

Présentation de l'UE Méthodologie en biologie et mathématiques

L'UE est structurée en 12 séances dont la dernière est consacrée à l'évaluation orale.

Divers documents seront déposés au cours du semestre sur eCampus

Méthodologie en Biologie et Mathématiques

Vous y trouverez, entre autres, l'ensemble des articles proposés pour l'évaluation orale.

Equipe pédagogique :

Natalia Conde e Silva (natalia.conde-e-silva@universite-paris-saclay.fr) – biologie

Gaëlle Lelandais (gaelle.lelandais@universite-paris-saclay.fr) – biologie/mathématiques

Alain Virouleau (alain.virouleau@universite-paris-saclay.fr) – mathématiques

Modalités de contrôle des connaissances et des compétences

Session 1 : 60% CC + 20% EEF + 20% EOF

Session 2 : 60% CC (R) + 40% EEF

Quelques règles à lire attentivement (et à suivre !...) :

1. L'existence du contrôle continu rend obligatoire la présence en séance. Un étudiant absent sans justification (ABI) à une séance ne sera pas autorisé à rendre le devoir correspondant à cette séance.

Par contre, en cas d'absence justifiée (ABJ), l'étudiant a la possibilité de rendre le devoir. La note ne sera comptabilisée dans la moyenne que si elle est favorable à l'étudiant.

2. Tout retard non justifié dans la remise des devoirs demandés dans le cadre du contrôle continu sera sanctionné.

3. L'absence de note de contrôle continu ne permettra pas la présentation de l'article.

4. Un étudiant absent injustifié (ABI) à 3 séances sera DEF à l'UE.

5. Un étudiant absent injustifié (ABI) à la présentation orale finale ou à l'examen de synthèse sera DEF à l'UE.

Calendrier des séances 2024-25

		séances "Biologie"	séances "Maths"
Séance 1	Démarche scientifique	11/10/2024	08h15 - 10h15
Séances 2 et 3	Synthèse de document	18/10/2024	08h15 - 12h30
Séance 4	Analyses de données 1	08/11/2024	08h15 - 10h15
Séance 5	Analyses statistiques 1	08/11/2024	10h30 - 12h30
Séance 6	Analyses de données 2	22/11/2024	08h15 - 10h15
Séances 7 et 8	Analyses statistiques 2 et 3	26/11/2024	13h30 - 17h45
Séance 9	Analyse d'article 1	29/11/2024	08h15 - 10h15
Séance 10	Analyse d'article 2	06/12/2024	08h15 - 10h15
Séance 11	Analyses statistiques 4	06/12/2024	10h30 - 12h30
Evaluation	Soutenances	en janvier 2025, à confirmer	

Séance : Démarche scientifique

Objectifs de la séance

Identifier comment se construit la science
Connaître le principe et les étapes de la démarche scientifique
Connaître le principe et les étapes de la démarche expérimentale
Expliquer le principe de la communication scientifique

Déroulé de la séance

- Activité 1 « Bris de glace » : Qu'est-ce que la Science ?
- Activité 2 : La démarche scientifique
<http://www.cea.fr/multimedia/Pages/videos/culture-scientifique/science-societe/definition-demarche-scientifique.aspx>
- Activité 3 : La démarche expérimentale
<https://www.youtube.com/watch?v=xS4x7n7o9zI>
- La communication scientifique

RESEARCH PAPER

A MITE insertion abolishes the AP3-3 self-maintenance regulatory loop in apetalous flowers of *Nigella damascena*

Natalia Conde e Silva^{1,4}, Martine Leguilloux¹, Arnaud Bellec², Nathalie Rodde³, Juliette Aubert^{1,4},
Domenica Manicacci¹, Catherine Damerval¹, Helene Berges^{2,4} and Yves Deveaux^{1,4}

¹ Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Évolution-Le Moulon, IDEEV, 12 route 128, 91190 Gif-sur-Yvette, France

² French Plant Genomic Resource Center, INRA-CNRGV, 24 Chemin de Bordeaux-Flouze-Auzouville, CS 52627, 31326 Castanet Tolosan Cedex, France

³ Present address: DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France

⁴ Present address: Inrae Agriculture, One Kendall Square Building 800/700, Cambridge, MA, 02130, USA

* Correspondence: natalia.conde-e-silva@universite-paris-saclay.fr or yves.deveaux@universite-paris-saclay.fr

Received 19 July 2022; Editorial decision 2 December 2022; Accepted 12 December 2022

Editor: Madaleno Bartlett, University of Massachusetts, Amherst, USA

Abstract

MADS-box transcription factors are important regulators of floral organ identity through their binding to specific motifs, termed CarG, in the promoter of their target genes. Petal initiation and development depend on class A and B genes, but MADS-box genes of the APETALA3 (AP3) clade are key regulators of this process. In the early diverging eudicot *Nigella damascena*, an apetalous [T] morph is characterized by the lack of expression of the *NdAP3-3* gene, with its expression being petal-specific in the wild-type [P] morph. All [T] morph plants are homozygous for an *NdAP3-3* allele with a Miniature Inverted-repeat Transposable Element (MITE) insertion in the second intron of the gene. Here, we investigated to which extent the MITE insertion impairs regulation of the *NdAP3-3* gene. We found that expression of *NdAP3-3* is initiated in the [T] morph, but the MITE insertion prevents its positive self-maintenance by affecting the correct splicing of the mRNA. We also found specific CarG features in the promoter of the *NdAP3-3* genes with petal-specific expression. However, they are not sufficient to drive expression only in petals of transgenic *Arabidopsis*, highlighting the existence of *Nigella*-specific cis/trans-acting factors in regulating AP3 paralogs.

Keywords: APETALA3, *Arabidopsis*, CarG motif, floral mutant, MADS-box gene, MITE, *Nigella damascena*, petal, self-regulatory loop.

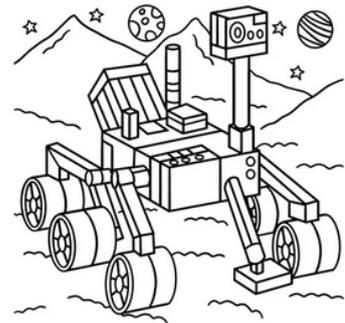
Introduction

Flowers present a large diversity of shape and ornamentation, and transcription factor families have played an important role in this evolution (Wils and Kaufmann, 2017). Flowers are composed of fertile organs, androecium and gynoecium, surrounded by a set of sterile organs, the perianth. While flowers with undifferentiated perianth are composed of visually almost identical organs, flowers with a differentiated perianth present a first whorl of sepals, which protect the fertile organs during development, and a second whorl of petals that play a major role in attracting pollinators. Remarkably, perianth

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Experimental Biology. All rights reserved. For permissions, please email: journals.permissions@oup.com



Documents pour les séances de « Biologie »



Séances : Synthèse de documents ***Identifier et hiérarchiser des informations***

Préparation du travail de synthèse de documents :

Lire les documents 1 et 2

Une pile microbienne à plantes pour l'électricité de demain

Pays-Bas : Plant-e, ou quand les plantes d'eau produisent du courant

Souligner directement dans le texte les informations importantes

Identifier les différentes parties qui vous semblent d'un intérêt majeur (en proposant des titres, sous-titres...)

Le but est de vous proposer un outil/une méthode **pour mieux synthétiser, compléter vos cours** et vos lectures (rédaction de fiches de synthèse par ex.). **Dans un futur proche, cela vous aidera dans la rédaction de vos compte-rendus de TP ou mémoires de stages** (notez que vous aurez un stage obligatoire dès l'année de LDD2 MSV). Vous devrez synthétiser un ensemble d'articles scientifiques de plusieurs pages chacun, pas nécessairement en accord les uns les autres etc... **Cette séance représente donc un exercice à petite échelle de ce type de travail (car nous ne travaillerons que sur la base deux/trois documents uniques).**

Faire la synthèse de plusieurs textes :

Une synthèse de documents doit posséder les qualités suivantes :

- * La synthèse doit être rédigée en un texte, ce n'est pas seulement un plan.
- * Ce texte doit avoir un titre et 4-5 mots-clés caractéristiques.
- * Ce texte doit être concis, clair, éviter les redondances et les répétitions.
- * Ce texte doit refléter fidèlement les idées des différents auteurs. Cette synthèse n'est pas une évaluation ou un commentaire personnel des textes.
- * Ce texte doit être ordonné et suivre un plan structuré.
- * Ce texte doit être cohérent avec des liens entre les idées présentées.
- * Ce texte doit reformuler correctement les documents d'origine dans un style clair. La reformulation (réécriture avec des mots différents, recherche de synonymes) doit respecter les règles de concision et de cohérence. Éviter le simple collage des phrases-clés du texte (et le plagiat !).

Comment procéder ?

1- Caractériser les différents documents

Titre, auteur(s), date de rédaction/publication, nature du document, idées/informations majeures

2- Analyse de chaque document

a- le caractère du texte

Lire intégralement le texte. Dans cette première lecture, identifier le caractère du texte, l'intention générale de l'auteur. Noter les passages non compris.

b- le thème, la thèse

Quel est le problème posé ? Quelles sont les idées principales du texte ? Quelles sont les idées secondaires ? Quels résultats ou quelles solutions l'auteur propose-t-il ?

c- l'organisation du texte

Les paragraphes vous renseignent sur l'organisation du texte, les arguments sont en général accompagnés d'exemples, leur caractère concret permet de les repérer.

Les phrases de liaisons montrent l'articulation des idées. Entourer ces mots de liaison (de plus, néanmoins, cependant, donc...).

Ajouter dans la marge les connecteurs logiques (ce que l'on comprend mais qui n'est pas formulé explicitement par l'auteur).

3- Synthèse des documents sous forme de tableau

	Document 1	Document 2
titre		
nature		
année		
auteur		
origine		
idée 1		
idée 2		
....		

Mettre en valeur dans le tableau, par un jeu de couleur par exemple, les points communs et différences/divergences.

4- construire le plan de la synthèse

Pour chaque idée nouvelle introduite, indiquer de quel(s) document(s) elle provient en utilisant le numéro du document entre crochet (ex : [1])

Ne pas oublier de référencer les documents utilisés en fin de synthèse dans une partie nommée bibliographie. Attention de bien faire correspondre le numéro du document et le renvoi utilisé dans le texte.

Vérification de la qualité du résumé/de la synthèse :

- procéder à une vérification croisée des informations entre celles trouvées dans la synthèse et celles des différents documents. Rien qui ne se trouve dans votre texte ne doit provenir d'ailleurs que des documents de départ.

- vérifier :

- le respect du sens
- la reformulation personnelle
- le fait que votre texte soit compréhensible sans avoir lu les documents d'origine
- l'orthographe (ou faites vérifier par quelqu'un d'autre !) et l'absence de répétitions
- le nombre de mots

Travail à rendre noté

Rédiger une synthèse (600 mots +/- 5%) à partir des documents 3 à 5. Ne pas oublier de donner un titre et 4 mots-clés, et de mentionner le nombre de mots et la bibliographie (qui ne rentre pas dans le décompte du nombre de mots).

Document 1

Pays-Bas: Plant-e, ou quand les plantes d'eau produisent du courant

Par Nicolas DELAUNAY, Le Point International

Publié le 03/06/2015

Fondée en 2009, Plant-e commercialise et perfectionne un système imaginé par des chercheurs de l'université de Wageningen et breveté en 2007. Ce système permet de produire de l'électricité pour peu que l'on dispose de plantes poussant dans un milieu saturé en eau, que ce soit à grande échelle, comme dans une mangrove, une rizière, un marais ou une tourbière, ou simplement dans un pot ou dans un jardin.

"Il n'en est qu'à ses débuts et beaucoup de choses doivent encore être grandement améliorées, mais son potentiel est énorme", soutient Jacqueline Cramer, professeur en innovation durable à l'université d'Utrecht et ancienne ministre en charge de l'Environnement aux Pays-Bas. "Si ce système devient assez performant, on peut imaginer fournir en électricité des zones reculées ou même en installer dans nos villes et campagnes pour produire de l'énergie verte", poursuit-elle.



Figure 1 : La scientifique hollandaise Marjolein Helder, co-fondatrice de Plant-e, à Wageningen, le 5 février 2015 © NICOLAS DELAUNAY AFP

La technologie est liée à la photosynthèse, via laquelle la plante produit notamment de la matière organique. Cette matière est utilisée par la plante, mais l'excédent est rejeté dans le sol via les racines, autour desquelles vivent des micro-organismes qui s'en nourrissent et libèrent des électrons. En plaçant des électrodes en carbone à proximité des racines, on "récolte" ces électrons et on peut générer du courant. Produire de l'électricité grâce aux plantes n'est pas un concept neuf, "mais ici, nous n'avons pas besoin d'endommager la plante, c'est un système non intrusif", souligne Mme Helder. Seul hic, si l'eau gèle ou s'évapore, le système arrête de fonctionner. "Mais il suffit de rajouter de l'eau ou que la glace fonde, et c'est reparti", assure Mme Helder. Cela étant, "il y a pas mal de régions du monde qui n'ont pas ce problème".

- Le chemin est encore long –

Plant-e vend actuellement son système sous la forme de petites plaques en plastique de 50 centimètres de côté qui s'assemblent les unes aux autres en intégrant et les plantes et la technologie. Il est destiné à être installé dans des lieux publics ou sur des toits d'immeubles, notamment. Pour couvrir une surface de 100 m², il en coûtera 60.000 euros. Mais le produit phare est en développement : un système en forme de tube qui pourra être directement immergé dans un milieu saturé en eau.

Le champ d'application potentiel est vaste, mais l'Asie du sud-est, qui regorge de rizières, mangroves et autres zones humides et où l'accès à l'électricité est difficile, apparaît comme un terrain particulièrement adapté. Selon des chiffres de la Banque mondiale, moins de la moitié de la population du Cambodge (31%) et de la Birmanie (49%) ont accès à l'électricité. Le Bangladesh (55%) ou le Laos (66%) font un peu mieux.

Le chemin est pourtant encore très long pour Plant-e, qui survit pour l'instant surtout grâce à des subsides: les coûts doivent être réduits et, surtout, l'efficacité améliorée. Actuellement, une installation du système sur 100 m² permet de recharger un smartphone, d'allumer une série d'ampoules LED ou d'alimenter une borne wifi, estime Mme Helder. Dans "quelques années", Plant-e espère produire 2.800 kWh par an avec la même surface, soit environ 80% des besoins en électricité d'un ménage néerlandais moyen (2,2 personnes).

Pour tester la technologie à plus grande échelle, deux systèmes de 100 m² chacun - d'un coût total de 120.000 euros - ont été installés sur un pont et dans un pôle économique avec l'appui des pouvoirs locaux. "Nous voulions aider au développement de cette technique qui a un potentiel énorme", explique Bas Boeker, directeur de projet au sein de l'organisation gérant une partie de l'immobilier de l'Etat. Les premiers résultats sont encourageants et les obstacles ne sont pas toujours ceux auxquels on s'attend: sur un des systèmes, des ampoules LED placées sur le rail de sécurité d'un pont enjambant une autoroute ont été vandalisées et détruites.

Document 2

Une pile microbienne à plantes pour l'électricité de demain ?

Quentin Manguit, Futura

Publié le 27/11/2012

Des plantes vivantes pourraient être utilisées pour la production de bioélectricité ! Un prototype de pile microbienne vient d'être réalisé, alimenté par du glucose issu de la photosynthèse d'une plante. Les marais ou les rizières du globe pourraient de cette manière devenir des centrales électriques, ainsi que nos toits !



Figure 1 : Dans certains pays en voie de développement, la mise en place de piles microbiennes à plantes serait aussi rentable que l'installation de panneaux photovoltaïques. Il s'agit d'un point positif pour le développement de cette filière. © Rémy Saglier-Doubleray, Flickr, cc by nc sa 2.0

Marjolein Helder de l'université de Wageningen (Pays-Bas) et David Strik pourraient avoir trouvé une source d'énergie originale. Ces chercheurs ont utilisé la photosynthèse pratiquée par des végétaux supérieurs pour nourrir des bactéries produisant des électrons. Le projet est certes toujours expérimental, mais des tests se sont montrés concluants. D'ici quelques années, estiment-ils, des habitations pourraient être alimentées par des plantes cultivées sur les toits !

Les plantes ont recours à la photosynthèse pour fabriquer des sucres à partir d'eau, de gaz carbonique (CO₂) et de lumière. Or, 40 à 70 % de ces sucres ne sont pas utilisés par ces organismes. Ils sont donc rejetés dans l'environnement par les racines, pour le plus grand plaisir des bactéries du sol. Celles-ci dégradent ces composés pour se fournir en énergie. C'est sur cette étape que les chercheurs ont décidé d'agir.

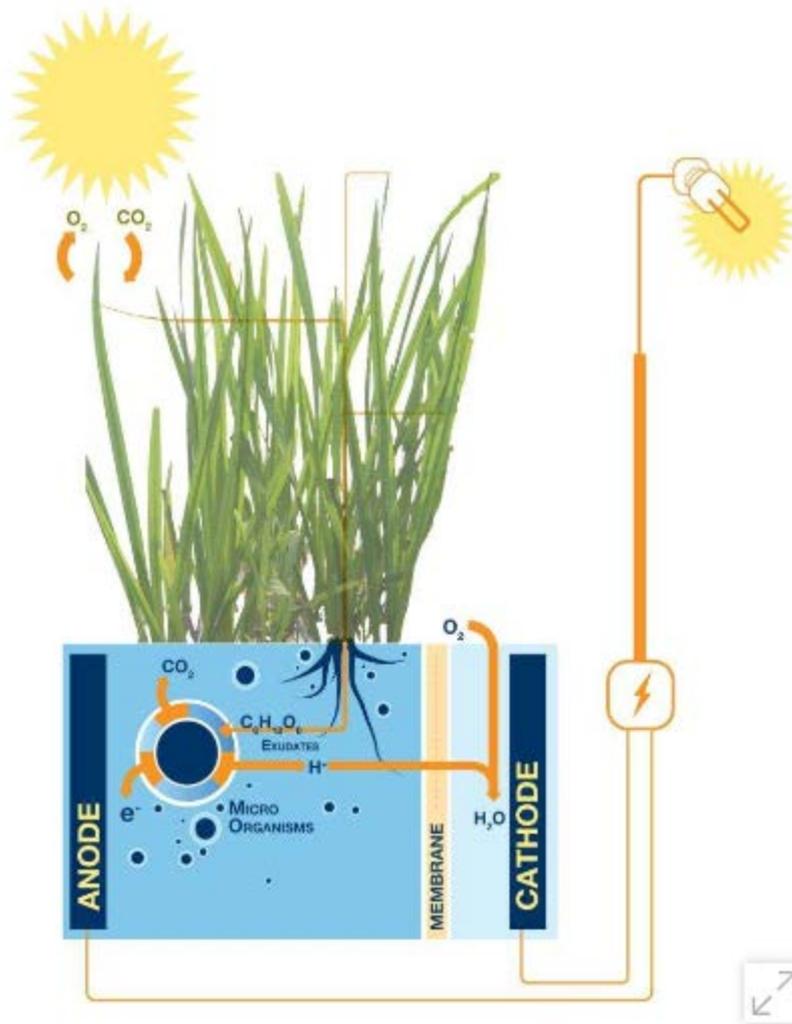


Figure 2 : Principe de fonctionnement d'une pile microbienne à plantes. Les sucres (C₆H₁₂O₆) produits par la photosynthèse des plantes sont dégradés par les micro-organismes présents dans le milieu (Micro-Organisms). Ils produisent en retour du CO₂, des protons (H⁺) et des électrons (e⁻) captés par l'anode. En se déplaçant vers la cathode, ces charges génèrent un courant électrique. Au niveau de la cathode, les protons qui ont migré à travers une membrane réagissent avec les électrons et le dioxygène de l'air (O₂) pour donner de l'eau (H₂O). © PlantPower.eu

Une pile microbienne alimentée par des plantes

Des végétaux ont été mis en culture dans un milieu contenant des micro-organismes. En dégradant les exsudats (les sucres libérés), ces bactéries produisent du CO₂, des protons (H⁺) et des électrons récupérables par une anode placée à proximité des racines. La cathode est quant à elle fixée à l'intérieur d'un second compartiment séparé du premier par une membrane perméable aux protons. La différence de potentiel entre les deux milieux engendre un courant électrique. Au final, les protons arrivés dans le second compartiment par diffusion vont réagir

avec des molécules de dioxygène (O₂) et des électrons issus de la cathode pour former de l'eau (H₂O).

Lors de tests, la production a atteint 0,4 watt par mètre carré (W/m²) de plantes en cours de croissance, soit plus que le courant généré par diverses autres piles microbiennes exploitant la fermentation de la biomasse. Dans le futur, la productivité du système pourrait atteindre 3,2 W/m². Un toit plat de 100 m² fournirait alors suffisamment d'électricité à l'année pour alimenter une habitation (soit en moyenne 2.500 kWh/an en France). Depuis 2009, le projet est développé par Plant-e, une spin-off créée par les deux chercheurs.

Pas de conflit pour l'exploitation des terres agricoles

La pile microbienne à plantes (ou Plant-MFC pour Plant Microbial Fuel Cell) pourrait prochainement être installée sur des toits plats ou dans certains pays en voie de développement. En effet, le dispositif peut être enterré au sein de zones humides exploitées par l'agriculture sans en gêner l'utilisation, par exemple dans des rizières, ou encore en milieu marécageux. Détail intéressant, la pile fonctionnerait avec une grande variété de plantes. Au final, elle se révèle donc non polluante, discrète et durable.

Cependant, quelques détails doivent encore être améliorés avant le développement à grande échelle de ce procédé. Les chercheurs se plaignent notamment de la trop grande quantité de matériaux qu'ils doivent encore utiliser pour construire les électrodes. Des solutions seraient déjà envisagées. En tout cas, ce moyen de production d'électricité a de quoi rencontrer un certain succès. À suivre donc...

Document 3

Retour vers le futur : une brève histoire de l'exploration de Mars

Cyril Szopa

Professeur des Universités, Exobiologiste au Laboratoire Atmosphères Modélisation et Observations Spatiales (LATMOS), Université de Versailles Saint-Quentin-en-Yvelines (UVSQ) – Université Paris-Saclay
Publié: 7 juin 2021,

<https://theconversation.com/retour-vers-le-futur-une-breve-histoire-de-l-exploration-de-mars-159385>

L'atterrissage de Perseverance, rover de la NASA, en février 2021, puis celui de Zhurong de l'agence spatiale chinoise en mai 2021, sont venus grandir le nombre d'engins motorisés à la surface de la planète rouge.

Il y a aujourd'hui six véhicules mobiles, appelés rovers, à la surface de Mars – Sojourner, Spirit, Opportunity, Curiosity, Perseverance, Zhurong – dont trois seulement sont en activité. Les cinq premiers ont été déployés à la surface de Mars par la NASA depuis la fin des années 1990 et viennent d'être rejoints par le rover de la mission chinoise Tianwen-1. Ce dernier est un rover géologue, et il permettra de mieux comprendre la structure et les roches de la planète rouge. Perseverance, lui, lance une épopée fantastique : faire revenir sur Terre un peu de Mars.

Pour comprendre pourquoi ces missions sont conçues ainsi et le futur de l'exploration martienne, retour vers le futur.

Les ancêtres des rovers sont des sondes fixes, les « atterrisseurs »

Les toutes premières sondes à se poser avec succès à la surface de Mars étaient deux atterrisseurs fixes appelés « Viking », de la mission du même nom. Elles ont atterri en 1976. L'ambition de cette mission, impulsée par un des fondateurs de l'exobiologie, Carl Sagan, était forte, car elle visait à détecter des activités biologiques liées à de la vie bactérienne supposée être présente à la surface de Mars.

Malheureusement, en dépit de son succès technique, les mesures réalisées par la sonde Viking n'ont pas permis de mettre en évidence de telles signatures. Elles ont en revanche fourni les premières mesures sur les conditions de surface, en particulier des données météorologiques sur plusieurs années. Celles-ci ont beaucoup fait progresser notre connaissance de l'environnement à la surface de Mars, ce qui a permis de préparer les missions ultérieures.

Néanmoins, l'absence de résultats concernant la vie, associée à une série d'échecs de missions lancées vers Mars dans les années 1980 et 1990, a longtemps repoussé les autres tentatives, jusqu'en 1997.

Le premier rover atterrit sur Mars au tournant du XXI^e siècle

Le petit robot Sojourner de la NASA, de la taille d'une voiture radioguidée, avait pour but de montrer la faisabilité de commander et faire évoluer un rover sur Mars : c'était un démonstrateur technologique. Il n'embarquait alors qu'un seul instrument scientifique et n'a roulé que quelques mètres à la surface de Mars.

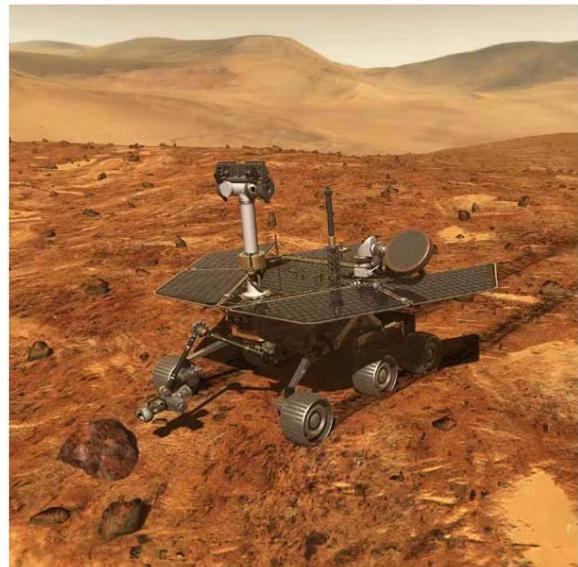
Fort de cet apprentissage, la NASA a ensuite développé les Mars Exploration Rovers, appelés Spirit et Opportunity, qui sont désormais célèbres. Spirit et Opportunity ont arpenté la surface de Mars pendant plusieurs années à partir de 2004. Ces véhicules mesuraient 1,5 mètre de haut et pesaient environ 200 kilogrammes. Spirit a évolué pendant 7 ans environ dans le cratère

Goussev et a été arrêté, car il ne pouvait plus se déplacer et plusieurs de ses instruments étaient en panne.

Opportunity a lui fonctionné pendant près de 15 ans dans la région de *Terra Meridiani*. Ce dernier a parcouru plus de 45 kilomètres à la surface de Mars, ce qui restera un record pour plusieurs années encore (voire plusieurs décennies selon les longévités des rovers actuels). La mission a été arrêtée en février 2019 suite à une tempête de poussière inattendue qui n'a pas permis au robot de recharger ses batteries via ses panneaux solaires.



Le cratère Goussev, ou Gusev, où le rover Spirit a évolué pendant 7 ans. NASA/JPL



Un modèle du rover Spirit, vue d'artiste. NASA/JPL

Les missions des rovers Spirit et Opportunity

Une des grandes missions historiques des rovers martiens est l'étude de l'eau à la surface de la planète.

La plupart des sondes orbitales qui ont atteint et étudié Mars au XX^e siècle avaient des capacités d'observation permettant de faire de l'imagerie et un peu de mesures spectrales, mais l'essentiel des indices d'eau récoltés depuis l'espace étaient des indices géomorphologiques (des formes de terrain observées à la surface pouvant être en lien avec des écoulements liquides).

Malheureusement, ces observations ne permettaient pas de savoir à quelles périodes pourraient avoir eu lieu ces écoulements, ni leur durée. La sonde orbitale Mars Global Surveyor de la NASA a confirmé cet état de fait à la fin des années 1990 et a montré également que l'évolution climatique à la surface a été complexe. Mais l'influence de l'eau sur cette évolution climatique et son interaction avec l'atmosphère ne pouvaient être comprises avec ce type d'observation à distance. Il devenait nécessaire d'envoyer des rovers pour obtenir des informations directement à la surface, au plus près des roches qui ont été formées, ou transformées, par l'eau liquide.

Spirit et Curiosity ont découvert des roches sédimentaires et des sels, qui montrent que l'eau a coulé et perduré pendant de longues périodes, et ce sur les deux sites d'exploration des rovers pourtant séparés de plusieurs milliers de kilomètres. Ceci laisse penser que l'eau a dû être abondante et largement répandue à la surface de Mars.

De plus, ces rovers ont montré l'importance d'être mobiles dans l'exploration, car les découvertes qu'ils ont permises sur l'eau liquide ont été réalisées hors de leur site d'atterrissage. Ceci a fini de convaincre la NASA, et d'autres agences spatiales aujourd'hui, que les rovers étaient des éléments essentiels dans le cadre d'une stratégie d'exploration de surface planétaire en dépit de leur plus grande complexité de développement (en comparaison avec les atterrisseurs).

Le rover Curiosity, symbole d'une évolution de l'exploration de Mars

En 2004, la NASA a donc retenu la mission Mars Science Laboratory pour envoyer vers Mars un rover d'exploration encore plus imposant – Curiosity.

La stratégie scientifique et exploratoire évolue vers le schéma actuel de l'exploration de Mars établi par la NASA : après avoir étudié une partie de la géologie martienne à la surface, et avoir montré l'existence d'eau liquide pérenne par le passé avec les Mars Exploration Rovers, l'objectif fixé à Curiosity était de montrer que la surface de Mars a pu être « habitable ». Derrière ce terme, on entend que Mars a pu réunir en un même lieu et au même moment les ingrédients que l'on considère nécessaires à la vie pour se développer et perdurer, à savoir : de l'eau liquide, une forme d'énergie et les composants nécessaires à former les matériaux biologiques. Les plus emblématiques en sont les molécules organiques, mais il ne faut pas oublier des composants minoritaires comme le phosphore par exemple.

Ce terme d'habitabilité, au « conditionnel », a émergé pour signifier que si ces conditions sont réunies, elles ne suffisent pas à garantir que la vie est forcément apparue.

L'objectif principal fixé à Curiosity était donc de montrer si le cratère Gale, dans lequel le rover a été envoyé, a pu être habitable. Si on sait que l'énergie solaire était bien évidemment disponible, il fallait au rover trouver des indices d'eau pérennes passés, à l'instar des Mars Exploration Rovers, mais également mettre en évidence la présence de molécules organiques et d'éléments mineurs. Pour cela, Curiosity a été fabriqué beaucoup plus grand que ses prédécesseurs : plus de 2 mètres de haut. Il embarque une dizaine d'instruments d'analyses, ainsi qu'un système de prélèvement de roches martiennes avec une foreuse fonctionnant sur le principe d'une perceuse à percussion.

Curiosity détecte de la matière organique

De plus, de par sa taille, le robot utilise de l'énergie électrique produite par un petit réacteur nucléaire ce qui lui permet de fonctionner comme il l'entend sans dépendre de la recharge de ses batteries par des panneaux solaires.

Grâce à ses performances hors norme, non seulement le robot a mis en évidence de nombreux indices géologiques et minéralogiques de la présence d'eau liquide pérenne dans le cratère par le passé, mais il a également permis la détection de matière organique endogène qui était recherchée sans succès depuis les missions Viking. Cette matière a été détectée par l'instrument Sample Analysis at Mars – limitation des conditions expérimentales sur Mars : le mode d'analyse utilisé et la présence de composés inorganiques inattendus dans le sol n'ont permis détecter que des molécules organiques ayant réagi avec des espèces inorganiques. Ces réactions ont produit des espèces chlorées, dont le chlorobenzène emblématique de ces analyses, et des espèces organiques soufrées.

À partir de ces espèces, nous tentons de remonter à la nature réelle de la matière organique présente dans l'échantillon, mais c'est un travail d'investigation long et fastidieux. Ce qu'il est possible de dire à ce jour, à partir de l'analyse de plusieurs échantillons collectés sur le trajet du rover jusqu'à aujourd'hui, c'est que cette matière organique est présente en très faible concentration dans la couche superficielle de la surface dans laquelle Curiosity prélève les échantillons (dans les 10 premiers centimètres).

Ceci tendrait à montrer l'influence forte des conditions qui règnent à la surface de Mars sur la pérennité de la matière organique : la présence d'oxydants comme les perchlorates et de rayonnements ionisants en quantité à la surface peut mener à l'évolution et la destruction de ces molécules.

Néanmoins, à travers cette détection directe de matière organique, Curiosity a prouvé que le cratère Gale a pu être habitable par le passé.

Le futur rover Rosalind Franklin de l'ESA cherchera des molécules organiques enfouies plus profondément

L'Agence Spatiale Européenne, qui développe la mission Exomars en partenariat avec l'agence spatiale russe Roscosmos, a décidé de forer plus en profondeur, jusqu'à 2 m environ, pour y rechercher des échantillons qui ont pu être protégés des conditions agressives de la surface [d'après des modèles théoriques](#). Nous ne saurons qu'à la suite de l'analyse des premiers échantillons, en 2023, si tel est le cas, et nous serons alors capables d'évaluer l'influence des conditions de surface en fonction de la profondeur.

De plus, l'ESA a sélectionné plusieurs instruments qui apporteront des informations complémentaires sur la minéralogie et le contenu en matériaux organiques des échantillons, ce qui permettra de plus facilement reconstruire le puzzle des matériaux présents dans les échantillons, en cas de détection de matière organique.

Document 4

Mars : découverte de molécules organiques inédites par Curiosity

Dorian De Schaepe meester

Publié le 05/11/2021

<https://www.futura-sciences.com/sciences/actualites/curiosity-mars-decouverte-molecules-organiques-inedites-curiosity-94635/>

Lors d'une collecte d'échantillons du rover Curiosity sur Mars, en 2017, un dysfonctionnement dans la manœuvre aurait permis de découvrir de nouvelles molécules organiques encore inédites à la surface de la Planète rouge.

En presque dix ans de présence sur Mars, le rover Curiosity n'a pourtant pas fini de surprendre. Une étude publiée dans Nature Astronomy le 1er novembre rapporte les résultats d'analyses menées sur des échantillons prélevés par l'astromobile en 2017. Lors de l'opération de prélèvement, les roches martiennes auraient été stockées dans un réceptacle contenant un liquide, destiné à mener des expériences nommées en anglais wet experiment (littéralement : « expérimentation mouillée »). La réaction chimique induite aurait permis d'observer la présence de molécules organiques contenues dans les échantillons

Deux nouveaux types de molécules organiques

Arrivé dans le cratère Gale en août 2012, Curiosity s'est immédiatement lancé dans la recherche de traces de vie passée, neuf ans avant Perseverance et Zhurong. Par le biais d'études chimiques, physiques et géologiques, le rover avait pour tâche de détecter ces biosignatures. En 2018, des roches sédimentaires prélevées dans le cratère avaient été entreposées dans l'instrument Sample Analysis at Mars (SAM) afin d'analyser leurs compositions. Certaines roches contenaient ainsi plusieurs molécules telles que du carbone, du sulfure, du benzène ou encore du thiophène. Cette découverte avait alors suscité l'intérêt de la communauté scientifique, pointant vers de possibles biosignatures.



L'instrument SAM, inséré dans le « ventre » de Curiosity, permettant de mener diverses expériences sur les échantillons prélevés. © Nasa, JPL-Caltech

En 2017, Curiosity connaissait un dysfonctionnement lors d'une opération de prélèvement de sables dans les dunes de Bagnold. Les ingénieurs ont donc sciemment disposé les échantillons dans des tubes contenant un agent liquide nommé N-méthyl-N-(triméthylsilyl)trifluoroacétamide, ce dernier permettant ainsi de mener des analyses chimiques lorsqu'il est mélangé à d'autres composants chimiques. Pas de biosignature pour les chercheurs de la Nasa, mais la découverte d'ammoniac et d'acide benzoïque. Ce dernier, dérivé du benzène, est présent sur Terre dans certaines plantes telles que la canneberge ou le cacaoyer. L'ammoniac est généralement observé sous forme de gaz, émanant de l'écorce terrestre et d'autres sources telles que la fermentation bactérienne ou certains hydrocarbures tels que le pétrole et le charbon

Si ces molécules ne sont pas la marque certaine de traces de vie sur Mars il y a 3 milliards d'années, à l'époque où de l'eau était présente sur la Planète rouge, elles pourront en revanche renseigner les scientifiques sur l'évolution géologique de Mars.

Un trio de robots à la recherche de la vie sur Mars

Curiosity devrait continuer à sillonner les terres désolées de Mars à la recherche de biosignatures. Les chercheurs ont attesté de l'efficacité de cette méthode, sans affirmer clairement si elle sera répétée ou améliorée, l'étude expliquant que « la dérivation effectuée [...] s'est révélée être un puissant outil étendant notre domaine de recherche de molécules organiques ou biologiques ».



Photographie du rover Zhurong, dans la région d'Utopia Planitia. © CNSA

En parallèle du travail effectué par Curiosity, Perseverance et Zhurong, arrivés en février et mai 2021, s'affairent aussi à rechercher des traces de vie passée sur Mars. Les roches forées par Perseverance démontrent que de l'eau était bien présente sur Mars il y a 3 milliards d'années grâce à la découverte de sels minéraux dans le cratère Jezero. Concernant Zhurong, la Chinese National Space Agency (CNSA) n'a pas annoncé d'avancée majeure dans la quête aux biosignatures, mais grâce au prolongement officiel de sa mission, l'astromobile continue d'explorer les plaines d'Utopia Planitia avant que son autonomie ne soit épuisée.

Document 5

Les robots mobiles sur Mars : des moyens irremplaçables d'étude

Pierre Thomas

Laboratoire de Géologie de Lyon / ENS de Lyon

D'après l'article publié le 08/01/2021

<https://planet-terre.ens-lyon.fr/ressource/robots-mobiles-Mars.xml>

Pourquoi aller sur Mars, et pourquoi avec des robots mobiles ?

Mais pourquoi étudier Mars plus que Vénus ou Mercure par exemple ? C'est en 1965 avec la sonde Mariner 4 (simple survol) et surtout en 1972 avec la mission Mariner 9 (mise en orbite) qu'on a commencé à vraiment découvrir la surface de Mars. Cette première mission orbitale, puis toutes celles qui ont suivi ont montré que Mars, par certains côtés, ressemblait plus à la Terre qu'à la Lune, la seule référence qu'on avait à l'époque. En plus des calottes polaires qu'on connaissait depuis longtemps (elles ont été découvertes en 1666 par Cassini) et une atmosphère, on voyait des failles et des volcans géants, des lits de rivières... Les rivières étaient asséchées, les volcans semblaient éteints, mais cette "inactivité" semblait relativement récente (géologiquement parlant), c'est-à-dire postérieure à la fin du bombardement intense des premiers temps du système solaire. Ce qu'on voyait posait plein de questions purement scientifiques, dont une qui va plus loin que la seule soif de connaissances à cause de ses implications "philosophiques" : au temps où les volcans crachaient et les rivières coulaient, Mars a-t-il été habitable, voire habité, ne serait-ce que par les plus simples des micro-organismes possibles ? Il ne faut pas oublier que, quand de l'eau coulait sur Mars, la vie naissait sur Terre. Et si une forme de vie est née sur Mars, cette planète ne serait-elle pas encore habitée dans quelques recoins à l'abri du rayonnement solaire et des peroxydes qu'il engendre ?

Pour répondre à ces questions, les études « à distance » depuis l'orbite (et il y en a eu beaucoup), si elles sont nécessaires, ne suffisent pas. Pour compléter les apports des sondes en orbite, de leurs caméras, de leur spectromètres (optiques, infrarouge, gamma...), il faut que des géologues aillent sur place indirectement par robots interposés, aillent « sur le terrain » pour employer leur jargon professionnel. Indirectement sur le terrain, ils vont faire le travail dont ils ont l'habitude : analyser l'état actuel de l'affleurement à toutes les échelles et reconstituer ce qui s'est passé sur ce site et à telle ou telle époque. Sur Terre, les géologues font ce travail soit pour des raisons de recherche fondamentale (quelle est l'origine de l'Himalaya, par exemple, où trouver des traces de vie les plus vieilles possibles) soit pour des raisons très appliquées (trouver du pétrole ou des minerais, par exemple, ou étudier la stabilité d'un versant...). Sur place, le travail du géologue terrestre consiste certes à reconnaître les roches (la première et souvent la seule chose à laquelle pense un étudiant en géologie lors de sa première sortie sur le terrain), mais aussi à étudier la géométrie des affleurements pour retracer leur dynamique sédimentaire ou volcanique, à étudier leurs éventuelles déformations, altération..., à déterminer là où il va faire des mesures, et enfin à choisir les roches qu'il va prélever et ramener au laboratoire pour des analyses ultérieures.

Le robot Perseverance de la mission Mars 2020 a pour but ultime de sélectionner des échantillons, de les prélever et de les mettre « en dépôts », pour que de futures missions (à la fin des années 2020 / au début des années 2030 si tout se passe comme prévu) viennent les rechercher pour les ramener sur Terre. Ramener des échantillons sur Terre est en effet nécessaire pour analyser ultra-finement les éventuelles molécules carbonées qui se trouveraient dans les échantillons, déterminer si ces molécules ont un "intérêt" prébiotique, ou même si elles sont d'origine biologique, dater les roches par radiochronologie... Les instruments capables de faire cela sont en effet trop massifs pour être spatialisés dans l'état actuel de techniques et des lanceurs. Mais ces échantillons à ramener sur Terre ne doivent pas être choisis "au hasard"

contrairement à ceux ramenés de la Lune en décembre 2020 par la mission chinoise Chang'e-5 et prélevés sur place, là où s'est posé la sonde Chang'e-5. La collecte des échantillons martiens à ramener se fera "intelligemment". Elle se fera là où les analyses morphologiques, puis minéralogiques et chimiques faites in situ suggèreront qu'ils sont "intéressants". Et les analyses détaillées in situ, morphologiques et/ou chimiques ne se feront que là où des géologues, via les caméras du robot, penseront que les conditions géologiques (eau, argile...) auront été propices à de la chimie prébiotique, voire à une forme de vie, et à leurs préservations.

Dans le cas de Mars, on peut aller sur le terrain « par procuration » avec des robots, mais on pourra aussi y aller « en vrai » avec des missions habitées dans un futur plus ou moins lointain. Et un bon géologue sur Mars peut faire bien plus qu'un robot. Par exemple, un géologue équipé d'un véhicule semblable à la jeep lunaire des missions Apollo 15 à 17 aurait pu en moins d'une semaine obtenir la quasi-totalité des résultats qu'Opportunity a mis 14 ans à obtenir. Et si, en plus, il ramène des échantillons lors de son retour sur Terre... Mais si et quand les problèmes "techniques" des missions habitées seront résolus, quand il s'agira de choisir entre une (des) mission(s) habitée(s) ou une (des) mission(s) robotisée(s), il se posera le problème du rapport "bénéfices scientifiques escomptés / couts financiers" entre ces deux options. En effet, on peut estimer le cout d'une mission habitée à 50 à 100 fois celui d'une mission robotisée. Ne vaudrait-il pas mieux, si on n'a que des buts purement scientifiques, explorer avec des robots 100 sites soigneusement choisis plutôt qu'en étudier un seul par des géologues en chair et en os ? Mais si les futurs explorateurs martiens ont le moyen de se déplacer sur très de longues distances à la surface de Mars, cela change les données du problème. Et aux buts scientifiques, se surajouteront forcément des buts politiques...

Les anciennes questions et les résultats déjà acquis

Cela fait donc presque 50 ans que l'on sait que de l'eau a coulé sur Mars. Mais cette eau était-elle pérenne ou très épisodique, douce ou salée, neutre, acide ou basique... ? Si la vie terrestre peut habiter dans des milieux aux conditions extrêmes (acides, basiques, haute température..., cf. Les extrémophiles dans leurs environnements géologiques - Un nouveau regard sur la biodiversité et sur la vie terrestre et extraterrestre), la chimie prébiotique terrestre semble nécessiter des pH neutres ou légèrement basiques mais semble impossible en milieu acide (cf. L'origine de la vie sur Terre vue par un géologue : quoi de neuf depuis 2015 ?). Et la vie terrestre nécessite une certaine "constance".

Spirit s'est posé en 2004 au fond d'un cratère à fond plat. Les images prises depuis l'orbite laissaient penser (il y a plus 16 ans) que le fond de ce cratère était recouvert de sédiments déposés dans un ancien lac. L'étude de ces sédiments devait permettre de reconstituer les conditions de l'époque de leur dépôt. Fatale erreur ! Les observations et analyses faites sur place ont montré que ce cratère était rempli d'anciennes laves et cendres basaltiques, qui avaient recouvert les éventuels sédiments lacustres. Mais, aussi bien les analyses chimiques (faites automatiquement par le robot) que l'interprétation de la géométrie des affleurements d'après les photographies faites au sol, ont montré que ces éruptions volcaniques avaient parfois eu lieu sous l'eau d'un lac ou dans des terrains gorgés d'eau.

Opportunity (en 2004) et *Curiosity* (en 2012) se sont posés, eux, sur des terrains faits de roches sédimentaires bien stratifiées comme on l'avait déterminé (sans erreur cette fois) avec les sondes en orbite. En analysant les milliers de photographies faites durant leurs trajets et les analyses chimiques et minéralogiques effectuées sur place, on a pu reconstituer dans quels milieux et paysages s'étaient déposées ces roches il y a 3 à 4 milliards d'années (Ga) et ce qui leur était arrivé depuis. Dans ce passé lointain, la région de Mars où s'est posé Opportunity correspondait

à une plaine parfois recouverte de dunes éoliennes faites de sable basaltique, parfois inondées sous une faible tranche d'eau (quelques centimètres), eau temporaire et alors parcourue de courants déposant du sable. Cette eau était salée et très acide. Curiosity s'est posé dans une région successivement occupée, dans ce même passé lointain, par deux épisodes lacustres (lacs plus ou moins profonds et temporaires), séparés par une longue période d'émersion avec érosion. Les eaux de ces lacs étaient salées, mais non acides. Les sites où a roulé Curiosity pendant ses trois premières années correspondaient aux bords du lac le plus récent, bords parfois sous l'eau, parfois en position littorale avec des dunes éoliennes. Puis, au bout de 3 ans, Curiosity a atteint des couches sous-jacentes, qui correspondaient au centre du lac le plus vieux. Ces deux lacs successifs parfois s'asséchaient temporairement et étaient remplacés par des plaines vaseuses ou des champs de dunes. Dans l'ancienne vase, Curiosity a trouvé des traces de matière carbonée, mais trop dégradée et en teneur trop faible pour savoir si elle avait ou non une origine biologique. Après le dépôt de ces sédiments, aussi bien dans la région explorée par Opportunity que dans celle explorée par Curiosity, de l'eau circulait en profondeur dans le sous-sol, en altérait les roches et déposait des minéraux dans les fractures.

Pour choisir les affleurements à étudier en détail, puis pour tirer ces conclusions, les géologues des équipes scientifiques qui se trouvaient derrière les caméras et autres instruments des robots devaient avoir une très bonne culture géologique et une solide expérience du terrain.



Source - © 2013 NASA / JPL-Caltech / MSSS

Images brute et annotée montrant la géométrie de détail des couches de grès de Yellowknife Bay.

Ce dispositif de strates de 2e ordre inclinées contenues dans des strates de 1er ordre plus épaisses et horizontales est appelé "stratifications entrecroisées". Un tel dispositif est caractéristique de sable déposé par des courants, coulant dans ce cas de droite à gauche.

Et Perseverance...

Avec des vues orbitales et d'autres prises par Perseverance (robot de la mission Mars 2020 qui doit se poser le 18 février 2021 sur la planète rouge) avec des caméras à grand angle, les ingénieurs et les géologues choisiront les sites à visiter et le trajet à prendre pour y aller. Avec les caméras à plus petit champ, donc à meilleure résolution, et avec le spectro-imageur à distance SuperCam, les géologues choisiront quelle(s) partie(s) du site étudier à haute résolution, « par contact ». Au contact de l'affleurement et grâce à un bras porte-outils, la spectrométrie X à haute résolution, la fluorescence UV et la diffusion Raman permettront aux ingénieurs et aux géologues de connaître les compositions chimique, minéralogique et « organique ». En fonction de ces résultats, les géologues décideront ou non de prélever des mini carottes à ramener sur Terre dans une dizaine d'années. Comme Perseverance ne pourra recueillir au maximum que 43 carottes d'une dizaine de grammes chacune, les géologues n'auront pas intérêt à se tromper dans leur choix au risque de rapporter des « cailloux sans intérêt ».

Séances : Analyse de données scientifiques

Pensez à apporter des ciseaux, de la colle et une calculatrice !

A- Exploitation de données expérimentales

1- Objectifs :

- acquérir des notions de base de traitement des données
- développer un esprit critique vis à vis de données obtenues ou d'autres données déjà traitées

2- Les différents types de données :

a- Les données peuvent être classées en catégories = les variables

- Variables discontinues (Compter les blancs et les noirs, répondre à un questionnaire)
- Variables continues (mesures, poids ...) Chiffres après la virgule, si instruments très précis

b- Elles ne possèdent pas toutes le même niveau d'information (richesse d'information différente)

- Données qualitatives :
 - Binaires (1/0 ; oui/non ; +/-, mort/vivant)
 - Nominales (état civil, Homme/Femme, profession)
- Données ordinales ou semi-quantitatives
 - Donne une information plus riche ; ici on classe (marches d'escalier, rangs, grades...)
 - Pas de moyenne possible...
- Données quantitatives :
 - discontinues : dénombrement (donc nombres entiers comme le nombre d'enfants dans une famille par ex)
 - continues : mesures (problème de l'unité et de la précision, par ex nb de chiffres après la virgule)

3- Traitement et analyse critique des données

a- Traitement des données = trouver la représentation la plus judicieuse

La représentation choisie n'entraîne-t-elle pas de perte d'information ?

Obtient-on les mêmes conclusions en changeant la représentation ?

Comment représenter ces données ? (proposer différentes formes de représentation) ; le résultat obtenu est-il significatif ?

* Exemple :

Vous travaillez pendant vos vacances dans une grande enseigne de sport. Votre responsable vous demande de réaliser une étude de marché des prix des vélos pour enfant en Île-de-France. Vous devez lui présenter le résultat de votre recherche sous forme de graphique. Prix d'un VTT pour enfant dans tous les magasins d'île de France

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

La moyenne est-elle suffisante pour comparer des données ?

- ⇒ Ecart-types
- ⇒ Tri des données

Prix d'un VTT pour enfant dans les mêmes magasins classés en fonction de la distance par rapport à l'enseigne où vous travaillez.

distance (en km)	01-50	51-100	101-150	151-200	201-250
prix des VTTs	93	67	71	73	80
	104	62	98	62	71
	97	66	69	91	80
	105	57	75	68	69
	99	62	69	72	72
	97	52	74	74	75
	105	68	75	76	77
	109	68	78	79	79
	97	62	79	82	83
	101	65	85	88	89

Comment intégrer cette nouvelle information dans le cadre de la demande de votre responsable ?

B- Mise en forme et analyse de données expérimentales

L'analyse de documents scientifiques telle qu'elle est proposée ici nécessite des réponses aux questions posées dans chaque figure. Il s'agit alors de mettre en œuvre une démarche dans l'analyse aboutissant à la réponse demandée, et pas seulement de donner la réponse, donc de justifier cette réponse. La réponse s'articulera autour de deux attitudes : j'observe que... je déduis que.... Il ne faut pas oublier cependant que nous sommes des biologistes !

Ces séances constituent une préparation à l'analyse d'article.

L'obésité

Véritable problème de santé publique, l'obésité est une maladie chronique, facteur de nombreuses maladies graves et qui raccourcit l'espérance de vie des personnes atteintes. Accéder à une meilleure compréhension des causes et des mécanismes biologiques conduisant à l'obésité est aujourd'hui un des plus grands enjeux de la recherche. Comme toutes les maladies chroniques, l'obésité devient en effet irréversible lorsqu'elle est installée : prévenir son développement est donc primordiale si l'on veut enrayer l'épidémie mondiale.

L'objet de cette étude (inspirée des données d'un article publié dans le journal PNAS en 2005 : Wang *et al*, 2005, PNAS 102, 18011-18016) est d'obtenir une preuve directe d'une relation entre le régime alimentaire et l'obésité.

Lors de votre stage de L2, vous étudiez l'effet d'un régime hyperlipidique sur le poids des rats et la structure des adipocytes.

Vous disposez de 2 lots de 10 rats :

1° lot : lot témoin nourri avec un régime à 4% de matières grasses

2° lot : lot nourri avec un régime à 60% de matières grasses

Dans un premier temps, **la masse de ces rats** a été suivie pendant 20 semaines après leur naissance. Les résultats sont reportés dans le tableau 1. L'apparence de ces rats à l'issue des 20 semaines est illustrée dans sur la photo ci-dessous.

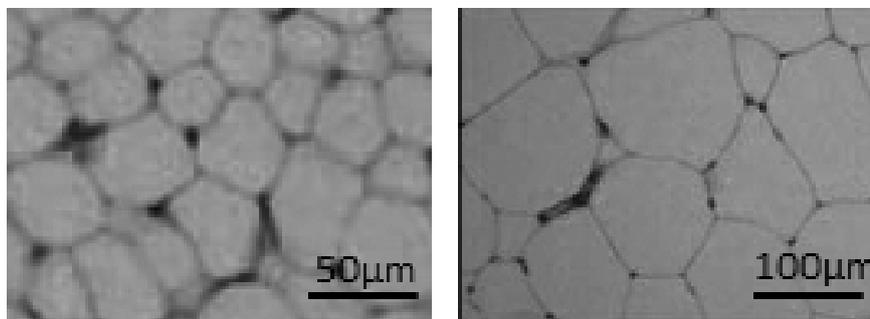


Photographie (à 20 semaines) d'un rat témoin, soumis à un régime alimentaire à 4% de matières grasses (gauche), et d'un rat soumis à un régime alimentaire à 60% de matières grasses (droite).

Tableau 1 : suivi du poids (en g) des rats témoins et des rats soumis à un régime à 60% en matière grasse

	Semaine n°					
	1	2	5	10	15	20
Lot témoin	145,2	163,5	360,8	436,3	569,4	620,1
	143,8	209,3	339,4	442,1	534,9	594,4
	138,6	210,9	379,4	468,1	507,1	663,6
	159,4	208,1	355,9	476,9	546,5	612,8
	102,9	233,6	327,6	438,1	493,1	639,1
	160,7	211,9	362,7	473,9	519,6	608,7
	152,1	209,7	323,9	489,4	577,2	691,4
	150,6	199,9	381,2	468,5	519,4	666,3
	173,6	208,3	383,7	405,7	508,2	626,7
	153,8	207,9	307,1	460,3	566,8	602,4
Lot soumis à un régime à 60% M.G.	156,1	206,5	469,3	509,8	674,1	765,4
	151	215,8	433,6	546,3	739,5	798,1
	144,9	208,4	399,9	589,1	764,8	866,4
	176,8	210,4	403,6	556,7	693,4	810,2
	143,9	218,9	415,9	539,6	715,7	805,6
	142,8	209,8	376,5	482,9	712,5	834,6
	126,1	196,4	405,3	551,2	669,4	781,4
	153,4	231,6	428,1	549,3	666,6	842,4
	102,4	192,9	415,7	526,9	745,1	771
	149,6	215,5	398,4	562,9	705,6	832,8

Dans un second temps, une étude histologique a été réalisée afin de comparer le tissu adipeux des 2 populations de rats. Ci-dessous, une photo représentative de ces deux tissus (témoin à gauche et régime à 60% de matières grasses à droite).



- Construire une figure composite illustrant l'ensemble de ces résultats (voir annexes, doc 6).
- Décrire et interpréter ces résultats.

C- Lecture et interprétation d'images

Nous vivons dans un monde d'images. Comment les lire, les interpréter ?

Dans le domaine scientifique, on rencontre différents types d'images (photographies, schémas, représentations graphiques, protocoles expérimentaux...). Le premier « travail » sur une image est de comprendre à quel type elle appartient, comment elle a été obtenue. Il faut ensuite l'observer attentivement et réfléchir sur ce qu'elle véhicule comme information. Dans ce travail, on peut bénéficier de l'aide d'une légende, elle-même précédée d'un titre. Il est nécessaire dans ce cas de lire attentivement le titre et la légende et de voir s'ils sont ou non en adéquation avec ce que l'image montre.

Application : Lecture et interprétation de diverses images (annexes, doc7).

D- Application : Gestion et analyse de données biologiques

« Étude du volume des nématodes »

Dans le cadre de votre stage obligatoire de fin de LDD2, vous venez d'intégrer une équipe de recherche de l'Institut de Biologie Intégrative de la Cellule situé sur le campus du CNRS à Gif-sur-Yvette. Votre équipe d'accueil développe une collaboration avec un groupe de chercheurs japonais mené par Yasumi Ohshima au Département des Sciences de la Vie de l'Université de Sojo.

L'objet de cette collaboration est d'identifier les facteurs environnementaux et génétiques qui peuvent contrôler le volume du corps d'un nématode. Les nématodes sont des animaux bilatériens protostomiens de forme effilée, en général de petite taille, de l'ordre du millimètre. Les nématodes qui sont analysés dans cette étude se nourrissent de bactéries.

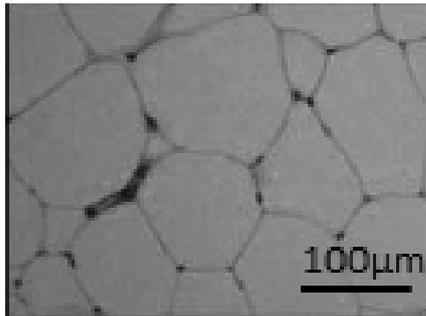
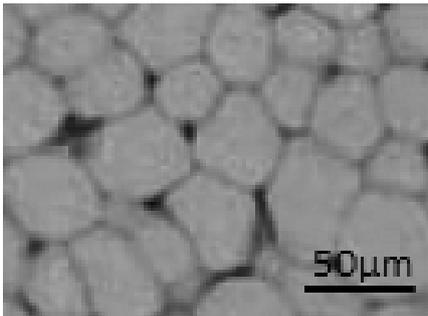
L'équipe japonaise a mesuré le volume des individus chez 4 espèces différentes de nématodes (en nanolitre, nL) :

- *Rhabditoides regina* (*R.regina*)
- *Pellioditis typical* (*P.typical*)
- *Oscheius dolichuroides* (*O.dolichuroides*)
- *Oscheius sp* (*O. sp.*)

Les mesures ont été effectuées sur des nématodes mâles (M) et femelles (F) nourris avec deux souches différentes de bactéries *Escherichia coli* : OP50 et HB101. Les résultats de ces mesures sont listés dans les tableaux 1 et 2 et sont aussi illustrées par les photos numérotées de 1 à 8

Votre équipe en France a généré des données sur l'espèce *C. elegans* : les mesures de volume ont été réalisées chez la souche de référence (contrôle) et chez deux souches de mutants où les gènes *daf-2* et *dpy-5* sont mutés. Ces données sont présentées dans le tableau 3 et par les photos 9 à 11.

Présentez et analysez ces résultats et exposez vos conclusions dans un document dont le format sera compréhensible de tous : sous forme d'un compte-rendu expérimental.



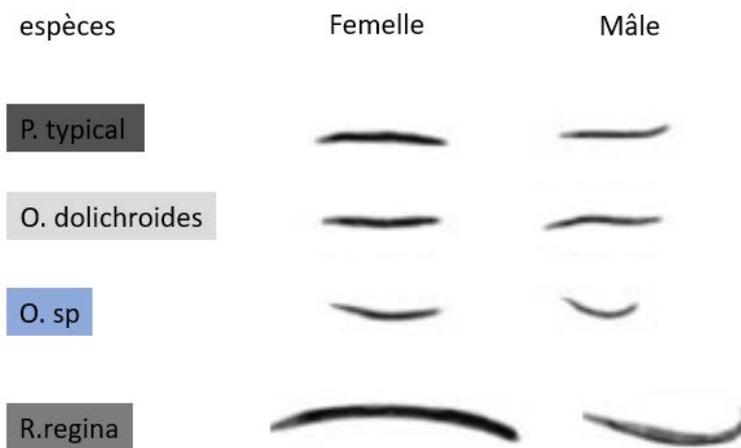
Etude de l'influence de la souche de bactéries ingérées et du sexe sur le volume du corps chez 4 espèces de nématodes (résultats obtenus par vos collaborateurs japonais).

Tableau 1

espèces	P. typical	P. typical	P. typical	P. typical	R.regina	R.regina	R.regina	R.regina
sexe	F	F	M	M	F	F	M	M
bactéries	OP50	HB101	OP50	HB101	OP50	HB101	OP50	HB101
volume en nL	6,62	11,93	2,62	2,34	66,17	94,22	14,75	23,04
	5,41	9,82	3,14	4,03	61,61	77,04	20,49	19,28
	5,39	10,08	4,85	4,04	77,23	85,44	18,93	18,83
	7,31	10,14	3,11	4,16	70,45	89,67	10,67	17,15
	7,51	7,65	2,88	3,12	44,84	71,82	22,49	13,88
effectifs	5	5	5	5	5	5	5	5
variance	1,02	2,32	0,78	0,62	148,54	83,70	22,52	11,12
ecart-type	1,25	1,56	1,15	1,2	15	19	4,1	3,2
mediane	6,62	10,08	3,11	4,03	66,17	85,44	18,93	18,83
somme	32,24	49,62	16,6	17,69	320,3	418,19	87,33	92,18
moyenne	6,87	10,12	3,78	3,94	74,6	87	18,1	18,3

Tableau 2

espèces	O. dolichroides	O. dolichroides	O. dolichroides	O. dolichroides	O.sp	O.sp	O.sp	O.sp
sexe	F	F	M	M	F	F	M	M
bactéries	OP50	HB101	OP50	HB101	OP50	HB101	OP50	HB101
volume en nL	3,6	5,82	1,24	1,97	1,56	4,16	0,96	0,93
	2,64	5,25	1,33	1,78	1,36	3,52	0,78	0,96
	2,63	6,39	1,27	1,74	1,37	3,22	0,99	0,91
	4,21	6,31	1,07	1,88	1,66	3,63	0,72	0,82
	4,79	7,19	1,35	2,03	1,7	4,12	1,01	1,13
effectifs	5	5	5	5	5	5	5	5
variance	0,91	0,52	0,01	0,02	0,03	0,40	2,83	2,74
ecart-type	0,68	1,22	0,15	0,23	0,56	0,52	0,11	0,12
mediane	3,6	6,31	1,27	1,88	1,56	3,63	0,96	0,93
somme	17,87	30,96	6,26	9,4	7,65	18,65	4,46	4,75
moyenne	3,43	5,62	1,31	1,84	2,18	3,43	0,92	0,94



Photos 1 à 8

Tous les animaux de ces photographies ont été élevés sur des bactéries *E. coli* OP50.

Etude du rôle de deux gènes dans le contrôle du volume du corps chez le nématode *Caenorhabditis elegans* (résultats obtenus au sein de votre équipe d'accueil).

Tableau 3

génotype	contrôle	contrôle	daf-2	daf-2	dpy-5	dpy-5
sexe	F	F	F	F	F	F
bactéries	OP50	HB101	OP50	HB101	OP50	HB101
volume en nL	4,92	7,22	4,77	5,05	3,71	5,57
	5,51	8,33	4,57	4,38	1,81	8,22
	5,04	9,15	4,69	4,52	2,47	5,48
	5,96	9,22	4,29	5,27	3,27	6,32
	5,31	13,38	4,41	4,99	4,93	6,64
effectifs	5	5	5	5	5	5
variance	0,17007	7,67818667	0,03888	0,14137	1,42892	1,32871
ecart-type	1,04	1,39	0,38	0,42	0,63	1,09
mediane	5,31	9,15	4,57	4,99	3,27	6,32
somme	26,74	47,3	22,73	24,21	16,19	32,23
moyenne	5,81	9,54	4,44	4,92	3,02	6,24

Photographie d'un *C. elegans* adulte femelle contrôle



Photographie d'un *C. elegans* adulte femelle *daf-2*



Photographie d'un *C. elegans* adulte femelle *dpy-5*



Tous les animaux de ces photographies ont été élevés sur des bactéries *E. coli* OP50.

A préparer pour la séance "Analyse des données II" :

- Choix de 2 questions à traiter et construction des figures permettant d'y répondre.
 - Pour répondre aux 2 questions choisies, vous devrez présenter deux figures dont au moins l'une doit être composite, c'est-à-dire qu'elle comprendra plusieurs graphiques, des photos...
 - Vous n'oublierez pas de donner un titre et une légende explicative.
 - Traitement des données : utilisez de préférence le logiciel Excel pour traiter les données issues des travaux présentés.
- Des conseils vous seront donnés à la prochaine séance pour améliorer votre travail.

Questions au choix :

- 1- Chez les nématodes nourris avec la souche OP50, la taille varie-t-elle en fonction de l'espèce et du sexe ?
- 2- Selon les espèces, la souche de bactérie ingérée a-t-elle un impact sur le volume des nématodes mâles ?
- 3- Selon les espèces, la souche de bactérie ingérée a-t-elle un impact sur le volume des nématodes femelles ?
- 4- Quel est l'impact des mutations *daf2* et *dpy5* sur le volume des nématodes femelles nourries avec chacune des 2 souches bactériennes ?

Travail (noté) à rendre

Vous rédigerez un compte-rendu dans lequel vous analyserez les résultats expérimentaux, en ciblant deux questions.

Ce document devra comprendre :

- un titre : un vrai titre, surtout pas : «Analyses de données» ou «CR de méthodologie»
- une brève introduction précisant les objectifs de l'étude et les 2 questions posées
- un matériel et méthodes décrivant l'étude réalisée et les outils utilisés pour analyser les données
- une partie résultats/interprétation permettant de répondre aux questions posées
- une partie conclusion/perspectives.

Pour la partie résultats/interprétation, vous devrez présenter deux figures dont au moins l'une doit être composite c'est à dire qu'elle comprendra plusieurs graphiques, des photos...

Vous n'oublierez pas de donner un titre et une légende explicative à la figure.

Les données de chaque figure seront décrites puis interprétées.

Pour rédiger le document, vous ferez appel à l'ensemble des connaissances acquises jusqu'à présent concernant l'analyse des données.

Quelques conseils concernant la structure d'un compte-rendu (CR) de TP...

TITRE

Il est généralement donné lors de votre TP, sur votre poly ou oralement par votre enseignant.

INTRODUCTION

Présenter le sujet : définir le contexte en apportant les connaissances déjà acquises et clairement exposer la/les question(s) posée(s) dans votre étude. = c'est L'OBJECTIF DU TP.

Expliquer la DEMARCHE que vous allez utiliser pour répondre à l'objectif posé, en précisant succinctement le type d'expérimentation / techniques que vous allez mettre en oeuvre dans la suite du CR.

Un conseil : si possible, il est préférable de rédiger cette partie en dernier car cela permet de voir toutes les notions nécessaires à présenter pour que le lecteur puisse suivre votre étude.

MATERIEL ET METHODES

Présenter le protocole succinctement mais précisément sans recopier le poly de TP (faire préciser à l'enseignant ce qu'il attend de vous à ce sujet), c'est-à-dire :

- présenter le matériel si besoin (les souches de bactéries, les variétés de plantes utilisées...)
- faire apparaître les différentes étapes et méthodes (en précisant le principe des techniques) ainsi que leur but en précisant les concentrations / quantités / volumes, les temps d'incubation / de réaction / de migration... de manière à ce que le lecteur ait toutes les infos nécessaires pour refaire les mêmes manip que vous, dans les mêmes conditions pour vérifier que les résultats obtenus sont reproductibles et donc fiables.

Il est possible dans un CR de TP de formaliser cela sur forme de schéma, tableau...

RESULTATS :

UNE REGLE D'OR : description PUIS interprétation

- Organiser logiquement la **présentation** des différents résultats (selon l'ordre chronologique des manip s'il y a une logique à cet ordre, selon les groupes d'échantillons analysés, selon les problématiques s'il y en a plusieurs...) :
- faire ressortir le plan par des sous-titres....
- organiser vos mesures dans des tableaux, les représenter sous forme graphique (si mesures répétées, faire figurer les écarts-types pour avoir une idée de l'erreur sur la mesure), chercher la représentation la plus « parlante » pour faciliter l'analyse
- organiser vos photos, dessiner et légèder rigoureusement vos observations
- pour chaque figure, inscrire un titre et une légende

Commencer par **décrire** vos résultats précisément

Donner **ensuite vos commentaires, interprétations et conclusions** pour les différentes manip.

CONCLUSIONS (et PERSPECTIVES) :

Faire un court récapitulatif des conclusions obtenues précédemment.

Avez-vous répondu à votre objectif initial ?

Qu'est-ce que cela vous apporte par rapport à ce que vous saviez déjà ?

Essayer d'élargir la discussion. Quelles sont les autres questions que vos résultats soulèvent ?

Vers quelle autre problématique aimeriez-vous continuer ?

Séances : Structure et analyse d'un article scientifique

Objectifs :

- découverte et observation d'un article scientifique.
- analyse approfondie d'un article scientifique en vue de sa restitution sous forme d'exposé oral avec un support visuel (type Présentation).

1- Présentation de la structure d'un article scientifique

Mais d'abord : Qu'est-ce qu'un article scientifique ?
 Pourquoi et comment publie-t-on un article scientifique ?
 Quelles sont les parties qui le composent ?
 Quels sont les objectifs de chacune des parties ?

Les règles qui régissent l'écriture d'un article sont universellement reconnues et se sont imposées d'elles-mêmes aux scientifiques pour communiquer.

Considérez l'article analysé (annexes, doc 6) pour répondre à ce questionnaire.

Titre d'un article

Choisir un bon titre pour un article est un exercice difficile. Les chercheurs réfléchissent toujours longuement avant de se déterminer sur le titre de leurs articles, car le titre est le premier élément que le lecteur perçoit. Autrement dit, c'est le titre qui « attire » le lecteur.

Autrefois, le titre des articles scientifiques était peu informatif. Pour donner un exemple : « *Contribution à l'étude des requins* ». Un tel titre est fort vague. On perçoit bien qu'il s'agit de requins, mais on ne sait pas si l'article se situe au niveau systématique ou physiologique. On n'a non plus aucune information sur ce que l'article apporte comme nouvelle donnée scientifique. De nos jours, on n'utilise plus ce type de formule, mais des titres qui donnent une idée claire sur le contenu. Autrement dit, il faut que le titre délivre le « message » de l'article d'une manière très précise et concise.

Discussion/Conclusion

Écrire et critiquer un article scientifique nécessite de connaître les objectifs et les standards de qualité que chaque partie de l'article doit atteindre. Pour cela il est nécessaire de pratiquer l'écriture et la réflexion scientifique dans le cadre que constitue la structure d'un article scientifique. La première approche pour vous est d'appliquer ces règles dans les compte-rendus de travaux pratiques, dans les rapports de stage, et dans la lecture régulière d'articles scientifiques.

2-Analyse de l'article et création d'un exposé oral avec un support visuel

Utilisez les connaissances acquises au cours de la séance « *Analyse de données expérimentales* » pour comprendre et analyser l'article, initier éventuellement une critique.

Il faut noter

- 1) Que chaque partie est à la fois indépendante et reliée aux autres,
- 2) Que ceci constitue la logique structurale de l'article,
- 3) Que l'« *entrée* » dans l'article peut se faire de façons très différentes selon l'intérêt et la logique du lecteur : lecture du titre et du résumé, lecture des méthodes et des figures avec les légendes, lecture de la conclusion...

A préparer pour "analyse d'un article scientifique II"

Au cours de cette séance, après avoir travaillé en groupe sur l'introduction et les objectifs du travail décrit dans l'article (auteurs et adresses donnés ci-dessous), chaque sous-groupe d'étudiants (3 à 5 étudiants) prendra en charge l'analyse d'une partie des expériences (protocole, résultats obtenus, interprétation).

*Chaque sous-groupe doit préparer, en lien avec la figure étudiée (**travail noté**) :*

- 2 diapositives de type Powerpoint présentant la question posée, la démarche expérimentale utilisée, les résultats et interprétations.

- l'oral de présentation relatif à ces deux diapositives (réalisé par chaque membre du groupe)

- 1 ou 2 phrases de conclusion qui pourront être incluse dans une diapositive de conclusion générale de l'article.

A la fin de cette séance II :

*Le titre de l'article n'est pas donné. Chaque sous-groupe d'étudiants proposera un **titre et 3 mots-clés** à l'issue de la séance.*

Auteurs de l'article analysé et présenté en groupe :

Tamashiro, Wakayama, Akutsu, Yamazaki, Lachey, Wortman, Seeley, D'Alessio, Woods, Yanagimachi and Sakai

Ou en abrégé :

Tamashiro, *et al.*

TITRE

Le titre de l'article n'est pas donné. Chaque étudiant, en guise d'exercice, proposera un titre à l'issue de la séance.

KELLIE L.K. TAMASHIRO^{1,3}, TERUHIKO WAKAYAMA³, HIDENORI AKUTSU³, YUKIKO YAMAZAKI³,
JENNIFER L. LACHEY¹, MATTHEW D. WORTMAN¹, RANDY J. SEELEY¹, DAVID A. D'ALESSIO²,
STEPHEN C. WOODS¹, RYUZO YANAGIMACHI³ & RANDALL R. SAKAI¹

¹Department of Psychiatry and ²Division of Endocrinology,
University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

³Department of Anatomy and Reproductive Biology, Institute for Biogenesis Research, John A. Burns
School of Medicine, University of Hawaii, Honolulu, Hawaii, USA

Correspondence should be addressed to R.R.S.; email: randall.sakai@uc.edu

RÉSUMÉ

Le clonage des mammifères à partir de cellules somatiques a été réussi chez plusieurs espèces, et ses applications potentielles basiques, cliniques et thérapeutiques sont actuellement testées dans beaucoup de directions. Déterminer les effets à long terme du clonage sur la descendance des animaux clonés est crucial si l'on considère les applications futures de la technique. Bien que le développement à terme des animaux clonés à partir de cellules somatiques adultes ait été obtenu, des problèmes apparus chez ces clones montrent que le procédé peut produire des animaux qui ne sont pas identiques phénotypiquement aux organismes donneurs des noyaux. Nous avons utilisé un modèle souris pour tirer profit de la brièveté de son cycle biologique et de sa durée de vie. Nous montrons ici que le poids élevé des femelles B6C3F1 clonées reflète un accroissement du tissu grasseux associé à une plus grande taille corporelle, et que ces souris partagent beaucoup de caractéristiques liées à l'obésité. Nous montrons aussi que le phénotype obèse n'est pas transmis aux descendants issus de croisements entre mâles et de femelles clonés.

INTRODUCTION

Des animaux mâles et femelles ont pu être générés à partir de cellules somatiques dans de nombreuses espèces¹⁻⁸. Plusieurs types de cellules somatiques peuvent conduire au développement complet d'individus⁹⁻¹⁵. Bien que l'attention se porte aujourd'hui principalement sur les applications cliniques de cette méthode, les avancées rapides et l'intérêt grandissant pour les techniques de clonage suscitent des interrogations quant à sa sûreté. En effet, les portées clonées peuvent présenter des problèmes de santé parfois létaux^{2,3,16}. Cependant, entre technique de clonage elle-même et re-programmation des noyaux de cellules somatiques vers un état zygotique, la cause exacte de ces aberrations n'est toujours pas connue. De plus, certains travaux ont montré que les descendants clonés ne sont pas toujours parfaitement identiques à leur donneur¹⁷⁻²⁴. Les applications proposées par cette technologie dépendent de la fiabilité des méthodes de clonage à reproduire des génotypes et des phénotypes spécifiques chez les descendants.

Nous avons précédemment étudié la croissance postnatale et le développement comportemental des souris clonées²⁵. Bien que ne présentant pas d'altérations comportementales majeures, ces souris

étaient substantiellement plus lourdes que les animaux contrôles de même âge (Fig. 1). De plus, l'analyse de cinq générations de clones adultes a montré que tous les descendants présentaient une augmentation équivalente du poids corporel non cumulative d'une génération à l'autre²⁰. A la naissance, aucune différence de poids n'a été mise en évidence entre les portées issues du clonage et les souris contrôles issues de fécondation naturelle et culture d'embryons. Cependant, ces deux groupes de souris présentaient un poids corporel supérieur à celui de leurs géniteurs. Les clones issus de cellules souches embryonnaires sont également plus lourdes à la naissance^{26,27}. Toutefois, les cellules souches embryonnaires exposées subissent probablement des modifications dues à leur exposition prolongée aux conditions de culture qui précèdent le clonage en lui-même. De même, les bovins et brebis clonés présentent un « syndrome du gros nouveau-né » qui se caractérise par une élévation des poids de naissance²⁸, un phénotype qui a récemment été attribué à des modifications épigénétiques survenant au cours de la culture des embryons²⁹. La différence du poids entre les souris clonées et les souris contrôles s'installe entre les 8^e et 10^e semaines post-natales. De même, les souris issues d'embryons cryo-préserverés (congelés) sont plus lourdes que leurs contrôles quand elles atteignent l'âge adulte³⁰. L'ensemble de ces données indique que soit la technique employée soit les

processus biologiques mis en œuvre au cours du clonage induisent une augmentation du poids corporel des portées par rapport à leurs « parents ».

Pour mieux comprendre ce phénomène, nous avons examiné le phénotype de souris clonées provenant de différentes souches (B6C3F1 et B6D2F1) et utilisé des conditions et des groupes contrôles spécifiques afin d'analyser les caractéristiques métaboliques et comportementales de ces souris ainsi que la transmission héréditaire de ce phénotype. Plusieurs indicateurs de l'obésité ont été analysés dont la composition de la carcasse et les concentrations plasmatiques d'insuline et de leptine afin de distinguer une augmentation du poids corporel d'une réelle obésité. Nous montrons que les souris clonées adultes sont non seulement plus lourdes mais qu'elles présentent aussi des altérations métaboliques et comportementales caractéristiques de l'obésité. Nous avons également examiné le phénotype des souris descendant de croisements naturels entre souris B6C3F1 mâles et femelles clonées. Nous avons ainsi observé que le phénotype obèse des souris clonées n'était pas transmis à la descendance.

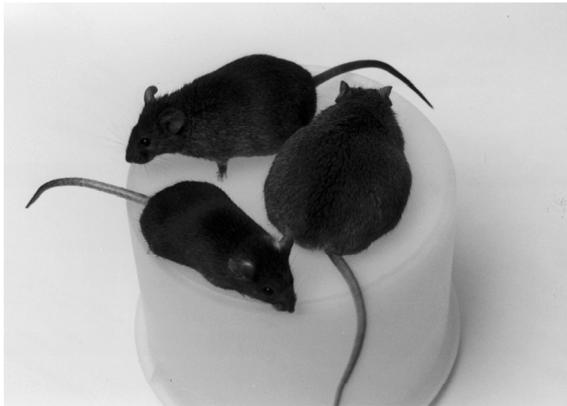


Fig. 1 Cloned and control mice. An adult cloned B6C3F1 mouse (right) and strain- and age-matched representatives of the two control groups of mice used in this study: stock control (foreground) and IVEM mouse (top left).

RÉSULTATS

Poids corporel, prise alimentaire et restriction alimentaire

À la naissance, le poids corporel des souris clonées n'était pas différent de celui des souris contrôles ayant subi des manipulations *in vitro* et des conditions de cultures comparables (« *in vitro* embryo-manipulated », IVEM). Cependant, les souris de ces deux groupes étaient plus lourdes que les souris contrôles « stock » (Fig. 2a). Les courbes de croissance des souris clonées et IVEM étaient identiques jusqu'à environ 8 semaines (Fig. 2b). Puis, les souris clonées devenaient significativement plus lourdes que les souris IVEM. Passé l'âge de 8 semaines, les souris clonées et IVEM conservaient un

poids plus élevé que les souris contrôles « stock » (encart Fig. 2b). Cette différence a persisté tout au long de l'expérience.

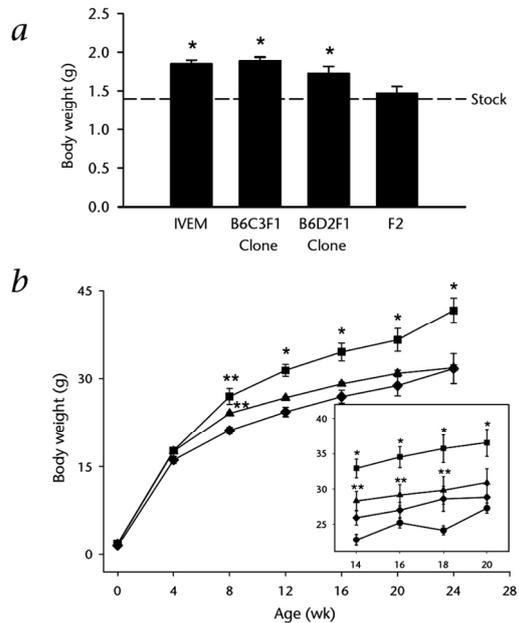


Fig. 2 Mouse body weights. **a**, Birth weights of IVEM, cloned and F2 mice. *, $P < 0.05$, versus stock and F2. **b**, Body weights of clones (■), IVEM (▲) and F2 mice (◆). Growth curves are identical through 8–10 wk of age, at which time clones begin gaining weight more rapidly than IVEM mice. Body weights of F2 offspring over the same time period are significantly less than those of clones. *, $P < 0.05$, versus IVEM and F2; **, $P < 0.05$, versus F2. Inset, comparison of clone (■), IVEM (▲) and F2 (◆) offspring with normally bred and raised age-matched stock mice (●) at 14–20 wk of age. Clones and IVEM mice are significantly heavier than female stock mice, but F2 offspring are of similar weight. *, $P < 0.05$, versus IVEM, F2 and stock; **, $P < 0.05$, versus stock.

Nous avons mesuré l'apport alimentaire *ad libitum* (à volonté) des souris pour déterminer si l'augmentation de masse corporelle des clones est dû à une hyperphagie. Bien que les clones soient plus gros que les souris témoins et IVEM, ils n'étaient pas hyperphagiques. Les clones et les témoins ont mangé à peu près la même quantité de nourriture mais ont mangé plus que les souris IVEM. Cependant, puisque les clones sont significativement plus gros que les souris témoins et IVEM, nous avons également calculé leur apport alimentaire en terme de masse corporelle maigre, puisque les tissus adipeux sont métaboliquement inertes. Les clones mangent significativement moins par gramme de masse corporelle maigre que les témoins, qui mangent la même quantité que les souris IVEM (Fig. 3).

Pour évaluer la réponse des clones à un important stress hypocalorique, nous avons privé les souris de nourriture, mais pas d'eau, pendant 24 heures. Les souris appartenant aux trois groupes ont perdu le même pourcentage de base de masse corporelle durant la période de privation. Quand la nourriture a été réadministrée, toutes les souris appartenant aux trois groupes ont compensé leur perte de masse corporelle en augmentant leur consommation (Fig. 3). Les souris témoins ont mangé plus de nourriture (par rapport à

leur masse corporelle maigre) que les souris appartenant aux deux autres groupes au cours des conditions *ad libitum* et de « réadministration ».

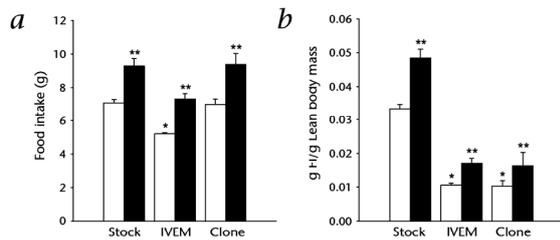


Fig. 3 Food intake. Food intake was measured daily for all groups as adults (1 y of age). Mean 24-h intake is expressed as grams of total intake (a) and gram/gram of lean body mass (b) on a typical day (*ad libitum* feeding; open bars) and after a 24-h food deprivation (filled bars). **a**, Clones eat approximately the same amount of food as stock controls as adults. **b**, After adjustment for lean body mass, cloned mice are shown to eat significantly less per gram of lean body mass than stock controls and the same amount as IVEM mice. After a 24-h food deprivation, all groups have lost the same percentage of body weight. Food intake over the subsequent 24-h period increased significantly for all groups, indicating that clones are capable of 'defending' body weight loss as well as IVEM and stock mice. *, $P < 0.05$, versus stock; **, $P < 0.05$, versus *ad libitum*.

Réponse à l'administration de MTII et de leptine

Nous avons évalué la réponse des clones au MTII, bien connu pour inhiber la prise de nourriture, afin de déterminer si un dysfonctionnement du système de signalisation de la mélanocortine est impliqué dans l'augmentation de masse corporelle des clones. MTII réduit la prise de nourriture au cours des six premières heures du cycle de l'obscurité chez les clones (figure 4a). Chez les souris IVEM et contrôles, une petite diminution non significative s'est produite. Des résultats similaires d'une augmentation de la sensibilité des clones au MTII ont été obtenus en augmentant la période de privation de nourriture 18 heures avant l'administration de MTII (données non présentées). Nous avons également examiné la réponse des clones à un signal anorexigène (leptine), et nous avons constaté que l'administration de leptine a diminué la prise alimentaire plus fortement chez les clones que chez les souris IVEM et témoins (Fig. 4b).

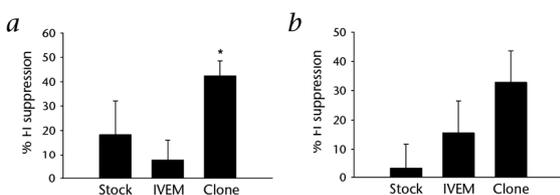


Fig. 4 Effect of MTII and leptin on food intake. **a**, MTII (100 nM), a melanocortin agonist, significantly suppresses food intake (F) of cloned mice (*, $P < 0.05$, $n = 7-9$). A smaller, nonsignificant decrease occurs in IVEM and stock mice. **b**, Leptin (5 $\mu\text{g/g}$ body weight) suppresses food intake more in clones than in IVEM and stock mice. Data represent percent suppression of food intake compared with vehicle over 6 h (MTII) and 4 h (leptin) after administration; $n = 7-9$.

Analyse des hormones du plasma et de la carcasse

Pour déterminer si l'augmentation de la masse corporelle est due à une augmentation du tissu adipeux ou de la masse corporelle maigre, nous avons mesuré la composition du corps des clones en analysant leur carcasse. Les clones ont significativement plus de matière grasse dans leur

carcasse que les souris IVEM ou témoins. Ainsi, les clones sont plus gros et ont aussi plus de tissu adipeux que les souris des deux autres groupes témoins (tableau 1). De plus, nous avons mesuré la concentration de leptine, d'insuline et de corticostéone dans le plasma. Les souris clonées ont plus de leptine et d'insuline dans le plasma. Mais contrairement à certains modèles d'obésité, les concentrations plasmatiques de corticostéone chez les clones ne diffèrent pas de celles des contrôles.

Influence du *background* (« fond » génétique) de la souche sur l'obésité des clones

Pour déterminer l'influence du *background* de la souche sur le phénotype d'obésité des clones, nous avons également généré des clones avec un *background* B6D2F1 [dérivé d'un croisement C57Bl/6 (femelle) et DBA/2 (mâle)]. À huit semaines, le poids de ces clones étaient indissociables de ceux des souris témoins B6D2F1 (témoin : 22 g (réf. 31) à 23 g (réf. 32); clones B6D2F1 : $22,9 \pm 0,2$ g). Toutefois, alors que la masse corporelle de souris témoins B6D2F1, âgées de 18 mois, atteint 30 g³¹, celle des souris clonées, au même âge, est de $42,4 \pm 3,5$ g en moyenne.

Croisement des clones mâles et femelles

Afin de déterminer si le phénotype d'obésité des clones est génétiquement transmis par les lignées germinales, les clones mâles B6C3F1 dérivant d'un noyau cellulaire de fibroblaste ont été accouplés avec les clones femelles B6C3F1 dérivant d'un noyau cellulaire de cumulus. La taille moyenne des portées ($n = 4$) a été de $5,8 \pm 2,4$ souriceaux. Cette taille de portée a peut-être été diminuée par le fait que les souris utilisées pour générer la descendance F2 étaient âgées de 12-14 mois. D'autres recherches concernant la fécondité de clones mâles et femelles ne montrent pas de différence dans la fécondité ou dans la taille des portées²³. Le poids du corps de la descendance (F2) est plus faible que celui des clones (Fig 2b) et le poids du placenta est au niveau de la norme (T.W., observation personnelle). Une comparaison des souris F2 et des souris témoins au même âge est montré dans l'encadré de la Fig. 2b.

DISCUSSION

La production d'animaux clonés à partir de cellules somatiques d'adulte a ouvert de nouveaux champs d'investigation en recherche aussi bien fondamentale que clinique. Les perspectives de l'utilisation du clonage chez l'homme à des fins thérapeutiques ou reproductives ont mis en évidence la nécessité de développer des études complètes sur des animaux clonés. Bien que la production de clones à partir de différents types de cellules somatiques ait été réalisée avec succès, la sécurité et les éventuels

effets biologiques secondaires doivent être étudiés. Par ailleurs, les études de la descendance d'animaux clonés appartenant à d'autres espèces que la souris sont rendues difficiles, à cause d'une part du faible nombre de descendants, et d'autre part des difficultés techniques propres à la technique de clonage. De plus, la détermination des effets biologiques tout au long de la vie d'un animal et sur des générations successives constitue un obstacle dans les études de clonage d'espèces domestiques. Malgré la réussite du clonage de grands ongulés, tels que le mouton, la chèvre, la vache ou le cochon, à notre connaissance, une seule étude a porté sur un grand nombre d'animaux clonés adultes et aucune n'a étudié le taux de graisse ou rapporté des indices d'obésité. Même si des études complètes doivent à la fin être réalisées sur de grands animaux pour déterminer les effets à long terme du clonage à partir de cellules somatiques sur leur santé, la souris reste un modèle privilégié pour étudier les effets du clonage sur la santé aussi bien à court terme qu'à long terme.

Le suivi de la croissance post-natale et du comportement de souris clonées à partir de cellules somatiques d'adulte a déjà été réalisé. Bien qu'il n'y ait pas de modification majeure du comportement, le poids des souris clonées était significativement plus important à partir de l'âge de 8-10 semaines par rapport à des individus témoins. Dans nos travaux, nous avons élargi l'étude de ces souris clonées pour définir de manière complète leur prise de nourriture et leur métabolisme. Une augmentation du poids n'est pas obligatoirement associée à une obésité et, inversement, une absence de changement de poids n'exclut pas la présence d'une obésité sous-jacente. Grâce à notre étude, nous avons pu démontrer l'obésité réelle des souris clonées. Le poids plus élevé de ces souris est directement dû à une augmentation de la proportion de tissu adipeux. De plus, ces souris clonées présentent plusieurs autres caractéristiques communes aux humains obèses telles que l'hyperinsulinémie et l'hyperleptinémie. De façon intéressante, le phénotype d'obésité et la taille augmentée du placenta des souris clonées ne sont pas transmis à leur descendance issue d'une reproduction naturelle.

Les souris clonées de la lignée B6C3F1 ont un poids significativement plus élevé à la naissance par rapport aux souris non clonées de la même lignée mais ne présentent pas de différence de poids par rapport aux souris dites « IVEM ». Cependant, étant donné que les souris témoins sont issues de portées de taille plus importante que celle des souris clonées et des souris IVEM, la différence de poids observée à la naissance pourrait être le reflet de la différence de taille des portées (*Note : lorsqu'il y a un petit nombre d'individus au sein d'une portée, ces individus ont tendance à avoir un poids naturellement plus élevé*

par rapport aux individus issus d'une portée plus grande).

Table 1 Terminal parameters

	Stock (n = 7)	IVEM (n = 7)	Clone (n = 9)
Terminal body weight (g)	28.9 ± 0.7	39.3 ± 1.3*	49.7 ± 3.9**
% fat tissue	14.4 ± 1.4	19.5 ± 1.9	25.2 ± 2.8*
% lean tissue	27.5 ± 0.4	34.7 ± 1.9*	32.9 ± 2.1*
% water	58.1 ± 1.2	45.8 ± 1.4*	41.9 ± 2.8*
Plasma measures:			
Leptin (ng/ml)	2.1 ± 0.5	9.6 ± 1.7*	14.6 ± 1.7**
Insulin (pM)	12 ± 2	26 ± 2*	32 ± 5*
Corticosterone (µg/dl)	8.23 ± 3.20	3.83 ± 0.76	4.54 ± 0.72

Data are expressed as mean ± s.e.m. *, $P < 0.05$, compared with stock controls; **, $P < 0.05$, compared with stock and IVEM mice.

C'est à partir du stade adulte que la différence de poids des clones par rapport aux souris non clonées commence à être significative. Pourtant, contrairement à nos attentes, au stade adulte, les clones ne se nourrissent pas plus que les souris témoins. Etant donné que la prise alimentaire a été mesurée au stade adulte, après que les souris clonées soient devenues obèses, nous ne pouvons pas exclure la possibilité que les souris étaient hyperphagiques avant le début de l'accélération de la prise de poids. En réponse à un traitement hypocalorique, les clones perdent un poids comparable aux témoins et compensent cette perte par une augmentation de prise de nourriture équivalente à celle des souris non clonées. L'activité globale des clones est similaire à celle des souris IVEM à 2, 4 et 6 mois, indiquant ici que la prise de poids n'est pas liée à une activité physique moins importante. Par conséquent, la différence de poids observée pourrait être due à une dépense énergétique basale moins importante chez les souris clonées.

Différents types de souris obèses sont déficientes dans les mécanismes de signalisation dépendants d'une hormone, la mélanocortine. En particulier, la leptine réduit la prise de nourriture et le poids de l'animal en stimulant les neurones qui synthétisent la pro-opiomélanocortine dans le noyau arqué de l'hypothalamus, ces neurones projetant leur axone dans le noyau paraventriculaire (et dans d'autres sites du cerveau) et agissant sur les récepteurs à la mélanocortine MC-4. L'administration de MTII, un agoniste de la mélanocortine, réduit la prise de nourriture chez les clones B6C3F1 mais pas chez les souris témoin ni IVEM. Ce résultat est étonnant compte tenu du fait que les souris clonées sont plus grosses que les souris témoins et qu'elles ont donc reçu une dose moins importante de MTII que les sauvages (quantité de MTII administré / poids de la souris). Les souris clonées sont également plus sensibles à un apport exogène de leptine. Par conséquent, bien que certaines formes d'obésité montrent une sensibilité réduite à la mélanocortine, les souris clonées obèses semblent avoir une sensibilité accrue à cette molécule. Comme les clones

sont hyperleptinémiques, la conséquence serait que la leptine endogène chez les souris clonées supprimerait de manière chronique l'appétit et entraînerait la réduction du poids de ces dernières grâce à la signalisation dépendante de la mélanocortine. Par conséquent, les souris clonées pourraient être encore plus obèses sans l'action catabolique de leur leptine endogène. L'obésité des souris clonées ne peut donc pas être corrélée à un défaut dans la signalisation leptine-mélanocortine, et doit être due à d'autres processus.

Comme il a été signalé dans des articles antérieurs, nous avons trouvé que les générations de souris obtenues par des procédés de culture *in vitro* d'embryons sont plus grosses que celles qui sont obtenues par des croisements naturels. Nos résultats indiquent que la technique de culture *in vitro*, mais aussi la cellule somatique utilisée pour le transfert de noyau, sont par certains aspects impliquées dans le développement de l'obésité des clones. Le noyau de la cellule somatique donneuse doit être reprogrammé correctement pour initier un développement embryonnaire normal. Certains gènes sont alors modifiés et reçoivent une empreinte parentale de nature épigénétique. Bien que les processus de reprogrammation et d'empreinte ne soient pas complètement compris (10-39), on peut penser qu'un manque ou des défauts de ces processus auront pour résultat une mauvaise expression des gènes et donc des anomalies de développement. Certaines de ces anomalies pourront être fatales, entraînant la mort de l'individu, d'autres pourront produire des phénotypes qui seront alors différents de celui du donneur, comme ici l'obésité.

Une étude récente montre que l'empreinte épigénétique d'un gène (*Xsist*) présente dans des cellules somatiques peut être reprogrammée durant le clonage (40). Une baisse de l'expression de gènes soumis ou non à l'empreinte dans le placenta de fœtus clonés, comparée à celle de fœtus normaux, a été démontrée suggérant que le phénotype observé, un agrandissement du placenta, pourrait être indépendant du phénomène d'empreinte génomique (41). En revanche, d'autres études ont montré que des gènes spécifiques de certains tissus sont méthylés de façon différentielle ou pas dans des fœtus de souris clonés à partir de noyaux de cellules du cumulus (24). De plus des différences d'expression de gènes soumis à empreinte épigénétique ont été montrées à l'origine de phénotype dans des fœtus de souris clonés à partir de noyaux de cellules souches embryonnaires (26,27). Il a été de plus suggéré que les technologies qui utilisent la culture d'embryons pouvait produire des modifications épigénétiques affectant le développement (42) ou la croissance après la naissance. Par exemple, un changement épigénétique du gène codant pour le récepteur 2 du facteur de croissance apparenté à l'insuline, est associé à un

phénotype de croissance fœtale accrue (29). Il n'est pas encore clairement démontré si un tel mécanisme peut être directement responsable des faibles taux de réussite du clonage ou des phénotypes anormaux observés parmi les adultes issus d'un clonage.

Pour déterminer si l'obésité des souris adultes clonées est attribuable à des événements génétiques ou épigénétiques, nous avons produit des portées de souris B6C3F2 en croisant des mâles et des femelles B6C3F2 clonés. Les générations F2 issues de ces croisements grossissent moins que les F1, et même moins que les F1 non clonées et cultivées *in vitro*, ceci pendant la période de deux mois considérée. De plus, les F2 présentent un placenta de taille normale contrairement aux F1.

L'obésité observée dans les souris clonées est un phénotype qui n'est pas transmis par la lignée germinale, ce qui indique que les modifications épigénétiques issues de la technique de clonage sont effacées, corrigées durant la gamétogenèse. Les recherches futures devront déterminer quand et où, comment ces modifications sont apportées et quelles méthodes devront être appliquées pour les circonscrire. L'âge des donneurs, le type cellulaire utilisé comme source des noyaux transférés peuvent être la cause des différents phénotypes observés dans les clones. Bien que beaucoup de types cellulaires issus de donneurs d'âges différents permettent de cloner avec succès chez les mammifères, il est nécessaire pour la compréhension des résultats de déterminer exactement les conséquences de l'utilisation de ces différents types cellulaires.

En résumé, nous avons caractérisé un phénotype d'obésité des souris clonées. Cette découverte a des implications directes dans la compréhension et pour l'application de la technique de clonage chez les mammifères. Bien que les souris clonées soient obèses, elles ne transmettent pas ce phénotype à leur descendance au cours de croisements naturels. Ceci indique en retour que ce sont des modifications épigénétiques, plutôt que génétiques, qui pourraient être responsables des faibles taux de succès de clonage et/ou des phénotypes aberrants dans les animaux clonés survivants. Comme le taux de réussite de clonage à partir de cellules d'adulte est bas indépendamment de la technique utilisée, peu d'individus clonés sont obtenus et l'obtention d'une génération avec des traits spécifiques désirés par clonage passe par une reproduction naturelle.

MÉTHODES

Animaux. Toutes les souris ont été maintenues individuellement dans des cages en polycarbonate (18,5 x 29 x 13 cm) avec de la nourriture et de l'eau *ad libitum*, à moins que cela ne soit indiqué autrement. Elles ont été maintenues dans des pièces thermostatées dont l'humidité était contrôlée, avec un

cycle de 14 h de lumière et de 10 h d'obscurité, l'éclairage débutant à 5 h du matin. Les protocoles de maintenance et de traitement des animaux ont été revus et approuvés par les instances de contrôle des Universités de Hawaïi et de Cincinnati.

Les clones de souris femelles ont été produits par microinjection du noyau d'une cellule de cumulus adulte provenant de la lignée B6C3F1 dans un ovocyte énucléé provenant d'une souris B6D2F1, comme décrit précédemment². Neuf femelles clonées ont été sélectionnées de manière aléatoire pour cette étude.

Des souris IVEM ont été produites en tenant compte d'un grand nombre de facteurs inhérents à la technique de clonage : manipulation des embryons, culture *in vitro* d'embryons, transfert d'embryons dans des mères CD-1 pseudo-gestantes, tailles réduites des portées, mise bas par césarienne, et échanges de parents nourriciers. Dix femelles C57Bl/6 et 10 mâles C3H/He ont été obtenus du *National Cancer Institute* et croisés pour produire des souris hybrides B6C3F1. Pour contrôler les manipulations d'embryons et la culture *in vitro*, des embryons au stade 1 cellule ont été collectés à partir d'oviductes de souris le jour après leur accouplement, et cultivés *in vitro* pendant 20 h jusqu'à ce qu'ils atteignent le stade deux cellules. Pour contrôler la taille des portées, et l'utilisation de mères porteuses, 2-3 embryons ont été transférés dans les oviductes de chacune de 20 mères porteuses CD-1 pseudo-gestantes afin de favoriser l'implantation de 1-2 embryons dans chacune de ces femelles. Actuellement, seulement 1 ou 2 embryons clonés atteignent le terme du développement chez les mères porteuses, et ces portées sont délivrées par césarienne, car la parturition n'est pas initiée naturellement en raison du petit nombre de fœtus par portée. C'est pourquoi les souriceaux IVEM (hybrides B6C3F1) sont délivrés par une césarienne au jour 19,5 après conception, et introduits dans des portées de mères CD-1 allaitantes pour leur développement. Nous avons sélectionné de manière aléatoire 7 souris femelles de ce groupe pour notre étude (souris IVEM). Notre second groupe contrôle (stock) comportait 7 femelles B6C3F1 obtenues du *National Cancer Institute*. Toutes les expériences impliquaient des souris de 10-12 mois, à moins que cela ne soit autrement spécifié. Toutes les souris provenaient de la lignée B6C3F1, à moins que cela ne soit aussi stipulé.

Poids corporel, prise de nourriture et réponse à la privation de nourriture. Les poids des souris ont été contrôlés chaque semaine de la naissance à 8 semaines, puis deux fois par semaine. La prise de nourriture basique a été déterminée sur une période de 4 jours de nourrissage *ad libitum*. Toutes les souris ont été nourries par du milieu *Laboratory Rodent Diet 5001* (PMI Feeds, Richmond, Indiana). Pour mesurer

la réponse à une carence nutritive, les souris ont été privées de nourriture, mais non d'eau, pendant 24 h. Ensuite la nourriture a été de nouveau fournie *ad libitum*. La prise de nourriture et le poids corporel ont été mesurés toutes les heures pendant 6 h, et ensuite à 24 h.

Administration de MTII. Les souris ont été privées de nourriture 4 h avant le début de la phase d'obscurité, puis ont reçu une injection intrapéritonéale de 100 nM de MTII (Phoenix Pharmaceuticals, Mountain View, California) dissous dans 100 μ l d'eau désoxygénée stérile, ou 100 μ l d'eau (contrôle) 30 min avant la phase d'obscurité. La nourriture a été rendue accessible au début de la phase d'obscurité, et la prise de nourriture mesurée chaque heure pendant les 6 premières heures, puis après 24 h.

Administration de leptine. La leptine (5 μ g/g de masse corporelle dans une solution saline physiologique; Calbiochem, La Jolla, California) ou la solution contrôle sans leptine ont été administrés par injection intrapéritonéale comme décrit pour l'administration de MTII.

Analyse des carcasses et contrôles plasmatiques. Les souris ont été privées de nourriture pendant 18 h avant d'être tuées par inhalation de CO₂. Le sang collecté a été centrifugé à 4°C. L'insuline plasmatique a été mesurée par radioimmunoessai (RIA) en utilisant un antiserum reconnaissant l'insuline de rongeur⁴³. La leptine plasmatique a été mesurée en utilisant un RIA commercial (Linco, St. Charles, Missouri). La corticostérone plasmatique a été mesurée par RIA au moyen d'un anticorps de lapin réagissant contre la corticostérone (B3-163 : Endocrine Sciences, Tarzana, California). Des procédés standards ont été utilisés pour l'analyse des carcasses. Les carcasses ont été déshydratées dans un lyophilisateur (Labconco Freezone 18) jusqu'à ce que les poids des carcasses diffèrent de moins de 0,1g par rapport au poids mesuré le jour précédent. La masse grasseuse a ensuite été déterminée par extraction à l'éther de pétrole, et la masse aqueuse, la masse grasseuse et la masse maigre (sans eau) ainsi déterminées.

Génération des souris F2. Des souris B6C3F2 ont été générées par des croisements naturels entre des mâles et des femelles B6C3F1 clonés. Les poids corporels ont été contrôlés hebdomadairement jusqu'à l'âge de 8 semaines, et ensuite deux fois par semaine pendant 6 mois.

RÉFÉRENCES

1. Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J. & Campbell, K.H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810-813 (1997).
2. Wakayama, T., Perry, A.C.F., Zuccotti, M., Johnson, K.R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369-374 (1998).

3. Wakayama, T. & Yanagimachi, R. Cloning of male mice from adult tail-tip cells. *Nature Genet.* **22**, 127–128 (1999).
4. Kato, Y. *et al.* Eight calves cloned from somatic cells of a single adult. *Science* **282**, 2095–2098 (1998).
5. Galli, C., Duchi, R., Moor, R. & Lazzari, G. Mammalian leukocytes contain all the genetic information necessary for the development of a new individual. *Cloning* **1**, 161–170 (1999).
6. Wells, D.N., Misica, P.M. & Tervit, H.R. Production of cloned calves following nuclear transfer with cultured adult mural granulosa cells. *Biol. Reprod.* **4**, 996–1005 (1999).
7. Kubota, C. *et al.* Six cloned calves produced from adult fibroblast cells after long-term culture. *Proc. Natl. Acad. Sci. USA* **97**, 990–995 (2000).
8. Polejaeva, I.A. *et al.* Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature* **407**, 86–90 (2000).
9. Coleman, A. Somatic cell nuclear transfer in mammals: progress and applications. *Cloning* **1**, 185–200 (1999/2000).
10. Prather, R.S., Kuholzer, B., Lai, L. & Park, K.-W. Changes in the structure of nuclei after transfer to oocytes. *Cloning*, **2**, 117–122 (2000).
11. Solter, D. Mammalian cloning: advances and limitations. *Nature Rev. Genet.*, **1**, 199–207 (2000).
12. Tsunoda, Y. & Kato, Y. The recent progress on nuclear transfer in mammals. *Zoolog. Sci.* **17**, 1177–1184 (2000).
13. Wakayama, I. & Yanagimachi, R. Mouse cloning with nucleus donor cells of different age and type. *Mol. Reprod. Dev.* **58**, 376–383 (2001).
14. Yamazaki, Y. *et al.* Assessment of the developmental totipotency of neural cells in the cerebral cortex of mouse embryo by nuclear transfer. *Proc. Natl. Acad. Sci. USA* **98**, 14022–14026 (2001).
15. Yanagimachi, R. Cloning: experience from the mouse and other animals. *Mol. Cell. Endocrinol.* (in the press).
16. Renard, J.P. *et al.* Lymphoid hypoplasia and somatic cloning. *Lancet* **353**, 1489–1491 (1999).
17. Ogura, A. *et al.* Production of male cloned mice from fresh, cultured, and cryopreserved immature sertoli cells. *Biol. Reprod.* **62**, 1579–1584 (2000).
18. Shiels, P.G. *et al.* Analysis of telomeres lengths in cloned sheep. *Nature* **399**, 316–317 (1999).
19. Evans, M.J. *et al.* Mitochondrial DNA genotypes in nuclear transfer-derived cloned sheep. *Nature Genet.* **23**, 90 (1999).
20. Wakayama, T. *et al.* Cloning of mice to six generations. *Nature* **407**, 318–319 (2000).
21. Lanza, R.P. *et al.* Extension of cell life-span and telomere length in animals cloned from senescent somatic cells. *Science* **288**, 665–669 (2000).
22. Steinborn, R. *et al.* Mitochondrial DNA heteroplasmy in cloned cattle produced by fetal and adult cell cloning. *Nature Genet.* **25**, 255–257 (2000).
23. Tian, X.C., Xu, J. & Yang, X. Normal telomere lengths found in cloned cattle. *Nature Genet.* **26**, 272–273 (2000).
24. Ohgane, J. *et al.* DNA methylation variation in cloned mice. *Genesis* **30**, 45–50 (2001).
25. Tamashiro, K.L.K., Wakayama, T., Blanchard, R.J., Blanchard, D.C. & Yanagimachi, R. Postnatal growth and behavioral development of mice cloned from adult cumulus cells. *Biol. Reprod.* **63**, 328 (2000).
26. Humpherys, D. *et al.* Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95–97 (2001).
27. Eggan, K. *et al.* Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl. Acad. Sci. USA* **98**, 6209–6214 (2001).
28. Young, L.E., Sinclair, K.D. & Wilmot, I. Large offspring syndrome in cattle and sheep. *Rev. Reprod.* **3**, 155–163 (1998).
29. Young, L.E. *et al.* Epigenetic change in *IGF2R* is associated with fetal overgrowth after sheep embryo culture. *Nature Genet.* **27**, 153 (2001).
30. Dulioust, E. *et al.* Long-term effects of embryo freezing in mice. *Proc. Natl. Acad. Sci. USA* **92**, 589–593 (1995).
31. Turturro, A. *et al.* Growth curves and survival characteristics of the animals used in the Biomarkers of Aging Program. *J. Gerontol.* **54**, B492–501 (1999).
32. The Jackson Laboratory. *JAX Notes #468 442–444* (The Jackson Laboratory, Marketing Communications, Bar Harbor, Maine, 1997).
33. Baguisi, A. *et al.* Production of goats by somatic cell nuclear transfer. *Nature Biotechnol.* **17**, 456–461 (1999).
34. Lanza, R.P. *et al.* Cloned cattle can be healthy and normal. *Nature* **294**, 1893–1894 (2001).
35. Seeley, R.J. *et al.* Melanocortin receptors in leptin effects. *Nature* **390**, 349 (1997).
36. Woods, S.C., Seeley, R.J., Porte, D. & Schwartz, M.W. Signals that regulate food intake and energy homeostasis. *Science* **280**, 1378 (1998).
37. Schwartz, M.W., Woods, S.C., Porte, D., Seeley, R.J. & Baskin, D.G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
38. Walker, S.K., Hartwich, K.M. & Seamark, R.F. The production of unusually large offspring following embryo manipulation: concepts and challenges. *Theriogenology* **45**, 111–120 (1996).
39. Kikyo, N. & Wolffe, A.P. Reprogramming nuclei: insights from cloning, nuclear transfer and heterokaryons. *J. Cell Sci.* **113**, 11–20 (2000).
40. Eggan, K. *et al.* X-Chromosome inactivation in cloned mouse embryos. *Science* **290**, 1578–1581 (2000).
41. Inoue, K. *et al.* Faithful expression of imprinted genes in cloned mice. *Science* **295**, 297 (2002).
42. Young, L.E. & Fairburn, H.R. Improving the safety of embryo technologies: Possible role of genomic imprinting. *Theriogenology* **53**, 627–648 (2000).
43. Einsnick, J.W., Laschansky, E.C., Vogel, R.E. & D'Alessio, D.A. Effect of somatostatin-28 on dynamics of insulin secretion in perfused rat pancreas. *Diabetes* **40**, 1163–1169 (1991).

Séance : Présentation orale d'un article scientifique

La qualité d'une présentation orale passe par la démonstration d'un esprit de synthèse, de précision, de rigueur et d'esprit critique. La pédagogie avec laquelle les données sont présentées est aussi déterminante.

Objectif

L'objectif de la présentation est que chacun des étudiants du groupe comprenne l'essentiel de l'article. Pour atteindre cet objectif, il n'est pas nécessaire de rentrer dans tous les détails de l'article.

Présentation

Il faut tenir compte des considérations suivantes pour votre présentation

Quel est le sujet de l'article ?

Pourquoi s'intéresser à ce sujet ?

Quel est l'objectif de cet article ?

Quel est le message de cet article ?

Pour le contenu de l'article

Décrire le contenu de façon synthétique.

Expliquer dans quel contexte scientifique les expériences décrites ont été effectuées.

Quelle démarche a-t-elle été adoptée ? Pourquoi ? Comment a-t-elle été mise en œuvre ?

Quels sont les résultats obtenus ?

Quelles sont les principales conclusions ?

Critique de l'article

La démarche et les résultats vous semblent-ils rigoureux ?

Manque-t-il des informations ?

Des expériences manquent-elles ?

Perspectives

Un bon article scientifique ouvre toujours des perspectives.

Estimez-vous que tel est le cas pour l'article analysé ?

Cet article incite-t-il à entreprendre de nouvelles recherches ?

Quelles sont les perspectives données par les auteurs ?

Plus difficile ! Quels travaux imagineriez-vous à la suite de la lecture de l'article ?

Instructions et conseils

Le support de votre présentation se fera sur *Powerpoint*, mais vous pourrez aussi utiliser le tableau en supplément si besoin (annexes, doc 9 et 10)...

Utilisez votre connaissance de la structure d'un article scientifique. Si vous connaissez les objectifs attendus de chaque partie, vous savez où trouver les réponses aux questions ci-dessus ! Mais vous savez aussi reconnaître si une des parties n'atteint pas les objectifs fixés, ce qui est un bon début d'esprit critique...

Réussir une présentation avec powerpoint

Les participants

- Entendent plus rapidement qu'ils ne lisent
 - Voient plus vite qu'ils n'entendent
 - Saissent mieux une image simplifiée qu'un texte écrit
 - Lisent de gauche à droite, de haut en bas et dans le sens des aiguilles d'une montre (dans nos cultures)

Compte tenu de ceci, voici quelques recommandations pour préparer et utiliser une présentation powerpoint.

Concevoir un support visuel	Powerpoint est un support visuel qui ne remplace pas la communication. Un diaporama doit être conçu visuellement, un peu comme un tableau. Développez un seul message par diapositive que l'on retrouve dans le titre. Chaque diapositive doit être auto-suffisante. Gardez des règles de composition stables (taille des titres, fond de page) mais variez la présentation (type de schéma, organisation des mots-clés). Pour que les visuels aient un effet positif sur l'attention, il faut éviter de rajouter des distracteurs : sons et images sans utilité réelle, fonds de page trop complexe, répétition de logos ou titres, animations tout azimut, etc.
Bien occuper l'espace	Utilisez au plus 50% de l'espace. Le « blanc » est important pour la compréhension et l'attention. Mettez l'information principale au centre ou aux intersections des lignes pour tenir compte des habitudes d'exploration visuelle.
Limiter le texte	Limitez les textes à l'essentiel. Ils ne remplacent pas le discours. La personne qui lit de longs textes perd son auditoire. Ecrivez des citations quand elles sont importantes et qu'elles supportent le propos. Les mots sont un support à l'image (mots-clés, repères, etc.) Contentez-vous au maximum de 10 lignes pour faciliter la lecture et la rétention de l'information.
Faciliter la lecture	Évitez d'utiliser plus de 2 polices différentes : une pour les titres, l'autre pour les textes. Choisissez des polices sans serif (ou sans empattement) car elles sont plus lisibles. Réservez les italiques pour renforcer un mot et évitez de souligner ou ombrer les mots qui deviennent alors moins lisibles à distance. La taille des caractères doit augmenter avec la grandeur de la salle : il faut éviter de descendre en dessous de 24 points. Utilisez les tailles de caractère pour hiérarchiser l'information (en plus des éléments visuels) et les caractères gras pour améliorer la visibilité. Pour les schémas et flèches, pensez à épaissir les traits et contours (au moins 3 pt).
Mettre en valeur les graphiques	Le titre du graphique doit dire aux participants ce qu'ils doivent chercher : une évolution, une tendance, une relation entre 2 variables, etc... Choisissez le type de graphique en fonction de l'idée à faire passer : par exemple, camemberts pour des proportions ou courbes pour une évolution dans le temps. Éliminez toutes les informations qui ne sont pas utiles à la compréhension : légendes, grilles de fond, etc. Agrandissez les informations principales et renforcez l'épaisseur des traits. Éventuellement, vous pouvez attirer l'attention sur une zone importante avec une flèche.
Exploiter la couleur	Les couleurs augmentent la motivation à lire, aident à structurer l'information et peuvent améliorer la mémorisation. Utilisez les couleurs pour souligner, créer des catégories ou une atmosphère. Dans la mesure du possible, limitez vous à 3 - 4 couleurs maximum le noir compris. Créez un code couleur qui associe toujours le même type d'information aux mêmes couleurs. Utilisez de préférence des couleurs froides pour le fond et des couleurs chaudes pour le texte. Les couleurs de texte doivent être bien contrastées avec la couleur du fond : soit un fond pâle avec des caractères foncés, soit un fond foncé avec des caractères clairs. La relation entre les couleurs est plus importante que les couleurs elles-mêmes. Le contraste fond bleu avec texte jaune est très ergonomique, mais beaucoup vu... Ne négligez pas la valeur symbolique ou émotionnelle des couleurs. Elles deviennent plus sombres quand projetées dans une grande salle. Pensez aussi au 10% de daltoniens....
Doser les animations	Les animations permettent de passer du simple au complexe, de doser l'information, d'illustrer des évolutions dans le temps ou de simuler des processus ou phénomènes. Elles peuvent aussi être utilisées pour mettre l'accent sur une partie d'image ou de schéma. Cependant, elles doivent être utilisées avec une intention pédagogique. Le trop d'animation agit comme un distracteur et détourne l'attention de l'auditoire. Limitez à 2 les effets de transition et reprenez les effets les plus simples (apparaître, balayage de gauche à droite).
Présenter avec powerpoint	Pendant l'exposé, laissez du temps aux participants pour « rentrer dans l'image », 1 à 2 secondes sans parler. Dites bien ce que vous voulez montrer avec le support et si nécessaire donnez des repères de compréhension : légendes, origines, etc. Rappelez-vous que la durée de vie d'une diapositive est de 2 minutes. N'embarquez pas dans une présentation de 15 minutes avec plus de 10 diapositives. Pour les exposés plus longs, prévoyez des interruptions ou pauses sans powerpoint toutes les 15 minutes.

Et pour conclure

Résistez à la tentation d'utiliser toutes les possibilités du logiciel. Restez bien centrés sur le contenu et les objectifs de l'exposé.
Montrez moins de diapositives et des diapositives moins chargées que vous ne le feriez spontanément.
Essayez le tout avant votre intervention et prévoyez un plan B sous forme de transparents.

Préparer et donner une communication orale

1. PRÉPARER ET PLANIFIER L'EXPOSÉ

Choisir le titre	<ul style="list-style-type: none"> Choisissez un titre concis, fidèle et représentatif du sujet que vous allez traiter. Préférez un titre accrocheur qui réveille la curiosité. Pour le rendre attractif, utilisez la forme interrogative ou une formule provocante.
Délimiter les contenus	<ul style="list-style-type: none"> Commencez par faire l'inventaire de la matière à disposition pour traiter votre sujet. Sélectionnez ensuite les contenus en fonction du public, du temps à disposition et des objectifs de votre exposé. Ne cherchez pas à être exhaustif par rapport à votre sujet. Préférez la mise en valeur d'un point de vue personnel, de résultats pertinents, de convictions fortes ou de contributions originales.
Formuler les objectifs	<ul style="list-style-type: none"> Les objectifs résument l'idée force ou le message essentiel que vous souhaitez transmettre à votre public. Que doit-il retenir à la fin de votre exposé ? Les objectifs représentent la colonne vertébrale de la communication et aident ensuite à construire l'exposé. Dans un congrès, les objectifs se rapportent à la contribution scientifique ou rendent compte d'une pratique.
Organiser l'exposé	
Introduire le sujet	<ul style="list-style-type: none"> L'introduction doit être brève (1 à 2 minutes). Elle est censée délimiter le sujet et les contours de votre intervention. Commencez par saluer votre public et établir le contact. L'introduction doit ensuite indiquer les objectifs de l'exposé ainsi que le plan de présentation. C'est aussi l'occasion d'explicitier les idées forces qui vont être développées au cours de l'exposé. Motivé aussi votre auditoire en annonçant les perspectives et les bénéfices apportées par votre intervention.
Construire le sujet	<ul style="list-style-type: none"> Organisez une chronologie et une structure entre les notions abordées. L'enchaînement la plus logique consiste à procéder du simple au complexe, des généralités vers les détails pointus. Faites ressortir les notions essentielles ou les principaux résultats. Mettez bien l'accent sur l'apport personnel et méfiez-vous des contextualisations trop longues qui font ombre au sujet central. Évitez aussi l'exposé encyclopédique. Pensez « structure narrative » pour rendre le développement plus attractif.
Conclure	<ul style="list-style-type: none"> Pour terminer l'exposé, faites une synthèse des idées forces ou de vos arguments clés. Indiquez les éléments de réponses que vous pouvez apporter à vos questions de recherche ou hypothèses. Mentionnez les nouvelles pistes de réflexion ou les retombées. N'oubliez pas de remercier le public de son attention.

Le manque de temps est l'ennemi classique du conférencier. Répétez l'exercice avant sachant qu'en réalité cela demandera toujours plus de temps. Surveillez aussi le temps qui passe pendant votre intervention.

2. S'EXPRIMER EN PUBLIC

Notre vie nous prépare en général beaucoup mieux à participer à une discussion qu'à donner un exposé. Il faut donc apprendre et apprivoiser les habiletés de communication. Soyez attentif au fait que plus de 50% de votre message passe par votre attitude envers les participants.

La voix	<ul style="list-style-type: none"> Ne lisez jamais de texte, mais parlez aux gens. Parlez fort, pas trop vite, avec expressivité (intonation et rythme). Libérez votre énergie. Ne redoutez pas quelques secondes de silence. Évitez la répétition de mots inutiles Heu ! Alors ! Bien ! N'est-ce pas ! etc....
Le regard	<ul style="list-style-type: none"> Établissez un contact visuel avec les participants. Les regardez, si possible chacun, en alternant aléatoirement. Évitez de regarder le plafond, le sol, vos documents, l'écran de votre ordinateur, ou toujours les mêmes personnes.
Le corps	<ul style="list-style-type: none"> Adoptez une position stable, solide et calme. Restez toujours bien face au public. Évitez les balades et les bougettes inutiles ou de tourner le dos au public en voulant montrer des informations projetées au tableau.
Les gestes	<ul style="list-style-type: none"> Montrez ce que vous dites en utilisant un pointeur, un index ou une baguette. Appuyez vos propos par des gestes et mouvements appropriés. Évitez autant que possible les gestes répétitifs comme de manipuler sans cesse un stylo, de remuer les clés dans la poche, ou encore de remettre en place sans arrêt une mèche de cheveu rebelle. Les gestes d'auto-contact ne sont pas recommandés.

Petit plus à la thématique

Le poster scientifique : un autre moyen de communication

Le mode de communication est utilisé en colloque et sert de support pour engager la discussion avec d'autres scientifiques.

Conception du poster

Quand on conçoit un poster, il faut d'abord savoir comment ce poster sera évalué. Cette évaluation repose sur les réponses aux questions suivantes :

- Le poster est-il facile à lire, plaisant à voir ?
- Attire-t-il l'attention ?
- Comment l'information est-elle présentée : trop, pas assez, compacte, dispersée ?
- La lecture du poster fait-elle adhérer à la démonstration ?

Présentation du poster

- Le poster doit être lisible en 5 minutes --> bien choisir ce qui doit être prioritairement montré.
- Il doit contenir un minimum de texte (800 mots, un langage adapté à l'audience).
- Il doit être informatif mais jamais exhaustif (concision et efficacité).
- Il doit être attractif pour l'œil et facilement interprétable et lisible à distance (une structure claire).

Contenu du poster

- Le titre doit être court, spécifique, attractif.
- Introduction : doit contenir le minimum d'information, doit replacer dans le contexte, être juste suffisante pour éveiller l'intérêt.
- Matériel et méthodes (ou l'approche expérimentale) : éviter les détails.
- Résultats : des images, des figures et des tableaux avec un minimum d'interprétation, en veillant à la cohésion texte-figure.
- Conclusion : un bref rappel de l'hypothèse, des résultats avant la conclusion finale.
- Ne pas oublier les références !

Construction du poster

La structure doit être MODULAIRE, et impliquer :

- Un nombre limité de modules.
- Des modules simples et construits efficacement (une figure ou des phrases de texte).
- Des modules disposés sur un fond discret et de bon goût.
- Un agencement logique et clair des modules, en fonction de leur importance.

La « fabrication » du poster peut être artisanale (rare en sciences) ou réalisée sur ordinateur.

Utilisation des statistiques pour les sciences de la vie

Gaëlle LELANDAIS

gaelle.lelandais@univsite-paris-saclay.fr



Ce document est mis à disposition selon les termes de la licence Creative Commons Attribution – Partage dans les mêmes conditions 4.0 International (CC BY-SA 4.0).

Table des matières

PARTIE 1 : POURQUOI UTILISER LES STATISTIQUES POUR LES SCIENCES DE LA VIE ?	6
I. Présentation des cas d'études.....	6
1) Contexte de travail dans un laboratoire pharmaceutique : « le médicament est-il efficace ? ».....	6
2) Contexte de travail dans un casino : « les jetons sont-ils équilibrés ? ».....	6
3) Premiers résultats.....	7
4) Bilan.....	7
II. Problématique statistique.....	8
III. Rappels de définitions.....	8
1) Population et unités statistiques.....	8
2) Variables.....	9
IV. Expérience et variable aléatoire.....	10
1) Définitions.....	10
2) Convention de notations.....	11
V. Notion d'échantillonnage.....	11
PARTIE 2 : CALCULS DE PROBABILITES EN STATISTIQUES	12
I. Pourquoi des probabilités en statistiques ?.....	12
II. Variables aléatoires quantitatives discrètes.....	12
1) Loi de probabilité.....	12
2) Fonction de répartition.....	13
3) Espérance et variance.....	14
III. Variables aléatoires quantitatives continues.....	16
1) Loi(s) normale(s).....	16
2) Loi normale centrée – réduite.....	19
3) Calculs de probabilités avec la loi normale.....	20
IV. Théorème central limite.....	21
PARTIE 3 : ÉCHANTILLONNAGE ET ESTIMATION	23
I. Définition statistique de l'échantillon.....	23
1) Étude d'une variable aléatoire quantitative.....	23
2) Moyenne d'un échantillon.....	23
3) Représentations histogrammes de la variable aléatoire M	24
4) Fonction de densité de probabilité associée à la moyenne d'un échantillon.....	25
II. Estimation statistique.....	26
1) Principe.....	26
2) Qualité d'un estimateur.....	26
3) Estimation ponctuelle.....	27
4) Estimation par intervalle de confiance.....	28
III. Cas des petits échantillons ($n \leq 30$).....	28
IV. Cas des variables aléatoires qualitatives (2 classes).....	29
	3

1) Rappel du second cas d'étude.....	29
2) Fonction de densité de probabilité de la proportion d'un échantillon.....	29
3) Estimation ponctuelle.....	29
4) Intervalle de confiance.....	30
V. Exemples d'applications.....	30
PARTIE 4 : LES TESTS D'HYPOTHESES	31
I. Introduction aux tests d'hypothèses.....	31
1) Rappel des cas d'étude.....	31
2) Illustration de l'approche décisionnelle des tests d'hypothèses.....	32
II. Définition d'une règle de décision.....	33
1) Objectif.....	33
2) Evaluation d'une règle de décision.....	33
III. Théorie des tests statistiques.....	35
1) Hypothèse nulle et hypothèse alternative.....	35
2) Définitions des risques d'erreurs.....	35
3) Notion de "statistique du test".....	36
4) Application à la comparaison d'une moyenne et d'une norme (1^{er} cas d'étude).....	37
5) Deroulement d'un test d'hypothèses.....	37
IV. Comparaison de deux moyennes.....	39
V. Cas des petits échantillons.....	39
1) Distribution de la variable aléatoire étudiée.....	39
2) Égalité des variances.....	39
3) Statistique du test.....	40
4) Utilisation de la loi de Student.....	40
PARTIE 5 : ÉTUDE DES VA QUALITATIVES MULTI-CLASSES	42
I. Exemple de cas d'étude.....	42
II. Test d'hypothèses du χ^2 (Chi2).....	43
1) Principe.....	43
2) Hypothèses.....	43
3) Statistique du test.....	43
4) Règle de décision.....	44
III. Pour aller plus loin.....	45
PARTIE 6 : ÉVALUER LES PERFORMANCES D'UNE ANALYSE STATISTIQUE	46
I. Précision d'un intervalle de confiance.....	46
II. Degré de signification d'un test statistique.....	46
III. Notion de puissance d'un test statistique.....	47
IV. Simulations appliquées au cas d'étude 2 (logiciel R).....	48
1) Combien de fois faut-il lancer un jeton ?.....	48
2) Combien de jetons faut-il lancer ?.....	48
EXERCICES D'APPLICATIONS	50
I. Partie 1.....	50
1) Exercice 1.....	50
	4

2) Exercice 2	50
3) Exercice 3	51
4) Exercice 4	51
II. Partie 2	52
1) Exercice 1	52
2) Exercice 2	52
3) Exercice 3	53
III. Partie 3	54
1) Exercice 1	54
2) Exercice 2	55
IV. Partie 4	56
1) Exercice 1	56
V. Partie 5	58
1) Exercice 1	58
2) Exercice 2	59
REFERENCES BIBLIOGRAPHIQUES	60
DOCUMENTS ANNEXES	61

Partie 1 : Pourquoi utiliser les statistiques pour les sciences de la vie ?

I. Présentation des cas d'études

- 1) Contexte de travail dans un laboratoire pharmaceutique : « le médicament est-il efficace ? »

Une entreprise pharmaceutique développe un médicament destiné à traiter l'hypertension artérielle. Le protocole expérimental consiste à mesurer la tension artérielle systolique (en cm Hg) deux fois chez des patient(e)s : une première fois avant le traitement (valeur de référence) et une deuxième fois après 6 semaines de prise du médicament.

- o L'objectif est de mettre en place un protocole d'étude statistique destiné à évaluer l'efficacité du nouveau médicament.

- 2) Contexte de travail dans un casino¹ : « les jetons sont-ils équilibrés ? »

Un casino propose le jeu « pile/face » à ses client(e)s. Des centaines de jetons sont livrés au casino et la personne qui dirige l'établissement souhaite vérifier qu'ils sont équilibrés, soit : $Pr(\text{pile}) = Pr(\text{face})^2$. Le casino possède une machine qui lance des jetons et détecte automatiquement le côté gagnant.

- o L'objectif est de mettre en place un protocole d'étude statistique destiné à contrôler les nouveaux jetons. En particulier, il faut déterminer 1) combien de fois lancer un jeton et 2) combien de jeton lancer.

¹ Ce deuxième contexte de travail n'est pas en relation direct avec les sciences de la vie. Toutefois, il est utile pour discuter le cas particulier de l'étude des variables qualitatives à deux classes, rencontrées fréquemment en biologie.

² La probabilité d'obtenir le côté « Pile » est égale à la probabilité d'obtenir le côté « Face », et égale à $\frac{1}{2}$.

3) Premiers résultats

Une personne malade prend le médicament développé par l'entreprise pendant 6 semaines. Sa tension artérielle¹ est de 15 cm Hg avant le traitement et 12 cm Hg après le traitement (soit une diminution de 3 cm Hg).

➤ **Quelle conclusion est-il possible d'obtenir à partir de ce résultat ? Discuter la notion de REPRODUCTIBILITE en biologie.**

9 personnes supplémentaires prennent le médicament pendant 6 semaines. Pour chacune, les valeurs de la tension artérielle sont mesurées avant et après le traitement :

Personne	1	2	3	4	5	6	7	8	9	10
Avant	15	15	16	18	18	18	18	15	19	18
Après	12	13	16	16	14	15	17	14	16	16
Écart	3	2	0	2	4	3	1	1	3	2

Tableau 1 : Mesures de la tension artérielle (en cm Hg) de 10 personnes malades, avant et après la prise du médicament. Les écarts correspondent à la différence : tension artérielle « avant » moins tension artérielle « après » le traitement. La valeur moyenne des écarts est de 2,1 et la variance 1,29 (soit un écart type de 1,13). Des valeurs d'écarts positives sont attendues, si le traitement est efficace.

➤ **Quelle conclusion est-il possible d'obtenir à partir de ces résultats ?**

4) Bilan

- **Ce qui est certain** : des différences sont observées entre les mesures de la tension artérielle après et avant traitement.
- **Ce qui est hypothétique** : l'origine de ces différences (efficacité du médicament ou bien hasard d'échantillonnage ?).

¹ Notez que les valeurs données tout au long de ce polycopié ont été choisies pour illustrer des concepts statistiques. Leur pertinence d'un point de vue médical peut-être contestable ©.

II. Problématique statistique

A partir d'un ensemble d'observations, l'objectif est de déterminer :

- la part (obligatoire) du hasard,
- la part (éventuelle) de l'effet testé (efficacité du médicament par exemple).

○ Les statistiques permettent de calculer la probabilité d'observer uniquement par le hasard des écarts au moins aussi importants que ceux observés lors de l'étude. Ainsi : **Statistiques → Calculer la probabilité du hasard.**

Le raisonnement est alors le suivant :

- Si la probabilité d'observer uniquement par le hasard les données du Tableau 1 est grande, l'effet du médicament a peu de chances d'exister. En revanche,
- Si la probabilité d'observer uniquement par le hasard les données du Tableau 1 est petite, l'effet du médicament a de grandes chances d'exister¹.

➤ **Quels paramètres influencent le résultat d'une étude statistique ?**

III. Rappels de définitions

- 1) Population et unités statistiques

La population est l'ensemble des éléments qui forment le champ d'analyse d'une étude particulière. Elle est constituée d'un ensemble d'éléments appelés unités statistiques² (ou individus).

➤ **Quelles sont les populations étudiées dans les cas d'études ?**

¹ Cette notion est difficile à appréhender. Elle devra être reprise à l'issue du CHAPITRE « Tests d'hypothèses » (page 31).

² Le concept de population est général en statistiques. Il s'applique aux choses, aux événements (pas seulement aux êtres humains).

2) Variables

Les unités statistiques d'une population varient selon des caractéristiques appelées des variables. Les variables peuvent être quantitatives (discrètes/continus) ou qualitatives (2 classes ou plus).

➤ Quelles sont les variables étudiées dans les cas d'études ? Sont-elles quantitatives ou qualitatives ?

(a) Présentation des mesures d'une variable quantitative

• **Moyenne**

La moyenne d'une variable quantitative X est :

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$

Équation 1 : Formule de la moyenne d'une variable quantitative.

• **Variance**

La variance d'une variable quantitative X est :

$$Var = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left[\frac{\sum_{i=1}^n x_i}{n} \right]^2$$

Équation 2 : Formule de la variance d'une variable quantitative.

➤ Calculer les valeurs de la moyenne et de la variance associées aux données du Tableau 1.

➤ Que signifie une valeur faible de variance en termes de reproductibilité des mesures ?

• **Représentation histogramme**

Un ensemble de mesures d'une variable peut être représenté sous la forme d'un histogramme, avec en abscisse des intervalles de valeurs de la variable étudiée et en ordonnée les nombres de mesures (effectifs) observés dans ces intervalles.

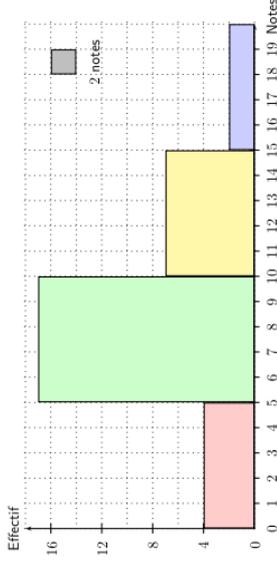


Figure 1 : Représentation histogramme du variable quantitative (« les notes des étudiants à l'examen »). Cette image est issue de http://mathematiques.daval.free.fr/LMG/pdf/Statprobas_2_series_continues.pdf.

IV. Expérience et variable aléatoire

1) Définitions

Une expérience aléatoire est une expérience qui, répétée dans des conditions apparemment identiques, peut donner des résultats différents (nommés événements).

➤ Quelles sont les expériences aléatoires réalisées dans les cas d'études ?

Une variable aléatoire est réalisée lorsqu'elle prend des valeurs en fonction du résultat d'une expérience aléatoire.

➤ Quelles sont les variables aléatoires¹ associées aux expériences aléatoires des cas d'études ?

¹ Décrire une variable aléatoire signifie donner d'une part son intitulé et d'autre part son type.

2) Convention de notations

Une variable aléatoire est notée en MAJUSCULES, tandis que les valeurs prises lors de la réalisation de cette variable aléatoire (c'est à dire les observations) sont notées en minuscules. Ainsi, les données du

Personne	1	2	3	4	5	6	7	8	9	10
Avant	15	15	16	18	18	18	18	15	19	18
Après	12	13	16	16	14	15	17	14	16	16
Écart	3	2	0	2	4	3	1	1	3	2

Tableau 1 (page 7) sont notées :

- X : variable aléatoire « écart en les mesures de la tension artérielle avant et après le traitement »,
- Et $\{x_1 = 3, x_2 = 2, x_3 = 0, x_4 = 2, x_5 = 4, x_6 = 3, x_7 = 1, x_8 = 1, x_9 = 3, x_{10} = 2\}$ les mesures de X dans l'échantillon (ici de taille 10).

V. Notion d'échantillonnage

Une étude statistique a pour objectif d'obtenir des connaissances sur l'ensemble de la population (le médicament sera-t-il efficace sur toutes les personnes malades ?). Une analyse complète de la population est souvent impossible (la population totale est inconnue, mal définie, ou bien elle comprend trop d'individus).

Un échantillonnage consiste à choisir parmi les éléments de la population un ensemble d'unités statistiques¹ pour lesquels des observations sont réalisées. Les unités statistiques choisies constituent un échantillon. Si l'échantillon est bien choisi, les observations permettent d'obtenir des connaissances sur la population.

➤ Comment constituer un échantillon informatif sur le plan statistique ?

¹ La définition d'une unité statistique est donnée page 11.

Partie 2 : Calculs de probabilités en statistiques

I. Pourquoi des probabilités en statistiques ?

La valeur prise par une variable aléatoire à l'issue d'une expérience aléatoire, ne peut pas être « prédite » avec certitude. Toutefois, il est possible de calculer la probabilité des différentes valeurs possibles. Également (voir page 8) :

- o Les statistiques permettent de calculer la probabilité d'observer uniquement par le hasard des écarts au moins aussi importants que ceux observés lors de l'étude. Ainsi : **Statistiques → Calculer la probabilité du hasard.**

II. Variables aléatoires quantitatives discrètes¹

- 1) Loi de probabilité

La loi de probabilité d'une variable aléatoire quantitative discrète est la fonction qui associe à chaque valeur x de la variable aléatoire X sa probabilité :

$$p(x) = P(X = x)$$

Équation 3 : Définition d'une loi de probabilité.

Les probabilités pour l'ensemble des valeurs x de X vérifient :

$$0 \leq p_i \leq 1 \text{ et } \sum_{i=1}^n p_i = 1$$

➤ La loi de Bernoulli et de la loi Binomiale sont des exemples de lois de probabilité. Les connaissez-vous ?

¹ Les résultats présentés dans ce paragraphe sont également applicables aux variables aléatoires qualitatives.

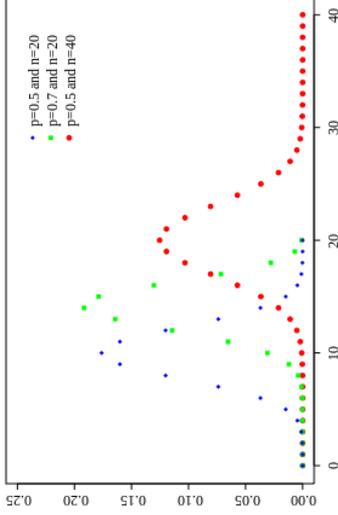


Figure 2 : Exemples de lois binomiales de variables aléatoires quantitatives discrètes. Ces lois binomiales sont définies par différentes valeurs de paramètres p et n . Les valeurs des variables aléatoires X sont en abscisse et les valeurs $P(X = x)$ sont en ordonnée. Cette image est issue de http://fr.wikipedia.org/wiki/Loi_binomiale.

2) Fonction de répartition

La fonction de répartition d'une variable aléatoire X quantitative discrète est la fonction qui donne la probabilité que X soit inférieure ou égale à une valeur x donnée :

$$F(x) = P(X \leq x)$$

Équation 4 : Définition d'une fonction de répartition. La variable aléatoire X est quantitative.

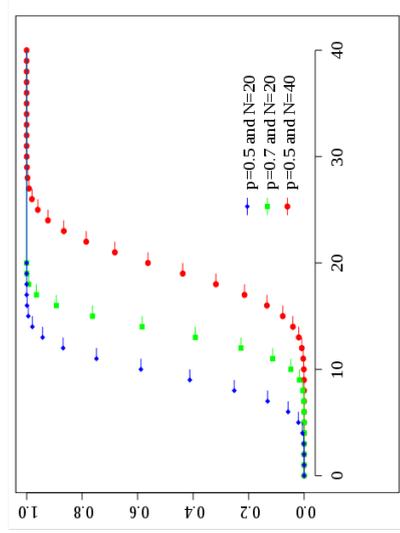


Figure 3 : Fonctions de répartition des variables aléatoires qui suivent les lois binomiales présentées Figure 2 (ci-dessus). Cette image est issue de http://fr.wikipedia.org/wiki/Loi_binomiale.

3) Espérance et variance

(a) Définitions

L'espérance d'une variable aléatoire X (discrète) est égale à la moyenne des valeurs de pondérées par leurs probabilités d'apparition :

$$E(X) = \sum_{i=1}^n x_i \times P(X = x_i)$$

Équation 5 : Formule de l'espérance. La variable aléatoire X est discrète.

➤ Quelle est la différence entre « espérance » et « moyenne » ?

La variance d'une variable aléatoire X (discrète) est égale à la somme des écarts à l'espérance, au carré et multipliée par les probabilités d'apparitions :

$$Var(X) = \sum_{i=1}^n (x_i - E(X))^2 \times P(X = x_i) = E(X^2) - (E(X))^2$$

Équation 6 : Formules de la variance. La variable aléatoire X est discrète.

➤ Pourquoi la racine carrée de la variance, appelée écart type, est souvent utilisée ?

(b) Quelques propriétés

Soient a et b deux constantes et X une variable aléatoire. On a :

$$E(aX + b) = aE(X) + b$$

Équation 7 : Propriété de l'espérance d'une variable aléatoire.

Soient X et Y deux variables aléatoires. L'espérance mathématique d'une somme est égale à la somme des espérances mathématiques. Tandis que l'espérance mathématique d'une différence est égale à la différence des espérances mathématiques :

$$E(X+Y) = E(X) + E(Y)$$

$$E(X-Y) = E(X) - E(Y)$$

Équation 8 : Propriétés de l'espérance de deux variables aléatoires.

Soient X et Y deux variables aléatoires indépendantes¹ :

$$E(X.Y) = E(X).E(Y)$$

Équation 9 : Propriétés de l'espérance de deux variables aléatoires indépendantes.

Soient a et b deux constantes et X une variable aléatoire :

$$Var(aX + b) = a^2 Var(X)$$

Équation 10 : Propriété de la variance d'une variable aléatoire.

Soient X et Y deux variables aléatoires indépendantes. La variance de leur somme est égale à la somme des variances. De même pour la différence :

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

Équation 11 : Propriétés de la variance de deux variables aléatoires indépendantes.

➤ Quelles significations peut-on donner à ces propriétés en termes de reproductibilité des résultats d'une étude statistique ?

¹ "l'indépendance est une notion probabiliste qualifiant de manière intuitive des événements aléatoires n'ayant aucune influence l'un sur l'autre" (Wikipedia, page "Indépendance (probabilités)").

III. Variables aléatoires quantitatives continues

1) Loi(s) normale(s)

(a) Représentation graphique

La loi normale s'applique à des variables aléatoires quantitatives continues définies entre $-\infty$ et $+\infty$. Elle est entièrement décrite par deux paramètres : la moyenne μ et la variance σ^2 .

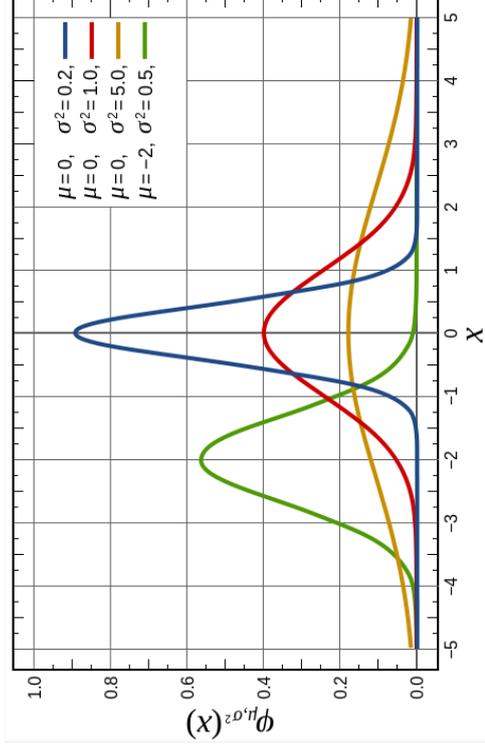


Figure 4 : Exemples de lois normales de variables aléatoires quantitatives continues. Ces lois normales sont définies par différents valeurs de paramètres μ et σ^2 . Les valeurs des variables aléatoires sont en abscisse et les valeurs $f(x)$ sont en ordonnée. Cette image est issue de http://fr.wikipedia.org/wiki/Loi_normale.

(b) Fonction de densité de probabilité

La densité de probabilité f d'une variable aléatoire quantitative continue est la probabilité calculée sur une variation faible (dx) de la variable aléatoire X :

$$P(x \leq X \leq x + dx) = f(x)dx$$

Équation 12 : Définition d'une fonction de densité de probabilité.

Dans le cas de la loi normale, la fonction de densité de probabilité s'écrit¹ :

$$f(x) = \frac{1}{\sqrt{\sigma^2 2\pi}} \exp \left[-\frac{1}{2\sigma^2} (x - \mu)^2 \right]$$

Équation 13 : Fonction de densité de probabilité de la loi normale.

(c) Notation

Une variable aléatoire quantitative continue qui suit une loi normale, par exemple de moyenne 12 et de variance 10 est notée : $X \sim N(\mu = 12, \sigma^2 = 10)$. Sa fonction de densité de probabilité est représentée ci-dessous par la courbe noire. La courbe verte correspond à la fonction de densité de probabilité d'une variable aléatoire de même moyenne (= 12) mais de variance plus élevée (= 20). A noter que puisque la loi normale est une fonction de densité de probabilité :

- Les valeurs de $f(x)$ sont toutes positives
- L'aire totale sous la courbe est égale à 1.

Également, si X est une variable quantitative continue, alors $P(X = x) = 0$.

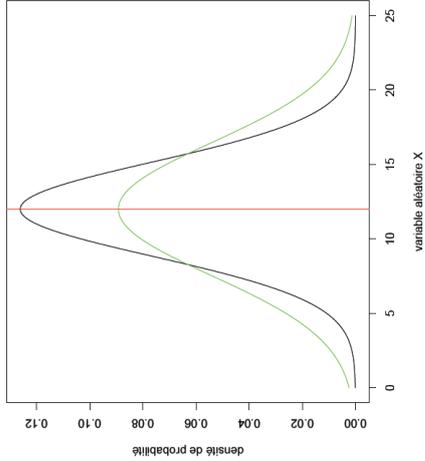


Figure 5 : Représentation de la loi normale de moyenne 12 et de variance 10 (en noire) et de la loi normale de moyenne 12 et de variance 20 (en vert). Une ligne verticale en $x = 12$ est représentée en rouge. Dans les deux cas les aires sous les courbes sont égales à 1.

¹ Cette formule est donnée seulement à titre indicatif. Elle est très peu utilisée dans la pratique.

(d) Fonction de répartition, fonction de densité et calculs de probabilités

La fonction de répartition $F(x)$ est une primitive de la densité de probabilité $f(x)$. La probabilité que X prenne une valeur comprise entre les bornes a et b est donc égale à :

$$P(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$$

Équation 14 : Relation entre fonction de répartition et fonction de densité de probabilité en termes de calcul de probabilité.

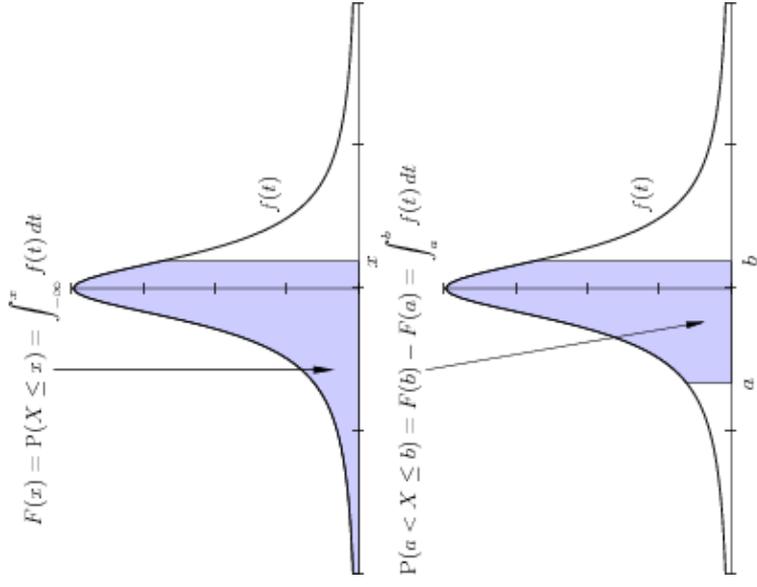


Figure 6 : Surface représentant une probabilité. Sur cet exemple la surface hachurée correspond à la probabilité que X prenne une valeur entre a et b . A noter qu'il est indifférent d'inclure ou d'exclure les bornes dans le calcul de probabilité d'un intervalle, lorsque la fonction de probabilité est continue. En effet $P(X = x) = 0$ si X est continue. Cette image est issue de http://fr.wikipedia.org/wiki/Densit%C3%A9_de_probabilit%C3%A9.

2) Loi normale centrée – réduite

(a) *Fonction de densité de probabilité*

La loi normale centrée - réduite est une loi normale particulière de moyenne 0 et de variance 1 (voir ci-dessous). Les variables aléatoires qui suivent une loi normale centrée réduite sont souvent notées Z .

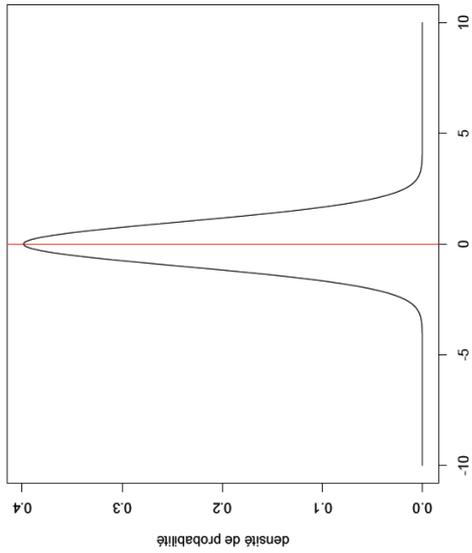


Figure 7 : Représentation de la loi normale centrée réduite. Les variables aléatoires qui suivent une loi normale centrée réduite sont souvent notées Z .

(b) *Fonction de répartition*

La fonction de répartition permet de calculer la probabilité que la variable aléatoire Z ait une valeur inférieure ou égale à une quantité quelconque z (voir page 13 et Figure 8 ci-dessous). Cette fonction de répartition est souvent notée ϕ :

$$\phi(z) = P(Z \leq z) \text{ avec } Z \sim N(0,1)$$

Équation 15 : Fonction de répartition de la loi normale centrée réduite.

➤ Pouvez-vous vérifier graphiquement la relation suivante : $\phi(-z) = 1 - \phi(z)$.

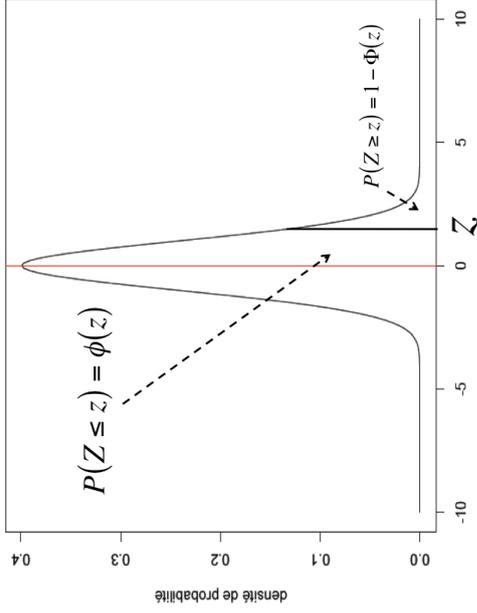


Figure 8 : Illustration d'un calcul de probabilité en utilisant la fonction de répartition ϕ de la loi normale centrée réduite.

3) Calculs de probabilités avec la loi normale

(a) *Lecture de la table de la loi normale*

Les valeurs de la fonction de répartition de la loi normale centrée réduite sont disponibles dans des tables statistiques, pour les valeurs positives de :

z	0.00	0.01	0.02	0.03	0.04
0.0	.5000	.5040	.5080	.5120	.5160
0.1	.5398	.5438	.5478	.5517	.5557
0.2	.5793	.5832	.5871	.5910	.5948
0.3	.6179	.6217	.6255	.6293	.6331
0.4	.6554	.6591	.6628	.6664	.6700
0.5	.6915	.6950	.6985	.7019	.7054
0.6	.7257	.7291	.7324	.7357	.7389

Figure 9 : Extrait de la table de la loi normale centrée-réduite (ou table de Gauss). En ligne et en colonne, les valeurs de Z sont notées. Les valeurs correspondantes $\Phi(z)$ sont dans le tableau. Ainsi si $z = 0,32$ alors $\Phi(z) = 0,6255$. Le reste de la table est présenté en Annexe.

(b) Procédure de « centrage/réduction » d'une variable aléatoire

Une variable aléatoire X qui suit une loi normale quelconque ($X \sim N(\mu, \sigma^2)$) peut être convertie en variable aléatoire Z qui suit une loi normale centrée réduite ($Z \sim N(0,1)$) en effectuant le changement de variable :

$$Z = \frac{X - \mu}{\sigma}$$

Équation 16 : Changement de variable associé à la procédure de « centrage / réduction » d'une variable aléatoire qui suit une loi normale quelconque.

La procédure de centrage – réduction permet de calculer des probabilités associées à une loi normale quelconque, pour diverses valeurs de μ et de σ^2 (il est toujours possible de se rapporter à une loi normale centrée réduite).

➤ Discuter la notion de « Z-Score » couramment utilisée en biologie.

(c) Applications

- Soit X une variable aléatoire qui suit une loi normale de moyenne $\mu = 4$ et la variance $\sigma^2 = 4$, calculer $P(X \leq 6)$?
- Calculer $P(1 \leq X \leq 3)$ si $X \sim N(\mu = 2, \sigma^2 = 4)$, A noter que pour ces calculs il est possible d'utiliser des logiciels statistiques tel que R¹.

IV. Théorème central limite

Le théorème central limite établit le lien entre la loi normale et une grande classe de lois de probabilité, quand le nombre d'observations augmente. C'est l'un des théorèmes les plus importants en statistiques.

¹ <http://cran.r-project.org/>

Énoncé du théorème :

Toute somme de n variables aléatoires indépendantes et qui suivent des lois de probabilité (ou des fonctions de probabilités) identiques, tend vers une variable aléatoire dont la loi converge vers la loi normale quand n augmente.

➤ Pourquoi la plupart des variables étudiées en biologie suivent une loi normale ?

Partie 3 : Échantillonnage et estimation

I. Définition statistique de l'échantillon

- 1) Étude d'une variable aléatoire quantitative¹

Soit X une variable aléatoire de moyenne μ et de variance σ^2 (dans la population). La création d'échantillon de taille n est la réalisation indépendante, n fois, de la variable aléatoire X . C'est une expérience aléatoire. Un échantillon est donc une variable aléatoire à n dimensions, notée : $(X_1, X_2, X_3, \dots, X_n)$.

➤ Illustrer cette notion par un schéma. Représenter d'une part la population et d'autre part les différents échantillons possibles, issus de cette population. Respecter les conventions de notations (les variables aléatoires sont notées en MAJUSCULES, tandis que les valeurs prises par une variable aléatoire, les observations, sont notées en minuscules).

- 2) Moyenne d'un échantillon

La moyenne des valeurs de X dans l'échantillon est également une variable aléatoire quantitative, notée :

$$M = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Équation 17 : Moyenne de la variable aléatoire X dans un échantillon de taille n .

L'expérience aléatoire qui définit un événement élémentaire m de M consiste à tirer n valeurs de X (de façon indépendante) puis à calculer la moyenne de ces valeurs (Équation 1, page 9).

¹ Cette définition statistique de l'échantillon est généralisable à tous les types de variables aléatoires.

- 3) Représentations histogrammes de la variable aléatoire M

Une approche par simulation numérique¹ appliquée au cas d'étude 1 permet d'obtenir les histogrammes suivants. Deux paramètres sont pris en compte : le nombre d'échantillons créés au hasard et la taille de ces échantillons (valeur n).

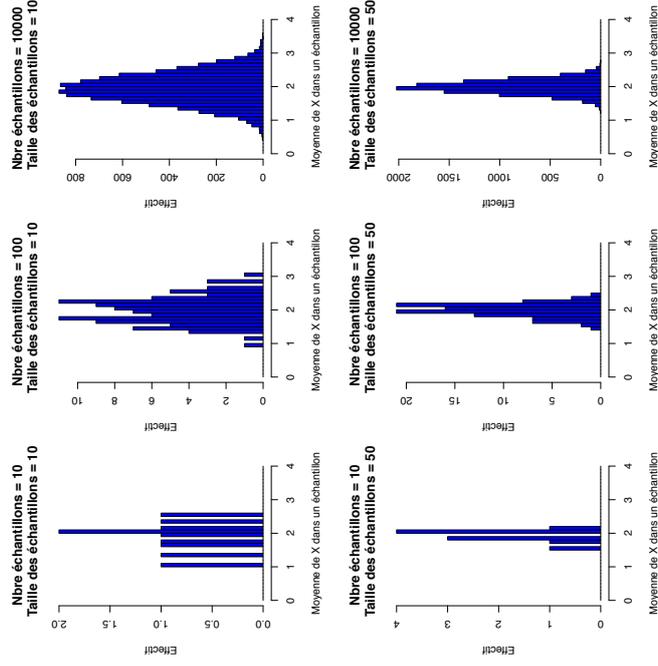


Figure 10 : Résultats de simulations numériques appliquées au cas d'étude 1. La variable aléatoire étudiée est X « écart entre les mesures de la tension artérielle avant et après traitement ». La procédure de création d'un échantillon consiste à tirer n valeurs de X dans l'ensemble $\{0; 1; 2; 3; 4\}$. La moyenne associée à ces valeurs est ensuite calculée et représentée en abscisse de ces graphiques. La procédure de création d'un échantillon est répétée plusieurs fois (10, 100 ou 10000 fois). Ainsi, plusieurs ensembles de valeurs moyennes sont obtenus et représentés en histogramme (voir la définition Figure 1, page 10).

➤ Commenter les résultats des simulations numériques. Quelle est la loi de probabilité de M (VA « moyenne de X dans un échantillon ») ? Quel impact à la taille des échantillons sur cette loi ?

¹ Ces simulations ont été réalisées avec le logiciel R <http://cran.r-project.org/>.

4) Fonction de densité de probabilité associée à la moyenne d'un échantillon

Les résultats des simulations montrent que la loi normale est la fonction de densité de probabilité associée à la variable aléatoire M . La caractérisation précise de cette loi normale nécessite de calculer l'espérance et la variance de M .

(a) *Espérance et variance de la moyenne d'un échantillon*

➤ Exprimer l'espérance et la variance de M , en fonction la moyenne μ et de variance σ^2 de X .

$$E(M) = \mu \quad \text{et} \quad \text{Var}(M) = \frac{\sigma^2}{n}$$

Équation 18 : Formules de l'espérance et de la variance de la variable aléatoire « moyenne de X dans un échantillon de taille n » (n étant le nombre d'observations de X). μ et σ^2 sont respectivement la moyenne et la variance de X .

(b) *Application du théorème central limite*

Les résultats obtenus par la simulation (voir paragraphe précédent) peuvent être « prédits » à l'avance par application du théorème central limite (voir page 21). Ainsi, la moyenne M suit (si n est suffisamment grand) la loi normale suivante :

$$M \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Équation 19 : Loi de probabilité de la variable aléatoire « moyenne de X dans un échantillon de taille n ». Cette formule est valable si n est suffisamment grand (dans la pratique > 30).

Ce résultat rend possible les calculs de probabilités¹ associées à la moyenne d'un échantillon, sans connaître la loi de probabilité de la variable aléatoire X initialement étudiée.

¹ Ce résultat permet de retrouver toutes les formules utilisées pour la réalisation d'analyses statistiques classiques (estimations et tests d'hypothèses). Il est important de bien le comprendre.

II. Estimation statistique

1) Principe

L'estimation statistique est la procédure d'utilisation des informations obtenues à partir d'un échantillon afin de déduire des résultats concernant l'ensemble de la population.

➤ Illustrer la notion d'estimation statistique en utilisant le(s) cas d'étude(s). Noter de manière différente les paramètres connus et les paramètres inconnus.

	Moyenne	Variance	Proportion
Population	μ	σ^2	π
Échantillon	m	Var	p

Tableau 2 : Conventions de notations de paramètres, classiquement utilisés en statistiques à l'Université Paris 7. Les valeurs des paramètres dans la population sont INCONNUES, tandis que les valeurs des paramètres dans les échantillons sont CONNUES (calculés à partir des observations).

2) Qualité d'un estimateur

Un estimateur (tel que la moyenne d'un échantillon) a peu de chances de proposer la valeur exacte du paramètre inconnu. Cela est dû à l'existence d'erreurs d'échantillonnages (une partie de la population a été omise). Pour qu'un estimateur fournisse des estimations utiles, il doit être « sans biais » et « convergent ». C'est le cas de la moyenne d'un échantillon.

(a) *Estimateur sans biais*

Un estimateur est dit « sans biais » si son espérance est égale à la valeur exacte du paramètre dans la population :

$$E(M) = \mu$$

Équation 20 : Propriété d'un estimateur sans biais, appliquée à la moyenne d'un échantillon. M est l'estimateur et μ le paramètre inconnu de la population.

➤ Relier cette propriété aux résultats des simulations présentées Figure 10, p. 24.

(b) *Estimateur convergent*

Un estimateur est dit « convergent » si sa variance tend vers 0 quand la taille n de l'échantillon augmente :

$$\lim_{n \rightarrow +\infty} \text{Var}(M) = 0$$

Équation 21 : Propriété d'un estimateur convergent, appliquée à la moyenne d'un échantillon. M est l'estimateur.

Ainsi, les estimations du paramètre μ obtenues à partir de plusieurs échantillons seront d'autant plus proches les unes des autres que la taille n des échantillons est grande.

➤ Relier cette propriété aux résultats des simulations présentées Figure 10, p. 24.

(c) *Applications à la variance*

➤ La variance d'un échantillon (Équation 2, page 9) est-elle un bon estimateur de la variance σ^2 de X dans la population ?

3) Estimation ponctuelle

L'estimation ponctuelle consiste à proposer une unique valeur pour le paramètre inconnu. Cette valeur est notée « $\hat{\mu}$ » :

$$\hat{\mu} = M \quad \text{et} \quad \hat{\sigma}^2 = \frac{n}{n-1} \text{Var}$$

Équation 22 : Estimations ponctuelles de la moyenne et de la variance d'une variable aléatoire quantitative. Les formules de m et Var sont présentées page 9.

➤ Les résultats obtenus à partir d'une procédure d'estimation ponctuelle sont-ils fiables ?

4) Estimation par intervalle de confiance

Construire un intervalle de confiance consiste à donner un intervalle de valeurs dans lequel le paramètre inconnu a une probabilité importante de se trouver. Dans le cas de la moyenne μ , on aura :

$$P(M - e \leq \mu \leq M + e) = 1 - \alpha$$

Équation 23 : Formule générale de l'intervalle de confiance de la moyenne. μ est le paramètre de la population à estimer et M un estimateur. α est le risque d'erreur associé au calcul de l'intervalle de confiance. Ce risque est dans la pratique fixé à 5%.

➤ Déterminer la valeur du paramètre e ? (dans le cas où σ^2 est connue)

$$\mu \in \left[M \pm z_{1-\alpha/2} \sqrt{\frac{\sigma^2}{n}} \right]$$

Équation 24 : Formule de l'intervalle de confiance de la moyenne d'une variable aléatoire quantitative. La valeur de $z_{1-\alpha/2}$ est lue dans la table de la loi normale centrée réduite. Elle dépend de α .

➤ Quelles sont les valeurs de $z_{1-\alpha/2}$ si $\alpha = 5\%$ (valeur classique), $\alpha = 1\%$ et $\alpha = 10\%$?

Rappel : $\Phi(z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$.

III. Cas des petits échantillons ($n \leq 30$)

Dans le cas où le nombre d'observations de la variable aléatoire est faible (petit échantillon), le théorème central limite (voir page 21) s'applique relativement mal. Si la variable aléatoire suit une loi normale et que la variance σ^2 est connue, la formule de l'Équation 24 est toujours valable. Cependant, si la variance n'est pas connue, son estimation ponctuelle devra être associée à l'utilisation de la loi de Student. L'intervalle de confiance s'écrit alors :

¹ Dans le cas où la variance n'est pas connue, l'estimateur ponctuel pourra être utilisé. La loi de Student remplacera la loi Normale.

$$\mu \in \left[M \pm t_{\alpha, (n-1), ddf} \sqrt{\frac{\hat{\sigma}^2}{n}} \right] \text{ avec } t_{\alpha, (n-1), ddf} \text{ qui est lue dans la table de Student}$$

Equation 25 : Formule de l'intervalle de confiance de la moyenne d'une variable aléatoire quantitative, dans le cas des petits échantillons (ddl étant un nombre de degré de liberté).

IV. Cas des variables aléatoires qualitatives (2 classes)

1) Rappel du second cas d'étude

Constituer un échantillon de taille n consiste (par exemple) à lancer n fois un jeton¹. La proportion (par exemple de « pile ») est un paramètre qui décrit l'échantillon (comme la moyenne dans le cas d'une variable aléatoire quantitative). On a : $p = \frac{1}{n} \sum_{i=1}^n x_i$.

2) Fonction de densité de probabilité de la proportion d'un échantillon

D'après le théorème central limite, la proportion P suit (si n est suffisamment grand) une loi normale :

$$P \sim N \left(\pi, \frac{\pi(1-\pi)}{n} \right)$$

Equation 26 : Loi de probabilité de la proportion d'une variable aléatoire qualitative à deux classes dans un échantillon. Dans la pratique, il faut pour appliquer ce résultat : $n \times \pi \geq 5$ et $n \times (1-\pi) \geq 5$.

➤ Quelle relation existe-t-il entre cette formule et la loi de Bernoulli ?

3) Estimation ponctuelle

L'estimation ponctuelle consiste à proposer une unique valeur pour le paramètre inconnu, ainsi :

$$\hat{\pi} = p$$

Equation 27 : Estimation ponctuelle de la proportion d'une variable aléatoire qualitative à deux classes.

¹ L'unité statistique est ici « un lancer du jeton », la VA est « le côté du jeton obtenu ».

4) Intervalle de confiance

L'intervalle de confiance consiste à donner un intervalle de valeurs dans lequel le paramètre inconnu a une probabilité importante de se trouver, ainsi :

$$\pi \in \left[P \pm z_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}} \right] \text{ avec } z_{1-\alpha/2} = 1,96 \text{ si } \alpha = 5\%$$

Equation 28 : Formule de l'intervalle de confiance de la proportion d'une variable aléatoire qualitative à deux classes.

V. Exemples d'applications

➤ Calculer les intervalles de confiances associés aux observations présentées ci-dessous en utilisant les 50 observations, puis seulement les 10 premières.

3 ; 2 ; 0 ; 2 ; 4 ; 3 ; 1 ; 1 ; 3 ; 2 ; 3 ; 1 ; -1 ; 1 ; -1 ; 0 ; 0 ; 4 ; 0 ; -1 ; 3 ; 4 ; 3 ; 2 ; 0 ; 2 ; -1 ; 0 ; 1 ; 1 ; 0 ; 3 ; 0 ; 0 ; -1 ; 0 ; 3 ; 0 ; 0 ; 3 ; 3 ; 4 ; 3 ; 0 ; 2 ; 1 ; 0 ; 2 ; 0 ; 0

Tableau 3 : Ensemble de mesures (ou observations) de la variable aléatoire étudiée dans le cas d'étude 1. Les valeurs écrites en rouge ont été présentées Tableau 1 (page 7). Le nombre d'observations est de 50, la moyenne dans l'échantillon est de 1,3 et la variance de l'échantillon est de 2,37 (soit une variance estimée de 2,42).

Partie 4 : Les tests d'hypothèses

I. Introduction aux tests d'hypothèses

Les tests d'hypothèses reposent sur un raisonnement différent de celui de l'estimation. Il s'agit de faire des hypothèses concernant les valeurs des paramètres d'une variable aléatoire X dans la population et ensuite d'utiliser les observations $(x_1, x_2, x_3, \dots, x_n)$ dans un échantillon pour choisir entre les hypothèses.

1) Rappel des cas d'étude

(a) Cas 1 : le médicament est-il efficace ?

- **Définition de la variable aléatoire** : la variable aléatoire X est quantitative (écart entre les mesures de la tension artérielle réalisées avant et après la prise du médicament pendant six semaines).
- **Paramètre de la population étudié** : moyenne μ de la variable aléatoire X dans la population de référence.

- **Hypothèse de travail** : le médicament est efficace si la moyenne $\mu > 0$.

(b) Cas 2 : le jeton est-il équilibré ?

- **Définition de la variable aléatoire** : la variable aléatoire est qualitative à deux classes (côté supérieur – pile ou face – du jeton après avoir été lancé).
- **Paramètre de la population étudiée** : proportion π de côtés « pile » obtenue dans la population de référence.

- **Hypothèse de travail** : le jeton est équilibré si la proportion $\pi = \frac{1}{2}$.

➤ Discuter les intervalles de confiance calculés au chapitre précédent. Le médicament est-il efficace ? Le jeton est-il équilibré ?

2) Illustration de l'approche décisionnelle des tests d'hypothèses

Soit X une variable aléatoire quantitative qui suit une loi normale de paramètres inconnus. L'objectif est de choisir entre deux hypothèses nommées H_A et H_B , à partir d'une observation x de X .

$$\text{Avec : } \begin{cases} H_A : X \sim N(\mu = 0; \sigma^2 = 4) \\ H_B : X \sim N(\mu = 4; \sigma^2 = 16) \end{cases}$$

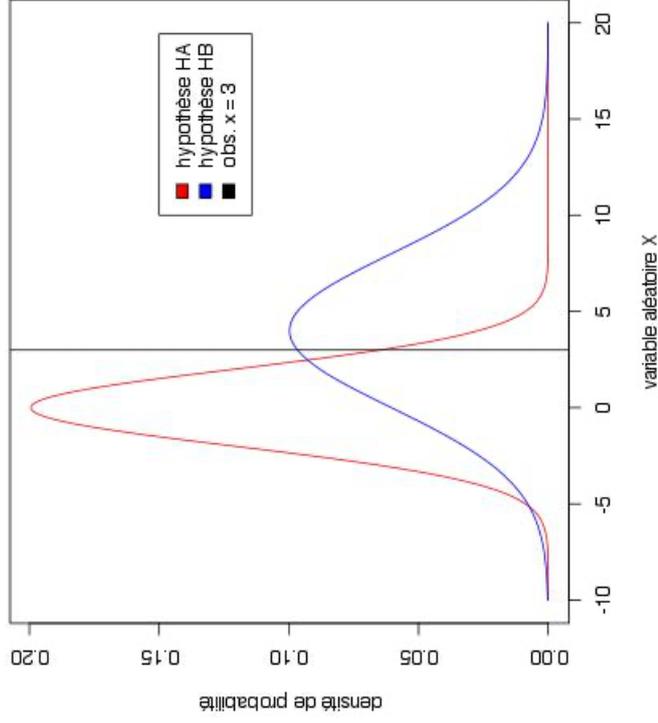


Figure 11 : Illustration de la mise en concurrence de deux hypothèses H_A et H_B . Une observation de X ($x = 3$) est notée par un trait vertical. Le choix entre les hypothèses repose sur une valeur de X observée.

➤ Quelle hypothèse choisir si l'observation de X est $x = 3$?

II. Définition d'une règle de décision

- 1) Objectif

Définir une « règle rationnelle » permettant de choisir entre les deux hypothèses, en fonction d'une observation quelconque x de la variable aléatoire X . Cette règle divisera l'ensemble des valeurs possibles de X en deux sous-ensembles :

- L'ensemble (R_A) des valeurs de X pour lesquelles il faudra choisir l'hypothèse H_A
- L'ensemble (R_B) des valeurs de X pour lesquelles il faudra choisir l'hypothèse H_B

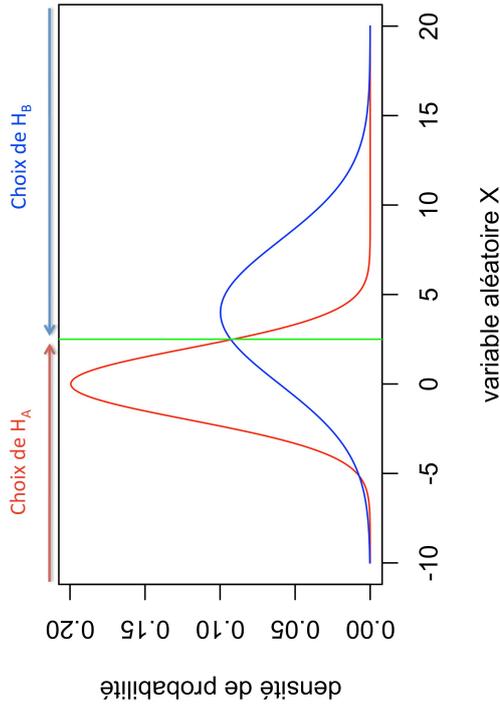


Figure 12 : Exemple de règle de décision applicable au choix des hypothèses H_A et H_B présentées Figure 11. Si $X \leq 2.5$, l'hypothèse H_A est retenue. Si $X > 2.5$, l'hypothèse H_B est retenue. La valeur $x = 2.5$ est à l'intersection entre les deux fonctions de densité de probabilité.

- 2) Évaluation d'une règle de décision

Évaluer une règle de décision consiste à calculer la probabilité de commettre une erreur sur chacune des hypothèses :

- **Erreur sur l'hypothèse H_A** : choisir H_B alors que H_A est vraie
- **Erreur sur l'hypothèse H_B** : choisir H_A alors que H_B est vraie

- **Évaluer la règle de décision** : $X \leq 2.5$ choisir H_A et $X > 2.5$ choisir H_B . Deux probabilités sont à calculer : 1) erreur sur l'hypothèse H_A et 2) erreur sur l'hypothèse H_B .

Il existe de multiples règles de décision. Certaines privilégient l'hypothèse H_A (le risque d'erreur sur H_A est faible), d'autres l'hypothèse H_B (le risque d'erreur sur H_B est faible). Le graphique ci-dessous représente les probabilités d'erreurs associées à une cinquantaine de règles de décision différentes.

Évaluation des règles de décision (fonction de x)

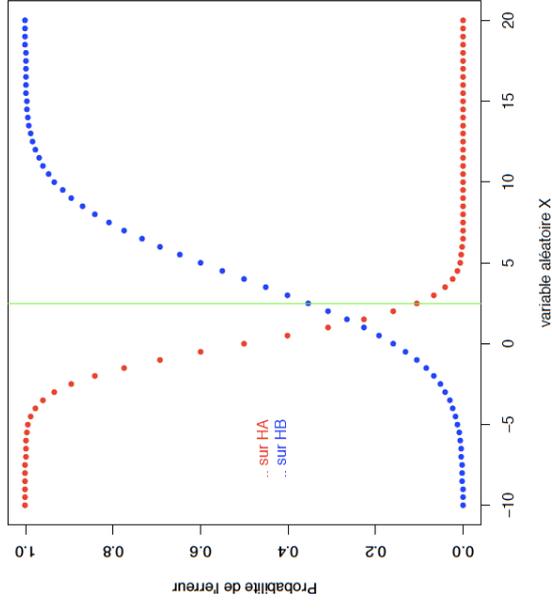


Figure 13 : Calculs des probabilités d'erreurs associées aux hypothèses H_A et H_B pour différentes règles de décisions. Le trait vertical vert représente la valeur $x = 2.5$, utilisée pour représenter la règle de décision de la Figure 11.

- **Discuter la notion de « bonne règle » de décision. Faut-il équilibrer le risque entre les deux hypothèses ? Faut-il privilégier une hypothèse par rapport à l'autre ?**

III. Théorie des tests statistiques

- 1) Hypothèse nulle et hypothèse alternative

Dans la pratique en statistiques (et en biologie) l'hypothèse selon laquelle seul le hasard est responsable des effets observés dans les échantillons est privilégiée. Cette hypothèse est nommée « **hypothèse nulle** » et notée H_0 . L'hypothèse en concurrence est nommée « **hypothèse alternative** » et notée H_1 . C'est l'hypothèse selon laquelle le hasard et l'effet biologique testé interviennent.

Hypothèse H_0 (nulle)	Hasard seul
Hypothèse H_1 (alternative)	Hasard + effet testé (efficacité du médicament par exemple)

Tableau 4 : Description des hypothèses mises en concurrence dans la théorie des tests en statistiques.

➤ **Reprendre les hypothèses nulles et alternatives associées aux deux cas d'études. Discuter les risques d'erreurs associées en termes de faux positifs et de faux négatifs.**

- 2) Définitions des risques d'erreurs

(a) *Erreur sur l'hypothèse H_0*

Cette erreur est nommée « risque de première espèce » et notée α . C'est la probabilité de choisir H_1 ¹ alors que H_0 est vrai.

(b) *Erreur sur l'hypothèse H_1*

Cette erreur est nommée « risque de deuxième espèce » et notée β . C'est la probabilité de choisir H_0 alors que H_1 est vraie.

¹ Il est de convention en statistiques de dire « rejeter H_0 » plutôt que « choisir H_1 ». Les notions restent toutefois équivalentes.

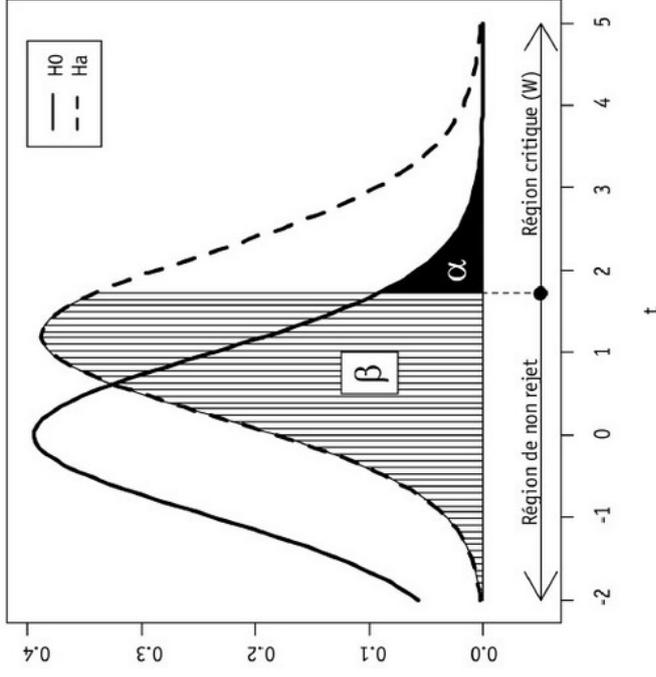


Figure 14 : Représentation des risques statistiques en relation avec la réalisation d'un test d'hypothèses. Le risque α est le risque de première espèce tandis que le risque β est le risque de seconde espèce (les définitions sont données dans le texte principal). Image issue du site Internet : <http://www.cairn.info/revue-staps-2007-3-page-49.htm> . H_0 représente l'hypothèse alternative, notée H_1 dans le cours. La valeur t est la « statistique du test », c'est à dire le paramètre qui est calculé à partir des observations pour choisir entre les hypothèses (voir ci-dessous).

La règle de décision est ensuite choisie de manière à garantir un risque d'erreur α très faible (généralement de 5%). C'est la raison pour laquelle l'hypothèse H_0 est dite « privilégiée » dans la théorie des tests statistiques.

- 3) Notion de "statistique du test"

Pour choisir entre les hypothèses H_0 et H_1 , les paramètres connus de l'étude statistique sont utilisés. Ces paramètres font nécessairement référence aux observations réalisées dans les échantillons. La statistique du test représente la combinaison des paramètres qui va être utilisée pour définir une règle de décision (par exemple "moyenne de X ", ou "écart entre les moyennes de X obtenues dans deux échantillons", etc.).

- 4) Application à la comparaison d'une moyenne et d'une norme (1^{er} cas d'étude)

- Soit M (moyenne de X dans un échantillon) la statistique du test. Déterminer la limite m_{lim} de M , définissant la règle de décision pour un risque α quelconque.
- Les observations présentées dans le Tableau 3 (page 30) sont-elles en faveur d'une efficacité du médicament ? (si $\alpha = 0.05$ et si $\alpha = 0.01$)

- 5) Déroulement d'un test d'hypothèses

(a) Formulation des hypothèses et définition de la statistique du test

Les hypothèses d'un test sont formulées en termes de paramètres relatifs à la population étudiée (ces paramètres sont inconnus). La statistique du test est formulée en termes de paramètres relatifs à (ou aux) échantillon(s).

(b) Vérification des conditions d'application

Les conditions d'application d'un test sont en relation directe avec la taille du ou des échantillons utilisés. Elles sont particulières dans le cas des petits échantillons ($n < 30$, voir page 39).

(c) Caractérisation de la fonction de densité de probabilité sous H_0

La loi normale est souvent applicable comme fonction de densité de probabilité. Sous H_0 les paramètres (moyenne et variance) sont généralement connus.

(d) Définition de la règle de décision

La règle de décision dépend d'une part de la statistique du test (paramètre noté en abscisse) et d'autre part du risque d'erreur de première espèce α (positionnement du trait vertical). Deux types de règles de décision s'utilisent, associées aux tests unilatéraux et aux tests bilatéraux (voir ci-dessous).

- **Test unilatéral**

La règle de décision comporte une seule zone de rejet de l'hypothèse H_0 au profit de l'hypothèse H_1 .

- **Test bilatéral**

La règle de décision comporte deux zones de rejets de l'hypothèse H_0 au profit de l'hypothèse H_1 . Le risque de première espèce α est alors partagé.

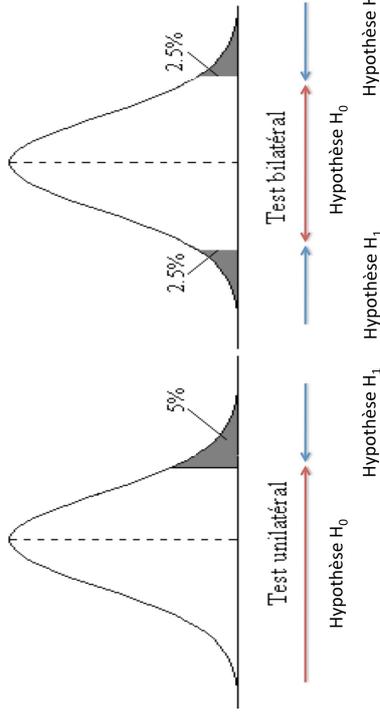


Figure 15 : Représentation des lois de probabilité associées à l'hypothèse nulle H_0 dans le cas d'un test unilatéral (à gauche) et bilatéral (à droite). Sur cet exemple, le risque de première espèce α est fixé dans les deux cas à 5%.

- Comment changerait la valeur m_{lim} de M (précédemment calculée) avec une règle de décision bilatérale ?

- **Conclusion**

La conclusion du test repose sur le calcul de la statistique du test à partir des observations de la variable aléatoire dans un échantillon.

IV. Comparaison de deux moyennes

Un nouvel échantillon de 50 patients est constitué. Ces patients ont reçu un second médicament dont l'efficacité doit être comparée au premier. Les nouvelles mesures de la variable aléatoire X sont présentées ci-dessous :

-1 ; 3 ; 3 ; 3 ; 0 ; 3 ; 0 ; -1 ; 3 ; 2 ; 1 ; -1 ; 3 ; 0 ; 1 ; 0 ; 3 ; 3 ; -1 ; 3 ; 1 ; 1 ; 3 ; 0 ; 2 ; 0 ; 2 ; 0 ; 1 ; 0 ; 2 ; 4 ; 2 ; -1 ; 1 ; 4 ; 0 ; 3 ; 0 ; 4 ; 0 ; 2 ; 2 ; 3 ; 2 ; 1 ; 4

Tableau 5: Ensemble de mesures (ou observations) obtenues sur des patients ayant reçu un second médicament. Le nombre d'observation est de 50. Leur moyenne est de 1,4 et l'estimation de la variance de 2,41.

- Ecrire les hypothèses du test statistique à réaliser. Les discuter en termes d'échantillons et de population(s).
- Soit $D = M^{E1} - M^{E2}$ la statistique du test. Déterminer la valeur de l'écart d_{lim} à partir de laquelle l'hypothèse nulle est rejetée (des effets différents sont associés aux médicaments).
- Quelle conclusion tirer à partir des observations présentées dans le Tableau 3 (page 30) et le Tableau 5 (ci-dessus) ?

V. Cas des petits échantillons

Dans le cas où un des échantillons est composé d'un nombre faible d'observations (généralement inférieur à 30), l'application des tests statistiques présentés ci-dessus nécessite des conditions d'applications particulières.

- 1) Distribution de la variable aléatoire étudiée

Le théorème central limite s'applique mal si $n < 30$ (voir page 21). Pour que la variable aléatoire « moyenne M d'un échantillon » suive une loi normale (voir page 25), il faut que la variable aléatoire X suive elle-même une loi normale.

- 2) Égalité des variances

L'estimation de la variance est d'autant plus difficile que la taille du (ou des) échantillon(s) disponibles est petite. Pour éviter de commettre des erreurs, les échantillons sont

généralement rassemblés. Cette étape n'est réalisable que si l'égalité des variances des populations dont sont extraites les échantillons est démontrée¹.

- 3) Statistique du test

Dans le cas des petits échantillons, la statistique du test à utiliser est présentée ci-dessous (Équation 29). La variance σ^2 est la variance commune, calculée après avoir vérifié l'égalité des variances associées aux deux petits échantillons comparés.

$$T = \frac{M_1 - M_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

Équation 29 : Statistique du test de comparaison de moyennes dans le cas de petits échantillons. Sous l'hypothèse H_0 , ce paramètre suit une loi de Student à $n_1 + n_2 - 2$ degré de liberté (ddl).

Si l'hypothèse H_0 est vraie, le paramètre T suit une loi de Student² à $(n_1 + n_2 - 2)$ ddl, ci-dessous).

- 4) Utilisation de la loi de Student

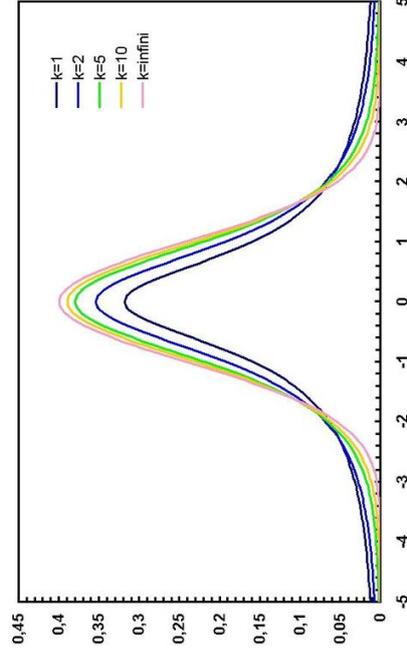


Figure 16 : Fonctions de densité de probabilité de la loi de Student, pour différentes valeurs de degrés de liberté (k). La loi de Student converge vers la loi normale quand k augmente. Cette image est issue du site internet http://fr.wikipedia.org/wiki/Loi_de_Student.

¹ Cela nécessite la réalisation d'un test statistique : le test de Fisher.

² La loi de Student a déjà été évoquée pour le calcul des intervalles de confiance de la moyenne dans le cas des petits échantillons (page 31).

- Réaliser un test de comparaison des moyennes en utilisant seulement les 10 premières valeurs des échantillons présentés Tableau 3 et Tableau 5. Représenter graphiquement la valeur t_{lim} , lue dans la table de Student et utilisée comme limite pour la règle de décision. Positionner également le risque α .
- Les résultats sont-ils différents de ceux obtenus avec l'ensemble des valeurs ?

Partie 5 : Étude des VA qualitatives multi-classes

I. Exemple de cas d'étude

Une étude statistique supplémentaire est réalisée dans le but d'évaluer le ressenti des patients vis à vis du traitement de leur hypertension artérielle. Les patients doivent indiquer s'ils ont ressenti les effets du traitement de manière « très bénéfique » ; « bénéfique » ; « neutre » ou « mauvais ». Des résultats sont présentés dans le tableau de contingence ci-dessous (Tableau 6).

- Définir la variable aléatoire de cette étude. Décrire la population et le (ou les) échantillon(s).
- Quelle serait les effectifs attendus si la répartition des patients dans les classes suivait une loi uniforme ?

Effet du traitement	Effectifs observés
Très bénéfique	12
Bénéfique	33
Neutre	52
Mauvais	10
Total	107

Tableau 6 : Tableau de contingence associé à l'étude de la variable aléatoire « effet du traitement ressenti ». Cette variable aléatoire est qualitative, à 4 classes. Un échantillon de 107 patients est disponible. Les effectifs observés dans chaque classe sont notés dans ce tableau.

L'analyse des variables aléatoires qualitatives multi-classes repose sur l'étude des tableaux de contingences dans lesquels des effectifs (nombres d'individus) dans chaque classe de la variable aléatoire étudiée.

- Donner d'autres exemples de variable aléatoire multi-classes.

II. Test d'hypothèses du χ^2 (Chi2)

1) Principe

Le test d'hypothèse du χ^2 permet d'évaluer si la répartition des individus dans les classes d'une variable aléatoire quantitative multi-classes est compatible avec la répartition attendue selon un modèle théorique (par exemple répartition uniforme).

2) Hypothèses

- H_0 : Le modèle théorique est vrai (c'est l'hypothèse nulle)
- H_1 : Le modèle théorique n'est pas vrai (c'est l'hypothèse alternative)

3) Statistique du test

Le test d'hypothèses du χ^2 utilise comme statistique du test, une mesure d'adéquation (ou critère du χ^2) entre les effectifs observés (notés O_i) et les effectifs théoriques (notés C_i), c'est à dire les effectifs attendus si le modèle théorique est vrai (hypothèse H_0). A noter que pour être utilisé, il faut que toutes les valeurs C_i soient supérieures à 5.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - C_i)^2}{C_i}$$

Équation 30 : Mesure d'adéquation utilisée dans le test d'hypothèses du χ^2 . Les O_i représentent les effectifs observés et les C_i les effectifs théoriques. k est le nombre de classes de la variable qualitative étudiée.

➤ Comment est la valeur de la mesure d'adéquation (faible ou forte) si le modèle théorique est vrai ? Pourquoi est-il important de réaliser une division par C_i dans le calcul du χ^2 ? Pourquoi faut-il que les C_i soient > 5 ? Que faire si cela n'est pas le cas ?

Si l'hypothèse H_0 est vraie, le critère du χ^2 suit une loi du Chi2 (voir ci-dessous Figure 17) pour un nombre de degrés de liberté (ddl) donné :

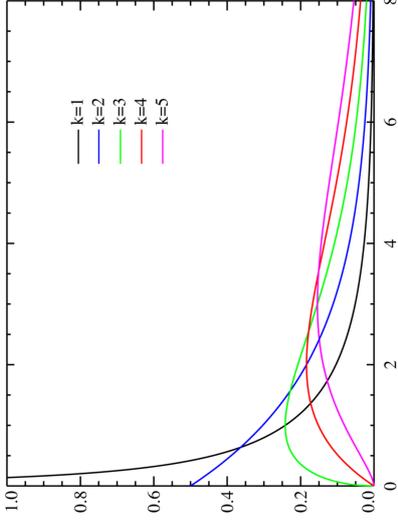


Figure 17 : Fonctions de densité de probabilités du Chi2, pour différentes valeurs de degré de liberté. L'axe des abscisses donne les valeurs de la VA et l'axe des ordonnées les valeurs correspondantes de la densité de probabilité. Cette image est extraite du site Wikipedia https://fr.wikipedia.org/wiki/Loi_du_%CF%87%87%CF%82%82.

4) Règle de décision

La règle de décision du test dépend alors du risque d'erreur de première espèce α , tel que :

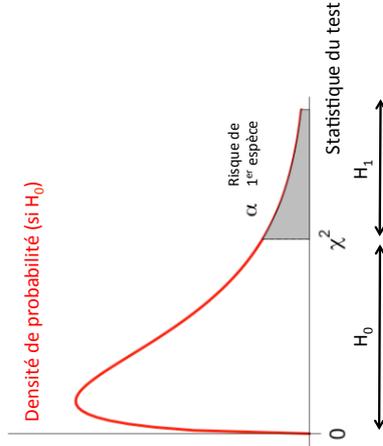


Figure 18 : Règle de décision associée à la réalisation du test du Chi2 (étude d'une VA qualitative multi-classe). La valeur limite peut être lue dans une table statistique (pour risque et un nombre de ddl donné).

- Appliquer le test du χ^2 aux observations présentées dans le Tableau 6 ; pour un modèle théorique de loi uniforme. Commenter les résultats.

III. Pour aller plus loin

Il existe une variante du test de χ^2 (ci-dessus) et qui consiste à comparer les répartitions d'individus dans les classes de la variable entre deux échantillons. Le modèle théorique consiste à considérer que les répartitions sont les mêmes.

Une illustration de ce test pourrait être de considérer de manière séparée, les hommes et les femmes qui ont participé à la précédente étude (voir ci-dessus, Tableau 6). La question serait alors "y a-t-il un ressenti des effets du médicament, différent entre les hommes et les femmes ?".

Effet du traitement	Femmes	Hommes	Total
Très bénéfique	9	3	12
Bénéfique	11	22	33
Neutre	10	42	52
Mauvais	9	1	10
Total	39	68	107

Tableau 7 : Tableau de contingence associé à l'étude de la variable aléatoire « jour de la naissance ». Cette variable aléatoire est qualitative, à 7 classes. Deux échantillons de 4231 individus et 4038 individus sont disponibles. Il s'agit cette fois de comparer la répartition des naissances entre les années 1 et 2.

- Appliquer le test du χ^2 aux observations présentées ci-dessus. Définir le modèle théorique, calculer les effectifs attendus (selon ce modèle), la statistique du test et conclure.

Partie 6 : Évaluer les performances d'une analyse statistiques

I. Précision d'un intervalle de confiance

L'intervalle de confiance permet d'évaluer la précision de l'estimation d'un paramètre statistique d'une population, à partir des observations d'un échantillon. Pour un risque d'erreur donné (par exemple $\alpha = 5\%$), plus l'intervalle de confiance est resserré autour de l'estimation ponctuelle du paramètre d'intérêt plus il est « précis ».

Dans le cadre d'une étude statistique, la précision d'un intervalle de confiance peut être utilisé de deux manières :

- Compte tenu d'une taille d'échantillon donnée, quel est le taux de précision de l'estimation obtenue ?
- Compte tenu d'un taux de précision souhaité, quelle est la taille d'échantillon nécessaire pour l'obtenir ?

- Quelle doit être la taille de l'échantillon pour réduire de moitié l'amplitude des intervalles de confiance calculés à partir des données du Tableau 3 (page 53).

II. Degré de signification d'un test statistique

Lorsqu'un test statistique aboutit au rejet de l'hypothèse H_0 , le degré de signification (nommé « p-value » en anglais) mesure la confiance qu'il est possible d'avoir en cette conclusion, par rapport à l'arbitraire de $\alpha = 5\%$. C'est la valeur limite ($< 5\%$) qui aurait pu être prise comme risque d'erreur sur H_0 , tout en conservant la conclusion de rejet de l'hypothèse H_0 .

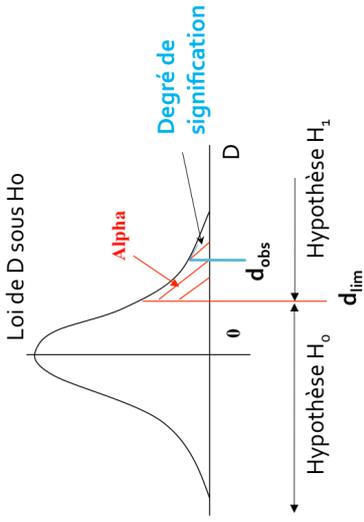


Figure 19 : Illustration du degré de signification associé à un test statistique. Le degré de signification est noté p (p -value en anglais). Plus sa valeur est faible, plus le test est dit « significatif ».

Ainsi, le calcul de la p value (dans le cas d'un test unilatéral) est :

$$p\text{Value} = \Pr(D > d_{\text{obs}})$$

Équation 31 : Formule utilisée pour calculer le degré de signification d'un test statistique. ici D est le paramètre utilisé pour la statistique.

- Calculer les degrés de signification associés aux tests statistiques précédemment réalisés. Discuter la notion de règle de décision, en quoi le choix d'une valeur fixe de risque α est-il arbitraire ?

III. Notion de puissance d'un test statistique

La puissance d'un test statistique est la probabilité de rejeter H_0 alors que H_1 est vraie (détecter un effet – par exemple efficacité d'un médicament ou truquage d'un jeton – alors que cet effet existe réellement). La puissance est directement reliée au risque de 2nd espèce β : $P = 1 - \beta$.

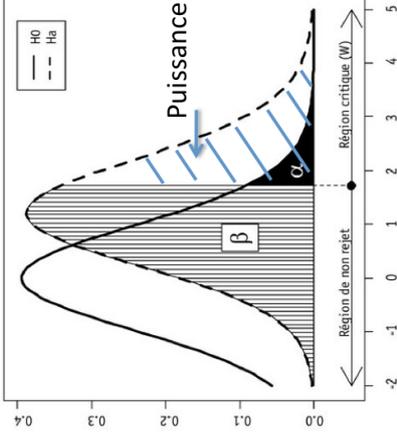


Figure 20 : Illustration de la puissance associée à la règle de décision d'un test statistique. La puissance est la probabilité de rejeter H_0 alors H_1 est vraie, elle est donc associée à la loi sous H_1 .

- Calculer les puissances des tests statistiques réalisés à partir des données du Tableau 3 et du Tableau 8 (page 53).

IV. Simulations appliquées au cas d'étude 2 (logiciel R)

- 1) Combien de fois faut-il lancer un jeton ?

- Discuter les graphiques présentés en annexes.

- 2) Combien de jetons faut-il lancer ?

- Discuter les graphiques présentés en annexes.