Brassage interchromosomique

Test-cross et adéquation à une loi de probabilité équirépartie.

Marc Dupin
Professeur de SVT au Lycée Malherbe, Caen.

Niveau : terminale S

Thème

Validation expérimentale d'une hypothèse théorique

En SVT, comme dans toutes sciences expérimentales, seuls des résultats semblables obtenus plusieurs fois dans les mêmes conditions sont valides.

Lors des travaux pratiques, les élèves sont amenés à exploiter non seulement leurs résultats mais aussi ceux des autres groupes.

Sujet : « *Le brassage interchromosomique est dû à la migration indépendante des chromosomes homologues de chaque paire lors de l'anaphase de la première division de méiose.* »

Travail : L'étude se fait à l'aide d'un élevage de drosophiles ou mouches du vinaigre. Ces mouches de petites tailles (quelques millimètres) sont observées à la loupe binoculaire. Les élèves sont amenés à identifier et compter les différents phénotypes*[1] issus d'un croisement de drosophiles, appelé test-cross (croisement d'un parent de génotype* inconnu avec un parent homozygote à allèles* récessifs pour les deux gènes étudiés).

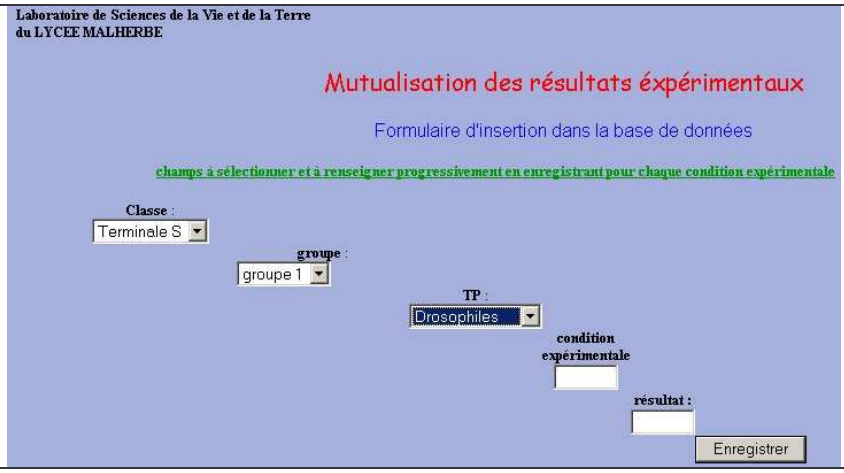
[1] Les étoiles renvoient au glossaire en fin du document.

Déroulement de la séquence : De manière à obtenir un nombre significatif d'individus étudiés, chaque binôme observe au moins une trentaine d'individus issus de l'élevage.

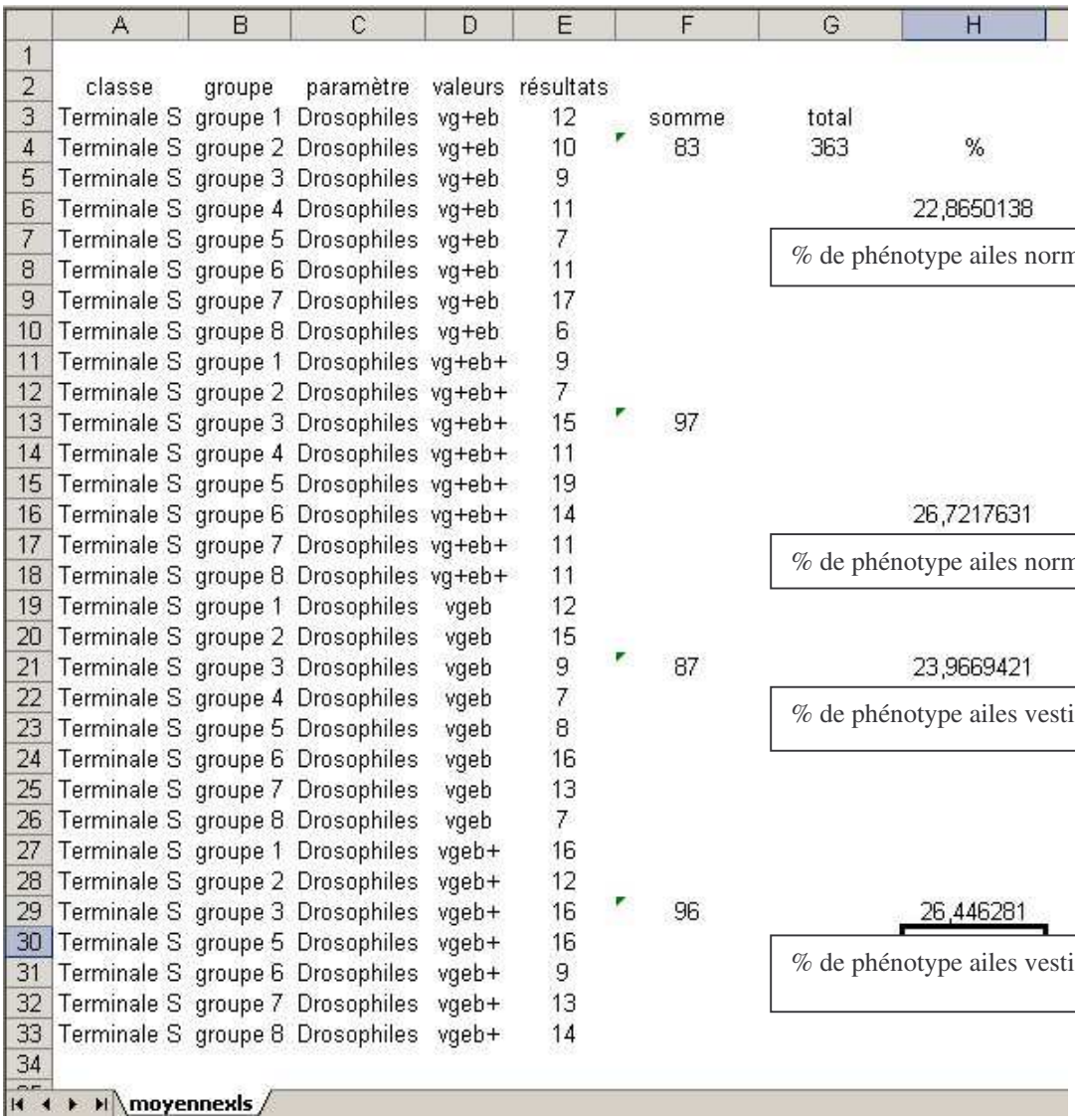
Phénotypes des parents du Test-cross	
Phénotype du parent 1 : Ailes longues, couleur du corps brun. 	Phénotype du parent 2 : Ailes vestigiales, couleur du corps noir (ébène). 

Cette observation permet de mettre en évidence l'existence de quatre phénotypes dont les pourcentages seront comparés.

Descendant à ailes vestigiales et corps brun	Descendant à ailes longues, corps brun	Descendant à ailes vestigiales, corps ébène.	Descendant à ailes Longues, corps ébène.
			

Les résultats de chaque groupe sont alors rentrés dans une base de données (mySQL) à l'aide d'un formulaire :	
L'ensemble des résultats est ensuite récupéré dans un fichier tableur (excel ou openoffice) par chaque binôme et traité : - classement par phénotype - calcul du pourcentage de chaque phénotype.	

Résultats expérimentaux



	A	B	C	D	E	F	G	H
1								
2	classe	groupe	paramètre	valeurs	résultats			
3	Terminale S	groupe 1	Drosophiles	vg+eb	12	somme	total	
4	Terminale S	groupe 2	Drosophiles	vg+eb	10	83	363	%
5	Terminale S	groupe 3	Drosophiles	vg+eb	9			
6	Terminale S	groupe 4	Drosophiles	vg+eb	11			22,8650138
7	Terminale S	groupe 5	Drosophiles	vg+eb	7			% de phénotype ailes normales corps noir
8	Terminale S	groupe 6	Drosophiles	vg+eb	11			
9	Terminale S	groupe 7	Drosophiles	vg+eb	17			
10	Terminale S	groupe 8	Drosophiles	vg+eb	6			
11	Terminale S	groupe 1	Drosophiles	vg+eb+	9			
12	Terminale S	groupe 2	Drosophiles	vg+eb+	7			
13	Terminale S	groupe 3	Drosophiles	vg+eb+	15	97		
14	Terminale S	groupe 4	Drosophiles	vg+eb+	11			
15	Terminale S	groupe 5	Drosophiles	vg+eb+	19			
16	Terminale S	groupe 6	Drosophiles	vg+eb+	14			26,7217631
17	Terminale S	groupe 7	Drosophiles	vg+eb+	11			% de phénotype ailes normales corps normal
18	Terminale S	groupe 8	Drosophiles	vg+eb+	11			
19	Terminale S	groupe 1	Drosophiles	vgeb	12			
20	Terminale S	groupe 2	Drosophiles	vgeb	15			
21	Terminale S	groupe 3	Drosophiles	vgeb	9	87		23,9669421
22	Terminale S	groupe 4	Drosophiles	vgeb	7			% de phénotype ailes vestigiales corps noir
23	Terminale S	groupe 5	Drosophiles	vgeb	8			
24	Terminale S	groupe 6	Drosophiles	vgeb	16			
25	Terminale S	groupe 7	Drosophiles	vgeb	13			
26	Terminale S	groupe 8	Drosophiles	vgeb	7			
27	Terminale S	groupe 1	Drosophiles	vgeb+	16			
28	Terminale S	groupe 2	Drosophiles	vgeb+	12			
29	Terminale S	groupe 3	Drosophiles	vgeb+	16	96		26,446281
30	Terminale S	groupe 5	Drosophiles	vgeb+	16			% de phénotype ailes vestigiales corps normal
31	Terminale S	groupe 6	Drosophiles	vgeb+	9			
32	Terminale S	groupe 7	Drosophiles	vgeb+	13			
33	Terminale S	groupe 8	Drosophiles	vgeb+	14			
34								

moyennexls

Ces résultats sont, « a l'œil nu » proches de 25% pour chaque phénotype, ce qui valide l'hypothèse d'une répartition aléatoire des couples d'allèles lors de la formation des gamètes (méiose).

Nous verrons ci-dessous comment analyser plus précisément cette proximité entre les fréquences obtenues et 25%.

Explication génotypique

Parent 1 de génotype inconnu $\frac{vg^+}{vg}\ \frac{eb^+}{eb}$ phénotype aile longue corps brun

Parent 2 homozygote récessif $\frac{vg}{vg}\ \frac{eb}{eb}$

Le parent 2 ne donne qu'un seul type de gamètes de génotype ($\underline{vg}$ $\underline{eb}$)
Les résultats font que le parent 1 doit élaborer quatre types de gamètes :
($\underline{vg^+}$ $\underline{eb^+}$) ; ($\underline{vg^+}$ $\underline{eb}$) ; ($\underline{vg}$ $\underline{eb^+}$) ; ($\underline{vg}$ $\underline{eb}$).

Le génotype du parent 1 doit donc être : $\frac{vg^+}{vg}\ \frac{eb^+}{eb}$

Il y a donc eu lors de la fabrication de ces gamètes un brassage des allèles, chacun se répartissant de manière équiprobable avec les deux autres (vg avec eb ou eb^+….)

NB : les deux barres utilisées dans l'écriture des génotypes symbolisent les 2 chromosomes de la paire sur laquelle sont portés les allèles.

Quelques notions de base de génétique

Dans une cellule diploïde, il y a deux allèles pour chaque gène : un allèle transmis par chaque parent. Les allèles transmis par les parents peuvent être identiques ou non. Chez l'être humain, chaque gène est présent en double exemplaire, l'un provenant du père, l'autre de la mère.
Dans une population, on peut avoir plusieurs allèles d'un gène, représentant plusieurs formes alternatives du même gène.
Si les allèles apportés par chaque parent sont identiques dans leur séquence nucléotidique, l'individu est homozygote pour ce gène. S'ils sont différents, l'individu est hétérozygote. Dans ce dernier cas, si l'un des deux allèles s'exprime et l'autre reste « muet », on dit que le premier est *dominant* et l'autre *récessif*. Les allèles dominants sont symbolisés par une lettre majuscule (ou un +), et les récessifs par une lettre minuscule.

La **méiose** est un processus se déroulant durant la gamétogénèse (spermatogenèse ou ovogenèse), c'est-à-dire durant l'élaboration des gamètes (les spermatozoïdes chez le mâle et les ovules chez la femelle). Elle a pour but de donner des cellules haploïdes (cellules contenant n chromosomes) à partir de cellules diploïdes (cellule contenant 2n chromosomes - chez l'homme, une cellule normale contient 2n = 23 paires de chromosomes alors qu'un gamète contient n = 23 chromosomes, soit un seul chromosome par paire).
En plus de ce rôle de division, la méiose a un rôle important dans le brassage génétique (mélange des allèles).
Chaque cellule va donc séparer son patrimoine génétique (contenu dans des chromosomes) en deux afin de ne transmettre que la moitié de ses allèles aux cellules filles.

_____

Modélisation

Dans le cross-test, on part d'une génération de mouches homozygotes, qu'on croise de telle sorte que chaque mouche ait un parent (vg, eb) et un parent (vg+,eb+). On obtient une génération F1 de mouches hétérozygotes, à ailes longues et corps brun.

On croise alors les éléments de cette population avec des homozygotes (vg, eb).

Dans le modèle classique, dans la nouvelle génération F2, la probabilité d'avoir des ailes longues est ½ et la probabilité d'avoir un corps brun est aussi ½.

On dira que les caractères ailes et couleurs sont indépendants si les probabilités de (vg,eb), (vg+,eb), (vg,eb+), (vg+,eb+) sont égales à ¼.[2]

Confrontation des données expérimentales avec le modèle

(vg+,eb)	(vg+,eb+)	(vg,eb)	(vg,eb+)
n_1=83	n_2=97	n_3=87	n_4=96
f_1=0,229	f_2=0,267	f_3=0,240	f_4=0,264

La question posée sur ces données peut être formulée ainsi :

Peut-on considérer que les 363 mouches proviennent d'une population où les 4 situations,* (vg,eb), (vg+,eb), (vg,eb+), (vg+,eb+)*, sont équiprobables ?

On peut se contenter, comme nous l'avons fait ci-dessus, de dire que les fréquences observées sont toutes proches de 1/4 : il s'agit là d'une vérification à l'œil nu, peu précise.
Nous allons considérer deux méthodes qui permettent de donner une réponse plus précise à la question posée.

Méthode 1 (méthode de validation par simulation)

Si l'hypothèse d'équiprobabilité est vérifiée, alors les fréquences $f_1,\ldots, f_4$ de chaque possibilité doivent être « proches » de 1/4. Il convient ici de quantifier cette notion de proximité des 4 fréquences avec ¼.

On pourrait mesurer l'écart entre (f_1, f_2, f_3, f_4) et (¼,¼,¼,¼) de différentes façons, par exemple par :

$$\delta = \left|f_1 - \frac{1}{4}\right| + \left|f_2 - \frac{1}{4}\right| + \left|f_3 - \frac{1}{4}\right| + \left|f_{\frac{4}{1}} - \frac{1}{4}\right|,$$

ou par :

$$d^2 = \left(f_1 - \frac{1}{4}\right)^2 + \left(f_2 - \frac{1}{4}\right)^2 + \left(f_3 - \frac{1}{4}\right)^2 + \left(f_4 - \frac{1}{4}\right)^2.$$

[2] **La définition stochastique d'indépendance de deux caractères sert ici, en génétique, à définir la notion de caractères indépendants.**

Nous choisirons la deuxième formule pour des raisons que nous donnerons ci-dessous, dans la méthode 2 (note 5).

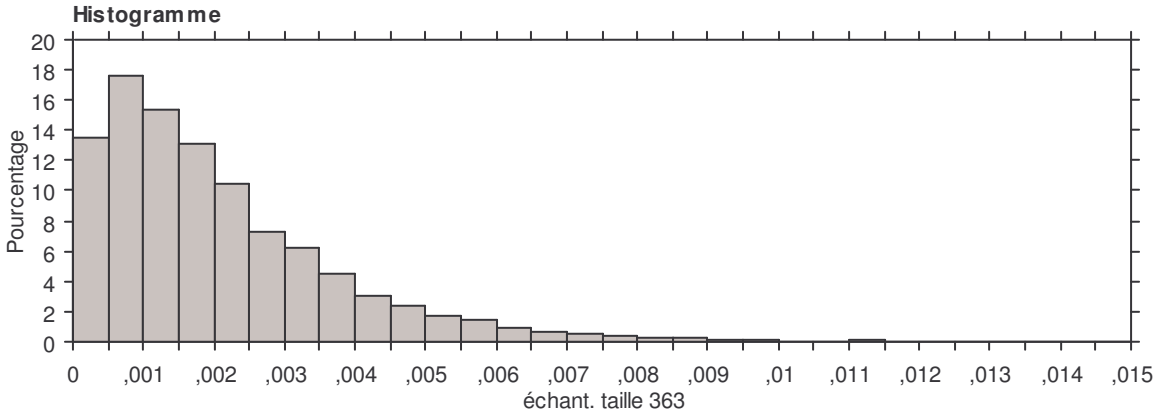
La valeur observée pour le second critère est :

$$d^2_{obs} = \left(\frac{83}{363} - \frac{1}{4}\right)^2 + \left(\frac{97}{363} - \frac{1}{4}\right)^2 + \left(\frac{87}{363} - \frac{1}{4}\right)^2 + \left(\frac{96}{363} - \frac{1}{4}\right)^2 \approx 0{,}0015$$

Mais que dire de 0,0015? Est-ce *petit* ou non ? *Petit* par rapport à quoi ? L'idée est ici de comparer 0,0015 aux valeurs de d^2 obtenues en tirant 363 fois une valeur dans un ensemble à 4 éléments, avec équiprobabilité de tirage de chaque élément lors de chaque tirage. Mais s'il est facile d'imaginer mentalement qu'on tire plusieurs séries de 363 nombres, il serait plus difficile de le faire à la main.

Nous allons dans cette optique simuler par ordinateur des échantillons de taille 363 d'une loi équirépartie sur les 4 éléments (vg,eb), (vg+,eb), (vg,eb+), (vg+,eb+), et pour chaque échantillon, calculer d^2. A l'aide d'un programme de simulation, nous avons construit 10 000 tels échantillons de taille 363. Les résultats des 10 000 valeurs ainsi simulées de d^2 sont consignés dans l'histogramme ci-dessous. On peut constater que 0,0015 est de l'ordre de grandeur de la plupart des valeurs simulées.

Nous déclarons que le modèle d'équiprobabilité est, au vu de ces 10 000 simulations, compatible avec les 363 données observées.[3]



Lecture : environ 14% des valeurs de χ^2 simulées sont entre 0 et 0,0005, environ 18% d'entre elles sont entre 0,0005 et 0,001.

Méthode 2 (théorique) [4]

[3] Si la valeur observée avait été 0,015, c'est-à-dire nettement supérieure aux 10 000 valeurs simulées, on aurait rejeté le modèle. Pour des valeurs observées moins extrêmes, se reporter à la méthode 2. On peut aussi consulter le document d'accompagnement des programmes mathématiques de terminale S,de 2002, page 145, en téléchargeant la partie concernant les probabilités et la statistique à l'adresse suivante :
http://www.ac-poitiers.fr/voir.asp?r=88

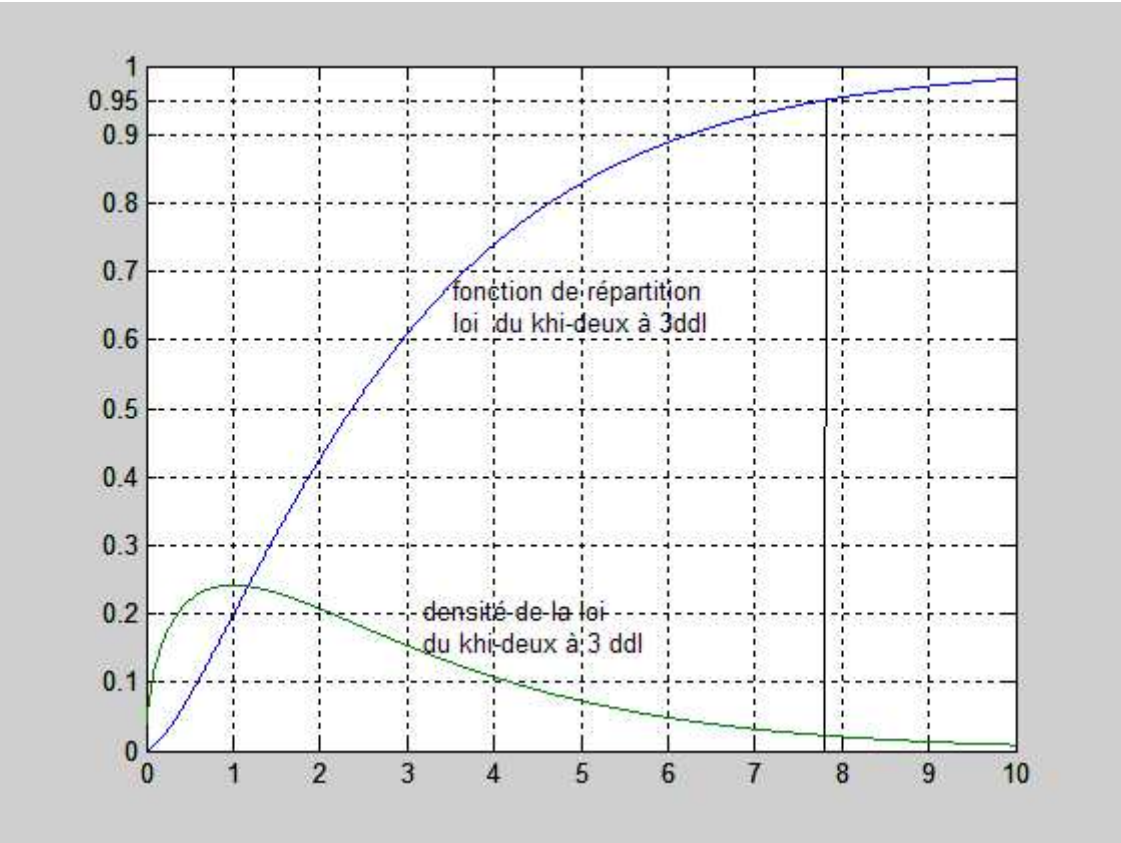
[4] à l'usage de ceux qui connaissent un peu de statistique

La méthode ci-dessus emploie des simulations, ce qui a quelques inconvénients, notamment celui de devoir refaire les simulations si on s'aperçoit par exemple qu'on a compté une mouche en trop ou qu'on en a oublié une.

On peut s'affranchir de cette difficulté en considérant, pour mesurer l'écart entre (f_1, f_2, f_3, f_4) et (¼,¼,¼,¼), la variable χ^2, dont la valeur sur un échantillon de taille n est :

$$\chi^2 = 4nd^2 = \frac{(n_1 - n/4)^2}{n/4} + \frac{(n_2 - n/4)^2}{n/4} + \frac{(n_3 - n/4)^2}{n/4} + \frac{(n_4 - n/4)^2}{n/4}$$

En effet, on démontre, dans le cadre de la théorie des probabilités, que la loi de cette variable aléatoire[5], dès que n est assez grand (ici dès que n est supérieur à 80) est bien approximée par une loi de probabilité appelée loi du khi-deux à 3 degrés de liberté[6]. La figure ci-dessous donne la densité et la fonction de répartition de cette loi.



Lecture : la probabilité que la valeur de χ^2soit inférieure ou égale à 0,95 est environ 7,8 .

Une règle de décision consensuelle en statistique est de se fixer a priori un nombre α appelé «risque de première espèce » et de calculer le nombre u tel que la probabilité que la valeur de χ^2soit inférieure ou égale à u vaille α, soit : :

$$\text{Prob}(\chi^2 \leq u) = \alpha$$

On procède alors ainsi :

[5] le terme aléatoire ici vient du fait que les valeurs de χ^2 varient selon l'échantillon considéré.
[6] Le nombre de degré de liberté est égal au nombre de possibilités,ici 4, diminué d'une unité.

-Si la valeur observée χ^2_{obs} est supérieure à u, alors on déclare le modèle incompatible avec les données. En fait, α est toujours choisi petit, et cette règle consiste à rejeter le modèle parce que la valeur observée χ^2_{obs} est dans un ensemble de valeurs jugées a priori « trop grandes ».

-Si la valeur observée χ^2_{obs} est inférieure ou égale à u, alors on déclare le modèle compatible avec les données. On accepte le modèle parce que la valeur observée χ^2_{obs} est dans l'ensemble de valeurs jugées « recevables » pour ce modèle.[7]

Si on prend[8], α=0,05, on trouve $u\approx$7,8.

La valeur observée, (χ^2_{obs}=2,21) est inférieure à 7,8, le modèle de l'équiprobabilité est donc déclaré compatible avec les données observées, au risque de première espèce 0,05.

On peut ici considérer que les 363 mouches constituent un échantillon d'une population où les 4 situations,* (vg,eb), (vg+,eb), (vg,eb+), (vg+,eb+)*, sont équiprobables .

Les données expérimentales recueillies ici constituent donc une validation de l'indépendance des deux caractères génétiques étudiés.

Glossaire

Génotype : ensemble de l'information génétique (allèles) d'un individu.
Phénotype : caractéristiques physiques et physiologiques d'un individu, résultant de son génotype et de son environnement
Méiose : ensemble de deux divisions cellulaires permettant la réduction du nombre de chromosomes dans les gamètes.
Allèle : variante donnée d'un gène au sein d'une espèce.

---

[7] La variable δ mentionnée ci-dessus ne se prête pas au même type de calculs théoriques, d'où notre choix de considérer plutôt d^2.

[8] Le choix de α est souvent l'objet de discussion. Mais en pratique, on le choisit toujours inférieur à 0,1 et plus il est petit, plus le seuil *u* correspondant est grand. Le nombre *u* correspondant à α=0,9 est 6,2 et la conclusion reste inchangée.