Statistique inférentielle

1 Echantillonage

La théorie de l'échantillonnage consiste à déterminer des propriétés sur des échantillons prélevés dans une population dont on connaît déjà des propriétés.
On ne considère ici que des échantillons aléatoires, c'est à dire constitués d'éléments pris au hasard dans une population.
Le tirage des éléments d'un échantillon peut être fait sans remise ; On dit qu'il est exhaustif. Sinon si le tirage est fait avec remise, on dit qu'il est non exhaustif ; dans ce cas les tirages sont indépendants.
Dans la plupart des cas, la population ayant un grand effectif, dans laquelle on tire une faible proportion d'éléments, on assimile un tirage sans remise à un tirage avec remise.

Propriété 1 :

Considérons une population ayant une certaine propriété avec une moyenne m et un écart-type σ.
Soit $\bar{X}$ la variable aléatoire qui à tout échantillon aléatoire prélevé avec remise et d'effectif n fixé, associe la moyenne de cet échantillon. Pour n suffisamment grand, $\bar{X}$ suit approximativement la loi normale $\mathcal{N}\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$.

Remarque

- n suffisamment grand quand $n \geqslant 30$.
- Si la population est elle-même normale, on peut utiliser ce résultat même si n est petit.
- Lorsque les échantillons de taille n sont prélevés sans remise dans une population d'effectif N, on peut utiliser le résultat précédent en prenant $\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-m}{N-1}}$ au lieu de $\dfrac{\sigma}{\sqrt{n}}$.
- il ne faut pas confondre l'écart type $\dfrac{\sigma}{\sqrt{n}}$ de la variable aléatoire qui prend pour valeurs les moyennes d'échantillons de taille n, et l'écart type d'un échantillon.

Propriété 2 :

Considérons une population dont un pourcentage p d'éléments possède une certaine propriété.
Soit F la variable aléatoire, qui à tout échantillon aléatoire prélevé avec remise d'effectif n fixé, associe le pourcentage d'éléments de cet échantillon possédant cette propriété. Pour n suffisamment grand, F suit approximativement la loi normale $\mathcal{N}\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$.

2 Estimation.

On considère une population-mère d'effectif N , dont les paramètres inconnus pour la variable X sont la moyenne M et l'écart-type σ.
On note, pour un échantillon donné, $\bar{x}_e$ sa moyenne, et σ_e son écart type.

Propriété 3 :

- on prend comme estimation de la moyenne : $m = \bar{x}_e$.
- on prend comme estimation de l'écart-type : $s = \sqrt{\frac{n}{n-1}}\sigma_e$.

De même on choisit la proportion f_e des éléments possédant une certaine propriété dans un échantillon prélevé aléatoirement dans une population comme meilleur estimation ponctuelle de la proportion inconnue p des éléments de cette population ayant cette propriété.

Propriété 4 :

Estimation par intervalle de confiance, pour une moyenne.
L'intervalle de confiance de M au seuil de risque de α est :

$$m - t_\alpha \frac{s}{\sqrt{n}} \leqslant M \leqslant m + t_\alpha \frac{s}{\sqrt{n}}$$

ou encore

$$m - t_\alpha \frac{\sigma_e}{\sqrt{n-1}} \leqslant M \leqslant m + t_\alpha \frac{\sigma_e}{\sqrt{n-1}}$$

Avec, comme valeur usuelles pour le seuil du risque :

- Si $\alpha = 10\%$, $t_\alpha = 1,645$.
- Si $\alpha = 5\%$, $t_\alpha = 1,96$.
- Si $\alpha = 1\%$, $t_\alpha = 2,58$.

Propriété 5 :

Estimation par intervalle de confiance, pour une proportion.
L'intervalle de confiance de p au seuil de risque de α est :

$$f_e - t_\alpha \frac{\sqrt{f_e(1-f_e)}}{\sqrt{n-1}} \leqslant p \leqslant f_e + t_\alpha \frac{\sqrt{f_e(1-f_e)}}{\sqrt{n-1}}$$

Avec, comme valeur usuelles pour le seuil du risque :

- Si $\alpha = 10\%$, $t_\alpha = 1,645$.
- Si $\alpha = 5\%$, $t_\alpha = 1,96$.
- Si $\alpha = 1\%$, $t_\alpha = 2,58$.

3 Tests statistiques

Définition 1 :

Un test statistique est une méthode permettant de prendre une décision à partir d'informations fournies par un échantillon.
Cette décision dépend donc de l'échantillon. Ainsi qu'elle que soit la décision prise, on court deux sortes de risques :

- Le risque dit de 1 ère espèce noté α , est la probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie en réalité : $\alpha = p(rejeter H0/H0 vraie)$.
- le risque dit de 2 nde espèce noté β, est la probabilité d'accepter l'hypothèse H_0 alors qu'elle est fausse en réalité : $\beta = p(accepter H0/H0 fausse)$.

Un test est bon si on arrive à minimiser α et β.