



Estimation du nombre d'animaux nécessaires

Hugues Jacobs.

Version française

English version below (page 5)

La réglementation sur l'expérimentation animale (décret 2013-118 transposition de la directive européenne 2010/63/UE), impose de veiller à n'utiliser que le nombre approprié d'animaux dans les protocoles proposés et mis en œuvre. Ainsi, la rubrique 3.3.5 « Stratégie d'expérimentation ou d'observation et approche statistique utilisée afin de réduire au minimum le nombre d'animaux, la douleur, la souffrance et l'angoisse, infligées et l'impact environnemental, le cas échéant – si une étude statistique est prévue, indiquez et justifiez les tests choisis : » du formulaire de demande d'autorisation de projet a pour but de recueillir les éléments permettant d'évaluer l'adéquation du projet avec les obligations réglementaires.

Le plus souvent, les projets de recherche proposés visent à démontrer l'existence d'un effet, au travers de tests statistiques. Cependant, l'évaluation rationnelle du nombre d'animaux nécessaires pour pouvoir effectivement observer cet effet escompté (s'il existe) apparaît peu maîtrisée et ne repose le plus souvent que sur l'empirisme et l'intuition. Or, cette démarche est au centre du processus de conception de l'expérience. Ici, nous proposons une méthode très simple, valable dans la grande majorité des cas, qui permettra de calculer sans difficulté une estimation du nombre d'animaux requis

1. La distribution Normale (N)

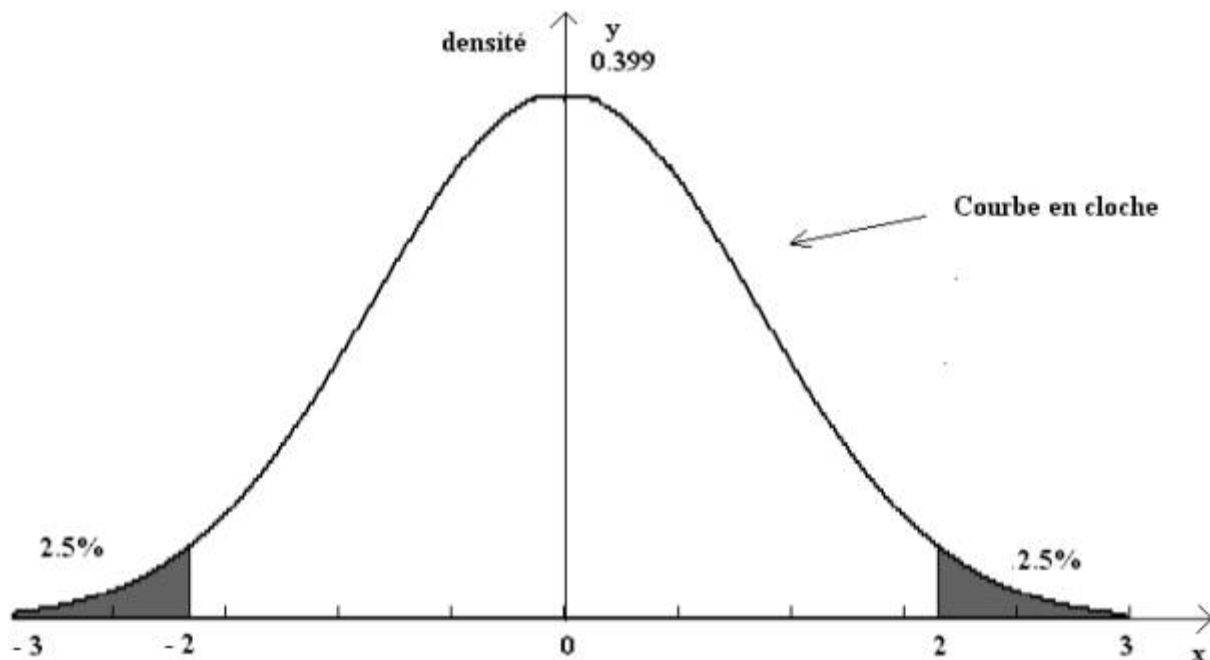


Figure 9.4 : densité de la loi normale de moyenne nulle et de variance un
(courbe en cloche)



Cette courbe de distribution représente la probabilité y de rencontrer une valeur x . Ainsi, elle est une représentation de la densité de probabilité d'une distribution de valeurs x . Elle peut être caractérisée par deux paramètres : un paramètre de position (la moyenne μ) et un paramètre de dispersion (l'écart-type σ). Il est donc important de noter que la connaissance de ces deux seuls paramètres est suffisante pour caractériser l'ensemble d'une distribution normale.

Cette distribution de valeurs $\mathcal{N}(\mu, \sigma)$ est absolument fondamentale dans la recherche biologique car elle est, de loin, la plus fréquemment rencontrée et il existe de nombreuses méthodes statistiques permettant d'estimer si la distribution des valeurs d'une expérience correspond à une telle distribution ou non.

On comprend que la conception d'une expérience repose sur la connaissance de la distribution des valeurs : suivent-elles une loi normale ? Quelles sont les valeurs de μ et σ ? Ces données ne peuvent être connues que grâce à des données bibliographiques, si elles existent ou grâce à la réalisation d'expériences préliminaires qui permettront de les estimer.

2. Les tests statistiques

Le plus souvent, on souhaitera déterminer si la moyenne des valeurs observées dans une situation est égale ou différente de la moyenne observée dans une autre condition. Typiquement, on aura des résultats d'animaux expérimentaux $[\mathcal{N}(\mu_1, \sigma_1)]$ à comparer aux résultats obtenus sur les animaux dits « contrôles » $[\mathcal{N}(\mu_0, \sigma_0)]$. Et pour comparer les moyennes μ_1 et μ_0 on utilisera un test statistique type ANOVA ou T-test. Le plus souvent nous n'aurons pas d'a priori sur le sens de la différence que nous souhaitons observer : nous ne sommes pas certain que les valeurs expérimentales seront plus grande ou plus petites par rapport aux contrôles, nous utiliserons donc d'un test bilatéral. Sinon, si nous sommes certains du sens de variation, nous utiliserons un test unilatéral.

La préoccupation majeure sera alors de ne pas conclure à l'existence d'un effet alors qu'en fait, il n'y en a pas (faux positifs : risque α ou erreur type I). Mais il existe un second risque dont il faut aussi tenir compte et qu'on oublie trop souvent : c'est le risque de conclure qu'il n'y a pas d'effet, alors qu'en fait il existe (faux négatifs : risque β ou erreur type II). Ce risque β est donc lié à la capacité du test à effectivement détecter une différence, il est donc lié à ce que l'on appellera la puissance du test (les tests unilatéraux sont plus puissants que les tests bilatéraux, mais ils reposent sur la contrainte forte de connaître avec certitude, à l'avance, le sens de la variation attendue).

Intuitivement, on sent bien que la puissance du test va être liée à la taille de l'effet que l'on souhaite mesurer, et que cette capacité à détecter un effet plus ou moins important va être lié à l'effectif de la cohorte d'animaux. Même sans entrer dans le calcul, on arrive donc à sentir qu'il existe un lien entre la puissance du test, la différence $|\mu_1 - \mu_0|$ que l'on souhaite détecter et l'effectif n de notre cohorte et que tout va être question de compromis.

$|\mu_1 - \mu_0|$ n'est pas une question strictement mathématique, c'est surtout ici une question scientifique, à savoir, quelle est la différence qui sera biologiquement significative. Il n'y a pas de règle, ce point est fixé



par l'expérimentateur, en fonction du contexte expérimental et de son expérience. C'est lui qui fixera ce seuil critique pour lequel la différence fait sens.

Le risque α c'est donc le risque que la différence que l'on observe soit un faux positif, ou en d'autres termes, qu'elle soit due au hasard. Le plus souvent, on fixe ce risque à 5%

Le risque β , c'est le risque de ne pas détecter une différence qui pourtant existe. Le plus souvent ce risque est donc ignoré. Ce qui signifie que l'on ne sait pas quelle chance on avait de détecter ce que l'on souhaitait observer. Fixons la puissance du test à 80%, ce qui est le standard généralement admis.

Ainsi risque α , β , seuil critique, et n sont liés et on sent bien que agir sur l'un, c'est modifier les autres. Il doit donc exister une fonction mathématique $F(\alpha, \beta, |\mu_1 - \mu_0|, n)$. Puisque l'on a fixé les trois premières variables, nous devrions pouvoir calculer n

3. quelques mots sur la taille de l'effet (d)

La taille de l'effet $|\mu_1 - \mu_0|$ va dépendre de la nature des valeurs étudiées et aura leur unité (cm, g, nombre de cellules/mm³, etc). Là encore, on sent bien que dans ce contexte il sera difficile de trouver une formule mathématique uniforme à tous les cas, il faut donc trouver un indicateur sans unité. Encore une fois, ce que l'on souhaite observer est une différence de moyenne μ ; le plus souvent la dispersion des valeurs autour de la moyenne sera considérée comme non modifiée et on aura $\sigma_0 = \sigma_1 = \sigma$. On pourra alors utiliser l'indicateur suivant :

$$d = \frac{|\mu_1 - \mu_0|}{\sigma}$$

4. Calcul de n

Considérant que nous avons des distributions normales, on admettra que $|\mu_1 - \mu_0| = Z\alpha \Sigma_0 + Z\beta \Sigma_1$ avec $Z\alpha$ et $Z\beta$ les valeurs x de N aux risques α et β et $\Sigma = \frac{\sigma}{\sqrt{n}}$

Nous avons fixé α (5%), dans ce cas, N nous donne une valeur $Z_\alpha = 1.96$.

Nous avons fixé β pour avoir une puissance de 0.8. Dans ce cas, N nous donne une valeur $Z_\beta = 0.845$.

Nous avons également considéré que les écart-types étaient égaux. Donc, $\Sigma_0 = \Sigma_1 = \Sigma = \frac{\sigma}{\sqrt{n}}$

Ainsi, dans le cas général d'un test sur des distributions normales $|\mu_1 - \mu_0| = Z\alpha \Sigma_0 + Z\beta \Sigma_1 \Leftrightarrow |\mu_1 - \mu_0|$

$$= \frac{\sigma}{\sqrt{n}} (Z\alpha + Z\beta) \Leftrightarrow n = \frac{\sigma^2}{|\mu_1 - \mu_0|^2} (Z\alpha + Z\beta)^2$$

Ce qui nous permet d'aboutir à la formule très simple : $n = \frac{7.86}{d^2}$



Puisque nous sommes sur un test bilatéral, il faudra encore multiplier le résultat par deux

$$n \approx \frac{16}{d^2}$$

Exemple d'applications

Vous étudiez la variable X dont la distribution suit une loi normale $N(\mu = 15, \sigma = 3)$. Vous appliquez un traitement T à vos animaux et vous pensez que ce traitement sera efficace s'il modifie la valeur de μ de 10%. Vous vous demandez combien d'animaux il vous faudra pour avoir 80% de chances de détecter une différence significative (si elle existe) à l'aide d'un test T de Student au risque de vous tromper de 5%

$$|\mu_1 - \mu_0| = 1.5 \text{ et donc } d = 1.5/3 = 0.5$$

En conséquence $n = 16/(0.5)^2 = 64$. Il vous faudra 64 animaux par groupe pour pouvoir détecter la différence escomptée.

Notez que si vous décidez que cette cohorte est trop importante, il est possible d'être moins ambitieux :

Si maintenant $|\mu_1 - \mu_0| = 3$ (20% de différence) alors $d = 1$ et $n = 16$. En diminuant votre sensibilité par deux, vous avez divisé la taille de votre cohorte par 4 !

Si maintenant vous êtes certain que votre variation sera une augmentation. Alors vous allez utiliser un test unilatéral. Dans ce cas $Z\alpha = 1.645$ et donc $(Z\alpha + Z\beta)^2 = 6.2$ et $n = 6.2/p^2$. Il vous faudra 25 animaux pour détecter une différence de 10% et 6 animaux pour détecter une différence de 20%

De manière intéressante, la formule peut également être utilisée pour calculer ce que l'on pourrait observer en considérant un nombre d'animaux fixé à l'avance :

Si l'on dispose maintenant de $n = 10$ animaux par groupe et que l'on souhaite faire un test bilatéral, alors

$$d = \sqrt{\frac{16}{10}} = 1.26 \text{ et } |\mu_1 - \mu_0| = 1.26 \times 3 = 3.78 \quad \Rightarrow \text{on sera en mesure de détecter une différence de l'ordre de 25\%}.$$



English version

Regulation of animal use in sciences (European directive 2010/63/UE, transposed in French decree 2013-118) enjoins scientists to estimate the appropriate number of animal they will need within protocol submission. Hence, item 3.3.5 “experimental strategy and statistics approaches to reduce to the minimum the number of animals, their anxiety and pain, and the impact on environment” of authorization project form is aimed to collect elements to evaluate whether the project is in line with the law.

Most of the time, submitted projects are aimed to demonstrate existence of effect through statistical tests; however, Evaluation of animal number needed to conclude appears poorly understood and is often relaying on empiricism and intuition, despite this evaluation is definitely central in experimental design. Here we propose a very simple method to estimate requested animal number to conclude; it will be applicable in most cases.

4. Normal distribution (N)

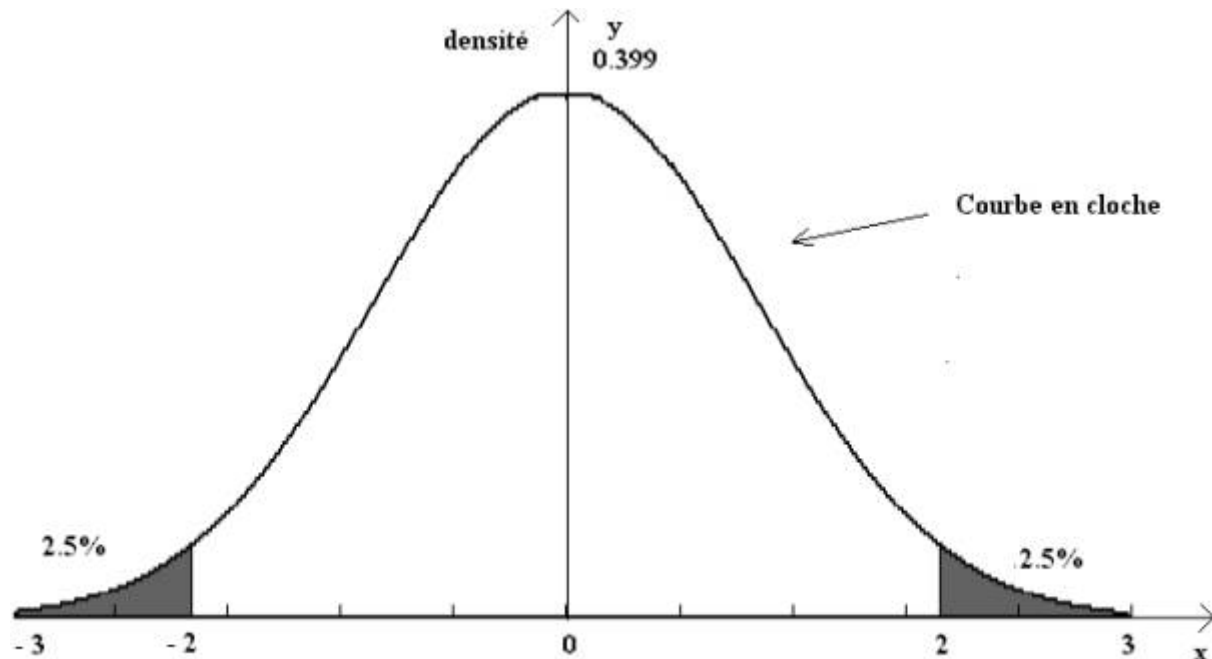


Figure 9.4 : densité de la loi normale de moyenne nulle et de variance un
(courbe en cloche)

This curve represents the probability y to encounter x value in a distribution. Therefore, it represents the probability density. This curve can be characterized by two parameters: position parameter (mean μ) and



dispersion parameter (standard deviation σ); it is crucial to notice that knowing these two parameters only is sufficient to define the whole normal distribution. This Normal distribution $N(\mu, \sigma)$ is absolutely central in biology because this is the distribution that will be encountered the most, by far; and it exist many methods to determine whether experimental value distributions feet with it, or not.

Now we understand that every experimental design needs to address whether value distributions correspond to a Normal distribution or not. What are μ and σ ? It is essential to know them, either from literature, previous experiments or preliminary experiment. Whatever, but you have to estimate them!

5. Statistic tests

Most of the time studies are aimed to compare means, and to determine whether it is change in one situation compared to another one. Usually, results obtained from one experimental group $[N(\mu_1, \sigma_1)]$ will be compared to results obtained from one "control" group $[N(\mu_0, \sigma_0)]$, and to make such comparison, we will use statistical tests such as ANOVA or T-test (for which it is mandatory to have a Normal distribution of values in each group). Most of the time, we won't assume we are sure of the orientation of the expected difference (increase or decrease): we will use bilateral (two-tailed) tests; otherwise we will use unilateral (one-tailed) tests.

Once the test chosen, One major concern will be to not fall into an inappropriate conclusion, and very often people will fear to conclude that there is an effect where there is no effect, actually (false positive, α or type I error). But it exists a second under considered risk to handle: to conclude that there is no effect where there is an effect (false negative, β or type II error). This β risk is linked to test's capability to detect a difference, it is therefore linked to so called test power (usually, one tailed tests are more powerful that two-tailed tests, but with major constraint that you must know orientation of variation before the experiment).

We can fell intuitively that power and expected size effect are linked, and that capability to detect more or less important effect will be linked to cohort size. Without even calculation, we can assume that it exists a link between power, $|\mu_1 - \mu_0|$, and number n of animals within our cohort, and that everything will be matter of compromises.

$|\mu_1 - \mu_0|$ parameter is not just mathematic, it is mostly a scientific concern: which difference will make sense at biological level? There is no rule there, this point is fixed by experimenter and will depend of experimental context (goals of the study) and knowledge.

α risk is the risk of making a false positive conclusion, you will accept to take. Most of the time, it is fixed to 5%

β risk is the risk to not detect an existing difference. Most of the time it is ignored, which means peoples do not know how many chance they actually had to observe what they expected. Let's fix power to 80% (chance to detect a difference, whether it exists), which is the admitted standard.



Hence, α and β risks, critical size effect, and n are linked, and we can feel that touching one will impact the others ; thus, it should exist a mathematical function $F(\alpha, \beta, |\mu_1 - \mu_0|, n)$, and since we have fixed all three first parameters, we should be able to calculate n .

6. Few words on size effect (d)

Effect size will depend on what values are analyzed and their unity (cm, g, number of cells/mm³, and so on). Again, it is easy to realize that because every value is different, it will be difficult to find a unique mathematical function, we must find a reporter with no dimension. We want to compare means, and most of the time standard deviation will not be modified, thus: $\sigma_0 = \sigma_1 = \sigma$ and will be able to use this reporter:

$$d = \frac{|\mu_1 - \mu_0|}{\sigma}$$

5. Cohort size (n) estimation

Considering we have Normal distributions, we will admit here that $|\mu_1 - \mu_0| = Z_\alpha \Sigma_0 + Z_\beta \Sigma_1$ with Z_α and Z_β values of N to the α and β risks and $\Sigma = \frac{\sigma}{\sqrt{n}}$

α was fixed to 5% this correspond to $Z_\alpha = 1.96$.

β was fixed to 20% to reach power = 0.8 (power = 1- β). This correspond to $Z_\beta = 0.845$.

Since we have considered standard deviation (σ) as equal, then $\Sigma_0 = \Sigma_1 = \Sigma = \frac{\sigma}{\sqrt{n}}$

Hence, for a test on Normal distributions: $|\mu_1 - \mu_0| = Z_\alpha \Sigma_0 + Z_\beta \Sigma_1 \Leftrightarrow |\mu_1 - \mu_0|$

$$= \frac{\sigma}{\sqrt{n}} (Z_\alpha + Z_\beta) \Leftrightarrow n = \frac{\sigma^2}{|\mu_1 - \mu_0|^2} (Z_\alpha + Z_\beta)^2$$

After few simplifications, we have this very simple formula for one tailed tests: $n = \frac{7.86}{d^2}$

For two-tailed tests, we just have to perform a 2x multiplication.

$$n \approx \frac{16}{d^2}$$



Examples of applications

We'd like to analyze the X variable that is distributed along the Normal distribution $N(\mu = 15, \sigma = 3)$. The objective is to test a treatment T and we will consider this treatment as efficient if it modifies μ of 10%, at least. The question is how many animals will we need to have 80% of chances to detect such significant difference (whether it exists) using a T-test with 5% of chances to accept a false positive conclusion?

$$|\mu_1 - \mu_0| = 1.5 \text{ and then, } d = 1.5/3 = 0.5$$

As immediate consequence, $n = 16/(0.5)^2 = 64$. We will need 64 animals per group to be able to detect the expected effect.

Note that if this cohort size is too important, it is possible to be less ambitious:

If now $|\mu_1 - \mu_0| = 3$ (20% difference) then $d = 1$ and $n = 16$. Very important conclusion there: a 2 fold decrease in sensibility corresponds to a 4 fold decrease in cohort size!

If now you are certain that your variation will be an increase, then you are allow to use a one tailed test. In this case, $Z\alpha = 1.645$ and then $(Z\alpha + Z\beta)^2 = 6.2$ and $n = 6.2/p^2$. You will need 25 animals to detect a 10% difference, and 6 animals to detect a 20% difference.

Importantly, it is noteworthy to notice that the formula can be used to calculate what it possible to detect considering a fixed cohort size:

If $n = 10$ animals per group and that we plan to use a two-tailed test, then $d = \sqrt{\frac{16}{10}} = 1.26$ et

$$|\mu_1 - \mu_0| = 1.26 \times 3 = 3.78 \Rightarrow \text{we will be able to detect a 25\% difference.}$$