Euréka! (p<.05)

~ ou ~

Les tests de signification statistique: problèmes et solutions

Dominic Beaulieu-Prévost
Département de psychologie
Université de Montréal

1

L'importance des statistiques en science

L'évaluation empirique d'hypothèses représente un élément central des recherches en psychologie et dans les domaines connexes.

Dans la très grande majorité des recherches, ce processus d'évaluation s'argumente autour d'un test de signification statistique.

D'une certaine façon, la procédure statistique EST le processus d'évaluation.

2

L'importance des statistiques en science

Ce modèle, basé sur le rejet ou non de l'hypothèse nulle, est utilisé depuis maintenant plus de 50 ans.

Par contre, des critiques de plus en plus nombreuses (p.e. Cohen, 1994; Kline, 2004) soulignent des failles majeures l'invalidant comme modèle d'évaluation d'hypothèse et mettant en doute sa capacité à répondre à la majorité de nos questions de recherche.

3

Plan d'exposé

1) Le modèle traditionnel

2) Les problèmes majeurs

3) Des solutions faciles

4

Le modèle traditionnel

5

Le modèle traditionnel

1) Définir les hypothèses concurrentes

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

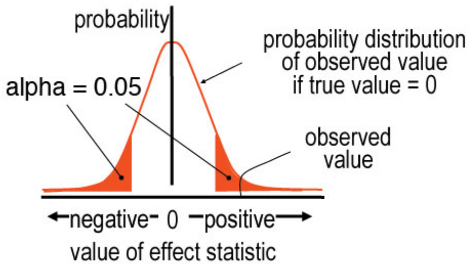
2) Choisir un critère de décision probabiliste (alpha)

Ex: $\alpha = 0,05$

6

Le modèle traditionnel

3) Comparer les données au modèle (H_0) et décider de rejeter ou non H_0 en comparant la valeur du *p* value à l'alpha.



7

Le modèle traditionnel

4) Et idéalement, tenir compte du risque d'erreur de type II par une évaluation de la puissance statistique.

	H_0 est vraie	H_0 est fausse
Rejeter H_0	Type I	Correct
Accepter H_0	Correct	Type II

8

Que peut-on savoir à partir d'un test statistique?

A. La probabilité que H_0 soit fausse.

B. La probabilité que H_1 soit vraie.

C. La probabilité de faire une erreur de type I si l'on rejette H_0.

D. La probabilité qu'une réplication expérimentale produise un résultat significatif (en calculant 1-*p*).

E. La probabilité que la décision (rejeter H_0 ou non) soit correcte.

F. La probabilité que les données soient le résultat du hasard.

G. La probabilité d'obtenir des résultats aussi extrêmes en supposant que H_0 soit vraie.

9

Que peut-on savoir à partir d'un test statistique?

Différencier $p(D/H)$ et $p(H/D)$

<u>Analogie</u>:

- Quelles sont les probabilités d'être mort si on a été pendu?
- Et quelles sont les probabilités d'avoir été pendu si on est mort?

10

Les problèmes majeurs

11

Les problèmes majeurs avec le modèle

1. La relation à N
2. L'improbabilité logique de H_0
3. Le manque de plausibilité de H_0
4. Conséquences...

12

Les problèmes majeurs avec le modèle

La relation à N (taille de l'échantillon)

p est relié à la fois à la taille de l'effet et à la taille de l'échantillon.

Conséquence 1

Pour une taille d'effet spécifique, *p* est un indicateur de la taille de l'échantillon.

Conséquence 2

Vous aurez TOUJOURS un résultat statistiquement significatif si la taille de votre échantillon est assez grande.

Les problèmes majeurs avec le modèle

L'improbabilité logique de H_0

La non-équivalence de H_0 et H_1

(hypothèses "ponctuelle" vs "par intervalle")

L'improbabilité d'une hypothèse ponctuelle sur une échelle continue.

La précision d'une hypothèse limite sa probabilité logique d'être vraie.

Conséquence

H_0 est TOUJOURS fausse ($1/\infty$), H_1 est TOUJOURS vraie et la probabilité d'erreur de type I est nulle ($1/\infty$).

Les problèmes majeurs avec le modèle

Le manque de plausibilité de H_0

La notion de bruit corrélationnel ambiant (Lykken, 1968) ou de "crud factor" (Meehl, 1990)

Les problèmes majeurs avec le modèle

Conséquences...

Nous essayons de rejeter une hypothèse que nous savons déjà être fausse!

C'est essayer d'augmenter notre confiance dans l'hypothèse "scientifique" en infirmant une hypothèse assurément fausse.

Des solutions faciles

Comment s'en sortir !?

Retourner à la base:

Qu'essayez-vous de faire?

- Construire un modèle? (estimation de paramètres)
- - ou -
- Évaluer un modèle? (test d'hypothèse)

Estimation de paramètres

L'intervalle de confiance, un outil de base

Le modèle de base: IC= M +/- Vc * ET

M= Taille de l'effet dans l'échantillon

Vc= Valeur critique pour le niveau de confiance précisé

ET= Erreur-type

Au lieu de simplement spécifier si la taille de l'effet est différente de zéro, les ICs nous donnent accès à :

- la taille de l'effet
- la précision de l'estimé

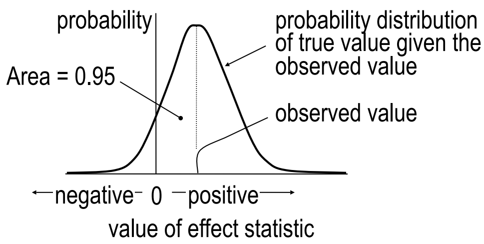
19

Estimation de paramètres

L'intervalle de confiance, un outil de base

Il est possible de calculer des ICs pour des:

- moyennes
- corrélations
- proportions
- différences



20

Estimation de paramètres

L'intervalle de confiance, un outil de base

Il est possible de calculer des ICs pour des:

- moyennes
- corrélations
- proportions
- différences

En fait, les ICs sont mathématiquement équivalents aux tests statistiques. Il est donc possible de les calculer pour tous les test statistiques existants!

21

Comment interpréter un IC

Selon une interprétation fréquentiste:

- L'intervalle des valeurs considérées équivalentes (étant donné l'erreur échantillonale) avec un niveau de confiance de 95%.
- Une estimation de la moyenne populationnelle avec un niveau de confiance de 95%.

Si on pouvait des ICs pour une infinité d'échantillons aléatoires provenant de la même population, la moyenne de la population serait incluse dans 95% d'entre eux.

Un niveau de confiance de 95% représente donc une *fréquence relative à long-terme* et non l'idée qu'il y a 95% de chances que la moyenne populationnelle soit incluse dans l'intervalle.

Comment interpréter un IC

Selon une approche bayésienne:

Il est possible de calculer un IC dans lequel il y a 95% de chances que la moyenne populationnelle soit incluse.

- Cela nécessite de calculer un intervalle de confiance à priori représentant les connaissances accumulées en plus des données de l'expérience. C'est un principe similaire à une méta-analyse.
- Mais si on postule un manque de connaissance à priori, l'IC bayésien coïncide avec son homologue fréquentiste. C'est ce que l'on appelle un *à priori agnostique* ou non-informatif.

Comment interpréter un IC

En résumé:

Les ICs traditionnels représente donc des intervalles qui ont 95% de chances d'inclure la moyenne populationnelle à la condition que l'on postule un *à priori agnostique*.

Par extension, on peut considérer la distribution reliée à l'IC comme la distribution des valeurs probables.

Les différents ICs

Formules et exemples d'intervalles de confiance:

1) Les ICs de moyennes

2) Les ICs de corrélations

3) Les ICs de proportions

4) Les ICs de différences

Les ICs de moyennes

IC(95%)= Moyenne +/- t_c * ET

ET= $\frac{\text{Écart-type}}{n^{1/2}}$

Pour connaître la valeur critique de t, voir un tableau à la fin d'un livre d'introduction aux statistiques.

Exemple:
M=10,00
ET=2,00
n=50 (dl=49)

Résultat:
$t_{c(49)}$=2,01
IC(95%)=10,00+/- 4,02

Sur SPSS, ces données sont accessibles par "*Descriptive*"

Les ICs de moyennes

IC(95%)= Moyenne +/- t_c * ET

ET= $\frac{\text{Écart-type}}{n^{1/2}}$

<u>IMPORTANT</u>
Vous devez normaliser vos données avant de calculer un IC si vous ne voulez pas perdre de précision. Pour les fins de l'interprétation, vous n'avez qu'à retransformer les limites de de l'IC en données originales.

Les transformation les plus utilisées (carré, cube, racine, ln,...)

Les ICs de corrélations

On ne peut pas calculer directement d'IC pour une correlation pcq la distribution des valeurs probables varie en fonction de la taille de la corrélation.

La solution la plus simple consiste à :

1) Transformer la corrélation en z' de Fisher

2) Calculer l'IC pour le z'

3) Retransformer les valeurs limites de l'IC en en corrélation

28

Les ICs de corrélations

1) Transformer la corrélation en z' de Fisher

$$z' = \frac{\ln(1+r) - \ln(1-r)}{2}$$

2) Calculer l'IC pour le z'

IC(95%)= z' +/- 1,96*ET

$$ET = \frac{1}{(n-3)^{1/2}}$$

3) Retransformer les valeurs limites de l'IC en en corrélation

$$r = \frac{e^{(2*z')} - 1}{e^{(2*z')} + 1}$$

29

Les ICs de corrélations

Heureusement, ces calculs se programment facilement sur Excel (voir http://www.projetmemoire.info pour un calculateur d'IC à télécharger).

À partir d'une feuille de calcul, vous n'avez qu'à entrer r et n.

Exercice:

Estimez les ICs (tous les r sont statistiquement significatifs)

r	n	IC(95%)
0,30	50	[à]
0,60	50	[à]
0,30	250	[à]

30

Les ICs de proportions

Pour tenir compte du fait qu'une proportion est calculée à partir d'un ratio de fréquences, il faut ajouter à l'IC une correction pour continuïté (C).

IC(95%)= p +/- (1,96*ET + C)

$$ET = \frac{p(1-p)}{n}$$

$$C = \frac{0,5}{n}$$

Cet IC est l'équivalent mathématique d'un test de chi-carré.

Voir http://www.projetmemoire.info pour un calculateur d'IC à télécharger.

31

Les ICs de proportions

Exemple:
Dans un test à l'aveugle fait avec un échantillon représentatif de 300 individus, 167 personnes (55,7%) ont préféré le produit A et 133 (44,3%) ont préféré le produit B. Que peut-on dire à propos des proportions dans la population?

Un chi-carré nous indique que :
- X^2= 3,63 et p=0,05
- Il y a statistiquement (mais de justesse) plus d'individus qui préfèrent A.

L'intervalle de confiance nous indique que :
- IC(95%)=[0,50 – 0,62]
- Entre 50% et 62% des individus préfèrent A.

32

Les ICs pour différences
(entre 2 groupes indépendants)

IC= M_2-M_1 +/- Vc*ET

ET= $(ET_1^2 + ET_2^2)^{1/2}$

Pour une différence de moyennes (équivalent d'un test-t)

IC(95%)= M_2-M_1 +/- t_c*ET_Δ

L'IC d'une différence de moyenne et l'erreur-type de la différence sont automatiquement produits lorsque vous faites un test-t sur SPSS.

33

Les ICs pour différences
(entre 2 groupes indépendants)

$IC= M_2-M_1$ +/- $Vc*ET$ $ET= (ET_1^2 + ET_2^2)^{1/2}$

Pour une différence de corrélations

$IC(95\%)= z_2'-z_1'$ +/- t_c*ET_Δ

$ET_\Delta= ((n_1-3)^{-1}+(n_2-3)^{-1})^{1/2}$

Voir http://www.projetmemoire.info pour un calculateur.

Pour une différence de proportions
- même principe

34

Les autres types d'ICs

Le principe est le même pour tous les autres types d'ICs possibles. Il est quand même préférable de vérifier si certains ajustements sont nécessaires avant d'utiliser un type d'IC pour la 1re fois.

Par exemple, un IC pour des mesures répétées sera plus petit que pour des groupes indépendants (même logique que pour les tests-t).

Il est à noter que les logiciels statistiques (p.e. SPSS) fournissent de plus en plus les ICs reliés aux tests que vous effectuez. Cela vous évite d'avoir à les calculer.

Il existe aussi plusieurs calculateurs d'ICs sur internet que vous pouvez trouver par Google (calculator confidence intervals ...)

35

Les tests d'hypothèse

Plusieurs tests d'hypothèse peuvent être effectués *sans calcul additionnel* à partir d'un IC.

Le principe est le même pour tous les types d'ICs. Il suffit de définir les hypothèses qui nous intéressent.

Les principales hypothèses d'intérêt

1) L'hypothèse nulle

2) L'hypothèse d'un effet substantiel

3) L'hypothèse d'un effet néfaste

4) L'hypothèse d'un effet non-pertinent

36

La logique des tests d'hypothèse

<u>Le principe de falsification</u>

L'interprétation actuelle en statistique:

- Falsifier une hypothèse fausse et unique.

Retour aux sources (Popper, 1963)

- Quelles hypothèses doit-on falsifier?
  (notions d'intérêt, de probabilité et de plausibilité)
- Le principe des fortes inférences (Platt, 1964):
  L'importance des hyp. concurrentes pour un processus de falsification efficace.

L'hypothèse nulle

Quoique ce soit rarement pertinent (H_0 étant improbable comme toutes les hypothèses ponctuelles), il est possible d'effectuer un test d'hypothèse nulle à partir d'un IC .

Le test est statistiquement significatif si l'IC exclu zéro.

<u>Exemple</u> provenant d'une différence de moyennes

IC(95%)= -1,32 à 3,47

En regardant l'IC on peut conclure que les données sont probables en supposant que l'effet est nul (p>0,05).

L'hypothèse d'un effet substantiel

La notion d'effet substantiel

Quantifier la taille minimal d'un effet substantiel (ou la taille maximale d'un effet non-substantiel) en fonction de :

- L'importance théorique de l'effet
- L'application prévue du phénomène
- Le coût potentiel d'une intervention
- Et, minimalement, la sensibilité de l'échelle!

L’hypothèse d’un effet substantiel

Lorsque l’*effet substantiel minimal* est quantifié, il peut être comparé à l’IC.

Si la valeur de la *limite inférieure de l’IC* est supérieure à l’*effet substantiel minimal*:

p(présence d’effet) > 0,95 $H_{présence}$ corroborée
p(absence d’effet) < 0,05 $H_{absence}$ falsifiée

Si la valeur de la *limite supérieure de l’IC* est inférieure à l’*effet substantiel minimal*:

p(présence d’effet) < 0,05 $H_{présence}$ falsifiée
p(absence d’effet) > 0,95 $H_{absence}$ corroborée

40

L’hypothèse d’un effet substantiel

Souvent, deux tailles d’effet peuvent être quantifiées:

- La taille minimale pour considérer que l’effet est *théoriquement intéressant*;
- La taille minimale pour considérer que l’effet a une *application pratique*.

41

L’hypothèse d’un effet substantiel

Exemple (corrélation)
Effet observé: r=0,40
Nombre de cas: n=100
IC: 0,22 ≤ r ≤ 0,55

Effet théoriquement intéressant: r ≥ 0.10 (c.-à-d. 1% de v. e.)
Effet cliniquement intéressant: r ≥ 0.30 (c.-à-d. 9% de v. e.)

Conclusions
p(présence d’effet théorique) > 0,95 $H_{présence}$ corroborée
p (absence d’effet théorique) < 0,05 $H_{absence}$ falsifiée
p (présence d’effet clinique) < 0,95 $H_{présence}$ non-corroborée
p (absence d’effet clinique) > 0,05 $H_{absence}$ non-falsifiée

42

L’hypothèse d’un effet substantiel

De façon plus claire, on peut dire que trois conclusions sont possibles pour chaque hypothèse:
1) L’hypothèse est corroborée (p > 0,95)
2) L’hypothèse est falsifiée/infirmée (p < 0,05)
3) L’hypothèse est indéterminée (0,05 < p < 0,95)

<u>En reprenant l’exemple précédent, on a:</u>

$p_{\text{(effet théorique)}} > 0{,}95$ $H_{\text{théorique}}$ corroborée
$0{,}05 < p_{\text{(effet clinique)}} < 0{,}95$ H_{clinique} indéterminée

La notion d’indétermination répond à la question de la puissance statistique.

L’hypothèse d’un effet néfaste

La possibilité d’un effet néfaste ou contraire à nos attentes (p.e. une mauvaise thérapie) peut être évaluée en utilisant l’*opposé de l’effet substantiel minimal*.

◆C’est une façon d’évaluer les *risques d’erreur importante* qui sont encourus (p.e. une mauvaise thérapie).

Si la valeur de la *limite inférieure de l’IC* est supérieure à l’*effet néfaste minimal*:

$p_{\text{(présence d'effet néfaste)}} < 0{,}05$ $H_{\text{néfaste}}$ falsifiée ☺

Si la valeur de la *limite supérieure de l’IC* est inférieure à l’*effet néfaste minimal*:

$p_{\text{(présence d'effet néfaste)}} > 0{,}95$ $H_{\text{néfaste}}$ corroborée ☹

L’hypothèse d’un effet non-pertinent

Un effet non-pertinent est un effet dont la taille se situe entre l’*effet néfaste minimal* et l’*effet substantiel minimal*.

Si l’IC est *totalement inclus* dans cet intervalle:

$p_{\text{(non-pertinence)}} > 0{,}95$ $H_{\text{non-pertinence}}$ corroborée

Si l’IC est *totalement exclu* de cet intervalle:

$p_{\text{(non-pertinence)}} < 0{,}05$ $H_{\text{non-pertinence}}$ falsifiée

◆Si on compare deux moyennes, cela revient à évaluer la l’hypothèse que les des moyennes sont équivalentes. C’est donc un <u>*test d’équivalence*</u> (Rogers, Howard & Vessey, 1993).

L'hypothèse d'un effet non-pertinent

Un effet non-pertinent est un effet dont la taille se situe entre l'*effet néfaste minimal* et l'*effet substantiel minimal*.

Si l'IC est *totalement inclus* dans cet intervalle:
$p_{\text{(non-pertinence)}} > 0{,}95$ $\quad$ $H_{\text{non-pertinence}}$ corroborée

Si l'IC est *totalement exclu* de cet intervalle:
$p_{\text{(non-pertinence)}} < 0{,}05$ $\quad$ $H_{\text{non-pertinence}}$ falsifiée

◆Si on compare deux moyennes, cela revient à évaluer la l'hypothèse que les des moyennes sont équivalentes. C'est donc un <u>*test d'équivalence*</u>.

46

Autres hypothèses

En utilisant le même principe, vous pouvez tester n'importe quelle autre hypothèse. tant quelle est formulée en intervalle.

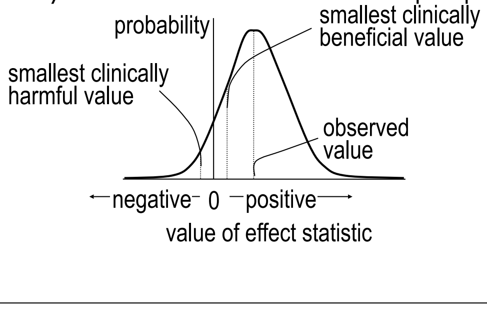
Les hypothèses par intervalle se testent selon le modèle présenté pour l'hypothèse de non-pertinence.
- ◆Vous pouvez, par exemple, évaluer l'hypothèse selon laquelle une corrélation se situerait entre 0,20 et 0,50.

Les hypothèses ponctuelles (déconseillées) se testent selon le modèle présenté pour l'hypothèse nulle.

47

L'estimation d'impact:
Une approche synthèse

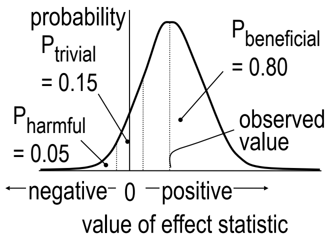
Au lieu de simplement évaluer les hypothèses par rapport au critère du 5% / 95%, il est possible de trouver les probabilités que ces hypothèses soient valides.
◆Ce type d'analyse est particulièrement approprié pour des méta-analyses cherchant à orienter des décisions pratiques.



48

L'estimation d'impact:
Une approche synthèse

Au lieu de simplement évaluer les hypothèses par rapport au critère du 5% / 95%, il est possible de trouver les probabilités que ces hypothèses soient valides (Hopkins, 2004).



49

L'estimation d'impact:
Une approche synthèse

<u>Pour les corrélations et les proportions</u>
Vous n'avez qu'à utiliser un tableau de distribution de z ou un calculateur électronique pour faire correspondre la valeur du z à l'aire sous la courbe.

<u>Pour les moyennes</u>
Vous devez utiliser un calculateur électronique pour faire correspondre la valeur du t et les degrés de liberté à l'aire sous la courbe.

Voir http://www.sportsci.org/resource/stats/index.html pour un calculateur à télécharger (http://www.sportsci.org/resource/stats/xcl.xls).

50

L'estimation d'impact:
Une approche synthèse

<u>Exemple</u>
r=0,30
n=50
I.C.= 0,02 - 0,53
$r_{(\text{effet substantiel minimal})}$=0,20

<u>Résultat</u>
$p_{(\text{effet substantiel})}$= 0,77
$p_{(\text{effet non-pertinent})}$= 0,23
$p_{(\text{effet néfaste})}$= 0,00

51

En résumé...

Mettez l'emphase sur :

- la taille de l'effet
- la précision de l'estimation (I.C.)

Opérationalisez vos hypothèses au lieu d'évaluer une H_0 inutile.

Comprenez ce que vous mesurez

Décidez ce qui est pertinent

... et jouez avec vos données !

Ressources additionnelles

1. La version PDF de cette présentation, disponible à www.projetmemoire.info
2. Le livre *Beyond significance testing* de Rex Kline (2004)