La méthode CMR
(Capture Marquage Recapture)

I Le principe

Dans une **population de N individus**,

▪ on capture et on **Marque M individus**.

▪ On obtient ainsi une proportion $\dfrac{M}{N}$.

▪ Puis on **Capture c individus** et on compte le nombres d'individus déjà marqués, c'est à dire ceux que l'on a **Recapturés** : on note ***r* ce nombre**).

▪ On obtient alors une nouvelle proportion : $\dfrac{r}{c}$

▫ On suppose que le milieu est clos : Pas de départ, pas d'arrivée, pas de naissance ni de mort d'individus

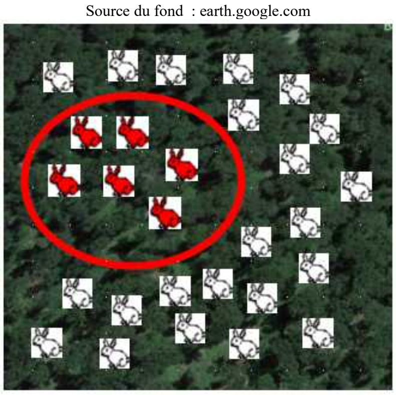
▫ On suppose également que toutes les captures sont indépendantes : Chaque individus a les mêmes chances d'être capturé et le fait de capturer un individu n'influence pas les chances d'en capturer un autre ou de le recapturer.

Dans ces conditons, on sait que $\dfrac{M}{N} = \dfrac{r}{c}$ et donc

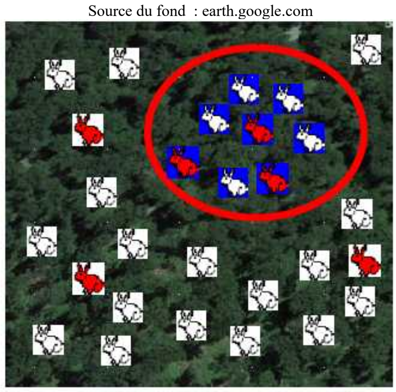
$\boxed{N = \dfrac{c \times M}{r}}$ N est appellé : indice de Lincoln Petersen

Exemple n°1. ***On souhaite connaître le nombre de lapins dans une parcelle de forêt.***

Source du fond : earth.google.com



Source du fond : earth.google.com



$M =$ 

$c =$  +  et $r =$ 

$$N = \frac{c \times M}{r}$$

Remarque n°1.

Si vous vous amusez à compter tous les lapins de l'exemple, vous constaterez que « ça ne marche pas» … Bien oui, c'est un schéma et pas la réalité.

Sur ce schéma, la deuxième condition n'est pas remplie : notre deuxième échantillon n'est pas représentatif (la population étudiée est trop petite pour que la méthode soit pertinente).

II Une question de confiance

N'oublions pas que nous sommes dans le domaine des statistiques :

En effectuant une capture on prélève un échantillon et cet échantillon ne répresente peut-être pas correctement la population.

On pourrait utiliser un grand nombre d'échantillons afin de pouvoir faire une moyenne mais dans la pratique cela coute cher...

Pour compenser cela, on utilise les intervalles de confiance …

Connaissance n°1 Déterminer un Intervalle de Confiance : IC

Pour un échantillon de taille n, on note f_{obs} la fréquence observée (pour nous c'est $\frac{r}{c}$).

On a alors :

$$\boxed{IC = \left[f_{obs} - \epsilon \ ; \ f_{obs} + \epsilon\right]}$$

avec $\epsilon = k\sqrt{\dfrac{f_{obs}(1-f_{obs})}{n}}$ et $k = 1{,}96$ pour un niveau de 95%
$k = 2{,}58$ pour un niveau de 99%

Remarque n°2.

ϵ représente la marge d'erreur et elle dépend de fortement de n :
plus n est grand plus ϵ est petit.

Exemple n°2. On souhaite estimer le nombre de lapins dans la forêt entière.

On capture 800 individus, on les marque puis on les relâche.
On procède à une seconde capture de 1000 individus. On compte alors 250 individus recapturés.

▪ En notant N le nombre de lapins total, on a $N = \dfrac{800 \times 1000}{250} = 3200$
On peut estimer à 3200 le nombre total de lapins.

▪ Ok mais quelle confiance peut accorder à ce résultat ?

Supponsons que nous voulions un niveau de confiance de 95 %.
On a $f_{obs} = \dfrac{250}{1000} = 0{,}25$

$\epsilon = 1{,}96 \times \sqrt{\dfrac{0{,}25(1-0{,}25)}{1000}} \approx 0{,}0268$ (1,96 car 95 %….)

▫ Nous avons ici une marge d'erreur d'environ 2,68 %

On obtient : $IC \approx [0{,}25 - 0{,}0268 \ ; \ 0{,}25 + 0{,}0268] = [0{,}2235 \ ; \ 0{,}2768]$

On en déduit que le nombre de lapins est compris entre $\dfrac{800}{0{,}2768} \approx 2890$ et $\dfrac{800}{0{,}2235} \approx 3780$ avec un niveau de confiance de 95 %.

Remarque n°3.

Pour comparaison, la marge d'erreur pour l'exemple est d'environ 18 %, c'est beaucoup...