SES-Colibri

(Version 2.0)

Vous vous intéressez à **la taille des effets** :
- lors d'une recherche personnelle,
- lors de la lecture d'un article,

Mais vous ne disposez que du résultat d'un test classique (*t* de Student ou *F*).

Avec **SES-Colibri**, **saisissez uniquement** :
- la valeur du *F* (ou du *t)*,
- les degrés de liberté (*ddl*),
- l'effectif total (*N*_total).

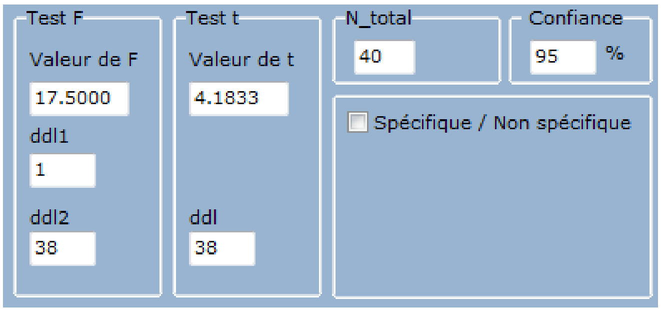
SES-Colibri calcule :
- les **indices** (calibrés/standardisés) de taille d'effet (*Eta*², *R*², *d* de Cohen, *Omega*²...),
- les **intervalles de confiance** sur ces indices calibrés ou sur un indice « brut » (différence de moyenne ou effet d'un contraste).

> *La seule prise en compte du résultat du test classique conduit souvent à des erreurs d'interprétation. Lors de la comparaison de moyennes, un résultat « significatif » est souvent interprété, à tort, comme la preuve que les moyennes diffèrent significativement (sous entendu de manière importante ou digne d'intérêt) et un résultat « non significatif » est souvent interprété comme la preuve d'une absence de différence entre les moyennes.*
> *Le calcul de la taille de l'effet aide à interpréter correctement le résultat du test et permet de compléter l'analyse classique.*

Diffusion gratuite : Delta-Expert : www.delta-expert.com
Auteur : Denis Corroyer (denis.corroyer@parisdescartes.fr)

Premier essai

Lancer **SES-Colibri**.
Des valeurs sont affichées à titre d'exemple :
$F = 17.50$, $ddl1 = 1$, $ddl2 = 38$, $t = 4.1833$ ($t^2 = F$), N_total = 40, Confiance = 95%.



Toutes ces valeurs seront bien entendu modifiées par l'utilisateur.

Pour un premier essai, cliquer sur « Calculer » :

SES-Colibri affiche :
- le résultat du test classique : p = .000163,
- plusieurs lignes correspondant à différents indices.

Prenons l'exemple du rapport de corrélation *Eta*². On obtient :
- le *Eta*² partiel observé (obs) dans l'échantillon : 32%
- les deux limites de l'intervalle de confiance sur cet indice : *IC* 95% = [9% ; 50%],
- les valeurs limites indicatives d'un effet faible et d'un effet fort : 4% et 16% pour *Eta*².

Résultat du test (F or t) classique :
p = .000163

IC pour Différence ou contraste sur des moyennes

Indice (partiel)	**obs**	Indice_Référence	**IC_inf**	**IC_sup**	IC_Référence	faible si <	fort si >
Eta² = R²	**32 %**	Cohen(1973)	**9 %**	**50 %**	Smithson(2003) p.43-44	4%	16%

Nous verrons plus loin, sur plusieurs exemples, comment interpréter ces valeurs.

Cliquer sur le menu Affichage – Estimateurs ponctuels (indices corrigés) pour obtenir d'autres indices.

Les indices de taille d'effet

Les indices calibrés

SES-Colibri fournit par défaut les indices suivants : Rapport de corrélation (*Eta*² et *Eta)*, Coefficient de détermination (R^2 et *R)*, rapport signal-bruit (f^2 et *f*).
En cliquant sur le menu Affichage – Estimateurs ponctuels (indices corrigés) on en obtient d'autres : *Omega*² et *Omega*, f^2_hat, *f*_hat, *EC* de Rouanet. Comme on pourra le vérifier sur les différents exemples, ces différents indices conduisent, pour l'essentiel, aux mêmes conclusions.
Dans le cas d'une comparaison à un degré de liberté, on obtient également le *d* de Cohen et sa version corrigée, le *g* de Hedges, ainsi que le coefficient de corrélation *r*.

Précisons enfin que tous ces indices sont des indices « partiels ».

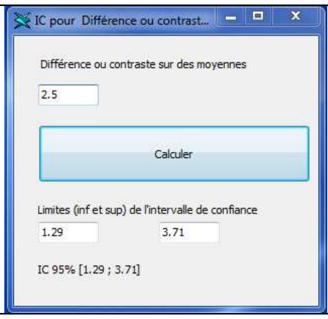
Les indices bruts

Dans le cas d'une comparaison à un degré de liberté, **SES-Colibri** permet également d'obtenir l'intervalle de confiance sur la différence de deux moyennes ou sur l'effet d'un contraste.

- Cliquer sur le bouton suivant :

IC pour Différence ou contraste sur des moyennes

- Indiquer la différence des moyennes des deux groupes : ici 2,
- Cliquer sur « Calculer »,
- Le logiciel indique les deux valeurs limites de l'intervalle et l'écriture usuelle de l'*IC* :
IC 95% = [1.03 ; 2.97].



Les valeurs repères

SES-Colibri propose deux valeurs repères permettant de se prononcer sur l'importance de l'effet (faible/modéré/fort). Ces valeurs s'appuient sur les valeurs proposées par Cohen (1988) et Rouanet (1996).

Le calcul des intervalles de confiance

Pour le calcul des intervalles de confiance sur les effets calibrés, **SES-Colibri** s'appuie principalement sur deux auteurs, Smithson et Steiger (Smithson, 2003, Steiger, Fouladi, 1997, Steiger, 2004). D'autres références sont données à la fin de ce document.

Précisions pour certains indices

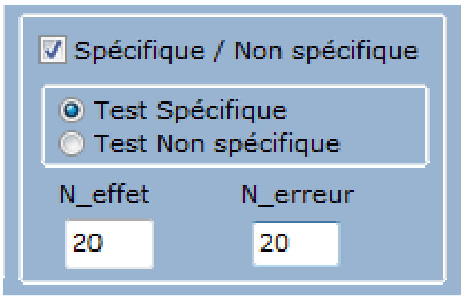
Pour le calcul du *d* de Cohen et du *g* de Hedges, les calculs sur ne sont valables que dans le cas de la comparaison de groupes équilibrés (de mêmes effectifs).

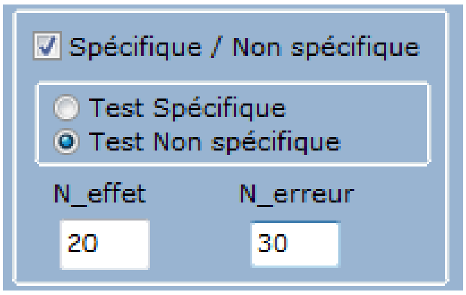
Pour le coefficient de corrélation de Bravais-Pearson (*r*) il n'est pas possible de déterminer le signe de la corrélation observée à partir du résultat du test. Dans le cas où la corrélation observée est négative, il faut inverser les signes, et l'ordre, des limites de l'intervalle de confiance (*IC*_inf et *IC*_sup) calculées par **SES-Colibri**.

Test Spécifique / non spécifique

Dans le cas où la comparaison porte sur un sous-ensemble des observations, plusieurs tests peuvent être envisagés. Il faut sélectionner l'option « Tests spécifique / non spécifique » pour indiquer à **SES-Colibri** quel test a été effectué et indiquer :
- le nombre de sujets pris en compte dans la comparaison (N_effet),
- le nombre de sujets pris en compte pour le terme d'erreur (N_erreur).

Par exemple si on cherche à comparer deux groupes de 10 sujets alors que les données comportent trois groupes de 10 sujets, on saisira 20 et 20 pour le test spécifique, 20 et 30 pour le test non spécifique.



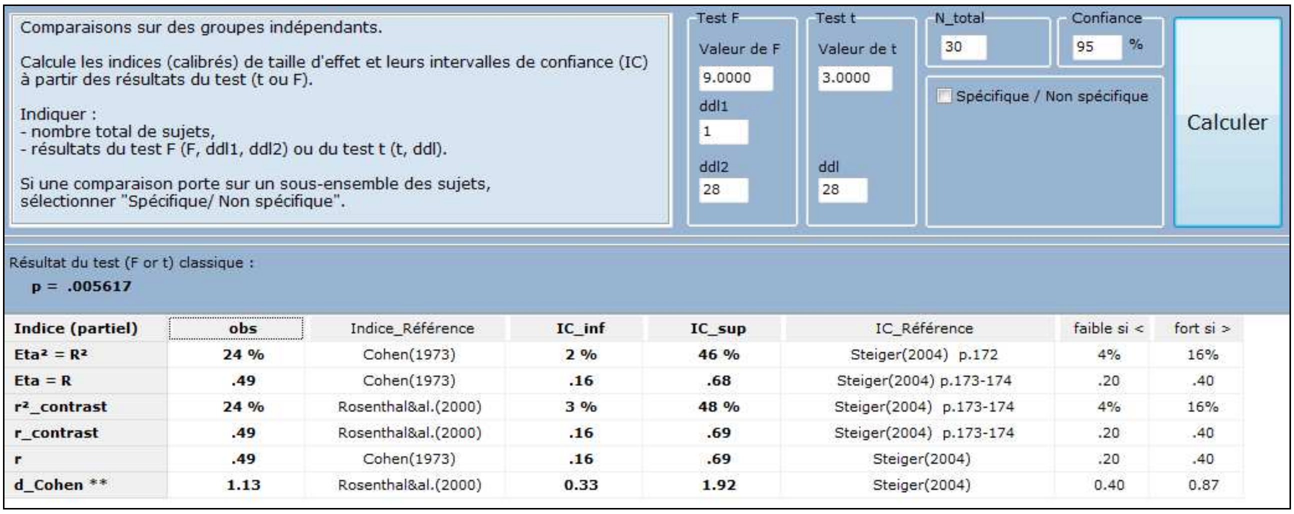


Interprétation des résultats

Voyons tout d'abord deux exemples :
- une situation (exemple 1) où le test est significatif mais où il n'est pas possible de conclure à un effet important,
- une situation (exemple 2) où le test est non significatif mais où il n'est pas possible de conclure à un effet faible.

Exemple 1 : Test significatif – Effet observé important - Effet parent indéterminé.

On compare deux groupes indépendants d'un effectif total de 30. On a obtenu, via un logiciel statistique : $F(1 ; 28) = 9.0$ ou $t(28) = 3.0$.
Saisir ces différentes valeurs dans **SES-Colibri.** Cliquer sur le bouton Calculer



Indice (partiel)	obs	Indice_Référence	IC_inf	IC_sup	IC_Référence	faible si <	fort si >
Eta² = R²	24 %	Cohen(1973)	2 %	46 %	Steiger(2004) p.172	4%	16%
Eta = R	.49	Cohen(1973)	.16	.68	Steiger(2004) p.173-174	.20	.40
r²_contrast	24 %	Rosenthal&al.(2000)	3 %	48 %	Steiger(2004) p.173-174	4%	16%
r_contrast	.49	Rosenthal&al.(2000)	.16	.69	Steiger(2004) p.173-174	.20	.40
r	.49	Cohen(1973)	.16	.69	Steiger(2004)	.20	.40
d_Cohen **	1.13	Rosenthal&al.(2000)	0.33	1.92	Steiger(2004)	0.40	0.87

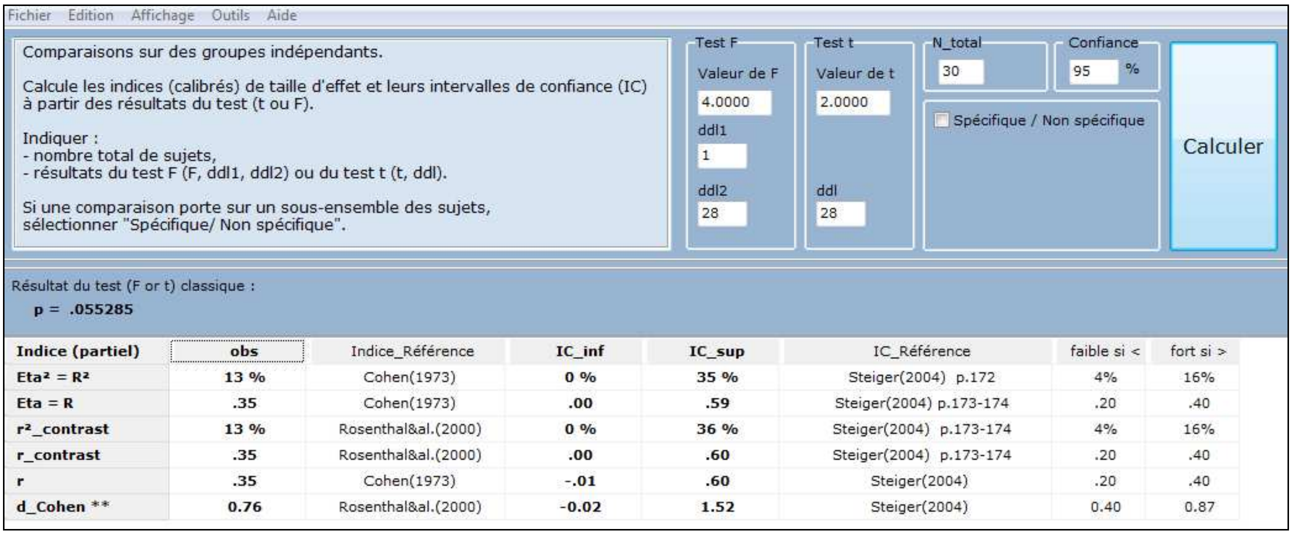
SES-Colibri rappelle le résultat du test classique (*F* ou *t*) :
$p = .005617$ (le test est significatif).

Mais, quel que soit l'indice retenu (prenons l'exemple de *Eta*²) on constate que :
- certes l'effet observé est important ($Eta^2 = 24\ \% > 16\%$)
- mais il existe une grande incertitude sur l'effet parent. En effet l'intervalle de confiance sur cet indice (*IC* 95% = [2% ; 46%]) recouvre des valeurs faibles (< 4%), modérées et importantes (> 16%). Il n'est donc pas possible de conclure à un effet parent important comme auraient pu le suggérer une lecture erronée du test (significatif) et du *Eta*² observé (important).

Remarque : on peut constater qu'on parvient à la même conclusion quel que soit l'indice utilisé.

Exemple 2 : Test non significatif – Effet observé modéré – Effet parent indéterminé.

Prenons un deuxième exemple (cf. copie d'écran ci-dessous) où l'on compare deux groupes indépendants avec un effectif total de 30. On a *F* (1 ; 28) = 4.0 ou, de manière équivalente, *t*(28) = 2.0.
On saisit ces différentes valeurs dans **SES-Colibri**, ainsi que le niveau de confiance (ici 95%).



Indice (partiel)	obs	Indice_Référence	IC_inf	IC_sup	IC_Référence	faible si <	fort si >
Eta² = R²	13 %	Cohen(1973)	0 %	35 %	Steiger(2004) p.172	4%	16%
Eta = R	.35	Cohen(1973)	.00	.59	Steiger(2004) p.173-174	.20	.40
r²_contrast	13 %	Rosenthal&al.(2000)	0 %	36 %	Steiger(2004) p.173-174	4%	16%
r_contrast	.35	Rosenthal&al.(2000)	.00	.60	Steiger(2004) p.173-174	.20	.40
r	.35	Cohen(1973)	-.01	.60	Steiger(2004)	.20	.40
d_Cohen **	0.76	Rosenthal&al.(2000)	-0.02	1.52	Steiger(2004)	0.40	0.87

Quel que soit le test utilisé (*F*(1 ; 28) = 4.0 ou *t*(28) = 2.0) le résultat est non significatif (*p* = .055285). Ce résultat peut suggérer au lecteur non averti un effet parent nul ou faible. Qu'en est-il en fait ?

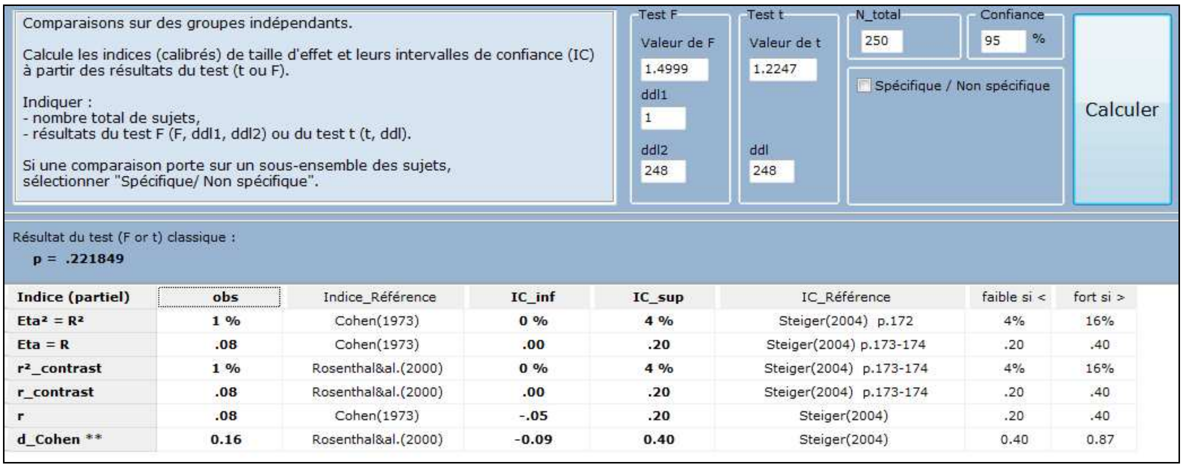
Prenons l'exemple de l'indice *d* de Cohen (qui peut varier entre -infini et +infini). On constate que :
- l'effet observé n'est pas nul, ni faible, mais modéré (0.40 < *d* de Cohen < 0.87),
- en ce qui concerne l'effet parent (*IC* 95% = [-0.02 ; 1.52]) il est certes possible qu'il soit nul (0 est dans l'intervalle de confiance) mais il est également possible qu'il soit important (*IC* contient des valeurs supérieures à 0.87). Il n'est donc pas possible de conclure à un effet parent nul ou faible comme aurait pu le suggérer le test non significatif.

Quand peut-on conclure à un effet faible ?

Les deux exemples suivant montrent deux situations où il est effectivement possible de
- conclure à un effet faible, suite à un test non significatif. (exemple 3),
- conclure à un effet important suite à un test significatif (exemple 4).

Exemple 3 – Test non significatif – Effet observé faible - Effet parent faible

On a recueilli cette fois (cf. copie d'écran ci-dessous) des données sur un grand effectif (250 sujets au total). Le test classique pour la comparaison de deux groupes indépendants donne $F(1\ ;48) = 1.4999$ ou, de manière équivalente, $t(248)= 1.2247$. On indique à **SES-Colibri** ces valeurs ainsi que le niveau de confiance (95%).



Indice (partiel)	obs	Indice_Référence	IC_inf	IC_sup	IC_Référence	faible si <	fort si >
Eta² = R²	1 %	Cohen(1973)	0 %	4 %	Steiger(2004) p.172	4%	16%
Eta = R	.08	Cohen(1973)	.00	.20	Steiger(2004) p.173-174	.20	.40
r²_contrast	1 %	Rosenthal&al.(2000)	0 %	4 %	Steiger(2004) p.173-174	4%	16%
r_contrast	.08	Rosenthal&al.(2000)	.00	.20	Steiger(2004) p.173-174	.20	.40
r	.08	Cohen(1973)	-.05	.20	Steiger(2004)	.20	.40
d_Cohen **	0.16	Rosenthal&al.(2000)	-0.09	0.40	Steiger(2004)	0.40	0.87

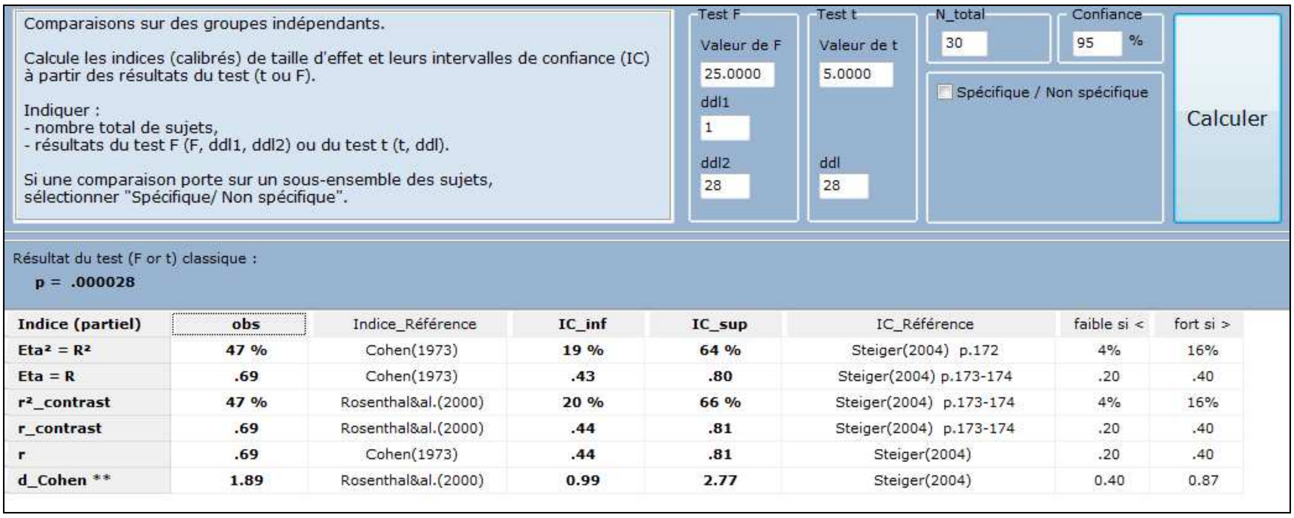
Le test classique est non significatif ($p = .22$). A la différence de l'exemple précédent, on pourra cette fois conclure à un effet parent faible, non pas sur la base du résultat à ce test, mais après l'analyse de la taille des effets.

Prenons l'exemple du *d* de Cohen :
- l'effet observé apparaît faible (*d_Cohen* = 0.16 < 0.40),
- il est également possible de conclure à un effet parent faible. En effet l'intervalle de confiance sur cet indice (*IC* 95% = [-0.09 ; +0.40]) ne comprend que des valeurs faibles (< 0.40 en valeur absolue).

Quand peut-on conclure à un effet important ?

Exemple 4 – Test significatif – Effet observé et effet parent importants

Revenons à un exemple avec un effectif total de 30 sujets.
On a obtenu, avec un autre logiciel, un $F(1 ; 28)$ égal à 25.0 cette fois (soit $t(28) = 5.0$).
On saisit ces différentes valeurs dans **SES-Colibri**, ainsi que le niveau de confiance (95% par défaut).



Indice (partiel)	obs	Indice_Référence	IC_inf	IC_sup	IC_Référence	faible si <	fort si >
Eta² = R²	47 %	Cohen(1973)	19 %	64 %	Steiger(2004) p.172	4%	16%
Eta = R	.69	Cohen(1973)	.43	.80	Steiger(2004) p.173-174	.20	.40
r²_contrast	47 %	Rosenthal&al.(2000)	20 %	66 %	Steiger(2004) p.173-174	4%	16%
r_contrast	.69	Rosenthal&al.(2000)	.44	.81	Steiger(2004) p.173-174	.20	.40
r	.69	Cohen(1973)	.44	.81	Steiger(2004)	.20	.40
d_Cohen **	1.89	Rosenthal&al.(2000)	0.99	2.77	Steiger(2004)	0.40	0.87

On a obtenu un test significatif (p = .000028). Ce test permet simplement de conclure qu'il existe bien, dans la population, une différence entre les deux conditions.

Considérons le coefficient de corrélation multiple, R (qui peut varier entre 0 et 1) tel qu'on l'obtiendrait avec une régression simple. On constate que :
- l'effet observé est important (R = .69 > .40).
- de même, à la différence de l'exemple 1, on peut conclure cette fois à un effet parent important. En effet l'intervalle de confiance sur cet indice (*IC* 95% = [.43 ; .80]) ne comprend que des valeurs de R importantes (> .40).

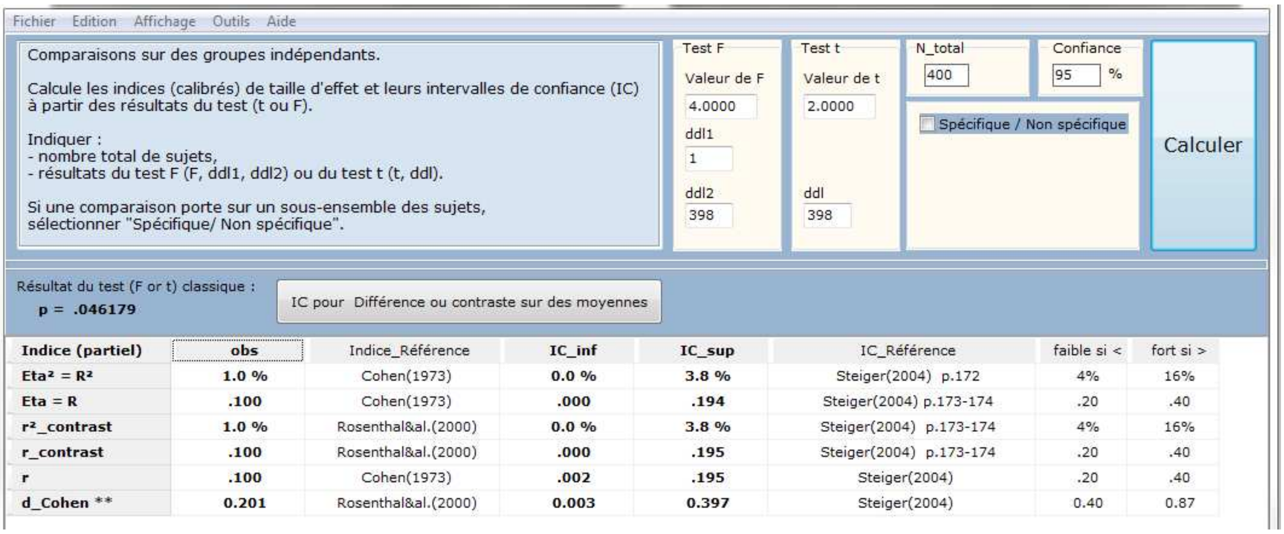
Remarque : On peut constater, comme pour les autres exemples, que la conclusion sera la même quel que soit l'indice utilisé.

Un test significatif et un effet faible !

Exemple 5 – Test significatif – Effet observé et effet parent faibles

Ce dernier exemple devrait définitivement montrer, nous semble-t-il, qu'un test significatif ne permet pas de conclure à un effet important.

En effet il est possible d'obtenir un test significatif et conclure, dès lors qu'on poursuit l'analyse, à un effet faible.



Indice (partiel)	obs	Indice_Référence	IC_inf	IC_sup	IC_Référence	faible si <	fort si >
Eta² = R²	1.0 %	Cohen(1973)	0.0 %	3.8 %	Steiger(2004) p.172	4%	16%
Eta = R	.100	Cohen(1973)	.000	.194	Steiger(2004) p.173-174	.20	.40
r²_contrast	1.0 %	Rosenthal&al.(2000)	0.0 %	3.8 %	Steiger(2004) p.173-174	4%	16%
r_contrast	.100	Rosenthal&al.(2000)	.000	.195	Steiger(2004) p.173-174	.20	.40
r	.100	Cohen(1973)	.002	.195	Steiger(2004)	.20	.40
d_Cohen **	0.201	Rosenthal&al.(2000)	0.003	0.397	Steiger(2004)	0.40	0.87

Cette situation, à première vue paradoxale, peut se produire dans le cas de grands échantillons. Dans cet exemple, on a recueilli des données sur un échantillon de 400 sujets.

Le test est significatif (F(1 ;398) = 4.00, p = .046).
L'effet observé est faible (*Eta*² = 1%). Il en est de même, semble-t-il, de l'effet parent car l'*IC* [0.0% ; 3.8%] ne comprend que des valeurs faibles (< 4%).
L'examen des autres indices conduit à la même conclusion.

Une dernière remarque :

Le test significatif conduit à la conclusion d'effet parent différent de 0. Or l'intervalle de confiance (à 95%) sur *Eta*² inclut la valeur 0, ce qui conduit cette fois à conclure qu'il est possible que l'effet parent soit nul.

En fait, pour les indices non signés (*Eta*², *R*², *Eta*, *R*…) *IC* doit être calculé avec le niveau de confiance 90% pour garder la cohérence avec le résultat du test.
On trouve alors, pour *Eta*² : IC 90% = [0.01% ; 3.22%]. La conclusion d'effet faible est toujours présente mais l'*IC* exclut un effet nul.

Références

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2 ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cohen, J. (1973). Eta-squared and partial Eta-squared in fixed Factor ANOVA designs. Educational Psychological Measurement, 33, 107-112.

Corroyer, D., & Rouanet, H. (1994). Sur l'importance des effets et ses indicateurs dans l'analyse statistique des données. Année Psychologique, 94(4), 607-624.

Corroyer, D., & Wolff, M. (2003). L'Analyse Statistique des Données en Psychologie; Concepts et Méthodes de base. Paris: Armand Colin (Cursus).

Fleishman, A.I. (1980). Confidence intervals for correlation ratios. Educational and Psychological Measurement, 40, 659-670.

Hays, W.L. (1994). Statistics (5th edition), Harcourt Brace College Publishers: Orlando.

Hedges, L. V. (1981). Distribution theory for Glass's Estimator of effect size and related estimators. Journal of Educational Statistics, 2, 107–128.

Lecoutre, B., & Poitevineau, J. (2000). Aller au-delà des tests de signification traditionnels: vers de nouvelles normes de publication. Année psychologique, 100, 683-713.

Rosenthal, R., Rosnow, R., L., & Rubin, D. B. (2000). Contrasts and Effect Sizes in Behavioral Research - A Correlational Approach. Cambridge: Cambridge University Press.

Rouanet, H. (1996). Bayesian methods for assessing importance of effects. Psychological Bulletin, 119(1), 149-158.

Smithson, M. (2003). Confidence intervals. London: Sage.

Steiger, J.H. (2004). Beyond the F Test : Effect Size Confidence Intervals and Tests of Close Fit in the Analysis of Variance and Contrast Analysis. Psychological Methods, 9 (2), p.164-182.

Steiger, J.H. & Fouladi, R.T. (1997). Noncentrality Interval Estimation and the Evaluation of Statistical Models. in Harlow, Mulaik, Steiger. What if there were no significance tests ? Mahwah, NJ: Erlbaum.

Tiwari, R., C., & Yang, J. (1997). Algorithm AS 318 : An efficient recursive algorithm for computing the distribution function and non-centrality parameter of the non-central *F*-distribution. The Journal of the Royal Statistical Society; Applied Statistics, 46, 418-413.