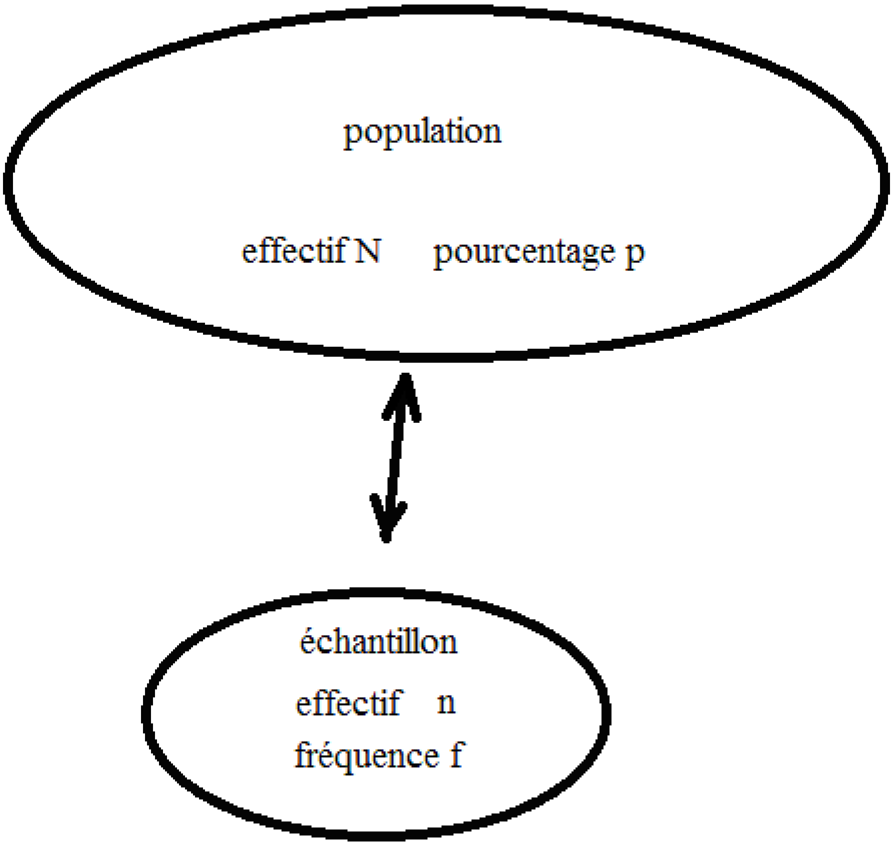
ECHANTILLONNAGE, ESTIMATION ET PRISE DE DECISION

Echantillonnage et estimation

La théorie de l'échantillonnage consiste, connaissant des propriétés d'une population, à déterminer des propriétés d'échantillons prélevés dans la population.
En réalité, on est plus souvent confronté au problème inverse, celui de l'estimation: déduire des informations sur la population totale à partir de renseignements donnés par des échantillons.
Le tirage des éléments d'un échantillon aléatoire peut être exhaustif (sans remise); dans ce cas la composition de l'urne est modifiée à chaque tirage: les tirages ne sont donc pas indépendants.
Sinon le tirage est non exhaustif (avec remise); dans ce cas la composition de l'urne n'est pas modifiée à chaque tirage: les tirages sont donc indépendants.
Dans la plupart des cas où la population a un grand effectif N dont on tire un petit nombre d'éléments *n* on assimile un tirage sans remise à un tirage avec remise.



Intervalle de fluctuation asymptotique d'une fréquence, au seuil de 95%, avec une loi normale

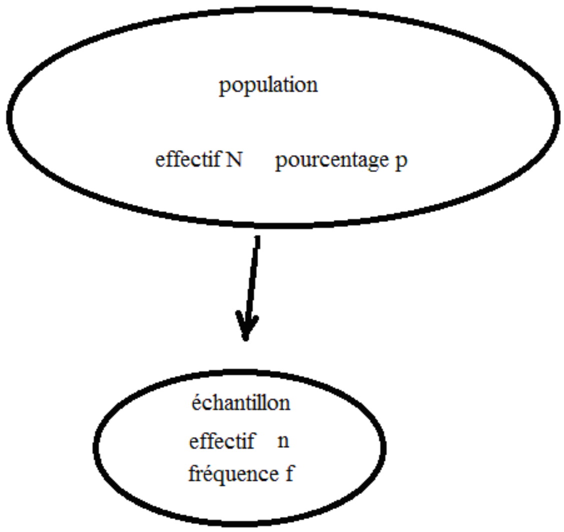
Considérons une population d'effectif N dont un pourcentage p d'éléments possède un certain caractère.
On prélève au hasard des échantillons de taille n et on mesure pour chacun la fréquence f des éléments possédant ce caractère.
E_1 : effectif n , fréquence f_1 ... E_k : effectif n , fréquence f_k
Soit F la variable aléatoire qui à tout échantillon de taille n associe la fréquence du caractère observé.

Lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$: $p\left(F \in \left[p - 1{,}96\sqrt{\dfrac{p(1-p)}{n}}\,;\, p + 1{,}96\sqrt{\dfrac{p(1-p)}{n}}\right]\right) = 0{,}95$

L'intervalle $\left[p - 1{,}96\sqrt{\dfrac{p(1-p)}{n}}\,;\, p + 1{,}96\sqrt{\dfrac{p(1-p)}{n}}\right]$ est appelé intervalle de fluctuation asymptotique de F au seuil de 95%.
Autrement dit, lorsqu'on prélève un échantillon de taille n en respectant les conditions données il y a 95% de chances que la fréquence dans cet échantillon soit dans l'intervalle :

$$\left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}}\,;\, p + 1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$$

Parmi les échantillons de taille n , 95% ont une fréquence compris dans l'intervalle donné



ECHANTILLONNAGE, ESTIMATION ET PRISE DE DECISION

Intervalle de fluctuation d’une fréquence, au seuil de 95%, avec une loi binomiale

Si les conditions ne sont pas respectées, l’intervalle de fluctuation asymptotique à environ 95% d’une fréquence correspondant à la réalisation sur un échantillon aléatoire de taille *n*, d’une variable aléatoire X de loi binomiale, est l’intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ défini par :

a est le plus petit entier naturel tel que $p(X \leq a) > 0{,}025$
b est le plus petit entier naturel tel que $p(X \leq b) \geq 0{,}975$

p.300 : 54, 57, 58	p.299 : 52, 55

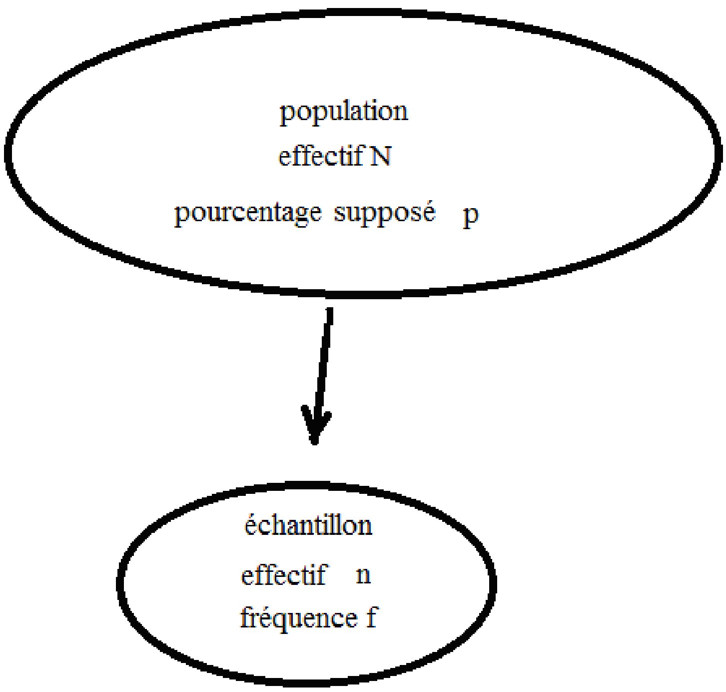
Prise de décision

Soit *p* la proportion supposée du caractère dans la population et *f* la fréquence connue du caractère dans l’échantillon de taille *n*.

Si *f* appartient pas à l’intervalle de fluctuation asymptotique $\left[p-1{,}96\sqrt{\frac{p(1-p)}{n}}\,;\,p+1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$, on ne rejette pas, au seuil de 5%, l’hypothèse selon laquelle la proportion dans la population est *p*. Dans le cas contraire, on la rejette.

On peut également utiliser l’intervalle de fluctuation issu de la loi binomiale : $\left[\frac{a}{n};\frac{b}{n}\right]$ défini par :

a est le plus petit entier naturel tel que $p(X \leq a) > 0{,}025$
b est le plus petit entier naturel tel que $p(X \leq b) \geq 0{,}975$



p.303 : 60, 62	p.271 : ER5, p.302 : 59, 61

Intervalle de confiance d'une proportion .

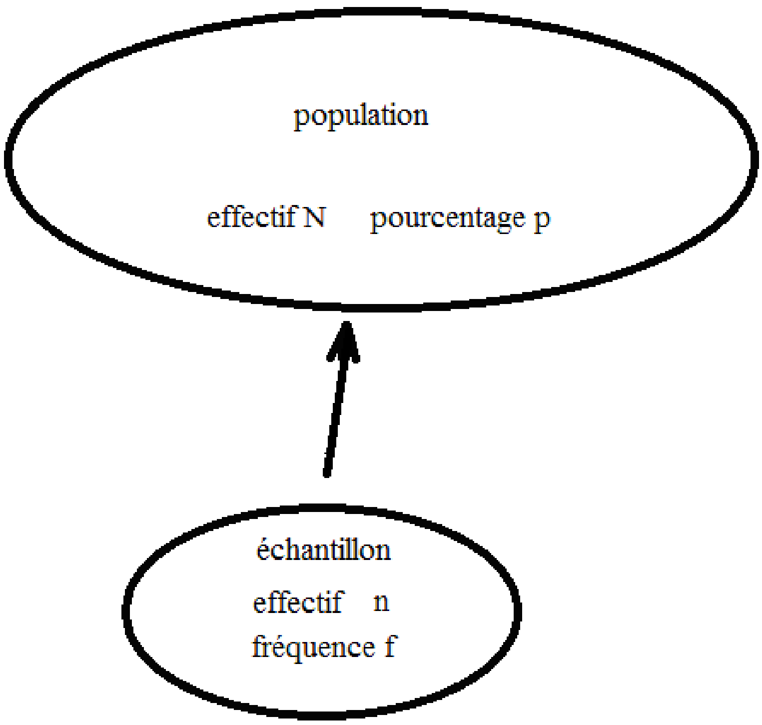
A l'aide d'un échantillon, nous allons définir, avec un coefficient de confiance choisi à l'avance, un intervalle de confiance de la proportion p des éléments de la population possédant un certain caractère.

Lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$: $p\left(p \in \left[f - 1{,}96\sqrt{\dfrac{f(1-f)}{n}} \, ; \, f + 1{,}96\sqrt{\dfrac{f(1-f)}{n}}\right]\right) = 0{,}95$

L'intervalle $\left[f - 1{,}96\sqrt{\dfrac{f(1-f)}{n}} \, ; \, f + 1{,}96\sqrt{\dfrac{f(1-f)}{n}}\right]$ est appelé intervalle de confiance de p au niveau de confiance de 95%.

Autrement dit, lorsqu'on prélève un échantillon de taille n en respectant les conditions données il y a 95% de chances que la proportion dans la population soit dans l'intervalle :

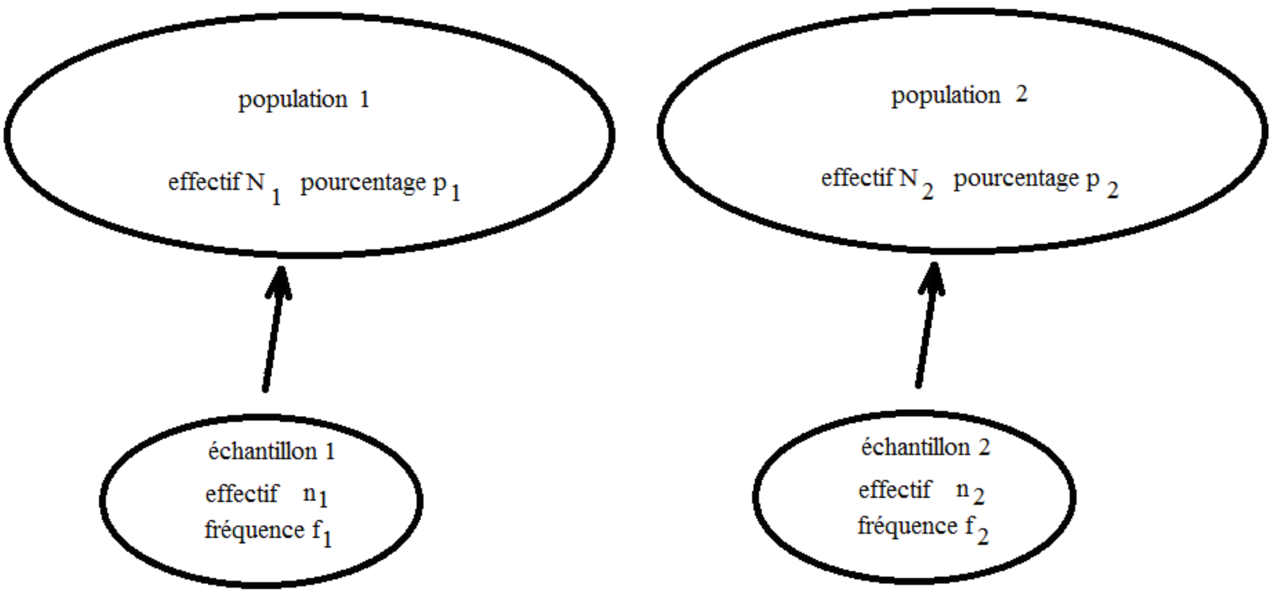
$$\left[f - 1{,}96\sqrt{\frac{f(1-f)}{n}} \, ; \, f + 1{,}96\sqrt{\frac{f(1-f)}{n}}\right]$$



p.305 : 65, 66, 67	p.304 : 64, 68

Comparaison de deux populations.

Considérons un caractère observé sur deux populations différentes P_1 et P_2. Les proportions (inconnues) dans ces deux populations sont appelées p_1 et p_2.
Considérons un échantillon de fréquence f_1 provenant de P_1 et un échantillon de fréquence f_2 provenant de P_2.
Pour savoir si les proportions p_1 et p_2 sont significativement différentes au seuil de 95%, on détermine les intervalles de confiance de chacun.
Si les deux intervalles à 95% sont disjoints, on considère (au risque de 5%) que les proportions p_1 et p_2 sont différentes.
Dans le cas contraires, on considère (au risque de 5%) que les proportions p_1 et p_2 sont égales.



p.305 : 69, 70	p.275 : ER6, p.305 : 68

Devoir Maison

p.306 : 71	