Exercices MS-GERAA

Le présent document propose une série d'exercices relatifs aux notions de génétique vues lors du cours Z00L2003 (partim « génétique »). Ces exercices couvrent plus ou moins les matières vues au cours, et sont à l'image des problèmes qui vous seront proposés lors de l'évaluation du mois de juin. Les solutions sont fournies dans la deuxième partie du document, mais nous vous suggérons d'essayer de résoudre les problèmes en ne recourant aux solutions qu'au cas où vous êtes « bloqués » sur un problème. Nous restons évidemment à votre disposition pour répondre à des questions sur cette matière, au besoin.

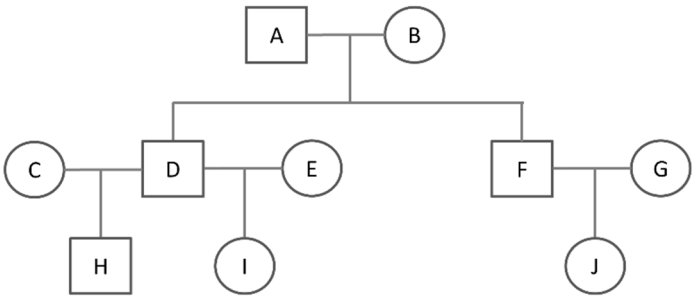
Bon travail !

Tom Druet et Frédéric Farnir

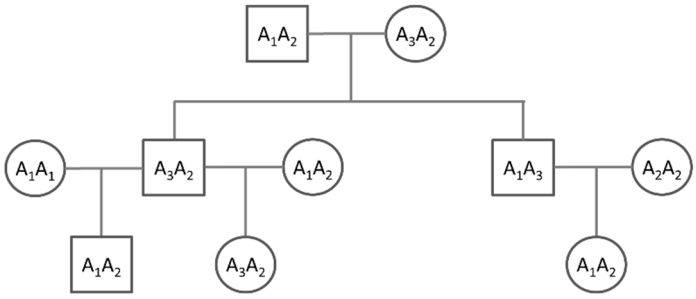
Questions

1) Parenté

1.1) Dans la généalogie ci-dessous, les lettres représentent les individus.

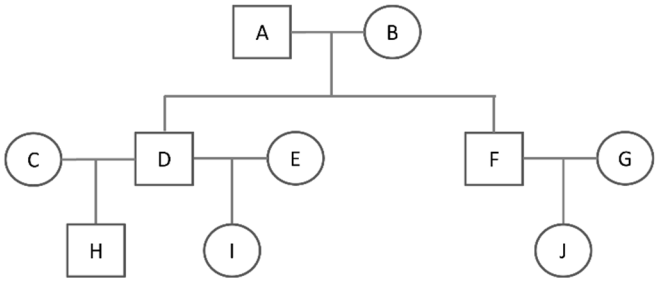


La seconde représentation indique les génotypes à un locus pour la même généalogie.



Sur base de ces éléments, décrivez les relations IBD et IBS entre les allèles des individus de la dernière génération (les descendants). Décrivez également les relations IBD et IBS entre les allèles des grands-parents. Enfin, indiquez les relations IBD et IBS entre les grands-parents et la génération actuelle.

1.2) Calculez les coefficients de parenté entre les individus (H,A), (C,E), (H,I), (H,F), (H,G), (H,J).



2) Calculs d'héritabilité via une régression ou une ANOVA

On dispose des poids à 1 an de 20 couples mère - fille dans l'espèce pangasius. On souhaite estimer l'héritabilité de ce caractère (poids à un an). Si on suppose que les animaux sont dans des conditions d'environnement comparables (c'est le cas pour le sexe et l'âge), utilisez une régression pour obtenir une estimation de l'héritabilité et expliquez le raisonnement.

Données

Mère	Fille	Mère	Fille	Mère	Fille	Mère	Fille
0.859	0.879	0.909	1.250	0.844	1.194	1.137	1.529
1.164	1.239	1.088	1.122	1.108	1.477	1.248	0.881
1.190	0.653	0.722	1.008	0.972	0.667	1.004	0.966
1.210	1.339	0.929	0.813	1.104	0.841	0.821	0.512
.0928	0.691	1.036	1.100	1.356	0.820	1.102	0.914

3) Index de sélection dans différentes configuration et précision de l'index

3.1) L'héritabilité d'un caractère est 0.40, sa répétabilité 0.60 et la valeur de c est égale à 0.10.

Cinq individus ont des performances propres corrigées pour la moyenne égales à 5, 10, -6, -11 et 2. Quel est le coefficient b de l'index de sélection ? Quelles sont les valeurs génétiques obtenues en utilisant l'index de sélection ? Quel individu a la meilleure valeur génétique ? Quelle est la précision de son index ? Est-elle meilleure que celle des autres individus ?

3.2) L'héritabilité d'un caractère est 0.10, sa répétabilité 0.15 et la valeur de c est égale à 0.15.

Deux stratégies sont envisagées pour obtenir la valeur génétique d'un animal. Soit à partir de la valeur de trois performances répétées, soit à partir des performances de 10 demi-frères. Quelle stratégie permettra d'obtenir le meilleur progrès génétique ?

3.3) L'héritabilité d'un caractère est 0.20, sa répétabilité 0.30 et la valeur de c est nulle.

Deux stratégies sont envisagées pour obtenir la valeur génétique d'un animal. Soit à partir de la moyenne des parents, soit à partir des performances de deux plein-frères. Quelle stratégie permettra d'obtenir de maximiser la réponse à la sélection ?

3.4) L'héritabilité d'un caractère est 0.40, sa répétabilité 0.60 et la valeur de c est égale à 0.10.

Pour les individus, A, B et C, on dispose des informations suivantes :

A) Performance propre = +10,

B) Moyenne des parents = +10,

C) Moyenne d'un groupe de 5 descendants = +10.

Calculez les valeurs génétiques des trois individus. Laquelle de ces valeurs est la plus précise ? Quel individu allez-vous sélectionner ?

4) Calcul de l'héritabilité sur base d'une série de données de descendants

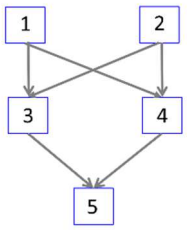
5 descendants de 4 reproducteurs ont été pesés. Les résultats (en grammes) sont les suivants :

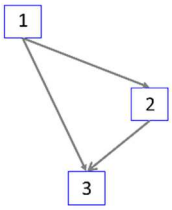
Repro 1	Repro 2	Repro 3	Repro 4
350	407	435	383
349	389	363	422
436	433	323	442
368	450	382	415
400	372	409	402

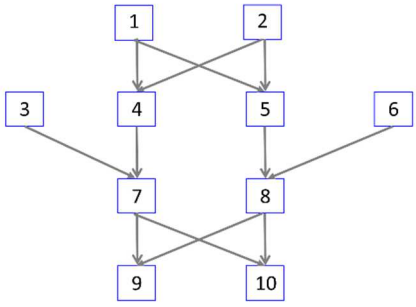
On demande d'estimer la variance génétique du poids de cette espèce, ainsi que l'héritabilité.

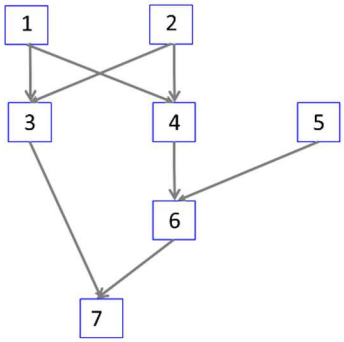
5) Consanguinité dans un pedigree

Calculez les coefficients de consanguinité de l'individu de la dernière génération pour chacune des quatre généalogies suivantes.









6) Consanguinité dans une population de taille constante

60 individus (diploïdes) sont sélectionnés pour créer un noyau d'une lignée de poissons (30 mâles et 30 femelles). Ces individus forment la génération G_0 dans l'expérience, et ils sont supposés non-apparentés. Les générations suivantes (G_1, G_2, G_3, ...) sont formées par l'union d'un père et d'une mère de la génération précédente. On considèrera qu'il y a 30 mâles et 30 femelles dans chaque génération. Après combien de générations la consanguinité dans cette expérience dépassera-t-elle 10% ? Faites le calcul de manière théorique et par simulation.

(**Truc** : dans le cas considéré, $\Delta F = \frac{1}{8*N_m} + \frac{1}{8*N_f}$ où N_m, N_f sont les nombres de mâles et de femelles, respectivement)

7) Identification génomique des parents

7.1) Dans le tableau suivant, les génotypes sont codés 0, 1 et 2 pour AA, AB et BB (par exemple). On cherche à identifier les parents d'un individu I_1. On connaît les génotypes de quatre pères et mères potentiels, soit P_1, P_2, P_3 et P_4 et M_1, M_2, M_3 et M_4. Chaque colonne représente un marqueur.

I_1	0	1	0	1	1	2	0	1	2	1
P_1	1	1	2	0	2	1	1	0	0	2
P_2	1	2	1	0	2	1	0	1	1	2
P_3	1	2	1	0	1	1	0	1	1	2
P_4	0	1	0	1	1	1	2	1	2	2
M_1	0	0	0	2	2	1	0	0	1	0
M_2	2	0	0	1	1	0	0	0	1	0
M_3	0	0	0	2	2	1	0	0	1	0
M_4	0	1	1	0	1	1	0	0	1	2

7.2) Un descendant a un génotype {0,2,1,1,0} (soit {AA, BB, AB, AB, AA}) et on veut savoir si un individu de génotype {1,2,2,0,1} pourrait être son père. Calculez la probabilité de ces génotypes si il s'agit du père sachant que les fréquences des allèles A sont {0.10, 0.90, 0.50, 0.30, 0.50} pour chacun des marqueurs.

8) Génétique factorielle - Hardy-Weinberg

Un locus EpEc est associé à l'épaisseur des écailles. Deux allèles coexistent à ce locus (E et e). La présence de l'allèle E conduit à un épaississement des écailles. Par conséquent, les individus de génotype ee ont des écailles fines, ceux de génotypes Ee ont des écailles d'épaisseur moyenne et ceux qui ont un génotype EE ont des écailles épaisses. Si, dans un échantillon aléatoire, on a 50 individus avec des écailles fines et 20 avec des écailles moyennes, combien devrait-on avoir d'individus avec des écailles épaisses pour que les proportions génotypiques soient en équilibre d'Hardy-Weinberg ?

Résolutions

Problème 1.1

Les allèles identiques sont soit IBD, soit IBS. Ils sont IBD s'ils sont hérités du même ancêtre.

Dans la dernière génération, l'allèle A_3 est distinct des autres (ni IBD, ni IBS). L'allèle A_2 de I vient de E et n'est donc pas IBD (IBS avec les autres A_2). De même, l'allèle A_2 de J vient de G. Donc aucun des allèles A_2 ne sont IBD (ils sont IBS). L'allèle A_1 de H vient de C et aucun allèle A_1 n'est donc IBD (ils sont IBS entre eux).

Par définition, les allèles des grands-parents ne sont pas IBD. Les deux allèles A_2 sont donc IBS.

L'allèle A_2 de H est IBD avec l'allèle A_2 de A.

L'allèle A_3 de I est IBD avec l'allèle A_3 de B.

L'allèle A_1 de J est IBD avec l'allèle A_1 de A.

Les autres allèles identiques sont IBS.

Problème 1.2

Les coefficients de parenté peuvent s'obtenir par la méthode des chemins ou par la méthode tabulaire. Dans ce second cas, il faut trier les individus et l'on obtient de relations de parenté (deux fois le coefficient de parenté). Dans une petite généalogie, la méthode des chemins est assez simple : on compte le nombre d'ancêtres reliant les deux individus d'intérêts (soit N) et le coefficient est 0.5^N. S'il y a plusieurs chemins possibles, on somme les probabilités :

Exemple, pour (H,A) on a un chemin de trois individus H-D-A, soit $0.5^3 = 0.125$.

Pour (C,E), il n'y a pas d'ancêtres communs, soit 0.

Pour (H,F), on a deux chemins : H-D-A-F et H-D-B-F, soit $2 \times 0.5^4 = 0.125$.

(H,I) = 0.125 / (H,G) = 0 / (H,J) = 0.0625.

Problème 2

Il a été vu dans le cours que la régression des données des filles (Y) sur les données de leur mère (X) conduisaient à un coefficient de régression valant la moitié de l'héritabilité.

Si on effectue le calcul du coefficient de régression de Y sur X, on obtient :

$$b = 0.1078$$

Par conséquent, l'estimation de l'héritabilité est: $h^2 = 2 * b = 0.2156$

Problème 3.1

Il s'agit d'un index sur performances propres, la valeur de b est donc l'héritabilité et la précision, identique pour tous, vaut h (0.63). Les valeurs estimées sont 2, 4, -2.4, -4.4 et 0.8. Le deuxième individu a la valeur génétique estimée la plus élevée.

Problème 3.2

Il s'agit d'abord d'un index sur une moyenne de performances propres (par exemple, diapo 65), la valeur de b est donc $\frac{n*h^2}{1+(n-1)*\rho}$ avec n = 3, h² = 0.10 et ρ = 0.15 et la précision la racine carrée de cette valeur, soit 0.48.

Le second index est sur la moyenne de 10 demi-frères (par exemple, diapo 67) et la précision s'obtient :

$$r_{AI} = \sqrt{\frac{n}{4 * (n + k)}}$$

La valeur de k est (4 – h²) / h², soit 39. La précision vaut alors 0.23.

Problème 3.3

Il faut calculer les précisions d'index et les comparer. Si vous voulez être précis, vous pouvez mentionner les autres composantes du progrès génétique.

Dans le premier cas, la précision de la moyenne des parents est la racine carrée de (h² / 2), soit 0.32 (diapo 68).

Pour la moyenne de deux plein-frères (avec c = 0), on utilise la diapo 70 :

$$r_{AI} = \sqrt{\frac{n * 0{,}25 * h^2}{1 + (n - 1) * 0{,}5 * h^2}}$$

Et la précision vaut 0.30.

Problème 3.4

Pour A, I = 10 * 0.40, soit 4. Précision 0.63 (racine (h²)

Pour B, I = 10 * 0.40, soit 4. Précision 0.45 (racine (h² / 2))

Pour C, I = 10* 2 * 5 / (5 + k) et k = (4 – 0.40) / 0.40, soit 9. I = 10 * 2 * 5 /14 = 7.14. Précision est racine (5 / (5 + k)), soit 0.60.

Problème 4

On reconnait une analyse de la variance aléatoire (les 4 reproducteurs sont un échantillon de la population des reproducteurs) à un facteur. On peut effectuer les calculs classiques de l'analyse de la variance, pour obtenir le traditionnel tableau :

Source	Sommes de carrés	Degrés de liberté	Carrés moyens
Modèle	4525	3	1508.33
Erreur	18888	16	1180.50
Total	23413	19	1232.26

Le carré moyen de l'erreur (CME) estime σ_E^2 et vaut : $s_E^2 = 1180.50$

Il est dit dans les notes que l'estimation de la variance génétique σ_A^2 se faisant via un calcul impliquant les sommes de carrés : $s_A^2 = \frac{4*(CMM-CME)}{5} = 262.27$

Par conséquent, l'estimation de l'héritabilité vaudra : $\hat{h}^2 = \frac{s_A^2}{s_A^2+s_E^2} = 0.182$

Problème 5

Les coefficients de consanguinité peuvent s'obtenir par la méthode des chemins ou par la méthode tabulaire (dans ce cas-là, il faut trier les individus). Le coefficient de parenté est également égal au coefficient de parenté des parents ou à la moitié de la relation génétique additive des parents. Dans chacun des exemples, seuls les individus de la dernière génération sont consanguins.

A) Individu 5, F = 0.25

B) Individu 3, F = 0.25

C) Individus 9 et 10, F = 0.0625

D) Individu 7, F = 0.125

Problème 6

Résolution théorique

La formule fournit facilement : $\Delta F = \frac{1}{240} + \frac{1}{240} = \frac{1}{120}$

Par ailleurs, on sait que, par définition : $\Delta F = \frac{F_{t+1}-F_t}{1-F_t} \Rightarrow F_{t+1} = F_t + (1 - F_t) * \Delta F$

Comme $F_0 = 0$, on peut en déduire les valeurs de F_1, F_2, F_3, …

On obtient : $F_1 = 0.0833$, $F_2 = 0.0166$, $F_3 = 0.0248$, …, $F_{12} = 0.0955$, $F_{13} = 0.1031$ (avec excel)

Résolution par simulation

Voir le document Excel attaché.

Problème 7.1

Le premier père peut être exclu à partir des marqueurs 3 et 9 (homozygotes opposés, incompatibles). Bien que le 4ème père ressemble beaucoup à I_1, il est incompatible au 7ème marqueur. Les pères P_2 et P_3 sont tous les deux compatibles.

Les mères M_1, M_3 et M_4 sont toutes les trois compatibles. La deuxième mère est incompatible aux marqueurs 1 et 6.

Il reste donc deux pères compatibles et trois mères. Pour éliminer d'autres parents, il faut tester des couples de parents.

La mère M_4 peut être éliminée parce qu'elle n'est pas compatible au marqueur 4 avec les deux pères restant et au marqueur 10 avec tous les pères. En effet, I_1 est hétérozygote au marqueur 10 et ne peut recevoir que des B des pères (tous BB), donc la mère doit avoir au moins un allèle A.

Les deux pères P_2 et P_3 sont identiques sauf au marqueur 5. L'individu est AB, les deux mères restantes sont BB et le père ne peut pas être BB, ce qui élimine P_2.

Enfin M_1 et M_3 sont identiques à tous les marqueurs et ne peuvent être distinguées. Il faudrait plus de marqueurs. Notez, qu'il existe des logiciels pour vous aider et que l'on fait appel à des probabilités.

Problème 7.2

Il y a incompatibilité entre les deux individus si, pour un ou plusieurs marqueurs, le père est AA (0) et le fils BB (2), ou si le père est BB (2) et le fils AA (0). On constate que cette situation n'arrive pas pour les 5 marqueurs testés. Au vu de ces génotypes, la probabilité que le second individu soit le père du premier n'est pas nulle. On peut calculer la valeur de cette probabilité. Pour cela, on partira de l'hypothèse que les 5 marqueurs sont indépendants, ce qui est le cas s'ils ont par exemple situés sur des chromosomes différents. Dans ce cas, la probabilité globale (c'est-à-dire pour les 5 marqueurs) est le produit des probabilités à chaque marqueur individuel. Effectuons le calcul pour le premier marqueur : nous devons calculer la probabilité Pr(Fils = AA | Père = AB). Pour arriver à un fils AA, il faut que le père lui ait donné un allèle A (Probabilité = 0.5) et que le fils ait hérité du côté maternel d'un allèle A également (Probabilité = f(A) = 0.10). La probabilité vaut donc :

Pr(Fils = AA | Père = AB) = 0.5*0.1 = 0.05

On répète ce type de calculs pour les 5 marqueurs, et on multiplie les 5 probabilités obtenues pour obtenir la probabilité finale. Le calcul est donc :

- Marqueur 1 : Pr(Fils = AA | Père = AB) = 0.5*0.1 = 0.05
- Marqueur 2 : Pr(Fils = BB | Père = BB) = 1.0*(1.0 - 0.9) = 0.10
- Marqueur 3 : Pr(Fils = AB | Père = BB) = 1.0*0.5 = 0.50
- Marqueur 4 : Pr(Fils = AB | Père = AA) = 1.0*(1.0 - 0.3) = 0.70
- Marqueur 5 : Pr(Fils = AA | Père = AB) = 0.5*0.5 = 0.25

La probabilité vaut donc :

Pr = 0.05*0.10*0.50*0.70*0.25 = 0.0004375

Si plusieurs pères étaient possibles, on comparerait cette probabilité à celles obtenues avec les autres pères pour voir quel père est le plus vraisemblable.

Problème 8

On appelle X le nombre (inconnu) d'individus EE. Il y a donc, au total, (50 + 20 + X) individus dans l'expérience, et 2*(50 + 20 + X) allèles.

On peut comptabiliser les fréquences relatives d'allèles e et E :

$$f_e = \frac{50 * 2 + 20 * 1 + X * 0}{2 * (50 + 20 + X)} = \frac{120}{140 + 2 * X}$$

$$f_E = \frac{50 * 0 + 20 * 1 + X * 2}{2 * (50 + 20 + X)} = \frac{20 + 2 * X}{140 + 2 * X}$$

Si on est à l'équilibre d'Hardy-Weinberg, la fréquence relative des ee devrait être f_e^2. La fréquence absolue attendue s'obtient en multipliant la fréquence relative par l'effectif total, ce qui donne :

$$(70 + X) * \left(\frac{120}{140 + 2 * X}\right)^2 = 50$$

On peut résoudre facilement cette équation du second degré, pour obtenir $X = 2$.

On peut vérifier que l'équilibre est alors respecté :

$$N_{ee} = 72 * \left(\frac{120}{144}\right)^2 = 50$$

$$N_{Ee} = 72 * 2 * \frac{120}{144} * \frac{24}{144} = 20$$

$$N_{EE} = 72 * \left(\frac{24}{144}\right)^2 = 2$$