LOI BINOMIALE

Objectifs

1-Comprendre la transformation d'une variable discrète binaire en une variable quantitative (numérique) de Bernoulli

2-Calculer les paramètres d'une variable de Bernoulli

3-Calculer la probabilité de constater une fréquence observée par rapport à une fréquence attendue avec la loi binomiale,

4-Exprimer une situation épidémiologique en paramètres de la loi binomiale.

Plan :

I-Variable de Bernoulli :

1.1. Transformation de la variable qualitative binaire en variable de Bernoulli.

1.2. Paramètres de la variable de Bernoulli : Moyenne, variance, écart type.

II- Loi binomiale :

2.1. De la loi de Bernoulli à la loi binomiale

2.2. Loi Binomiale

2.3.Loi Binomiale cumulée

2.4. Conditions d'application de la loi binomiale

Conclusion

Références bibliographiques

I.VARIABLE DE BERNOULLI

Une variable qualitative binaire ou dichotomique est un cas particulier de la variable nominale à 2 classes.

Exemples : Sexe (Homme/femme), Etat de santé (Malade/sain), survie (Vivant/décédé), Tabagisme (Fumeur/non-fumeur), ….

Considérons une expérience et soit A un évènement défini sur cette expérience.

Tout résultat de l'expérience est une réalisation de A ou de son contraire.

Lorsqu'on s'intéresse ainsi à l'évènement **A** ou à son contraire **Ā** ; la réalisation de l'expérience est appelée *épreuve de Bernoulli*

1.1. Transformation d'une variable qualitative binaire en variable quantitative de Bernoulli

Exemple : Répartition des élèves d'une école selon leur statut de vaccinal

vaccinés

Non	0
Non	0
Oui	1
Oui	1
Oui	1
Oui	1
Non	0
Non	0
Non	0
Oui	1
Oui	1
Oui	1
Oui	1
Oui	1
Non	0
Non	0
Oui	1
Non	0
Oui	1
Oui	1
Non	0
Oui	1
Oui	1
Oui	1
Oui	1
Non	0
Oui	1
Non	0
Oui	1
Oui	1

Dans notre exemple parmi les 30 sujets, on peut utiliser les modalités de la variable statut vaccinal Oui/Non, comme on peut utiliser les modalités 1/0

Vacciné= Oui =1(19 vaccinés))

Vacciné = Non = 0 (11 Non vaccinés)

Nous transformons la variable qualitative en variable quantitative (numérique) à 02 modalités 1 et 0.

Oui =1 - Non = 0 :

C'est la variable de BERNOULLI.

D'où il est possible maintenant de calculer les paramètres : la moyenne et la variance.

1.2. Paramètres de la variable de Bernoulli :

1.2.1. Moyenne d'une variable de Bernoulli

$m = \frac{\sum x}{N}$, $\sum$ x = 19, N = 30, $m = \frac{19}{30}$ = 0,633

1.2.2. Variance d'une variable de Bernoulli

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$$

$$or\ x = 1\ ou\ x = 0\ , donc\ \sum x^2 = \sum x$$

$$\sigma^2 = \frac{\sum x}{N} - \left(\sum \frac{x}{N}\right)^2 = \frac{\sum x}{N}\left(1 - \frac{\sum x}{N}\right)$$

D'où la formule de variance de la variable de Bernoulli s'écrit :

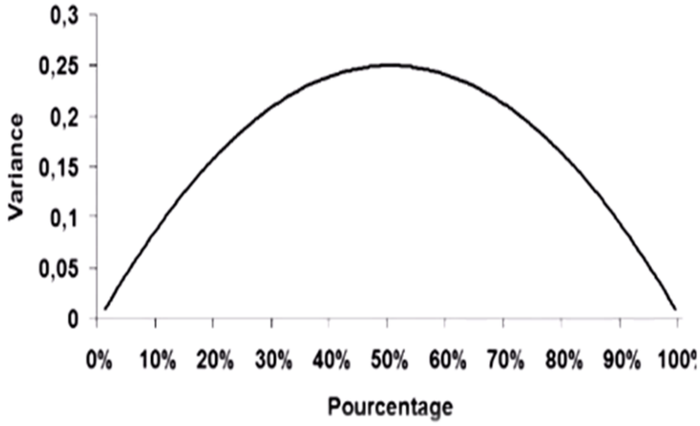
$$\boldsymbol{\sigma^2 = p(1 - p)}$$

$$\boldsymbol{\sigma^2 = \frac{\sum x^2 - \left(\frac{\sum x}{N}\right)^2}{N} = \frac{19 - \frac{19^2}{30}}{30}}$$

Dans notre exemple :

$$\sigma^2 = p(1 - p) = 0{,}633(1 - 0{,}633)\ ,\ \ \sigma^2 = 0{,}23$$

1.2.3. Variance d'une variable de Bernoulli en fonction du pourcentage



Pour un % de 0,5, la variance 0,5 * 0,5 = 0,25, est au maximum**.**
Quel que soit les autres % qu'on étudie, le couple p (1-p) sera toujours < 0,25.
Plus le % est faible, ou bien ; plus le % est grand, la variance sera faible

1.2.4. Ecart type d'une variable de Bernoulli

$$\boldsymbol{\sigma = \sqrt{p(1-p)}}$$

D'où on peut associer à une épreuve de Bernoulli, une variable X telle que
Pr (X) =1 quandA s'est réalisée

Pr (X) =0 quand Ā s'est réalisée

La loi de X est donc définie par :

Pr (X =1) = p

Pr (X=0) = 1-p

X est appelée variable de Bernoulli et sa loi de probabilité loi de Bernoulli avec

E(X) =p

V(X)= p*(1-p)

II.LOI BINOMIALE

Elle permet de calculer la probabilité d'observer un nombre donné d'évènements de type binaire (Succès/échec**, Malade/**Non Malade**, Vacciné/**Non vacciné **……).**
Exemples : un nombre de succès dans une série de tentatives
un nombre de malades parmi une série d'individus

2.1. De la variable de Bernoulli à la variable binomiale

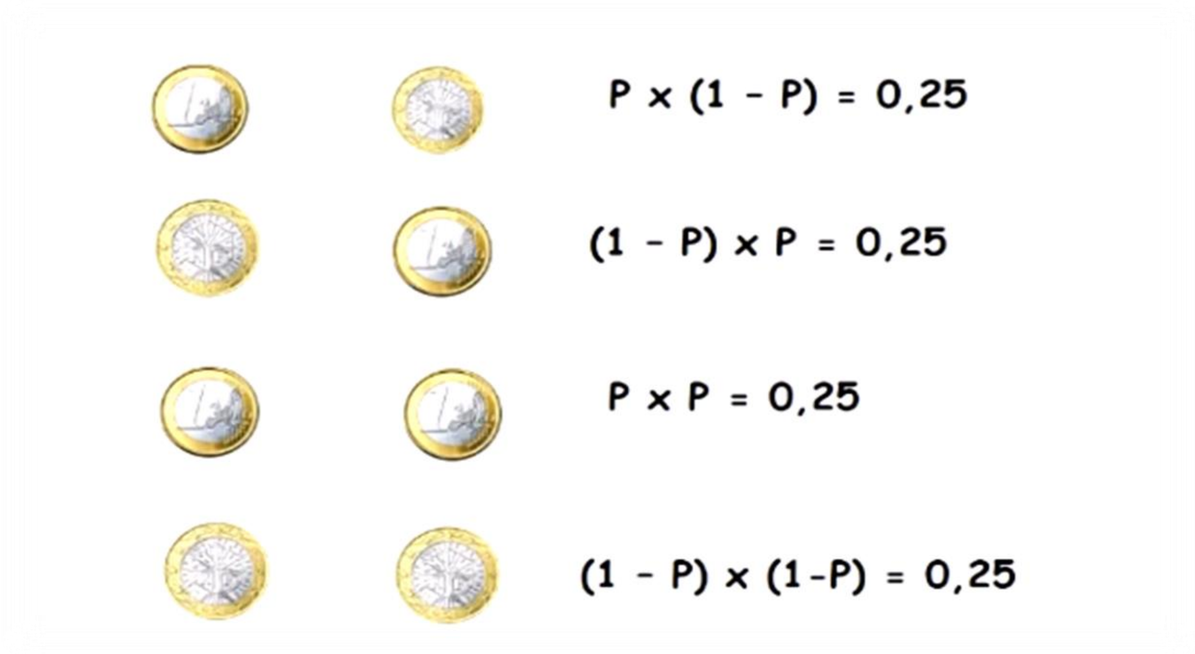
2.1.1. Raisonnement général et situation de jeu du hasard :

<u>Soit une pièce de monnaie</u>

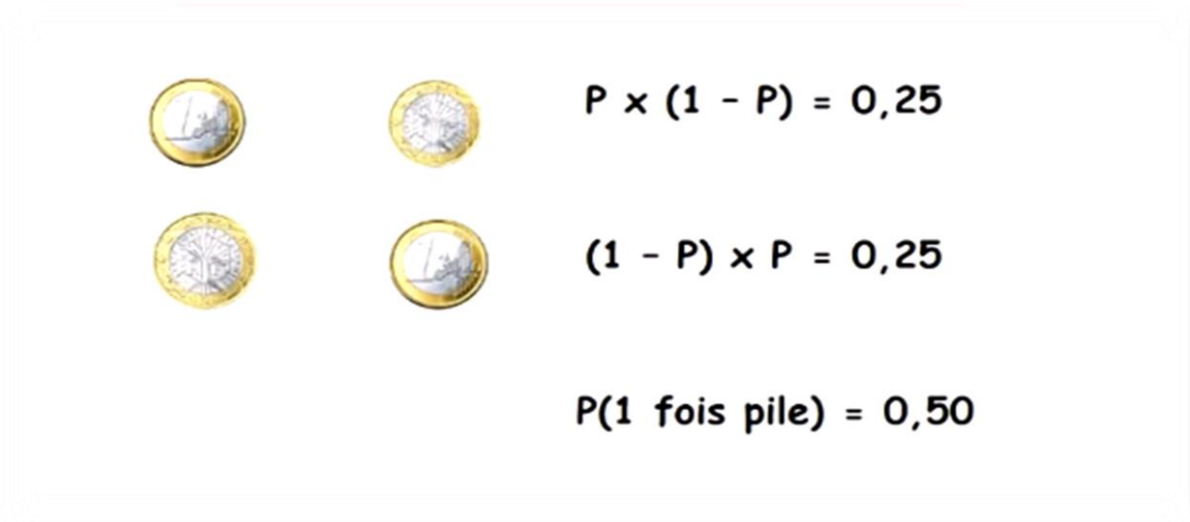
Si on jette la pièce **1 fois**, on pourrait tomber sur les 2 possibilités suivantes : soit**Pile** avec **P=50%**, soit **Face** avec **1-P= 50%**

<u>Nombre de jets : n =2</u>

Si on jette la pièce 2 fois, on pourrait tomber sur les possibilités suivantes :



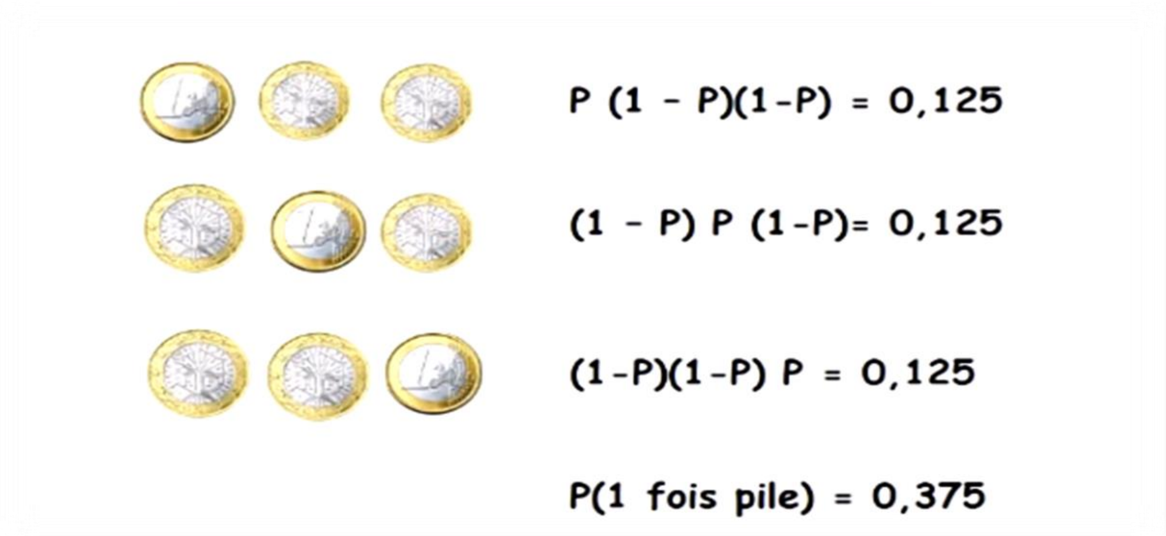
Maintenant, *si on s'intéresse* seulement à la **probabilité** d'avoir **une fois Pile**, quand <u>on jette la pièce 2 fois de suite</u>, on aura les 2 situations suivantes :



On a donc : **La probabilité** d'obtenir une fois Pile lorsqu'on jette la pièce 2 fois de suite est **: 0,25 + 0,25 = 0,5**

Nombre de jets : n =3

Si l'on intéresse à la probabilité d'avoir**une fois Pile,** quand on jette la pièce **3fois**, on aura les 3 possibilités suivantes :



*On a donc :***La probabilité**d'obtenir une**fois Pile** lorsqu'on jette la pièce **3** fois de suite est : **0,125 + 0,125 + 0,125 = 0,375**

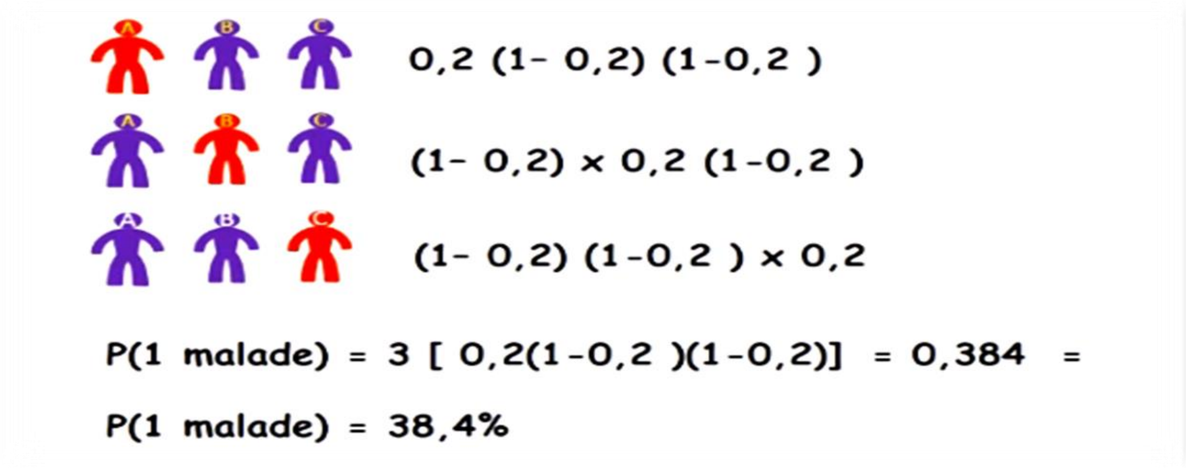
2.1.2.Variable binaire de type épidémiologique :

Exemple : Soit dans une population donnée, **la probabilité d'être malade** d'une maladie donnée est **P=0,2** et la probabilité d'un sujet d'être **sain est 1-P = 0,8**

Dans cet exemple, on va s'intéresser à la probabilité de trouver **<u>1 sujet</u>** malade dans un échantillon **<u>de 3 sujets</u>**, sachant que la probabilité de la maladie dans la population est de **0,2 ou 20%**

<u>Nombre de jets = Taille de l'échantillon</u>

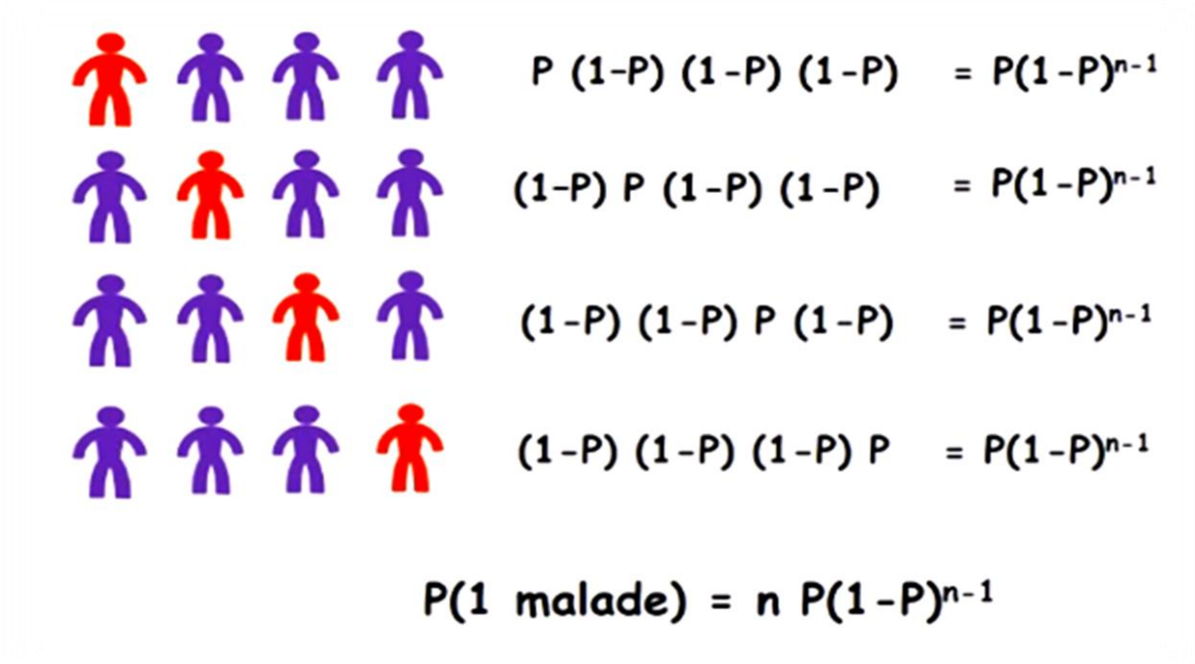
n= 3, Pile =malade



Il y a 38,4 chances sur 100 de trouver un sujet malade dans un échantillon de **3 sujets**, sachant que la probabilité générale de la maladie est de **20%**.

Si on généralise :

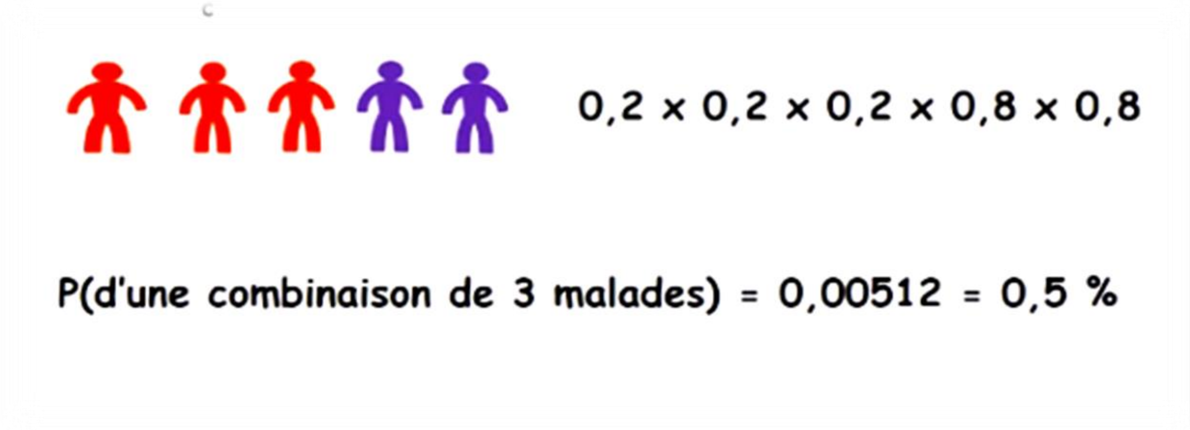
<u>Taille de l'échantillon : n</u>



<u>Taille de l'échantillon : n =5</u>

P(3 malades)

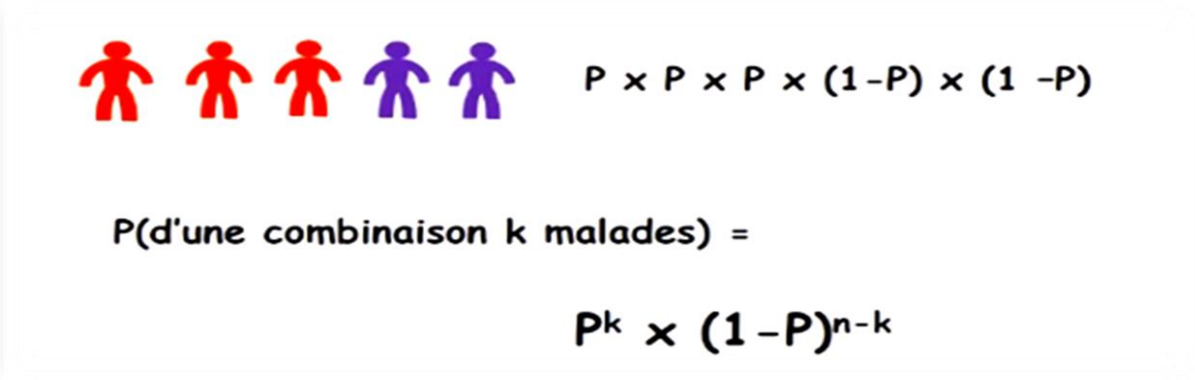
Probabilité d'une combinaison de 3 malades



Dans cet exemple, on va s'intéresser à la probabilité de trouver **<u>3 sujets</u>** malades dans un échantillon **<u>de 5 sujets</u>**, sachant que la probabilité de la maladie dans la population est de 0,2 **ou 20%**

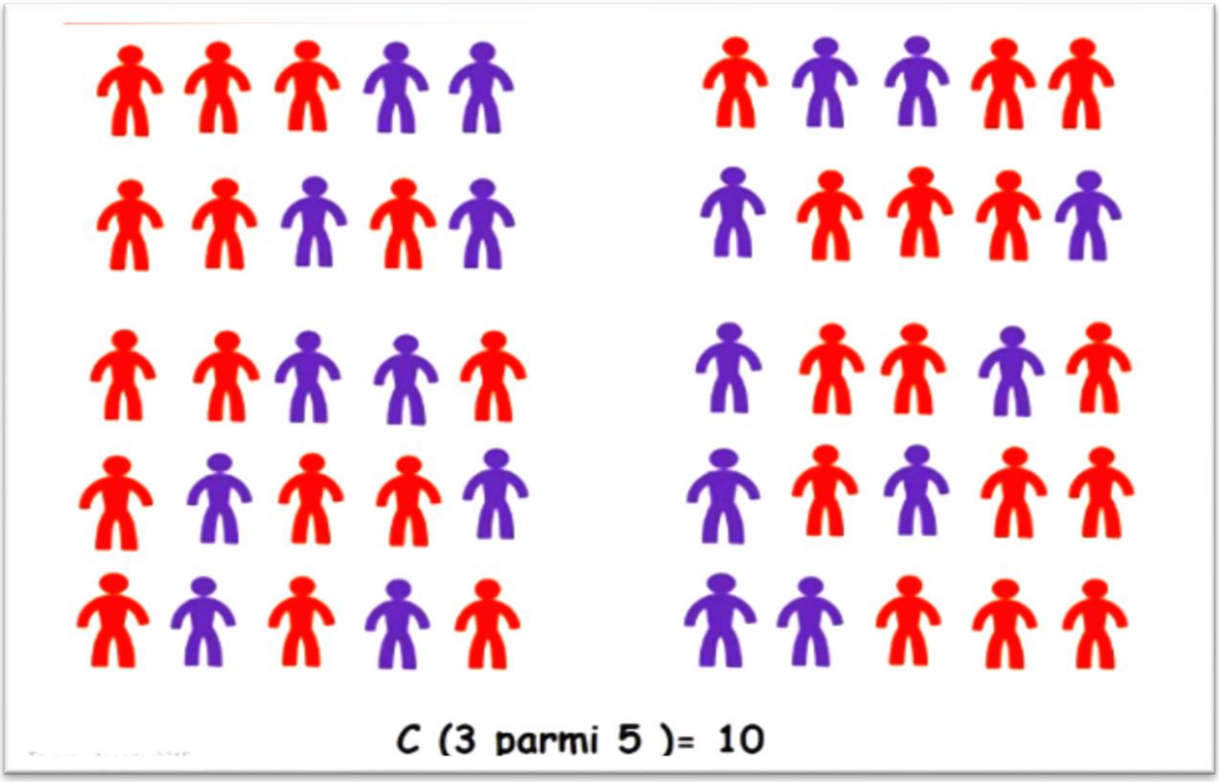
<u>Taille de l'échantillon : n</u>

<u>P(k malades)</u>



Il faut déterminer le nombre de combinaisons de 3 malades pris parmi 5 sujets :

Combinaisons de 3 malades parmi 5 sujets



A savoir que le nombre de combinaisons possibles de k malades parmi n sujets est de :

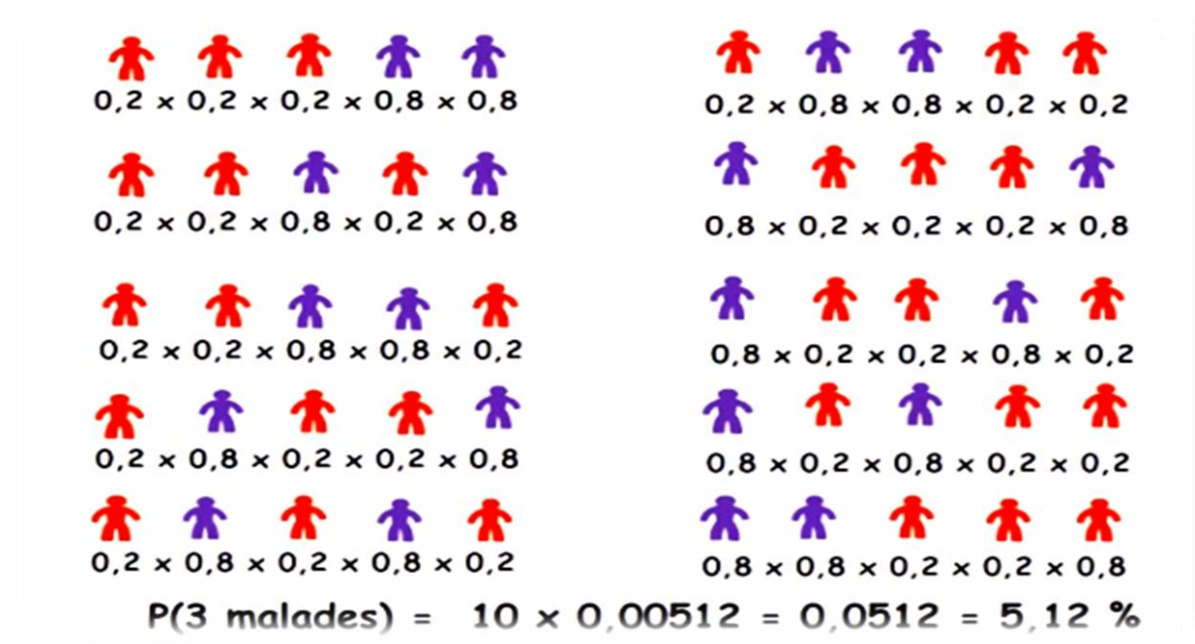
$$C_n^k = \frac{n!}{k!(n-k)!}$$

Dans notre exemple, il est de :

$$\boldsymbol{C_5^3} = \frac{5!}{3!(5-3)!} = \frac{5\times4\times3\times2}{3\times2\times(2)} = \frac{120}{12} = 10$$

Il va falloir calculer la probabilité **d'être malade parmi 5 sujets dans les 10 combinaisons possibles**

<u>Probabilités de 3 malades parmi 5 sujets</u>



2.1.3. Formule de la loi binomiale :

Considérons une suite de n épreuves de Bernoulli identiques et indépendantes
Appelons k le nombre de réalisations de l'évènement A (Nombre de Pile) parmi n épreuves.
k est une réalisation d'une variable K qui peut prendre les valeurs 0,1,2,3,…n (intervalle fini)
On montre que la loi de probabilité de la variable K est donnée par la formule

$$\mathbf{Pr(K = k) = \frac{n!}{k!(n-k)!} P^k (1-P)^{n-k}}$$

$$\mathbf{Pr(k) = C_n^k P^k (1-P)^{n-k}}$$

Autre écriture : $\boldsymbol{P(X = x) = C_n^x * p^x * q^{n-x}\ x \epsilon \{0 \dots n\}}$

x= k

-La loi binomiale dépend de 2 paramètres n et p. On la note souvent en abrégé B(n,p)
-Si p=0,5, la relation se simplifie, Pr(X=k) = $C_n^k * 0,5^k * 0,5^{n-k}$ = $C_n^k * 0,5^n$

2.1.4. Moyenne et Variance de la loi binomiale

E(X) = np

V(X)= np(1-p) -$\sigma = \sqrt{np(1-p}$

2.2. Exemple d'application à l'épidémiologie

Raisonnement général : Quelle est

1) La probabilité d'observer **k** individus possédant une caractéristique donnée,
2) Dans un échantillon de **n** individus,
3) Tirés d'une population où la proportion P de la caractéristique est connue ?

Exemple: étude sur les déterminants des troubles au sommeil chez l'enfant de 2 à 3 ans : Sachant que :

-Le pourcentage d'enfants avec troubles du sommeil dans la population d'étude : P=17%

Taille de l'échantillon : n =10

Nombre de troubles du sommeil : k = 4

Question : Cet échantillon provient- il de la population d'étude ?

Ou bien : *Quelle* est La probabilité d'observer 4 malades dans un échantillon de 10 sujets choisis au hasard dans une population où la fréquence de la maladie est de 17% ?

(si probabilité élevée, échantillon provient de la population
si la probabilité faible, échantillon ne provient pas de la population d'étude)

Paramètres de la loi binomiale

Caractéristique = troubles du sommeil	**k**	**4**
Taille de l'échantillon	**n**	**10**
Proportion de sujets porteurs de la caractéristique dans la population	**P**	**0,17**

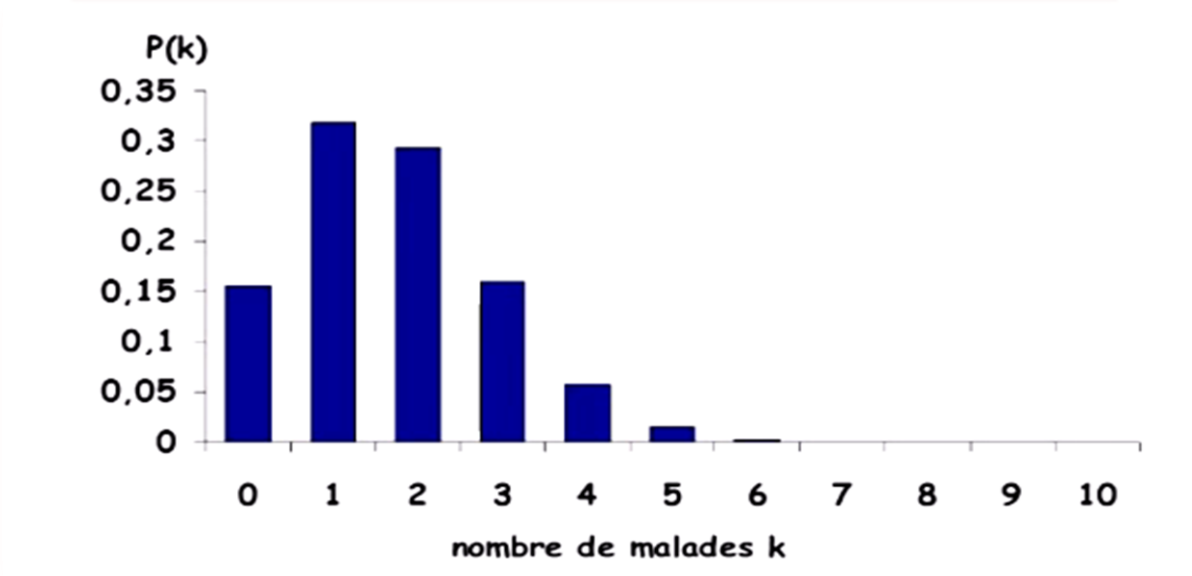
$$\mathrm{Pr(k)} = \frac{\mathrm{n!}}{\mathrm{k!(n-k)!}}\mathrm{P^k}\,(1-P)^{n-k}$$

$\mathrm{Pr\ (4\ malades)} = \frac{10!}{4!(10-4)!}\,0{,}17^4\,(1\text{-}0{,}17)^{\ 10-4}$ -**Pr (4malades) = 0,057 = 5,7 %**

2.3. Loi Binomiale cumulée

Si la probabilité d'être malade est de 17% et Si on observe un échantillon de 10 sujets
En appliquant la formule de calcul de probabilité selon la loi binomiale selon le nombre de malades ou la valeur de (k) , on aura :

Probabilité de n'observer aucun malade Pr (0) = 0,155
Probabilité de n'observer1 malade Pr (1) =0,318
Probabilité de n'observer 2 malades Pr (2) = 0,293
Probabilité de n'observer 3 malades Pr (3) =0,160
Probabilité de n'observer 4 malades Pr (4) = 0,057
Probabilité de n'observer 5 malades Pr (5) =0,014
Probabilité de n'observer 6 malades Pr (6) = 0,0024
Probabilité de n'observer 7 malades Pr (7) =0,00028
Probabilité de n'observer 8 malades Pr (8) = 0,00002
Probabilité de n'observer 9 malades Pr (9) =0,000001
Probabilité de n'observer 10 malades Pr (10) = 0,00000002
La somme des probabilités pour tout l'échantillon est égale à 1.



Distribution de la probabilité d'obtenir k malades dans un échantillon de 10 sujets issus d'une population ou la fréquence de la maladie est de 17%

Propriété cumulative de la loi Binomiale

Souvent la question posée n'est pas simplement la probabilité d'observer une seule valeur k , mais la probabilité d'observer :

1) Une valeur au plus égale à k
2) Une valeur au moins égal à k

1) Pr (valeur au plus égale à k)= Pr(0) + Pr (1)+ ….+ Pr (k-1) + Pr (k)

Exemple 1 : Quelle est la probabilité d'observer au maximum 3 malades dans un échantillon de 10 sujets choisis au hasard dans une population ou la fréquence de la maladie est de 17%

Pr (3 malades maxi) = Pr (0) + Pr (1)+ Pr (2) + Pr (3) =

= 0,155 +0,318 + 0,293 + 0,160 = 0,926

2) Pr (valeur au moins égale à k)=

Pr (k) +Pr(k+1) +…. + Pr (k_n)= 1 – Pr (valeur $< k$)

Exemple 2 : Quelle est la probabilité d'observer au moins 4 malades dans un échantillon de 10 sujets choisis au hasard dans une population au la fréquence de la maladie est de 17% ?

Pr (4 malades min) = Pr (4) + Pr (5)+….. Pr (10)

= 0,.057 + 0,014 + ……+ 0,00000002

Ou bien

= 1- Pr (3 malades maxi)

Pr (3) =0,926 d'où1- Pr (3 malades maxi) = 1 - 0,926= 0,074= 7,4%

2.4. Conditions d'application de la loi binomiale

Pour utiliser la loi binomiale, il faut que :

1. la variable étudiée soit de type binaire,
2. les tentatives soient indépendantes les unes des autres (situation de jeu) ou bien que l'échantillon soit tiré au sort,
3. chaque événement ait la même ait la même probabilité de succès au cours de chaque tentative ou bien tous les individus de la population étudiée aient la même chance d'être tiré au sort,
4. la taille n de l'échantillon soit négligeable par rapport à la taille N de la population (n/N< 10%).

Références bibliographiques :

-Admane O. Hoang KY-Ouakli N. Statistiques (cours et exercices) pour les étudiants du tronc commun biomédical. Office des Publications Universitaires 1998

- Ancelle T.Statistique. Epidémiologie. Editions Maloine 2002.

-Thierry ANCELLE. 2014*(YouTube https://www.youtube.com/c/ThierryAncellecelle consulté le Novembre 2016)*