Chapitre 4 :

Les distributions d'échantillonnage

D'une façon générale, la distribution d'échantillonnage caractérise les fluctuations d'échantillonnage de toute statistique (moyenne, proportion variance) calculée sur tous les échantillons possibles de même taille.

<u>I- Distribution d'échantillonnage d'une moyenne :</u>

Si nous prélevons tous les échantillons possibles de même taille d'une population, la moyenne arithmétique (X) calculée sur chaque échantillon variera d'un échantillon à l'autre. Cependant, ces moyennes auront des valeurs autour d'une valeur centrale qui est celle de la moyenne de la population.

Démonstration

$$\overline{X} = \frac{\sum_{1}^{n} xi}{n}$$

$$E(\overline{X}) = \frac{\sum_{1}^{n} E(xi)}{n} = \frac{\sum_{1}^{n} m}{n} = \frac{n.m}{n} = m$$

$$V(\overline{X}) = \frac{\sum_{1}^{n} V(xi)}{n^2} = \frac{\sum_{1}^{n} \sigma^2}{n^2} = \frac{n.\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Conclusion

Si on prélève un échantillon aléatoire de taille n, d'une population (infinie ou d'une population finis échantillonnage avec remis) dont les éléments possède caractère X qui suit une loi probabilité de moyenne E(X)=m et de variance V(X)= σ^2 alors la moyenne d'échantillon $\overline{X}$ suit une loi de probabilité de moyenne :

E($\overline{X}$)=E(X)=m

Et de variance:

Var ($\overline{X}$)= (σ^2 /n)

Remarque : Lorsque l'échantillonnage s'effectue sans remise à partir d'une population finie de taille N, on doit apporter une correction à Var ($\overline{X}$)

Var ($\overline{X}$) = (σ^2 /n) * [(N-n)/(N-1)] = (σ^2 /n) *[1- ((n-1)/(N-1))] .

(n-1/N-1) ≈ le taux de sondage(n /N), → le facteur de correction peut être ignoré si le taux de sondage est inférieur à 10%.

Théorème central limite :

On sait que X suit une variable aléatoire sans pour autant connaître cette loi de probabilité. Pour identifier cette loi il faut recourir au théorème central limite

Théorème central limite (TCL) :

Si on prélève un échantillon aléatoire de taille n, d'une population (infinie ou d'une population finie et échantillonnage avec remise) dont les éléments possèdent un caractère X qui suit une loi de probabilité de moyenne E(X)=m et de variance V(X)= σ^2 alors la distribution d'échantillonnage de la variable aléatoire $\overline{X}$ tend à se rapprocher d'une loi normale de moyenne E($\overline{X}$)=m et de variance V($\overline{X}$)= σ^2 /n

Ainsi $\overline{X} \longrightarrow N(m, \frac{\sigma}{\sqrt{n}})$

On remarque que l'écart type de $\overline{X}$ est en fonction de l'écart type de la population σ . Un problème peut ainsi se poser lorsqu'on ne connaît pas cet écart type de la population. Dans ce cas il faut estimer cet écart type par celui de l'échantillon (S) :

$$S = \sqrt{\frac{\sum_{1}^{n}(Xi - \overline{X})^2}{n-1}}$$

Cependant, dans ce cas, la loi de probabilité peut changer selon la taille de l'échantillon. Trois cas sont alors possibles :

1er cas :

σ connu alors la loi de probabilité sera :

$$\overline{X} \longrightarrow N(m, \frac{\sigma}{\sqrt{n}}) \text{ d'où } \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \longrightarrow N(0,1)$$

$2^{ème}$ cas :

σ inconnu et n≥30 alors la loi de probabilité sera :

$$\overline{X} \longrightarrow N(m, \frac{S}{\sqrt{n}}) \text{ d'où } \frac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \longrightarrow N(0,1)$$

$3^{ème}$ cas :

σ inconnu et n<30 alors la loi de probabilité sera :

$$\frac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \longrightarrow T(n-1)$$

II-La distribution d'échantillonnage de la différence de deux moyennes :

La distribution d'échantillonnage de la différence de deux moyennes d'un caractère X mesuré sur deux échantillons extraits de deux populations différentes en supposant que les écarts type de ce caractère au niveau de chaque population sont connus.

Si on désigne par m_1 et m_2 les moyennes de cette variable X sur deux populations 1 et 2, et par σ_1 et σ_2 ses écarts types, et si on tire un échantillon d'une taille n_1 de la première population et un échantillon d'une taille n_2 de la deuxième population et on calcule les moyennes empiriques de ces deux échantillon ($\overline{X}_1$ et $\overline{X}_2$), on aura alors :

$$\overline{X}_1 \longrightarrow N(m_1, \frac{\sigma_1}{\sqrt{n_1}})$$

$$\overline{X}_2 \longrightarrow N(m_2, \frac{\sigma_2}{\sqrt{n_2}})$$

Puisque la différence de deux lois normales est une loi normale, la statistique X = $\overline{X}_1$ - $\overline{X}_2$ est une loi normale de moyenne m = $m_1 - m_2$ et d'écart type

$$\sigma = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{d'où :}$$

$$\frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

III- Distribution d'échantillonnage d'une proportion :

On considère une population caractérisée par une proportion p d'individus possédant un certain caractérise qualitatif. Si on prélève au hasard différents échantillons de cette population on remarque que la proportion f d'individus possédant ce caractère qualitatif varie d'un échantillon à l'autre, on peut donc lui attribuer une variable aléatoire.

On démontre que pour une taille d'échantillon n $\geq$30 :

$$f \to N(p, \sqrt{\frac{pq}{n}}) \ \ d'où \ \ \frac{f-p}{\sqrt{\frac{pq}{n}}} \to N(0,1)$$

IV-La distribution d'échantillonnage de la différence de deux proportions :

La différence de deux proportions ($f_1 - f_2$) relatives à un même caractère mesuré sur deux échantillons n_1 et n_2 ayant chacun une taille supérieure à 30 et extraits de deux populations différentes suit une loi normale d'une moyenne égale à la différence des deux proportions réelles de ce caractère dans les deux population $p_1 - p_2$ et d'écart type

$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ d'où la statistique :

$$\frac{(f_1 - f_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \to N(0,1)$$