Gestion des déplacements : évaluations d'impact et tests de matériel : calcul de la taille des échantillons, application aux cas simples et usuels

Patrick Olivero

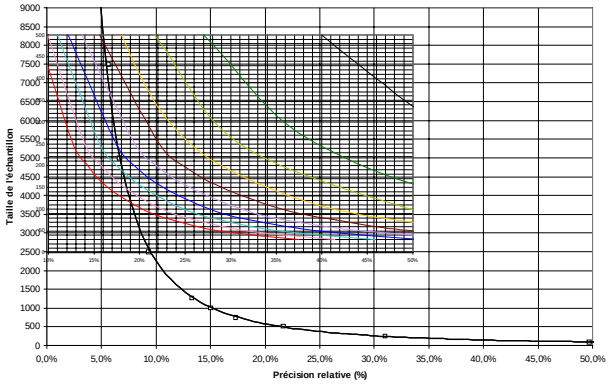
CETE du Sud-Ouest
12, av. Edouard Belin
31400 Toulouse
Tel. : +33 (0) 5 62 25 97 70
Fax : +33 (0) 5 62 25 97 99
E-Mail : zelt@equipement.gouv.fr

Gestion des déplacements

Evaluations d'impact et tests de matériel

Calcul de la taille des échantillons

Application aux cas simples et usuels



Patrick Olivero

(CETE du Sud-Ouest / DAT/ZELT)

Version 2.1

Avril 2001

Remerciements

M Jean Peybernard, chercheur au LCPC et professeur de statistique à l'ENTPE,

M. Nour-Eddin El Faouzi, statisticien, chercheur au laboratoire LICIT (INRETS / ENTPE),

ont bien voulu effectuer une lecture critique d'une première ébauche de ce document.

Leurs conseils ont permis de nombreuses clarifications et simplifications, et nous les remercions vivement pour cette contribution.

Sommaire

Avant-propos

Depuis plusieurs années, plusieurs travaux méthodologiques ont eu pour ambition de fournir aux techniciens des outils leur permettant de réaliser, ou de piloter, les études d'évaluation des matériels et systèmes d'exploitation.

Dans le domaine urbain et périurbain, ces travaux ont souvent été réalisés à l'initiative du CERTU, en particulier au sein d'un groupe de travail sur l'évaluation des opérations d'exploitation dites "SDER de niveau 1".[1]

Il est apparu à ce groupe de travail qu'un minimum de connaissances statistiques était nécessaire aux techniciens pour réaliser les plans d'expérience et interpréter les résultats. Pour actualiser et renforcer cette compétence, un stage de formation *Méthodes statistiques pour l'exploitation de la route*, assuré par M. Jean Peybernard (LCPC), a été organisé, et une première session a eu lieu en 2000.[2]

En complément, le CERTU a demandé à la ZELT de rédiger une note technique sur le calcul de la taille des échantillons, écueil sur lequel se heurtent parfois les expérimentateurs lorsqu'ils élaborent les plans d'expérience.

Le problème a été traité ici d'une manière pragmatique, c'est-à-dire en fournissant des méthodes ou outils (abaques) permettant de traiter la plupart des cas courants.

En particulier, nous n'avons pas développé le cas des petits échantillons et nous nous sommes placés délibérément dans l'hypothèse d'une taille d'échantillon supérieure à 30.

Nous avons conservé un découpage du problème en 2 sous-ensembles : mesure de la moyenne d'une variable continue d'une part ; mesure d'une fréquence (proportion) d'autre part. Nous n'ignorons pas que, moyennant certaines restrictions d'emploi, ces 2 situations peuvent être traitées de manière analogue par emploi de la loi normale ; toutefois, il nous a semblé préférable de fournir, pour les proportions, des outils développés à partir de la loi binomiale dont les conditions d'application sont très larges et non bridées par la condition usuelle np>20.

Nous avons privilégié ici l'utilisation d'abaques. Les outils informatiques qui ont permis leur établissement ont été fournis au CERTU.

---

[1] Groupe de travail dirigé successivement par JP Mizzi et F. Kunkel.

[2] D'autres sessions de ce stage sont programmées ou le seront. Contacter le CERTU à cet effet.

1 Généralités sur les problèmes de l'échantillonnage

La réalisation d'un test de matériel, ou d'une évaluation d'impact a généralement pour objectif de fournir des informations sur une variable (ou plusieurs) caractéristique du phénomène étudié ; par exemple[3] :

- Temps de parcours des véhicules entre deux points A et B.
- Pourcentage de véhicules franchissant un feu au rouge.
- Taux de détection d'un capteur.
- Taux de fausses alarmes d'un système de DAI. [4]
- Fréquence de passage d'un bus à son arrêt commercial.
- etc.

La nature des phénomènes qui nous occupent fait qu'il est impossible (indépendamment même de toutes considérations logistiques) de procéder à une étude exhaustive de la population concernée ; on procède donc par échantillonnage.

Plus précisément : l'expérimentateur choisit une période de temps, ou un nombre d'individus, qui sera le support des mesures ; le nombre d'individus observés constitue l'échantillon ; il doit être tel (autant que faire se peut) que les caractéristiques de la variable étudiée, établies sur l' échantillon, représentent également les caractéristiques de la population réelle.

Or, la satisfaction de cet « espoir » n'est pas certaine ; elle dépend de plusieurs facteurs, parmi lesquels :

- La pertinence et la qualité des mesures effectuées.
- La représentativité de l'échantillon.
- La taille de l'échantillon.

[3] Nous nous limitons ici, et ce sera le cas dans tout ce document, à des exemples qui concernent la gestion des déplacements.

[4] DAI : Détection Automatique d'Incidents.

1.1 La qualité des mesures

Nous ne développons pas ce point, qui relève du savoir-faire de l'expérimentateur et de la pertinence des méthodes de mesure qu'il utilise. Nous nous limitons à quelques considérations générales qui n'épuisent pas le problème :

- Lorsque des enquêtes par interview ou questionnaire sont utilisées, on prendra garde au biais dit « de complaisance », qui est la tendance du sujet interrogé à se positionner sur les réponses qu'il suppose attendues. Il est parfois possible de faire des hypothèses crédibles sur l'occurrence de ce biais[5]. Le biais contraire peut être qualifié de « biais militant » : il consiste à privilégier l'utilité supposée de la réponse au détriment de son objectivité. Ce dernier biais doit être appréhendé lors de la constitution de l'échantillon et relève du problème de la représentativité, problème traité plus loin.

- Lorsqu'un appareillage technique est utilisé pour effectuer des mesures, l'impression intrinsèque de la mesure est une donnée qu'il ne faut pas confondre avec l'incertitude résultant de la taille de l'échantillon. La première viendra toujours s'ajouter à la seconde. En d'autres termes, les bornes de l'intervalle de confiance seront également associées à un intervalle de précision résultant de la qualité métrologique de l'appareil.

- La précision d'un appareil de mesure peut également dépendre de facteurs extrinsèques éventuellement non maîtrisables : par exemple la météo, la luminosité, l'environnement électromagnétique, le savoir-faire de l'opérateur, etc.

- La précision du résultat peut parfois être entachée par des imprécisions (voire des erreurs de manipulation) résultant non pas de la mesure mais du traitement qui est fait de cette mesure. Par exemple : supposons que l'on étudie la répartition des temps de parcours, sur un trajet donné AB en milieu urbain, et que les données recueillies soient des dates de passage en A et en B ; supposons en outre que l'on dispose d'un logiciel capable d'apparier les dates de passage d'un véhicule donné et d'en déduire un temps de parcours[6]. Ce logiciel doit être capable d'éliminer les temps de parcours « anormaux » qui sont ceux de véhicules ayant effectué un arrêt de longue durée sur le trajet AB, arrêt dont on peut supposer qu'il n'est pas une conséquence des conditions de trafic, mais relève du libre choix du conducteur ou de circonstances fortuites (s'arrêter pour faire un achat ; s'arrêter pour cause de panne ou d'incident ; etc.). Le traitement des données doit donc être précédé par une phase de validation et de détection des valeurs aberrantes.

---

5 Des analyses de ce type ont été effectuées par la ZELT dans les travaux du programme européen CENTAUR. Cf. référence [6].

6 Cet exemple n'est pas fortuit : cette méthode et ces outils sont ceux utilisés par la ZELT, méthode dite « ZELT-PSION ».

1.2 La représentativité de l'échantillon

1.2.1 Généralités

Ce problème, important[7], relève lui aussi du savoir-faire de l'expérimentateur et de sa connaissance de la population étudiée. Il peut nécessiter le recours à des compétences plus spécialisées, qui sont celles d'organismes spécialisés dans les sondages. La position du problème est simple et n'appelle pas de longs développements : est-ce que l'échantillon possède des caractéristiques proches de celles de la population ?

Quelques exemples :

- On s'intéresse à la vitesse moyenne des véhicules sur une section donnée de VRU[8], et on ne mesure que la vitesse des véhicules circulant sur la voie de droite : on a toutes chances d'avoir un résultat par défaut.
- On s'intéresse aux temps de parcours de véhicules en les mesurant par insertion de véhicules dans le flot : l'échantillon obtenu ne sera représentatif que si le conducteur-enquêteur s'est astreint à un mode de conduite proche de celui de l'ensemble des véhicules.
- On s'intéresse au temps de parcours en milieu urbain, mais la période de mesures inclut des épisodes non récurrents (manifestations sociales, intempéries non usuelles, etc.).
- On s'intéresse au taux de violation d'un feu rouge mais on n'effectue les mesures que pendant des périodes de trafic dense : échantillon non représentatif car la probabilité de violation du rouge est certainement une fonction décroissante du taux d'occupation (hors saturation).
- On s'intéresse à la rentabilité d'un carrefour (nombre de véhicules écoulés par seconde de vert) mais les conditions de trafic pendant la période de mesure ne correspondent pas au champ optimum d'utilisation de l'algorithme de régulation.
- Etc.

Dans les exemples cités ci-dessus, des réflexions qui relèvent du simple « bon sens » permettent d'éviter les écueils.

Il n'en est pas de même pour les expériences utilisant des enquêtes (interviews ou questionnaires) : le problème est plus complexe, et nous allons nous y attarder quelque peu.

---

[7] L'essentiel de ce document est consacré à la détermination de la taille de l'échantillon. Ce chapitre relatif à la représentativité de l'échantillon est en quelque sorte un développement annexe succinct.

[8] VRU : Voie Rapide Urbaine.

1.2.2 Cas particulier des enquêtes d'opinion

1.2.2.1 Généralités

Les enquêtes d'opinion sont fréquemment utilisées dans des problèmes relatifs à la gestion des déplacements, pour apprécier l'acceptabilité d'un système par les usagers.

Par exemple :

- Lisibilité et compréhension de l'information routière.
- Opinion sur l'utilité, ou l'utilisation, d'un système quelconque.
- Analyse de facteurs expliquant les choix modaux.
- Opinion sur la qualité du service rendu.
- Enquêtes destinées à prévoir le comportement des usagers.
- Etc.

A notre sens, l'expérimentateur doit clairement choisir entre l'un ou l'autre des deux objectifs suivants :

L'objectif est d'avoir un avis d'expert

Dans ce cas l'échantillon sera constitué au sein du sous-ensemble de la population le plus directement concerné par le problème étudié.

Exemple : on veut étudier l'opinion de la population sur la qualité des aménagements destinés aux vélos ; si l'échantillon est constitué par une partie quelconque de la population, le taux de cyclistes pratiquants sera faible ; les réponses assises sur la pratique réelle du vélo seront minoritaires ; elles seront « noyées » dans la masse des réponses moins pertinentes émanant de cyclistes occasionnels, voire de non-cyclistes. Pour obtenir des réponses « d'expert », il faut constituer l'échantillon dans une population particulière, interviewée in situ (c'est-à-dire sur un vélo), ou constituée à partir d'un fichier d'association de cyclistes. Au sein de cette sous-population représentative, on pourra alors admettre que l'on effectue un sondage aléatoire, c'est-à-dire que tous les individus sont représentatifs.

L'objectif est d'avoir une opinion de l'ensemble de la population

Dans ce cas la précaution à prendre est de s'assurer que l'échantillon possède des caractéristiques représentatives de la population totale ou, plus précisément, que les variables que l'on contrôle au sein de l'échantillon sont celles qui sont susceptibles d'avoir une incidence sur les réponses fournies. En toute rigueur, ce problème est impossible à résoudre. En effet, ce n'est que lorsque l'enquête sera effectuée, que l'on pourra analyser l'ensemble des caractéristiques et déterminer celles qui sont pertinentes (c'est-à-dire qui sont explicatives des réponses) et qu'il aurait fallu contrôler. Fort heureusement, on peut souvent faire des hypothèses crédibles sur la nature des variables à contrôler, et construire l'échantillon en utilisant la méthode dite « des quotas », brièvement décrite ci-dessous.

1.2.2.2 Aperçu sur la méthode des quotas

La méthode des quotas est en fait la succession de 4 étapes distinctes :

1. On choisit les variables, dites « variables de contrôle » que l'on suppose corrélées avec les variables statistiques que l'on veut étudier.

Exemple : on étudie l'acceptabilité par l'usager d'un système de péage urbain et on fait l'hypothèse que les variables de contrôle sont le niveau de revenu, le lieu d'habitat et le lieu de travail. Ceci signifie que l'on suppose que l'acceptabilité du péage est expliquée par ces 3 variables. On peut se tromper : par exemple, il se pourrait que, à niveau de revenu identique, la population jeune soit plus réticente au principe du péage que la population plus âgée (ou le contraire) ; dans ce cas, il aurait fallu ajouter l'âge comme variable de contrôle supplémentaire. Dans d'autres cas, le sexe, la catégorie socioprofessionnelle (CSP), les préférences politiques, etc. doivent être prises en compte. Dans le doute, on pourra ajouter des variables de contrôle dont on n'est pas sûr de la pertinence ; mieux vaut cet excès de précaution que l'inverse. Toutefois, la multiplication des variables de contrôle rend plus difficile (et plus coûteuse) la constitution de l'échantillon, et il faut garder une juste mesure en la matière.

2. On recherche, sur la base des données statistiques existantes, la répartition des variables choisies dans la population totale. Pour des variables telles que l'âge, le sexe, le lieu d'habitat, la CSP, les données sont souvent accessibles à l'INSEE ou dans des banques de données publiques. Pour d'autres variables (le revenu par ménage, la préférence politique, etc.) on est moins assuré de disposer de données fiables et récentes.
3. On construit l'échantillon en multipliant le taux de sondage par le pourcentage d'occurrence de chacune des variables de contrôle dans la population totale (cf. exemple ci-dessous).

4. On réalise l'enquête en respectant les pourcentages calculés ci-dessus.

1.2.2.3 Exemple

Cet exemple est fictif. Ne pas s'attacher à la vraisemblance des valeurs numériques, mais aux principes de la méthode.

On souhaite effectuer une enquête au 1/100° sur l'utilisation des transports en commun pour les déplacements domicile-travail, auprès de la population active de plus de 15 ans d'une grande agglomération. On suppose que le lieu d'habitat et le lieu de travail sont des variables de contrôle ; on a un doute sur le caractère explicatif de l'âge et du sexe.

On découpe le périmètre urbain en n zones ; supposons qu'il y ait 3 zones : centre-ville, périphérie du centre, banlieue, que nous désignons dans ce qui suit par A, B et C.

La taille de l'agglomération est suffisamment importante pour que l'on puisse disposer des données suivantes (ou les estimer de manière fiable) :

Répartition des types de trajets effectués par les actifs de plus de 15 ans pour leurs déplacements domicile vers travail									
AA	AB	AC	BA	BB	BC	CA	CB	CC	Total
20%	10%	5%	10%	5%	5%	30%	10%	5%	100%

Les données de l'INSEE, et autres banques de données fournissent les données suivantes :

Nombre d'actifs de plus de 15 ans : 350 000 actifs, dont 60% d'hommes et 40% de femmes.

Répartition par âge

- De 15 à 24 ans : 15%.
- De 25 à 34 ans : 40%.
- de 35 à 59 ans : 40%.
- 60 ans et plus : 5%.

Le sondage étant effectué au 1/100°, on veut disposer d'un échantillon de 3500 personnes.

Compte tenu des données qui précèdent on demandera à l'organisme chargé de réaliser les enquêtes de constituer l'échantillon au plus proche de ce qui suit :

- Cible : actifs habitant et travaillant dans l'agglomération et âgés de plus de 15 ans.
- Sexe : 2100 hommes ; 1400 femmes.
- Répartition par âge :
  - De 15 à 24 ans : 525.
  - De 25 à 34 ans : 1400.
  - de 35 à 59 ans : 1400.
  - 60 ans et plus : 175.
- Répartition des trajets domicile-travail :

AA	AB	AC	BA	BB	BC	CA	CB	CC	Total
700	350	175	350	175	175	1050	350	175	3500

On conçoit aisément que la constitution d'un échantillon répondant à ces critères soit affaire de spécialistes.

1.3 La taille de l'échantillon

Cette question est détaillée dans les chapitres qui suivent. Elle peut être résumée par la question suivante :

« Quelle taille doit-on donner à l'échantillon pour disposer d'une estimation satisfaisante des caractéristiques de la population étudiée avec une précision acceptable? ».

2 Enjeux et difficultés du dimensionnement des échantillons

2.1 Enjeux

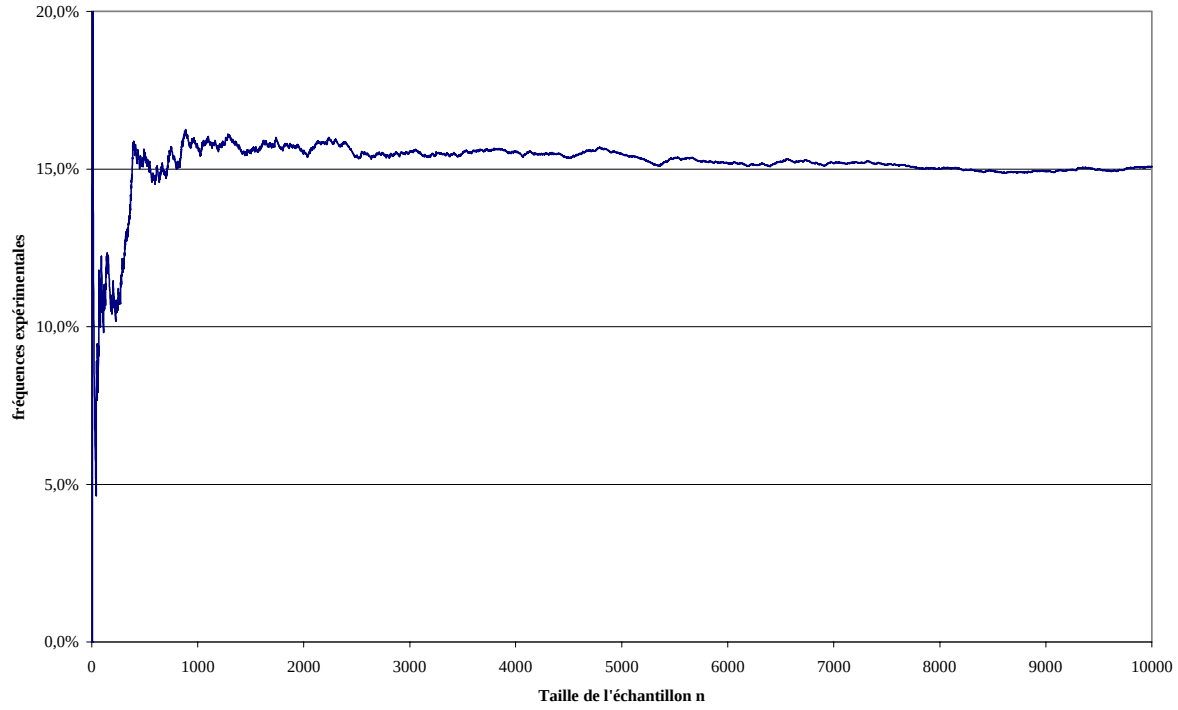
Pour illustrer l'importance de la taille de l'échantillon, nous avons simulé une population dont les individus représentent la réponse d'un capteur au passage d'un véhicule. La variable attachée à cet événement est une variable discrète pouvant prendre l'une ou l'autre des valeurs 0 (le véhicule n'est pas détecté) ou 1 (le véhicule est détecté).

La valeur vraie du taux de non détection dans cette population est de 15,1 %. La répartition des non détections est aléatoire. L'exemple a pour but de montrer comment varie l'estimation du taux de détection quand on fait croître la taille de l'échantillon.

Supposons que l'on réalise une expérience destinée à mesurer le taux de détection dans cette population en "tirant" un échantillon de taille n.

La figure qui suit montre quel serait le résultat obtenu expérimentalement en fonction de la taille de n :

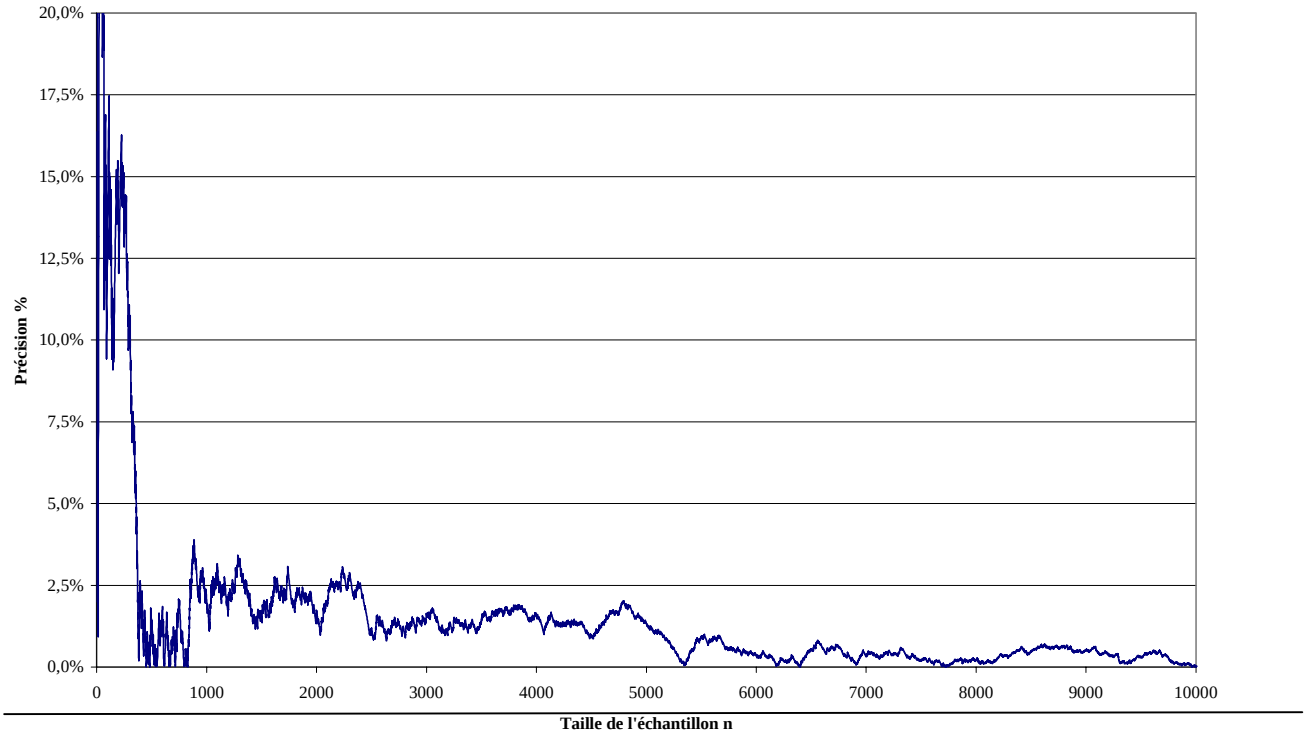
Figure 1 : variation de l'estimation du taux de non détection en fonction de la taille de l'échantillon (exemple)



Si on appelle « précision de la mesure[9] » la valeur absolue du demi-écart relatif entre la fréquence expérimentale et la valeur vraie, on a l'évolution suivante, en fonction de n :

9 Ou "justesse de la mesure".

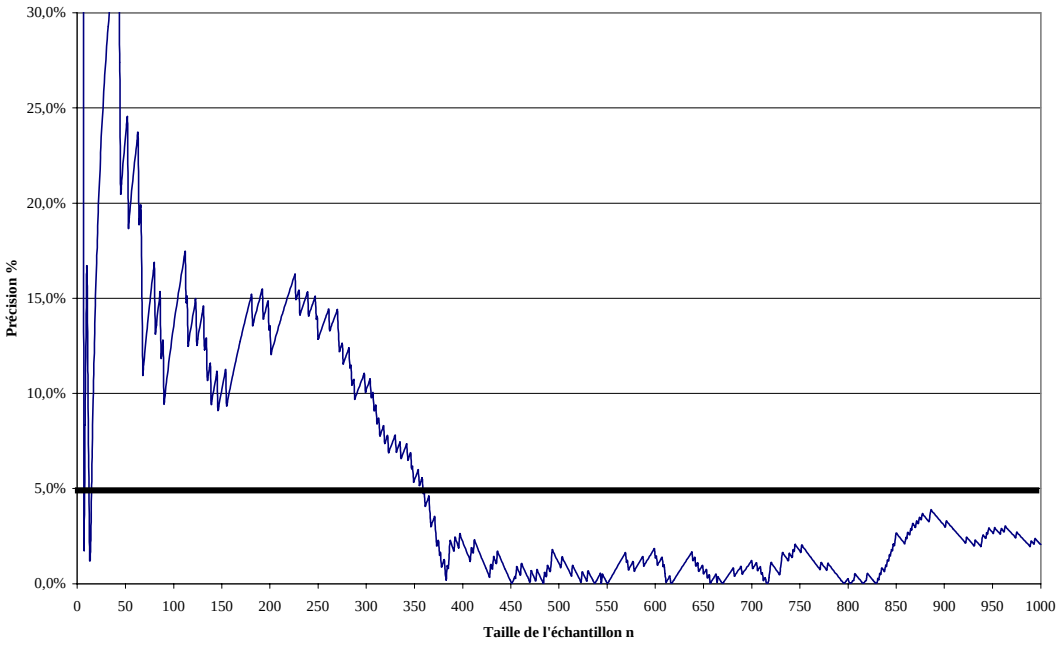
Figure 2 : variation de la précision de la mesure en fonction de la taille de l'échantillon (exemple)



Supposons que l'on accepte de se limiter à une précision de 5%. La taille minimale de l'échantillon correspond à l'effectif au delà duquel la précision (au sens où nous l'avons définie plus haut) est stabilisée au dessous de 5%.

Pour préciser cette valeur, nous nous intéressons ci-dessous à l'intervalle [0, 1000].

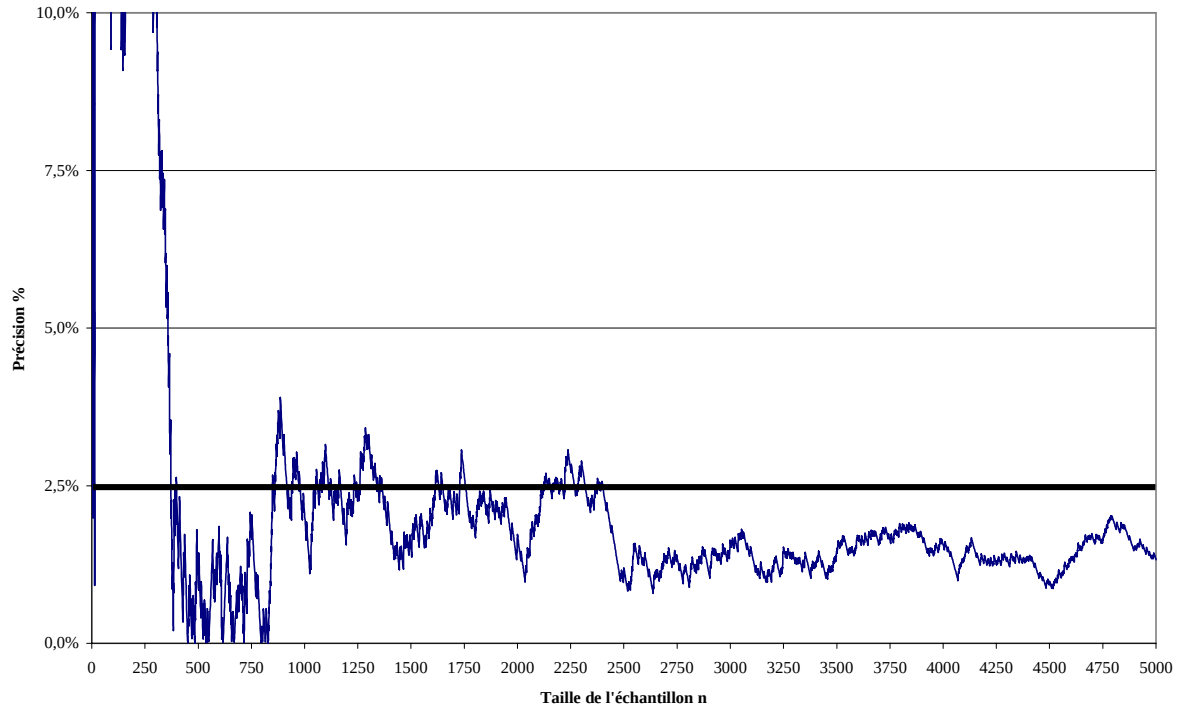
Figure 3 : taille d'échantillon nécessaire pour une précision de 5% (exemple)



On voit qu'il est nécessire que l'échantillon contienne 375 individus pour atteindre la précision souhaitée.

Si par contre on voulait atteindre une précision de 2,5%, il faudrait un échantillon d'environ 2400 individus, comme le montre le graphe ci-dessous :

Figure 4: taille d'échantillon nécessaire pour une précision de 2,5 % (exemple)



Les valeurs numériques présentées plus haut ne sont pas extrapolables à d'autres populations. Par contre les tendances sont généralisables (elles correspondent d'ailleurs au sentiment intuitif de tout un chacun) :

- Lorsque la taille de l'échantillon croît, la fréquence expérimentale se rapproche de la fréquence réelle et la précision de la mesure s'améliore.
- Une taille d'échantillon trop faible ne permet pas de conclure avec une précision raisonnable.

2.2 Difficultés

Les développements qui précèdent ont été possibles car, s'agissant d'une population parfaitement décrite, on connaissait a priori le résultat, c'est à dire la fréquence réelle.

Bien entendu, ce n'est jamais le cas puisque le but de l'expérience est, précisément, de déterminer cette fréquence.

Dans la pratique l'expérimentateur peut fixer ses propres contraintes en matière de précision attendue mais ne peut pas choisir « au hasard » la taille de l'échantillon qui lui permettra de satisfaire ces contraintes. **Il souhaite donc pouvoir prédéterminer la taille de l'échantillon**.

Précisons d'emblée que ce problème est, en toute rigueur, impossible à résoudre. En effet sa résolution suppose connues des valeurs qui sont l'enjeu de l'expérience. Dans tous les cas, on est amené à faire des hypothèses sur l'ordre de grandeur des résultats que l'on va obtenir.

En d'autres termes, on n'est jamais assuré a priori d'avoir correctement dimensionné un échantillon. Ce n'est qu'a posteriori que l'on pourra vérifier l'adéquation des hypothèses des hypothèses sur les ordres de grandeur.

Mais cette évidence ne doit pas être décourageante : d'une part car il est souvent possible de faire des hypothèses crédibles ; d'autre part car un mauvais dimensionnement de l'échantillon ne rend pas forcément caduque l'expérience : il modifie, dans un sens ou dans l'autre, la qualité de l'estimation. Si la précision est meilleure que celle espérée, l'échantillon aura été dimensionné trop largement et le seul regret que pourra avoir l'expérimentateur est d'avoir été trop « luxueux ». Dans le cas contraire, la taille de l'échantillon aura été sous-estimée et il est de la responsabilité de l'expérimentateur de décider si la qualité de l'estimation reste acceptable.

2.3 Champ de l'étude

L'étude présentée ici est limitée à deux situations simples dont nous estimons qu'elles correspondent à la majorité des problèmes rencontrés par le praticien de l'exploitation de la route :

- La détermination de la taille d'un échantillon pour la mesure de la moyenne d'une variable continue.
- La détermination de la taille d'un échantillon pour la mesure de la fréquence d'une variable discrète pouvant prendre deux états.

La première situation sera illustrée par l'exemple d'une expérience destinée à mesurer des temps de parcours de véhicules.

La seconde par l'exemple d'une expérience destinée à mesurer le taux de détection d'un capteur.

2.4 Quelques définitions préalables

2.4.1 Variable discrète, variable continue

- Une variable discrète est une variable dont le domaine de définition comprend un nombre fini de valeurs, ou un nombre infini de valeurs dénombrables. Exemples : la variable caractérisant l'occurrence ou la non occurrence de la détection par un capteur est une variable discrète pouvant prendre deux valeurs ; la variable caractérisant le nombre d'incidents réels détectés par un système de DAI entre l'occurrence de deux fausses détections successives est une variable discrète pouvant prendre une infinité de valeurs dénombrables (1, 2, …n) ; etc.

- Une variable continue est une variable dont le domaine de définition est un intervalle continu. Exemple : la vitesse des véhicules, les temps de parcours, etc.

Dans la pratique, la frontière entre variable discrète et variable continue est perméable. Par exemple : le temps de parcours est une variable continue. Mais si on mesure ces temps avec une précision de la seconde, on peut aussi considérer que c'est une variable discrète dont l'intervalle de définition est infini et dénombrable (le nombre de secondes).

2.4.2 Echantillonnage indépendant, échantillonnage exhaustif

Un échantillonnage est dit « indépendant » (ou « avec remise »), si le fait de tirer un individu dans la population totale ne modifie pas la probabilité qu'ont les autres individus d'être tirés.

Dans le cas contraire, il est dit « exhaustif » (ou « sans remise »).

Un exemple simple permettra de comprendre ces notions :

Supposons qu'une urne contienne 100 boules, dont 10 sont noires, les autres blanches. Au départ, la probabilité de tirer une boule noire est égale à 1/10. Supposons que le premier tirage soit une boule blanche et que l'on ne remettre pas cette boule dans l'urne. Au deuxième tirage, la probabilité de tirer une boule noire aura augmenté : elle deviendra égale à 10/99. Si au contraire le premier tirage est une boule noire et que l'on n'effectue pas de remise, la probabilité de tirer une boule noire au 2° tirage a diminué et devient égale à 9/99. Si par contre on remet systématiquement la boule tirée dans l'urne, la probabilité de tirer une boule noire reste constante et égale à 1/10.

Dans tout ce qui suit, nous n'envisageons que des tirages indépendants. On admettra en effet, pour reprendre les deux exemples cités plus haut :

1. Que le fait pour un véhicule d'être ou de n'être pas détecté ne modifie pas la probabilité qu'ont les autres véhicules d'être ou de n'être pas détectés.

2. Que le fait pour un véhicule d'avoir mis un temps t_i pour aller de A à B, ne modifie pas la loi de distribution des probabilités de t pour les autres véhicules.

2.4.3 Niveau de confiance, niveau de risque

Nous appelons ***niveau de confiance*** (ou seuil de confiance) d'un événement quelconque, la probabilité attachée à l'estimation d'un paramètre de la population. Nous notons **$(1 - \alpha)$** ce seuil de confiance, avec **$0 \leq \alpha \leq 1$** et α la probabilité de l'événement contraire (probabilité de conclure à tort), appelé **niveau de risque.**

3 Taille de l'échantillon dans le cas de la mesure de la moyenne d'une variable continue

3.1 Base théorique : loi de la moyenne d'un gros échantillon

La loi de probabilité de la moyenne arithmétique d'un échantillon tiré avec remise dans une population de moyenne m et d'écart-type σ, peut être assimilée à une loi normale de moyenne m et d'écart-type σ/√n, quelle que soit la distribution de l'échantillon, si la taille de l'échantillon est suffisante[10].

Supposons l'expérience effectuée : on a tiré un échantillon de taille n dont la moyenne arithmétique est $\bar{x}$ et l'écart-type **s'**.

La théorie montre que la moyenne arithmétique μ est toujours un estimateur sans biais de la moyenne réelle m[11] et que, dans le cas de tirages indépendants, un estimateur sans biais de l'écart-type de la population est :

$$s = s'\sqrt{\frac{n}{n-1}}$$

L'intervalle de confiance bilatéral symétrique de la moyenne, pour un niveau de risque α est alors donné par :

$$x \pm u_{1-\alpha/2} s \sqrt{n}$$

Dans cette expression $u_{(1-\alpha/2)}$ est la valeur de la variable centrée réduite correspondant au seuil de probabilité $(1-\alpha/2)$. Cette valeur est disponible dans les tables et tableurs usuels.

On trouvera en annexe 1 une table donnant $u_{(1-\alpha/2)}$ en fonction de α, pour α variant entre 0,01 et 0,1.

[10] Dans la pratique on admet que la taille de l'échantillon doit être supérieure à 30. C'est l'hypothèse que nous faisons dans tout ce chapitre.

[11] Un estimateur d'une caractéristique quelconque d'une population est dit « sans biais » s'il est toujours centré sur la valeur réelle de cette caractéristique dans la population. C'est le cas pour la moyenne arithmétique de l'échantillon. Ce n'est pas le cas pour l'écart-type : l'écart-type d'un échantillon est un estimateur biaisé de l'écart-type de la population réelle.

3.2 Méthode

3.2.1 Choix d'un niveau de risque accepté

L'expérimentateur doit choisir préalablement un niveau de risque accepté α.

α représente la probabilité de conclure à tort que la moyenne réelle de la population est comprise dans l'intervalle de confiance calculé à partir de l'échantillon (dans la pratique on choisit souvent $\alpha = 5\%$, ce qui signifie qu'en moyenne on se trompe une fois sur 20).

On admettra dans tout ce qui suit que le risque est partagé, c'est-à-dire que la probabilité d'être inférieur à la borne inférieure de l'intervalle de confiance est égal au risque d'être supérieur à la borne supérieure de cet intervalle, soit : $\alpha/2$.

3.2.2 Choix d'une précision relative

Nous appelons dans ce qui suit ***précision relative,*** la valeur :

$$I=\frac{u_{1-\alpha/2}s\sqrt{n}}{\bar{x}}$$

I, exprimée en pourcentage, a une interprétation immédiate : c'est la demi-étendue de l'intervalle de confiance rapportée à la moyenne[12]. Par exemple : le fait de choisir I = 10 % signifie que l'on souhaite un intervalle de confiance dont l'amplitude soit :

$$\bar{x} \pm 10\% \ \bar{x}$$

L'expérimentateur doit choisir la valeur I de la précision relative souhaitée.

Posons **c = s / $\bar{x}$** = coefficient de variation ; il vient :

$$n=\left(\frac{cu_{(1-\alpha/2)}}{I}\right)^2$$

n représente la taille minimale de l'échantillon permettant de disposer d'une précision **I**, avec un niveau de risque α.

Si on examine cette expression, on voit qu'elle comporte un terme u connu si α est connu, un terme I choisi par l'expérimentateur **et un rapport c = s / $\bar{x}$ inconnu**.

On touche là à la difficulté **incontournable** déjà signalée qui réside dans le fait que le calcul nécessite la connaissance de données qui ne peuvent pas être rigoureusement connues

[12] Attention : ne pas confondre l'incertitude relative avec le risque α.

avant l'expérience. La prédétermination de la taille de l'échantillon nécessite impérativement une évaluation de l'ordre de grandeur du coefficient de variation **c**.

3.2.3 Détermination d'un ordre de grandeur du rapport c = s / $\bar{x}$ et calcul de n

Pour déterminer l'ordre de grandeur de c, les outils disponibles sont en nombre limité. Nous en évoquons 3 :

1. L'ordre de grandeur est connu par l'état de l'art ou par des expériences antérieures : on utilisera le rapport c déduit de ces données disponibles.
2. On effectue un test préalable de calibrage en relevant une trentaine de temps de parcours et on utilise comme valeur approchée de calcul la valeur c mesurée sur cet échantillon de calibrage.
3. La situation la plus favorable est celle où le recueil de données est automatique et ne nécessite pas la mise en œuvre de moyens coûteux. On peut alors se permettre de débuter l'expérience « au fil de l'eau » et d'effectuer un suivi de l'évolution du coefficient c au fur et à mesure de la croissance de l'échantillon. Lorsque c est à peu près stabilisé, on utilise cette valeur pour dimensionner l'échantillon, c'est-à-dire fixer le terme de l'expérience.

Lorsqu'on dispose d'un ordre de grandeur de c, par un moyen quelconque, on procède au calcul de n.

Nous avons vu plus haut que :

$$n = \left(\frac{c u_{(1-\alpha/2)}}{I} \right)^2$$

Le calcul de n ne pose donc aucune espèce de difficulté.

3.3 Exemple

On veut réaliser une expérience destinée à mesurer des temps de parcours de véhicules entre 16h et 19h. Pour déterminer l'ordre de grandeur de c, on effectue une mesure de calibrage en mesurant 30 temps de parcours sur le site, entre 17h et 18h. Ces 30 mesures fournissent les temps suivants (en secondes) :

1110	992	884	999	770	993
1109	952	869	1057	975	947
1036	960	1026	1063	1180	746
735	783	1033	963	988	771
722	791	1278	911	1025	971

La moyenne de cet échantillon est égale à : 955 secondes et son écart-type est égal à 136 secondes.

- Estimateur sans biais de la moyenne : 955s.
- Estimateur sans biais de l'écart-type : 136.√30/29 = 138

Une valeur approchée du coefficient c est donc : 138 / 955 = 0,145. On utilise cette valeur pour prédimensionner l'échantillon, avec les choix suivants : 1 - α = 0,99 et I = 5%.

Avec :

- C = 0,145
- I = 0,05
- $u_{(1-\alpha/2)} = u_{0,995} = 2,58$

soit : **n = 56** (cf. formule de calcul au § précédent).

Cette valeur étant assez faible, on peut se permettre d'augmenter la précision du résultat en augmentant la taille de l'échantillon. On a cherché quel était « le prix à payer » pour cela :

Précision relative	Taille
5%	56
4%	87
3%	156
2%	350

Une taille n = 100 semble un compromis raisonnable. L'expérience a été conduite avec cette taille d'échantillon et a conduit aux résultats suivants :

- Moyenne = 994 s.
- Ecart-type = 148 secondes. Estimateur sans biais = 149s.

On peut maintenant calculer la précision relative I :

$$\mathbf{I}=\frac{\mathbf{u_{1-\alpha/2}s\sqrt{n}}}{\bar{\mathbf{x}}}$$

avec :

$u_{(1-\alpha/2)} = u_{0,05} = 2,58$
$s = 149\ s$
$\bar{x} = 994\ s$
$n = 100$

soit : I% = 3,9 %

soit : **956 s < m < 1032 s.**

4 Taille de l'échantillon dans le cas de la mesure d'une fréquence

4.1 Base théorique

On s'intéresse ici au calcul de la taille de l'échantillon lorsque la variable est une variable discrète binaire.

L'exemple qui nous servira de support est la détermination expérimentale du taux de détection d'un capteur[13].

4.1.1 Notations et définitions

L'événement élémentaire est constitué par le passage d'un véhicule sur le capteur.

La variable est l'état du capteur :

> *X=1 : le véhicule est détecté.*
> *X=0 : le véhicule n'est pas détecté.*

On appellera parfois « tirage »l'occurrence d'un événement.

On désignera par p la probabilité (inconnue) de non-détection[14] dans la population totale. La probabilité de détection est donc 1-p.

Dans l'échantillon de taille n nous désignerons par f la fréquence observée expérimentalement, et par k le nombre d'événements lui correspondant.

On a donc **un estimateur de p donné par f = k/n**.

4.1.2 Principes de la méthode

Dans son principe général, la méthode est similaire à celle utilisée pour prédéterminer l'échantillon destiné à estimer la moyenne d'une variable continue. Elle en diffère par les moyens de calcul à utiliser.

1. L'expérimentateur doit choisir préalablement un niveau de risque accepté α. On admettra ici aussi que le risque est partagé ($\alpha/2$).

2. Soient p_1 et p_2 les bornes de l'intervalle de confiance de la proportion inconnue p, dont un estimateur est f. Nous appelons dans ce qui suit ***précision relative de la mesure*** la valeur : $\mathbf{I\%=\frac{(p_2-p_1)/2}{f}}$, ou f est la fréquence mesurée. Il s'agit donc de la demi-amplitude

---

[13] Plus exactement : du taux de non-détection, complément à 1 du taux de détection. Le principe de l'expérience est, par exemple, le suivant : un observateur relève et date tous les passages de véhicules sur le capteur ; on compare ces relevés avec les données fournies par le capteur. Le taux de non-détection est le pourcentage de véhicules ayant effectivement franchi le capteur mais qui n'ont pas été détectés. On ne s'intéresse pas ici aux fausses détections qui constituent un problème différent.

[14] Le lecteur pourra transposer sans peine l'exemple à tous types de phénomènes binaires. Il réservera la notation p à la probabilité de l'état dont la probabilité d'occurrence est la plus faible.

de l'intervalle de confiance, rapportée à la fréquence mesurée[15]. **L'expérimentateur doit choisir la valeur I de la précision relative souhaitée**.

3. L'expérimentateur doit enfin faire une hypothèse sur l'ordre de grandeur de la valeur de f qu'il mesurera expérimentalement[16]. On ne peut, là encore, qu'approcher un ordre de grandeur :

- Soit par l'état de l'art ou les informations fournies par l'industriel.
- Soit par un test de calibrage préalable.
- Soit « au fil de l'eau », comme indiqué au §3.2.3

Dans ce qui suit, nous appellerons parfois la fréquence f, « **fréquence-cible** ».

4.2 Cas général (traitement du problème par la loi binomiale)

4.2.1 Position du problème

On assimile l'expérience à une série de tirages indépendants (cf. § 2.4.2). Comme nous l'avons déjà indiqué, cela signifie que le fait qu'un véhicule soit, ou ne soit pas, détecté ne modifie pas la probabilité qu'ont les autres véhicules d'être, ou de ne pas être, détectés. Dans ces conditions la loi de densité de probabilité de la fréquence est **une loi binomiale** dont les caractéristiques sont :

$$\mathbf{Moyenne = p}$$

$$\mathbf{Ecart - type = \sqrt{\frac{p(1-p)}{n}}}$$

On trouve, dans la littérature, des tables et abaques souvent limitées à des tailles d'échantillon inférieures à 100 (par exemple dans [1]), plus rarement utilisables pour des échantillons de taille supérieure (c'est le cas dans [4]).

Exemple :

- Soit f la fréquence cible, f = 0,4.
- Soit I la précision relative, I = 25%.²
- On choisit 1-α =0,95, avec un risque partagé (intervalle de confiance bilatéral).

On cherche donc la taille minimale n de l'échantillon permettant de satisfaire à l'inéquation :

$$0,3 < p < 0,5$$

On utilise l'abaque Ib (page 52) de [1] ; on trouve n=100 environ[17]

---

[15] On notera que le caractère discret de la loi binomiale implique que la valeur φ ne correspond généralement pas exactement à la valeur centrale de l'intervalle de confiance.

[16] Cette hypothèse est le pendant de l'hypothèse faite sur c dans le cas d'une variable continue.

4.2.2 Abaques ZELT

La ZELT a établi ses propres abaques de calcul pour traiter le problème dans le cas où la loi binomiale peut être utilisée[18], c'est-à-dire :

> *Lorsque elle s'applique à un tirage avec remise.*
>
> *Lorsqu'elle s'applique à un tirage sans remise à condition que la fraction prélevée ne représente pas plus de 10% de la population totale.*

4.2.2.1 Justification théorique

Sur un échantillon de taille n, la probabilité que la fréquence expérimentale soit égale à une valeur p, c'est-à-dire que le nombre de véhicules non-détectés soit k = np, est égale à :

$$\mathbf{Pr\,(k)=C_n^k p^k (1-p)^{n-k}}$$

Les limites inférieures et supérieures de l'intervalle de confiance (respectivement p_1 et p_2), pour un seuil de confiance 1-α, sont les solutions des équations suivantes[19] :

$$\sum_{j=k}^{n} C_n^j p_1^j (1-p_1)^{n-j} = \alpha/2$$

$$\sum_{j=0}^{k} C_n^j p_2^j (1-p_2)^{n-j} = \alpha/2$$

Le principe du calcul réalisé par la ZELT pour l'établissement des abaques consiste, pour une fréquence cible donnée k/n, et pour 1-α =0,95, à établir les courbes n=f(I%). Ce calcul a été rendu possible par le développement d'un programme spécifique sous environnement DELPHI 4.[20]

4.2.2.2 Abaques

On trouvera ci-après un jeu d'abaques établis pour des valeurs de la fréquence-cible variant de 5% à 50% (inclus) par pas de 5% (soit 10 abaques numérotés de 1 à 10).

Un abaque supplémentaire (numéroté 11) précise le domaine des échantillons de taille moyenne (taille inférieure à 500).

On rappelle que ces abaques sont établis pour 1-α =0,95.

---

[17] Dans la limite de la précision de lecture sur l'abaque. Lecture assez difficile, car cet abaque n'a pas été construit pour fournir n mais pour fournir un intervalle de confiance.

[18] Elle peut l'être dans les problémes dont nous traitons ici (exemple : taux de détection d'un capteur) car la population dont est extrait l'échantillon est quasiment infinie.

[19] Voir justification de ces relations en [1], p. 42 ou en [4], p. 260 et suivantes.

[20] Ce programme permet, si nécessaire, de dresser les abaques qui correspondent à d'autres valeurs de 1- α.

Utilisation des abaques n°1 à 10 :

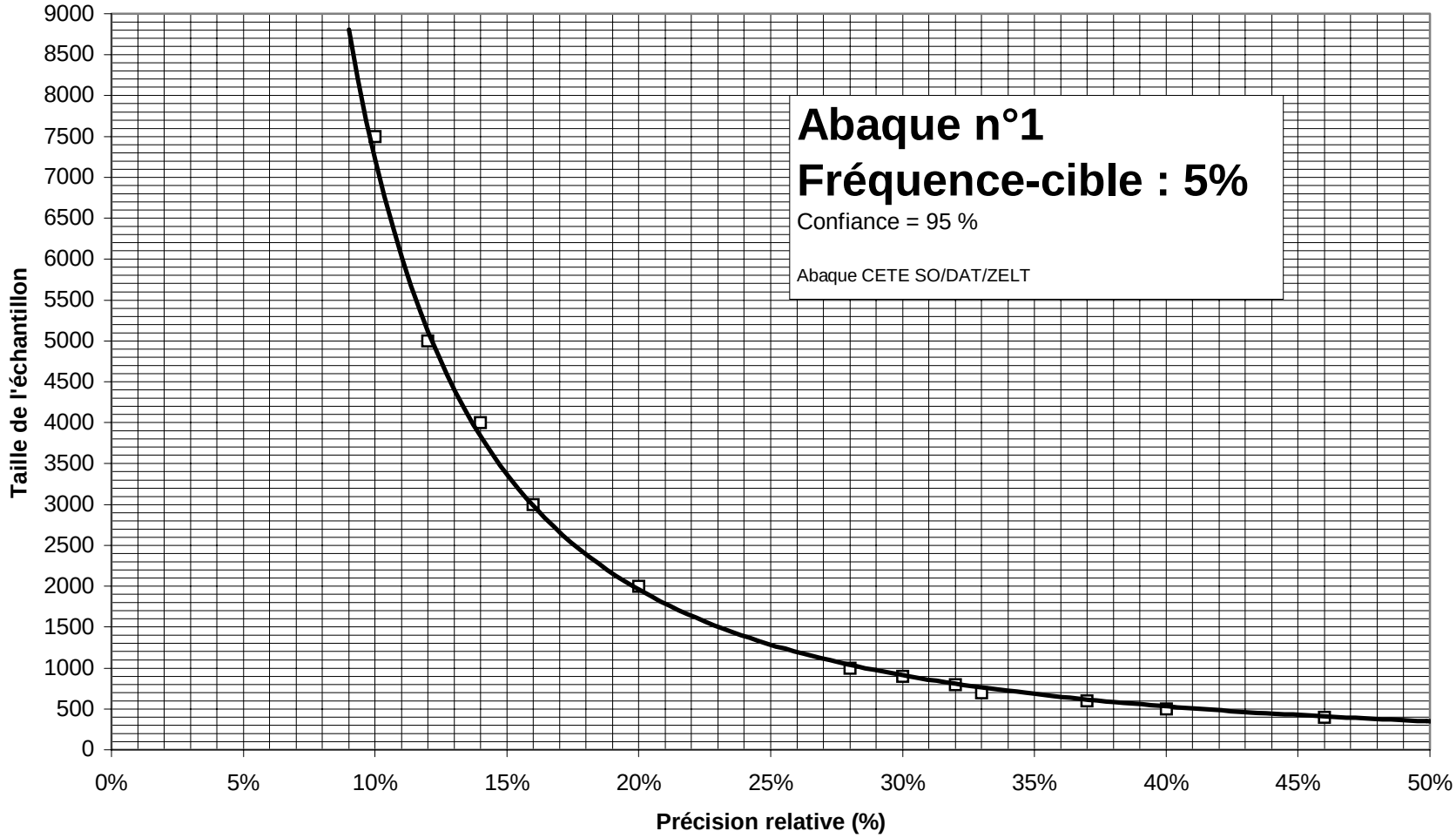
L'abaque à utiliser est déterminé par la fréquence-cible. Nous rappelons que la fréquence-cible représente l'ordre de grandeur prévisible de la fréquence dans l'échantillon.

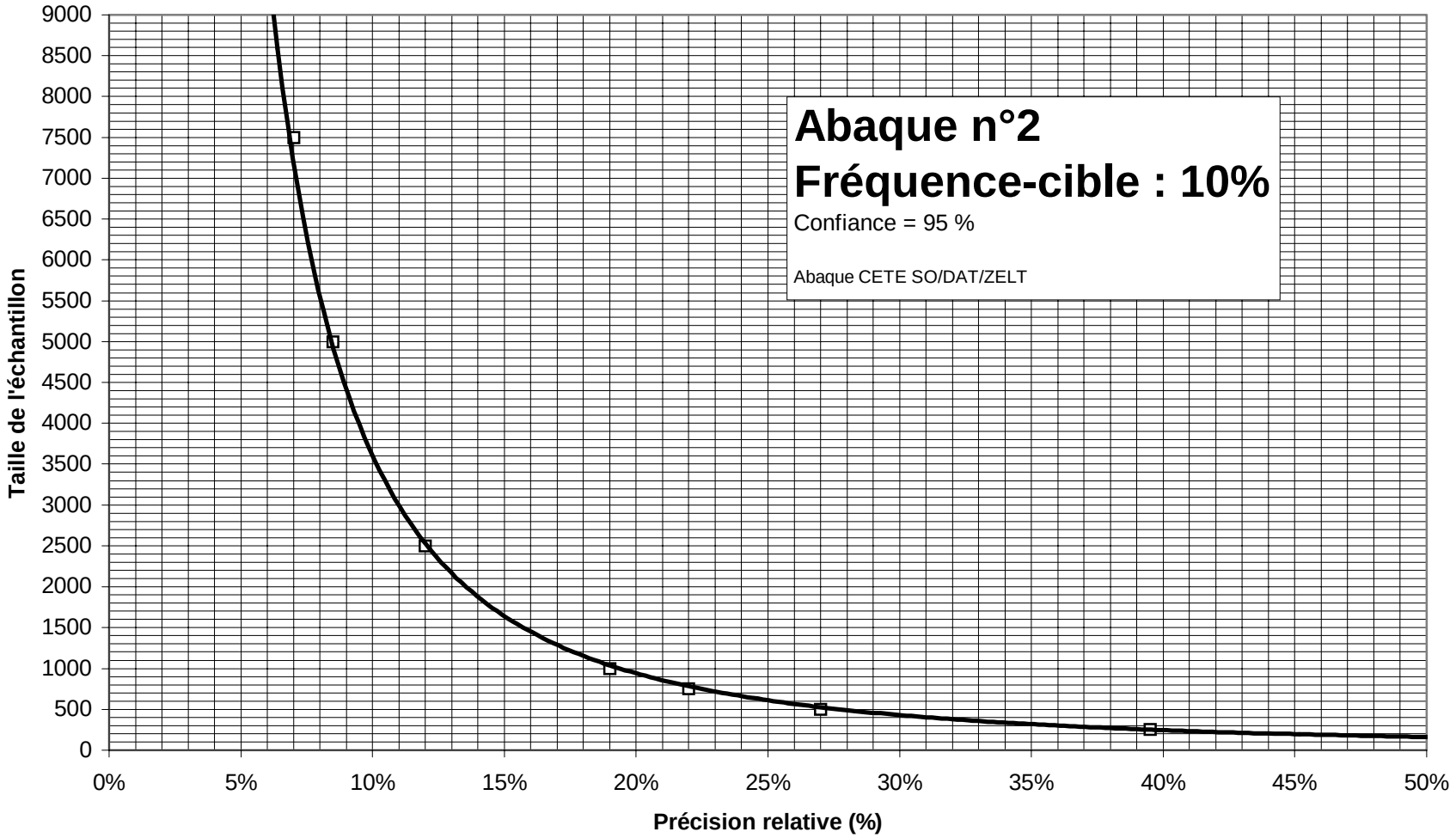
L'entrée dans l'abaque est constituée par la précision relative I %, en abscisse. Nous rappelons que nous avons appelé "précision relative" le rapport entre la demi-amplitude de l'intervalle de confiance et la moyenne mesurée sur l'échantillon.

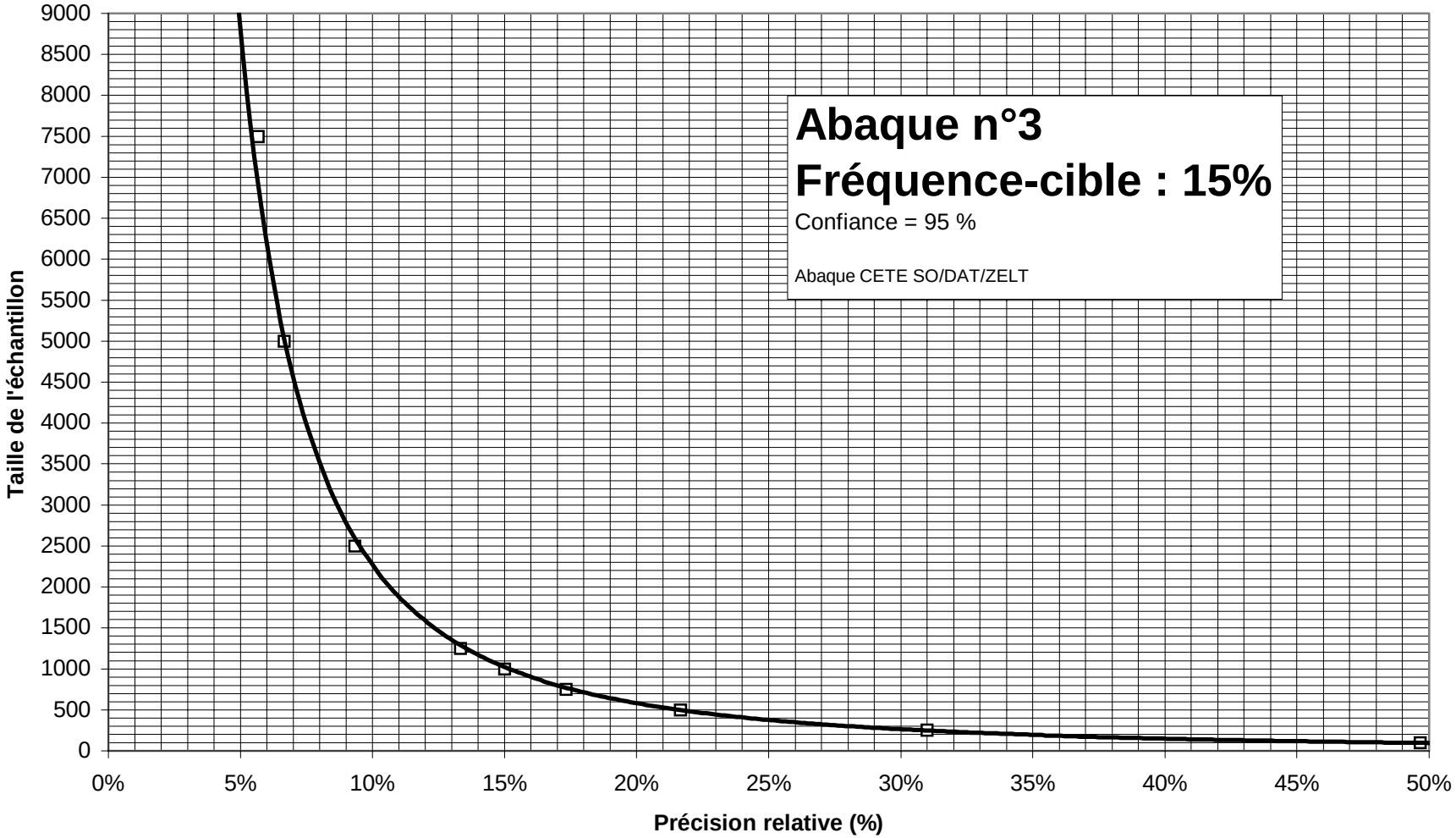
La taille de l'échantillon permettant de satisfaire I%, pour la fréquence-cible considérée et pour $1-\alpha = 0,95$ est lue en ordonnée.

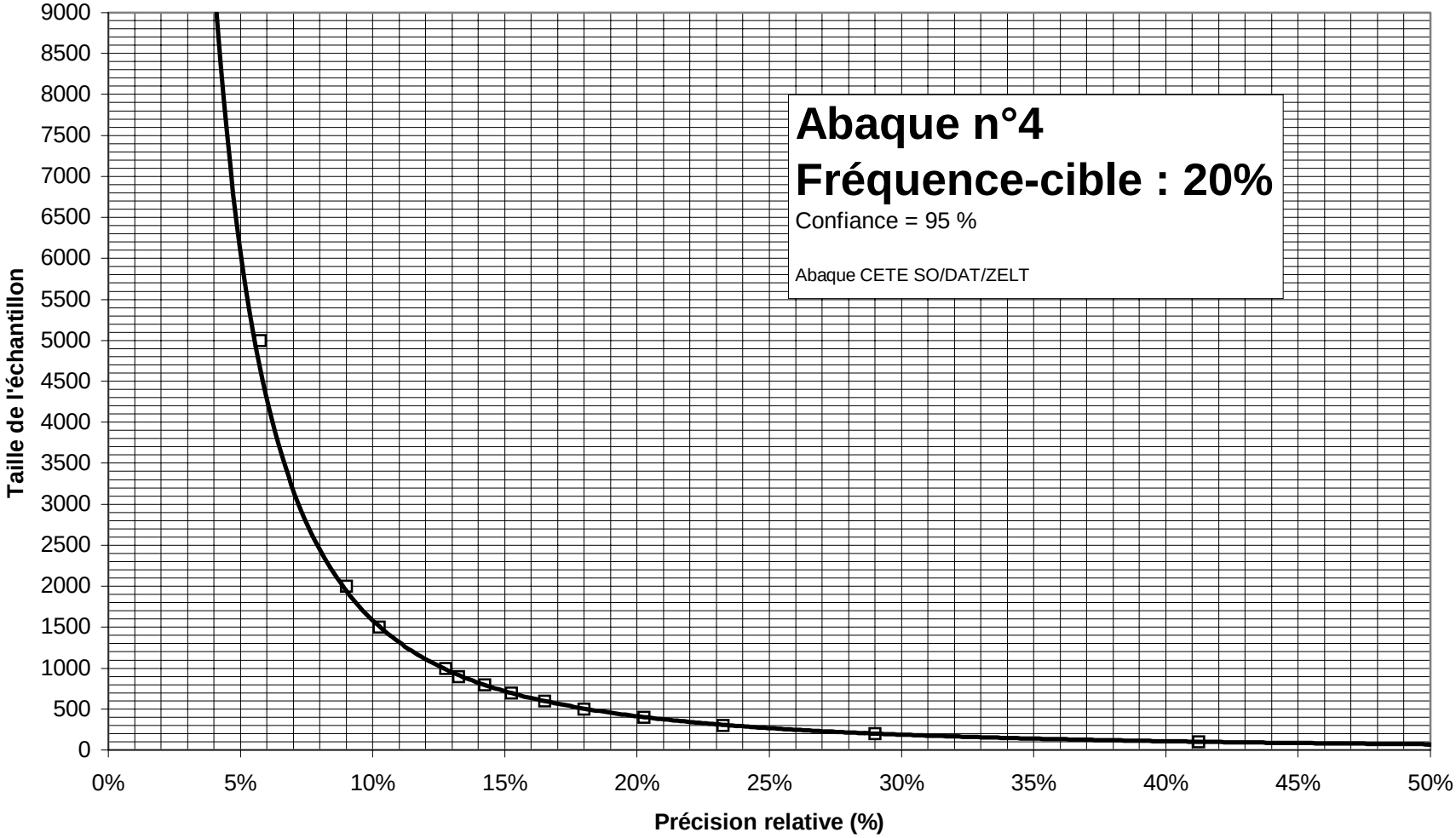
Utilisation de l'abaque N°11 :

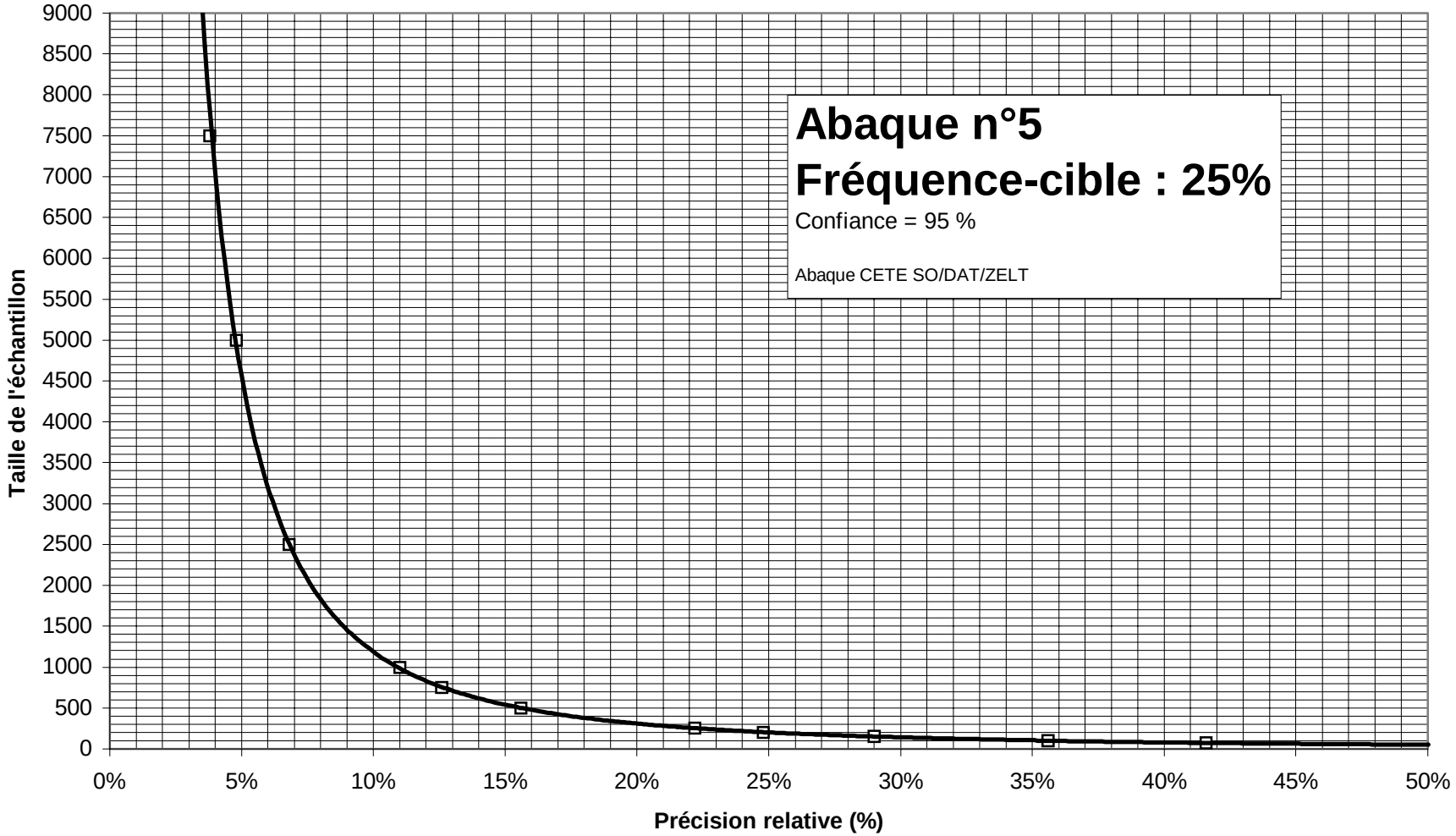
L'abaque n°11 est un récapitulatif des abaques précédents pour la zone n <500. Elle n'apporte pas d'information supplémentaire par rapport aux abaques 1 à 10, mais procure une meilleure lisibilité.

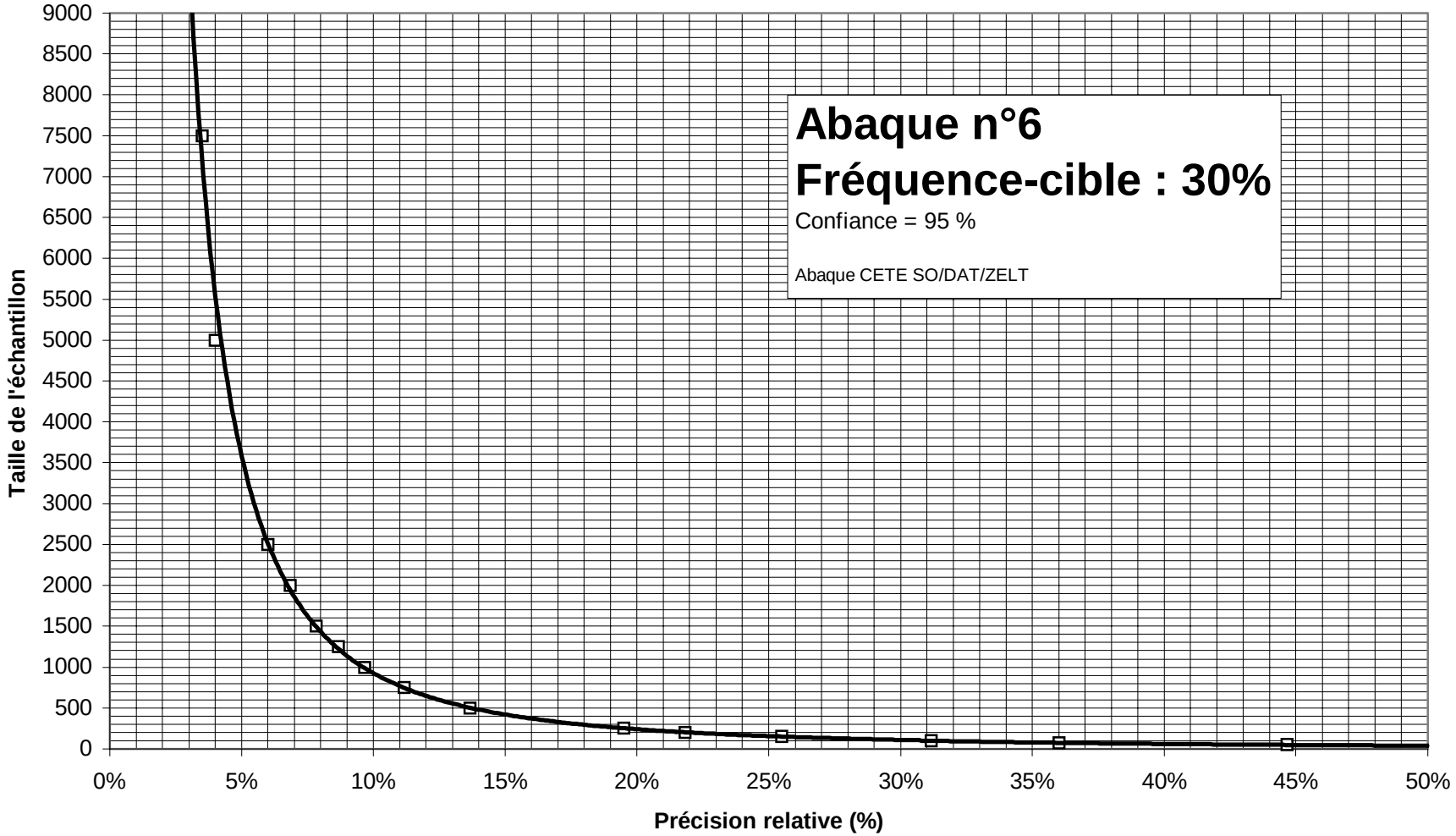

Abaque n°1
Fréquence-cible : 5%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)

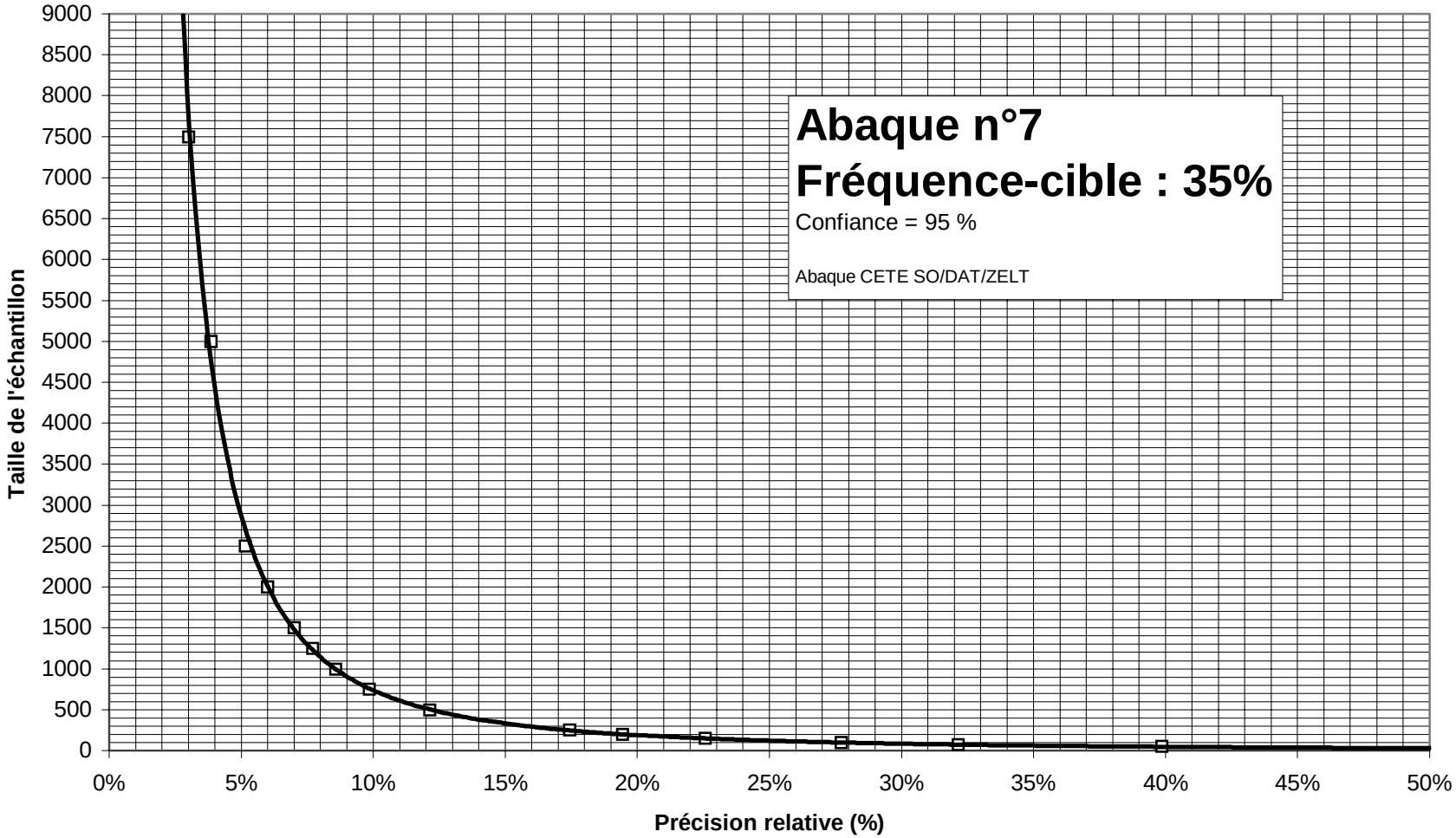

Abaque n°2
Fréquence-cible : 10%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)

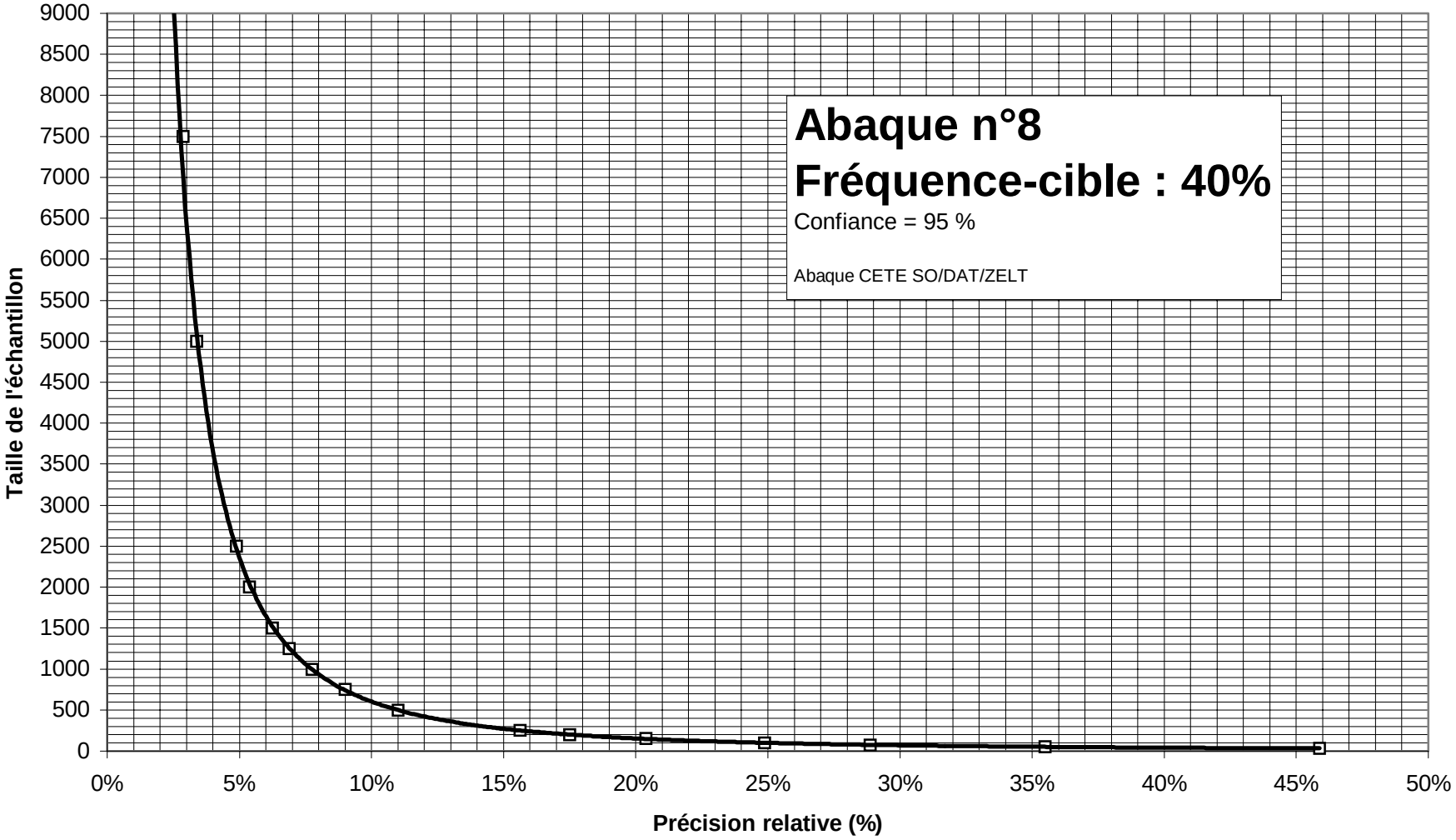

Abaque n°3
Fréquence-cible : 15%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)

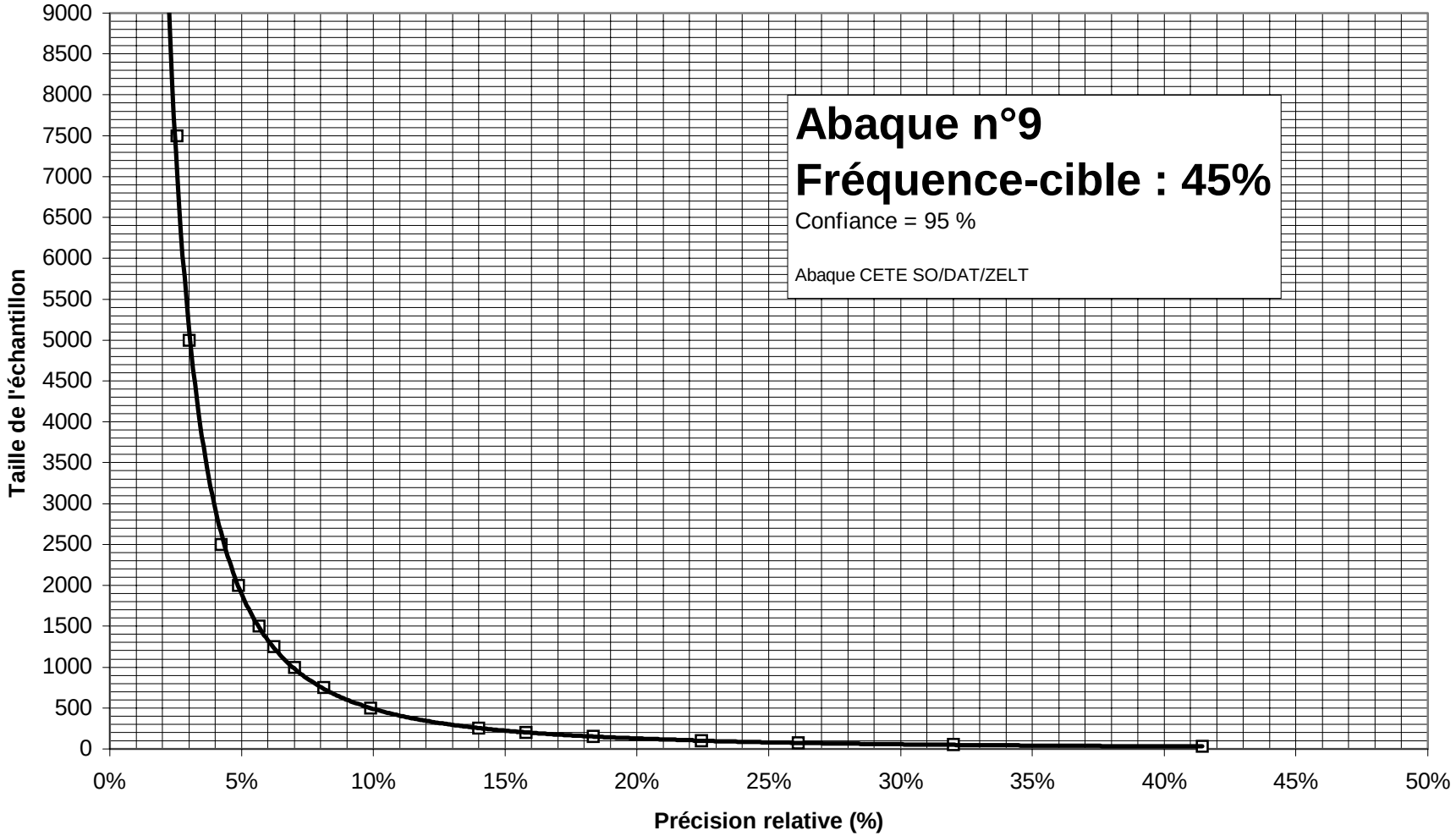

Abaque n°4
Fréquence-cible : 20%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)

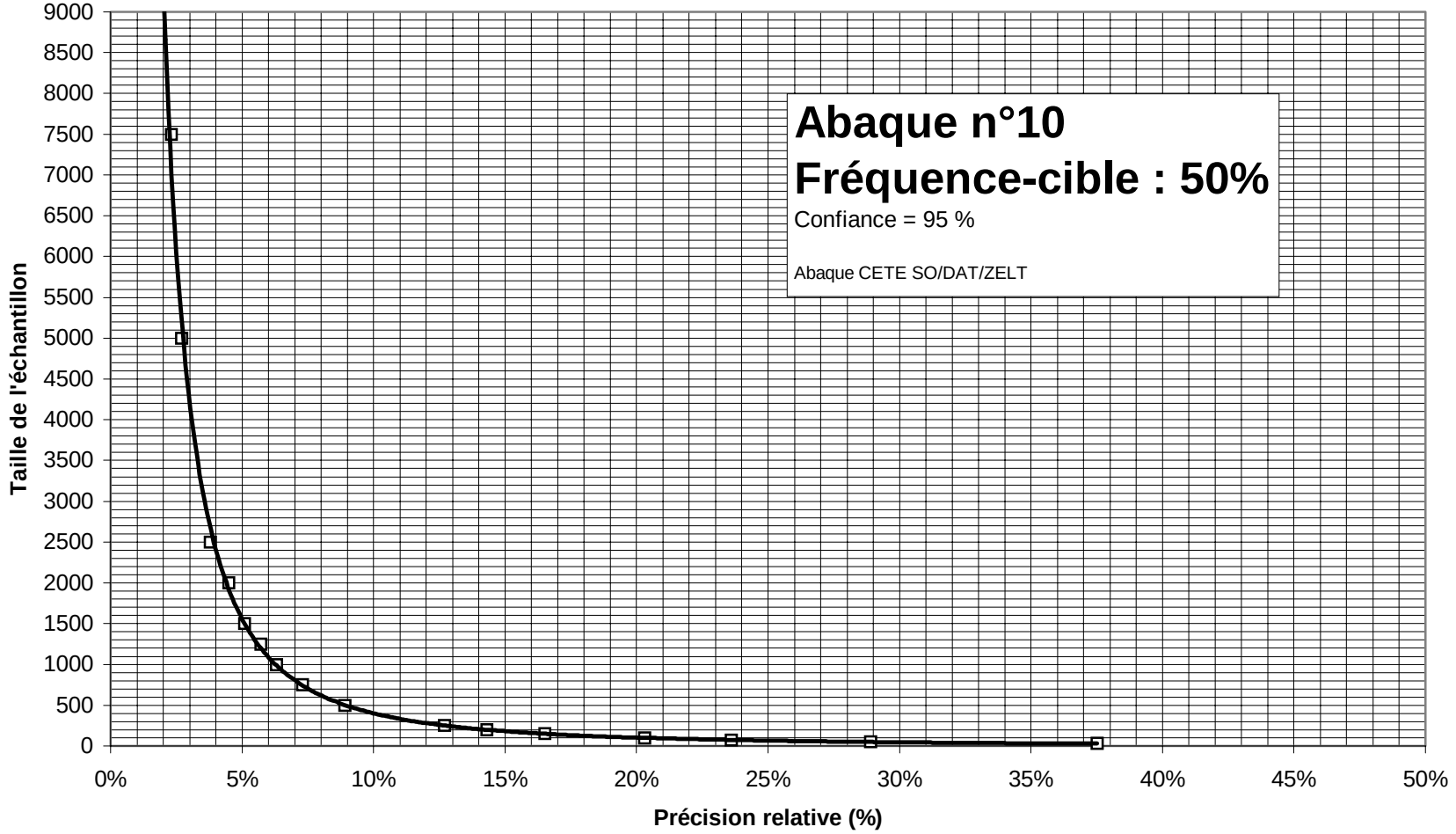

Abaque n°5
Fréquence-cible : 25%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)


Abaque n°6
Fréquence-cible : 30%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)

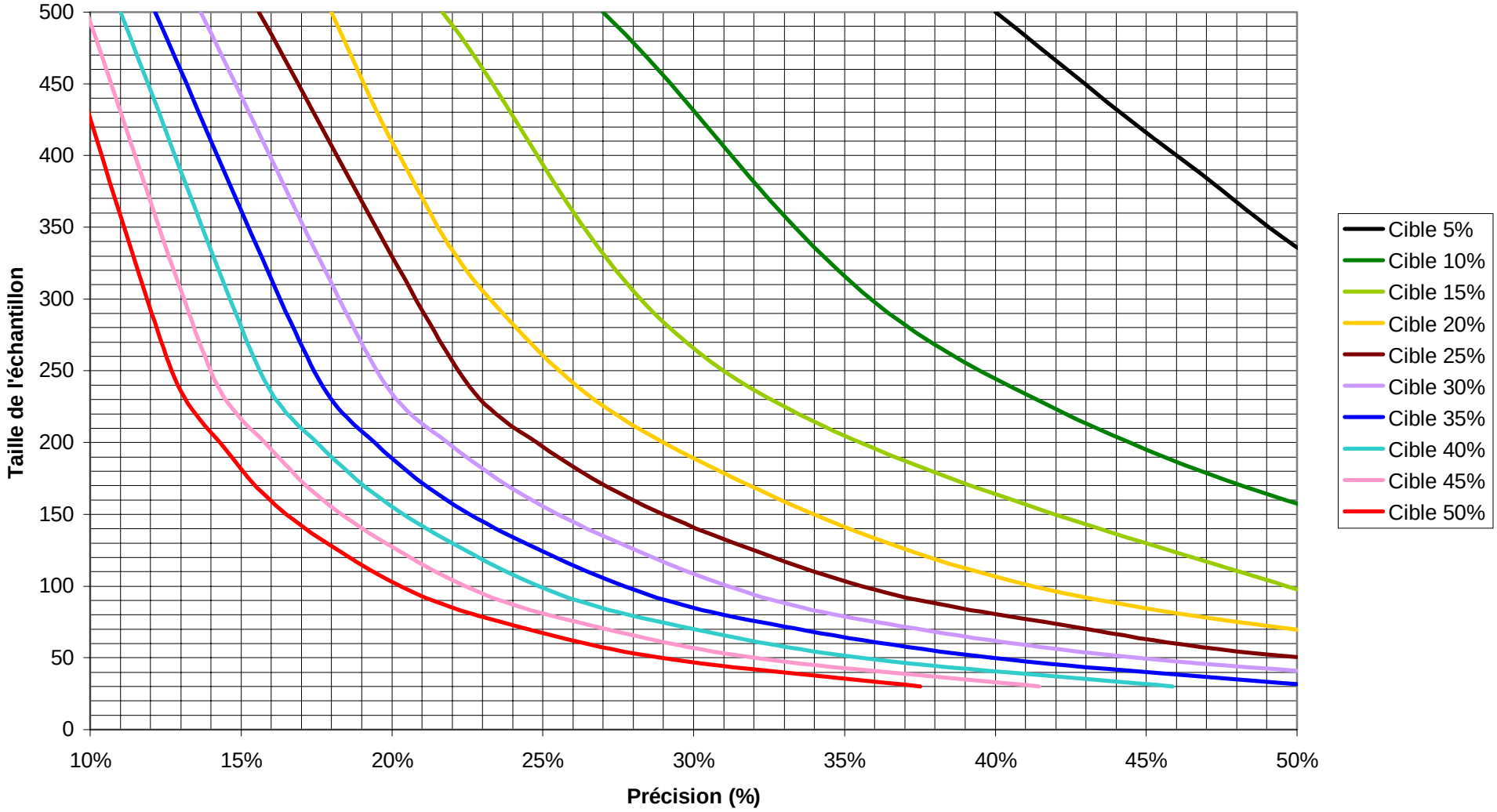

Abaque n°7
Fréquence-cible : 35%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)


Abaque n°8
Fréquence-cible : 40%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)


Abaque n°9
Fréquence-cible : 45%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)


Abaque n°10
Fréquence-cible : 50%
Confiance = 95 %
Abaque CETE SO/DAT/ZELT
Taille de l'échantillon
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
Précision relative (%)

Abaque n°11 : domaine des échantillons de taille moyenne



4.2.2.3 Exemples

Exemple 1 :

Nous reprenons l'exemple déjà utilisé page 23.

- Soit f la fréquence cible, f = 0,4.
- Soit I la précision relative, I = 25%.
- On choisit 1-α =0,95, avec un risque partagé (intervalle de confiance bilatéral).

L'abaque 11 fournit la valeur n=100.

Exemple 2 :

On teste le taux de non-détection d'un capteur.

On souhaite une précision relative de 10%.

Les indications fournies par le constructeur laissent présager une fréquence-cible de non détection de l'ordre de 10%.

L'abaque n°2 montre qu'il faudra disposer d'un échantillon de 3400 véhicules au minimum.

Exemple n°3 :

Au milieu de la nuit, le taux de franchissement de feux rouges peut être de l'ordre de 15%.

Si on veut établir ce résultat avec une incertitude relative de 20%, l'abaque n°3 montre qu'il faudra utiliser un échantillon de 600 véhicules au minimum.

Remarque : rappelons que ceci signifie que l'on vise un résultat dont l'ordre de grandeur est :

12% < f < 18% avec une confiance de 95%.

Exemple n°4 :

Pendant l'heure fluide diurne, le taux de franchissements illicites n'est que de 5% environ. Pour établir le résultat avec la même précision relative que ci-dessus, c'est-à-dire pour pouvoir affirmer avec une confiance de 95% que 4% < f < 5%, l'abaque n°1 montre qu'il faudra utiliser un échantillon de 2000 véhicules.

Exemple n°5 :

On évalue l'efficacité d'un guidage par PMV[21] en observant, dans un échantillon de n véhicules, le comportement des usagers : suivi ou non des recommandations incitant à utiliser un itinéraire de délestage.

On estime, au vu d'expériences similaires, que le taux de non-respect est élevé, de l'ordre de 45%.

On cherche un ordre de grandeur, et on se satisfait d'une précision médiocre, de l'ordre de 25%. L'abaque n° 11 fournit : n = 80 véhicules.

[21] PMV : panneau à messages variables.

Exemple n°6 :

Exemple traité, par une autre méthode, dans la référence [3], chap. IV, § 4.5.4 :

"Pour estimer une proportion de l'ordre de 0,2, par un intervalle bilatéral symétrique à 0,95 d'amplitude +/- 0,03 il faut un échantillon de taille au moins égale à 683".

Calcul par abaque ZELT : l'amplitude visée correspond à une précision relative I% = 15%. L'abaque n°4 fournit, pour cette valeur de I%, n = 700.

Exemple n°7 :

Adaptation d' exemples issus de [3], chap. IV, § 4.5.4 :

- On estime que dans une fabrication, le pourcentage de pièces défectueuses est de l'ordre de 25%. Quelle est la taille n de l'échantillon permettant de vérifier ce résultat avec une précision relative de l'ordre de 45% ? L'abaque 11 fournit n = 65 pièces.
- Même question avec un pourcentage de pièces défectueuses de l'ordre de 15%, et une précision relative de l'ordre de 33% : l'abaque 11 fournit n = 225 pièces.

4.3 Approximations de la loi binomiale

4.3.1 Position du problème

On trouve dans la littérature de nombreux développements sur l'approximation de la loi binomiale par la loi normale ou par la Loi de Poisson.

Ces approximations se justifient pour des raisons pratiques, ces deux lois étant beaucoup plus faciles à manier que la loi binomiale.

Pour les problèmes qui nous occupent ici, on peut se dispenser d'utiliser de telles approximations si l'on dispose des abaques qui viennent d'être présentées. En effet, ces abaques s'appuient exclusivement sur la loi binomiale, alors que les approximations demandent une certaine rigueur dans leur domaine d'application :

- On admet que la loi binomiale peut être assimilée à une loi normale lorsque la taille de l'échantillon est grande, et la fréquence p pas trop petite. En pratique l'approximation est utilisable lorsque le produit np est supérieur à 20.
- On admet que la loi binomiale peut être assimilée à une loi de Poisson lorsque la taille de l'échantillon est grande, et la fréquence p faible. En pratique l'approximation est utilisable lorsque le produit np est égal à quelques unités ou, plus généralement, quand p <0,1.

Nous ne développons pas ici en détail la théorie de ces approximations. En effet, les abaques décrits plus haut permettent de se libérer des contraintes de calcul, et il ne nous semble pas, dans ces conditions, qu'il y ait un avantage quelconque à substituer à la loi binomiale une approximation.

Nous nous contentons d'évoquer l'approximation par la loi normale qui est d'un emploi très fréquent [22]. Nous rappelons les conditions d'emploi : produit np > 20.

4.3.2 Approximation par la loi normale

On montre que lorsque le produit np est suffisamment grand, la loi binomiale de la fréquence tend vers une loi normale de moyenne p et d'écart-type $\sigma = \sqrt{p(1-p)/n}$.

Dans ces conditions, on montre que la taille n de l'échantillon est donnée par[23] :

$$n \geq \frac{(u_{1-\alpha/2})^2 (1-p)}{I^2 p}$$

[22] On trouvera des développements théoriques plus complets dans les références bibliographiques [3], [4] et [7].

[23] Mêmes notations que dans les paragraphes qui précèdent. On retrouve ici la relation présentée au § 3.2.3.

Exemple

Nous reprenons l'exemple déjà utilisé page 23.

- Soit f la fréquence cible, f = 0,4.
- Soit I la précision relative, I = 25%.
- On choisit 1-α =0,95, avec un risque partagé (intervalle de confiance bilatéral).

La relation ci-dessus donne : $n= (1,96)^2(0,6/0,4)/(0,25)^2 = 92$.

Le produit np est égal à 37 (donc > 20), ce qui légitime l'emploi de l'approximation normale.

On avait trouvé n=100 en utilisant l'abaque de la référence [1] et en utilisant l'abaque ZELT.

Références

Nota : nous avons limité les références aux documents que nous avons effectivement utilisés pour établir ce document. Il ne saurait s'agir d'une bibliographie en matière de calculs statistiques, d'autant moins, comme nous l'avons indiqué plus haut, que seuls des cas simples ont été traités ici.

1. *Tables statistiques*, CISIA-CERESTA, 1997.

2. *Aide-mémoire statistique*, CISIA-CERESTA, 1999.

3. J.Peybernard, *Méthodes statistiques pour l'exploitation de la route*, formation LCPC-CERTU[24], 2000.

4. B. Grais, *Méthodes statistiques, 2° volume : techniques statistiques***,** Dunod, 1998.

5. P. Bailly, *Statistique descriptive*, Presses Universitaires de Grenoble (PUG), 1999.

6. *Le projet CENTAUR à Toulouse*, collectif, coordination ZELT, édition Mairie de Toulouse, 298 p., mars 2000.

7. N.E El Faouzi, *Dimensionnement d'un système de recueil du temps de parcours fondé sur les véhicules traceurs*, Note technique, LICIT (ENTPE-INRETS), août 2000.

[24] Support de formation remis aux stagiaires. Non disponible hors stages.

Annexe : Table u(1-α/2) en fonction de α

α	$u_{(1-\alpha/2)}$
0,01	2,575834515
0,015	2,43238901
0,02	2,326341928
0,025	2,241395123
0,03	2,170090738
0,035	2,108354238
0,04	2,053748176
0,045	2,004653652
0,05	1,959961082
0,055	1,918879207
0,06	1,880789569
0,065	1,845255611
0,07	1,811913535
0,075	1,780463208
0,08	1,750686351
0,085	1,722382876
0,09	1,695398169
0,095	1,669591256
0,1	1,644853