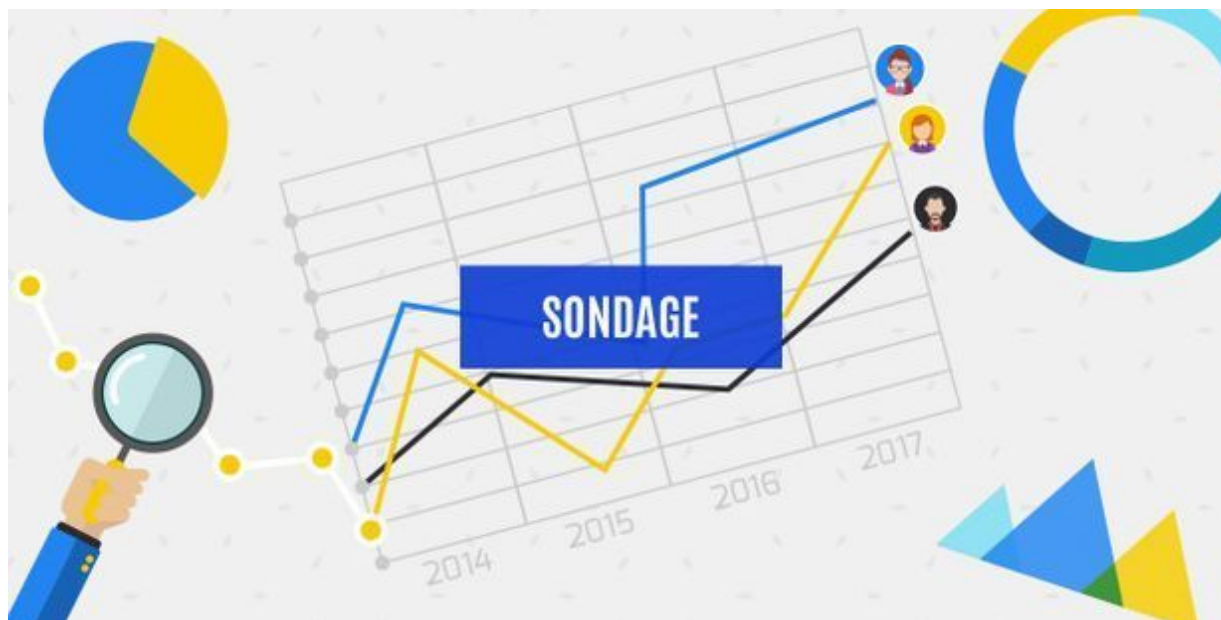


Créer un sondage

Introduction

Qu'est ce qu'un sondage ?

- Moyen simple pour **recueillir des renseignements** et **connaitre les avis** des personnes interrogées
 - Plus compliqué à mettre en place qu'il n'y parait.....



Définissez les objectifs du sondage

- Toutes les questions que vous poserez viseront à apporter une réponse à votre **objectif**
- Exemple de questions : *un patron veut connaître la satisfaction des employés*
 - *Dans quelle mesure êtes-vous satisfait dans votre travail ? Répondez en utilisant une échelle de 1 à 10*
 - *Vrai ou faux ? Je me réveille chaque matin avec l'idée que mon travail a un sens*
- Conseil : après l'élaboration du sondage, reprendre les questions une à une pour éliminer celles qui ne réalisent pas directement l'objectif de votre sondage

Comment obtenir des réponses honnêtes?

- Une **réponse sincère** est parfois difficile à avoir
- Exemple du patron :
 - *si vos employés pensent que leur sincérité aura des conséquences sur leur carrière, comme perdre leur réputation, ils ne vont pas donner de réponses sincères*
- Besoin d'ajuster le sondage pour obtenir des résultats honnêtes par exemple, en le rendant **anonyme**

Quelle méthode ?

- Enquêtes par téléphone, entretiens en tête-à-tête, sondages par courrier, sondages électroniques....
- Chaque méthode présente des **inconvénients** et des **avantages** que vous devez examiner en fonction de votre **budget**, du **personnel**...
 - *Entretiens en tête-à-tête : très chers et très longs mais résultats plus fiables*
 - *Questionnaires électroniques : sujets à distorsions mais plus économiques à administrer, pour plus de répondants*
- Conduire vos enquêtes sur des supports différents pour croiser les informations

Taille de l'échantillon

Qu'est ce qu'un échantillon?

- **Expérience** : on lance plusieurs fois un dé et on note si la face supérieure est un 4 ou un autre nombre
- La **valeur théorique** pour obtenir un 4 : $1/6$
- *La mise en défaut de cette expérience, permet d'affirmer s'il est raisonnable de penser que le dé est pipé*
- En réalisant l'expérience un certain nombre de fois (on crée un **échantillon**), on mesure la **fréquence observée** d'apparition du 4.
 - Si la fréquence observée et la valeur théorique sont trop "éloignées" (dépassent un seuil fixé) alors on considère que le dé est pipé.
 - Sinon, on considère que le dé n'est pas pipé

Qu'est ce qu'un échantillon ?

- Un **échantillon** de **taille n** est constitué des résultats de **n répétitions indépendantes** de la **même expérience** sur l'ensemble des personnes sur lesquels porte l'étude statistique (la **population**)
- Exemple: *Si, sur l'ensemble des puces produites par une entreprise en une semaine, on en prélève 200, cet ensemble de 200 puces constitue un échantillon de taille 200 de la population de toutes les cartes à puce produites en une semaine*

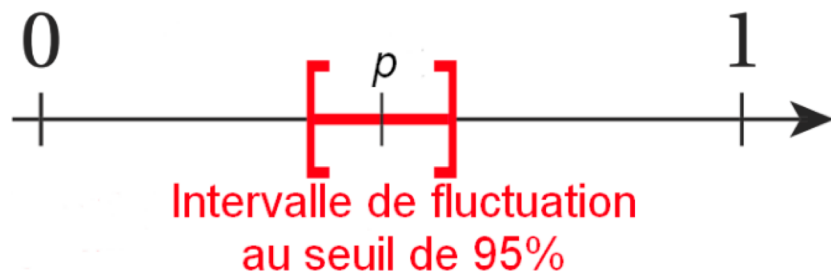


2 cas

- On connaît la proportion théorique : on calcule **l'intervalle de fluctuation**
- On connaît la fréquence observée : on calcule **l'intervalle de confiance**

Intervalle de fluctuation

- On suppose que 22% des puces sont défectueuses. La **proportion théorique** p est donc égale à 22%.
- On prélève un **échantillon de taille** 200 parmi cette production et on compte le nombre de puces défectueuses. Ce nombre est égal à 41.
- La **fréquence observée** f est égale à $41/200=0,205$.
- L'**intervalle de fluctuation** de la fréquence p des puces au seuil de 95%, est un intervalle de centre 0,22 tel que les fréquences observées se trouvent dans cet intervalle pour 95 % des échantillons de taille 200



$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$$

$$[0,15; 0,29]$$

Exercice

- Deux entreprises A et B recrutent dans un bassin d'emploi où il y a autant de femmes que d'hommes (exactement 50 % de femmes), avec la contrainte du respect de la parité.
 - Dans l'entreprise A, il y a 100 employés dont 43 femmes (soit 43%).
 - Dans l'entreprise B, il y a 2500 employés dont 1150 femmes (soit 46 %).
- 46 % est plus proche de 50% que 43. Si on admet la parité, il est vrai que B est plus proche que A.
- *Peut-on alors affirmer que l'entreprise B respecte mieux la parité que l'entreprise A ?*

Correction

- *Est-ce que la fréquence observée f se situe dans l'intervalle de fluctuation au seuil de 95%.*
- La proportion théorique p est égale à 0,5 (50% de femmes).
- Pour l'entreprise A:
 - La taille de l'échantillon n est égale à 100.
 - La fréquence observée f est égale à 0,43.
 - L'intervalle de fluctuation au seuil de 95% de $p=0,5$ est $IF_A=[0,4;0,6]$ donc $f \in IF_A$
- Pour l'entreprise B:
 - La taille de l'échantillon n est égale à 2500.
 - La fréquence observée f est égale à 0,46.
 - L'intervalle de fluctuation au seuil de 95% de $p=0,5$ est $IF_B=[0,48;0,52]$ donc $f \notin IF_B$
- Conclusion : La proportion de 46% pour l'entreprise B s'observe dans moins de 5% des échantillons de taille 2500. On peut rejeter l'hypothèse que l'entreprise B respecte la parité. Par contre, pour l'entreprise A, on peut accepter cette hypothèse.

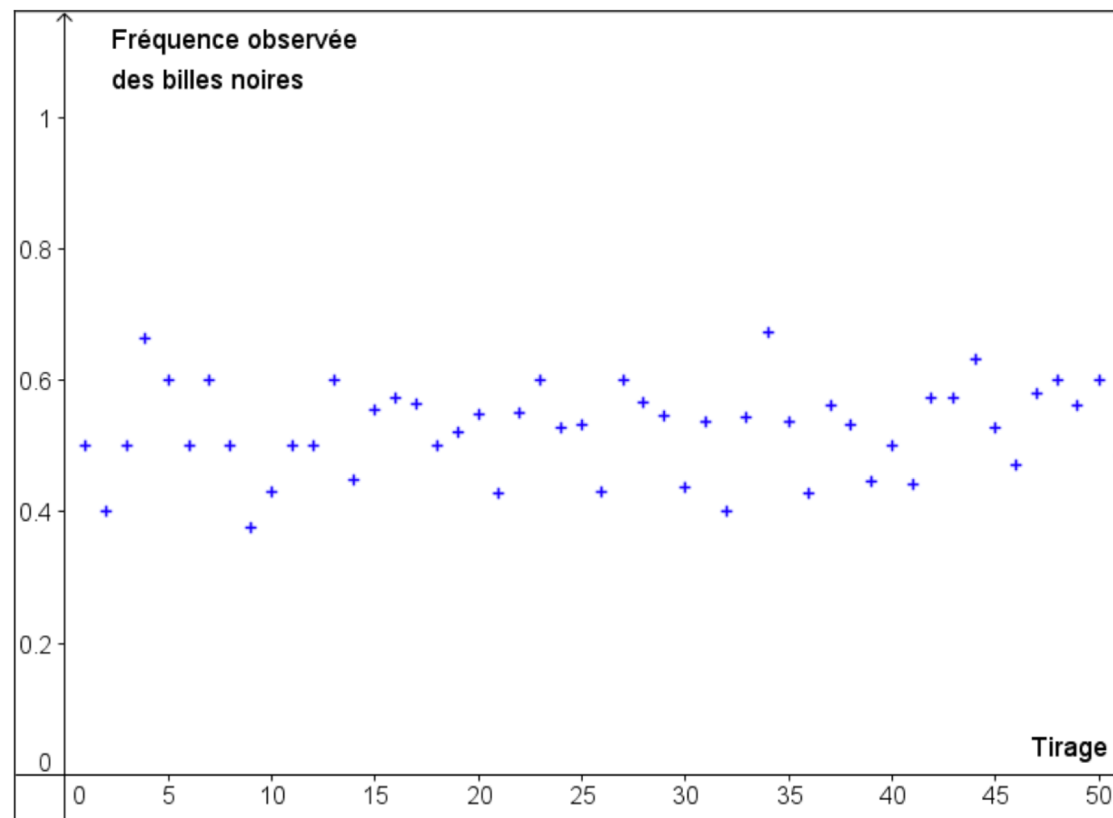
Intervalle de confiance

- Un jeu consiste à tirer 100 billes d'un sac contenant 300 billes noires et 300 billes blanches
- L'expérience peut être simulée avec un tableur afin d'effectuer rapidement un grand nombre de tirages
- Pour cet échantillon de taille 100, on compte le nombre de billes noires et on calcule la fréquence observée f
- On pourrait ainsi vérifier que, dans 95 % des cas, la fréquence des billes noires dans l'échantillon appartient à l'intervalle:

$$\left[0,5 - \frac{1}{\sqrt{100}} ; 0,5 + \frac{1}{\sqrt{100}} \right] \text{ soit : } [0,4 ; 0,6] \text{ où } p = 0,5 \text{ et } n = 100.$$

Intervalle de confiance

- Nuage de points des fréquences observées des billes noires pour 50 tirages



Intervalle de confiance

- Soit p la proportion théorique tel que $0,2 < p < 0,8$.
- On considère la fréquence observée f pour un échantillon donné de taille $n > 25$.

- L'intervalle de confiance de p au niveau 0,95

$$I_C = \left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}} \right]$$

- 95 % des intervalles de confiance associés aux échantillons de taille n possibles ayant comme fréquence observée f contiennent la proportion théorique p

Exercice

1. Avant les élections, le candidat A commande un sondage effectué sur 250 personnes. 138 personnes interrogées déclarent avoir l'intention de voter pour le candidat A. *Le candidat A peut-il espérer être élu ?*
2. Le candidat A commande un second sondage effectué sur 1000 personnes pour lequel 538 personnes déclarent avoir l'intention de voter pour lui. *Le candidat A peut-il espérer être élu?*

Correction

- 1) Soit p la proportion théorique d'électeurs pour le candidat A.

- La fréquence observée est égale à $f=138/250=0,552$

- L'intervalle de confiance de p au seuil de 0,95 est: $I_c = \left[0,552 - \frac{1}{\sqrt{250}}; 0,552 + \frac{1}{\sqrt{250}} \right]$

de façon approchée $[0,49; 0,62]$

- Or $0,49 < p < 0,62$.

- Il est donc possible que le candidat A ne soit pas élu

car p peut être inférieur à 0,5

Correction

- 2) Soit p la proportion théorique d'électeurs pour le candidat A.

- La fréquence observée est égale à $f=538/1000=0,538$

- L'intervalle de confiance de p au seuil de 0,95 est: $I_C = \left[0,538 - \frac{1}{\sqrt{1000}}; 0,538 + \frac{1}{\sqrt{1000}} \right]$

de façon approchée $[0,51; 0,57]$

- Or $0,51 < p < 0,57$.

- p est forcément plus grand que 0,5

donc le candidat A est sûr d'être élu

Quelle taille d'échantillon?

- **Marge d'erreur :**

- Pourcentage indiquant dans quelle mesure les résultats de votre sondage sont susceptibles de refléter l'opinion de la population globale
- Plus la marge d'erreur est faible, plus vous êtes susceptible d'obtenir une réponse exacte pour un niveau de donné

- **Niveau de confiance :**

- Pourcentage indiquant le degré de certitude avec lequel la population va choisir une réponse située entre deux valeurs données
- Par exemple, un niveau de confiance de 95 % signifie que vous pouvez être certain(e) à 95 % que les résultats se situent dans l'intervalle de confiance

Quelle taille d'échantillon?

- La taille d'échantillon minimale pour l'obtention de résultats significatifs pour un événement et un niveau de risque fixé :

$$\text{Taille de l'échantillon} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)}$$

N = taille de la population

e = marge d'erreur (pourcentage sous forme décimale)

z = cote z

p = valeur en pourcentage

Exemple, si vous avez trouvé que, lors d'un sondage précédent, 75 % des clients ont déclaré être satisfaits de votre produit et que vous envisagez de réaliser une nouvelle fois ce sondage, vous pouvez utiliser p = 0,75 pour calculer la taille de l'échantillon

Quelle taille d'échantillon?

- La taille d'échantillon minimale pour l'obtention de résultats significatifs pour un événement et un niveau de risque fixé :

$$\text{Taille de l'échantillon} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)}$$

N = taille de la population

e = marge d'erreur (pourcentage sous forme décimale)

z = cote z

p = valeur en pourcentage

- La cote z correspond au nombre d'écart types d'une proportion donnée par rapport à la moyenne.

Niveau de confiance souhaité	cote z
80 %	1,28
85 %	1,44
90 %	1,65
95 %	1,96
99 %	2,58

Quelle taille d'échantillon?

Taille de la population	Taille d'échantillon par marge d'erreur		
	3 %	5 %	10 %
500	345	220	80
1 000	525	285	90
3 000	810	350	100
5 000	910	370	100
10 000	1 000	385	100
Plus de 100 000	1 100	400	100

Calculateur en ligne : <https://fr.surveymonkey.com/mp/sample-size-calculator/>

Type de questions

Choisir le type des questions

- Pour recueillir des informations :
 - sur une idée connue, posez des **questions fermées**
 - *Quelle est votre activité favorite en ligne ?*
 - (a) *Le réseautage social*
 - (b) *Les forums*
 - (c) *Le shopping*
 - sur une nouvelle idée, posez des **questions ouvertes**
 - *Décrivez votre première visite à un Apple Store*
Réponse :
 - Compromis : posez **questions semi-fermées**
 - *Quelle est votre activité favorite en ligne ?*
 - (a) *La messagerie instantanée*
 - (b) *Le réseautage social*
 - (c) *Les forums*
 - (d) *Autres*

Choisir le type des questions

- Echelle de Likert
 - Question à **échelle de satisfaction**
 - Question à **échelle de notation**



A votre entrée dans la boutique, avez-vous été satisfait(e) de l'accueil ? *

Non, pas du tout

Oui, tout à fait

0 1 2 3 4 5 6 7 8 9 10

Ne se prononce pas

Problèmes liés à l'échelle de Likert

- **Position du milieu de l'échelle**
- Exemple
 - pas du tout d'accord
 - **un peu d'accord**
 - moyennement d'accord
 - tout à fait
- La position de "un peu d'accord" vers la gauche de l'échelle, vers "pas du tout...", constitue un **indice fortuit** en contradiction avec le sens du libellé
- Etre un peu d'accord n'est pas assimilable à un désaccord, même léger
- Il devient alors difficile de prédire la réponse, selon que le répondant se fie à la position ou au texte, ce qui diminue la fiabilité de l'échelle
- Autres indices fortuits : modification de la police ou de la taille des caractères, variations de couleur...

Problèmes liés à l'échelle de Likert

- La fiabilité d'échelles comportant un nombre variable de points :
 - 3 points (contre / neutre / pour),
 - 5 points (désaccord total / désaccord / neutre / accord / accord total),
 - 7 points (désaccord total / désaccord / léger désaccord / neutre / léger accord / accord / accord total)
 - 11 points (note de 0 à 10).
- Recommandation : échelles en 7 points, sans ajout d'une valeur "Ne sait pas"

Outils

Conduire un sondage



A vous de jouer....