Intervalles de confiance sur 3 indices d'association pour un tableau de contingence *2 x 2*

Note technique - Denis Corroyer – 09/06/2006
Laboratoire de Psychologie Environnementale – Université Paris 5 – CNRS (UMR 8069)

Référence : Agresti (2002) – *Categorical Data Analysis*, Hoboken : John Wiley & Sons.

> *Remarque : seuls sont présentés ici les intervalles de confiance « standards » pour de grands échantillons (avec approximation normale). On se reportera à Agresti pour d'autres modes de calcul de ces intervalles de confiance et pour le calcul des intervalles de confiance exacts.*

Notations

En ligne une variable *G* définissant des groupes (g1 et g2)
En colonne une variable observée à deux modalités notées 1 et 0 (pouvant désigner par exemple des réussites et échecs). On s'intéresse aux « réussites » notées 1.

Effectifs conjoints : *n1g* désigne le nombre de 1 dans un groupe; *n1g1* désigne le nombre de 1 dans le groupe *g1*.
Effectifs marginaux : *ng* désigne l'effectif d'un groupe; *ng1* désigne l'effectif total du groupe *g1*.

	1	*0*	*Total*
g1	*n1g1*	*n0g1*	*ng1*
g2	*n1g2*	*n0g2*	*ng2*

Deux statistiques de base

Pourcentage (ou proportion ou fréquence) = *Pg1* = *n1g / ng*
Chances (ou *Odds*) = *Og1* = *n1g / n0g*

Exemple : Dossier CIVILITE

	1	*0*	*Total*
g1	222	144	366
g2	254	260	514

pg1 = n1g1 / ng1 = 61 %
pg2 = n1g2 / ng2 = 49 %

og1 = n1g1 / n0g1 = 1.54
og2 = n1g2 / n0g2 = 0.98

Trois indices d'association

Différence des Pourcentages, noté DP

DP = Pg1 -Pg2

Rapport des Pourcentages ou Risque relatif, noté RP

RP = Pg1 / Pg2

Odds Ratio ou Rapport des chances ou Cross-product ratio, noté OR

OR = Og1 / Og2

Exemple : Dossier CIVILITE

DP = 11 pts
RP = 1.2
OR = 1.6

Intervalle de confiance sur la différence des pourcentages (*DP*)

cf. Agresti, p. 72-77, 102, 110, 410-411

Distribution d'échantillonnage :

$$Moy\,(DP) = 0$$
$$Ety\,(DP) = \sqrt{\frac{Pg1\left(1 - Pg1\right)}{ng1} + \frac{Pg2\left(1 - Pg2\right)}{ng2}}$$

D'où $IC_{\alpha} = DP_{obs} \pm z_{\alpha/2} \times Ety(DP)$

Intervalle de confiance sur le rapport des pourcentages (*RP*)

cf. Agresti, p. 73, 77

Pour assurer une meilleure convergence de la distribution d'échantillonnage vers une distribution normale, on calcule la transformée de l'indice *RP* par la fonction logarithme (ln)

Distribution d'échantillonnage de Log *RP* :

$$Moy\,(Log\ RP) = 0$$
$$Ety\,(Log\ RP) = \sqrt{\frac{1 - Pg1}{Pg1 \times ng1} + \frac{1 - Pg2}{Pg2 \times ng2}}$$

D'où $IC(\log RP)_{\alpha} = Log\ RP_{obs} \pm z_{\alpha/2} \times Ety(Log\ RP_{obs})$

Les valeurs limites de l'IC sur les log de *RP*, ainsi obtenues, sont ensuite traduites en valeurs de *RP* par la fonction *exp*, inverse de la fonction log.

Intervalle de confiance sur l'odds ratio (*OR*)

cf. Agresti, p. 71, 77-78, 99-102, 255, 256

Comme pour l'indice *RP*, on obtient une meilleure convergence de la distribution d'échantillonnage vers une distribution normale en calculant la transformée de l'indice *OR* par la fonction logarithme (ln)

Distribution d'échantillonnage de Log *OR* :

$$Moy\,(Log\ OR) = 0$$
$$Ety\,(Log\ OR) = \sqrt{\frac{1}{n1g1} + \frac{1}{n0g1} + \frac{1}{n1g2} + \frac{1}{n0g2}}$$

D'où $IC(\log OR)_{\alpha} = Log\ OR_{obs} \pm z_{\alpha/2} \times Ety(Log\ OR_{obs})$

Les valeurs limites de l'IC sur les log de *OR*, ainsi obtenues, sont ensuite traduites en valeurs de *OR* par la fonction *exp*, inverse de log.

Exemple : Dossier CIVILITE

Effectifs observés

	1	0	Total
g1	222	144	366
g2	254	260	514

Alpha = 5%
Z alpha = 1.96

pg1 = 60.7%
pg2 = 49.4%

og1 = 1.54
og2 = 0.98

DP

DP = 11.2%
Ety (*DP*) = 0.03373947
Demi IC = 0.06612811

IC (*DP*) = [4.6% ; 17.9%]

RP

RP = 1.2
Log (*RP*) = 0.20493305
Ety (Log *RP*) = 0.06134938
Demi IC (Log *R*P) = 0.12024251
IC (Log RP) = [0.08 ; 0.33]

IC (RP) = [1.1 ;1.4]

OR

OR = 1.6
Log (*OR*) = 0.45621145
Ety (Log *OR*) = 0.13867989
Demi IC (Log *OR*) = 0.27180742
IC (Log *OR*) = [0.18 ; 0.73]

IC (*OR*) = [1.2 ; 2.1]