Chapitre III Echantillonnage- Estimation.

I. Echantillonnage

1. Distribution d'echantillonnage de la moyenne :

Soit une population de taille N, de moyenne m et d'écart-type σ. On extrait de cette population une serie d'échantillons de taille n. Chacun de ces échantillons a une moyenne $\overline{x_i}$

Les moyenne ainsi obtenues constituent une distribution d'échantillonage de moyenne $\overline{X}$, en fait $\overline{X}$ est une variable aléatoire dont les $\overline{X_i}$ sont des réalisation.

Ainsi on note $\mu_{\overline{X}}$ et $\sigma_{\overline{X}}$ la moyenne et l'écart type de $\overline{X}$.

On pose alors $\mu_{\overline{X}} = m$.

Si le tirage est exhaustif (sans remise) $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} \sqrt{\dfrac{N-n}{N-1}}$.

Si $N = \infty$ ou si le tirage est non-exhaustif (avec remise) on a :

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Si $n << N$ (n est petit (négligeable)) devant N on ne fera pas de distinction entre tirage exhaustif et non exhaustif

Car dans ce cas $\dfrac{N-n}{N-1} \simeq 1$.

- Si la population totale suit une loi normale alors $\overline{X}$ aussi suit une loi normale.

- Si $n \geqslant 30$ alors $\overline{X}$ suit une loi normale quelque soit la loi de la population.

2. Distribution d'échantillonnage des fréquences :

La probabilité de réalisation d'un événement donné dans une population de taille N est p.

Pour chaque échantillon de taille n, on détermine la proportion f_i de réalisation de l'événement donné.

La population et les échantillons suivent les lois binômiales $B(N, p)$ et $B(n, f_i)$ respectivement.

La moyenne μ_f et l'écart-type σ_f de la distribution d'échantillonage des fréquences valent :

- $\mu_f = p$; $\sigma_f = \sqrt{\dfrac{p(1-p)}{n}}$ si le tirage est non exhaustif ou si la population est infinie. et

- $\mu_f = p$ et $\sigma_f = \sqrt{\dfrac{p(1-p)}{n}} \sqrt{\dfrac{N-n}{N-1}}$ si le tirage est exhaustif

- Si $n \geqslant 30$ la distribution d'échantillonnage des fréquence approche la loi normale

3. Distribution d'échantillage de différence de moyennes :

Soit deux populations P_1, P_2 de moyennes respectives m_1, m_2 et d'écarts-type σ_1 et σ_2. On extrait de chacune des deux populations des échantillons de taille n_1 de P_1 et des échantillons de taille n_2 de P_2.

On s'interesse à la distribution de la différence des moyenne des deux populations. :

$$\mu_{\bar{X}_1 - \bar{X}_2} = m_1 - m_2 .$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma^2}{n_2}}$$ (tirage non exhaustif)

-4- Distribution de la variance :

Chaque échantillon de taille n de la population a une variance donnée par la formule $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_i)^2$

alors $$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

On a $$E(S^2) = \frac{n-1}{n} \sigma^2$$

$$V(S^2) = \frac{n-1}{n^3} \left[(n-1)\mu_4 - (n-3)\sigma^4 \right]$$

$$Cov(\bar{X}, S^2) = \frac{n-1}{n^2} \mu_3$$

II. Estimation :

1. Intervalle de confiance de la moyenne :

On veut estimer la moyenne m d'une population à l'aide d'un échantillon aléatoire

(Si la population suit une loi normale, ou si la taille de l'échantillon $n \geq 30$) alors

$\overline{X}$ suit une loi normale $N(m, \sigma_{\overline{X}})$

avec $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$ si le tirage est non exhaustif

$\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$ si le tirage est exhaustif

Donc $\dfrac{\overline{X} - m}{\sigma_{\overline{X}}}$ suit une loi normale $N(0,1)$

$$P\left(\left|\frac{\overline{X}-m}{\sigma_{\overline{X}}}\right| < u_\alpha\right) = \alpha \leftarrow \quad \alpha \text{ représente une probabilité donnée}$$

$$P\left(-u_\alpha < \frac{\overline{X}-m}{\sigma_{\overline{X}}} < u_\alpha\right) = \alpha$$

donc $m \in \left]\overline{X} - u_\alpha \sigma_{\overline{X}}, \overline{X} + u_\alpha \sigma_{\overline{X}}\right[$ qui est appelé intervalle de confiance.

Si par exemple $\alpha = 0,95$ on a : $u_\alpha = 1,\cancel{65} = 1,96$.

$u_\alpha = 1,96$

2. Intervalle de confiance d'une fréquence :

Si $n > 30$, $B(N,p)$ et $B(n,f)$ peuvent être approximées par une loi normale. et dans ce cas.

$$\frac{f-p}{\sigma_f} \text{ elle aussi suit une loi } N(0,1)$$

$$P\left(\left| \frac{f-p}{\sigma_f} \right| < u_\alpha \right) = \alpha \longleftarrow$$ α probabilité donnée

$$P\left(-u_\alpha < \frac{f-p}{\sigma_f} < u_\alpha \right) = \alpha .$$

Si $\sigma_f = \sqrt{\frac{f(1-f)}{n}}$ on a alors

$$p \in \left] f - u_\alpha \sqrt{\frac{f(1-f)}{n}} , f + u_\alpha \sqrt{\frac{f(1-f)}{n}} \right[$$

$(\alpha = 0{,}95 , \quad u_\alpha = 1{,}96)$.

Exercice 1

La moyenne des notes d'une épreuve de mathématiques de 300 étudiants est égale à 9,8. L'écart type est donné par 3,68

Trouver la probabilité qu'un échantillon aléatoire exhaustif de notes de 40 étudiants extrait de l'ensemble ait une moyenne :

1/ Comprise entre 10 et 13

2/ inférieure à 10

même questions si l'échantillons est non exhaustif

Solution I/ Echantillon exhaustif.

$n = 40$, $N = 300$, $m = 9,8$, $\sigma = 3,68$

$n = 40 > 30$ La distribution d'échantillonnage de la moyenne suit une distribution normale. de moyenne $\mu_{\bar{X}} = m = 9,8$

et d'écart type : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{3,68}{\sqrt{40}} \sqrt{\frac{300-40}{300-1}} = 0,543$

1/ $P(10 < \bar{X} < 13) = P\left(\frac{10-\mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{13-\mu_{\bar{X}}}{\sigma_{\bar{X}}}\right)$

$$\frac{10-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{10-9,8}{\frac{3,68}{\sqrt{40}}\sqrt{\frac{300-40}{300-1}}} = 0,37$$

$$\frac{13-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{13-9,8}{\frac{3,68}{\sqrt{40}}\sqrt{\frac{300-40}{300-1}}} = 5,89$$

Donc

$$P(10<\bar{x}<13) = P\left(0{,}37 < \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < 5{,}89\right)$$

$$= F(5{,}89) - F(0{,}37)$$

$$= 1 - 0{,}6443$$

$$= 0{,}3556.$$

2/ $$P(\bar{X}<10) = P\left(\frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < \frac{10-9{,}8}{0{,}543}\right)$$

$$= F(0{,}37)$$

$$= 0{,}6443$$

2/ Echantillon non exhaustif

$\mu_{\bar{x}} = m$ et $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3{,}68}{\sqrt{40}}$.

et on refait la même Procédure on trouve

$$P(10<\bar{x}<13) = 0{,}3668$$

$$P(\bar{x}<10) = 0{,}6331$$

Exercice 2 : La taille moyenne d'un échantillon aléatoire de 40 personnes extrait d'une population de 780 individus est de 1,70 m. L'écart type pour toute la population vaut 24 Cm. Trouver l'intervalle de confiance pour la taille moyenne de la population à 95%.

Solution :

$n = 40 > 30$. $\bar{X}$ suit une loi normale.

$\sigma = 0{,}24$, m est à déterminer dans un intervalle de confiance avec $\alpha = 0{,}95$.

$$\sigma_{\bar{X}} \begin{cases} \frac{\sigma}{\sqrt{n}} \text{ tirage non exhaustif } = \frac{0{,}24}{\sqrt{40}} = 0{,}038 \\ \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \text{ tirage exhaustif } \frac{0{,}24}{\sqrt{40}}\sqrt{\frac{780-40}{780-1}} = 0{,}037 \end{cases}$$

$$P\left(\left|\frac{\bar{X}-\cancel{m}}{\sigma_{\bar{X}}}\right| < u_\alpha\right) = 0{,}95 \Longrightarrow u_\alpha = 1{,}96.$$

$$-1{,}96 < \frac{\bar{X}-\cancel{m}}{0{,}038} < +1{,}96.$$

tirage non exhaustif

$$1{,}70 - 0{,}038 \times 1{,}96 < \cancel{m} < 1{,}70 + 0{,}038 \times 1{,}96$$

$$1{,}62 < m < 1{,}\cancel{70}7$$

tirage exhaustif

$$1{,}70 - 0{,}037 \times 1{,}96 < m < 1{,}70 + 0{,}037 \times 1{,}96$$

$$1{,}66 < m < 1{,}77.$$

Exercice 3 : 500 étudiants se présentent à un examen de mathématiques. Un échantillon aléatoire de 38 notes donne une moyenne égale à 8,65 et un écart type egale à 2,82

Trouver l'intervalle de confiance pour la moyenne des notes des 500 étudiants à :

1/ à 90%   2/ à 95%   3/ à 99%.

Solution : Ici $N=500$, $n=38$ on peut supposer que

$n << N$ et $n > 30$.

On ne connait pas σ alors on considère que 2,82 est une bonne estimation de σ.

et on a : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2,82}{\sqrt{38}} = 0,46$.

$$P\left(\left| \frac{\bar{X}-m}{\sigma_{\bar{X}}} \right| < u_\alpha \right) = 0,9 \longrightarrow u_\alpha = 1,64$$

$$7,9 < m < 9,40.$$

$$P\left(\left| \frac{\bar{X}-m}{\sigma_{\bar{X}}} \right| < u_\alpha \right) = 0,95 \longrightarrow u_\alpha = 1,96.$$

$$7,75 < m < 9,55$$

$$P\left(\left| \frac{\bar{X}-m}{\sigma_{\bar{X}}} \right| < u_\alpha \right) = 0,99 \longrightarrow u_\alpha = 2,58$$

$$7,46 < m < 9,84$$