Puissance et taille d'échantillon

Quels sont les points essentiels à considérer ?

Elise Dupuis Lozeron - Angèle Gayet-Ageron

Service d'Epidémiologie clinique
Unité d'Appui méthodologique - CRC

25 février 2019

Concept de puissance statistique

- Le concept de puissance vient de la théorie des tests statistiques
- La puissance d'une étude est en fait **la puissance du test statistique** effectué sur **l'outcome principal**
- C'est un **calcul théorique**, réalisé avec des valeurs fixées avant la réalisation de l'étude et sous les conditions d'application du test statistique choisi

Rappel sur les tests statistiques

- Le but d'un test statistique est de tester une **hypothèse nulle** H_0
- L'hypothèse nulle concerne un paramètre de la *population*
- L'**hypothèse nulle** H_0 est celle que l'on veut rejeter. Elle exprime souvent l'absence d'effet. Par exemple : l'effet du traitement A est le même que l'effet du traitement B
- L'**hypothèse alternative** H_a est celle que l'on aimerait accepter. Par exemple : l'effet du traitement A est différent de l'effet du traitement B

Rappel sur les tests statistiques

- Deux types d'erreurs :
  - L'erreur de type I : rejeter H_0 quand elle est vraie. Faux positif. La probabilité de commettre cette erreur est égale à α
  - L'erreur de type II : ne pas rejeter H_0 quand elle est fausse, c'est à dire quand H_a est vraie. Faux négatif. La probabilité de commettre cette erreur est égale à β

Rappel sur les tests statistiques

Décision	H_0 vraie	H_a vraie
Rejeter H_0	Erreur type I (α)	1- β
Ne pas rejeter H_0	1-α	Erreur type II (β)

- La **puissance** d'un test statistique est égale à 1- β. C'est la probabilité de rejeter H_0 quand elle est fausse (et que H_a est vraie)

Exemple

- Essai randomisé contrôlé en double aveugle pour comparer la prise de corticostéroïdes inhalés avec un placebo chez des enfants d'âge scolaire.
- Outcome primaire: VEMS moyen évalué à la fin du follow-up
- Le VEMS des deux groupes est comparé au moyen d'un test t
- H_0 : le VEMS moyen de la population des enfants d'âge scolaire est le même dans les deux groupes
- H_a : le VEMS moyen de la population des enfants d'âge scolaire n'est pas le même dans les deux groupes

(Exemple tiré de Medical statistics at a glance, Petrie and Sabin.)

Exemple

- Dans le groupe traité avec des corticostéroïdes $n_1 = 50$, le VEMS moyen $\bar{x}_1 = 1.64\mathrm{l}$, sd $s_1 = 0.29\mathrm{l}$
- Dans le groupe placebo $n_2 = 48$, le VEMS moyen $\bar{x}_2 = 1.54\mathrm{l}$, sd $s_2 = 0.25\mathrm{l}$
- La différence estimée entre les moyennes des deux groupes est égale à $1.64 - 1.54 = 0.10\mathrm{l}$.
- La statistique de test t vaut:

$$t = \frac{1.64 - 1.54}{\sqrt{\frac{0.29^2}{50} + \frac{0.25^2}{48}}} = 1.831$$

- La p-valeur vaut $p = 0.067$

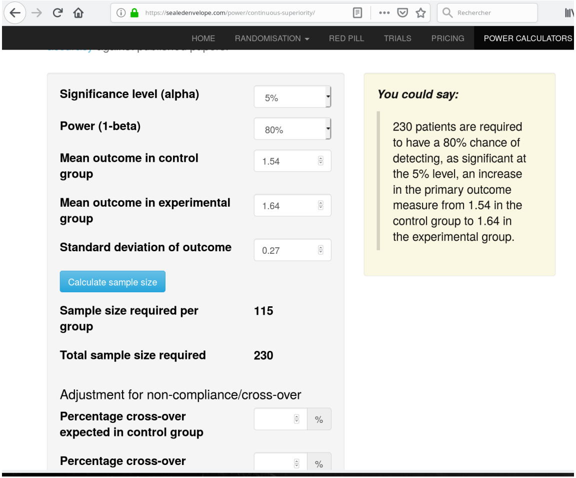
Exemple

- La statistique de test et la p-valeur quantifient l'évidence contre H_0
- Dans cet exemple il n'y a pas suffisamment d'évidence en faveur du traitement
- Mais attention: "absence of evidence is not evidence of absence"
- Il se peut qu'en fait une vraie différence existe mais que la puissance de l'étude soit insuffisante

Exemple

- Si une différence de 0.10l entre les deux groupes est jugée cliniquement importante, combien d'enfants aurait-il fallu inclure pour avoir une puissance suffisante ?
- https://sealedenvelope.com/

Exemple


https://sealedenvelope.com/power/continuous-superiority/
Rechercher
HOME
RANDOMISATION
RED PILL
TRIALS
PRICING
POWER CALCULATORS
Significance level (alpha)
5%
Power (1-beta)
80%
Mean outcome in control group
1.54
Mean outcome in experimental group
1.64
Standard deviation of outcome
0.27
Calculate sample size
Sample size required per group
115
Total sample size required
230
Adjustment for non-compliance/cross-over
Percentage cross-over expected in control group
%
Percentage cross-over
%
You could say:
230 patients are required to have a 80% chance of detecting, as significant at the 5% level, an increase in the primary outcome measure from 1.54 in the control group to 1.64 in the experimental group.

Exemple

- Il y a un certain nombre de facteurs qui influencent la puissance d'un test statistique:
  - La taille de l'échantillon
  - La différence à détecter
  - Le niveau de significativité (α)
  - La variabilité des observations

Facteurs influençant la puissance

- http://www.statstudio.net/free-tools/power-analysis/
- Une valeur de puissance fixée détermine une taille d'échantillon donnée et inversément

Perte de suivi

- Il arrive très souvent que des patients soient perdus de vu avant l'évaluation de l'outcome primaire
- Il est donc courant d'anticiper cette perte de suivi et d'augmenter la taille d'échantillon calculée par un facteur égal à $1/(1 - P_{loss})$, P_{loss}, étant la proportion hypothétique de patients perdus de vu
- La puissance statistique pour une taille d'échantillon projetée est aussi réduite en cas de non-observance ("non-compliance"), c'est à dire quand une partie des patients assignés au groupe traitement arrête le traitement ou quand une partie des patients assignés au groupe contrôle commence le traitement

Comparaison de 2 proportions

- Formule pour calculer la taille d'échantillon pour une étude comparant deux proportions :

$$N = \frac{2 \times \bar{p}(1-\bar{p})(z_{\alpha/2} + z_{\beta})^2}{(p_1 - p_2)^2}$$

- p_1 est la proportion de patients pour lesquels l'outcome vaut 1 dans le groupe 1
- p_2 est la proportion de patients pour lesquels l'outcome vaut 1 dans le groupe 2
- $\bar{p} = \frac{p_1+p_2}{2}$ est la proportion moyenne de patients pour lesquels l'outcome vaut 1 dans les 2 groupes
- $z_{\alpha/2}$ et z_{β} sont des constantes déterminées par α et β

Comparaison de 2 proportions

$$N = \frac{2 \times \bar{p}(1-\bar{p})(z_{\alpha/2} + z_{\beta})^2}{(p_1 - p_2)^2}$$

- N augmente quand $(p_1 - p_2)$ diminue
- N augmente quand p_1 (ou p_2) est proche de 0.5

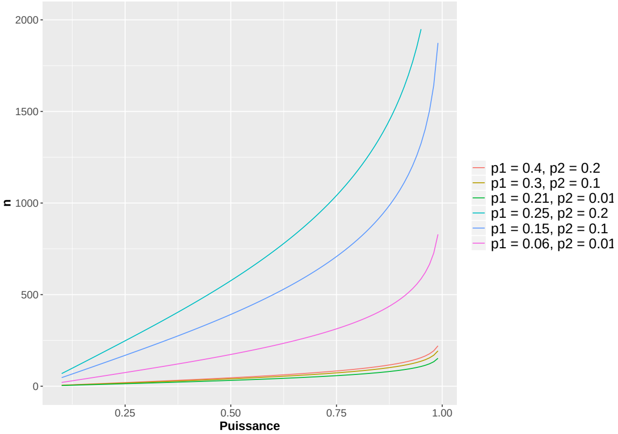
Comparaison de 2 proportions

$$N = \frac{2 \times \bar{p}(1-\bar{p})(z_{\alpha/2}+z_{\beta})^2}{(p_1-p_2)^2}$$

- $z_{\alpha/2}$ et z_{β} sont des constantes déterminées par α et β
- Valeurs de $(z_{\alpha/2}+z_{\beta})^2$ en fonction de α et β

1-β	70%	80%	90%	95%	99%
α=5%	6.2	7.8	10.5	12.9	18.4
α=1%	9.6	11.7	14.9	17.8	24.1

Comparaison de 2 proportions



- Le choix de la valeur de base est aussi important que le choix de la différence à détecter
- Attention n augmente de manière exponentielle avec la puissance et non de manière linéaire

Exemple

- Effet de l'aspirine comparée au placebo sur la diminution de la proportion d'évènements cardio-vasculaires
- $$N = \frac{2 \times \bar{p}(1-\bar{p})(z_{\alpha/2} + z_{\beta})^2}{(p_1 - p_2)^2}$$
- Avec $p_1 = 0.45$ et $p_2 = 0.25$, soit une différence à détecter de 20% et une puissance de 80%, n = 89, soit 178 patients au total
- Avec $p_1 = 0.25$ et $p_2 = 0.05$, soit une différence à détecter de 20% et une puissance de 80%, n = 50 par bras, soit 100 patients au total

Comparaison de deux moyennes

- Formules pour calculer la taille d'échantillon pour une étude comparant deux moyennes

$$N = \frac{2 \times \sigma^2 (z_{\alpha/2} + z_\beta)^2}{(\bar{x}_1 - \bar{x}_2)^2}$$

- $\bar{x}_1$ est l'outcome moyen dans le groupe 1
- $\bar{x}_2$ est l'outcome moyen dans le groupe 2
- σ est l'écart-type de l'outcome
- $z_{\alpha/2}$ and z_β sont des constantes déterminées par α et β

Comparaison de deux moyennes

$$N = \frac{2 \times \sigma^2 (z_{\alpha/2} + z_\beta)^2}{(\bar{x}_1 - \bar{x}_2)^2}$$

- La taille d'échantillon augmente avec σ
- La taille d'échantillon augmente avec $(\bar{x}_1 - \bar{x}_2)$

Cas particulier des données continues pairées

- Par définition, il n'y a qu'un seul groupe
- L'écart-type de la différence dépend aussi de la corrélation entre les deux mesures:

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2 * \mathrm{Cov}(X; Y)$$

- Si la corrélation est positive, alors $\mathrm{Cov}(X; Y)$ est positive et la $\mathrm{Var}(X - Y)$ est plus petite que si les données n'étaient pas pairées
- Conséquence : lorsque les données sont pairées le nombre nécessaire de sujets est plus petit !

Marche à suivre du calcul de la taille d'échantion (RCT de supériorité)

1. Définir l'outcome primaire, l'hypothèse à tester et le test statistique utilisé pour comparer les interventions
2. Déterminer la différence à détecter, la valeur de base (valeur de l'outcome dans le bras contrôle) et la variabilité des observations
3. Choisir α et $1 - \beta$
4. Calculer la taille d'échantillon (formules, tables, logiciels, etc. . .)

Considérations générales concernant le calcul de la taille d'échantillon

- La puissance, et donc la taille d'échantillon nécessaire pour atteindre une telle puissance, est liée à un **test statistique**
- Ce test statistique est celui qui est effectué sur **l'outcome principal** de l'étude
- Pour un calcul de taille d'échantillon basé sur la puissance d'un test statistique il faut toujours connaître la valeur postulée sous H_0 et sous H_a (différence à détecter) ainsi que la variabilité des observations

Considérations générales concernant le calcul de la taille d'échantillon

- On peut aussi parfois calculer la taille d'échantillon nécessaire pour obtenir une certaine **largeur de l'intervalle de confiance** autour de l'estimation du paramètre d'intérêt (ex : sensibilité)
- Si l'analyse principale de l'étude est un **modèle statistique** (régression logistique, modèle de Cox. . .) la puissance est celle du test de significativité de l'intervention/exposition
- Pour certains modèles, on peut calculer la taille d'échantillon nécessaire pour **ajuster le modèle sur un nombre défini de variables**
- Pour des designs complexes, il peut être nécessaire de faire des **simulations**