

CONSTRUCTION DES DONNÉES

1. L'OBSERVATION :

Processus incluant l'attention volontaire et l'intelligence, orienté par un objectif terminal ou organisateur, et dirigé sur un objet pour en recueillir des informations.

Ce processus ne s'apparente pas à un mécanisme d'impression photographique dans la mesure où il requiert "une concentration élective de l'activité mentale comportant une augmentation de l'efficacité sur un secteur déterminé et l'inhibition des activités concurrentes" (Piéron, H. in Dictionnaire de la langue pédagogique de Foulquié P. p 40)

"Observer annonce ou suppose un grand travail de l'esprit. On observe ce qu'on étudie ; on remarque ce qui frappe. (...) On observe, cherchant. (...) On remarque quelquefois comme on rencontre, par hasard." (Lafaye, B., *Dictionnaire des synonymes*, Hachette, 1869)

L'observation comme méthode d'investigation de la personne est parfois opposée à la méthode des tests, malgré leur complémentarité.

Cattell (*Handbook of multivariate experimental psychology*, Chicago, Rand Mac Nally, 1966) évoque l'observation comme le processus qui caractériserait la phase inductive de toute recherche.

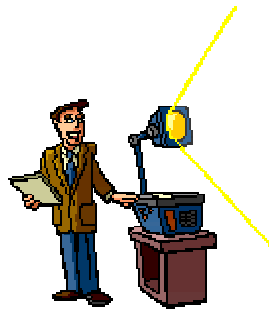
Campbell (*Experimental and quasi-experimental designs for research on teaching, Handbook of research on teaching*, Chicago, Rand Mac Nally, 1963) insiste sur *l'observation provoquée* en tant que processus planifié et provoqué pour confirmer ou infirmer valablement une hypothèse.

En nous appuyant sur les travaux de De Ketele (**28**), nous rapportons ci-après les principaux paramètres qui peuvent caractériser le processus d'observation mis en œuvre dans une perspective de recherche.

systématique	Observation Répétabilité Standardisation Technicité	non systématique
<p><i>Dans la phase exploratoire et inductive, une grande flexibilité est requise pour ne pas négliger des variables significatives en relation avec le problème étudié.</i></p> <p><i>Dans la phase de validation d'une hypothèse, la systématisation doit accrue afin de permettre la répétition des procédures employées dans des conditions semblables.</i></p> <p><i>Le degré de systématisation s'articule sur les paramètres suivants :</i></p> <ul style="list-style-type: none"> - <i>recours à des systèmes de signes</i> - <i>recours à des systèmes de catégories</i> - <i>recours à des systèmes exhaustifs ou non-exhaustifs</i> - <i>recours à des systèmes unidimensionnels ou multidimensionnels</i> - <i>découpage ou non des unités à observer par l'observateur</i> - <i>échantillonnage ou non des unités à observer</i> - <i>recours à des variables qualitatives nominales ou ordinales pour expliciter les attributs.</i> 		

Introspective autospective	Posture de l'observateur	allospective
<i>L'observateur s'observe comme sujet agissant ou observe la situation avec laquelle il interagit. Recours à des techniques introspectives rétroactives (enregistrement audio ou vidéo, vidéoscopie)</i>		<i>L'observateur est une personne distincte du sujet agissant observé et l'objet de l'observation n'est pas une situation avec laquelle il interagit.</i>
Narrative	Pertinence du choix des attributs et de leurs signes. Pertinence du choix des éléments constitutifs du récit	Attributive
<i>Centrée sur le déroulement des actions, la succession des états qui accompagnent l'action. L'attention est portée sur le déroulement qui forme un tout et qui conduit à un récit</i>		<i>Fondée sur des "propositions conçues comme affirmant ou niant une qualité d'un sujet".(proposition d'inhérence) ou fondée sur la relation sujet-prédicat., l'observation consiste alors repérer l'absence ou la présence d'objets, de caractéristiques de ces objets observés.</i>
Objet perceptif	L'attribution d'une signification (inférence) est d'autant plus faible que l'objet à observer est de nature perceptive et d'autant plus grande qu'il est de nature évaluative	Objet évaluatif
<i>L'objet considéré par l'observateur comme un objet de nature perceptive, le conduit à énoncer ce qu'il voit ou ce qu'il entend sans se soucier des significations des faits perceptifs.</i>		<i>L'objet considéré par l'observateur comme un objet de nature évaluative, le conduit à énoncer des intentions, des motifs, des sentiments. Le processus d'attribution se fonde sur les effets du comportement, les relations entre les éléments de la situation, sur l'expérience personnelle de l'observateur.</i>
<i>Il est remarquable que lorsque le niveau d'inférence augmente :</i> <ul style="list-style-type: none"> - le niveau d'accord inter-observateur tend à diminuer ; - la formation à l'observation est plus coûteuse en temps et en effort ; - la spécification et l'opérationnalisation des indicateurs d'observation est plus exigeante ; - la validité est plus questionnée dans la mesure où elle s'appuie sur des éléments observés qui ne sont pas précisés ensuite. 		
Notation immédiate, synchrone	Prise en compte de la durée qui sépare le moment de l'observation de celui de la transcription. Problème de la mémoire.	Notation différée, asynchrone
<i>Dans les deux situations le recours à des techniques permettant de réaliser une observation rétrospective s'avère d'une grande utilité (vidéoscopie)</i>		

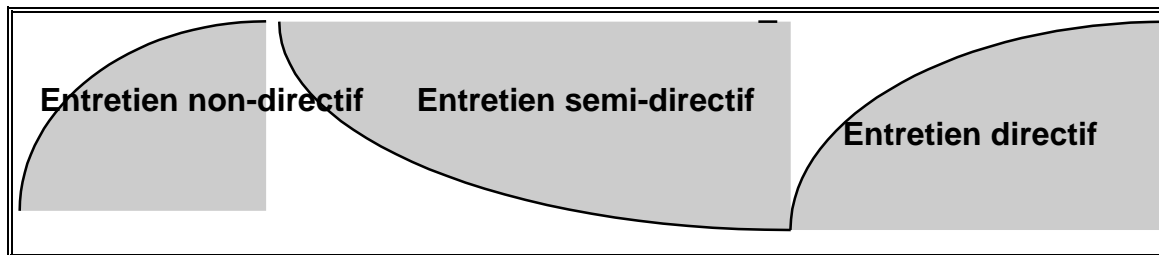
Situation habituelle	Où observer ? Dans quelle condition ?	Situation de laboratoire
Situation manipulée	Quel degré de manipulation choisir ?	Situation non manipulée
Observateur "passif"	Quel est le degré d'implication de l'observateur ?	Observateur "actif"
<i>Selon le degré d'implication de l'observateur dans la situation, on pourra parler d'observation participante passive ou au contraire d'observation participante active. Il convient de prendre compte la mesure dans laquelle l'observateur joue des rôles et tient une place susceptible de modifier radicalement la situation observée.</i>		
Observateur "perçu"	Problème de la réactivité des sujets à l'observation	Observateur "non perçu"
Observation longitudinale		Observation transversale
<i>Vise à recueillir des données sur un objet tout au long d'une période plus ou moins longue</i>	<i>Voir le modèle des antécédents, concomitants et conséquents qui permet de tirer parti de l'observation longitudinale (De Ketele Sciences de l'éducation, 1976, IX, 1)</i>	<i>Vise à recueillir des données sur un objet à un moment précis dans une situation plus ou moins circonscrite.</i>



2. L'entretien

Choisir L'ENTRETIEN, c'est opter pour certaines conditions méthodologiques :

- un rapport verbal entre le chercheur et la personne interrogée (informateur),
- un entretien provoqué par le chercheur
- un entretien à des fins de recherche
- un entretien basé sur l'utilisation d'un guide d'entretien
- un entretien dans une perspective intensive



- L'entretien directif mené sur la base d'un questionnaire
 - L'entretien non directif articulé uniquement sur un thème général qu'il s'agit de faire explorer par la personne interrogée.
 - L'entretien semi-directif pour permettre à la personne interrogée, à la fois, de structurer lui-même sa pensée autour de l'objet d'étude de se limiter au thème étudié en éliminant, par la directivité des questions et des reformulations, des considérations par lesquelles la personne se laisse entraîner et en requérant des approfondissements sur des points qu'elle n'aurait pas développés par- elle-même.
- Les stratégies d'intervention du chercheur dans la méthode de production des données fondée sur l'entretien :
- Le plan de l'entretien :**
- guide d'entretien : ensemble des points qu'il souhaite explorer
 - mode d'intervention : l'usage du guide, ce qu'il attend comme discours, ses attitudes dans l'entretien

Que faire...

Avant

Début

Pendant

Fin

Après

...l'entretien ?

3. Le récit de vie :

Il constitue une méthode de production de données mêlant

- approche biographique par référence à différentes séquences temporelles de la vie de la personne

et

- approche thématique par référence à l'objet d'étude

Cette **approche biographique**.

vise à permettre de:

- sortir de l'opposition individu et société,

- saisir les rapports dialectiques ou de circularité entre point de vue subjectif de l'homme et son inscription dans l'objectivité d'une histoire,

- saisir, dans une perspective interactionniste, les subjectivités, de comprendre comment la conduite est sans cesse remodelée pour prendre en compte les attentes des autres,

- saisir ce qui échappe aux statistiques, aux régularité objectives dominantes, aux déterminismes macrosociologiques,

- reconnaître une valeur sociologique au savoir individuel,

Si le **récit de vie** reste une histoire singulière, unique, il n'en constitue pas moins une histoire qui individualise l'histoire sociale collective d'un groupe, en tant que produit et expression de ce groupe.

4. Le questionnaire

4.1. La mise au point d'un *questionnaire* nécessite de prendre en compte :

Forme :

- quelle architecture générale donner au questionnaire ?
- comment et dans quel ordre transcrire et disposer les questions ?
- comment gérer les articulations, les filtres, les branchements ?
- sous quelles formes recueillir les réponses ?
- quelle forme adopter pour rendre possible une saisie informatique ?

Contenu :

- comment et avec quel vocabulaire doit s'adresser au répondant ?
- quelle formulation adopter pour chaque question ?
- quelles explications des termes employés doit-on fournir ?
- comment aborder les questions qui touchent à des thématiques sensibles ?
- comment ne pas lasser, braquer, influencer le répondant ?

Le mode d'enquête : administration du questionnaire

administré par l'intermédiaire du chercheur :

- face à face
- téléphone

auto-administré

- télécopie
- courrier électronique
- courrier postal
- formulaire sur site Web

4.2. Forme des questions :

4.2.1. Question fermée

➤ question dont les modalités ou les valeurs de réponse sont imposées en nombre fini et même restreint

- réponse unique à choisir parmi deux modalités type : oui /non

- réponse unique à choisir parmi plus de deux modalités ordonnées ou non :

exemple de question à modalités ordonnées :

Que pensez-vous de la présence d'un cours sur les N.T.I.C. appliquées à la recherche et à la formation en sciences de l'éducation ?

formulation 1

1=tout à fait contre	2=plutôt contre	3=plutôt pour	4=tout à fait pour
----------------------	-----------------	---------------	--------------------

formulation 2

tout à fait contre(-2)	plutôt contre(-1)	moyen(0)	plutôt pour(+1)	tout à fait pour(+2)
------------------------	-------------------	-----------------	-----------------	----------------------

- réponses multiples à choisir en nombre précisé ou non parmi plusieurs modalités

- réponse = rangement ordonné par préférence (*exemple* : cela revient à choisir une réponse parmi les n! rangements possibles dans le cas où n items sont proposés, qu'ils doivent être tous pris en compte et que les ex æquo ne sont pas admis)

4.2.2.- Questions semi-ouvertes

exemple

quelle est votre profession ?

type A type B type C Autre = précisez

4.2.3. Question filtre

exemple

Remplissez cette condition ?

si **oui** allez aux questions du bloc A

si **non** allez aux questions du bloc B

4.2.4. - Questions ouvertes

variables quantitatives : nombre

variables qualitatives : réponse textuelle

4.3. La modalité : non-réponse

Faut-il la proposer ou non ?

question formulation 1

A votre avis l'Etat devrait-il augmenter le coût de l'inscription à l'université et doter chaque étudiant d'un matériel informatique multimédia, ou rester dans la situation actuelle en incitant chaque étudiant à s'équiper par lui-même?

question formulation 2

A votre avis l'Etat devrait-il augmenter le coût de l'inscription à l'université et doter chaque étudiant d'un matériel informatique multimédia, ou rester dans la situation actuelle en incitant chaque étudiant à s'équiper par lui-même ou encore vous n'avez pas d'avis sur cette question ?

4.4. Des effets remarquables...

Effet d'ordre d'énumération des modalités

Tendance à choisir la première des modalités proposées

Effet de la modalité médiane

Dans le cas de question à modalités ordonnées (variable ordinale), tendance à répondre en choisissant une position centrale des possibilités offertes représentées par la valeur ou la modalité médiane d'une variable-question fermée :

<i>Tout à fait en désaccord</i>	<i>Plutôt en désaccord</i>	<i>Point de vue mitigé</i>	<i>Plutôt d'accord</i>	<i>Tout à fait d'accord</i>
<i>Jamais</i>	<i>rarement</i>	<i>quelquefois</i>	<i>souvent</i>	<i>toujours</i>

Effet de formulation :

Ces trois formulations ne conduisent pas au même :

Formulation 1

Pensez-vous que fumer doive être interdit dans les lycées ?

Formulation 2

Pensez-vous que fumer doive être autorisé dans les lycées ?

Formulation 3

Pensez-vous que dans les lycées fumer doive être ... ? interdit autorisé

Effet du vocabulaire:

Biais dû à la difficulté que pose l'emploi du vocabulaire non usuel.

Biais dû à l'erreur de compréhension.

Biais dû à l'ambiguïté des termes utilisés

Effet de la longueur du questionnaire :

Biais de l'esprit fatigué
Biais de la nécessité de prendre du temps pour répondre

Effet de halo :

Orientation des réponses à une question par les réponses aux questions précédentes

Effet de la tendance au conformisme social de l'époque :

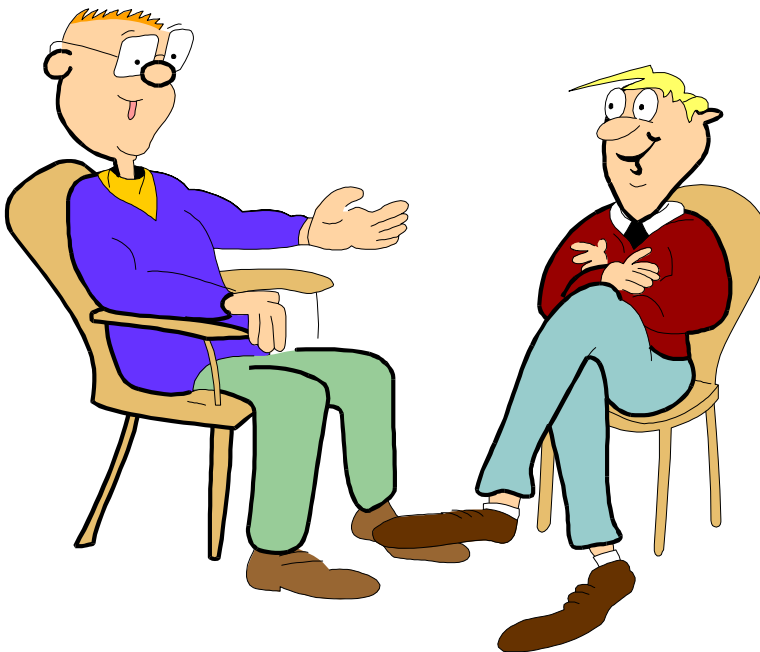
Les réponses correspondent à l'image socialement admise à laquelle il convient d'être conforme.

Effet de la réaction de prestance :

Les réponses correspondent à l'image valorisante que la personne souhaite donner d'elle.

Effet de la tendance à l'acquiescement :

Les réponses privilégient le oui plutôt que le non



4.5. Les questions sensibles et procédures des réponses aléatoires

méthode de Warner

Respecter l'anonymat

Résumé de l'article de Jean-Claude Régnier paru dans le n°16 du bulletin du CLERSE de novembre 1992:

Les instruments de raisonnement humain fournis par la statistique et le calcul des probabilités peuvent parfois nous rendre de grand service dans des travaux d'investigation lorsque ceux-ci reposent sur des données recueillies au moyen de questions adressées directement à des personnes dont il importe absolument de préserver l'anonymat.

Par quels dispositifs est-il possible d'accéder à des informations inscrites dans le jardin secret des individus?

Comment peut-on procéder pour respecter ce secret?

L'article qui suit se propose de reprendre les démarches de Warner¹ (1965) traitant de cette question et de les mettre en œuvre dans le domaine d'investigations en Sciences de l'Education.

Nombre de problématiques abordées dans le cadre des Sciences de l'Education peuvent ou pourraient s'appuyer sur des données appartenant au jardin secret des individus concernés par les investigations. Nous pouvons aussi penser que la multiplication des sondages, les propos qu'ils induisent dans les médias, la corrélation plus ou moins implicite qui est établie avec l'informatique et sa puissance de stockage et de traitement d'informations privées mettent ces individus dans un état de méfiance préjudiciable quant à la fiabilité des données requises.

Le législateur a lui-même établi des garde-fous au travers d'une législation fondée sur les deux thèmes "informatique" et "liberté".

Ainsi est-il pertinent d'imaginer des dispositifs permettant de recueillir des informations en préservant de manière absolue l'anonymat des personnes interrogées et de leur associer des traitements statistiques adéquats.

Identifions quelques thèmes où cet anonymat est un préalable absolu :

- la fraude aux examens universitaires,
- l'usage de "drogue" lors de la préparation d'examens universitaires,
- la pratique du vol de documents, de livres dans les centres de documentation,
- le "piratage" informatique,
- la mise au point de logiciels "virus" et les stratégies élaborées pour leur propagation,
- l'effet des campagnes d'information relatives à la question du SIDA

Tous ces thèmes peuvent donner lieu à des problématiques inscrites dans le cadre des Sciences de l'Education. Or force est de constater qu'elles impliquent la prise en compte de données intimes aux individus et requérant la préservation du secret.

A titre d'exemple pour exposer la procédure choisie, nous opterons pour une étude du comportement de fraude lors d'examens universitaires.

Un enseignant ayant eu l'écho de quelques rumeurs sur des pratiques de copiages lors d'examens universitaires pourra avoir envie d'évaluer la part de réalité de ce fait. Il pourra

¹ Cf l'ouvrage de Lebart et alii "Traitement des données statistiques" Dunod 1982 pages 74 à 76

Cet ouvrage rapporte que S.L. Warner en 1965 propose une stratégie dans un article intitulé "Randomized Response: a Survey Technique for Eliminating Evasive Answer Bias" J.A.S.A. vol 60 pp 63-69

Un numéro de «International Statistic Review» (aout 1976) contient les communications relatives à cette question lors d'un congrès à Varsovie en septembre 1975.

s'intéresser à l'occurrence de ces pratiques, aux stratégies employées, aux représentations qui lui sont associées et surtout à la motivation qui détermine le recours à la fraude ainsi qu'à la prise de conscience des divers risques encourus.

Pour la clarté de cet exposé, nous nous réduisons à la question:

Avez-vous fraudé au moins une fois aux examens de l'année universitaire précédente?

La question initiale appelle alors deux modalités de réponse:

A_1 = « *j'ai fraudé au moins une fois* »

A_2 = « *je n'ai jamais fraudé* »

Nous excluons ici l'hypothèse d'un comportement de mensonge systématique. Nous proposons le protocole suivant :

« Je souhaite recueillir une information relative à la fraude aux examens universitaires pour infirmer ou confirmer des rumeurs à ce propos.

Avez-vous fraudé au moins une fois aux examens de l'année universitaire précédente ?

oui

non

Toutefois pour préserver le secret total de votre réponse, je vous propose la démarche suivante :
Voici une urne contenant 20 boules numérotées de 1 à 20.

Dans l'isoloir, vous tirez une boule, vous prenez connaissance de son numéro et vous la remettez dans l'urne.

Si le numéro tiré est inférieur ou égal à 5, je vous demande de répondre par OUI (c'est vrai) ou NON (c'est faux) à l'assertion

A_1 = « *j'ai fraudé au moins une fois* »

Si le numéro tiré est supérieur ou égal à 6, je vous demande de répondre par OUI (c'est vrai) ou NON (c'est faux) à l'assertion

A_2 = « *je n'ai jamais fraudé* »

Mentionnez votre réponse sur le petit carton mis à votre disposition et mettez ce carton dans la seconde urne.

La constitution de l'échantillon des étudiants à interroger soulève un problème concret de faisabilité :

on peut procéder à un tirage aléatoire² de plus d'une centaine d'étudiants parmi ceux de l'université concernée à partir des numéros d'inscription et les inviter par courrier à venir répondre à l'enquête à la manière des consultations électorales.

Comment pourrions-nous interpréter les abstentions, c'est à dire ceux qui ne viendront pas s'exprimer ?

Comment pouvons-nous modéliser ce paramètre lié au comportement qui se traduit par venir et s'exprimer ou ne pas venir ?

Quel biais peut introduire ce paramètre ?

Par cette démarche, nous recueillons un ensemble de n bulletins portant les réponses OUI ou NON sans savoir si elles se rapportent à A_1 ou à A_2 .

Le secret de chacun est donc respecté.

C'est alors que le recours à un raisonnement probabiliste va nous aider à obtenir une estimation de la proportion *des étudiants ayant fraudé au moins une fois*.

² à l'aide d'une table de nombres au hasard

Cette fois nous devons introduire un peu plus de formalisme mathématique en espérant ne pas induire un découragement chez le lecteur non initié.

<i>notation</i>	<i>information</i>	<i>état</i>
π	proportion réelle des étudiants ayant fraudé au moins une fois dans la population	inconnue
θ	probabilité d'être invité à répondre à A_1	connue ³
$1 - \theta$	probabilité d'être invité à répondre à A_2	connue
ρ	probabilité que l'étudiant interrogé ait répondu OUI à l'une ou l'autre des assertions A_1 ou A_2	inconnue

<i>notation</i>	<i>événement</i>	<i>probabilité</i>
A_1	répondre à l'assertion A_1	$P(A_1) = \theta$
A_2	répondre à l'assertion A_2	$P(A_2) = 1 - \theta$
«OUI»	répondre OUI à A_1 ou A_2	$P(\text{«OUI»}) = \rho$
$(\text{«OUI»}/A_1)$	répondre OUI sachant qu'il s'agit de répondre à A_1	$P(\text{«OUI»}/A_1) = \pi$
$(\text{«OUI»}/A_2)$	répondre OUI sachant qu'il s'agit de répondre à A_2	$P(\text{«OUI»}/A_2) = 1 - \pi$

L'événement «OUI» est réalisé par les éventualités qui elles-mêmes réalisent soit l'événement «OUI» et A_1 soit l'événement «OUI» et A_2 .

$$\text{«OUI»} = (\text{«OUI» et } A_1) \text{ ou } (\text{«OUI» et } A_2)$$

Ainsi en terme de probabilités :

$$P(\text{«OUI»}) = P(\text{«OUI» et } A_1) + P(\text{«OUI» et } A_2)$$

Le recours à la notion de "probabilité conditionnelle" nous permet d'obtenir les relations suivantes :

$$P(\text{«OUI» et } A_1) = P(\text{«OUI»}/A_1) P(A_1)$$

$$P(\text{«OUI» et } A_2) = P(\text{«OUI»}/A_2) P(A_2)$$

$$P(\text{«OUI»}) = P(\text{«OUI»}/A_1) P(A_1) + P(\text{«OUI»}/A_2) P(A_2)$$

$$P(\text{«OUI»}) = \pi\theta + (1 - \pi)(1 - \theta) = \rho$$

En transformant cette relation nous extrayons une formule fournissant la valeur qui nous préoccupe à savoir celle de π

$$\pi = \frac{\rho + \theta - 1}{2\theta - 1}$$

Si maintenant nous tenons la variable "répondre à l'assertion" comme une variable aléatoire, nous avons alors à faire à une variable de Bernoulli puisqu'il n'y a que deux modalités «OUI» et «NON».

événements	«OUI»	«NON»
probabilité	ρ	$1 - \rho$

En interrogeant les n étudiants de l'échantillon, nous pouvons définir la variable aléatoire X qui décrit le nombre de "oui" émis parmi les n réponses.

³ La valeur est déterminée par la procédure aléatoire choisie: par exemple la composition de l'urne selon la proportion des tirages favorables à A_1 et le mode de tirage selon qu'il s'agit d'un tirage avec ou sans remise.

X est une variable binomiale de paramètres n et ρ : $X = B(n; \rho)$

Nous connaissons:

- l'espérance mathématique de X : $E(X) = n\rho$
- la variance de X : $\sigma^2(X) = V(X) = n\rho(1 - \rho)$

Nous pouvons définir la variable aléatoire Y caractérisant la proportion de réponses "oui" dans l'échantillon.

$$Y = \frac{X}{n} \text{ avec } E(Y) = \rho \text{ et } \sigma^2(Y) = V(Y) = \frac{\rho(1-\rho)}{n}$$

Ceci nous conduit à procéder à une estimation ponctuelle de la valeur ρ par utilisation d'une variable aléatoire : la statistique $\hat{\rho} = \frac{X}{n} = Y$ dont les caractéristiques sont $E(\hat{\rho}) = \rho$ et $V(\hat{\rho}) = \frac{\rho(1-\rho)}{n}$

A partir de cette relation, nous pouvons construire un estimateur $\hat{\pi}$ de la valeur π inconnue.

$$\text{estimateur de } \pi = \hat{\Pi} = \frac{Y + \theta - 1}{2\theta - 1}$$

Nous connaissons là encore:

- l'espérance mathématique de $\hat{\pi}$: $E(\hat{\pi}) = \frac{E(Y) + \theta - 1}{2\theta - 1} = \pi$
- la variance de $\hat{\pi}$: $V(\hat{\pi}) = \frac{1}{(2\theta - 1)^2} V(\hat{\rho}) = \frac{1}{(2\theta - 1)^2} \left(\frac{\rho(1-\rho)}{n} \right)$

De cette relation, nous parvenons à une expression en fonction de π ,

$$V(\hat{\Pi}) = \frac{\pi(1-\pi)}{n} + \frac{\theta(1-\theta)}{n(2\theta-1)^2}$$

Dans notre exemple, nous supposons que notre échantillon est de $n = 150$ individus extraits aléatoirement d'une population dont l'effectif total N est inconnu.

Le dépouillement a fourni les résultats suivants:

84 "oui" ont été dénombrés dans l'urne

notation	événement	probabilité
A_1	répondre à l'assertion A_1	$P(A_1) = \frac{5}{20}$
A_2	répondre à l'assertion A_2	$P(A_2) = \frac{15}{20}$
«OUI»	répondre OUI à A_1 ou A_2	estimation ponctuelle $P(\text{«OUI»}) = \frac{84}{150}$

Ceci nous donne donc une estimation ponctuelle de la proportion des étudiants *ayant fraudé au moins une fois* au sein de la population entière.

$$\text{estimation de } \pi = \frac{\frac{84}{150} + \frac{5}{20} - 1}{2 \frac{5}{20} - 1} = \frac{0,19}{0,5} = 0,38$$

Ainsi il y aurait 38% d'étudiants *ayant fraudé au moins une fois*

En raison de l'effectif important de l'échantillon, nous pouvons approcher la variable

$$Z = \frac{\hat{\Pi} - E(\hat{\Pi})}{\sqrt{V(\hat{\Pi})}} = \frac{\hat{\Pi} - \pi}{\sqrt{V(\hat{\Pi})}} \text{ par la variable de Laplace-Gauss LG}(0;1).$$

La variance $V(\pi)$ est inconnue mais nous pouvons l'estimer par sa valeur maximale correspondant à la valeur $\pi = 0,5$.

$$V(\pi) \approx \frac{1}{4n} + \frac{\theta(1-\theta)}{n(2\theta-1)^2} = \frac{1}{600} + \frac{0,1875}{37,5} \approx 0,00666$$

$$\sigma(\pi) \approx 0,0816$$

Enfin nous pouvons obtenir l'intervalle de confiance⁴ à 95% suivant:

$$(0,38) - 1,96\sigma(\pi) < \pi < (0,38) + 1,96\sigma(\pi)$$

c'est à dire: $0,2199 < \pi < 0,5400$

Ainsi peut on s'attendre avec une confiance de 95% à une proportion π d'étudiants

ayant fraudé au moins une fois comprise entre 22% et 54%. Notons que la variance $V(\pi)$ inconnue peut aussi être estimée par la valeur correspondant à la valeur $\pi = 0,38$ issue de l'estimation ponctuelle.

$$V(\pi) \approx \frac{1}{4n} + \frac{\theta(1-\theta)}{n(2\theta-1)^2} = \frac{0,2356}{150} + \frac{0,1875}{37,5} \approx 0,00657$$

$$\sigma(\pi) \approx 0,0810$$

Ce qui nous permet d'obtenir l'intervalle de confiance à 95% suivant:

$$(0,38) - 1,96\sigma(\pi) < \pi < (0,38) + 1,96\sigma(\pi)$$

c'est à dire:

$$\boxed{0,2211 < \pi < 0,5388}$$

⁴ une "fourchette" d'estimation de la proportion p

5. Méthode de production de données et NTIC

- Enquêtes par courrier électronique
- Enquêtes par site WEB

<i>Avantages :</i>	<i>Inconvénients:</i>
méthode très rapide et peu coûteuse, court délai de réponse Va directement au répondant	Ne convient pas à l'échantillonnage aléatoire Impossibilité de contrôler l'entrée des données (idem enquête par poste)
Le répondant peut remplir le questionnaire au moment qui lui convient	Les questionnaires doivent être brefs, en raison des limites des systèmes de courrier électronique des répondants
Moindre intrusion qu'avec l'enquête par téléphone	Impossibilité de traiter des questions par grille ou des enchaînements
Possibilité de communiquer avec des groupes cibles habituellement difficiles à joindre	Nécessité d'inclure des mesures incitatives efficaces pour obtenir un bon taux de réponse

