GUIDE D'INTRODUCTION À LA PUISSANCE STATISTIQUE

AVEC LE LOGICIEL G*POWER 3

Samuel Guay, B.A. (c)
Joel Gagnon, Ph.D. (c)
Sébastien Gélinas, Ph.D. (c)
Michael Cantinotti, Ph.D., professeur

SESAME
SERVICE DE SOUTIEN ET D'ENTRAIDE EN MÉTHODOLOGIE APPLIQUÉE

Département de psychologie

Université du Québec à Trois-Rivières

Mars 2016 – Version #1

Si vous avez des suggestions pour améliorer ce document, s'il vous plait, contactez Samuel Guay à l'adresse samuel.guay@uqtr.ca

Intéressé(e) à en apprendre davantage sur la puissance statistique au travers d'une capsule multimédia ? Consultez le site du SESAME à l'adresse :

www.sesame-savoir.com -> Capsules

1. Pourquoi rédiger un guide d'introduction à la puissance statistique ?

Bien que plusieurs outils de calculs de puissance pour soutenir la planification de recherches soient disponibles, ces outils demeurent encore trop peu utilisés ou utilisés de façon sous-optimale. Ce guide se veut une introduction à l'analyse de puissance statistique afin de vous inciter à en faire usage lors de vos recherches en psychologie.

Comme *G*Power*, un logiciel librement disponible, exécutera les calculs pour vous, ce guide n'est pas destiné à détailler les formules mathématiques permettant d'effectuer les différentes analyses de puissance. L'accent est ainsi mis sur une compréhension appliquée et pratique de la puissance statistique. Ce guide vise trois objectifs principaux :

1) Vous aider à comprendre l'utilité et l'importance de la puissance statistique.
2) Vous présenter les principales étapes à effectuer lors d'une analyse de puissance *a priori* de base.
3) Vous proposer une façon de présenter les résultats d'une analyse de puissance *a priori* dans une étude, ainsi que dans le cadre d'une demande de bourse ou de subvention.

2. Cadre théorique

La puissance statistique représente la probabilité qu'un test statistique va effectivement conclure qu'un phénomène (d'une taille prédéfinie) est présent lorsque l'hypothèse nulle est réellement fausse (Cohen, 1988). Autrement dit, il s'agit de la probabilité qu'un test rejette l'hypothèse nulle lorsque celle-ci est fausse dans la population visée par l'étude (Aberson, 2010).

Concrètement, la puissance indique au chercheur la probabilité que l'effet à l'étude apparaisse statistiquement significatif dans un échantillon lorsque cet effet existe effectivement dans la population. Par exemple, si votre analyse de puissance *a priori* indique une puissance de 0,80 (soit 80 %) et que vous réalisez votre étude une multitude de fois, vous devriez détecter une différence statistiquement significative dans 80 % des cas lorsque l'effet est présent dans la population. En contrepartie, pour 20 % des cas, l'effet ne sera pas mis en évidence, même s'il y a réellement un effet au sein de la population qui vous intéresse.

En résumé, si l'effet recherché est réellement présent dans la population, la puissance constitue la probabilité de détecter l'effet mesuré à partir de l'échantillon, de façon statistiquement significative. Il s'agit donc d'une probabilité informant le chercheur de ses chances de détecter significativement l'effet qu'il veut mettre en évidence. Plus la probabilité est élevée, plus le chercheur a de probabilités de mettre l'effet en évidence lorsque celui-ci existe réellement.

Importance et utilité de la puissance

Déjà en 1962, Cohen alarmait ses collègues à l'effet qu'un problème de puissance existait dans les publications qu'il avait recensées au sein d'un périodique. La puissance moyenne était de 48 % pour détecter une taille d'effet moyenne. Presque 50 ans plus tard, la même tendance est constatée parmi plusieurs domaines de recherche, notamment la psychiatrie, la psychologie sociale et les neurosciences (Aberson, 2010; Button et al., 2013). Dans le domaine de la neuropsychologie clinique, par exemple, un maigre 50 % de puissance moyenne a été trouvé pour mettre en évidence une taille d'effet moyenne (Bezeau & Graves, 2001). En d'autres mots, les études avaient environ une chance sur deux d'obtenir un résultat statistiquement significatif lorsqu'un effet existait dans la population concernée par les études.

Il est important de préciser que la puissance est largement influencée par la taille de l'effet à l'étude. Comme chaque domaine possède ses propres tailles d'effet, il faut interpréter la puissance en fonction de celles-ci, et non seulement en fonction des critères de Cohen (1988), qui constituent de simples conventions. Par exemple, lorsqu'on considère les tailles d'effet typiquement retrouvées dans le domaine de la neuropsychologie clinique, soit de grandes tailles d'effet selon le barème de Cohen (1988), la puissance moyenne augmente à 95 % comparativement à 50 % pour des tailles d'effet modérées (Bezeau & Graves, 2001). Cette notion est abordée plus en profondeur dans la section traitant de la taille d'effet.

En somme, même s'il semble que les chercheurs considèrent de plus en plus l'importance de la puissance statistique en recherche, ces derniers ne l'utilisent encore que trop rarement lors de la phase de planification du devis de recherche pour augmenter leurs chances de détecter un phénomène (Descôteaux, 2007). Certains chercheurs concluent même qu'intégrer des cours sur la notion de puissance statistique, ainsi que sur la façon de faire des analyses *a priori* s'avère une nécessité afin d'offrir une formation de qualité aux futurs chercheurs (Abraham & Russell, 2008).

Commençons d'abord par examiner à l'aide d'un exemple les conséquences qu'un manque de puissance peut engendrer.

- Imaginez que vous vous préparez à mener une étude en voulant recruter trop peu de participants. Les résultats ne généreront probablement aucune différence statistiquement significative parce que l'analyse de puissance de votre étude révèle une puissance *a priori* de 30 %. Autrement dit, vous n'avez que trois chances sur 10 que votre effet soit mis en évidence de façon statistiquement significative lorsque l'effet existe réellement dans la population.

- Décider de réaliser cette expérience malgré tout, en recrutant trop peu de participants, pourrait générer plusieurs impacts négatifs :
  - <u>Concernant vos finances :</u>
    - Vous risquez de débourser de l'argent pour les compensations financières et les assistants de recherche pour une étude qui ne découvrira probablement rien.
    - Si l'argent provient de fonds publics, vous avez la responsabilité de ne pas mener d'études qui sont vouées à l'échec. Dans le contexte de ressources financières limitées, il faut en faire bon usage !

  - <u>Concernant votre temps :</u>
    - Non seulement vous risquez de perdre votre temps, mais vous risquez également de faire perdre leur temps aux participants, ce qui représente un gaspillage de ressources et un enjeu éthique. Tout le temps et les efforts investis de près comme de loin pour l'étude devront être réinvestis pour refaire l'expérience avec une puissance suffisante.
    - De plus, un sujet d'étude particulier pourrait sembler inintéressant à des chercheurs qui lisent les résultats non significatifs et qui concluraient à tort qu'il n'y a rien à rechercher dans ce domaine. En raison d'un manque de connaissances sur les enjeux de puissance, ce sujet d'étude pourrait être abandonné sans raison valable tout simplement parce qu'aucune étude avec une puissance adéquate n'a été publiée.

  - <u>Concernant l'éthique :</u>
    - Outre le temps et l'argent, s'il y a des risques liés l'étude, cela reviendrait à exposer les participants à ces risques inutilement, puisque les chances de découvrir un effet sont plutôt minces.
    - Tirer les conclusions sur vos résultats à l'aide d'une pièce de monnaie est l'équivalent d'avoir une puissance de 50 %. Est-ce éthique de mener une étude qui autant de puissance que de tirer ses résultats à pile ou face ? D'après Cohen (1988), le meilleur compromis serait d'avoir au moins une puissance de 0,80 avant de débuter une étude.

De plus, des études faibles en puissance produisent plus de résultats contradictoires et inconsistants dans la littérature. Ces études faibles en puissance tendent à surestimer les tailles d'effets. Comme elles sont faibles en puissance, les chances de trouver une grosse taille d'effet sont plus grandes que de trouver une petite taille d'effet (Ioannidis, 2005). Comme les éditeurs de revue préfèrent habituellement les résultats statistiquement significatifs, ce sont la plupart du temps de grandes tailles d'effet qui se retrouvent dans la littérature. La littérature s'avère

donc moins consistante comparativement à la situation où les études seraient réalisées avec une puissance adéquate.

En effet, l'interprétation des résultats non statistiquement significatifs est difficile, particulièrement lorsqu'il manque de la puissance statistique. Cela peut donner lieu au phénomène communément appelé le *File Drawer Problem* en anglais, terme créé par Rosenthal (1979). Ce phénomène survient lorsque les résultats d'études non statistiquement significatifs sont mis de côté et remisés au fond du classeur du chercheur en raison du biais de publication. Autrement dit, ces études ne sont pas publiées, car les chercheurs et les éditeurs de journaux ont souvent une préférence pour les résultats statistiquement significatifs. Subséquemment, comme les méta-analyses ne tiennent fréquemment compte que des résultats publiés dans des revues avec comité de pairs et que ceux-ci sont majoritairement significatifs, ceci peut engendrer une surestimation des tailles d'effet (Ware & Munafò, 2014).

Bref, le manque de puissance augmente le biais de publication et génère de la discordance dans la littérature ou un effet d'inflation des tailles d'effet réelles. Cet effet d'inflation entraine une difficulté à répliquer les études antérieures. De plus, les études réplicatives ont tendance à obtenir de plus petites tailles d'effet, puisque les découvertes régressent vers la taille d'effet réelle (phénomène de régression vers la moyenne). Donc, plus la puissance est faible, plus la probabilité d'obtenir un résultat statistiquement significatif concernant l'effet observé est faible également (Ioannidis, 2005).

Les bénéfices associés à la réalisation d'une analyse de puissance *a priori*

Lorsque l'on parle de puissance *a priori*, cela signifie que l'estimation de la puissance est réalisée avant le début de la collecte de données. Il s'agit d'une étape préliminaire importante lors de la phase de planification d'une étude. En réalisant une étude avec une puissance adéquate, vous aurez une plus grande probabilité d'obtenir des résultats statistiquement significatifs. En plus de contribuer à une augmentation fiable des connaissances dans littérature scientifique, cette dernière deviendra plus consistante. De plus, une étude avec suffisamment de puissance permettra d'éviter la majorité des problèmes qu'un manque de puissance peut engendrer !

La principale fonction de l'analyse de puissance *a priori* consiste à indiquer le nombre de participants nécessaires afin de détecter un effet d'une taille prédéfinie. Il s'agit donc d'informer le chercheur sur le nombre de participants requis pour obtenir des résultats statistiquement significatifs. Toutefois, si le nombre de participants est limité, effectuer une analyse de puissance peut s'avérer efficace pour justifier la réalisation ou non d'une étude en indiquant la puissance en fonction d'un nombre donné de participants, d'une taille d'effet prédéfinie et du seuil alpha retenu. Cette analyse de puissance est légèrement différente de l'analyse *a priori* et porte le nom d'analyse de puissance de type « sensibilité » (Faul, Erdfelder, Lang, & Buchner, 2007).

- Par exemple, si le budget permet de tester seulement 40 participants pour tenter de mettre en évidence un phénomène donné, et que l'analyse de puissance indique une puissance de 50 %, il faut se demander si ça vaut la peine de réaliser l'étude. Il faudrait plutôt examiner des moyens d'améliorer la sensibilité de l'étude. Certains moyens pour augmenter la puissance existent sans pour autant augmenter le nombre de participants. Par exemple, utiliser un devis à mesures répétées (ou avec blocs) au lieu d'utiliser des groupes indépendants (lorsque cela est possible). Voir Lipsey (1990) pour plus de détails sur la relation entre la sensibilité et la puissance.

Concepts essentiels à comprendre

Posséder une bonne compréhension des concepts essentiels est important, d'une part, pour être capable d'évaluer la puissance éventuelle d'une étude, d'autres parts, pour être en mesure d'évaluer les paramètres requis pour atteindre un niveau de puissance adéquat lors de la préparation d'une recherche. Surtout, il est important de comprendre ces concepts pour mener avec rigueur une recherche.

Comme les tests d'hypothèses sont basés sur des probabilités conditionnelles et que la puissance découle de ces tests, un résumé du test d'hypothèse, ainsi que des erreurs de type I et II est indispensable pour développer une bonne compréhension de la notion de puissance statistique.

- Le test d'hypothèse s'avère utile quand l'étude de la population entière est irréaliste et que le chercheur doit plutôt étudier un échantillon de cette population (Blais, 1991).
    - Puisque seulement une partie de la population est étudiée, il y a nécessairement une erreur d'échantillonnage lors de l'estimation du paramètre à partir de la statistique de l'échantillon. Il s'agit de la différence entre la « vraie » valeur qui aurait été obtenue si la population entière avait été étudiée et celle obtenue à partir de l'échantillon. Plus la taille de l'échantillon augmente, plus l'erreur d'échantillonnage diminue.
    - Le test d'hypothèse est utilisé afin de prendre une décision pour inférer la valeur du paramètre dans la population à partir de la statistique de l'échantillon étudié (c.-à-d., présence ou absence de l'effet d'intérêt). La valeur résultante du test est la probabilité d'obtenir la statistique observée lorsque l'hypothèse nulle est vraie. Le test d'hypothèse se réfère à deux hypothèses.
        - L'hypothèse nulle (H_0) affirme qu'aucun effet n'est présent dans la population. Ce sera cette hypothèse qui est testée.
        - L'hypothèse alternative (H_1) affirme la présence d'un effet. Autrement dit, c'est l'hypothèse du chercheur. Le chercheur retiendra cette hypothèse seulement s'il est en mesure de rejeter l'hypothèse nulle.

- **Alpha** (α)
  - Le test d'hypothèse permet d'établir quelle hypothèse est la plus appropriée pour conclure sur l'état de la population d'intérêt.
    - Y a-t-il présence d'un effet ou non ?
  - Le test d'hypothèse se base sur les probabilités conditionnelles pour déterminer la probabilité d'obtenir le résultat observé dans l'échantillon alors que l'hypothèse nulle est vraie.
    - Il s'agit du fameux *p* tant omniprésent en recherche quantitative !
  - Pour prendre la décision, la valeur du *p* sera alors comparée au seuil alpha.
    - α est généralement fixé à 0,05 en psychologie.
    - Un seuil alpha de 0,05 signifie que nous assumons un risque de 5 % de nous tromper lors du rejet de l'hypothèse nulle (H_0).
      - Si nous répétions une infinité de fois une expérience, nous risquons de rejeter à tort l'hypothèse nulle dans 5 % des cas.
  - Par contre, une valeur *p* [statistiquement] non significative, supérieure à alpha ne signifie pas qu'il n'y a pas de différence dans la population. Cela indique plutôt qu'aucune preuve suffisante de l'existence d'un effet n'a pu être établie à partir de l'échantillon étudié. Cela peut survenir en raison d'une erreur d'échantillonnage trop importante, des erreurs de mesures qui ont artificiellement minimisé l'effet recherché ou à un manque de puissance. Il est donc important d'avoir suffisamment de puissance pour écarter cette dernière raison.

En comparant la valeur alpha au seuil alpha, il n'y a que deux décisions possibles qui s'offrent au chercheur :

  - Soit le **non-rejet de l'hypothèse nulle** ou le **rejet de l'hypothèse nulle** (qui mène à retenir l'hypothèse alternative).
    - En réalité, ces décisions sont associées à quatre scénarios possibles qu'on retrouve dans les quatre cases blanches dans le tableau ci-dessous) :

Tableau 1 - Décisions possibles lors d'un test d'hypothèse

	RÉALITÉ DANS LA POPULATION	
DÉCISION du chercheur (basée sur l'échantillon)	**H_0 est vraie** (aucun effet n'existe dans la population)	**H_0 est fausse** (un effet existe dans la population)
H_0 retenue ($p > 0,05$) Elle ne peut être rejetée Résultats n'indiquent aucune présence d'un phénomène	**Décision adéquate** Aucune évidence pour rejeter H_0 $(1 - \alpha)$	**Erreur de type II** (Erreur β) Conserver H_0 à tort
Rejet d'H_0 ($p < 0,05$) **(H_1 retenue)** Résultats suggèrent la présence d'un phénomène	**Erreur de type I** (Erreur α) Rejeter H_0 à tort	**Décision adéquate** Puissance $(1 - \beta)$

- **Erreur de type I (erreur α)**
  - Cette erreur consiste à rejeter à tort l'hypothèse nulle. Nous pouvons également dire que nous avons détecté un effet dans notre échantillon alors qu'en réalité, il n'y a pas d'effet au sein de la population.
  - On contrôle pour ce type d'erreur en ajustant le seuil α. Plus on diminue le seuil alpha, moins on a de risques de faire cette erreur, puisque la probabilité de rejeter à tort l'hypothèse nulle diminue.
    - Ajuster un seuil alpha de 0,05 à 0,01 par exemple (mais ceci va entraîner une diminution de la puissance statistique).
- **Erreur de type II (ou erreur β)**
  - Cette erreur consiste à conserver l'hypothèse nulle alors qu'elle est fausse. En d'autres mots, cela consiste à ne pas avoir détecté un effet alors qu'il y a réellement un effet dans la population. Cette erreur est inversement proportionnelle à la puissance !
    - La puissance et l'erreur β sont donc complémentaires.
      - 1 = β + puissance, soit la puissance = 1 – β !
    - Effectivement, les deux définitions sont opposées. La puissance, quant à elle, est la probabilité qu'un test statistique rejette l'hypothèse nulle lorsque l'hypothèse nulle est fausse, pour un effet d'une taille prédéfinie.

Taille d'effet

- Le test d'hypothèse indique seulement s'il y a présence ou non d'un phénomène alors que la taille d'effet nous informe sur l'ampleur de ce phénomène.
  - La taille d'effet se définit comme étant la statistique qui quantifie à quel point le résultat d'un traitement (au sens large) diffère de l'hypothèse nulle, soit <u>l'importance de l'ampleur de l'effet</u> (Vacha-Haase & Thompson, 2004).
  - C'est un indicateur de la signification clinique :
    - L'ampleur de la différence entre des groupes ou la force de la relation entre des variables.

- Taille d'effet standardisée :
  - Pour comparer plusieurs tests entre eux, nous devons avoir des tailles d'effet standardisées. Standardiser une taille d'effet signifie que sa valeur est indépendante des unités de mesure utilisées pour un test donné.
  - Ce sont donc des tailles d'effet indépendantes des échelles de mesure originales. Ainsi, il est possible de comparer les différents résultats provenant d'études basées sur des instruments différents à l'aide des tailles d'effet standardisées.

- 2 groupes de tailles d'effet standardisées sont couramment utilisés :
  - Mesure de la différence de moyennes entre plusieurs groupes :
    - Par exemple, le *d* de Cohen.
  - Mesure de la force d'association entre deux variables :
    - Par exemple, le *r*, le r^2 ou le η^2 (êta carré).

4. Démarche d'une analyse de puissance *a priori*

- Définition de l'analyse de puissance *a priori*
  - Comme vous le savez déjà, il s'agit de mener une analyse de puissance avant même de collecter les données sur l'échantillon. Il a été mentionné plus haut que sa principale utilité est d'estimer le nombre de participants requis pour détecter de façon statistiquement significative un phénomène. Quatre éléments sont indispensables lors d'un calcul de puissance *a priori*. À noter que d'autres éléments viennent influencer la puissance, tels que la « directionnalité » des tests statistiques, le type de mesure, la fidélité des scores, etc.

- Influence des principaux éléments permettant le calcul (« P.A.NN.E. »)
  - **P**uissance (P.) [Figure 1]
    - Comme la puissance est complémentaire à bêta, plus bêta est petit, plus la puissance augmente. C'est logique, car la puissance est égale à $1 - \beta$.
  - **A**lpha (A.) [Figure 2]
    - Plus la valeur d'alpha diminue, plus il s'avère difficile de rejeter l'hypothèse nulle ($\alpha = 0{,}05 > \alpha = 0{,}01$). Donc, plus alpha est conservateur, plus la puissance diminue. Au contraire, plus alpha est libéral, plus la puissance augmente.
      - Toutefois, en psychologie, α est généralement gardé constant à 0,05.
  - **N**ombre « ***N*** » de participants (NN.) [Figure 3]
    - Plus le nombre de participants augmente, plus la puissance augmente.
    - Dans le cas d'un calcul de puissance *a priori*, c'est ce nombre qui est déterminé afin de connaitre le nombre de participants à recruter pour avoir une puissance d'au moins 0,80.
  - Taille d'**e**ffet (E.) [Figure 4]
    - Plus la taille d'effet augmente, plus la puissance augmente également.
    - Donc, lorsque le nombre de participants est gardé constant, la puissance augmente plus la taille d'effet augmente.

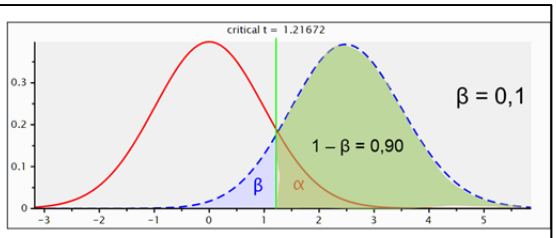
  - Les figures à la page suivante illustrent comment chaque concept influence la puissance tout en gardant les 3 autres concepts constants pour un test *t* à groupes indépendants unilatéral. L'aire sous la courbe (en vert) représente la puissance. Il est donc possible d'observer ses variations.

Figure 1

Lorsque β ↑,
la puissance ↓

1 – β = Puissance !

Il est recommandé d'avoir une puissance d'au moins 0,80
(β de 0,2)



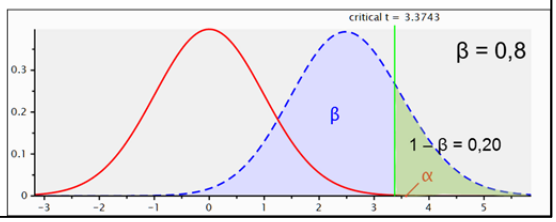
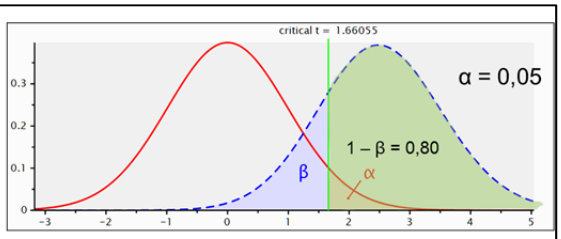


Figure 2

Lorsque α ↑, la puissance ↑

Plus α est libéral, plus la puissance augmente

En psychologie, α est généralement gardé constant à 0,05



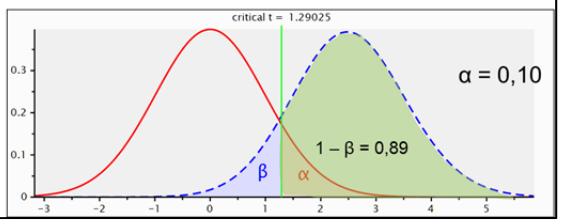
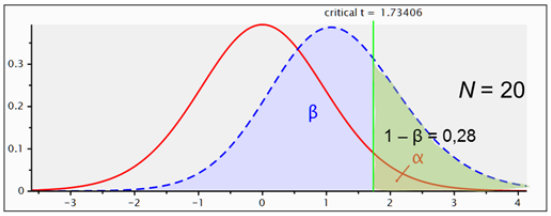


Figure 3

Plus le *N* augmente, plus la puissance augmente



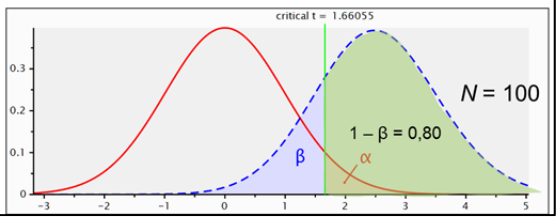
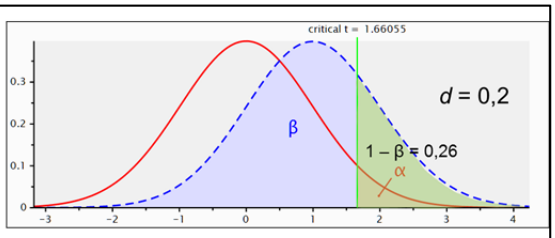
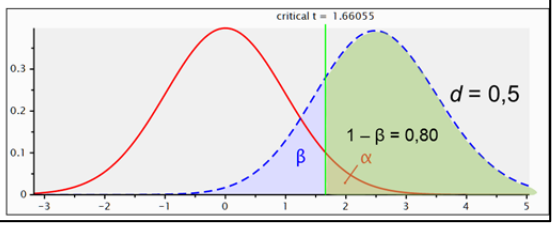


Figure 4

Plus la taille d'effet est grande, plus la puissance augmente





En raison de la relation mathématique qui les unit, il est possible de calculer l'un des quatre éléments de la P.A.NN.E. en connaissant les trois autres (Aberson, 2010).

- En effet, la puissance statistique s'obtient en fonction du seuil alpha, du nombre de participants et de la taille d'effet.
- Ainsi, pour trouver le nombre de participants (**NN.**) requis, il faut donc spécifier :
  - La puissance (**P.**)
    - Une puissance d'au moins 0,80 est habituellement considérée minimalement acceptable en psychologie (Cohen, 1988).
      - Il s'agit du meilleur compromis entre l'erreur bêta et l'erreur alpha (Aberson, 2010).
  - Alpha (**A.)**
    - Normalement en psychologie, le seuil est de 0,05 (mais cette convention peut être modifiée si la situation le justifie).
  - Taille d'effet (**E.**)
    - C'est la statistique la plus difficile est à estimer réalistement. Il faut donc s'armer de patience et de rigueur pour arriver à nos fins.
    - Pistes pour estimer concrètement la taille d'effet recherchée :
      - Le choix du test statistique détermine le type de taille d'effet !
        - Cherche-t-on une différence de moyenne ou une force de relation ?
      - Se baser sur la littérature existante
        - Quelles sont tailles d'effet rapportées dans la littérature dans notre domaine ?
      - Effectuer une étude pilote pour obtenir une taille d'effet préliminaire.
    - Si aucune étude sur notre sujet n'a été réalisée et que l'étude pilote n'est pas envisageable, les critères de Cohen (1988) peuvent servir de guide pour évaluer l'importance de la taille d'effet.
      - Toutefois, Thompson (2007) nous met en garde :
        - Le barème de Cohen ne devrait pas être appliqué à l'aveuglette. Il faut également se pencher sur la question de l'importance pratique dans notre contexte. Il faut se questionner sur la taille d'effet minimale qui possède une pertinence sur le plan clinique.

Taille d'effet (Cohen, 1988)	*d*	*r*
Petite	0,20	0,1
Modérée	0,50	0,3
Grande	0,80	0,5

- Lorsque vous avez défini ces trois critères, il ne reste qu'à ouvrir *G*Power* !

3. Mise en pratique avec le logiciel *G*Power*

Dans le domaine des neurosciences, la puissance moyenne des études recensées par Button et ses collègues (2013) est de seulement 20 % pour détecter une taille d'effet modérée. Dans le cadre de l'étude que nous projetons accomplir, nous voulons utiliser une technique de neurostimulation non-invasive. La neurostimulation peut se décrire comme étant toute stimulation exogène artificielle susceptible d'induire un changement fonctionnel dans l'activité neuronale du cerveau. Elle peut être de différentes natures, notamment de nature électrique ou électromagnétique. Par exemple, votre cellulaire modifie l'activité électrique de votre cerveau lorsque vous l'utilisez sur l'oreille (Ferreri et al., 2006) !

La technique que nous voulons utiliser est la stimulation transcrânienne par courant direct (*t*DCS), mais de multiples résultats sont contradictoires dans la littérature. Particulièrement lorsque la stimulation est appliquée au cortex préfrontal, toute stimulation semble provoquer des effets contraires. Il semble que stimuler le cortex dorsolatéral préfrontal de la même manière diminue ET augmente les capacités de la mémoire de travail parmi les études recensées (Tremblay et al., 2014). Étant donné que certains résultats s'opposent, il est important de mener une étude avec suffisamment de puissance pour s'assurer de la véracité de nos résultats et ainsi contribuer à la littérature avec plus de fiabilité.

Notre étude utilisera la *t*DCS dans le but d'améliorer les habiletés mathématiques des étudiants au baccalauréat en psychologie. La première étape est donc de conduire une analyse de puissance *a priori* afin d'avoir une idée du nombre de participants nécessaire pour déceler un effet significatif si l'effet existe réellement.

Notre devis est expérimental, les participants seront assignés aléatoirement soit à la condition *t*DCS, soit à la condition contrôle. Dans la condition *t*DCS, les participants devront résoudre des problèmes d'arithmétiques pendant 20 minutes tout en étant stimulés alors que dans la condition contrôle, les participants ne devront résoudre que des problèmes d'arithmétiques pendant 20 minutes. Voici les étapes que nous avons effectuées lors de la préparation de l'étude :

A. Quelle est notre hypothèse ? Il est important de s'interroger dès le départ, car nos tests statistiques seront effectués en fonction de notre hypothèse et ils influenceront la démarche pour faire le calcul de puissance.
- Il est attendu que les participants de la condition *t*DCS performent mieux et plus rapidement que les participants de la condition contrôle.
- Les tests utilisés pour vérifier nos hypothèses seront par conséquent des tests *t* à groupes indépendants unilatéraux. À noter qu'un test unilatéral s'avère plus puissant qu'un bilatéral. Il est donc avantageux, sur le plan de la puissance, d'émettre des hypothèses unidirectionnelles lorsque cela est plausible et pertinent.

B. Les hypothèses ont été émises en fonction de notre recension de la littérature. Lors de celle-ci, nous en avons profité pour repérer les différentes tailles d'effet obtenues par les autres chercheurs.

- Étude de Hauser, Rotzer, Grabner, Mérillat et Jäncke (2013) à propos de la vitesse : $d = 0{,}44$
- Étude de Grabner, Rütsche, Ruff et Hauser (2015) à propos de la précision : $d = 0{,}58$ et $0{,}85 = 0{,}715$ en moyenne.
- Nous préconisons l'approche de faire une moyenne des différentes tailles d'effet trouvées, mais d'autres façons plus sophistiquées existent.

C. À quelle valeur le seuil de significativité est-il fixé ?

- Généralement, le seuil alpha est gardé constant à 0,05 en psychologie.

D. Quelle puissance désirons-nous ? (Puissance minimalement acceptable)

- Comme dit précédemment, nous devrions convoiter une puissance minimale de 0,80.

E. Combien de participants prévoyons-nous recruter ?

- Il est désormais temps d'ouvrir *G*Power* et d'y entrer les données.
- La Figure 4 est une démonstration des étapes suivantes :

1. D'abord, sélectionner la bonne famille de tests.
   - ✓ Dans l'exemple, il s'agit de la famille des tests *t*.
2. Ensuite, choisir le type spécifique de tests.
   - ✓ Des tests *t* à groupes indépendants unilatéraux seront effectués.
3. Sélectionner le type d'analyse de puissance.
   - ✓ Il s'agit d'une analyse de puissance *a priori*.
4. Spécifier le seuil de significativité (**A.**)
   - ✓ 0,05 dans ce cas-ci.
5. Spécifier la puissance désirée (**P.**)
   - ✓ 0,80 pour les raisons énoncées auparavant.
6. Spécifier le ratio entre les groupes.
   - ✓ Il est préférable d'avoir le même nombre dans chaque groupe, alors le ratio = 1.
7. Entrer la taille d'effet que vous avez estimé obtenir lors de l'étude (**E.**)
   - ✓ Dans l'exemple, en tenant compte des études antérieures, nous estimons à 0,44 la taille d'effet attendue concernant la rapidité et à 0,715 pour la précision des réponses.
8. Appuyer sur *Calculate* pour que le logiciel exécute le calcul.
9. Regarder attentivement le nombre de participants total pour l'échantillon.
   - ✓ Il s'agit du nombre de participants (**NN.**) minimum pour obtenir une puissance de 0,80.
10. Si jamais vous avez plus d'une taille d'effet pour différentes hypothèses ou vous n'êtes pas certains de l'estimation de la taille d'effet, plusieurs scénarios peuvent être élaborés en cliquant sur ce bouton.
    - ✓ Voir Figure 5.

Figure 4

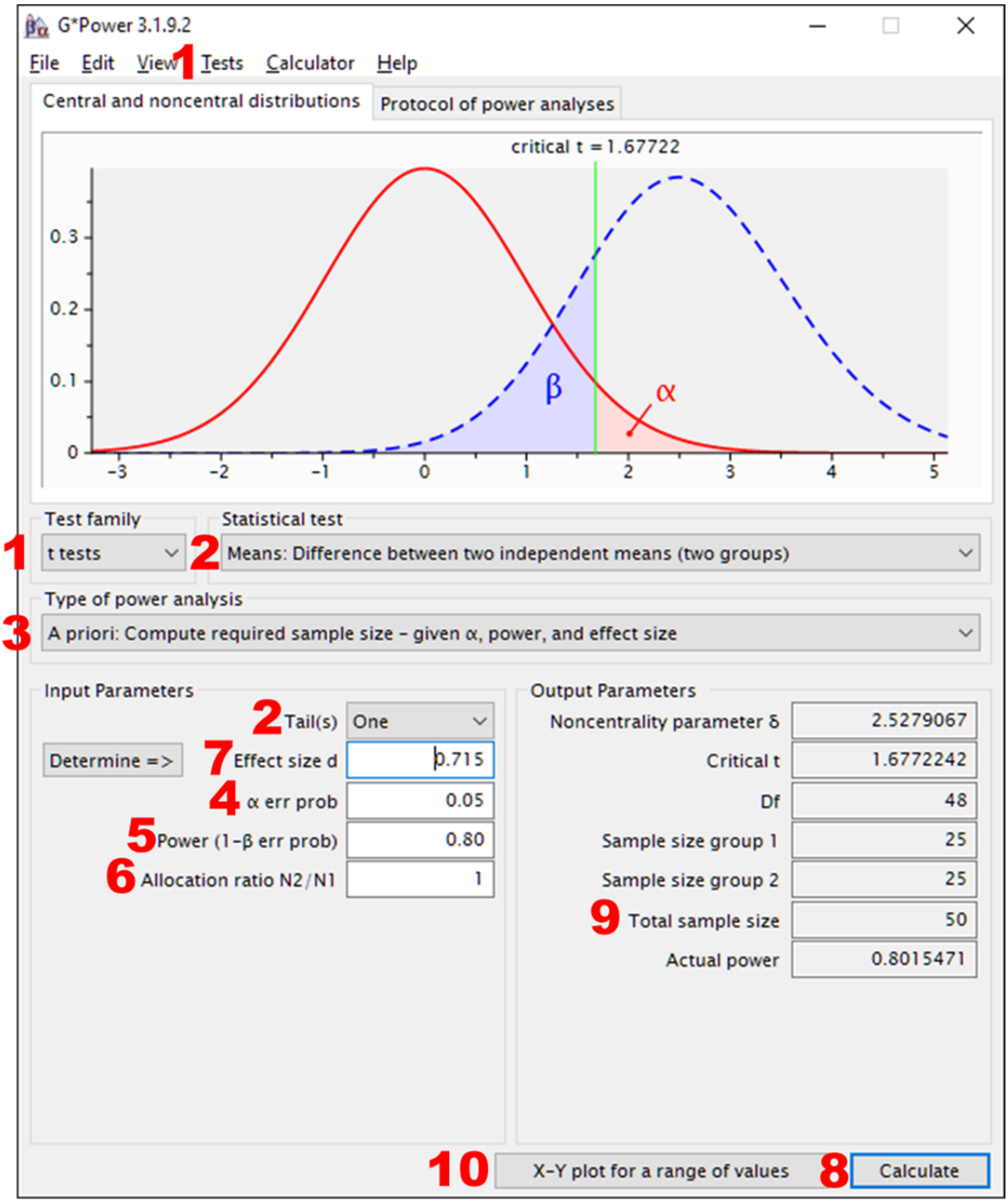
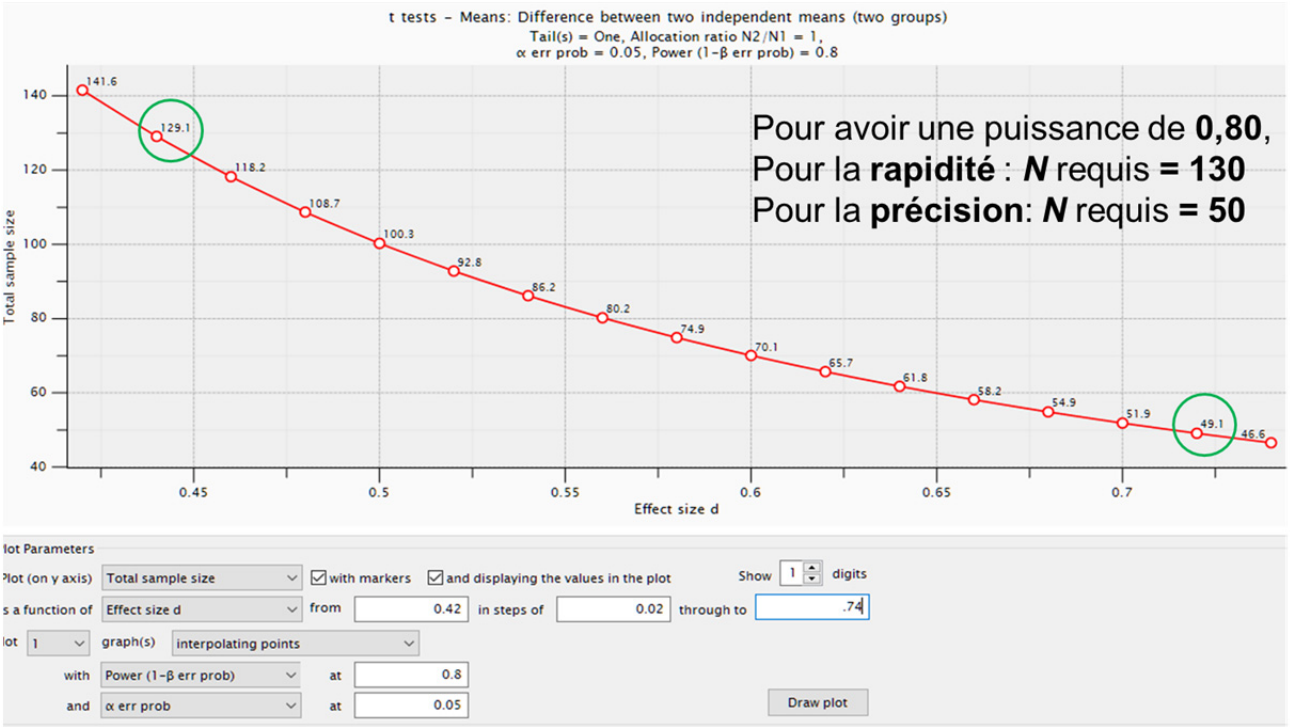

G*Power 3.1.9.2
File Edit View 1 Tests Calculator Help
Central and noncentral distributions
Protocol of power analyses
critical t = 1.67722
β
α
Test family
1 t tests
Statistical test
2 Means: Difference between two independent means (two groups)
Type of power analysis
3 A priori: Compute required sample size – given α, power, and effect size
Input Parameters
2 Tail(s) One
Determine =>
7 Effect size d 0.715
4 α err prob 0.05
5 Power (1-β err prob) 0.80
6 Allocation ratio N2/N1 1
Output Parameters
Noncentrality parameter δ 2.5279067
Critical t 1.6772242
Df 48
Sample size group 1 25
Sample size group 2 25
9 Total sample size 50
Actual power 0.8015471
10 X-Y plot for a range of values
8 Calculate

Figure 5



10. Si jamais vous n'êtes pas certain(e) de l'estimation de la taille d'effet, plusieurs scénarios peuvent être élaborés.
    - Voici les scénarios possibles en fonction de la taille d'effet.
    - En ayant une taille d'effet de 0,44, il faudrait au moins 130 participants pour obtenir une puissance de 0,80 pour l'hypothèse sur la rapidité.
    - Pour l'hypothèse sur la précision, la taille d'effet est de 0,72, alors seulement 50 participants seraient nécessaires.

Lorsqu'il y a une grande disparité entre le nombre de participants requis pour les différentes hypothèses, il faut les considérer dans l'ordre d'importance. Par exemple, dans ce cas-ci, nous considérons l'hypothèse de la précision plus importante que celle de la rapidité (qui est une hypothèse plutôt secondaire).

- À partir de ces scénarios, nous savons qu'il faudrait au moins 50 participants pour obtenir une puissance de 0,80 si la taille d'effet observée est de 0,715.
- Si la taille d'effet observée est plus petite, un plus grand nombre de participants sera requis.

Ensuite, il ne reste plus qu'à présenter le résultat de l'analyse au style APA.

4. Présentation d'une analyse de puissance au style APA

Il est important de présenter une analyse de puissance afin de démontrer la rigueur investie lors de la préparation d'une expérience. De plus, lorsque bien présentée, l'analyse de puissance *a priori* augmentera certainement la valeur de vos demandes de subventions ou de bourses. Effectivement, faire une analyse de puissance *a priori* démontre votre rigueur en recherche.

D'ailleurs, il a été constaté que parmi les études recensées dans le domaine de la psychologie de la santé, celles étant subventionnées avaient une meilleure puissance que celles qui ne l'étaient pas (Maddock & Rossi, 2001). Pour expliquer ce fait, les auteurs ont émis l'hypothèse que pour obtenir une subvention, les chercheurs doivent fournir des analyses de puissance *a priori* dans les documents de demandes. Les études subventionnées ont été préparées avec rigueur et en portant une attention particulière à la puissance. En conséquence, ces études sont plus en mesure de détecter la présence d'un effet significativement que les études qui n'ont pas été préparées avec une analyse de puissance a priori.

Dans un article scientifique ainsi que dans une demande de subvention, la présentation des résultats d'une analyse de puissance *a priori* se fait dans la section méthode. Il n'est pas nécessaire d'y aller en détail, l'important est de mentionner les éléments essentiels soit : le nombre de participants nécessaire, la puissance attendue, comment vous vous y êtes pris pour estimer la taille d'effet, et le seuil de significativité statistique. Voici un exemple basé sur les résultats de l'analyse de puissance faite précédemment :

> « Pour déterminer la taille de l'échantillon requise de la présente étude, nous avons recensé les résultats obtenus par d'autres chercheurs. La taille d'effet a été estimée à partir des résultats obtenus par Hauser et al. (2013) et Grabner et al. (2015) qui avaient obtenu respectivement un *d* de Cohen de 0,44 pour l'hypothèse de la rapidité et, en moyenne, 0,715 pour l'hypothèse de la précision. L'hypothèse de la précision étant plus importante, une analyse de puissance a priori effectuée à l'aide du logiciel *G*Power* 3.1 (Faul et al., 2007) a indiqué qu'afin d'obtenir une puissance de 80 %, un minimum de 50 participants est nécessaire dans le but de détecter une taille d'effet entre modérée et grande de 0,715 au seuil de signification $p < 0,05$. »

Prendre en compte l'attrition dans l'analyse de puissance

Dans cet exemple, l'attrition (ou la perte de participants en cours d'étude) n'est pas prise en compte. Cependant, dans certains devis de recherche, il arrive que l'attrition puisse devenir un problème surtout si l'on choisit le nombre de participants nécessaire de façon conservatrice. Afin de s'assurer d'une puissance suffisante sachant le taux d'attrition moyen présent dans les études antérieures (lorsque l'information est disponible), il suffit de prendre en compte cette variable lors du

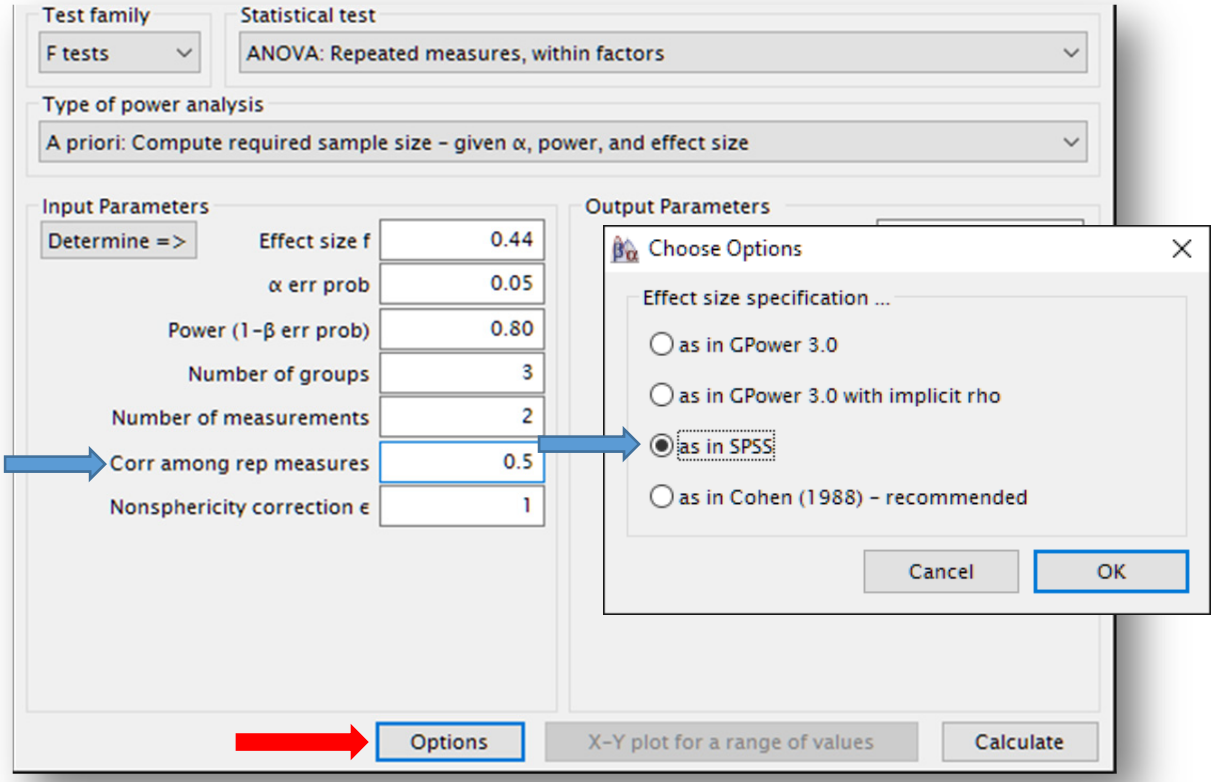
calcul de puissance. Par exemple, les études antérieures ont décelé un taux d'attrition variant entre 10 % et 20 %. Dans l'optique d'être le plus conservateur possible et ainsi s'assurer d'une puissance adéquate, il convient de choisir le plus haut taux d'attrition. Il suffit ensuite de multiplier ce pourcentage (20 %) à l'échantillon total obtenu grâce à l'analyse de puissance (dans ce cas-ci 50) : 20 % de 50 = 10, donc l'échantillon idéal serait de 60 participants soit 30 personnes par groupes. Cette information peut donc être ajoutée à la présentation de l'analyse de puissance ainsi :

> « Pour déterminer la taille de l'échantillon requise de la présente étude, nous avons recensé les résultats obtenus par d'autres chercheurs. La taille d'effet a été estimée à partir des résultats obtenus par Hauser et al. (2013) et Grabner et al. (2015) qui avaient obtenu respectivement un *d* de Cohen de 0,44 pour l'hypothèse de la rapidité et, en moyenne, 0,715 pour l'hypothèse de la précision. L'hypothèse de la précision étant plus importante, une analyse de puissance a priori effectuée à l'aide du logiciel *G*Power* 3.1 (Faul et al., 2007) a indiqué qu'afin d'obtenir une puissance de 80 %, un minimum de 50 participants est nécessaire dans le but de détecter une taille d'effet entre modéré et grande de 0,715 au seuil de signification $p < 0,05$ <u>en prenant en compte une attrition de 20 %, un total de 60 participants (soit 30 par groupe) est nécessaire.</u> »

5. Précision concernant l'ANOVA à mesures répétées

Dans le cas d'un calcul de puissance *a priori* pour une analyse de variance à mesures répétées, il est important d'apporter une précision. Par défaut, *G*Power* en plus des informations vues précédemment (puissance, taille d'effet, alpha), demande le nombre de temps de mesures et la corrélation entre les mesures. L'indice problématique dans cette situation est la corrélation entre les mesures. En effet, *G*Power* prend pour acquis que la corrélation entre les mesures n'est pas prise en compte dans le calcul de la taille d'effet alors qu'en réalité, lorsque la taille d'effet est calculée à partir du logiciel SPSS, la corrélation entre les mesures est prise en compte. En conséquence, il est essentiel d'indiquer à *G*Power* que la taille d'effet que vous utilisez pour votre calcul de puissance prend en compte l'indice de corrélation entre les mesures. En omettant cette étape, le calcul de puissance indiquera de très petites tailles d'échantillons qui n'auront aucun lien avec la réalité. Pour ce faire, il suffit de cliquer sur « option » située en bas de la fenêtre et sélectionner « *as in SPSS* » (voir Figure 6). Une fois cette manœuvre effectuée, l'indice de corrélation entre les mesures disparaitra et il vous sera ainsi possible d'effectuer votre calcul de puissance en étant confiant des résultats que vous obtiendrez.

Figure 6



6. Outils et références utiles pour les analyses de puissance

SÉSAME : Service de soutien et d'entraide en méthodologie appliquée, Département de psychologie, Université du Québec à Trois-Rivières & Département des sciences de la santé, Université du Québec à Chicoutimi.
http://www.sesame-savoir.com

G*Power 3 : Programme gratuit pour effectuer les analyses de puissance
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. http://www.gpower.hhu.de/en.html

Applied power analysis for the behavioral sciences : Excellent livre qui couvre les analyses de puissance en lien avec une variété de tests statistiques.
Aberson, C. L. (2010). *Applied power analysis for the behavioral sciences*. Londres: Routledge.

Excellent site pour convertir différentes tailles d'effet standardisées entre elles
http://www.psychometrica.de/effect_size.html#fvalue

The New Statistics: Estimation and Research Integrity: Atelier méthodologique à visionner qui est divisé en six parties sur les « nouvelles statistiques » à utiliser.
http://www.psychologicalscience.org/index.php/members/new-statistics

Références

Aberson, C. L. (2010). *Applied power analysis for the behavioral sciences*. Londres: Routledge.

Abraham, W. T., & Russell, D. W. (2008). Statistical power analysis in psychological research. *Social and Personnality Psychology Compass, 2*(1), 283-301.

Bezeau, S., & Graves, R. (2001). Statistical power and effect sizes of clinical neurospsychology research. *Journal of Clinical and Experimental Neuropsychology, 23*(3), 399-406.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2e ed.). Hillsdale, NJ: Erlbaulm.

Descôteaux, J. (2007). Statistical power: An historical introduction. *Tutorials in Quantitative Methods for Psychology, 3*(2), 28-34.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Ferreri, F., Curcio, G., Pasqualetti, P., Gennaro, L. D., Fini, R., & Rossini, P. M. (2006). Mobile phone emissions and human brain excitability. *Annals of Neurology, 60*, 188-196.

Hauser, T. U., Rotzer, S., Grabner, R. H., Mérillat, S., & Jäncke, L. (2013). Enhancing performance in numerical magnitude processing andmental arithmetic using transcranial direct current stimulation (tdcs). *Frontiers in Human Neuroscience, 7*, 244.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124.

Lipsey, M. W. (1990). *Design sensitivity : Statistical power for experimental research*. Newbury Park, CA: Sage Publications.

Maddock, J. E., & Rossi, J. S. (2001). Statistical power of articles published in three health psychology-related journals. *Health Psychology, 20*(1), 76-78.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44*(5), 423-432.

Tremblay, S., Lepage, J.-F., Latulipe-Loiselle, A., Fregni, F., Pascual-Leone, A., & Théoret, H. (2014). The uncertain outcome of prefrontal tdcs. *Brain Stimulation, 7*, 773-783.

Ware, J. J., & Munafò, M. R. (2014). Significance chasing in research practice: Causes, consequences and possible solutions. *Addiction, 110*, 4-8.