Objectif : Un laboratoire dispose de 200 échantillons de sang et doit déterminer ceux infectés par un agent pathogène.

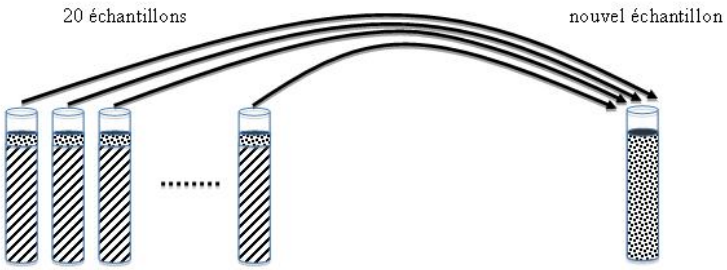
On souhaite minimiser le nombre de tests à réaliser.

Les tests sont supposés être tous parfaitement fiables. On sait que la probabilité qu'un individu soit infecté est égale à 0,05.

Partie A :

Dans un premier temps, on répartit ces 200 échantillons en 10 groupes de 20 échantillons. Pour chaque groupe, on effectue le protocole suivant :

1. On réalise un prélèvement sur chacun des 20 échantillons ;
2. On mélange ces prélèvements. On crée ainsi un nouvel échantillon qu'on analyse ;
3. Si le résultat est négatif, c'est qu'aucun des 20 échantillons n'est infecté, et si le résultat est positif, on analyse alors chaque échantillon.



1. Déterminer la probabilité que les 20 échantillons ne soient pas infectés. En déduire celle de l'événement A : « Au moins un des 20 échantillons est infecté ».

   La probabilité qu'aucun des 20 échantillons testés ne soit infecté est égale à $(1-0,05)^{20} \approx 0,36$.

   Et donc, la probabilité demandée (celle de l'évènement contraire) est environ égale à 0,64.

2. Soit X la variable aléatoire égale au nombre de réalisations de A dans les 10 groupes. Montrer que X suit une loi binomiale et préciser ses paramètres.

   X suit une loi binomiale de paramètres 10 (10 groupes à tester) et 0,64 (la probabilité que le groupe soit infecté est 0,64).

3. Y est la variable aléatoire égale au nombre total d'analyses effectuées. Justifier que Y = 10 + 20 X.

   On fait de toute façon 10 analyses (les 10 groupes).

   Chaque fois qu'un groupe est infecté, on analyse les 20 échantillons composants ce groupe. Ainsi, si 3 groupes sont infectés, cela nécessitera 3×20 analyses en plus. Plus généralement, X étant le nombre de groupes infectés, on aura à faire $20 \times$ X analyses en plus.

   Au total, on a bien 10+20 X analyses à faire ; on a donc (avec la notation choisie) : Y = 10 + 20 X.

4. Calculer alors E(X), puis E(Y). Comment interprétez-vous ce dernier résultat ?

   X suit une loi binomiale ; l'espérance de cette loi est donnée par la formule E(X)=np ce qui donne ici : $E(X)=10 \times 0,64 = 6,4$.

   Comme Y=10+20 X, on a : E(Y)=E(10+20 X)=E(10)+20 E(X)=10+20 E(X) ce qui donne : $E(Y) = \approx 138,3$.

   Cela veut dire qu'avec cette méthode, on fera en moyenne 138,3 analyses (au lieu des 200 initiales !)

Partie B :

Il est peut-être possible d'améliorer encore l'efficacité de cette méthode. En effet, les 200 échantillons auraient pu être répartis autrement ...

1. Adapter la démarche de la partie A pour une répartition en 50 groupes de 4. Que constatez-vous ?

   La probabilité qu'aucun des 4 échantillons testés ne soit infecté est égale à $(1-0,05)^4 \approx 0,81$.

   La probabilité de l'évènement contraire) est environ égale à 0,19.

   X suit une loi binomiale de paramètres $n = 50$ et $p = 0,19$; son espérance est environ 17,2.

   On a alors : Y = 50 + 4 X ce qui donne une espérance d'environ : $E(Y) \approx 50 + 4 \times 17,2 \approx 118,8$: cette répartition est encore plus efficace que la précédente puisqu'en moyenne, on aura besoin de 118,8 analyses de sang pour analyser les 200 échantillons de départ.

2. Écrire toutes les répartitions possibles en groupes de même effectif, et indiquer pour chacune le nombre moyen de tests nécessaires. Conclure.

   Les résultats sont consignés dans le tableau suivant :

répartition						
nbre de groupes	nb éch par gpe	total	P(aucun échantillon infecté)	P(au moins 1 échantillon infecté)	E(X)	E(Y)
1	200	200	0,00004	0,99996	1,0000	200,9930
2	100	200	0,00592	0,99408	1,9882	200,8159
4	50	200	0,07694	0,92306	3,6922	188,6110
5	40	200	0,12851	0,87149	4,3574	179,2976
8	25	200	0,27739	0,72261	5,7809	152,5221
10	20	200	0,35849	0,64151	6,4151	138,3028
20	10	200	0,59874	0,40126	8,0253	100,2526
25	8	200	0,66342	0,33658	8,4145	92,3159
40	5	200	0,77378	0,22622	9,0488	85,2438
50	4	200	0,81451	0,18549	9,2747	87,0988
100	2	200	0,90250	0,09750	9,7500	119,5000
200	1	200	0,95000	0,05000	10,0000	210,0000

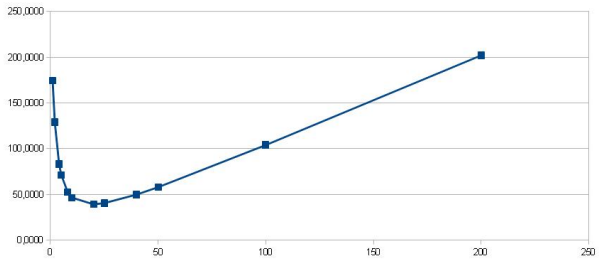
Partie C :

Généralisation :

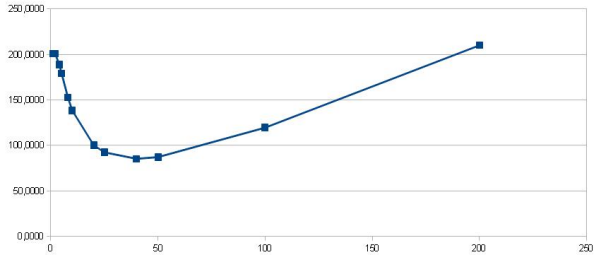
A partir du nombre initial d'échantillons et de la probabilité d'infection, on peut obtenir à l'aide d'un tableur, la répartition en groupes de même effectif permettant de minimiser le nombre moyen de tests.

indications :
- Reprendre le travail précédent et le généraliser à toutes les répartitions possibles de 200 échantillons en groupes ;
- Utiliser le tableur pour « automatiser » les calculs faits précédemment ;
- Se donner la possibilité de modifier la valeur de p (probabilité qu'un individu soit infectée) initialement égale à 0,05 ;
- Construire une représentation graphique donnant le nombre moyen d'analyses en fonction des répartitions, pour une valeur de p donnée ;
- Modifier les valeurs de p et observer les allures de la courbe obtenue ;
- Donner une conclusion générale.

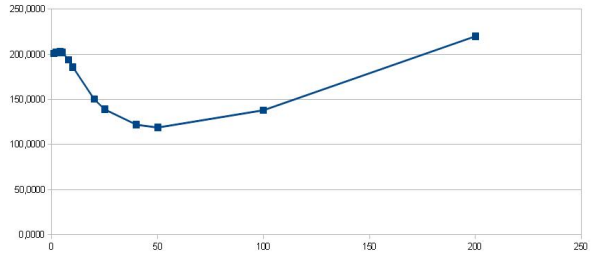
On présente ici les courbes représentant le nombre moyen d'analyses en fonctions du nombre de groupes pour différentes valeurs de la probabilité qu'un individu soit infecté.



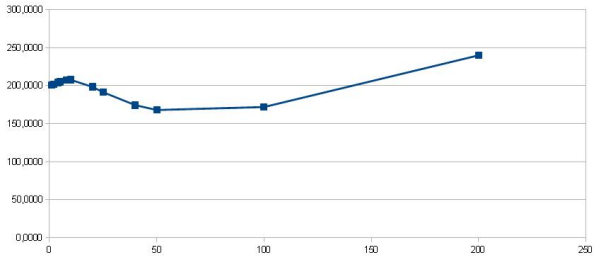
$p = 0,01$



$p = 0,05$



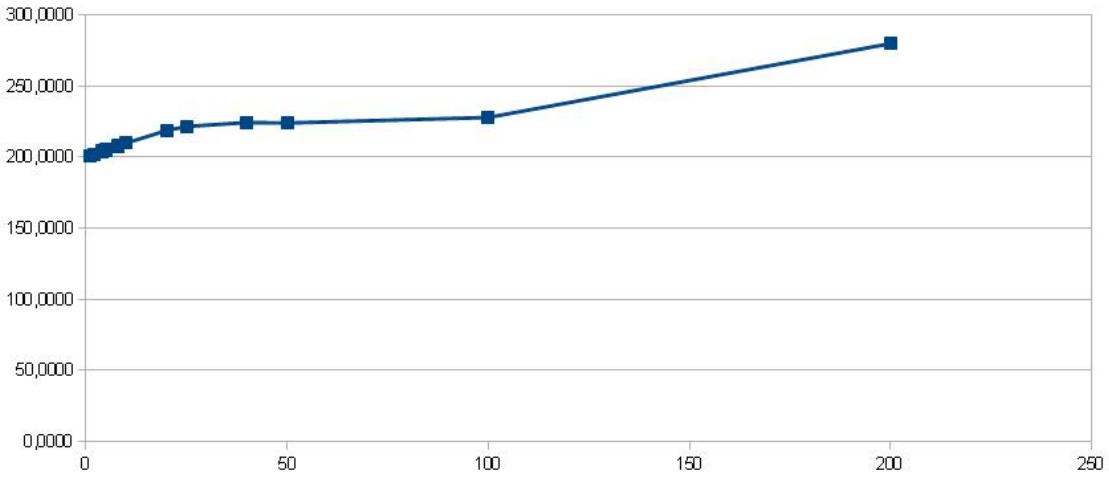
$p = 0,1$



$p = 0,2$

On remarque que pour ces faibles valeurs de p, on a intérêt à utiliser le procédé précédent. Plus p est petit, plus on a intérêt à faire des groupes importants (ça se comprend puisque le risque qu'un groupe soit infecté est d'autant plus rare).

En revanche, dès que la probabilité devient importante, la technique n'est plus efficace ; il vaut mieux faire les 200 analyses :



$p = 0,4$