# Tp1 – Sondages

Alain LATOUR

15 janvier 2021

## Avant-propos

Vous réalisez aujourd'hui votre première séance de travail pratique (TP) du cours de sondages. Le travail fait durant les séances de TP est évalué. Il vous faut donc être actif et participatif.

Pour celui-ci vous travaillerez en binôme. Pour connaître votre « partenaire binomial » voir sur le site :
https ://www-ljk.imag.fr/membres/Alain.Latour/Cours/M1/SOND/germes/germe.html

**Le compte rendu est à rendre le 18 janvier 2021**

## 1 Contexte

Lors d'une enquête, une question sensible doit être traitée avec une attention particulière. Imaginons une enquête dont un des objectifs serait d'estimer une proportion d'individus dans une classe correspondant à une situation un peu *« délicate. »* Par exemple, on souhaiterait estimer la proportion d'individus faisant usage de drogue. Ce serait une maladresse que de poser directement la question : *« Faites-vous usage de drogue ? »*) Les répondants pourraient préférer ne pas répondre.

Citons à ce propos (Lohr, 2019, p. 540) :

> *Dans l'enquête nationale auprès des ménages sur l'abus de drogues on interroge les répondants sur leur consommation de marijuana et de cocaïne. Certains répondants considèrent ces questions comme intrusives ; d'autres craignent que le fait de fournir des renseignements exacts puisse les exposer à des pénalités (par exemple, ils craignent que la déclaration de leur revenu réel dans le cadre d'une enquête puisse entraîner des pénalités pour le paiement insuffisant de l'impôt sur le revenu). Certaines personnes peuvent protéger leurs renseignements personnels en refusant de répondre au sondage ou à des questions précises, tandis que d'autres peuvent donner des réponses inexactes à des questions de nature délicate.*

Il existe quelques astuces permettant de rassurer le répondant et d'estimer la proportion en question. Pour plus de détails, voir (Lohr, 2019, Section 15.4), Fox et Tracy (1986) et (Cochran, 1977, Section 13.17).

# 2 Estimation d'une proportion dans le contexte d'une question délicate

## 2.1 Méthode de Warner

L'interviewer dispose d'une pile de fiches de deux types, voir Warner (1965). Sur certaines est inscrite la proposition *« Je fais usage de drogue »* (fiche A), et sur les autres *« Je ne fais pas usage de drogue »* (fiche B). Supposons que $\theta$ soit la proportion de fiches sur lesquelles figure la proposition *« Je fais usage de drogue »* . La proportion de fiches avec l'affirmation : *« Je ne fais pas usage de drogue »* est alors de $1-\theta$.

Le répondant tire au hasard une fiche dans cette pile. Sans que l'interviewer ne la voit et sans que le répondant lui communique la question, ce dernier doit simplement dire si la proposition inscrite sur la fiche choisie s'applique à lui ; il répond par *oui* ou *non*. Le but du sondage est d'estimer $\pi_A$, la proportion des gens qui font usage de drogue.

Faisons un premier calcul qui nous conduira assez rapidement à une estimation de $\pi_A$ : évaluons la probabilité d'avoir *« oui »* comme réponse :

$$\Pr\{\text{« oui »}\} = p_{\text{oui}} = \theta\pi_A + (1-\theta)(1-\pi_A) = (2\theta-1)\pi_A + (1-\theta)$$

On peut alors résoudre cette équation pour trouver $\pi_A$.[1] Utilisons Maxima :

```
(% i10) p[oui] = (2*theta-1)*pi[A]+(1-theta) ;
        solve([%], [pi[A]]) ;
```

$$p_{\text{oui}} = \pi_A\,(2\theta-1) - \theta + 1 \qquad (\%\ \text{o9})$$

$$[\pi_A = \frac{\theta + p_{\text{oui}} - 1}{2\theta - 1}] \qquad (\%\ \text{o10})$$

Sous une forme peut-être plus attrayante :

$$\pi_A = \frac{p_{\text{oui}} - (1-\theta)}{2\theta - 1}, \qquad \theta \neq \tfrac{1}{2}$$

Explicitons maintenant l'estimateur. Soit $X$ le nombre de personnes dans un échantillon de taille $n$ tiré avec remise qui répondent « *oui* » après avoir lu la fiche. L'estimateur sans biais et à vraisemblance maximale de la proportion de *« oui »* est : $X/n$. Il semble raisonnable d'estimer la proportion qui nous intéresse par :[2]

$$\hat{\pi}_{AW} = \frac{X/n - (1-\theta)}{2\theta - 1}, \qquad \theta \neq \tfrac{1}{2}$$

La variable $X$ nous est familière. Il s'agit d'une variable binomiale ou hypergéométrique selon que le choix des personnes interrogées se fait avec ou sans remise. Dans le premier cas la probabilité d'un succès est constante et vaut $(2\theta-1)\pi_A + (1-\theta)$. Nous développons la méthode pour ce cas...

L'astuce, s'il en est une est d'écrire :

$$\hat{\pi}_{AW} = -\frac{(1-\theta)}{2\theta-1} + \frac{1}{n(2\theta-1)} \times X$$

1. L'équation est très simple à résoudre. Nous utilisons Maxima uniquement pour nous familiariser avec le logiciel.
2. On dénote l'estimation par $\hat{\pi}_{AW}$ car elle est donnée par la méthode de Warner.

Il apparaît alors que $\hat{\pi}_{AW}$ est une transformation affine de $X$ :

$$\hat{\pi}_{AW} = a + b \times X$$

avec $a = -(1-\theta)/(2\theta-1)$ et $b = 1/(n(2\theta-1))$.

Nous savons que dans ce cas :

$$\begin{aligned} \mathrm{E}[\hat{\pi}_{AW}] &= a + b\mathrm{E}[X] \\ \mathrm{Var}[\hat{\pi}_{AW}] &= b^2\,\mathrm{Var}[X] \end{aligned}$$

On vérifie sans difficulté que l'estimateur est sans biais : $\mathrm{E}[\hat{\pi}_{AW}] = \pi_A$. Sa variance est :

$$\begin{aligned} \mathrm{Var}[\hat{\pi}_{AW}] &= \frac{1}{n^2(2\theta-1)^2}\,\mathrm{Var}[X] \\ &= \frac{p_{\text{oui}}(1-p_{\text{oui}})}{n(2\theta-1)^2} \\ &= \frac{(\theta - \pi_{AW}\,(2\theta-1))\,(\pi_{AW}\,(2\theta-1) - \theta + 1)}{n\,(2\theta-1)^2} \end{aligned}$$

## 2.2 Méthode de la question complémentaire innocente

Une autre façon de faire est de remplacer l'affirmation *« Je ne fais pas usage de drogue »* par une affirmation auxiliaire qualifiée parfois *« d'innoncente. »* Par exemple : *« Je suis gaucher »* ou *« Je suis né au mois de mai. »* On utilise des questions pour lesquelles on connait la vraie proportion ou que l'on peut estimer facilement,Horvitz et collab. (1967). On connait la proportion de gens qui sont gauchers. En France, elle est de $\alpha = 12,7\%$. La proportion de gens nés en mai est $\alpha = 31/365 = 8,49\%$.

La probabilité qu'une personne interrogée réponde par l'affirmative est alors :

$$\Pr\{\text{« oui »}\} = \theta\pi_A + (1-\theta)\alpha = p_{\text{oui}}$$

En isolant $\pi_A$, cela nous conduit à [3]

$$\hat{\pi}_{AI} = \frac{X/n - \alpha(1-\theta)}{\theta}$$

Tout comme l'estimateur de Warner, cet estimateur est sans biais. [4] Sa variance est :

$$\begin{aligned} \mathrm{Var}[\hat{\pi}_{AI}] &= \frac{1}{n^2\theta^2}\,\mathrm{Var}[X] \\ &= \frac{p_{\text{oui}}(1-p_{\text{oui}})}{n\theta^2} \end{aligned}$$

**Remarque :** La valeur de $p_{\text{oui}}$ dépend de la procédure. Dans la première, elle est : $p_{\text{oui}} = \theta\pi_A + (1-\theta)(1-\pi_A)$ tandis que dans la seconde, elle est $p_{\text{oui}} = \theta\pi_A + (1-\theta)\alpha$. On note aussi une différence au niveau de la variance des estimateurs. Cela signifie qu'une approche peut être plus efficace qu'une autre. Cela dépend, d'une part, du *« paramétrage »* de la procédure, c'est-à-dire des valeurs $\theta$ et possiblement $\alpha$ et, d'autre part, de la valeur de $\pi_A$ qui n'est pas connue puisque l'on cherche à l'estimer.

On montre que (voir Dowling et Schactman (1975)) que $\mathrm{Var}[\hat{\pi}_{AI}] \leqslant \mathrm{Var}[\hat{\pi}_{AW}]$ quels que soient $\pi_A$ et $\alpha$ pour des valeurs de $\theta$ supérieures à un peu plus d'un tiers.

3. On dénote l'estimation par $\hat{\pi}_{AI}$ car la méthode utilise une question « innocente. »
4. On laisse en exercice au lecteur le soin de le vérifier.

## 2.3 Programmation

On peut facilement simuler en R chacune de ces méthodes.

Script 1 – Simulation de la première méthode

```
1    # Choix de la question
2    question <- sample(c(1,2),n,prob=c(theta,1-theta),replace=TRUE)
3    # Choix du répondant
4    repondant <- sample(c(1,2),n,prob=c(P,1-P),replace=TRUE)
5    #
6    nb.oui <- sum(question-repondant==0)
```

Dans le script 1, pour une taille et une proportion fixées, $n$ et $\theta$, on détermine à la ligne 2 quelle affirmation est choisie pour chacun des répondants. Dans le vecteur `question`, on trouve une suite de 1 et de 2. Évidemment, si c'est 1, l'individu sera confronté à la première affirmation, sinon, à la seconde. À la ligne 4, on détermine à quelle classe appartient le répondant : à la classe 1 avec probabilité `P`, à la seconde avec probabilité $1 - $ `P`. Sous l'hypothèse où tous répondent honnêtement à l'affirmation, le nombre de « oui » est évalué à la ligne 6.

On peut donc répéter cette expérience un grand nombre fois et se faire une idée assez précise de la distribution de $\hat{\pi}_{AW}$. Les exercices qui suivent permettront de vous familiariser avec la méthode et de connaître les propriétés de cet estimateur. On a écrit pour vous une fonction qui fait le travail :

```
1 questDelicateMeth1 <- function(n, P, theta, n.Sim, p.neg=0)
```

Dans cet appel, `n.Sim` est le nombre de fois où on refait l'expérience d'interroger $n$ individus en utilisant la méthode 1. Le paramètre `p.neg` est la proportion de gens de la classe $A$ qui refusent de répondre à l'affirmation délicate. Par défaut, cette valeur est 0 et correspond au fait que le répondant a une complète confiance en la méthode. Cette fonction retourne un vecteur donnant les `n.Sim` valeurs estimées de $\hat{\pi}_{AW}$.

Pour la seconde méthode, on dispose d'une autre fonction :

```
1 questDelicateMeth2 <- function(n, P, theta, alpha, n.Sim, p.neg=0)
```

On a ajouté le paramètre `alpha` qui est la probabilité qu'un individu réponde positivement à la seconde affirmation. Cette fonction retourne, elle aussi, un vecteur donnant les `n.Sim` valeurs estimées de $\hat{\pi}_{AW}$.

## 2.4 Exercices

1. Considérez un acte illégal commis par 10% de la population ($\pi_A = 1/10$). Si tous les répondants répondent honnêtement, simuler 1 000 fois la méthode 1 avec $n = 500$ et $\theta = 0{,}8$.
   (a) Quelle est la moyenne estimée et quel est l'écart type estimé de $\hat{\pi}_{AW}$ ?
   (b) Quelles sont les valeurs théoriques correspondantes ?
   (c) Comment se compare la variance de cette méthode avec celle où l'on poserait directement la question dans le contexte où tous les répondants répondent honnêtement ?
   (d) Tracez un histogramme des valeurs observées de $\hat{\pi}_{AW}$.
   (e) Commenter le tout...

2. Reprendre le travail avec la méthode 2 en utilisant $\alpha = 0,217$. Pour ce faire, la `questDelicateMeth2` est disponible.
3. Dans l'exercice précédent, supposons que tous les répondants répondent honnêtement avec l'une des deux méthodes décrites, mais qu'avec la méthode où l'on pose la question directe, certains répondants ayant commis l'acte illégal le nient. Avec $n = 500$, laquelle des méthodes donne une $\text{EQM}(\hat{\pi}_A)$ plus petite que la méthode directe si (i) 0%, (ii) 10%, (iii) 20% et (iv) 30% de ceux qui ont commis l'acte le nient sous la méthode (a) ?
   **Remarque** Pour tout estimateur $\hat{\theta}$ de $\theta$ on a :

$$\text{EQM}(\hat{\theta}) = \text{Var}\left[\hat{\theta}\right] + \text{biais}^2,$$

   où $\text{biais} = \text{E}\left[\hat{\theta}\right] - \theta$.

# 3 Échantillonnage classique

## 3.1 Remarques préalables

La théorie des sondages a été développée afin de répondre aux questions d'inférences dans un contexte où la population cible est finie. Afin d'illustrer certaines nuances et fixer les notations, considérons la population :

| N° de l'individu | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Valeur de la variable | 3 | 6 | 24 | 27 | 30 | 36 | 51 | 57 |

Cette population, qui n'a d'intérêt que d'un point de vue pédagogique, est de taille $N = 8$. La variable aléatoire représentant l'observation de la valeur correspondant à un individu choisi avec équiprobabilité dans cette population pourrait se décrire de la manière suivante :

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $y_i$ | 3 | 6 | 24 | 27 | 30 | 36 | 51 | 57 |
| $p_i$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |

Cette variable a respectivement pour moyenne et variance :

$$\overline{y}_{\mathcal{U}} = \frac{1}{N}\sum_{i=1}^{N} y_i = 29,25$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y}_{\mathcal{U}})^2 = 318,9375$$

En sondages on travaille plutôt avec la variance dite corrigée :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y}_{\mathcal{U}})^2 = 364,5$$

On divise par $N-1$ et non par $N$. On remarque que $S^2 = \sigma^2 \times N/(N-1)$. On utilise $S^2$ car cela simplifie l'écriture des formules. Dans les cours des statistiques élémentaires on indique que la variance de la moyenne échantillonnale est

$$\sigma^2_{\overline{y}} = \frac{\sigma^2}{n}.$$

En sondage, quand nous utilisons un échantillon aléatoire simple sans remises de taille $n$ provenant d'une population finie de taille $N$, la variance est

$$\sigma^2_{\overline{y}} = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}.$$

L'élégance de la formule tient en partie dans le fait que l'on introduit le facteur de correction pour population finie, $\left(1 - \frac{n}{N}\right)$ dans lequel on retrouve $\frac{1}{n}$, la fraction échantillonnale. En pratique cette variance est estimée par

$$\sigma^2_{\overline{y}} = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}.$$

où $s$ est l'écart type échantillonnale. Les exercices qui suivent vous permettront de mettre de l'ordre dans ces différentes notions.

4. On peut apprendre beaucoup de choses en étudiant l'ensemble de tous les échantillons possibles d'une population donnée. Évidemment, cela n'est possible que si $N$, la taille de la population n'est pas trop grande. Prenons par exemple une population comptant 8 individus pour lesquels les valeurs d'une variables sont :

| N° de l'individu | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Valeur de la variable | 3 | 6 | 24 | 27 | 30 | 36 | 51 | 57 |

Intéressons-nous aux échantillons de taille 3. Il y en a $\binom{8}{3} = 56$. En R, il est assez facile de générer tous ces échantillons, voir le Script 2.

Script 2 – Tous les échantillons de taille 3 d'une population de 8 individus

```
> options(OutDec = ",")
> pop <- c(3, 6, 24, 27, 30, 36, 51, 57)
> N <- length(pop)
> n <- 3
> (n.ech <- choose(N,n))
[1] 56
> ech.all <- cbind(choix <- t(matrix(combn(1:N,n),n,n.ech)),(matrix(↘
→pop[choix],n.ech,n)))
> rownames(ech.all) <- seq(1:nrow(ech.all))
> colnames(ech.all) <- c(paste("i",1:n,sep=""),paste("pop[i",1:n,"]"↘
→,sep=""))
> head(ech.all)
  i1 i2 i3 pop[i1] pop[i2] pop[i3]
1  1  2  3       3       6      24
2  1  2  4       3       6      27
3  1  2  5       3       6      30
4  1  2  6       3       6      36
5  1  2  7       3       6      51
6  1  2  8       3       6      57
```

L'exécution de ce script produit la trame de donnée `ech.all` dont les six premiers éléments sont affichés de la ligne 11 à 17. S'inspirer de ce script pour répondre aux questions suivantes.

(a) Calculer $\overline{y}_{\mathcal{U}}$ et $S$ pour la population.

(b) Générer une trame de données contenant tous les échantillons possibles.

(c) À partir cette trame de données, en créer une autre contenant pour chaque échantillon $\overline{y}$, $s$, la borne inférieure et la borne supérieure de l'intervalle de confiance donné par $\overline{y} \pm 1,96\hat{\sigma}_{\overline{y}}$. Finalement, y ajouter une variable `incl` indiquant si l'intervalle de confiance recouvre la vraie valeur de la moyenne. La trame de données débute de la manière suivante :

| | ybar | s | b.inf | b.sup | incl |
|---|---|---|---|---|---|
| 1 | 11 | 11,35782 | 0,8391437 | 21,16086 | 0 |
| 2 | 12 | 13,07670 | 0,3014103 | 23,69859 | 0 |
| 3 | 13 | 14,79865 | -0,2390710 | 26,23907 | 0 |
| 4 | 15 | 18,24829 | -1,3251646 | 31,32516 | 1 |
| 5 | 20 | 26,88866 | -4,0549579 | 44,05496 | 1 |

(d) Vérifier numériquement que $\mu_{\overline{y}} = \overline{y}_{\mathcal{U}}$ et que $\sigma_{\overline{y}} = \sqrt{1-f}\,\dfrac{S}{\sqrt{n}}$.

(e) Est-ce que $s$ est un estimateur sans biais de $S$ ?

(f) Déterminer la probabilité de commettre dans l'estimation de la moyenne une erreur

i. de plus de 2 unités ;

ii. de plus de 5 unités ;

iii. de plus de 25 %.

(g) Quelle est la probabilité de se tromper de plus de 20 % dans l'estimation de l'écart type $S$ de la population ?

(h) Quel est le niveau de confiance de l'intervalle de confiance donné par la formule $\overline{y} - 2\hat{\sigma}_{\overline{y}} \leqslant \overline{y}_{\mathcal{U}} \leqslant \overline{y} + 2\hat{\sigma}_{\overline{y}}$ ?

(i) Quel est le niveau de confiance de l'intervalle de confiance donné par la formule $\overline{y} - 3\hat{\sigma}_{\overline{y}} \leqslant \overline{y}_{\mathcal{U}} \leqslant \overline{y} + 3\hat{\sigma}_{\overline{y}}$ ?

(j) Quel est le niveau de confiance de l'intervalle de confiance donné par la formule $\overline{y} - 2\sigma_{\overline{y}} \leqslant \overline{y}_{\mathcal{U}} \leqslant \overline{y} + 2\sigma_{\overline{y}}$ ?

5. Considérer une population dont la distribution des pointures de chaussures ($y$) est la suivante : [5]

| $i$ | $y_i$ | $f_i$ | $i$ | $y_i$ | $f_i$ |
|---|---|---|---|---|---|
| 1 | 36 | 0,02 | 12 | $41\frac{1}{2}$ | 0,04 |
| 2 | $36\frac{1}{2}$ | 0,03 | 13 | 42 | 0,04 |
| 3 | 37 | 0,04 | 14 | $42\frac{1}{2}$ | 0,03 |
| 4 | $37\frac{1}{2}$ | 0,06 | 15 | 43 | 0,03 |
| 5 | 38 | 0,10 | 16 | $43\frac{1}{2}$ | 0,02 |
| 6 | $38\frac{1}{2}$ | 0,12 | 17 | 44 | 0,02 |
| 7 | 39 | 0,13 | 18 | $44\frac{1}{2}$ | 0,01 |
| 8 | $39\frac{1}{2}$ | 0,10 | 19 | 45 | 0,01 |
| 9 | 40 | 0,07 | 20 | $45\frac{1}{2}$ | 0,01 |
| 10 | $40\frac{1}{2}$ | 0,05 | 21 | 46 | 0,01 |
| 11 | 41 | 0,05 | 22 | $46\frac{1}{2}$ | 0,01 |

Prélever 1 000 d'échantillons de taille 5 et tâcher de répondre aux questions suivantes du mieux que possible, par simulation. La population est considérée infinie.

---

Script 3 – Commande pour générer un échantillon aléatoire simple avec remise de taille $n$ de cette population

```
# Descriptif de population
# Valeurs
y <- c(36,  36.5 , 37,  37.5 , 38,  38.5 , 39,  39.5 , 40,  40.5 , ↘
→41,  41.5 , 42,  42.5 , 43,  43.5 , 44,  44.5 , 45,  45.5 , 46,  ↘
→46.5)
# Probabilités
e <- c(0.02, 0.03, 0.04, 0.06, 0.10, 0.12, 0.13, 0.10, 0.07, 0.05, ↘
→0.05, 0.04, 0.04, 0.03, 0.03, 0.02, 0.02, 0.01, 0.01, 0.01, 0.01, ↘
→0.01)
# Un échantillon
echt <- sample(y,n,replace=TRUE,prob=e)
```

---

(a) Dans quelle mesure vos simulations confirment-elles que $\overline{y}$ est un estimateur sans biais de $\overline{y}_{\mathcal{U}}$ ?

(b) Calculer la variance des $\overline{y}$. Théoriquement, que devrait être cette variance ? Énoncer le théorème qui justifie votre réponse.

(c) Lorsque la taille de l'échantillon est grande, la distribution de la variable $\overline{y}$ est à peu près symétrique et d'allure assez proche d'une distribution normale. Est-ce le cas lorsque $n = 5$ ? (Faire un graphique).

(d) Si la distribution des $\overline{y}$ est approximativement normale, la probabilité que $\overline{y}$ s'éloigne de plus de 1,96 écarts-types de la moyenne est de 5%. Quelle est cette

5. Le point de Paris est l'unité de mesure utilisée dans l'industrie de la chaussure. Il a été établi vers 1800 en France. Un point de Paris correspond à 0,667 centimètre. On calcule la pointure en divisant la longueur intérieure de la chaussure (en centimètre) par le point de Paris. À la pointure 40,5 correspond donc à peu près à 27 cm de long.
Source : `https://fr.wikipedia.org/wiki/Pointures_et_tailles_en_habillement`

probabilité, en réalité ? (Il s'agit d'estimer cette probabilité — plus vous prélèverez d'échantillons plus votre estimation sera juste).

(e) L'intervalle de confiance calculé à partir de la formule $\overline{y} \pm 1,96s/\sqrt{n}$ est appelé *« intervalle de confiance à 95% »* car la probabilité de recouvrement est présumée être de 95%. Quelle est, en fait, la probabilité de recouvrement (c'est-à-dire, la probabilité que l'intervalle de confiance recouvre la vraie moyenne) ?

(f) Est-ce que vos simulations semblent indiquer que l'estimateur $s^2$ est sans biais pour $S^2$ ?

(g) La théorie statistique indique que dans certaines conditions l'intervalle de confiance calculé par la formule [6]

$$\frac{(n-1)\times s^2}{\chi^2_{\alpha/2;n-1}} < S^2 < \frac{(n-1)\times s^2}{\chi^2_{1-\alpha/2;n-1}}$$

est un intervalle de confiance à 95%. Estimer le niveau réel de cet intervalle de confiance.

## Références


Cochran, W. G. 1977, *Sampling Techniques*, 3e éd., Wiley series in probability and mathematical statistics, John Wiley Sons, New York, ISBN 047116240x, 428 p..

Dowling, T. A. et R. H. Schactman. 1975, «On the relative efficiency of randomized response models», *J. Amer. Statist. Assoc.*, vol. 70, p. 84–87, ISSN 0162-1459.

Fox, J. et P. Tracy. 1986, *Randomized Response : A Method for Sensitive Surveys*, no 58 dans Quantitative Applications in the Social Sciences, SAGE Publications, ISBN 9780803923096.

Horvitz, D. G., B. V. Shah et W. R. Simmons. 1967, «The unrelated question randomized response model», dans *Proceedings of the Social Statistics Section*, American Statistical Association, p. 65–72.

Lohr, S. L. 2019, *Sampling : Design and Analysis*, 2e éd., Advanced Series, CRC Press, Boca Raton, FL, ISBN 0367273415, 596 p..

Warner, S. L. 1965, «Randomized response : A survey technique for eliminating evasive answer bias.», *Journal of the American Statistical Association*, vol. 60, p. 63–69.


6. Dans cette formule, $\chi^2_{\beta;\nu}$ dénote la valeur qui fait que $\Pr(\chi^2_\nu > \chi^2_{\beta;\nu}) = \beta$. En R, elle se calcule par : `qchisq(1-beta,nu)`.