

# FICHE PRATIQUE : Constitution d'un échantillon



# CONSTITUTION D'UN ECHANTILLON : CONSEILS METHODOLOGIQUES

Auteur : Ihssane SLIMANI - (fiche élaborée dans le cadre du guide pratique d'audit de la fonction ACHAT)

L'auditeur détermine, pour chaque étape de l'audit nécessitant de recourir à un échantillon, la taille de celui-ci, l'unité d'échantillonnage (factures, commande, etc.) ainsi que ses principales caractéristiques.

## Etape 1 : choisir la méthode d'échantillonnage

- L'échantillon pourra alternativement être constitué sur une base statistique (échantillonnage aléatoire) ou non statistique (échantillonnage raisonné), en fonction des objectifs de l'audit :
  - dans le cas d'un **échantillon statistique**, l'auditeur souhaite obtenir une base raisonnable à partir de laquelle pourront être tirées des conclusions probantes concernant la population dans son ensemble. Il aura recours à ce type d'échantillonnage lorsqu'il entend faire un énoncé quantitatif au sujet d'une population plus grande en se fondant sur les résultats d'un échantillon ; cette approche n'est pertinente que pour une population assez importante (*cf. infra*)
  - **dans le cas d'un échantillon non statistique**, l'auditeur sélectionne l'échantillon en fonction d'une caractéristique donnée et n'a pas l'intention de tirer des conclusions au sujet d'autres cas que ceux qui sont dans l'échantillon.

L'échantillonnage non statistique ou raisonné donne à l'auditeur la flexibilité d'utiliser son opinion professionnelle pour sélectionner les éléments qui ont le plus besoin d'être testés. Le recours à ce type d'échantillon se justifie lorsque les énoncés généraux sont d'ordre qualitatif plutôt que quantitatif. Par exemple, on peut avoir l'intention de recenser les cas particuliers de dysfonctionnements possibles sans faire d'observation sur la fréquence avec laquelle ils se produisent.

*Pour les petites populations une démarche non statistique peut permettre d'obtenir des résultats significatifs : si sur 50 documents, on en a examiné 15 dont 10 sont irréguliers, on ne pourra déduire que 2/3 des documents sont irréguliers, mais on aura montré que 20% le sont effectivement. Dans un autre cas, si sur 50 documents, on en a examiné 15 tous réguliers cette fois-là, on ne pourra déduire que 100% des documents le sont mais on sera sûr que 30% sont réguliers, sans pouvoir exclure que 70% ne le soient pas.*

- Si l'approche statistique permet de soutenir des conclusions et des recommandations de manière plus crédible, sa constitution peut s'avérer difficilement compatible avec les contraintes de temps d'un audit. Elle doit notamment intégrer le risque d'échantillonnage (risque que la conclusion basée sur un échantillon soit différente de celle à laquelle il serait parvenu si l'ensemble de la population avait été soumis à la même procédure d'audit). Ce risque est notamment fonction :
  - **de la taille de la population**, à croiser avec les moyens dont dispose l'auditeur pour l'analyse, ainsi qu'avec la précision et la marge d'erreur acceptées de l'estimation. [Voir en annexe 1, le tableau pour déterminer automatiquement la taille de votre échantillon ;](#)
  - **de son caractère plus ou moins uniforme** au regard de la caractéristique observée. La conception d'un échantillon devient plus complexe dans le cas où les caractéristiques principales des membres de l'échantillon ne sont pas uniformément réparties dans toute la population (ex : petit nombre de commandes de montant élevé et grand nombre de commandes dont le montant relatif est très faible), rendant nécessaire le recours à des échantillons complexes (ex. échantillons aléatoires décomposés en strates de caractéristiques similaires).

*Cf. notamment sur ce point la norme ISA530 sur les sondages en audit.*

- Un croisement possible entre les deux approches peut consister à procéder en deux temps :
  - 1- Analyse de risques préalable permettant d'isoler une population « à risque ».** Cette analyse prendra en compte :
    - l'impact, ce qui peut notamment amener l'auditeur à la sélection de commandes d'un montant élevé ou encore de commandes de caractère stratégiques ;
    - la probabilité, ce qui peut conduire l'auditeur à constituer son échantillon à partir des points d'alerte les plus facilement et rapidement repérables aux différentes étapes de la procédure (cf. rubriques « points d'alerte dans chacune des fiches).
  - 2- Sélection d'un échantillon statistique**, destiné à être représentatif, non de la population dans son ensemble, mais du type de la population « à risque » isolé.

## **Etape 2 : constituer l'échantillon**

- Pour la constitution d'un échantillon probabiliste et la présentation des différentes techniques disponibles (aléatoire simple, aléatoire systématique, échantillonnage stratifié, l'échantillonnage en grappe à plusieurs degrés, etc.), l'auditeur pourra notamment se référer à l'annexe 2 de la présente fiche, ainsi qu'aux manuels d'utilisation des logiciels d'audit type IDEA, qu'il croisera utilement avec les supports existants de formation à l'audit.

# ANNEXE 1 : DETERMINATION DE LA TAILLE D'ECHANTILLON

Pour des raisons de coûts et de délai, il n'est pas toujours possible d'étudier l'ensemble de la population. Il est alors nécessaire de créer un échantillon représentatif de la population permettant ainsi de fournir une estimation aussi précise que possible d'une variable. La question de la taille de l'échantillon dépend de la contrainte de coût. Si cette contrainte est forte, l'échantillon correspond au rapport entre le budget global alloué à l'étude et le coût unitaire d'une enquête.

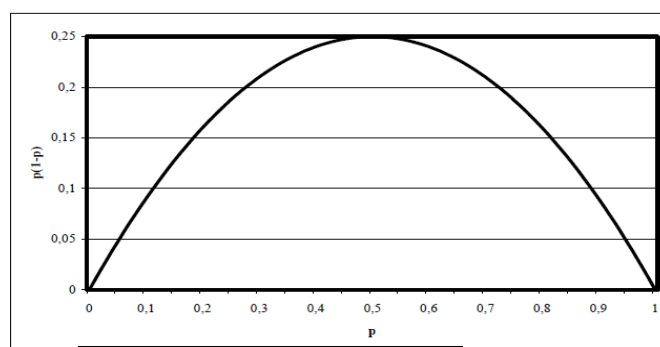
A l'inverse, si la contrainte de coût est faible, la taille de l'échantillon est liée à la précision souhaitée a priori de l'estimation. En effet, on ne peut pas dire qu'une certaine taille d'échantillon est suffisante pour étudier une population d'une certaine taille. La taille de l'échantillon se définit nécessairement par rapport au degré de précision exigé par l'enquêteur et au phénomène que l'on veut étudier.

De manière générale, plus l'échantillon est grand et plus l'estimation sera précise. La notion de précision est matérialisée par un seuil de confiance (en général 95%) et une marge d'erreur. Par exemple si l'on définit un seuil de confiance de 95% et une marge d'erreur de 2%, cela signifie que l'échantillon permettra d'extrapoler le résultat avec 5% de risques de se tromper de plus ou moins 2%.

La taille de l'échantillon dépend ainsi de plusieurs facteurs :

- La marge d'erreur acceptée, notée  $m$  ;
- Le risque maximum accepté, noté  $t$ . La valeur est issue d'une loi de probabilité (loi de Student pour une taille de population inférieure à 30, loi Normale pour une taille de population supérieure à 30)<sup>1</sup> ;
- La connaissance statistique préalable de la population enquêtée, paramètre inconnu qui peut correspondre à une estimation provenant d'études antérieures, de résultats d' « experts » et approximée, dans le cas d'une proportion, par  $p^*$  ( $1-p$ ). Si l'on n'a aucune piste quant à la valeur de  $p$ , il est admis de prendre la valeur qui donne la dispersion maximale (cf. graphique ci-dessous), à savoir  $p = 0,5$ .

La taille de l'échantillon, notée  $n$ , se calcule comme suit : 
$$n = \frac{t^2 \cdot p(1-p)}{m^2}$$



<sup>1</sup> La valeur de test respectivement de 2,58, 1,96 et 1,65 pour les seuils de confiance de 99%, 95% et 90%.

La taille de l'échantillon ne dépend ainsi pas de la taille de la population, sauf si celle-ci est très petite. En effet, lorsque la population est réduite, il est nécessaire d'appliquer un facteur correctif. La taille de l'échantillon devient :

$$n' = \frac{N \cdot n}{N + n}$$

Le tableau ci-dessous donne à titre d'illustration la taille d'échantillon minimale selon la taille de la population et la précision souhaitée, pour p valant 0,5.

Lorsque la population est très petite, par exemple N= 10 ou 15, la définition d'un échantillon n'a pas de sens statistiquement, il est nécessaire d'étudier l'ensemble des éléments pour avoir une estimation statistiquement représentative. Ce n'est que pour des populations supérieures à 100 que la notion d'échantillonnage a véritablement du sens statistiquement.

**Tableau : Taille d'échantillon en fonction de la taille de la population, de la précision et la marge d'erreur acceptées de l'estimation**

Niveau de confiance	Marge d'erreur	Taille de la population étudiée						
		10	15	30	50	100	150	200
99%	2%	10	15	30	50	98	145	191
	5%	10	15	29	47	87	123	154
	10%	10	14	26	39	63	79	91
95%	2%	10	15	30	49	96	142	185
	5%	10	15	28	45	80	108	132
	10%	10	13	23	33	49	59	65
90%	2%	10	15	30	49	95	138	179
	5%	10	15	27	43	73	97	115
	10%	9	13	21	29	41	47	51

Source : *Calculs mission*

Exemple de lecture du tableau : Si l'on considère une taille de population de 50 commandes à examiner avec un seuil de confiance de 95% (donc probabilité de 5% de se tromper), il faut une taille d'échantillon égale à 49 avec toutefois, une marge d'erreur de 2% (par exemple pour la moyenne du montant des commandes).

En revanche, si l'on accepte une marge d'erreur de 10% (par exemple pour la moyenne du montant des commandes), pour le même seuil de confiance (95%), la taille de l'échantillon à examiner n'est plus que de 33.

# ANNEXE 2 : RAPPEL DES PRINCIPALES METHODES DE CONSTITUTION DE L'ECHANTILLON

Il existe deux catégories de méthodes :

- **une méthode probabiliste** avec notamment la sélection par tirage aléatoire. Chaque individu statistique (i.e. un individu, une entreprise, une facture, etc.) a la même chance d'entrer dans l'échantillon. Parmi les méthodes probabilistes figurent :
  - l'échantillonnage aléatoire simple : facile à mettre en œuvre, il s'agit de tirer au hasard des individus qui ont tous la même probabilité d'être sélectionné.
  - l'échantillonnage aléatoire systématique : les individus de la population sont numérotés de 1 à N. On détermine l'intervalle d'échantillonnage k en divisant la population N par la taille de l'échantillon que l'on souhaite obtenir. On sélectionne un nombre qui correspond à l'origine choisie au hasard. Enfin, à partir de ce premier nombre, on sélectionne chaque kème individu.
  - l'échantillonnage stratifié : il s'agit de diviser la population en sous-ensembles homogènes, appelés strates, et de réaliser un sondage sur chacune des strates.
  - l'échantillonnage en grappe à plusieurs degrés : si la population est répartie sur M grappes, le 1er degré consiste à choisir un échantillon de M grappes et le 2nd degré à réaliser une enquête dans chacune des M grappes, soit auprès de tous les éléments, soit en constituant des échantillons. L'inconvénient est qu'il induit une perte de précision due à l'existence d'une similarité entre individus d'une même grappe.
- **une méthode non-probabiliste** : les individus statistiques sont sélectionnés en fonction de critères de répartition. Elle est notamment utilisée lorsqu'il n'est pas possible d'avoir une liste exhaustive de toutes les unités de sondage. Parmi les méthodes dites non-probabilistes figurent notamment :
  - l'échantillonnage par quotas : définir un nombre précis d'éléments à sélectionner pour diverses sous-populations ;
  - le volontariat, mais l'inconvénient est que cette technique crée un biais et n'est pas pertinente dans une démarche d'audit.