

Quelle taille d'échantillon faut-il prendre ? ...

Alain Morineau

CISIA•CERESTA

1, avenue Herbillon 94160 Saint-Mandé

La question revient régulièrement et dans des contextes souvent très différents. La réponse est rarement simple. L'objet de ce billet n'est pas de traiter le problème. Je donne simplement deux petits tableaux dans le but de montrer la nature des problèmes et pour fournir un *ordre de grandeur* dans un cas particulier.

Plaçons-nous dans le cas de la comparaison de deux proportions. On dispose de deux échantillons qu'on supposera de même taille pour simplifier. Il y a un échantillon d'hommes et un de femmes. Le taux des OUI est-il le même chez les hommes et chez les femmes?

Pour réaliser le test, on se fixe le seuil α d'erreur de première espèce, par exemple $\alpha = 0,05$. On rejettera à tort l'égalité des taux dans 5% des cas.

Pour aborder le problème des tailles des échantillons (tailles égales ici), on doit se fixer ce que l'on entend par des taux différents. On peut par exemple se fixer le changement de taux en pourcentage du taux: si le taux réel des OUI est de 10% et si on veut détecter un changement de 25% du taux, on se donne le moyen de décider d'une *différence de taux* dès que l'on observe 12,5% de OUI.

Enfin, pour calculer les tailles d'échantillon, il reste à se fixer le taux β d'erreur de seconde espèce que l'on accepte. Si on choisit $\beta = 0,20$ par exemple, on accepte à tort l'égalité des taux dans 20% des cas.

Avec ces prémisses, on peut maintenant calculer¹ les tailles (supposées égales) des échantillons. Par exemple, d'après les tableaux de la page suivante :

- Il faut des échantillons de taille supérieure à 1000 pour détecter un changement de taux de 5 points autour de 20% aux seuils $\alpha = 0,05$ et $\beta = 0,20$ (tableau 1. dernière ligne).
- Il suffit de quelque 700 observations si on accepte un seuil $\beta = 0,40$ (ibidem).
- Dans le cas où on déclare que le taux change s'il dépasse sa valeur de 50% (ce qui est moins

¹ Il existe des logiciels commerciaux pour réaliser ce type de calculs. Mais tout programme de calcul de probabilités classiques (binomiales, normales, etc) peut faire l'affaire, par exemple le petit logiciel "LeProbabiliste" édité par le CISIA•CERESTA.