu$  et d'écart-type  $\sigma$ .

On prélève un échantillon de taille  $n$ .

On relève la moyenne  $\overline{X}_n$  de l'échantillon.

$\overline{X}_n$  suit approximativement une loi  $\mathcal{N}$  de moyenne  $m$  d'écart-type  $\frac{\sigma}{\sqrt{n}}$

Si la population suit une loi normale, ce n'est pas une approximation.

**Exemple :** Une usine fabrique des tiges. En moyenne, le diamètre des tiges est de 20 mm, avec un écart type de 0,1 mm.

On prélève un échantillon de 100 tiges. Soit  $\overline{X}_{100}$  la moyenne des diamètres de cet échantillon. Quelle est la probabilité que  $19,99 \leq \overline{X}_{100} \leq 20,01$  ?

$\overline{X}$  suit la loi  $\mathcal{N}$  de moyenne 20 et écart-type  $\frac{0,1}{\sqrt{100}} = 0,01$ . La probabilité recherchée est donc :

$$p(19,99 \leq \overline{X} \leq 20,01) \simeq 0,6826$$

# Exercice 1

Une usine produit des disques en grande série. On appelle  $X$  la variable aléatoire qui à chaque disque, pris au hasard, associe son diamètre. On suppose que  $X$  suit la loi normale de moyenne  $\mu = 50$  et d'écart-type  $\sigma = 1$ .

On prélève au hasard des échantillons de 100 disques. On appelle  $\overline{X}_{100}$  la variable aléatoire correspondant à la moyenne des valeurs de l'échantillon.

- Quelle est la loi suivie par  $\overline{X}_{100}$  ?
- Calculez  $p(49,7 \leq \overline{X}_{100} \leq 50,3)$ .
- Déterminez  $a$  pour que  $p(\mu - a \leq \overline{X}_{100} \leq \mu + a) = 0,95$ .

## Exercice 2

Un traceur produit des cartouches ayant une forme de rectangle dont la longueur doit être de 10 cm. Pour tester le réglage du traceur, on prélève 100 feuilles. Soit  $X$  la variable aléatoire qui pour chaque feuille associe la longueur du cartouche. On suppose que  $X$  suit la loi normale de moyenne  $m = 10$  et d'écart-type  $\sigma = 0,03$ . On note  $\overline{X}_{100}$  la variable qui à chaque échantillon associe la moyenne des longueurs.

- Quelle est la loi suivie par  $\overline{X}_{100}$  ?
- Déterminez le réel positif  $h$  tel que 
$$p(m - h \leq \overline{X}_{100} \leq m + h) = 0,95.$$

## I. 2) Loi des fréquences

Dans une population la proportion  $p$  d'individus possèdent une propriété.  
On prélève aléatoirement un échantillon de taille  $n$ .

$F$  est la fréquence de la propriété dans l'échantillon  $\Rightarrow F$  suit

**approximativement** la loi  $\mathcal{N}$  de moyenne  $p$  et d'écart-type  $\sqrt{\frac{p(1-p)}{n}}$

## I. 2) Loi des fréquences

Dans une population la proportion  $p$  d'individus possèdent une propriété. On prélève aléatoirement un échantillon de taille  $n$ .

$F$  est la fréquence de la propriété dans l'échantillon  $\Rightarrow F$  suit

**approximativement** la loi  $\mathcal{N}$  de moyenne  $p$  et d'écart-type  $\sqrt{\frac{p(1-p)}{n}}$

**Exemple** : Lors d'un référendum, 53 % des électeurs vont voter OUI. On prélève un échantillon de 1 000 électeurs. Quelle est la probabilité qu'il y ait entre 52,5 % et 53,5 % d'électeurs votant OUI dans cet échantillon ?

## I. 2) Loi des fréquences

Dans une population la proportion  $p$  d'individus possèdent une propriété. On prélève aléatoirement un échantillon de taille  $n$ .

$F$  est la fréquence de la propriété dans l'échantillon  $\Rightarrow F$  suit

**approximativement** la loi  $\mathcal{N}$  de moyenne  $p$  et d'écart-type  $\sqrt{\frac{p(1-p)}{n}}$

**Exemple** : Lors d'un référendum, 53 % des électeurs vont voter OUI. On prélève un échantillon de 1 000 électeurs. Quelle est la probabilité qu'il y ait entre 52,5 % et 53,5 % d'électeurs votant OUI dans cet échantillon ?

Soit  $F$  la variable aléatoire représentant le pourcentage d'électeurs de l'échantillon votant OUI.  $F$  suit la loi  $\mathcal{N}$  de moyenne 0,53 et d'écart-type  $\sqrt{\frac{0,53 \times 0,47}{1000}} \simeq 0,0158$ . La probabilité recherchée est donc :

$$p(0,525 \leq F \leq 0,535) \simeq 0,2472$$

## Exercice 3

Une entreprise produit une grande série de pièces pour le bâtiment. Pour analyser la qualité de la production, on prélève au hasard un échantillon de  $n = 200$  pièces.

On suppose que le pourcentage de pièces défectueuses dans la production est de  $p = 4\%$ . Soit  $F$  la variable aléatoire qui à tout échantillon de  $n$  pièces associe la fréquence de pièces défectueuses dans cet échantillon.

- quelle est, approximativement, la loi suivie par  $F$ ? Précisez-en les paramètres.
- Quel est la probabilité que sur un échantillon donné, la fréquence de pièce défectueuses soit comprise entre  $3\%$  et  $5\%$ ?
- Déterminez le réel  $a$  tel que  $p(0,04 - a \leq F \leq 0,04 + a) = 0,95$ .
- Quelle valeur doit-on donner à  $n$  pour que  $p(0,03 \leq F \leq 0,05) = 0,99$

## Exercice 4

Dans une population, on constate qu'il y a 28 % de fumeurs. On prend un échantillon de 400 personnes.

- a) Quel est la probabilité d'avoir dans cet échantillon un pourcentage de fumeurs compris entre 26 % et 30 % de fumeurs ?
- b) Quelle est la probabilité d'avoir dans cet échantillon un pourcentage de fumeurs inférieur à 22 % ?

## II. Estimation

On a réalisé un échantillon. On connaît les paramètres de l'échantillon et on veut estimer ceux de la population.

## II. 1) Estimation ponctuelle

Dans un échantillon de taille  $n$  on a relevé la moyenne  $\bar{x}_e$  et l'écart-type  $\sigma_e$ .  
*ou bien* on a relevé la fréquence  $f_e$ .

## II. 1) Estimation ponctuelle

Dans un échantillon de taille  $n$  on a relevé la moyenne  $\bar{x}_e$  et l'écart-type  $\sigma_e$ .  
*ou bien* on a relevé la fréquence  $f_e$ .

1. La moyenne de la population est estimée par  $\bar{x}$ .
2. La fréquence dans la population est estimée par  $f$ .
3. L'écart-type  $\sigma$  dans la population est estimée par  $s = \sqrt{\frac{n}{n-1}}\sigma_e$ .

Le dernier montre que la mesure de  $\sigma_e$  sous-estime le  $\sigma$  réel : Il y a en moyenne moins de diversité dans un échantillon que dans la population entière !

## Exemple

À un examen auquel se présentent 100 000 candidats, 100 copies ont été corrigées. La moyenne sur ce paquet était de 12 avec un écart-type de 3. Dans ce paquet, 5 copies étaient blanches.

## Exemple

À un examen auquel se présentent 100 000 candidats, 100 copies ont été corrigées. La moyenne sur ce paquet était de 12 avec un écart-type de 3. Dans ce paquet, 5 copies étaient blanches.

On peut estimer que sur l'ensemble des copies :

- a) la moyenne était de 12,
- b) l'écart-type  $\sigma = \sqrt{\frac{100}{99}} 3 \simeq 3,015$ ,
- c) 5% de l'ensemble des copies étaient blanches.

Il arrive souvent que l'on approxime  $\sqrt{\frac{n}{n-1}} \approx 1$  dans l'estimation de  $\sigma$ . En effet, pour  $n$  assez grand, cela ne change pas grand chose.

## II. 2) Estimation d'une moyenne par intervalle de confiance

On se donne un taux de confiance, par exemple 95%. Ce taux de confiance correspond à un certain facteur  $t$ . Par exemple,  $t = 1,96$ .

L'intervalle de confiance de la moyenne est :

$$I_C = \left[ \bar{x}_e - t \frac{\sigma}{\sqrt{n}}; \bar{x}_e + t \frac{\sigma}{\sqrt{n}} \right]$$

$\bar{x}_e$  est la moyenne de l'échantillon considéré,  $\sigma$  est l'écart-type de la population (s'il est connu) où son estimation ponctuelle  $s$ .

## II. 2) Estimation d'une moyenne par intervalle de confiance

On se donne un taux de confiance, par exemple 95 %. Ce taux de confiance correspond à un certain facteur  $t$ . Par exemple,  $t = 1,96$ .

L'intervalle de confiance de la moyenne est :

$$I_C = \left[ \bar{x}_e - t \frac{\sigma}{\sqrt{n}} ; \bar{x}_e + t \frac{\sigma}{\sqrt{n}} \right]$$

$\bar{x}_e$  est la moyenne de l'échantillon considéré,  $\sigma$  est l'écart-type de la population (s'il est connu) où son estimation ponctuelle  $s$ .

Dans l'exemple précédent,  $\bar{x}_e = 12$  et  $\sigma \approx 3,015$ . Si on prend le taux de confiance 95 %, on aura  $t = 1,96$  et

$$I_C \approx \left[ 12 - 1,96 \frac{3,015}{\sqrt{100}} ; 12 + 1,96 \frac{3,015}{\sqrt{100}} \right] \approx [11,409 ; 12,591]$$

## Exercice 5

Une entreprise fabrique des tiges métalliques. Leur longueur et leur diamètre sont exprimés en millimètres.

On prélève au hasard un échantillon de  $n = 50$  tiges dans la production d'une journée.

Dans l'échantillon, on a obtenu une moyenne  $\mu_e = 9,99$  avec un écart-type  $\sigma_e = 0,19$ .

- À partir de ces informations, donnez une estimation ponctuelle de la moyenne  $\mu$  des diamètres des tiges produites dans une journée.
- Déterminez un intervalle de confiance de la moyenne  $\mu$  à 95%.
- On considère l'affirmation : «  $\mu$  appartient obligatoirement à l'intervalle de la question précédente ». Cette affirmation est-elle vraie ?

## Exercice 6

On a contrôlé le dosage d'un produit dans un mélange à la sortie d'une chaîne de conditionnement. On a prélevé un échantillon de 100 lots de 5 kg dans la production d'une journée. On a obtenu les résultats suivants, où  $P_i$  représente la masse de produit en grammes, et  $n_i$  l'effectif correspondant.

$P_i$	142	144	146	148	150	152	154	156	158	160
$n_i$	1	5	6	21	32	22	7	4	1	1

- Calculez  $\mu_e$  et  $\sigma_e$ .
- Déduisez-en une estimation ponctuelle de  $\mu$  et  $\sigma$  de masse de produit dans la production de la journée.
- Donnez un intervalle de confiance de la moyenne  $\mu$  avec le coefficient de confiance de 5%.
- même question avec un coefficient de confiance de 99 %.

## II. 3) Estimation d'une fréquence par intervalle de confiance

L'intervalle de confiance de la fréquence est :

$$I_C = \left[ f - t \sqrt{\frac{f(1-f)}{n-1}} ; f + t \sqrt{\frac{f(1-f)}{n-1}} \right]$$

$f$  est la fréquence du caractère dans l'échantillon considéré.

On note généralement  $p$  la fréquence dans l'ensemble de la population.

## II. 3) Estimation d'une fréquence par intervalle de confiance

L'intervalle de confiance de la fréquence est :

$$I_C = \left[ f - t \sqrt{\frac{f(1-f)}{n-1}} ; f + t \sqrt{\frac{f(1-f)}{n-1}} \right]$$

$f$  est la fréquence du caractère dans l'échantillon considéré.

On note généralement  $p$  la fréquence dans l'ensemble de la population.

Dans l'exemple du paquet de copies :  $n = 100$  copies,  $f = 5\%$  de copies blanches. On souhaite déterminer un intervalle de confiance de  $p$  au seuil de confiance de 95 %.

$$I_C = \left[ 0,05 - 1,96 \sqrt{\frac{0,05 \times 0,95}{99}} ; 0,05 + 1,96 \sqrt{\frac{0,05 \times 0,95}{99}} \right]$$

$$I_C \simeq [0,007 ; 0,093]$$

## Exercice 7

Dans une ville, la proportion  $p$  de ménages qui possèdent au moins un téléviseur est inconnue.

On interroge 400 ménages. 327 de ces ménages possèdent au moins un téléviseur.

- Donnez une estimation ponctuelle de  $p$ .
- Donnez une estimation de  $p$  par intervalle de confiance avec le risque 5%.
- Quelle taille d'échantillon aurait-il fallu prendre pour obtenir un intervalle de confiance à  $\pm 2\%$ , avec un taux de confiance de 99 % ?

### III. Test de validité

### III. 1) Principe général

Il s'agit de prendre une décision sur l'ensemble de la population à partir d'une étude sur des échantillons.

On émet une hypothèse : « la moyenne dans la population est ... » ou « la fréquence dans la population est ... ».

On doit accepter ou rejeter l'hypothèse.

**Méthode :** Avant le prélèvement d'un échantillon :

- a) On choisit l'hypothèse nulle  $H_0$  et l'hypothèse alternative  $H_1$ .
- b) On détermine l'intervalle de fluctuation  $I_F$  sous l'hypothèse  $H_0$  avec un seuil de risque  $\alpha$  donné.
- c) On énonce la règle de décision :
  - Si la moyenne / la fréquence de l'échantillon est dans  $I_F$ , on accepte  $H_0$
  - sinon on accepte  $H_1$  (donc on rejette  $H_0$ )

## III. 2) Test bilatéral

On teste l'affirmation : dans la population, la moyenne est

$$\mu = \mu_{\text{AFFIRMATION}}$$

Pour simplifier, dans ce cas, on suppose  $\sigma$  connu.

Au lieu de  $\mu$  ce pourrait être  $p$  pour une fréquence.

## III. 2) Test bilatéral

On teste l'affirmation : dans la population, la moyenne est

$$\mu = \mu_{\text{AFFIRMATION}}$$

Pour simplifier, dans ce cas, on suppose  $\sigma$  connu.

Au lieu de  $\mu$  ce pourrait être  $p$  pour une fréquence.

a)  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu \neq \mu_{\text{AFFIRMATION}}$

## III. 2) Test bilatéral

On teste l'affirmation : dans la population, la moyenne est

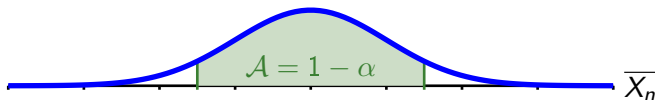
$$\mu = \mu_{\text{AFFIRMATION}}$$

Pour simplifier, dans ce cas, on suppose  $\sigma$  connu.

Au lieu de  $\mu$  ce pourrait être  $p$  pour une fréquence.

a)  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu \neq \mu_{\text{AFFIRMATION}}$

b)  $I_F$  au risque  $\alpha \rightarrow$  au taux de confiance  $1 - \alpha$ , supposant  $H_0$  vraie :



## III. 2) Test bilatéral

On teste l'affirmation : dans la population, la moyenne est

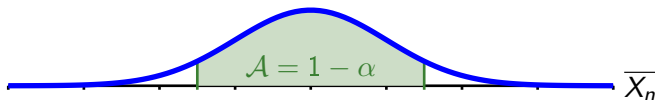
$$\mu = \mu_{\text{AFFIRMATION}}$$

Pour simplifier, dans ce cas, on suppose  $\sigma$  connu.

Au lieu de  $\mu$  ce pourrait être  $p$  pour une fréquence.

a)  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu \neq \mu_{\text{AFFIRMATION}}$

b)  $I_F$  au risque  $\alpha \rightarrow$  au taux de confiance  $1 - \alpha$ , supposant  $H_0$  vraie :



c) On énonce le test :

- $\mu_e \in I_F \Rightarrow$  on accepte  $H_0$ ,
- $\mu_e \notin I_F \Rightarrow$  on rejette  $H_0$ .

## III. 2) Test bilatéral

On teste l'affirmation : dans la population, la moyenne est

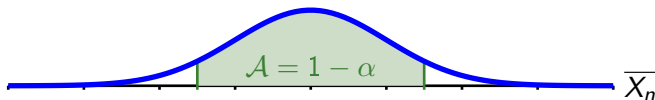
$$\mu = \mu_{\text{AFFIRMATION}}$$

Pour simplifier, dans ce cas, on suppose  $\sigma$  connu.

Au lieu de  $\mu$  ce pourrait être  $p$  pour une fréquence.

a)  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu \neq \mu_{\text{AFFIRMATION}}$

b)  $I_F$  au risque  $\alpha \rightarrow$  au taux de confiance  $1 - \alpha$ , supposant  $H_0$  vraie :



c) On énonce le test :

- $\mu_e \in I_F \Rightarrow$  on accepte  $H_0$ ,
- $\mu_e \notin I_F \Rightarrow$  on rejette  $H_0$ .

d) On réalise le test.

Une usine fabrique des roues dentées dont le diamètre **annoncé** est de 23,65 mm.

Un service qualité prélève un lot de 100 roues et obtient une moyenne  $\bar{\mu}_e = 23,644$  mm et un écart-type  $\sigma_e = 0,018$  mm.

Au risque de 5 %, quelle conclusion doit tirer le service qualité ?

Quand la conclusion est que l'on accepte  $H_0$ , il ne faut pas en déduire que forcément  $\mu = \mu_{AFFIRMATION}$  exactement. Tout ce que l'on peut dire, c'est que  $\mu$  n'est pas assez différent de  $\mu_{AFFIRMATION}$  pour que cela soit sensible. La sensibilité du test dépendra notamment de la taille de l'échantillon pris.

### III. 3) Test unilatéral

Dans le test précédent, on envisageait deux risques : que  $\mu$  soit sensiblement  $>$  ou  $<$  à la valeur  $\mu_{AFFIRMATION}$ .

Maintenant, on ne va envisager qu'un seul des deux risques.

### III. 3) Test unilatéral

Dans le test précédent, on envisageait deux risques : que  $\mu$  soit sensiblement  $>$  ou  $<$  à la valeur  $\mu_{\text{AFFIRMATION}}$ .

Maintenant, on ne va envisager qu'un seul des deux risques.

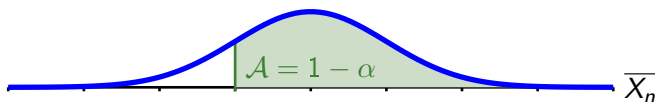
1.  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu < \mu_{\text{AFFIRMATION}}$

### III. 3) Test unilatéral

Dans le test précédent, on envisageait deux risques : que  $\mu$  soit sensiblement  $>$  ou  $<$  à la valeur  $\mu_{\text{AFFIRMATION}}$ .

Maintenant, on ne va envisager qu'un seul des deux risques.

1.  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu < \mu_{\text{AFFIRMATION}}$
2. Intervalle de fluctuation **unilatéral** au risque  $\alpha \rightarrow$  au taux de confiance  $1 - \alpha$ , supposant  $H_0$  vraie :

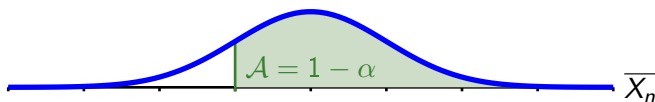


### III. 3) Test unilatéral

Dans le test précédent, on envisageait deux risques : que  $\mu$  soit sensiblement  $>$  ou  $<$  à la valeur  $\mu_{\text{AFFIRMATION}}$ .

Maintenant, on ne va envisager qu'un seul des deux risques.

1.  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu < \mu_{\text{AFFIRMATION}}$
2. Intervalle de fluctuation **unilatéral** au risque  $\alpha \rightarrow$  au taux de confiance  $1 - \alpha$ , supposant  $H_0$  vraie :



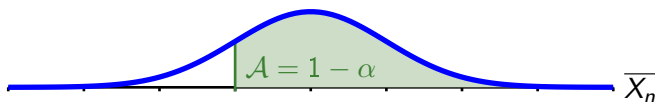
3. On énonce le test :
  - $\mu \in I_F \Rightarrow$  on accepte  $H_0$
  - $\mu \notin I_F \Rightarrow$  on accepte  $H_1$

### III. 3) Test unilatéral

Dans le test précédent, on envisageait deux risques : que  $\mu$  soit sensiblement  $>$  ou  $<$  à la valeur  $\mu_{\text{AFFIRMATION}}$ .

Maintenant, on ne va envisager qu'un seul des deux risques.

1.  $H_0 : \mu = \mu_{\text{AFFIRMATION}}$  et  $H_1 : \mu < \mu_{\text{AFFIRMATION}}$
2. Intervalle de fluctuation **unilatéral** au risque  $\alpha \rightarrow$  au taux de confiance  $1 - \alpha$ , supposant  $H_0$  vraie :



3. On énonce le test :
  - $\mu \in I_F \Rightarrow$  on accepte  $H_0$
  - $\mu \notin I_F \Rightarrow$  on accepte  $H_1$
4. On réalise le test.

## Exercice 9

Une entreprise a constaté qu'un certain nombre de ballons d'eau chaude qu'elle installe ont un défaut de fonctionnement. On s'intéresse à la fréquence de ballons installés ayant un défaut.

Suivant des statistiques plus anciennes, le service qualité annonce que la fréquence de défaut est de 5%. Mais on envisage le risque que cette fréquence soit  $> 5\%$ .

L'entreprise installe 304 ballons et constate que 18 ont un défaut.

Peut-on affirmer, au seuil de risque de 1% que  $p$ , la fréquence de défaut, est toujours de 5% ?

### III. 4) Risque de première espèce

Supposons que suite à un test, on rejette  $H_0$  alors que  $H_0$  est vraie. C'est le risque de première espèce, risque  $\alpha$ .

### III. 4) Risque de première espèce

Supposons que suite à un test, on rejette  $H_0$  alors que  $H_0$  est vraie. C'est le risque de première espèce, risque  $\alpha$ .

**Exemple** :  $p$  vaut vraiment 60 % mais je ne le sais pas. Je décide de prélever un échantillon de  $n = 200$  et je décide d'un test au risque  $\alpha = 5\%$  :

$$I_F \simeq [53,2\% ; 66,8\%]$$

### III. 4) Risque de première espèce

Supposons que suite à un test, on rejette  $H_0$  alors que  $H_0$  est vraie. C'est le risque de première espèce, risque  $\alpha$ .

**Exemple** :  $p$  vaut vraiment 60 % mais je ne le sais pas. Je décide de prélever un échantillon de  $n = 200$  et je décide d'un test au risque  $\alpha = 5\%$  :

$$I_F \simeq [53,2\% ; 66,8\%]$$

Se peut-il que dans un échantillon,  $f \notin I_F$ , me laissant croire que  $p \neq 60\%$  alors que  $p = 60\%$  en réalité ?

### III. 4) Risque de première espèce

Supposons que suite à un test, on rejette  $H_0$  alors que  $H_0$  est vraie. C'est le risque de première espèce, risque  $\alpha$ .

**Exemple** :  $p$  vaut vraiment 60 % mais je ne le sais pas. Je décide de prélever un échantillon de  $n = 200$  et je décide d'un test au risque  $\alpha = 5\%$  :

$$I_F \simeq [53,2\% ; 66,8\%]$$

Se peut-il que dans un échantillon,  $f \notin I_F$ , me laissant croire que  $p \neq 60\%$  alors que  $p = 60\%$  en réalité ?

**OUI!** Si  $p = 60\%$ , la probabilité que  $f \notin I_F$  est précisément  $\alpha = 5\%$ , c'est pour cela que l'on dit « Intervalle de fluctuation au risque  $\alpha$  »

### III. 5) Risque de seconde espèce

Supposons que suite à un test, on accepte  $H_0$  alors que  $H_0$  est fausse. C'est le risque de seconde espèce, risque  $\beta$ .

### III. 5) Risque de seconde espèce

Supposons que suite à un test, on accepte  $H_0$  alors que  $H_0$  est fausse. C'est le risque de seconde espèce, risque  $\beta$ .

**Exemple** :  $p$  ne vaut pas 60 % mais je fais l'hypothèse que c'est le cas. Je décide de prélever un échantillon de  $n = 200$  et je décide d'un test au risque  $\alpha = 5\%$  :

$$I_F \simeq [53,2\% ; 66,8\%]$$

### III. 5) Risque de seconde espèce

Supposons que suite à un test, on accepte  $H_0$  alors que  $H_0$  est fausse. C'est le risque de seconde espèce, risque  $\beta$ .

**Exemple** :  $p$  ne vaut pas 60 % mais je fais l'hypothèse que c'est le cas. Je décide de prélever un échantillon de  $n = 200$  et je décide d'un test au risque  $\alpha = 5\%$  :

$$I_F \simeq [53,2\% ; 66,8\%]$$

Se peut-il que dans un échantillon,  $f \in I_F$ , me laissant croire que  $p = 60\%$  alors que  $p \neq 60\%$  en réalité ?

### III. 5) Risque de seconde espèce

Supposons que suite à un test, on accepte  $H_0$  alors que  $H_0$  est fausse. C'est le risque de seconde espèce, risque  $\beta$ .

**Exemple** :  $p$  ne vaut pas 60 % mais je fais l'hypothèse que c'est le cas. Je décide de prélever un échantillon de  $n = 200$  et je décide d'un test au risque  $\alpha = 5\%$  :

$$I_F \simeq [53,2\% ; 66,8\%]$$

Se peut-il que dans un échantillon,  $f \in I_F$ , me laissant croire que  $p = 60\%$  alors que  $p \neq 60\%$  en réalité ?

**OUI!** Mais le calcul est plus délicat.

Il faut envisager, pour chaque valeur de  $p \neq 60\%$  possible, la probabilité que  $f \in I_F$ .