EXERCICE 4

1) **« Mondialisation : 52% des Français la redoutent »**
Cette information est tirée du journal « LIBERATION » du lundi 12/12/2005 et c'est le résultat d'un sondage réalisé par l'Institut BVA, par téléphone du 2 au 3 décembre, auprès d'un échantillon de 950 personnes de 18 ans et plus.
(8 points) Donner un intervalle d'estimation à 94% du pourcentage de français qui redoutent la mondialisation en détaillant vos calculs (hypothèse, problème, résolution).
Commentez le résultat ? (en français)

P = {Personnes de 18 ans et plus en 2005}

X = Variable aléatoire indicatrice de crainte de la mondialisation

$$X_1 = \begin{cases} 1 \quad si \quad redoute\ mondialisation & p\ inconnu \\ 0 \quad \sin on & q\ inconnu \end{cases}$$

$\mathcal{L}(X_1) = \mathcal{B}(1\,;p)$ Loi de Bernoulli avec p inconnu

p = Proportion de personnes de 18 ans et plus redoutant la mondialisation dans la population française en 2005

<u>Problème :</u> Estimer p

RESOLUTION : On décide d'interroger un échantillon de 950 étudiants : n=950

<u>A priori,</u>

Taux de sondage $= \dfrac{n}{N} = \dfrac{950}{N} < \dfrac{1}{10}$

Taux de sondage **petit,**

donc, $X_1, X_2, \ldots, X_{950}$, sont des variables aléatoires **indépendantes**.

$S_{950} = X_1 + X_2 + \ldots + X_{950}$ = Variable aléatoire. Estimateur du nombre de personnes de 18 ans et plus redoutant la mondialisation dans un échantillon de 950 personnes en 2005.

$F_{950} = \dfrac{S_{950}}{950}$ = Variable aléatoire. C'est l'estimateur de la proportion de personnes de 18 ans et plus redoutant la mondialisation dans un échantillon de 950 personnes en 2005.

= estimateur de p

$\mathcal{L}(F_{950}) = \mathcal{B}(950\,;p)$ avec p inconnu

p et q inconnus mais je connais une estimation de p mesurée dans un échantillon aléatoire que je suppose représentatif de P → f = 0,52

$$\left.\begin{array}{l} nf = 950 \times 0{,}52 = 494 \\ n(1-f) = 950 \times 0{,}48 = 456 \end{array}\right\} \geq 10$$

Dans les deux cas les estimations de n*p et n*q sont très nettement supérieures à 10 donc, on peut appliquer le théorème central limite :

$\mathcal{L}(F_{950}) = \mathcal{N}\left(p, \sqrt{\dfrac{pq}{950}}\right)$

Soit $\alpha = 0{,}94 \Rightarrow t_{0,06} = 1{,}8808$

$I_{94\%} = [\ p - t_{0,06} \times \sqrt{\dfrac{pq}{n}}\ ;\ p + t_{0,06} \times \sqrt{\dfrac{pq}{n}}\ ]$

$I_{94\%} = [\ p - 1{,}8808 \times \sqrt{\dfrac{pq}{950}}\ ;\ p + 1{,}8808 \times \sqrt{\dfrac{pq}{950}}\ ]$

J'ai une probabilité de 94 % que la vraie valeur de la proportion de personnes de 18 ans et plus redoutant la mondialisation dans la population totale en 2005 appartienne à cet intervalle **théorique**.

A posteriori, je tire **un** échantillon de 950 personnes de 18 ans et plus

$f_{950} = 0{,}52$: **une** estimation de p = proportion observée de personnes de 18 ans ou plus redoutant la mondialisation en 2005 dans mon échantillon aléatoire que je suppose représentatif, de taille 950.

$I_{94\%} = [f_{950} - 1{,}8808 \times \sqrt{\dfrac{f(1-f)}{950}}\ ;\ f_{950} + 1{,}8808 \times \sqrt{\dfrac{f(1-f)}{950}}\ ]$

$I_{94\%} = [0{,}52 - 1{,}8808 \times \sqrt{\dfrac{0{,}52 \times 0{,}48}{950}}\ ;\ 0{,}52 + 1{,}8808 \times \sqrt{\dfrac{0{,}52 \times 0{,}48}{950}}\ ]$

$I_{94\%} = [0{,}49\ ;\ 0{,}55]$

On a 94 % de probabilité (que la vraie valeur) qu'il y ait entre 49 et 55 % de Français de 18 ans et plus dans la population française redoutant la mondialisation en 2005. *En supposant que l'échantillon sur lequel repose l'estimation est représentatif.*

(3 points) Sans faire de calcul mais en donnant juste l'intervalle d'estimation à 94% et en commentant le résultat obtenu, donner l'intervalle d'estimation du pourcentage de salariés qui redoutent la mondialisation sachant que dans cette enquête, 55% des 480 salariés (ouvriers, employés, cadres) interrogés redoutent la mondialisation. Comparer les 2 intervalles.

P = {Salariés de 18 ans et plus en 2005}

Variable identique

Problème : Estimer p chez les salariés

RESOLUTION : n=480

A priori,

$\mathcal{L}(F_{480}) = \mathcal{B}(480\ ;\ p)$ avec p inconnu

On peut appliquer le théorème central limite

car $nf = 480 \times 0{,}55 = 264$ et $n(1-f) = 480 \times 0{,}45 = 216$ sont ≥ 10

donc $\mathcal{L}(F_{480}) = \mathcal{N}\left(p, \sqrt{\dfrac{pq}{480}}\right)$

$I_{94\%} = [\, p - 1{,}8808 \times \sqrt{\frac{pq}{480}}\, ; p + 1{,}8808 \times \sqrt{\frac{pq}{480}}\,]$

<u>A posteriori,</u>

$f_{480} = 0{,}55$: **une** estimation de p mesurée dans un échantillon aléatoire de 480 personnes. Cet échantillon est supposé représentatif.

$I_{94\%} = [f_{480} - 1{,}8808 \times \sqrt{\frac{f(1-f)}{480}}\, ; f_{480} + 1{,}8808 \times \sqrt{\frac{f(1-f)}{480}}\,]$

$I_{94} = [0{,}55 - 1{,}8808 \times \sqrt{\frac{(0{,}55 \times 0{,}45)}{480}}\, ; 0{,}55 + 1{,}8808 \times \sqrt{\frac{(0{,}55 \times 0{,}45)}{480}}\,]$

$I_{94} = [0{,}51 \, ; 0{,}59]$

On a 95 % de probabilité que la vraie valeur de la proportion de salariés redoutant la mondialisation dans la population totale en se situe dans cet intervalle entre 51 et 59 % (si l'échantillon retenu pour l'estimation est représentatif) .

COMMENTAIRE :

Les deux intervalles se chevauchent, on ne peut pas dire que les salariés redoutent plus la mondialisation que l'ensemble de la population. On pourrait réaliser une comparaison avec les non-salariés et voir si les 2 intervalles se chevauchent.

EXERCICE 5

Une enquête sur l'équipement informatique des étudiants a été menée en 2003. Environ 9500 étudiants au total étaient inscrits à l'Université Marc BLOCH cette année-là en premier et deuxième cycle.
Nous avons interrogé entre mars et juin 2003 des étudiants de Marc Bloch qui étaient présents dans les cours ou travaux dirigés de premier et deuxième cycle, soit 778 étudiants.
J'ai calculé le nombre d'années de retard (par rapport à l'âge « normal ») des étudiants inscrits en premier et deuxième cycle dans cette enquête. Sur les 778 étudiants concernés j'ai observé un retard moyen de 1,29 an avec un écart-type de 1,72 an.
Donnez l'estimation à 95% du nombre d'années de retard des étudiants dans la population des étudiants de Marc BLOCH. Commentez (en français).

P = {Etudiants de 1er et 2ème cycle de l'UMB}

X = Nombre d'années de retard

$\mathcal{L}(X) = \mathcal{L}(\mu, \sigma)$, Loi quelconque avec μ et σ inconnus

μ = Nombre moyen d'années de retard de l'ensemble des étudiants de 1er et 2e cycle de l'UMB en 2003.

<u>Problème :</u> Estimer μ

RESOLUTION : On décide d'interroger un échantillon de 778 étudiants : n=778

<u>A priori,</u>

Taux de sondage $= \frac{n}{N} = \frac{778}{9500} < \frac{1}{10}$

Taux de sondage **petit,**

donc, $X_1, X_2, \ldots, X_{778}$, sont des variables aléatoires **indépendantes** de loi quelconque $\mathcal{L}(\mu, \sigma)$

$\overline{X} = \frac{\sum X_i}{778}$ = nombre moyen **théorique** d'années de retard de l'ensemble des étudiants de 1[er] et 2e cycle de l'UMB en 2003.

= estimateur de μ

$\mathcal{L}(\overline{X})$ = ? car

- $\mathcal{L}(X)$ n'est pas une loi normale,
- l'écart-type σ de la population totale est inconnu

Mais, $n \geq 30$ (très grand) donc on peut utiliser le théorème central limite pour la normalité,

- avec $\frac{s}{\sqrt{n}}$ considéré comme une bonne estimation de $\frac{\sigma}{\sqrt{n}}$

$$\mathcal{L}(\overline{X}) = \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}}\right) \approx \mathcal{N}\left(\mu ; \frac{s}{\sqrt{n}}\right) = \mathcal{N}\left(\mu ; \frac{1{,}72}{\sqrt{778}}\right) = \mathcal{N}(\mu ; 0{,}06)$$

Soit $\alpha = 0{,}95 \Rightarrow t_{0,05} = 1{,}96$

$I_{95\%} = [\mu - t_{0,05} \times \frac{s}{\sqrt{n}} ; \mu + t_{0,05} \times \frac{s}{\sqrt{n}}]$

$I_{95\%} = [\mu - 1{,}96 \times 0{,}06 ; \mu + 1{,}96 \times 0{,}06]$

$I_{95\%} = [\mu - 0{,}12 ; \mu + 0{,}12]$

J'ai une probabilité de 95 % que la vraie valeur du nombre moyen d'années de retard de l'ensemble des étudiants de 1[er] et 2e cycle de l'UMB en 2003 appartienne à cet intervalle théorique.

A posteriori, je tire **un** échantillon de 778 étudiants

$\overline{x} = 1{,}29$: **une** estimation de μ = nombre moyen d'années de retard des étudiants de 1[er] et 2e cycle de l'UMB en 2003 dans mon échantillon aléatoire, que je suppose représentatif, de taille 778.

$I_{95\%} = [\overline{x} - t_{0,05} \times \frac{s}{\sqrt{n}} ; \overline{x} + t_{0,05} \times \frac{s}{\sqrt{n}}]$

$I_{95\%} = [1{,}29 - 0{,}12 ; 1{,}29 + 0{,}12]$

$I_{95\%} = [1{,}17 ; 1{,}41]$

On a 95 % de probabilité que le nombre moyen d'années de retard pour l'ensemble étudiants de 1[er] et 2e cycle de l'UMB en 2003 soit compris entre 1,17 et 1,41 an. En supposant l'échantillon tiré représentatif.

COMMENTAIRE :

Cette moyenne ne semble pas très élevée. Les étudiants interrogés sont les étudiants assidus présents en cours et en TD. Peut-être y'a-t-il un effet de sélection ?

De plus, il serait intéressant de comparer cette estimation avec le nombre moyen d'années de retard des étudiants des deux autres universités pour savoir s'il existe un effet de filière. En effet le retard peut signifier qu'une part plus importante d'étudiant arrive dans la filière après avoir échoué dans une autre filière ou après avoir redoublé avant d'entrer à l'Université. Le retard ne serait alors pas spécifique à la filière.