

Comment peut-on faire un calcul de puissance *a priori* ?

M. L. Delignette-Muller
VetAgro Sup

14 novembre 2018



Objectifs pédagogiques

- Savoir définir ce qu'on appelle un calcul de puissance *a priori*
- Comprendre sur quelles bases il est possible de faire un calcul de puissance *a priori*.
- Savoir réaliser un calcul de puissance à l'aide du logiciel **R** dans les trois cas classiques (comparaison de deux moyennes sur séries indépendantes, comparaison de deux moyennes sur séries appariées et comparaison de deux fréquences. **

** *savoir faire évalué uniquement en S6 après entraînement en TD*

Exemple introductif

Exemple du TD 1 : une étude est réalisée sur la prévention des strongyloses gastro-intestinales chez les bovins allaitants de race charolaise. On estime le poids moyen des animaux de 9 mois (variable que l'on notera PM) sur divers élevages allaitants que l'on peut séparer en deux groupes suivant qu'ils respectent ou non le protocole de traitement préventif (traitement correct ou incorrect).

Supposons que la distribution de la variable PM soit normale dans chacun des deux groupes d'élevage et d'écart type commun connu (hypothèse nécessaire pour réaliser facilement un calcul de puissance) $\sigma = 5$ kg.

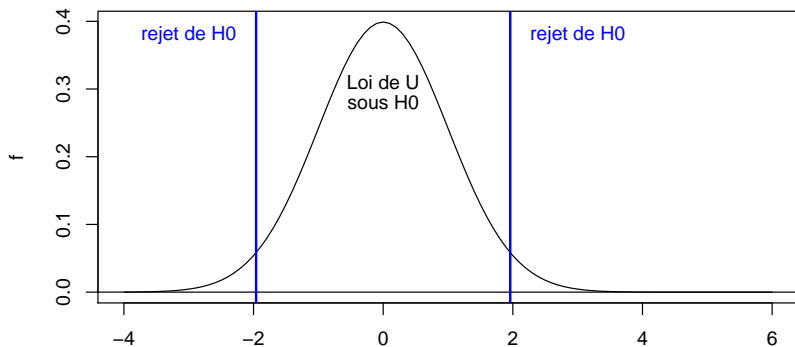
Imaginons que l'on dispose de deux groupes comprenant chacun 10 élevage ($n = 10$). Quelle est la probabilité de détecter une différence si celle-ci est égale à 7 ($\mu_1 - \mu_2 = 7$ kg) ?

Test et zone de rejet de H_0

Sous les conditions envisagées la variable $U = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \times \sqrt{\frac{2}{n}}}$ suit la loi

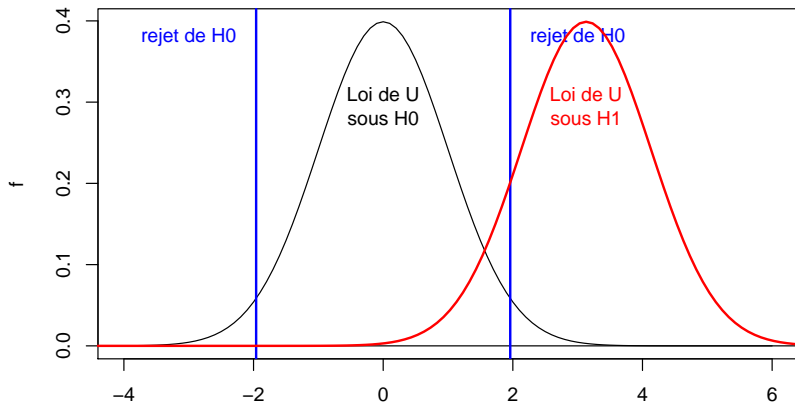
$N\left(\frac{\mu_1 - \mu_2}{\sigma \times \sqrt{\frac{2}{n}}}, 1\right)$ (résultat de statistique théorique).

On peut classiquement utiliser cette variable pour faire un test.



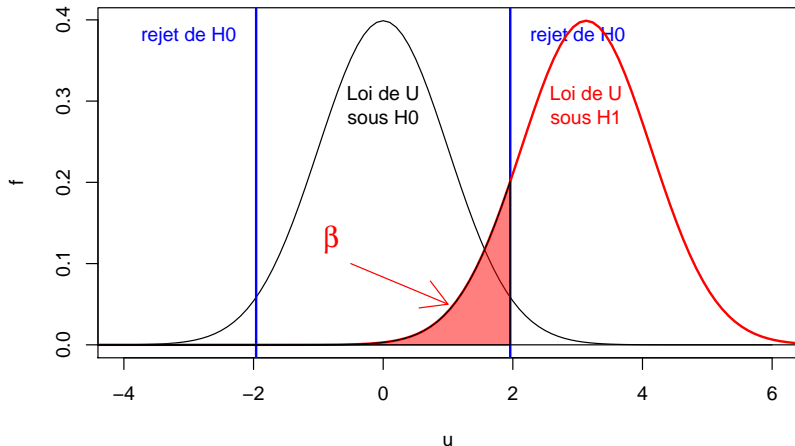
Distribution de U sous H_1

Distribution de U si $\mu_1 - \mu_2 = 7$ kg : loi normale d'écart type 1
 mais centrée en $\frac{\mu_1 - \mu_2}{\sigma \times \sqrt{\frac{2}{n}}} = \frac{7}{5 \times \sqrt{\frac{2}{10}}} = 3.13$



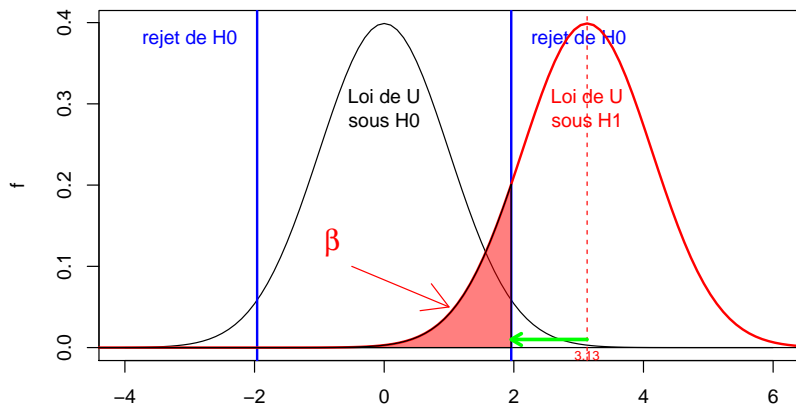
Visualisation du risque β

β = probabilité d'accepter H_0 si on est sous H_1



Calcul du risque β

β valeur de la fonction de répartition de la loi $N(0, 1)$ associée à un quantile de $1.96 - 3.13 = -1.17 \rightarrow$ environ 12% (correspondant à une puissance de 88%).



Calcul de l'effectif nécessaire pour atteindre une puissance donnée

Si on fixe la puissance (par ex. à 95%) et donc le risque β (à 5%) on calcule l'effectif n tel que

$1.96 - \frac{\mu_1 - \mu_2}{\sigma \times \sqrt{\frac{2}{n}}}$ corresponde au quantile à 5% de la loi $N(0, 1)$

(u_β égal à -1.645 pour $\beta = 5\%$).

C'est-à-dire l'effectif n tel que $1.96 - \frac{\mu_1 - \mu_2}{\sigma \times \sqrt{\frac{2}{n}}} = u_\beta$

Soit $n = \left(\frac{(1.96 - u_\beta) \times \sigma \sqrt{2}}{\mu_1 - \mu_2} \right)^2$ (ici $n > 13$ élevages par groupe).

On note que ce calcul d'effectif nécessite

- de fixer a priori la puissance $(1 - \beta)$,
- d'avoir une idée de la variabilité au sein de chaque groupe (σ)
- et de définir la différence que l'on souhaite se donner les moyens de mettre en évidence $\mu_1 - \mu_2$.

Retour sur l'exemple introductif

Exemple du TD 1 : une étude est réalisée sur la prévention des strongyloses gastro-intestinales chez les bovins allaitants de race charolaise. On estime le poids moyen des animaux de 9 mois (variable que l'on notera PM) sur divers élevages allaitants que l'on peut séparer en deux groupes suivant qu'ils respectent ou non le protocole de traitement préventif (traitement correct ou incorrect).

Supposons que la distribution de la variable PM soit normale dans chacun des deux groupes d'élevage et d'écart type commun connu (hypothèse nécessaire pour réaliser facilement un calcul de puissance) $\sigma = 5$ kg.

Combien doit-on échantillonner d'élevages dans chaque groupe pour avoir une probabilité d'au moins 90% de détecter une différence entre les moyennes si celle-ci est égale à 2 kg ($\mu_1 - \mu_2 = 2$ kg) ?

Calcul d'effectifs nécessaires à l'aide du logiciel R

$\mu_1 - \mu_2 = 2$, $\sigma = 5$ et $1 - \beta = 0.9$ ($\rightarrow n > 132$ élevages par groupe)

```
> power.t.test(n = NULL, delta = 2, sd = 5,
+             sig.level = 0.05, power = 0.9,
+             type = "two.sample")
```

Two-sample t test power calculation

```
      n = 132
  delta = 2
     sd = 5
sig.level = 0.05
  power = 0.9
alternative = two.sided
```

NOTE: n is number in *each* group

Calculs d'effectifs nécessaires dans les trois cas classiques

Il sera très facile de faire des calculs d'effectifs nécessaires avec le logiciel **R** à partir d'une **puissance fixée** et une **différence que l'on souhaite se donner les moyens de mettre en évidence prédéfinie** si l'on a une idée

- de l'écart **type au sein des groupes** pour comparer **deux moyennes sur deux séries indépendantes**,
- de l'écart **type des différences** pour comparer **deux moyennes sur deux séries appariées**,
- et de la valeur de **chacune des deux fréquences attendues** pour comparer **deux fréquences sur des séries indépendantes**.

Sans ces informations on ne peut pas faire de calcul d'effectifs nécessaires *a priori*