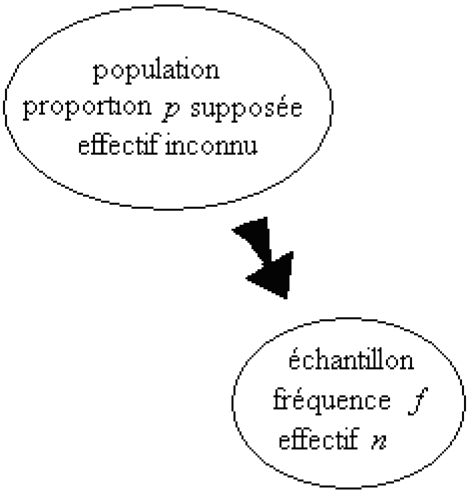
Echantillonnage

Echantillon : Lorsque dans une population, on prélève au hasard, successivement et avec remise *n* individus, on obtient un échantillon de taille *n*.

Nature du problème

Il s'agit de vérifier une information donnée sur une population d'effectif important à partir de l'étude d'un échantillon de quelques dizaines d'unités. Ce type de situation se rencontre fréquemment dans le monde industriel, car le plus souvent il n'est pas possible d'étudier la population entière.

Considérons une population où l'on suppose qu'une proportion *p* d'individus présente un certain caractère.
La méthode consiste à observer sur un échantillon de taille *n* issu de cette population la fréquence *f* d'individus présentant le caractère donné.



On considère que l'échantillon prélevé est représentatif de la population.
Dans ce cas si la proportion observée dans l'échantillon est trop éloignée de l'hypothèse, on considère que l'hypothèse de départ n'est pas correcte et on la rejette. Si la proportion observée dans l'échantillon est proche de l'hypothèse on ne peut pas la rejeter. Le but est donc de définir les valeurs limites permettant de prendre la décision.

Intervalle de fluctuation à 95%

Lorsqu'on observe un individu de l'échantillon, soit cet individu présente le caractère recherché (succès) soit il ne le présente pas (échec). L'observation d'un individu est donc une épreuve de Bernoulli.
L'hypothèse de départ étant à priori considérée comme bonne, la probabilité du succès correspond à la proportion p d'individus présentant le caractère dans la population.
On répète cette épreuve n fois. Les épreuves sont identiques et indépendantes. On est donc dans le cas d'un schéma de Bernoulli et la loi de probabilité est donc la loi binomiale $\boldsymbol{B}(n,p)$

L'intervalle de fluctuation à 95% de la fréquence observée f est l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ où :

n est l'effectif de l'échantillon
a est le plus petit entier tel que $p(X \leq a) > 2{,}5\%$
b est le plus petit entier tel que $p(X \leq b) \geq 97{,}5\%$

On a donc $p(a \leq X \leq b) \geq 95\%$ d'où le nom d'intervalle de fluctuation à 95%

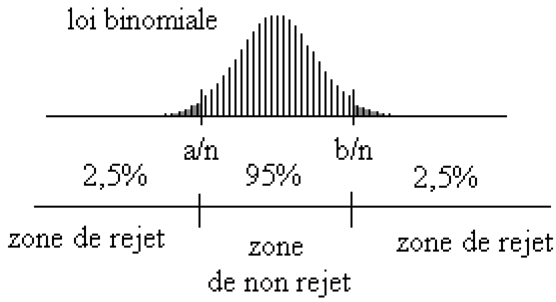
Cela signifie que si l'hypothèse de départ est correcte, on a une probabilité de 95% que la proportion f d'individus présentant le caractère dans l'échantillon tiré au hasard soit dans l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$.

Règle de décision :

Si $f \in \left[\frac{a}{n};\frac{b}{n}\right]$, on ne peut pas rejeter l'hypothèse de départ selon laquelle la proportion d'individus présentant le caractère donné dans la population est p.

Si $f \notin \left[\frac{a}{n};\frac{b}{n}\right]$, on rejette l'hypothèse de départ selon laquelle la proportion d'individus présentant le caractère donné dans la population est p.

On dit qu'on rejette l'hypothèse au risque de 5% car il se pourrait que l'échantillon ne soit pas représentatif et que l'hypothèse soit en fait bonne. On prend donc le risque de 5% de se tromper.



Remarque :

Lorsque $n > 25$ (n étant l'effectif de l'échantillon) et $0{,}2 < p < 0{,}8$ (p étant la proportion, dans la population, du caractère étudié) l'intervalle de fluctuation est proche de $\left[p-\sqrt{\frac{1}{n}};p+\sqrt{\frac{1}{n}}\right]$

p.329 : 6, 7, 8, p.332 : 13, 32 DM : 14 p.332	p.327, p.329, 5, p.334, p.335