

- Résumé
- 1 Qu'est-ce qu'est la puissance statistique?
- 2 Pourquoi vous en avez besoin ?
- 3 Les trois ingrédients de la puissance statistique
- 4 Formules clés pour calculer la puissance statistique
- 5 Quand croire votre analyse de puissance ?
- 6 Comment utiliser la simulation pour estimer la puissance ?
- 7 Comment changer votre design pour améliorer votre puissance ?
- 8 Analyse de puissance pour plusieurs traitements
- 9 Comment penser la puissance pour les designs par grappe (cluster) ?
- 10 Une bonne analyse de puissance facilite le pré-enregistrement

## Résumé

Ce guide<sup>1</sup> vous aidera à évaluer et à améliorer la puissance statistique de vos expériences. Nous nous concentrons sur les grandes idées et fournissons des exemples et des outils que vous pouvez utiliser dans R et Google Spreadsheets.

## 1 Qu'est-ce qu'est la puissance statistique?

La puissance est la capacité de distinguer le signal du bruit.

Le signal qui nous intéresse est l'impact d'un traitement sur certains résultats. L'éducation augmente-t-elle les revenus ? Les campagnes de santé publique diminuent-elles l'incidence des maladies ? Les observateurs internationaux peuvent-ils réduire la corruption gouvernementale ?

Le bruit qui nous préoccupe provient de la complexité du monde. Les résultats varient selon les personnes et les lieux pour une myriade de raisons. En termes statistiques, vous pouvez considérer cette variation comme l'écart type de la variable de résultat. Par exemple, supposons qu'une expérience utilise les taux d'une maladie rare comme résultat. Le nombre total de personnes touchées ne fluctue probablement pas énormément d'un jour à l'autre, ce qui signifie que le bruit de fond dans cet environnement sera faible. Lorsque le bruit est faible, les expériences peuvent détecter même de petits changements dans les résultats moyens. Un traitement qui réduirait l'incidence de la maladie de 1% serait facilement détecté, car les taux de base sont constants.

Supposons maintenant qu'une expérience utilise plutôt le revenu des sujets comme variable de résultat. Les revenus peuvent varier assez considérablement - dans certains endroits, il n'est pas rare que les gens aient des voisins qui gagnent deux, dix ou cent fois leur salaire quotidien. Lorsque le bruit est élevé, les expériences sont plus difficiles. Un traitement qui augmenterait les revenus des travailleurs de 1% serait difficile à détecter, car les revenus diffèrent tellement en premier lieu.

Une préoccupation majeure avant de se lancer dans une expérience est le danger d'un **faux négatif**. Supposons que le traitement ait réellement un impact causal sur les résultats. Il serait dommage de se donner autant de peine et de dépenser autant pour randomiser le traitement, collecter des données sur les groupes de traitement et de contrôle et analyser les résultats, pour que l'effet soit finalement noyé par le bruit de fond.

Si nos expériences ont une grande puissance statistique, nous pouvons être sûrs de ceci : s'il y a vraiment un effet de traitement, nous pourrions le voir.

## 2 Pourquoi vous en avez besoin ?

Les expérimentateurs se prémunissent souvent contre les faux positifs avec des tests de signification statistique. Après une expérience, nous craignons de conclure à tort qu'il y a un effet alors qu'il n'y en a pas vraiment.

L'analyse de puissance statistique pose la question inverse : en supposant qu'il y ait vraiment un effet de traitement et que vous deviez exécuter votre expérience un grand nombre de fois, à quelle fréquence obtiendrez-vous un résultat statistiquement significatif ?

Répondre à cette question nécessite des conjectures éclairées. Vous devrez fournir des suppositions quant à l'ampleur raisonnable de l'effet de votre traitement, combien de sujets répondront à votre enquête, combien de sujets votre organisation peut raisonnablement se permettre de traiter.

D'où viennent ces suppositions ? Avant qu'une expérience ne soit menée, il existe souvent une multitude de données de base disponibles. Les sujets seront-ils âgés, riches, éduqués ? Quelle a été l'ampleur du plus grand effet de traitement jamais établi pour votre variable dépendante ? Avec l'analyse de puissance statistique, vous pouvez voir à quel point la probabilité d'obtenir

des résultats significatifs est sensible aux changements de vos hypothèses.

De nombreuses disciplines se sont fixées sur une valeur de puissance cible de 0,80. Les chercheurs modifieront leurs designs et leurs hypothèses jusqu'à ce qu'ils puissent être sûrs que leurs expériences donneront des résultats statistiquement significatifs 80 % du temps. Bien que cette convention soit une référence utile, assurez-vous que vous êtes à l'aise avec les risques associés à un taux de réussite attendu de 80 %.

Attention : la puissance statistique compte beaucoup. Les résultats nuls d'études de faible puissance peuvent être difficiles à interpréter : n'y a-t-il vraiment aucun effet ? Ou l'étude n'est-elle tout simplement pas en mesure de le mesurer ? Les résultats positifs d'une étude de faible puissance peuvent également être trompeurs : à condition d'être statistiquement significative, une estimation d'une étude de faible puissance surestime probablement l'effet du traitement. Les études de faible puissance sont parfois basées sur des hypothèses trop optimistes ; une analyse de puissance convaincante rend ces hypothèses explicites et devrait vous protéger de la mise en œuvre de designs qui n'ont de manière réaliste aucune chance de répondre aux questions auxquelles vous souhaitez répondre.

## 3 Les trois ingrédients de la puissance statistique

Il existe trois grandes catégories de facteurs qui déterminent la puissance de votre expérience. Les deux premiers (l'efficacité du traitement et le bruit de fond) sont des choses que vous ne pouvez pas vraiment contrôler – ce sont les réalités de votre environnement expérimental. Le dernier facteur, le design expérimental, est la seule chose que vous pouvez contrôler – utilisez-le !

- **Efficacité du traitement.** Au fur et à mesure que l'efficacité de votre traitement augmente, la puissance statistique de votre expérience augmente. Cela a du sens : si votre traitement donnait à chaque sujet 1 000 000 \$, il ne fait aucun doute que nous pourrions discerner des différences de comportement entre les groupes de traitement et de contrôle. Souvent, cependant, nous ne contrôlons pas l'efficacité de nos traitements. Par exemple, les chercheurs impliqués dans l'évaluation du programme ne décident pas du traitement, ils sont censés évaluer le programme tel qu'il est.
- **Bruit de fond.** À mesure que le bruit de fond de vos variables de résultat augmente, la puissance de votre expérience diminue. Dans la mesure du possible, essayez de sélectionner des variables de résultat qui ont une faible variabilité. En termes pratiques, cela signifie comparer l'écart type de la variable de résultat à la taille de l'effet de traitement attendu - il n'y a pas de rapport magique à viser, mais plus les deux sont proches, mieux sera votre expérience. Dans l'ensemble, les chercheurs ne contrôlent pas le bruit de fond, et choisir des variables de résultat à bruit faible est plus facile à dire qu'à faire. De plus, de nombreux résultats que nous aimerions étudier sont intrinsèquement assez variables. De ce point de vue, le bruit est simplement à gérer du mieux que vous pouvez.
- **Design expérimental.** L'analyse de puissance statistique traditionnelle se concentre sur un élément (bien que très important) du design expérimental : le nombre de sujets dans chaque groupe expérimental. En termes simples, un plus grand nombre de sujets augmente la puissance. Cependant, il existe d'autres éléments du design expérimental qui peuvent augmenter la puissance : comment la randomisation est-elle menée ? Est-ce que d'autres facteurs seront contrôlés statistiquement ? Combien de groupes de traitement y aura-t-il et peuvent-ils être combinés pour certaines analyses ?

## 4 Formules clés pour calculer la puissance statistique

Les statisticiens ont dérivé des formules pour calculer la puissance de nombreux modèles expérimentaux. Elles peuvent être utiles pour calculer approximativement la taille de l'échantillon dont vous aurez besoin. Soyez prudent, cependant, car les hypothèses qui sous-tendent les formules peuvent parfois être obscures, ou pire encore, elles peuvent être fausses.

Voici une formule courante utilisée pour calculer la puissance<sup>2</sup>

$$\beta = \Phi \left( \frac{|\mu_t - \mu_c| \sqrt{N}}{2\sigma} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right)$$

- $\beta$  est notre mesure de puissance statistique. Parce que c'est la probabilité d'obtenir un résultat statistiquement significatif,  $\beta$  sera un nombre compris entre 0 et 1.
- $\Phi$  est la fonction de répartition de la distribution normale, et  $\Phi^{-1}$  est son inverse.
- $\mu_t$  est le résultat moyen dans le groupe de traitement. Supposons qu'il soit de 65.
- $\mu_c$  est le résultat moyen dans le groupe de contrôle. Supposons qu'il soit de 60.
- Ensemble, les hypothèses sur  $\mu_t$  et  $\mu_c$  définissent notre hypothèse sur la taille de l'effet du traitement :  $65-60=5$ .

- $\sigma$  est l'écart type des résultats. C'est ainsi que nous émettons des hypothèses sur le niveau de bruit de notre expérience - l'une des hypothèses que nous émettons est que  $\sigma$  est le même pour les groupes de traitement et de contrôle. Supposons  $\sigma = 20$
- $\alpha$  est notre niveau de signification – la convention dans de nombreuses disciplines est que  $\alpha$  doit être égal à 0,05.  $N$  est le nombre total de sujets. C'est la seule variable qui est sous le contrôle direct du chercheur. Cette formule suppose que chaque sujet ait 50 % de chances d'être dans le groupe de contrôle. Supposons que  $N = 500$ .

En utilisant la formule, nous trouvons que sous cet ensemble d'hypothèses,  $\beta = 0,80$ , ce qui signifie que nous avons 80 % de chances d'avoir un résultat statistiquement significatif avec ce design. Cliquez ici pour une Google spreadsheet (<https://docs.google.com/spreadsheets/d/117R4cqKkhX1MFqPIh7Yg2YzjHykxD7WsSLXqhEbD33I/edit#gid=0>) qui inclut cette formule. Vous pouvez copier ces formules directement dans Excel. Si vous êtes à l'aise avec R, voici le code pour le même calcul.

```
power_calculator <- function(mu_t, mu_c, sigma, alpha=0.05, N){
  lowertail <- (abs(mu_t - mu_c)*sqrt(N))/(2*sigma)
  uppertail <- -1*lowertail
  beta <- pnorm(lowertail- qnorm(1-alpha/2), lower.tail=TRUE) + 1- pnorm(uppertail- qnorm(1-alpha/2), lower.tail=FALSE)
  return(beta)
}
```

## 5 Quand croire votre analyse de puissance ?

D'un certain point de vue, l'idée d'analyser la puissance statistique n'a aucun sens. Vous voulez déterminer l'ampleur d'un effet de traitement, mais vous devez d'abord effectuer une analyse de puissance qui nécessite que vous connaissiez déjà l'effet de votre traitement et bien plus encore.

Ainsi, dans la plupart des analyses de puissance, vous voyez en fait ce qui se passe avec des nombres qui sont dans une certaine mesure inventés. La bonne nouvelle, c'est qu'il est facile de découvrir à quel point vos conclusions dépendent de vos hypothèses : faites simplement varier vos hypothèses et voyez comment les conclusions sur la puissance statistique varient.

Cela se voit plus facilement en pensant à la façon dont la puissance varie avec le nombre de sujets. Une analyse de puissance qui examine la puissance pour différentes tailles d'étude inclue simplement une plage de valeurs pour  $N$  et voit comment  $\beta$  change.

En utilisant la formule de la section 4, vous pouvez voir à quel point la puissance est sensible à toutes les hypothèses : la puissance sera plus élevée si vous supposez que l'effet du traitement est plus important, ou si vous êtes prêt à accepter un niveau  $\alpha$  plus élevé, ou si vous avez plus ou moins confiance dans le bruit de vos mesures.<sup>3</sup>

## Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

☐ Clustered Design?

☐ Binary Dependent Variable?

Significance Level

Alpha = 0.05 ▼

Treatment Effect Size

Standard Deviation of Outcome Variable

Power Target

Maximum Number of Subjects

Significance Level

Alpha = 0.05 ▼

Proportion (DV = 1) in Control Group

Proportion (DV = 1) in Treatment Group

Power Target

Maximum Number of Subjects

Significance Level

Alpha = 0.05 ▼

Treatment Effect Size

Standard Deviation of Outcome Variable

## 6 Comment utiliser la simulation pour estimer la puissance ?

La puissance est une mesure de la fréquence pour laquelle, compte tenu des hypothèses, nous obtiendrions des résultats statistiquement significatifs si nous devons mener notre expérience des milliers de fois. La formule de calcul de puissance se base sur des hypothèses et renvoie une solution analytique. Cependant, en raison des progrès de l'informatique moderne, nous n'avons pas à nous fier à des solutions analytiques pour l'analyse de puissance. Nous pouvons dire à nos ordinateurs d'exécuter littéralement l'expérience des milliers de fois et simplement compter la fréquence à laquelle notre expérience est significative.

Le code ci-dessous montre comment effectuer cette simulation dans R.

```

possible.ns <- seq(from=100, to=2000, by=40) # Les tailles d'échantillon que nous considérerons
stopifnot(all( (possible.ns %% 2)==0 )) ## On nécessite un nombre pair d'échantillons expérimentaux
powers <- rep(NA, length(possible.ns)) # objet vide pour collecter les estimations de simulation
alpha <- 0.05 # Niveau de signification standard
sims <- 500 # Nombre de simulations à réaliser pour chaque N
#### Boucle externe pour varier le nombre de sujets ####
for (j in 1:length(possible.ns)){ N <- possible.ns[j] # Choisit la j-ième valeur pour N
  Y0 <- rnorm(n=N, mean=60, sd=20) # résultat potentiel pour le contrôle
  tau <- 5 # hypothèse de l'effet du traitement
  Y1 <- Y0 + tau # résultat potentiel pour le traitement
  significant.experiments <- rep(NA, sims) # objet vide pour compter les expériences significatives

  #### Boucle interne pour mener des expériences "sims" fois pour chaque N ####
  Y0 <- rnorm(n=N, mean=60, sd=20) # résultat potentiel pour le contrôle
  tau <- 5 # Hypothèse de l'effet du traitement
  Y1 <- Y0 + tau # résultat potentiel pour le traitement
  for (i in 1:sims){
    ## Z.sim <- rbinom(n=N, size=1, prob=.5) # Faire une assignation aléatoire par tirage au sort
    Z.sim <- sample(rep(c(0,1),N/2)) # Faire une assignation aléatoire par tirage au sort en s'assurant de groupes de taille égale
    Y.sim <- Y1*Z.sim + Y0*(1-Z.sim) # Révéler les résultats selon l'assignation
    fit.sim <- lm(Y.sim ~ Z.sim) # Faire l'analyse (une simple régression)
    p.value <- summary(fit.sim)$coefficients[2,4] # Extraire les p-valeurs
    significant.experiments[i] <- (p.value <= alpha) # Déterminer la signification en accord avec p < 0.05
  }
  powers[j] <- mean(significant.experiments) # stocker le taux de réussite moyen (puissance) pour chaque N
}
powers

```

```

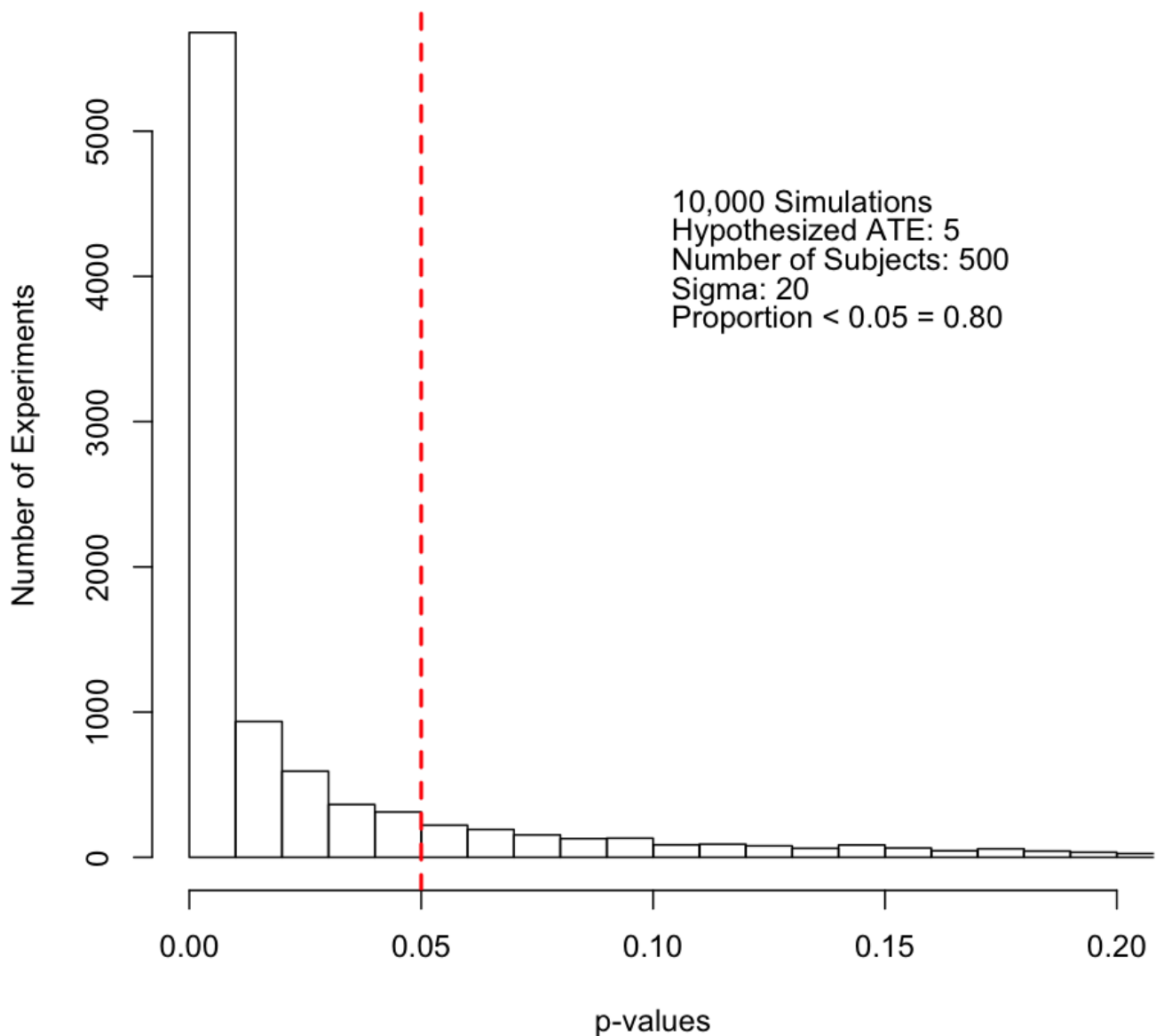
## [1] 0.262 0.334 0.454 0.466 0.544 0.522 0.636 0.658 0.748 0.818 0.816 0.808 0.794 0.920 0.920 0.896
0.936 0.924 0.952
## [20] 0.938 0.940 0.952 0.984 0.972 0.982 0.988 0.990 0.984 0.986 0.994 0.988 0.996 0.996 0.998 0.994
1.000 0.998 0.998
## [39] 0.998 1.000 0.998 0.998 1.000 1.000 1.000 1.000 1.000 1.000

```

Le code de cette simulation est disponible ici (<https://egap.org/resource/script-power-analysis-simulations-in-r/>). La simulation est un moyen beaucoup plus flexible et beaucoup plus intuitif de penser à l'analyse de puissance. Même les plus petits ajustements d'un design expérimental sont difficiles à capturer dans une formule (ajout d'un deuxième groupe de traitement, par exemple), mais sont relativement simples à inclure dans une simulation.

En plus de compter la fréquence à laquelle vos expériences sont statistiquement significatives, vous pouvez observer directement la distribution des p-valeurs que vous êtes susceptible d'obtenir. Le graphique ci-dessous montre que sous ces hypothèses, vous pouvez vous attendre à obtenir quelques p-valeurs dans la plage de 0.01, mais que 80% seront inférieures à 0.05.

## Simulated p-values



## 7 Comment changer votre design pour améliorer votre puissance ?

En ce qui concerne la puissance statistique, la seule chose que vous contrôlez est le design de l'expérience. Comme nous l'avons vu ci-dessus, un choix de design évident est le nombre de sujets à inclure dans l'expérience. Plus il y a de sujets, plus la puissance est élevée.

Cependant, le nombre de sujets n'est pas le seul choix de design qui a des conséquences sur la puissance. Il existe deux grandes catégories de choix de design qui sont particulièrement importantes à cet égard.

- **Choix de l'estimateur.** Utilisez-vous la différence des moyennes ? Allez-vous effectuer une transformation, comme un logit ou un probit ? Contrôlerez-vous les covariables ? Utiliserez-vous un estimateur robuste pour l'erreur type ? Tous ces choix feront une différence pour la signification statistique de vos résultats, et donc pour la puissance de votre expérience. Une façon simple d'y penser est d'imaginer quelle commande vous exécuterez dans R ou Stata après l'expérience ; c'est votre estimateur !
- **Protocole de randomisation.** Quel type de randomisation utiliserez-vous ? La randomisation simple donne à tous les sujets une probabilité égale d'être dans le groupe de traitement, puis effectue un tirage au sort (éventuellement pondéré) pour chacun. La randomisation complète est similaire, mais elle garantit qu'un nombre exact d'unités seront assignées

au traitement. La randomisation par bloc est encore plus puissante : elle garantit qu'un certain nombre d'unités au sein d'un sous-groupe seront assignés au traitement. Une assignation aléatoire restreinte rejette certaines assignations aléatoires sur la base d'un certain ensemble de critères — un manque d'équilibre par exemple. Ces différents types d'assignation aléatoire peuvent augmenter considérablement la puissance d'une expérience sans frais supplémentaires. Renseignez-vous sur les protocoles de randomisation ici (<https://egap.org/resource/10-things-to-know-about-randomization/>).

Il y a trop de choix à couvrir dans ce court article, mais consultez le code de simulation pour l'analyse de puissance pour savoir comment commencer. Pour vous donner une idée de l'approche de simulation, réfléchissez à la manière dont vous effectueriez une analyse de puissance si vous vouliez inclure des covariables dans votre analyse.

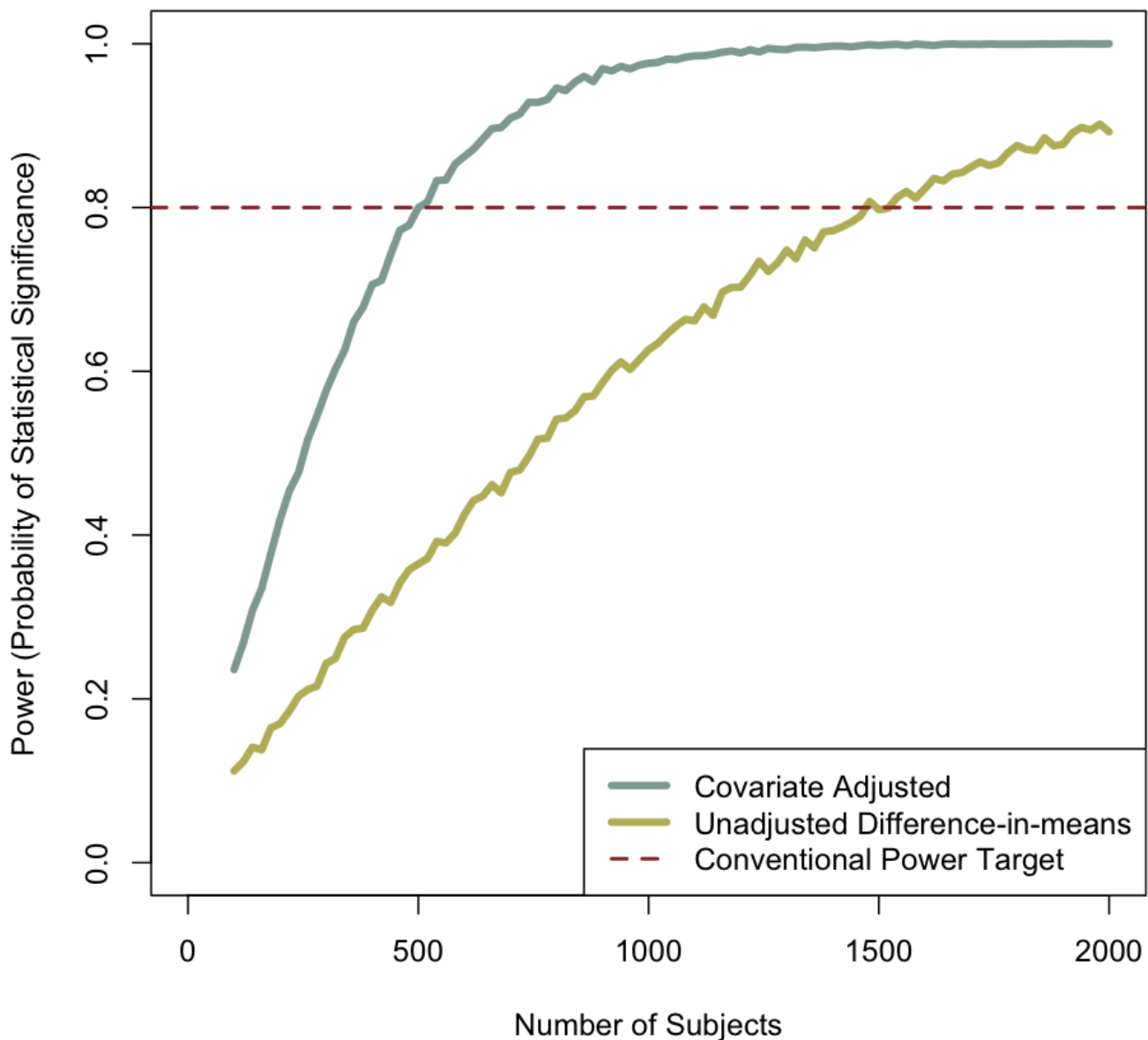
Si les covariables que vous incluez en tant que variables de contrôle sont fortement liées au résultat, alors vous avez considérablement augmenté la puissance de votre expérience. Malheureusement, la puissance supplémentaire qui accompagne l'inclusion des variables de contrôle est très difficile à capturer dans une formule compacte. Quasiment aucune des formules de puissance trouvées dans les manuels ou sur internet ne peut fournir des conseils sur ce que l'inclusion de covariables fera pour votre puissance.

La réponse est la simulation.

- Supposons que nous étudions l'effet d'une intervention éducative sur le revenu
- Supposons que nous ayons de bonnes données sur la relation entre deux covariables et le revenu : l'âge et le sexe. Dans cette économie, les hommes gagnent plus que les femmes et les personnes âgées gagnent plus que les jeunes.
- Exécutez une régression du revenu sur l'âge et le sexe et enregistrez les coefficients, en utilisant des données d'enquête préexistantes (mieux encore : utilisez les données de référence des futurs participants à votre expérience !)
- Générez de fausses données de covariables - N sujets au total, mais ventilés par âge et par sexe d'une manière qui reflète votre groupe de sujets expérimentaux.
- Générez de fausses données de contrôle - où le résultat est fonction de l'âge et du sexe selon vos estimations de régression
- Faire l'hypothèse d'un effet de traitement pour générer de fausses données de traitement
- Exécutez l'expérience 10 000 fois et enregistrez la fréquence à laquelle, à l'aide d'une régression avec contrôles, votre expérience s'avère significative.

Voici un graphique qui compare la puissance d'expériences avec et sans contrôle par les attributs de base. Le  $R^2$  de la régression reliant le revenu à l'âge et au sexe est assez élevé - environ 0,66 - ce qui signifie que les covariables que nous avons rassemblées (générées) sont hautement prédictives. Pour une comparaison approximative, sigma, le niveau de bruit de fond auquel le modèle non ajusté est confronté, est d'environ 33. Ce graphique montre que pour n'importe quel N, le modèle ajusté par les covariables a plus de puissance, à tel point que le modèle non ajusté aurait besoin de 1500 sujets pour réaliser ce que le modèle ajusté peut faire avec 500 sujets.

## Power Analysis: Hypothetical Treatment Effect = 5



Cette approche ne repose pas sur une formule pour trouver la probabilité d'obtenir un résultat statistiquement significatif : elle repose sur la force brute ! Et parce que la simulation vous permet de spécifier chaque étape du design expérimental, vous effectuez une analyse de puissance beaucoup plus nuancée que la simple considération du nombre de sujets.

## 8 Analyse de puissance pour plusieurs traitements

De nombreuses expériences emploient plusieurs traitements qui sont comparés à la fois les uns aux autres et à un groupe de contrôle. Cette complication supplémentaire change ce que nous entendons lorsque nous disons la “puissance” d’une expérience. Dans le cas d’un seul groupe de traitement, la puissance est simplement la probabilité d’obtenir un résultat statistiquement significatif. Dans le cas des traitements multiples, cela peut signifier plusieurs choses : A) la probabilité qu’au moins un des traitements soit significatif, B) la probabilité que tous les traitements soient significatifs (par rapport au contrôle) ou C) la probabilité que les traitements soient classés dans un certain ordre hypothétique et que ce classement soit statistiquement significatif.

Cette question des bras de traitement multiples est liée au problème des comparaisons multiples. Les tests de signification standard sont basés sur l’hypothèse que vous effectuez un seul test de signification statistique, et les p-valeurs dérivées de ces tests reflètent la probabilité sous l’hypothèse nulle de voir un effet de traitement aussi important (ou plus important). Si, toutefois, vous effectuez plusieurs tests, cette probabilité n’est plus correcte. Dans une série de tests, la probabilité qu’au moins un des tests soit significatif même lorsque l’effet réel est nul est plus élevée, essentiellement parce que vous avez plus de



tentatives. Une solution couramment citée (sinon couramment utilisée) consiste à utiliser la correction de Bonferroni : spécifiez le nombre de comparaisons que vous effectuerez à l'avance, puis divisez votre niveau de signification  $\alpha$  par ce nombre.

Si vous allez utiliser une correction de Bonferroni, les calculateurs de puissance standard seront plus compliqués à utiliser : vous devrez spécifier vos niveaux  $\alpha$  corrigés de Bonferroni et calculer la puissance séparée de chaque comparaison. Pour calculer la probabilité que tous les tests soient significatifs, multipliez toutes les puissances séparées ensemble. Pour calculer la probabilité qu'au moins un des tests soit significatif, calculez la probabilité qu'aucun ne le soit, puis soustrayez de un.

Ou vous pouvez utiliser la simulation. Un exemple de calcul de puissance fait dans R est disponible sur la page des simulations.

## 9 Comment penser la puissance pour les designs par grappe (cluster) ?

Lorsqu'une expérience doit assigner des groupes entiers de personnes à un traitement plutôt qu'individuellement, nous avons une expérience par grappe. Ceci est courant pour les expériences éducatives, où des classes entières d'enfants sont assignées au traitement ou au contrôle, ou dans l'économie du développement, où des villages entiers d'individus sont assignés au traitement ou au contrôle.

En règle générale, le découpage par grappe diminue votre puissance. Si vous pouvez éviter de regrouper vos traitements en grappe, c'est préférable pour la puissance. À moins que vous ne soyez confronté à des problèmes liés aux effets de débordement, à la logistique ou à l'éthique, réduisez la variation au niveau le plus bas possible.

Le meilleur scénario pour un design par grappe est lorsque la grappe dans laquelle se trouve un sujet fournit très peu d'informations sur leurs résultats. Supposons que les sujets soient assignés de manière aléatoire à des grappes - la grappe n'aiderait pas du tout à prédire les résultats. Si la grappe n'est pas prédictive du résultat, alors nous n'avons pas perdu trop de puissance à cause du découpage par grappe.

Là où le découpage par grappe cause vraiment des problèmes, c'est lorsqu'il existe une relation forte entre la grappe et le résultat. Pour prendre l'exemple des villages, supposons que certains villages soient, dans l'ensemble, beaucoup plus riches que d'autres. Les grappes pourraient alors être assez prédictives du niveau d'instruction. Le découpage par grappe peut réduire la taille effective de votre échantillon du nombre total d'individus au nombre total de grappes.

Il existe des formules qui peuvent vous aider à comprendre les conséquences du découpage par grappe — voir Gelman/Hill page 447-449 pour une discussion approfondie. Bien que ces formules puissent être utiles, elles peuvent également être assez lourdes à utiliser. L'idée principale est cependant simple : vous obtenez généralement plus de puissance en augmentant le nombre de grappes qu'en augmentant le nombre de sujets au sein des grappes. Mieux vaut avoir 100 grappes de 10 sujets que 10 grappes avec 100 sujets.

Encore une fois, une approche plus flexible de l'analyse de puissance lorsqu'il s'agit de grappes est la simulation. Voir la page de simulations ici (<https://TODO>), ou lisez Gelman/Hill page 450-453 pour une autre approche de simulation.

## 10 Une bonne analyse de puissance facilite le pré-enregistrement

Lorsque vous traitez de puissance statistique, vous vous concentrez sur ce que vous ne pouvez pas contrôler (le bruit) et ce que vous pouvez contrôler (le design). Si vous utilisez l'approche de simulation pour l'analyse de puissance, vous serez obligé d'imaginer à quoi ressembleront vos données et comment vous allez les gérer. Vous aurez la possibilité de spécifier à l'avance toutes vos intuitions et vos meilleures suppositions, afin de pouvoir lancer vos expériences avec des attentes claires sur ce qu'elles peuvent et ne peuvent pas montrer. C'est du travail, mais la bonne nouvelle est que si vous le faites vraiment, vous êtes sur la bonne voie pour mettre en place un plan de pré-analyse complet et enregistrable.

1. Auteur : Alex Coppock, 20 novembre 2013. Le guide est un document vivant et susceptible d'être mis à jour par les membres de EGAP à tout moment. Coppock n'est pas responsable des modifications ultérieures apportées à ce guide.↵
2. Formule reproduite de Gerber et Green 2012, page 93↵
3. Pour un outil supplémentaire de visualisation de puissance en ligne, consultez le blog R Psychologist (<http://rpsychologist.com/d3/NHST/>) de Kristoffer Magnusson.↵