

1^{ère} baccalauréat en Sciences Vétérinaires

Biostatistiques – Travaux dirigés

Séance de TD n°1 : Résolutions

1. Chez le porc, la couleur blanche de la robe (II: Landrace) domine le tacheté pie-noire (ii: Piétrain). Deux hybrides de première génération sont croisés entre-eux et donnent 7 descendants.

Quelle est la probabilité que l'on ait, parmi ceux-ci, moins de 3 individus blancs ?

RESOLUTION :

Hybride de première génération = Ii

Croisement de ces deux hybrides :

	I	i
I	II	Ii
i	Ii	ii

On a donc 25% de chance d'avoir II, 25% d'avoir ii et 50% d'avoir Ii.

Etant donné que la couleur blanche est dominante, tout individu ayant au moins un I sera blanc. Cela signifie que les individus II et Ii seront blancs. On a donc 75% de chances d'avoir un individu blanc après le croisement de deux hybrides de première génération.

La question est donc quelle est la probabilité d'avoir moins de 3 individus blancs sur 7 descendants sachant que la probabilité d'être blanc est de 75%.

On se trouve dans une distribution binomiale : distribution discrète (blanc ou non) avec une taille d'échantillon déterminée (N=7) et une prévalence connue (p = 0,75).

La probabilité d'avoir moins de 3 individus équivaut à calculer toutes les probabilités inférieures à 3 et à en faire la somme. En terme statistique cela s'écrit :

$$P(r < 3) = P(r = 0) + P(r = 1) + P(r = 2)$$

Donc calculons, chaque probabilité séparément.

Selon la formule de la loi binomiale :

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$P(0) = C_7^0 * 0,75^0 * 0,25^{7-0} = 1 * 1 * 0,25^7 = 0,000061$$

$$P(1) = C_7^1 * 0,75^1 * 0,25^{7-1} = 7 * 0,75 * 0,25^6 = 0,0012817$$

$$P(2) = C_7^2 * 0,75^2 * 0,25^{7-2} = 21 * 0,5625 * 0,25^5 = 0,01153$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

Donc, $P(r < 3) = P(r = 0) + P(r = 1) + P(r = 2) = 0,000061 + 0,0012817 + 0,01153 = 0,01288$

2. En race de Moyenne et Haute Belgique, on observe 3 phénotypes de robe: blanche, noire et bleue, correspondant respectivement aux génotypes NN, nn et Nn. Un taureau bleu est croisé avec 3 vaches bleues; quelle est la probabilité que, parmi les 3 produits, il y ait un blanc ?

REPONSE :

	N	n
N	NN	Nn
n	Nn	nn

Le croisement d'un taureau bleu avec une vache bleue entraîne comme probabilité pour les robes : 25% de NN (blanche), 25% de nn (noire) et 50% de Nn (bleue).

La probabilité d'être blanc après un croisement est donc de 25%.

On se trouve dans une distribution binomiale : distribution discrète (blanc ou non) avec une taille d'échantillon déterminée (N=3) et une prévalence connue (p = 0,25).

Selon la formule de la loi binomiale :

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$P(1) = C_3^1 * 0,25^1 * 0,75^{3-1} = 3 * 0,25 * 0,75^2 = 0,421875$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

La réponse est donc $P(r=1) = 0,421875$

3. D'un jeu de 32 cartes, X tire une carte et la replace dans le jeu. A son tour, Y tire une carte. Quelle est la probabilité que tous deux aient tiré la même carte ?

REPONSE :

Le tirage de Y est indépendant de celui de X.

La probabilité que Y tire une carte = $1/32$

La probabilité que X tire la même carte que Y = $1/32$.

Probabilité que Y ET X tire la même carte = $(1/32) * (1/32) * 32$ (car 32 cartes possibles) = $1/32$

$p = 1/32$

4. On jette deux fois consécutivement une paire de dés. Quelle est la probabilité d'avoir une fois 5 (somme obtenue avec les deux dés) ? Au moins une fois 5 ? Deux fois 5 ?

REPONSE :

a. Quelle est la probabilité d'avoir une fois 5 (somme obtenue avec les deux dés) ?

Tout d'abord, il faut calculer la probabilité d'avoir 5 sur une paire de dés lors d'un lancer. Les possibilités d'avoir 5 sont : 4 et 1, 1 et 4, 2 et 3, 3 et 2. Il y a donc 4 possibilités. Le nombre total de possibilités lors du lancer d'une paire de dés = $6 * 6 = 36$.

La probabilité d'avoir 5 sur une paire de dés lors d'un lancer = $4/36$

Ensuite, il faut calculer la probabilité d'avoir une fois 5 lors de 2 lancers. Ce sont deux événements successifs. Cela peut apparaître lors de 2 situations : (Avoir 5 lors du 1^{er} lancer ET ne pas avoir 5 lors du 2^{ème} lancer) OU (Ne pas avoir 5 lors du 1^{er} lancer ET avoir 5 lors du 2^{ème} lancer).

$P(5 \text{ lancer } 1) * P(\text{Pas } 5 \text{ lancer } 2) + P(\text{Pas } 5 \text{ lancer } 1) * P(5 \text{ lancer } 2) = 4/36 * 32/36 + 32/36 * 4/36 = 0,1975$

b. Probabilité d'avoir au moins une fois 5

La probabilité d'avoir au moins une fois 5 correspond à la probabilité d'avoir une fois 5 OU la probabilité d'avoir 2 fois 5 = $P(1x5) + P(2x5)$.

La probabilité d'avoir une fois 5 a été considérée à la question précédente et égale 0,1975.

La probabilité d'avoir deux fois 5 = $P(5 \text{ lancer } 1) \text{ ET } P(5 \text{ lancer } 2) = 4/36 * 4/36 = 0,01234567$

La probabilité d'avoir au moins une fois 5 = $0,1975 + 0,01234567 = \mathbf{0,20987654}$

c. Probabilité d'avoir deux fois 5

Cela a été calculé à la question précédente :

La probabilité d'avoir deux fois 5 = $P(5 \text{ lancer } 1) \text{ ET } P(5 \text{ lancer } 2) = 4/36 * 4/36 = \mathbf{0,01234}$

5. Un variant génétique a, dans une population, la fréquence de 5 pour mille. Combien d'animaux doit-on examiner pour avoir 9 chances sur 10 de rencontrer au moins une fois ce variant ?

REPONSE :

Deux possibilités :

A. On considère qu'on est dans une distribution binomiale : distribution discrète (on a le variant ou non), prévalence = $5/1000$, taille d'échantillon à calculer. L'inconnue ici est la taille d'échantillon. Mais on sait que la probabilité d'avoir au moins une fois ce variant = $9/10$. En termes statistique cela s'écrit : $P(r \geq 1) = 0,9$. Sachant que la somme des probabilités = 1, on peut en déduire que $P(r = 0) = 1 - P(r \geq 1)$ et donc que $P(r = 0) = 1 - 0,9 = 0,1$.

Selon la formule de la loi binomiale :

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$P(0) = C_n^0 * \frac{5}{1000}^0 * 0,995^{n-0} = 0,10$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

$$\Leftrightarrow P(0) = 1 * 1 * 0,995^n = 0,10$$

$$\Leftrightarrow 0,995^n = 0,10$$

$$\Leftrightarrow n * \log(0,995) = \log(0,10)$$

$$\Leftrightarrow n = \frac{\log(0,10)}{\log(0,995)}$$

$$\Leftrightarrow n = 459,364764$$

- B. On considère qu'on est dans une distribution de poisson : : distribution discrète (on a le variant ou non), prévalence petite : 5/1000 et taille d'échantillon certainement grande. L'inconnue ici est la taille d'échantillon. Mais on sait que la probabilité d'avoir au moins une fois ce variant = 9/10. En termes statistique cela s'écrit :
 $P(k \geq 1) = 0,9$.
 Sachant que la somme des probabilités = 1, on peut en déduire que $P(k = 0) = 1 - P(k \geq 1)$ et donc que $P(k = 0) = 1 - 0,9 = 0,1$.

Selon la formule de la loi de poisson:

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

$$P(0) = \frac{e^{-\mu} * \mu^0}{0!} = 0,10$$

$$\Leftrightarrow P(0) = \frac{e^{-\mu} * 1}{1} = 0,10$$

$$\Leftrightarrow P(0) = e^{-\mu} = 0,10$$

$$\Leftrightarrow -\mu = \ln(0,10)$$

$$\Leftrightarrow \mu = -\ln(0,10)$$

$$\Leftrightarrow \mu = 2,302585$$

Pour rappel, μ correspond à la moyenne et donc à $n * p$.

$$\mu = n * p$$

$$2,302585 = n * \frac{5}{1000}$$

$$n = \frac{2,302585}{\frac{5}{1000}} = 460,517$$

$$\Rightarrow n = \mathbf{460 \text{ animaux}}$$

6. Vous cherchez à déterminer si le nouveau test antigénique « rapide » pour la détection de la leucose féline (due au virus de l'immunodéficience féline) est suffisamment fiable, par rapport à un test de référence réalisé en laboratoire (« gold standard », pour lequel la sensibilité et la spécificité du test sont 100%). Sur les 300 chats testés, 8 ont présenté à la fois un test rapide et une analyse en laboratoire positifs. 12 ont eu un test positif en laboratoire et 10 ont eu un test rapide positif.

- Quelle est la prévalence de la maladie dans cet échantillon ?
- Quelles sont les spécificité et sensibilité de ce test ?
- Quelle est la Valeur Prédictive Positive ?

RESOLUTION

Nous pouvons organiser les éléments dont nous disposons sous forme de tableau :

	Porteurs	Sains	Total
Test +	8		10
Test -			
Total	12		300

A partir de ces informations, nous pouvons déduire le reste des valeurs de la table :

	Porteurs	Sains	Total
Test +	8	2	10
Test -	4	286	290
Total	12	288	300

Nous pourrions aussi considérer cette table d'un point de vue épidémiologique :

	Porteurs	Sains	Total
Test +	VP = 8	FP = 2	VP + FP = 10
Test -	FN = 4	VN = 286	FN + VN = 290
Total	VP + FN = 12	FP + VN = 288	N = 300

Avec VP = vrais positifs, FN = faux négatifs, FP = faux positifs et VN = vrais négatifs. N = effectifs totaux.

- La prévalence correspond au taux de porteurs (malades), c'est-à-dire au nombre d'individus porteurs sur l'ensemble de la population.

$$\text{Prévalence} = \frac{\text{Nombre d'individus porteurs}}{\text{Nombre d'individus total}} = \frac{VP + FN}{N} = \frac{12}{300} = 0,04$$

La prévalence de la maladie est donc de 4%.

- b) La sensibilité d'un test correspond à sa capacité à détecter (test +) un individu qui est malade. Il s'agit donc du calcul de probabilité suivant :

$$\begin{aligned} \text{Sensibilité} &= P(\text{test} + | \text{porteurs}) \\ &= \frac{\text{Nombre d'individus à test} + \text{et porteurs}}{\text{Nombre de porteurs}} = \frac{8}{12} = 0,666667 \end{aligned}$$

Si on raisonne d'un point de vue épidémiologique, la sensibilité correspond au taux de vrais positifs :

$$\text{Sensibilité} = \frac{VP}{VP + FN} = \frac{VP}{\text{Porteurs}} = \frac{8}{12}$$

La spécificité d'un test correspond à sa capacité à « ne pas détecter » comme malade (test –) un individu qui est sain. Cela correspond donc au taux de vrais négatifs. On peut donc la calculer de la manière suivante :

$$\begin{aligned} \text{Spécificité} &= P(\text{test} - | \text{sains}) = \frac{\text{Nombre d'individus à test} - \text{et sains}}{\text{Nombre de sains}} \\ &= \frac{VN}{VN + FP} = \frac{VN}{\text{Sains}} = \frac{286}{288} = 0,993056 \end{aligned}$$

Le test présente donc une sensibilité de 66,7% et une spécificité de 99,3 %.

- c) La Valeur Prédictive Positive (VPP) d'un test correspond à la probabilité d'observer un individu porteur sachant que son test est + c'est-à-dire la probabilité que la maladie soit vraiment présente lorsque le résultat du test est positif.

$$\begin{aligned} VPP &= P(\text{porteurs} | \text{test} +) = \frac{\text{Nombre d'individus à test} + \text{et porteurs}}{\text{Nombre d'individus à test} +} \\ &= \frac{VP}{VP + FP} = \frac{8}{10} = 0,8 \end{aligned}$$

Plus la VPP est faible et plus le test rapide indique un nombre important d'individus « malades » alors qu'ils sont en réalité sains (augmentation du nombre de faux positifs).

7. 60 % des œufs de parasites entament leurs divisions cellulaires dans un certain milieu de culture. Quelle est la probabilité que, sur 8 œufs, 6 et plus aient commencé à se diviser?

REPONSE :

On se trouve dans une distribution binomiale : distribution discrète à 2 possibilité (division cellulaire : oui vs non), avec une taille d'échantillon connue (N=8) et une prévalence (p =0,60).

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$\text{Pour rappel : } C_n^r = \frac{n!}{r!(n-r)!}$$

$$P(r \geq 6) = P(r = 6) + P(r = 7) + P(r = 8).$$

$$P(6) = C_8^6 * 0,6^6 * (1 - 0,6)^{8-6} = 0,2090188$$

$$P(7) = C_8^7 * 0,6^7 * (1 - 0,6)^{8-7} = 0,08957957$$

$$P(8) = C_8^8 * 0,6^8 * (1 - 0,6)^{8-8} = 0,016796$$

$$P(r \geq 6) = 0,31539456$$

8. Considérant des nichées de 10 porcelets et en admettant une égale proportion des sexes, quelle est la probabilité qu'une nichée au hasard contienne plus de 7 mâles ?

REPONSE :

On se trouve dans une distribution binomiale : distribution discrète à 2 possibilité (sexe nichée : mâles vs femelles), avec une taille d'échantillon connue (N=10) et une prévalence (égale proportion des sexes => p =0,50).

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$\text{Pour rappel : } C_n^r = \frac{n!}{r!(n-r)!}$$

$$P(r \geq 7) = P(r = 7) + P(r = 8) + P(r = 9) + P(r = 10).$$

$$P(7) = C_{10}^7 * 0,5^7 * (1 - 0,5)^{10-7} = 0,0439$$

$$P(8) = C_{10}^8 * 0,5^8 * (1 - 0,5)^{10-8} = 0,009765625$$

$$P(9) = C_{10}^9 * 0,5^9 * (1 - 0,5)^{10-9} = 0,0009765625$$

$$P(r > 7) = 0,05469$$

9. Des cas de maladies sont recensés dans une série d'étables, et ont donné les résultats suivants:

Animaux malades	Fréquence
0	2
1	9
2	26
3	20
4	16
5	8

On demande de calculer le nombre d'exploitations, le nombre moyen d'animaux malades, la variance, la déviation standard, le coefficient de variation. Si l'effectif moyen d'une ferme est de 15 animaux, calculez la probabilité qu'un animal pris au hasard dans une des fermes de l'expérience soit malade.

REPONSE :

$$\text{Nombre d'exploitation : } \sum \text{Fréquence} = 2 + 9 + 26 + 20 + 16 + 8 = 81$$

$$\text{Nombre moyen d'animaux malades} = \frac{\sum f \cdot X}{N} = \frac{(2 \cdot 0 + 9 \cdot 1 + 26 \cdot 2 + 20 \cdot 3 + 16 \cdot 4 + 8 \cdot 5)}{81} = \frac{225}{81} = 2,778$$

$$\text{Variance} = s^2 = \frac{\sum f \cdot X^2 - \frac{(\sum f \cdot X)^2}{\sum f}}{\sum f - 1} = \frac{(2 \cdot 0^2 + 9 \cdot 1^2 + 26 \cdot 2^2 + 20 \cdot 3^2 + 16 \cdot 4^2 + 8 \cdot 5^2) - \frac{225^2}{81}}{81 - 1} = 1,55$$

$$\text{Déviation standard} = s = \sqrt{s^2} = \sqrt{1,55} = 1,24499$$

$$\text{Coefficient de variation} = \frac{s}{\bar{x}} * 100 = \frac{1,24499}{2,778} * 100 = 44,8196\%$$

$$\text{Probabilité d'avoir un animal malade} = \frac{\text{Nombre animaux malades}}{\text{Nombre total animaux}} = \frac{(9 \cdot 1 + 26 \cdot 2 + 20 \cdot 3 + 16 \cdot 4 + 8 \cdot 5)}{81 \cdot 15} = \frac{225}{1215} = \frac{225}{1215} = 0,185185$$

10. Dans une solution homogène, la présence d'un virus à cloner est estimée à 1 par 100ml. Un expérimentateur, prélevant 10 ml, espère obtenir 0 ou 1 virus, mais pas plus. Quelle chance a-t-il d'atteindre ce but ? Quand aura-t-il plus de 3 virus ?

REPONSE :

On se trouve dans une distribution de Poisson : distribution discrète à deux possibilités (virus : présence vs absence), la taille d'échantillon est inconnue (on a aucune idée du nombre total de virus) et la prévalence est faible (moyenne : 1 virus par 100 ml).

Selon la formule de la loi de poisson:

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

On sait que la moyenne est de 1 virus par 100 ml. Ici, comme on prélève 10 ml, la moyenne sera de 1/10 virus.

a.) Probabilité d'obtenir 0 ou 1 virus ?

$$P(k = 0 \text{ ou } k = 1) = P(k = 0) + P(k = 1)$$

$$P(0) = \frac{e^{-0,1} * 0,1^0}{0!} = 0,904837418$$

$$P(1) = \frac{e^{-0,1} * 0,1^1}{1!} = 0,0904837418$$

$$P(k=0 \text{ ou } k=1) = \mathbf{0,99532116}$$

b.) Probabilité d'obtenir plus de 3 virus ?

$$P(k > 3) = 1 - P(k=3) - P(k=2) - P(k=1) - P(k=0)$$

$$P(2) = \frac{e^{-0,1} * 0,1^2}{2!} = 0,004524187$$

$$P(3) = \frac{e^{-0,1} * 0,1^3}{3!} = 0,00015$$

$$P(k > 3) = 1 - 0,00015 - 0,0045 - 0,09048 - 0,9048 = \mathbf{0,000003847}$$

11. La probabilité d'obtenir une génisse lors d'un vêlage est de 0,45. Quelle est la probabilité d'obtenir 7 génisses lors de 20 naissances ? Pourquoi cette probabilité ne vaut-elle pas 0,45⁷? Quand vaudra-t-elle cette valeur ?

REPONSE :

On se trouve dans une distribution binomiale : distribution discrète à deux possibilités (sexe naissance : génisse vs veau), la taille d'échantillon est connue (N=20) et la prévalence est de 0,45.

$$P(r) = C_n^r * p^r * q^{n-r}$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

a. Probabilité d'obtenir 7 génisses :

$$P(7) = C_{20}^7 * 0,45^7 * (1 - 0,45)^{20-7} = \mathbf{0,12207}$$

b. Pourquoi cette probabilité ne vaut pas 0,45⁷ ?

Cette probabilité ne vaut pas 0,45⁷ car il y a 20 naissances et qu'il faut prendre en compte les 13 cas où ce ne sont pas des génisses. De plus, il y a plusieurs possibilités d'avoir 7 génisses sur 20 naissances.

c. Quand cette probabilité vaudrait 0,45⁷ ?

La probabilité vaut 0,45⁷ si on cherche la probabilité d'avoir 7 génisses sur 7 naissances.

12. Une affection bovine, caractérisée par de la syndactylie et des problèmes respiratoires, est causée par un variant génétique rare (p=0,01). La tare est une tare récessive. Quelle est la probabilité d'observer 1 individu dans un pool aléatoire de 1000 individus ? Utilisez une distribution binomiale et une distribution de Poisson.

REPONSE :

a. Loi binomiale :

$$P(r) = C_n^r * p^r * q^{n-r}$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

Etant donné que la tare est récessive, la prévalence de la tare est égale à $p^2 = 0,01 * 0,01 = 0,001$

$$P(r = 1 | p=0,001, N=1000) = C_{1000}^1 * 0,001^1 * (1 - 0,001)^{1000-1} = \mathbf{0,09049}$$

b. Loi de poisson :

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

La moyenne d'une distribution discrète à deux possibilités = $n * p = 1000 * 0,0001 = 0,1$

$$P(k = 1 | \mu = 0,1) = \frac{e^{-0,1} * 0,1^1}{1!} = \mathbf{0,0904837}$$

13. Les cas de dysplasie de la hanche dans une race de chiens sont classifiés en 3 catégories : légère (30%), moyenne (60%) et sévère (10%). Quelle est la probabilité d'avoir 12 cas légers et 8 cas moyens dans un élevage de 20 chiens ?

REPONSE :

On se trouve dans une distribution multinomiale (et même trinomiale) : distribution discrète à plus de 2 possibilités (dysplasie hanche : légère vs moyenne vs sévère), la taille d'échantillon est connue et la prévalence de chaque catégorie de dysplasie de la hanche est connue également.

$$P(k; l; m) = \frac{n!}{k! l! m!} * p^k * q^l * r^m$$

Dans ce cas-ci, $n = 20$, $k = 12$, $l = 8$, $m = 0$, p = prévalence des cas légers = 0,30, q = prévalence des cas moyens = 0,6 et r = prévalence des cas sévères = 0,10.

$$P(r = 12, 8, 0) = \frac{20!}{12!8!0!} * 0,3^{12} * 0,6^8 * 0,1^0 = \mathbf{0,00112442}$$

14. Quelle est la probabilité d'avoir un chien de plus de 60 cm en prélevant 5 chiens dans deux races (10 chiens au total) dont 60 cm est le percentile 80 et 85 respectivement ?

REPONSE :

Pour avoir 1 chien sur 10 de plus de 60 cm, on a deux possibilités :

-avoir un chien de plus de 60 cm dans la race 1 et 0 chien de plus de 60 cm dans la race 2 = $P(r=1|race1)$ ET $P(r=0|race2) = P(r=1|race1) * P(r=0|race2)$

-avoir 0 chien de plus de 60 cm dans la race 1 et un chien de plus de 60 cm dans la race2 = $P(r=0|race1)$ ET $P(r=1|race2) = P(r=0|race1) * P(r=1|race2)$

Il faut faire la somme de ces deux événements pour avoir la probabilité totale.

Le calcul de la probabilité d'avoir, par exemple, 1 chien de plus de 60 cm lorsqu'on prend 5 chiens dans une race se calcule grâce à la distribution binomiale. En effet, nous sommes dans une distribution discrète à deux possibilités (taille du chien : plus de 60 cm vs moins de 60 cm) avec une taille d'échantillon connue : $N=5$ et une prévalence connue.

Pour rappel, le percentile 80 signifie que 80% des chiens ont une taille inférieure à 60 cm et donc 20% des chiens ont une taille supérieure à 60 cm. On peut donc en déduire les prévalences dans les deux races :

-race 1 : prévalence d'avoir un chien de plus de 60 cm = $1 - 0,80 = 0,20$

-race 2 : prévalence d'avoir un chien de plus de 60 cm = $1 - 0,85 = 0,15$

$$P(r) = C_n^r * p^r * q^{n-r}$$

Pour rappel : $C_n^r = \frac{n!}{r! * (n-r)!}$

- $P(r = 1 | \text{race 1}) = C_5^1 * 0,20^1 * (1 - 0,20)^{5-1} = 0,4096$
- $P(r = 0 | \text{race 1}) = C_5^0 * 0,20^0 * (1 - 0,20)^{5-0} = 0,32768$
- $P(r = 1 | \text{race 2}) = C_5^1 * 0,15^1 * (1 - 0,15)^{5-1} = 0,3915$
- $P(r = 0 | \text{race 2}) = C_5^0 * 0,15^0 * (1 - 0,15)^{5-0} = 0,4437$

Donc,

$$\begin{aligned} P(r = 1 | \text{race 1}) * P(r = 0 | \text{race 2}) + P(r = 0 | \text{race 1}) * P(r = 1 | \text{race 2}) \\ = 0,4096 * 0,4437 + 0,32768 * 0,3915 = 0,31003 \end{aligned}$$

15. Vous souhaitez éprouver un nouveau test de dépistage pour une maladie du chien, devant être prochainement utilisé en routine. Vous le testez sur un total de 300 chiens, pour lesquels la prévalence de la maladie observée est de 22%. Parmi les chiens en bonne santé, 13 présentent un test positif. Vous savez que la sensibilité du test est de 89%.

Déterminez (en arrondissant les effectifs calculés à l'entier) :

- a) La spécificité du test.
- b) La probabilité d'être sain sachant que le test est négatif (aussi appelée Valeur Prédictive Négative).
- c) La probabilité d'être malade si le test est positif. A quoi correspond cette probabilité ?

RESOLUTION

Nous pouvons organiser les éléments dont nous disposons sous forme de tableau.

Comme la prévalence de la maladie (taux de malades) est de 22%, le nombre d'individus malades sur 300 chiens correspond à $\text{prévalence} \times \text{effectifs totaux} = 0,22 \times 300 = 66$ chiens malades. Cela nous permet déjà de remplir partiellement le tableau.

	Malades	Sains	Total
Test +	VP	FP = 13	VP + FP
Test -	FN	VN = 221	FN + VN
Total	VP + FN = $0,22 \times 300 = 66$	FP + VN = 234	N = 300

Avec VP = vrais positifs, FN = faux négatifs, FP = faux positifs et VN = vrais négatifs. N = effectifs totaux.

Nous savons que la sensibilité du test est de 89% et que le nombre de malades est de 66 chiens. Or,

$$\begin{aligned} \text{Sensibilité} &= P(\text{test} + \mid \text{malades}) = \frac{\text{Nombre d'individus à test} + \text{et malades}}{\text{Nombre de malades}} \\ &= \frac{\text{Nombre d'individus à test} + \text{et malades}}{66} = 0,89 \end{aligned}$$

On peut aussi raisonner d'un point de vue épidémiologique, sachant que la sensibilité correspond au taux de vrais positifs :

$$\text{Sensibilité} = \frac{VP}{VP + FN} = \frac{VP}{66}$$

Donc :

$$\begin{aligned} \text{Nombre d'individus à test} + \text{et malades (VP)} &= \text{nombre de malades} \times \text{sensibilité} \\ &= (VP + FN) \times \text{sensibilité} = 66 \times 0,89 = 58,74 \end{aligned}$$

En arrondissant à l'entier, on obtient donc 59 individus malades et à test +. Cela nous permet de compléter la table.

	Malades	Sains	Total
Test +	VP = 59	FP = 13	VP + FP = 72
Test -	FN = 7	VN = 221	FN + VN = 228
Total	VP + FN = 66	FP + VN = 234	N = 300

- a) La spécificité d'un test correspond à sa capacité de « ne pas détecter » comme malade (test –) un individu qui est sain. Cela correspond donc au taux de vrais négatifs. On peut donc la calculer de la manière suivante :

$$\begin{aligned} \text{Spécificité} &= P(\text{test} - \mid \text{sains}) = \frac{\text{Nombre d'individus à test} - \text{et sains}}{\text{Nombre de sains}} \\ &= \frac{VN}{VN + FP} = \frac{VN}{\text{Sains}} = \frac{221}{234} = 0,944444 \end{aligned}$$

La spécificité du test est de 94,4%.

- b) La Valeur Prédictive Négative (VPN) d'un test correspond à la probabilité d'observer un individu sain sachant que son test est négatif, c'est-à-dire la probabilité que l'individu soit vraiment sain lorsque le résultat du test est négatif.

$$\begin{aligned} VPN &= P(\text{sains} \mid \text{test}-) = \frac{\text{Nombre d'individus à test - et sains}}{\text{Nombre d'individus à test -}} = \frac{VN}{VN + FN} \\ &= \frac{221}{228} = 0,969298 \end{aligned}$$

La VPN est ici de 96,9 %.

Remarque : Plus la VPN est faible et plus le test indique un nombre important d'individus « sains » alors qu'ils sont en réalité malades (augmentation du nombre de faux négatifs).

- c) La probabilité d'être malade si le test est positif, que nous pouvons noter $P(\text{malade} \mid \text{test} +)$ correspond à la Valeur Prédictive Positive (VPP) d'un test. En somme, il s'agit de déterminer la probabilité que la maladie soit vraiment présente lorsque le résultat du test est positif.

$$\begin{aligned} VPP &= P(\text{malade} \mid \text{test}+) = \frac{\text{Nombre d'individus à test + et malades}}{\text{Nombre d'individus à test +}} \\ &= \frac{VP}{VP + FP} = \frac{59}{72} = 0,819444 \end{aligned}$$

La VPP est donc de 81,9%.

16. Calculez de deux manières différentes la probabilité de tirer deux vaches malades en prenant 2 vaches au hasard dans une étable de 50 animaux où une maladie est présente à raison de 4 %.

REPONSE :

On se trouve dans la situation où il n'y a pas de remise (càd qu'après avoir pris une vache, on ne la remet pas dans le troupeau. La probabilité change donc d'un tirage à l'autre).

1^{ère} solution :

Dans ce cas, le tirage de la deuxième vache est dépendant de la première vache. On peut l'écrire sous cette forme :

Probabilité de tirer une vache malade ET probabilité de tirer une deuxième vache malade sachant qu'on a déjà une vache malade au premier tirage = $P(1 \text{ vache malade}) * P(2^{\text{ème}} \text{ vache malade} \mid 1^{\text{ère}} \text{ vache malade})$.

Comme on sait que la maladie est présente à 4%, cela signifie que 4% des 50 animaux sont malades = $50 * 0,04 = 2$ vaches sont malades dans le troupeau.

Lors du premier tirage, la probabilité de tirer une vache malade = $2/50$

Lors du second tirage, il reste 49 vaches et parmi ces 49 vaches, une seule est malade puisqu'on a déjà tiré une vache malade lors du premier tirage. La probabilité de tirer une 2^{ème} vache malade = 1/49.

On obtient donc :

$$P(1 \text{ vache malade}) * P(2^{\text{ème}} \text{ vache malade} | 1^{\text{ère}} \text{ vache malade}) = 2/50 * 1/49 = \mathbf{0,0008163}$$

2^{ème} solution :

Comme on se trouve dans une distribution discrète à deux possibilités et qu'il n'y a pas de remise, on se trouve dans une distribution hypergéométrique. On peut donc utiliser la formule de la loi hypergéométrique :

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 50, constitué de 2 sous-ensembles : N1 = nombre total de vaches malades = 2 et N2 = nombre total de vaches saines = 48. a1 = nombre d'individus qu'on souhaite tirer = 2 ; a = nombre d'individus total que l'on tire = 2 ; a2 = a-a1 = 0.

Ou vous faites le tableau à double entrée pour visualiser la situation :

	Tiré	Non tiré	
Malade	2	0	2
Sain	0	48	48
	2	48	50

$$P(2; 2; 2; 50) = \frac{C_2^2 * C_{48}^0}{C_{50}^2} = \mathbf{0,0008163}$$

17. Le dosage de substances radio-actives s'effectue en comptant le nombre de particules émises pendant un temps donné. Soit une source radio-active émettant, en moyenne, une particule par minute. Le comptage se poursuit pendant plusieurs centaines de minutes, pendant lesquelles les particules sont émises au hasard. Dans quel pourcentage de ces minutes:

- a) N'y aura-t-il aucune particule émise ?
- b) Y aura-t-il une particule émise ?
- c) Y aura-t-il deux particules ou plus émises ?

REPONSE :

On se trouve dans une distribution de Poisson : distribution discrète à deux possibilités (particules : présence vs absence), la taille d'échantillon est inconnue (on a aucune idée du nombre total de particules) et la prévalence est faible (moyenne : 1 particule par minutes).

Selon la formule de la loi de poisson:

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

$$a) P(k=0) = \frac{e^{-1} * 1^0}{0!} = 0,3679 \Rightarrow \text{en pourcentage} = 36,79\%$$

$$b) P(k=1) = \frac{e^{-1} * 1^1}{1!} = 0,3679 \Rightarrow \text{en pourcentage} = 36,79\%$$

$$c) P(k \geq 2) = 1 - P(k=1) - P(k=0) = 1 - 0,3679 - 0,3679 = 0,2642 \Rightarrow \text{en pourcentage} = 26,42\%$$

18. On ensemence une série de 20 tubes avec un même volume (0,1 ml) d'un liquide contenant des bactéries. Après incubation, 3 tubes sont restés stériles. Quelle est la concentration moyenne de ce liquide en bactéries?

REPONSE :

On se trouve dans une distribution de Poisson : distribution discrète à deux possibilités (bactéries : présence vs absence), la taille d'échantillon est inconnue (on a aucune idée du nombre total de bactéries).

Selon la formule de la loi de poisson:

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

Dans ce problème, ce que l'on cherche c'est la concentration moyenne = μ

On nous dit également que 3 tubes sur 20 sont stériles. On peut donc en déduire que la probabilité d'avoir 0 bactérie = 3/20. A partir de là, on peut retrouver la concentration moyenne :

$$P(0) = \frac{e^{-\mu} * \mu^0}{0!} = \frac{3}{20} = 0,15$$

$$\Rightarrow e^{-\mu} = 0,15$$

$$\Rightarrow -\mu = \ln(0,15)$$

$$\Rightarrow \mu = -\ln(0,15)$$

$$\Rightarrow \mu = 1,897$$

On a donc en moyenne 1,897 bactéries par 0,1ml de liquide.

19. Si la probabilité qu'un individu soit victime d'un choc sérique après injection d'un sérum donné est de 0,001, déterminez la probabilité que, parmi 2000 individus traités, soient victimes de ce choc:

- a) exactement 3 individus.**
- b) plus de 2 individus.**

REPONSE :

Nous avons deux possibilités pour résoudre cet exercice :

On se trouve dans une distribution discrète à 2 possibilités (choc sérique : oui vs non), la taille d'échantillon est connue et grande : $N = 2000$, la prévalence est petite : 0,001.

On peut donc utiliser la distribution binomiale ou la distribution de poisson (N grand et prévalence petite) :

- Loi de Poisson :

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

La moyenne d'une distribution discrète à deux possibilités = $n * p = 2000 * 0,001 = 2$

$$\text{a. } P(k = 3) = \frac{e^{-2} * 2^3}{3!} = \mathbf{0,1804}$$

$$\text{b. } P(k > 2) = 1 - P(k = 0) - P(k = 1) - P(k = 2)$$

$$P(k = 0) = \frac{e^{-2} * 2^0}{0!} = 0,135335$$

$$P(k = 1) = \frac{e^{-2} * 2^1}{1!} = 0,27067$$

$$P(k = 2) = \frac{e^{-2} * 2^2}{2!} = 0,27067$$

$$\Rightarrow P(k > 2) = 1 - P(k = 0) - P(k = 1) - P(k = 2) = 1 - 0,135335 - 0,27067 - 0,27067 = \mathbf{0,3233}$$

OU

- Loi binomiale :

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$\text{Pour rappel : } C_n^r = \frac{n!}{r! * (n-r)!}$$

$$\text{a. } P(r = 3) = C_{2000}^3 * 0,001^3 * (1 - 0,001)^{2000-3} = \mathbf{0,1805}$$

$$b. P(r > 2) = 1 - P(r = 0) - P(r = 1) - P(r = 2)$$

$$P(r = 0) = C_{2000}^0 * 0,001^0 * (1 - 0,001)^{2000-0} = 0,135199$$

$$P(r = 1) = C_{2000}^1 * 0,001^1 * (1 - 0,001)^{2000-1} = 0,2706705$$

$$P(r = 2) = C_{2000}^2 * 0,001^2 * (1 - 0,001)^{2000-2} = 0,2706705$$

$$\Rightarrow P(k > 2) = 1 - P(k = 0) - P(k = 1) - P(k = 2) = 1 - 0,135335 - 0,27067 - 0,27067 = \mathbf{0,3233}$$

20. Voici le poids de 20 chiens de race Berger Allemand, tous sexes confondus, exprimés en kg :

2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	3	3
9	8	0	5	5	3	1	0	6	7	8	7	5	3	1	9	8	8	4	5

On demande de calculer :

- la somme des écarts au carré
- la moyenne arithmétique
- la moyenne géométrique
- la moyenne harmonique
- la variance
- la déviation standard
- l'erreur standard
- le coefficient de variation
- la médiane
- le mode
- le 3e quartile
- le maximum
- le minimum
- l'étendue

REPONSE :

$$a) \sum (X - \bar{X})^2 = (29 - 32,6)^2 + (29 - 32,6)^2 + (30 - 32,6)^2 + (35 - 32,6)^2 + (35 - 32,6)^2 + (33 - 32,6)^2 + (31 - 32,6)^2 + (30 - 32,6)^2 + (36 - 32,6)^2 + (37 - 32,6)^2 + (38 - 32,6)^2 + (37 - 32,6)^2 + (35 - 32,6)^2 + (33 - 32,6)^2 + (31 - 32,6)^2 + (29 - 32,6)^2 + (28 - 32,6)^2 + (28 - 32,6)^2 + (34 - 32,6)^2 + (35 - 32,6)^2 = 212,8$$

$$b) \frac{\sum X}{N} = \frac{(29+28+30+35+35+33+31+30+36+37+38+37+35+33+31+29+28+28+34+35)}{20} = 32,6$$

$$c) \sqrt[20]{\prod X} = \sqrt[20]{29 * 28 * 30 * 35 * 35 * 33 * 31 * 30 * 36 * 37 * 38 * 37 * 35 * 33 * 31 * 29 * 28 * 28 * 34 * 35} = 32,4358$$

$$d) \frac{1}{MH} = \frac{1}{N} * \sum \frac{1}{X} = \frac{1}{20} * \left(\frac{1}{29} + \frac{1}{28} + \frac{1}{30} + \frac{1}{35} + \frac{1}{35} + \frac{1}{33} + \frac{1}{31} + \frac{1}{30} + \frac{1}{36} + \frac{1}{37} + \frac{1}{38} + \frac{1}{37} + \frac{1}{35} + \frac{1}{33} + \frac{1}{31} + \frac{1}{29} + \frac{1}{28} + \frac{1}{28} + \frac{1}{34} + \frac{1}{35} \right) = 0,030987$$

$$\Rightarrow MH = \frac{1}{0,030987} = 32,2715$$

$$e) s^2 = \frac{\sum (X - \bar{X})^2}{n-1} = \frac{212,8}{20-1} = 11,2$$

$$f) s = \sqrt{s^2} = \sqrt{11,2} = 3,3466$$

$$g) S_{\bar{X}} = \sqrt{\frac{s^2}{N}} = \sqrt{\frac{11,2}{20}} = 0,7483$$

$$h) CV = \frac{s}{\bar{X}} = \frac{3,3466}{32,6} * 100 = 10,2656\%$$

- i) Médiane : position de la valeur = $(20+1)/2 = 10,5$. La médiane est la valeur au milieu de la 10^{ème} et 11^{ème} valeur. On doit donc classer les valeurs par ordre croissant:

28 – 28 – 28 – 29 – 29 – 30 – 30 – 31 – 31 – **33** – **33** – 34 – 35 – 35 – 35 – 35 – 36 – 37 – 37 – 38

La 10^{ème} valeur = 33 ; la 11^{ème} valeur = 33 → La médiane = $(33+33)/2 = 33$

- j) Mode = valeur la plus fréquente. On voit au point i que l'on a 4 fois la valeur 35. C'est donc 35 le mode.

- k) 3^{ème} quartile = 75% des valeurs ont une valeur inférieure à celle-là. Pour trouver la position de la valeur = $(20+1)*0,75 = 15,75$. On doit donc classer les valeurs par ordre croissant:

28 – 28 – 28 – 29 – 29 – 30 – 30 – 31 – 31 – 33 – 33 – 34 – 35 – 35 – **35** – **35** – 36 – – – 37 – – – 37 – – – 38

On voit au point i que la 15^{ème} valeur = 35 ; la 16^{ème} = 35 → Le 3^{ème} quartile = 35

- l) Le maximum est 38 (voir classement par ordre croissant aux points i et k).

- m) Le minimum est 28 (voir classement par ordre croissant aux points i et k).

- n) Etendue = maximum – minimum = $38-28 = 10$

21. Voici un diagramme. Décrivez-le et réalisez le box plot de manière simplifiée (quartiles, médiane, moyenne, étendue, minimum & maximum).

8	5
8	00
7	6677
7	33444
6	66688889
6	223334
5	55667
5	33344
4	688
4	0112
3	
3	12

REPONSE :

La colonne de gauche correspond aux unités et la colonne de droite au décimale. Dans la colonne de droite, un chiffre = un individu. Dans ce graphique on a donc 45 données (car 45 chiffres dans la colonne de droite). La valeur maximum est la valeur la plus haute dans le graphique = 8,5. La valeur minimum est la valeur la plus basse = 3,1.

Etendue = Maximum – minimum = 8,5-3,1 = 5,4

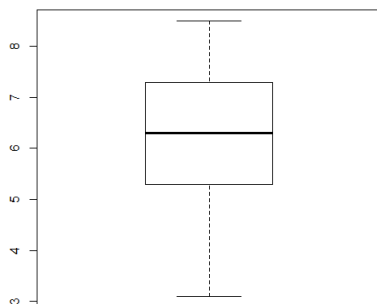
Moyenne = $\frac{\sum X}{N} =$

$\left(\begin{array}{l} 8,5 + 8 + 8 + 7,6 + 7,6 + 7,7 + 7,7 + 7,3 + 7,3 + 7,4 + 7,4 + 7,4 + 6,6 + 6,6 + 6,6 + \\ 6,8 + 6,8 + 6,8 + 6,8 + 6,9 + 6,2 + 6,2 + 6,3 + 6,3 + 6,3 + 6,4 + 5,5 + 5,5 + \\ 5,6 + 5,6 + 5,7 + 5,3 + 5,3 + 5,3 + 5,4 + 5,4 + 4,6 + 4,8 + 4,8 + 4 + 4,1 + 4,1 + 4,2 + 3,1 + 3,2 \end{array} \right) / 45$
6,11

Quartile 1 : Position de la valeur = $(n+1)/4 = 46/4 = 11,5$. Il faut trouver la valeur entre la 11^{ème} et la 12^{ème} position. 11^{ème} valeur = 5,3 ; 12^{ème} valeur = 5,3 → Quartile 1 = 5,3

Médiane : Position de la valeur = $(n+1)/2 = 46/2 = 23$. La 23^{ème} valeur est 6,3 → Médiane = 6,3

Quartile 3 : Position de la valeur = $(n+1)/4 * 3 = 34,4$. Il faut trouver la valeur entre la 34^{ème} et la 35^{ème} position. 34^{ème} valeur = 7,3 ; 35^{ème} valeur = 7,3 → Quartile 3 = 7,3



22. Considérons une expérience au cours de laquelle un dé et une pièce de monnaie sont jetés successivement (dans cet ordre). Etablissez la liste des événements possibles et attribuez-leur une probabilité. Déterminez la probabilité de chacun des événements suivants :

- a) face 6 du dé suivie de face sur la pièce
- b) chiffre pair sur le dé suivi de pile sur la pièce
- c) chiffre pair sur le dé
- d) pile sur la pièce de monnaie
- e) pile ou 6 sur le dé

REPONSE :

Liste des événements possibles :

Dé	Pièce	Probabilité
1	Pile	$=1/6 * 1/2 = 1/12$
2	Pile	$=1/6 * 1/2 = 1/12$
3	Pile	$=1/6 * 1/2 = 1/12$
4	Pile	$=1/6 * 1/2 = 1/12$
5	Pile	$=1/6 * 1/2 = 1/12$
6	Pile	$=1/6 * 1/2 = 1/12$
1	Face	$=1/6 * 1/2 = 1/12$
2	Face	$=1/6 * 1/2 = 1/12$
3	Face	$=1/6 * 1/2 = 1/12$
4	Face	$=1/6 * 1/2 = 1/12$
5	Face	$=1/6 * 1/2 = 1/12$
6	Face	$=1/6 * 1/2 = 1/12$

Le lancement du dé et de la pièce sont deux événements indépendants. La probabilité de l'un ET la probabilité de l'autre = le produit de leur probabilité.

- a) A partir du tableau, on voit qu'il n'y a qu'une possibilité. La probabilité = $P(6)*P(\text{face}) = 1/6*1/2 = 1/12$
- b) A partir du tableau, on voit qu'il y a 3 possibilités : $3*(1/12) = 3/12$
OU $P(\text{pair})*P(\text{pile}) = 3/6*1/2 = 3/12$
- c) A partir du tableau, on voit qu'il y a 6 possibilités : $6*(1/12) = 6/12 = 1/2$
OU $P(\text{pair}) = 3/6 = 1/2$
- d) A partir du tableau, on voit qu'il y a 6 possibilités : $6*(1/12) = 6/12 = 1/2$
OU $P(\text{pile}) = 1/2$
- e) A partir du tableau, on voit qu'il y a 7 possibilités : $7*(1/12) = 7/12$

OU le lancement du dé et de la pièce ne sont pas des événements mutuellement exclusifs.

Donc :

$$P(\text{pile}) + P(6) - P(\text{pile ET } 6) = \frac{1}{2} + \frac{1}{6} - (\frac{1}{2} * \frac{1}{6}) = \frac{8}{12} - \frac{1}{12} = \frac{7}{12}$$

23. Un particulier élève 15 lapins. Ne voulant pas vacciner ses lapins contre la myxomatose, il pense qu'en examinant deux lapins pris au hasard tous les jours, il détectera suffisamment tôt la maladie. La prévalence de la myxomatose dans sa région est de 0,13. Si son élevage est contaminé, après combien de jours détectera-t-il la maladie au seuil 99%?

REPONSE :

Solution simplifiée :

On suppose que la prévalence est 0,13, et donc la probabilité de prendre un lapin malade est $P(M) = 0,13$, et celle de prendre un lapin sain est $P(S) = 0,87$. Par conséquent, si on prend 2 lapins au hasard, la probabilité qu'ils soient sains est $P(SS) = 0,87^2$. La probabilité de non-détection après 1 jour est $1 - P(MS) - P(SM) - P(MM) = P(SS) = 0,87^2 = 0,7569$. Si on répète cela les jours suivants, la probabilité de ne pas détecter après n jours est $0,87^{2*n}$. Il faut choisir n pour que cette probabilité devienne inférieure à 0,01 : $0,87^{2*n} < 0,01 \Rightarrow 2*n*\log_{10}(0,87) < -2 \Rightarrow n > -2/(2*\log_{10}(0,87)) \Rightarrow n > 16,53$, soit $n = 17$ jours (qui est la réponse fournie dans le corrigé).

Solution plus complexe :

a) L'élevage est contaminé : il y a donc au moins 1 lapin contaminé dans l'élevage. La distribution est donc celle d'une distribution binomiale tronquée (notée BT), car la probabilité binomiale $B(0)$ d'avoir 0 atteint, qui serait non nulle selon une binomiale, doit être mise à 0. Comme la somme des probabilités de la distribution doit valoir 1, on divise toutes les probabilités par $1-B(0)$. La distribution est donc la suivante :

r	B[r]	BT[r]
0	0,12381943	0
1	0,2775263	0,31674555
2	0,29028613	0,33130857
3	0,18796305	0,21452547
4	0,0842593	0,09616659
5	0,02769903	0,03161338
6	0,00689823	0,00787307
7	0,00132528	0,00151256
8	0,00019803	0,00022601
9	2,3015E-05	2,6267E-05
10	2,0634E-06	2,355E-06
11	1,4015E-07	1,5995E-07
12	6,9805E-09	7,967E-09
13	2,4071E-10	2,7472E-10

14	5,1383E-12	5,8644E-12
15	5,1186E-14	5,8419E-14

b) Quelle est la probabilité de tirer 2 lapins sains $P(SS)$ dans l'exploitation ? Cela dépend du nombre de lapins malades dans l'exploitation... C'est évidemment beaucoup plus probable quand il y a peu de lapins malades que quand il y en a beaucoup ! On calculera cette probabilité comme suit :

$$P(SS) = P(SS, r=0) + P(SS, r=1) + \dots + P(SS, r=15)$$

$$= P(SS|r=0)*P(r=0) + P(SS|r=1)*P(r=1) + \dots + P(SS|r=15)*P(r=15)$$

$$= P(SS|r=1)*BT[1] + \dots + P(SS|r=15)*BT(15)$$

$P(SS|r=1)$ est la probabilité hypergéométrique de prendre 2 sains parmi les 14 sains (et 0 malade parmi le seul malade) $\Rightarrow C(2,14)*C(0,1)/C(2,15) = 13/15 = 0.8667$

$$P(SS|r=2) = C(2,13)*C(0,2)/C(2,15) = 0.7429$$

...

$$P(SS|r=13) = C(2,2)*C(0,13)/C(2,15) = 0.0095$$

$$P(SS|r=14) = P(SS|r=15) = 0 \text{ (il y a moins de 2 sains dans l'exploitation).}$$

Les calculs (avec excel) sont reproduits ci-dessous :

r	B[r]	BT[r]	Hypergeom	Produit
0	0,12381943	0	1	0
1	0,2775263	0,31674555	0,86666667	0,27451281
2	0,29028613	0,33130857	0,74285714	0,24611494
3	0,18796305	0,21452547	0,62857143	0,13484458
4	0,0842593	0,09616659	0,52380952	0,05037298
5	0,02769903	0,03161338	0,42857143	0,01354859
6	0,00689823	0,00787307	0,34285714	0,00269934
7	0,00132528	0,00151256	0,26666667	0,00040335
8	0,00019803	0,00022601	0,2	4,5203E-05
9	2,3015E-05	2,6267E-05	0,14285714	3,7525E-06
10	2,0634E-06	2,355E-06	0,0952381	2,2429E-07
11	1,4015E-07	1,5995E-07	0,05714286	9,1402E-09
12	6,9805E-09	7,967E-09	0,02857143	2,2763E-10
13	2,4071E-10	2,7472E-10	0,00952381	2,6164E-12
14	5,1383E-12	5,8644E-12	0	0
15	5,1186E-14	5,8419E-14	0	0

Somme 0,72254578

Donc, la probabilité de non-détection après 1 jour est 0,7225. Le même raisonnement que ci-dessus donne alors :

$n > -2/(\log_{10}(0,7225)) \Rightarrow n > 14,17$, soit $n = 15$ jours.

Ce résultat est plus cohérent avec l'énoncé, mais beaucoup moins facile que la première résolution...

24 Dans une exploitation laitière donnée, les vaches sont réformées après 6 lactations. L'exploitant souhaite savoir combien de veaux femelles chaque vache a donné au cours de sa vie dans l'exploitation. Il effectue donc une étude rétrospective sur les 10 dernières années, uniquement avec les vaches ayant eu 6 lactations. Il obtient les résultats suivants :

Nb veaux femelles sur les 6 lactations	Effectifs observés
0	4
1	12
2	25
3	30
4	19
5	8
6	2

- Sachant que le sex ratio vaut 50%, on demande de calculer la distribution binomiale théorique
- Calculez les moyenne, valeur p, variance et déviation standard observées
- Calculez les moyenne, variance et déviation standard théoriques
- Combien de veaux auront été produits par ces vaches ?

REPONSE :

a)

On se trouve dans une distribution binomiale : distribution discrète à deux possibilités (veaux femelles : oui vs non), la taille d'échantillon est connue ($N=6$) et la prévalence est de 0,50.

$$P(r) = C_n^r * p^r * q^{n-r}$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

Nb veaux femelles sur les 6 lactations	Effectifs observés	Probabilité théorique	Effectifs théoriques
0	4	$C_6^0 * 0,5^0 * (1 - 0,5)^{6-0} = 0,015625$	=Proba théorique*Nbre vache = 0,015625*100 = 1,56
1	12	$C_6^1 * 0,5^1 * (1 - 0,5)^{6-1} = 0,09375$	=0,09375*100 = 9,375
2	25	$C_6^2 * 0,5^2 * (1 - 0,5)^{6-2} = 0,234375$	= 0,234375*100= 23,4375
3	30	$C_6^3 * 0,5^3 * (1 - 0,5)^{6-3} = 0,3125$	=0,3125*100 = 31,25
4	19	$C_6^4 * 0,5^4 * (1 - 0,5)^{6-4} = 0,234375$	=0,234375*100 = 23,4375
5	8	$C_6^5 * 0,5^5 * (1 - 0,5)^{6-5} = 0,09375$	=0,09375*100 = 9,375
6	2	$C_6^6 * 0,5^6 * (1 - 0,5)^{6-6} = 0,015625$	=0,015625*100 = 1,5625

Nbre de vache = somme effectifs observés = 4+12+25+30+19+8+2 = 100

$$b) \text{ Moyenne } = \bar{X} = \frac{\sum f * X}{\sum f} = \frac{(4*0)+(12*1)+(25*2)+(30*3)+(19*4)+(8*5)+(2*6)}{(4+12+25+30+19+8+2)} = \frac{280}{100} = 2,8$$

$$\text{Variance} = s^2 = \frac{\sum f * X^2 - \frac{(\sum f * X)^2}{\sum f}}{\sum f - 1} = \frac{(4*0^2)+(12*1^2)+(25*2^2)+(30*3^2)+(19*4^2)+(8*5^2)+(2*6^2) - \frac{280^2}{100}}{100-1} = 1,7575$$

$$\text{Déviation standard} = s = \sqrt{s^2} = \sqrt{1,7575} = 1,3257$$

$$c) \text{ Moyenne } = \mu = \sum P_{th} * X = (0,015625 * 0) + (0,09375 * 1) + (0,234375 * 2) + (0,3125 * 3) + (0,234375 * 4) + (0,09375 * 5) + (0,015625 * 6) = 3$$

$$\text{Variance} = \sigma^2 = \sum P_{th} * (X - \mu)^2 = 0,015625 * (0 - 3)^2 + 0,09375 * (1 - 3)^2 + 0,234375 * (2 - 3)^2 + 0,3125 * (3 - 3)^2 + 0,234375 * (4 - 3)^2 + 0,09375 * (5 - 3)^2 + 0,015325 * (3 - 6)^2 = 1,50$$

$$\text{Déviation standard} = \sigma = \sqrt{\sigma^2} = \sqrt{1,50} = 1,224744$$

$$d) \text{ Nombre de veaux produits} = \sum f * X = (4 * 0) + (12 * 1) + (25 * 2) + (30 * 3) + (19 * 4) + (8 * 5) + (2 * 6) = 280$$

25. Un saumon sur 5000, capturés en Alaska, présente des parasites qui le rendent impropre à la consommation humaine. Trouvez la probabilité que, dans un bateau ayant pêché 1800 poissons, maximum deux poissons soient détruits suite à cette infestation.

REPONSE :

Pour résoudre cet exercice, vous avez le choix entre deux distributions :

- Distribution binomiale : distribution discrète à deux possibilités (parasite : présence vs absence), la taille d'échantillon est connue (N=1800) et la prévalence est de 1/5000.

$$P(r) = C_n^r * p^r * q^{n-r}$$

Pour rappel : $C_n^r = \frac{n!}{r!(n-r)!}$

La probabilité d'avoir maximum 2 poissons atteints = $P(r \leq 2) = P(r=0) + P(r=1) + P(r=2)$

$$P(r = 0) = C_{1800}^0 * \frac{1}{5000}^0 * \left(1 - \frac{1}{5000}\right)^{1800-0} = 0,697651$$

$$P(r = 1) = C_{1800}^1 * \frac{1}{5000}^1 * \left(1 - \frac{1}{5000}\right)^{1800-1} = 0,25120$$

$$P(r = 2) = C_{1800}^2 * \frac{1}{5000}^2 * \left(1 - \frac{1}{5000}\right)^{1800-2} = 0,0452$$

$$P(r \leq 2) = 0,697651 + 0,25120 + 0,0452 = \mathbf{0,994056}$$

- Distribution de Poisson : distribution discrète à deux possibilités (parasite : présence vs absence), la taille d'échantillon est grande (N=1800) et la prévalence est très petite (p=1/5000).

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

La moyenne d'une distribution discrète à deux possibilités = $n * p = 1800 * (1/5000) = 0,36$

La probabilité d'avoir maximum 2 poissons atteints = $P(k \leq 2) = P(k=0) + P(k=1) + P(k=2)$

$$P(k = 0) = \frac{e^{-0,36} * 0,36^0}{0!} = 0,697676$$

$$P(k = 1) = \frac{e^{-0,36} * 0,36^1}{1!} = 0,25116$$

$$P(k = 2) = \frac{e^{-0,36} * 0,36^2}{2!} = 0,0452$$

$$P(k \leq 2) = 0,697676 + 0,25116 + 0,0452 = \mathbf{0,99049}$$

26. Voici une table montrant la distribution des probabilités de vente de chevaux par mois dans un haras national :

Nb de chevaux vendus / mois	Probabilité
0	0,05
1	0,05
2	0,10
3	0,15
4	0,20
5	0,15
6	0,15
7	0,10
8	0,05

- Montrez que les probabilités ci-dessus correspondent bien à une distribution correcte
- Générez la distribution cumulée du nombre de chevaux vendus par mois
- En utilisant cette distribution cumulée, trouvez la probabilité que le nombre de chevaux vendus soit d'au moins 3 et moins de 7 en un mois
- Trouvez la probabilité que maximum 5 chevaux soient vendus en un mois
- Trouvez la variance et la déviation standard de cette distribution

REPONSE :

- Somme des probabilités =
 $0,05+0,05+0,10+0,15+0,20+0,15+0,15+0,10+0,05 = 1$
-

Nb de chevaux vendus / mois	Probabilité	Probabilité cumulée
0	0,05	0,05
1	0,05	= 0,05+0,05 = 0,10
2	0,10	= 0,10+0,10 = 0,20
3	0,15	= 0,20+0,15 = 0,35
4	0,20	=0,35+0,20 = 0,55
5	0,15	= 0,55+0,15 = 0,70
6	0,15	=0,70+0,15 = 0,85
7	0,10	= 0,85+0,10 = 0,95
8	0,05	=0,95 + 0,05 = 1,00

- $P(x \geq 3 \text{ ET } x < 7) = P(X=3)+P(X=4)+P(X=5)+P(X=6)=$
 $0,15+0,20+0,15+0,15 = 0,65$

OU en utilisant les probabilités cumulées :
 $P(X \leq 6) - P(X \leq 2) = 0,85 - 0,20 = 0,65$

- $P(X \leq 5) = P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5) =$
 $0,05+0,05+0,10+0,15+0,20+0,15 = 0,70$

OU en utilisant les probabilités cumulées :

$$P(X \leq 5) = 0,70$$

e) $\text{Variance} = \sigma^2 = \sum P * (X - \mu)^2$

$$\mu = \sum P * X = (0,05 * 0) + (0,05 * 1) + (0,10 * 2) + (0,15 * 3) + (0,20 * 4) + (0,15 * 5) + (0,15 * 6) + (0,10 * 7) + (0,05 * 8) = 4,25$$

$$\sigma^2 = \sum P * (X - \mu)^2 = 0,05 * (0 - 4,25)^2 + 0,05 * (1 - 4,25)^2 + 0,10 * (2 - 4,25)^2 + 0,15 * (3 - 4,25)^2 + 0,20 * (4 - 4,25)^2 + 0,15 * (5 - 4,25)^2 + 0,15 * (6 - 4,25)^2 + 0,10 * (7 - 4,25)^2 + 0,05 * (8 - 4,25)^2 = 4,1875$$

Déviation standard $\sqrt{\sigma^2} = \sqrt{4,1875} = 2,046$

27. De plus en plus d'employeurs utilisent le test psychologique comme une aide pour déterminer s'il existe de bonnes chances d'adaptation d'un candidat au sein de l'entreprise. Les données suivantes représentent les scores obtenus lors du test psychologique par un groupe de candidats.

Les scores sont groupés par classes :

Score	Nombre
41-50	5
51-60	7
61-70	10
71-80	16
81-90	11
91-100	9

On demande :

- a) le type de variables
- b) la moyenne
- c) la variance
- d) la déviation standard
- e) la classe modale
- f) la médiane
- g) le 1er quartile
- h) les fréquences relatives
- i) les fréquences cumulées

REPONSE :

- a) score = variable continue rendue discrète
Nombre = variable continue
- b) Comme on est avec une classe, la valeur de X = la moyenne de l'intervalle

$$\bar{X} = \frac{\sum f \cdot X}{\sum f} = \frac{(5 \cdot 45,5) + (7 \cdot 55,5) + (10 \cdot 65,5) + (16 \cdot 75,5) + (11 \cdot 85,5) + (9 \cdot 95,5)}{(5+7+10+16+11+9)} = \frac{4279}{58} = 73,7759$$

$$c) s^2 = \frac{\sum f \cdot X^2 - \frac{(\sum f \cdot X)^2}{\sum f}}{\sum f - 1} = \frac{(5 \cdot 45,5^2) + (7 \cdot 55,5^2) + (10 \cdot 65,5^2) + (16 \cdot 75,5^2) + (11 \cdot 85,5^2) + (9 \cdot 95,5^2) - \frac{4279^2}{58}}{58-1} = 225,0437$$

$$d) s = \sqrt{s^2} = \sqrt{225,0437} = 15,001$$

e) Classe modale = classe avec la fréquence la plus élevée. C'est la classe 71-80 qui a une fréquence de 16.

f) Position de la médiane = $(n+1)/2 = (58+1)/2 = 29,5$. La médiane se trouve donc entre la 29^{ème} et la 30^{ème} valeur (dans l'ordre croissant). La 29^{ème} et 30^{ème} valeur sont dans la classe 71-80. La médiane est donc 71-80.

g) Position du 1^{er} quartile (= 25% de la distribution a une valeur plus faible) = $(n+1)/4 = (58+1)/4 = 14,75$. Le 1^{er} quartile se trouve donc entre la 14^{ème} et la 15^{ème} valeur (dans l'ordre croissant). La 14^{ème} et 15^{ème} valeur sont dans la classe 61-70. Le 1^{er} quartile est donc 61-70.

h) Fréquence relative = fréquence absolue/taille échantillon

Score	Nombre	Fréquences relative
41-50	5	5/58 = 0,0862
51-60	7	7/58 = 0,1207
61-70	10	10/58 = 0,1724
71-80	16	16/58 = 0,2759
81-90	11	11/58 = 0,1896
91-100	9	9/58 = 0,1552

i)

Score	Nombre	Fréquences cumulées
41-50	5	5
51-60	7	=5+7 = 12
61-70	10	=12 + 10 = 22
71-80	16	=22 + 16 = 38
81-90	11	=38 + 11 = 49
91-100	9	=49 + 9 = 58

28. Voici un diagramme. Donnez-en le nombre de valeurs, la somme des valeurs, la moyenne et la variance.

6	5
6	00
5	6677
5	33444
4	66688889
4	223334
3	55667
3	33344
2	688
2	0112
1	
1	12

REPONSE :

La colonne de gauche correspond aux unités et la colonne de droite au décimale. Dans la colonne de droite, un chiffre = un individu. Dans ce graphique on a donc 45 données (car 45 chiffres dans la colonne de droite). La valeur maximum est la valeur la plus haute dans le graphique = 6,5. La valeur minimum est la valeur la plus basse = 1,1.

Nombre de valeurs = 45 = nombres de chiffres dans la colonne de droite.

Somme =

6,5+6+6+5,6+5,6+5,7+5,7+5,3+5,3+5,4+5,4+5,4+4,6+4,6+4,6+4,8+4,8+4,8+4,8+4,9+4,2+4,2+4,3+4,3+4,3+4,4+3,5+3,5+3,6+3,6+3,7+3,3+3,3+3,3+3,3+3,4+3,4+2,6+2,8+2,8+2,0+2,1+2,1+2,2+1,1+1,2 = 185

$$\text{Moyenne} = \bar{X} = \frac{\sum X}{N} = \frac{185}{45} = 4,111$$

$$\begin{aligned} \text{Variance} = s^2 = \frac{\sum (X - \bar{X})^2}{N-1} = & [(6,5 - 4,11)^2 + 2 * (6 - 4,11)^2 + 2 * (5,6 - 4,11)^2 + 2 * \\ & (5,7 - 4,11)^2 + 2 * (5,3 - 4,11)^2 + 3 * (5,4 - 4,11)^2 + 3 * (4,6 - 4,11)^2 + 4 * \\ & (4,8 - 4,11)^2 + (4,9 - 4,11)^2 + 2 * (4,2 - 4,11)^2 + 3 * (4,3 - 4,11)^2 + (4,4 - 4,11)^2 + \\ & 2 * (3,5 - 4,11)^2 + 2 * (3,6 - 4,11)^2 + (3,7 - 4,11)^2 + 3 * (3,3 - 4,11)^2 + 2 * \\ & (3,4 - 4,11)^2 + (2,6 - 4,11)^2 + 2 * (2,8 - 4,11)^2 + (2,0 - 4,11)^2 + 2 * (2,1 - 4,11)^2 + \\ & (2,2 - 4,11)^2 + (1,1 - 4,11)^2 + (1,2 - 4,11)^2] / (45 - 1) = 1,76 \end{aligned}$$

29. Voici 20 cadres d'une entreprise dont on note l'âge et le sexe : 34F, 49M, 27M, 63F, 33F, 29F, 45M, 46M, 30F, 39M, 42M, 30F, 48M, 35F, 32F, 37F, 48F, 50M, 48F, 61F. Quelle est la probabilité de choisir un cadre de sexe féminin ou de plus de 50 ans ? Ou répondant aux 2 critères ? Si deux cadres doivent être sélectionnés pour diriger ensemble un nouveau service, quelle est la probabilité que ce soit une femme de plus de 50 ans et un homme ?

REPONSE :

On va faire un tableau reprenant le croisement de la variable sexe (M vs F) et 50 ans (plus vs moins) :

	F	M	TOTAL
+ de 50 ans	2	0	2
- de 50 ans	10	8	18
TOTAL	12	8	20

Probabilité de choisir un cadre de sexe féminin ou de plus de 50 ans :

- A partir du tableau : tous les cas possibles de femme ou plus de 50 ans = 2 + 10 + 0 = 12 => probabilité = 12/20
- Avec les axiomes de probabilités : $P(F \text{ ou } + \text{ de } 50 \text{ ans}) = P(F) + P(+50 \text{ ans}) - P(F \text{ ET } + \text{ de } 50 \text{ ans}) = 12/20 + 2/20 - (P(F) * P(+50 \text{ ans} | F) = 12/20 + 2/20 - (12/20 * 2/12) = 12/20$

Probabilité de choisir un cadre de sexe féminin ET de plus de 50 ans :

- A partir du tableau : cela correspond à la case 2 => 2/20
- Avec les axiomes de probabilités : $P(F \text{ ET } + 50 \text{ ans}) = P(F) * P(+ 50 \text{ ans} | F) = 12/20 * 2/12 = 2/20$

Probabilité de choisir une femme de plus de 50 ans et un homme :

On a deux cas possibles : on choisit une femme de plus de 50 ans puis un homme OU on choisit un homme puis une femme de plus de 50 ans.

Soit on résout directement via les probabilités :

$$\Leftrightarrow P(F + 50 \text{ ans} ; 1^{\text{er}} \text{ tirage ET homme} ; 2^{\text{ème}} \text{ tirage}) \text{ OU } P(\text{homme} ; 1^{\text{er}} \text{ tirage ET } F + \text{ans} ; 2^{\text{ème}} \text{ tirage}) = 2/20 * 8/19 + 8/20 * 2/19 = 0,08421$$

Soit on résout via la loi hypergéométrique :

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 10, constitué de 2 sous-ensembles : N_1 = nombre total de femmes de plus de 50 ans = 2 et N_2 = nombre total d'hommes = 8. a_1 = nombre de femmes de plus de 50 ans qu'on souhaite choisir = 1 ; a = nombre d'individus total que l'on choisit = 2 ; $a_2 = a - a_1 = 1$.

$$P(1; 2; 2; 20) = \frac{C_2^1 * C_8^1}{C_{20}^2} = \mathbf{0,08421}$$

30. Une pièce de monnaie est jetée 3 fois. Quelle est la probabilité d'obtenir deux faces et un pile ? Quelle est la probabilité d'obtenir deux faces suivies d'un pile ? Etablissez la distribution binomiale et calculez la moyenne et la déviation standard de cette distribution.

REPONSE :

a.) Probabilité d'obtenir deux faces et un pile :

On peut utiliser la distribution binomiale : distribution discrète à 2 possibilités (résultat pièce : pile vs face), taille d'échantillon connue : $N=3$; prévalence connue : $p=0,5$

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$\text{Pour rappel : } C_n^r = \frac{n!}{r!(n-r)!}$$

$$P(r = 2) = C_3^2 * 0,5^2 * (1 - 0,5)^{3-2} = \mathbf{0,375}$$

b.) Probabilité d'obtenir deux faces suivies d'un pile :

Ici, il y a une notion d'ordre. Il n'y a donc qu'un cas possible : Face – Face – Pile

$$P(\text{Face}) = 0,5$$

$$P(\text{Pile}) = 0,5$$

$$\Rightarrow P(\text{Face}) * P(\text{Face}) * P(\text{Pile}) = 0,5 * 0,5 * 0,5 = \mathbf{0,125}$$

c.) Distribution binomiale :

Voici tous les cas possibles :

- | | |
|-------|-------|
| - PPP | - PFF |
| - PPF | - FPF |
| - PFP | - FFP |
| - FPP | - FFF |

$\Rightarrow$ 8 possibilités au total

Cas possible :	Nombre de Faces	Probabilité
PPP	0	1/8
PPF ou PFP ou FPP	1	3/8
PFF ou FPF ou FFP	2	3/8
FFF	3	1/8

$$\text{Moyenne : } \mu = \sum P * X = \left(\frac{1}{8} * 0\right) + \left(\frac{3}{8} * 1\right) + \left(\frac{3}{8} * 2\right) + \left(\frac{1}{8} * 3\right) = 1,5 \text{ Face}$$

$$\text{Déviation standard : } \sigma = \sqrt{\sum P * (X - \bar{X})^2} = \sqrt{\frac{1}{8} * (0 - 1,5)^2 + \frac{3}{8} * (1 - 1,5)^2 + \frac{3}{8} * (2 - 1,5)^2 + \frac{1}{8} * (3 - 1,5)^2} = 0,866$$

31. Calculer la probabilité de ne pas détecter la présence de chiens dysplasiques de la hanche dans un chenil de 25 Bergers allemands, sachant que la prévalence de la maladie dans cette race est de 0,15.

REPONSE :

On peut utiliser la distribution binomiale : distribution discrète à 2 possibilités (maladie : présence vs absence), taille d'échantillon connue : N=25 ; prévalence connue : p=0,15

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$\text{Pour rappel : } C_n^r = \frac{n!}{r! * (n-r)!}$$

$$P(r = 0) = C_{25}^0 * 0,15^0 * (1 - 0,15)^{25-0} = 0,0171978$$

32. On sait que la tuberculose aviaire a une fréquence de 1/5000 en Wallonie. Dans un élevage de 30000 volailles, calculez de deux manières différentes la probabilité d'avoir une volaille atteinte dans l'élevage.

REPONSE :

Pour résoudre cet exercice, vous avez le choix entre deux distributions :

- Distribution binomiale : distribution discrète à deux possibilités (tuberculose aviaire : présence vs absence), la taille d'échantillon est connue (N=3000) et la prévalence est de 1/5000.

$$P(r) = C_n^r * p^r * q^{n-r}$$

$$\text{Pour rappel : } C_n^r = \frac{n!}{r! * (n-r)!}$$

$$P(r = 1) = C_{3000}^1 * \frac{1}{5000} * \left(1 - \frac{1}{5000}\right)^{3000-1} = 0,0148666$$

- Distribution de Poisson : distribution discrète à deux possibilités (tuberculose aviaire : présence vs absence), la taille d'échantillon est grande (N=3000) et la prévalence est très petite (p=1/5000).

$$P(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

La moyenne d'une distribution discrète à deux possibilités = $n * p = 3000 * (1/5000) = 6$

$$P(k = 1) = \frac{e^{-6} * 6^1}{1!} = 0,0148725$$

33. Vous êtes vétérinaire rural, responsable de la médecine de troupeau de nombreuses exploitations. Vous travaillez énormément en prévention des pathologies et vous vous intéressez à une nouvelle méthode de dépistage d'une maladie respiratoire fréquente dans les élevages. La prévalence de la maladie est de 30%. Le test annonce une spécificité de 92,86% et une sensibilité de 88,36%.

- Quelle est la probabilité d'observer un individu à la fois malade et présentant un test + ? A quoi cela correspond cette probabilité d'un point de vue épidémiologique ?
- Quelle est la Valeur Prédictive Positive du test ?

RESOLUTION

Nous pouvons organiser les éléments dont nous disposons sous forme de tableau de pourcentages.

	Malades	Sains	Total
Test +	%VP	%FP	%VP + %FP
Test -	%FN	%VN	%FN + %VN
Total	%VP + %FN = 30%	%FP + %VN = 70%	100%

Avec %=pourcentage, VP = vrais positifs, FN = faux négatifs, FP = faux positifs et VN = vrais négatifs. N = effectifs totaux.

- La P(test+ et malades) correspond, épidémiologiquement parlant, au pourcentage de vrai positifs. Nous pouvons la déterminer à l'aide de la sensibilité du test. En effet,

$$\begin{aligned} \text{Sensibilité} = P(\text{test} + | \text{malades}) &= \frac{P(\text{test} + \text{ et malades})}{P(\text{malades})} = \frac{\%VP}{\%VP + \%FN} \\ &= 0,8836 \end{aligned}$$

Donc :

$$P(\text{test} + \text{et malades}) = \%VP = P(\text{malades}) * \text{sensibilité} = 0,3 * 0,8836 = 0,26508$$

Nous pouvons alors partiellement compléter le tableau :

	Malades	Sains	Total
Test +	%VP = 26,508%	%FP	%VP + %FP
Test -	%FN = 30% - 26,508% = 3,492%	%VN	%FN + %VN
Total	%VP + %FN = 30%	%FP + %VN = 70%	100%

- d) La Valeur Prédictive Positive (VPP) d'un test correspond à la probabilité d'observer un individu malade sachant que son test est positif c'est-à-dire la probabilité que la maladie soit vraiment présente lorsque le résultat du test est positif.

$$\begin{aligned} VPP = P(\text{malades} | \text{test}+) &= \frac{P(\text{test} + \text{et malades})}{P(\text{test}+)} \\ &= \frac{P(\text{test} + \text{et malades})}{P(\text{test} + \text{et malades}) + P(\text{test} + \text{et sains})} = \frac{\%VP}{\%VP + \%FP} \end{aligned}$$

Pour pouvoir calculer la VPP, nous avons besoin de connaître la P(test +). Celle-ci dépend de la P(test + et sains), c'est-à-dire du %FP. Or, %FP = (%FP+%VN)-%VN = 0,70 - %VN.

Nous savons que la spécificité du test est de 92,86%. La spécificité d'un test correspond à la probabilité d'avoir un test négatif si l'individu est sain. Cela correspond donc au taux de vrais négatifs. On peut donc la calculer de la manière suivante :

$$\begin{aligned} \text{Spécificité} = P(\text{test} - | \text{sains}) &= \frac{P(\text{test} - \text{et sains})}{P(\text{sains})} = \frac{\%VN}{\%VN + \%FP} = \frac{\%VN}{0,7} \\ &= 0,9286 \end{aligned}$$

$$\text{Donc } P(\text{test} - \text{et sains}) = \%VN = P(\text{sains}) * \text{spécificité} = 0,7 * 0,9286 = 0,65002$$

Nous pouvons dès lors terminer de compléter le tableau :

	Malades	Sains	Total
Test +	%VP = 26,508%	%FP = 70% - 65,002% = 4,998%	%VP + %FP = 31,506%
Test -	%FN = 3,492%	%VN = 65,002%	%FN + %VN = 68,494 %
Total	%VP + %FP = 30%	%FP + %VN = 70%	100%

Enfin,

$$VPP = \frac{P(\text{test} + \text{et malades})}{P(\text{test} +)} = \frac{\%VP}{\%VP + \%FP} = \frac{0,26508}{0,31506} = 0,841364$$

La VPP est donc de 84,1%.

34. Un éleveur achète un lot de 10 vaches. Par manque de temps, à l'arrivée du lot, il n'examine que 5 vaches prises au hasard pour voir s'il n'y a pas de problème particulier. Si dans le lot de 10 vaches, 2 présentent un problème quelconque, quelle est la probabilité qu'au moins une des 5 vaches examinées présente un problème ?

REPONSE :

On se trouve dans une distribution hypergéométrique : on sait exactement le nombre de vache étant atteinte et la probabilité varie à chaque tirage.

La probabilité qu'au moins une vache présente un problème = 1-probabilité que 0 vache ne présente un problème

On peut représenter la situation où aucune des vaches examinées ne présente de problème sous forme de tableau :

	Tiré	Non tiré	
Problème	0	2	2
Sain	5	3	8
	5	5	10

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 10, constitué de 2 sous-ensembles : N1 = nombre total de vaches ayant un problème = 2 et N2 = nombre total de vaches saines = 8. a1 = nombre d'individus ayant un problème qu'on choisit = 0 ; a = nombre d'individus total que l'on examine = 5 ; a2 = a-a1 = 5.

$$P(0; 2; 5; 10) = \frac{C_2^0 * C_8^5}{C_{10}^5} = 0,2222$$

Donc la probabilité d'avoir au moins une vache présentant un problème = 1 – 0,2222 = **0,7778**

35. On extrait au hasard une boule d'une boîte contenant 6 boules rouges, 4 boules blanches et 5 boules bleues. Déterminer la probabilité d'avoir une boule :

- a) Rouge.
- b) Blanche.
- c) Bleue.
- d) Pas rouge.
- e) Rouge ou blanche

Si on extrait successivement trois boules de la boîte, quelle est la probabilité pour que les trois boules soient extraites dans l'ordre rouge, blanche et bleue, quand chaque boule

- f) Est remise dans la boîte ?
- g) N'est pas remise dans la boîte ?

REPONSE :

Total de boules dans la boîte = $6+4+5=15$

a) $P(\text{Rouge}) = 6/15$

b) $P(\text{Blanche}) = 4/15$

c) $P(\text{bleue}) = 5/15$

d) $P(\text{Pas rouge}) = 1 - P(\text{rouge}) = 1 - 6/15 = 9/15$

OU

$P(\text{Pas rouge}) = P(\text{Blanche ou bleue}) = P(\text{Blanche}) + P(\text{bleue}) = 4/15 + 5/15 = 9/15$

e) $P(\text{Rouge ou Blanche}) = P(\text{Rouge}) + P(\text{Blanche}) = 6/15 + 4/15 = 10/15$

f) $P(\text{rouge}) * P(\text{blanche}) * P(\text{bleue}) = (6/15) * (4/15) * (5/15) = 0,03556$

g) La probabilité varie d'un tirage à l'autre :

$P(\text{rouge}) * P(\text{blanche} | \text{rouge au tirage précédent}) * P(\text{bleue} | \text{blanc au tirage précédent}) = (6/15) * (4/14) * (5/13) = 0,043956$

36. On joue deux fois avec un dé bien équilibré. Quelle est la probabilité d'obtenir 4, 5 ou 6 la première fois et 1, 2, 3 ou 4 la seconde fois ?

REPONSE :

1^{er} lancé : $P(4) \text{ OU } P(5) \text{ OU } P(6) = (1/6) + (1/6) + (1/6) = 3/6$

2^{ème} lancé : $P(1) \text{ OU } P(2) \text{ OU } P(3) \text{ OU } P(4) = (1/6) + (1/6) + (1/6) + (1/6) = 4/6$

$P(1^{\text{er}} \text{ lancé ET } 2^{\text{ème}} \text{ lancé}) = (3/6) * (4/6) = 0,33$

⇒ Les deux lancers sont indépendants l'un de l'autre.

37. Un sac contient 5 boules blanches et 2 boules noires ; un autre sac contient 3 boules blanches et 5 boules noires. Quand on extrait une boule dans chaque sac, quelle est la probabilité pour que

- a) Les deux boules soient blanches ?**
- b) Les deux boules soient noires ?**
- c) Une des boules soit blanche et l'autre noire ?**

REPONSE :

- a) Les deux boules soient blanches ?

$$1^{\text{er}} \text{ sac} : P(B) = 5/7$$

$$2^{\text{ème}} \text{ sac} : P(B) = 3/8$$

$$P(B | 1^{\text{er}} \text{ sac}) \text{ ET } P(B | 2^{\text{ème}} \text{ sac}) = (5/7) * (3/8) = \mathbf{0,267857}$$

- b) Les deux boules soient noires ?

$$1^{\text{er}} \text{ sac} : P(N) = 2/7$$

$$2^{\text{ème}} \text{ sac} : P(N) = 5/8$$

$$P(N | 1^{\text{er}} \text{ sac}) \text{ ET } P(N | 2^{\text{ème}} \text{ sac}) = (2/7) * (5/8) = \mathbf{0,17857}$$

- c) Une des boules soit blanche et l'autre noire ?

2 cas possible : soit blanche dans le premier sac ET noire dans le second, soit noire dans le premier sac ET blanche dans le second.

$$P(B|1^{\text{er}} \text{ sac}) \text{ ET } P(N|2^{\text{ème}} \text{ sac}) = (5/7) * (5/8) = 0,4464$$

$$P(N|1^{\text{er}} \text{ sac}) \text{ ET } P(B|2^{\text{ème}} \text{ sac}) = (2/7) * (3/8) = 0,107142$$

$$P(B|1^{\text{er}} \text{ sac}) \text{ ET } P(N|2^{\text{ème}} \text{ sac}) \text{ OU } P(N|1^{\text{er}} \text{ sac}) \text{ ET } P(B|2^{\text{ème}} \text{ sac}) = 0,4464 + 0,107142 = \mathbf{0,55357}$$

38. Un sac contient 2 boules blanches et 3 boules noires. Quatre personnes A, B, C et D tirent dans l'ordre indiqué une boule et ne la remettent pas dans le sac. La première qui tire une boule blanche reçoit 10 €. Calculer les espérances mathématiques des 4 personnes.

REPONSE :

Espérance mathématique = gain moyen

On peut représenter la situation sous forme d'un tableau :

Individu	Probabilité de tirer une boule blanche	Probabilité de tirer une boule blanche le 1 ^{er}	Gain	Espérance = probabilité*X
A	A étant le premier à tirer, lorsqu'il tire il y a toujours 5 boules au total et 2 boules blanches : 2/5	A étant le 1 ^{er} à tirer, il a 100% de chance de tirer le premier ! 1/1	10	$(2/5)*1*10 = 4$
B	B étant le deuxième à tirer, lorsqu'il tire une boule, il n'y a plus que 4 boules et si il tire une blanche cela signifie que A n'a pas tiré de blanche donc il en reste encore 2 dans le sac : 2/4	B étant le 2 ^{ème} à tirer, la probabilité qu'il tire la boule blanche en 1 ^{er} correspond au fait que A a tiré une noire : 3/5	10	$(2/4)*(3/5)*10 = 3$
C	C étant le troisième à tirer, lorsqu'il tire une boule, il n'y a plus que 3 boules et si il tire une blanche cela signifie que A et B n'ont pas tiré de blanche donc il en reste encore 2 dans le sac : 2/3	C étant le 3 ^{ème} à tirer, la probabilité qu'il tire la boule blanche en 1 ^{er} correspond au fait que A et B ont tiré une noire : $(3/5) * (2/4) = 3/10$	10	$(2/3) * (3/10) * 10 = 2$
D	D étant le quatrième à tirer, lorsqu'il tire une boule, il n'y a plus que 2 boules et si il tire une blanche cela signifie que A, B et C n'ont pas tiré de blanche donc il en reste encore 2 dans le sac : 2/2	D étant le 4 ^{ème} à tirer, la probabilité qu'il tire la boule blanche en 1 ^{er} correspond au fait que A, B et C ont tiré une noire : $(3/5) * (2/4) * (1/3) = 1/10$	10	$(2/2) * (1/10) * 10 = 1$

A = 4€, B= 3€, C=2€, D=1€

39. De combien de façons différentes 7 personnes peuvent-elles s'asseoir autour d'une table ronde si

- Elles peuvent s'asseoir à n'importe quelle place ?
- 2 personnes particulières ne doivent pas s'asseoir l'une à côté de l'autre ?

REPONSE :

a) = toutes les permutations possibles = $7!$ = **5040**

b) On peut tourner le problème dans le sens inverse : calculer toutes les possibilités que 7 personnes soient assises si 2 personnes particulières sont assises l'une à côté de l'autre. Ensuite, retirer cela à toutes les possibilités.

Quand 2 personnes sont assises l'une à côté de l'autre, il reste 5 personnes qui peuvent permuter entre elles $\Rightarrow 5!$ possibilités = 120 possibilités

Nombre de possibilités que 2 personnes particulières soient assises côte à côte = 7 places possibles * 2 (car elles peuvent permuter l'une avec l'autre) = 14 cas possibles

$\Rightarrow$ Nombre de façons différentes de s'asseoir pour les 7 personnes si 2 personnes particulières sont assises l'une à côté de l'autre = $14 \times 120 = 1680$

$\Rightarrow$ Nombre de façons différentes de s'asseoir pour les 7 personnes si 2 personnes particulières NE sont PAS assises l'une à côté de l'autre = $5040 - 1680 = 3360$

40. On extrait 5 cartes d'un jeu de 52 cartes bien battues. Quelle est la probabilité pour que

- 4 cartes soient des as ?
- 4 cartes soient des as et la 5e un roi ?
- 3 cartes soient des dix et 2 des valets ?
- l'on ait 3 cartes d'une même couleur et 2 d'une autre même couleur ?
- au moins une carte soit un as ?

REPONSE :

On considère dans ce problème qu'il n'y a pas de remise. La probabilité change donc à chaque tirage.

a) Nombre d'as = 4

On a 5 cas possibles de tirage en fonction de l'ordre du tirage:

As – As – As – As – Pas As = $(4/52) * (3/51) * (2/50) * (1/49) * (48/48)$

As – As – As – Pas As – As = $(4/52) * (3/51) * (2/50) * (48/49) * (1/48)$

As – As – Pas As – As – As = $(4/52) * (3/51) * (48/50) * (2/49) * (1/48)$

As – Pas As – As – As – As = $(4/52) * (48/51) * (3/50) * (2/49) * (1/48)$

Pas As – As – As – As – As = $(48/52) * (4/51) * (3/50) * (2/49) * (1/48)$

Et on fait la somme des 5 probabilités.

OU plus rapidement: $(4/52) * (3/51) * (2/50) * (1/50) * (48/48) * 5 = 0,0000185$

OU en passant par la loi hypergéométrique:

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 52, constitué de 2 sous-ensembles : N1 = nombre total d'as = 4 et N2 = nombre total de non as = 48. a1 = nombre d'as qu'on tire = 4 ; a = nombre de cartes tirées = 5 ; a2 = a-a1 = 1.

$$P(4; 4; 5; 52) = \frac{C_4^4 * C_{48}^1}{C_{52}^5} = \mathbf{0,0000185}$$

b) Dans ce cas une seule possibilité car on dit que la 5^{ème} doit être un roi. Il y a donc une notion d'ordre : As – As – As – As – Roi = $(4/52) * (3/51) * (2/50) * (1/49) * (4/48) = \mathbf{0,0000003078 = 3,078 * 10^{-7}}$

c) Pas de notion d'ordre. On a 10 possibilités :

VVVDD	VVDDV	DVDVV
VVDVD	VDVDV	DDVVV
VDVVD	DVVVDV	
DVVVD	VDDVV	

Nombre Valet = 4 et Nombre de Dix = 4

1 possibilités = $(4/52) * (3/51) * (2/50) * (4/49) * (3/48)$

$$\Rightarrow (4/52) * (3/51) * (2/50) * (4/49) * (3/48) * 10 = \mathbf{0,00000923}$$

OU en passant par la loi hypergéométrique:

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 52, constitué de 2 sous-ensembles : N1 = nombre total de valets = 4 et N2 = nombre total de dix = 4. a1 = nombre de valets qu'on tire = 3 ; a = nombre de cartes tirées = 5 ; a2 = a-a1 = 2.

$$P(3; 4; 5; 52) = \frac{C_4^3 * C_4^2}{C_{52}^5} = \mathbf{0,00000923}$$

d) Si on considère que la couleur = rouge ou noir.

Nombre de cartes d'une couleur = 26

Pas de notion d'ordre

On a 10 possibilités si les 3 cartes de même couleur sont les rouges et les 2 cartes de même couleur sont des noires :

RRRNN	RRNNR	NRNRR	RRNRN	RNRNR
NNRR	RNRRN	NRRNR	NRRRN	RNNRR

On a évidemment 10 possibilités également si les 3 cartes de même couleur sont les noires et les 2 cartes de même couleur sont des rouges.

⇒ 20 possibilités

$$1 \text{ possibilités} = (26/52) * (25/51) * (24/50) * (26/49) * (25/48)$$

$$\Rightarrow (26/52) * (25/51) * (24/50) * (26/49) * (25/48) * 20 = \mathbf{0,650}$$

OU en passant par la loi hypergéométrique:

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 52, constitué de 2 sous-ensembles : N1 = nombre total de rouge = 26 et N2 = nombre total de noire = 26. a1 = nombre de rouge qu'on tire = 3 ; a = nombre de cartes tirées = 5 ; a2 = a-a1 = 2.

$$P(3; 26; 5; 52) = \frac{C_{26}^3 * C_{26}^2}{C_{52}^5} = 0,325$$

Comme on a également la possibilité que les 3 cartes tirées de même couleur soit des noires et les 2 cartes de même couleur soient des rouges. La probabilité totale = 2* la probabilité calculée = 2*0,325 = **0,650**

e) On peut prendre le problème dans le sens inverse et faire 1 – probabilité de ne pas avoir d'as.

$$\text{Probabilité de ne pas avoir d'as} = (48/52) * (47/51) * (46/50) * (45/49) * (44/48) = 0,659$$

OU en passant par la loi hypergéométrique:

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a_2}}{C_N^a}$$

Dans notre cas, N = taille échantillon totale = 52, constitué de 2 sous-ensembles : N1 = nombre total de Non as = 48 et N2 = nombre total d'as = 4. a1 = nombre de non as à tirer = 5 ; a = nombre de cartes tirées = 5 ; a2 = a-a1 = 0 .

$$P(5; 48; 5; 52) = \frac{C_{48}^5 * C_4^0}{C_{52}^5} = 0,659$$

$$\Rightarrow \text{Probabilité d'avoir au moins un as} = 1 - 0,659 = \mathbf{0,3411}$$