

LES SONDAGES... ET LEURS DÉFAUTS

QUI NE SONT PAS CEUX QU'ON CROIT!

Antoine ROLLAND

Lyon - 30 avril 2018

QUI SUIS-JE ?

Antoine ROLLAND

- ▶ Etudes : Bac S, prépa, école d'ingénieur, agrégation de mathématique, thèse en informatique théorique
- ▶ Poste : Maître de conférence en statistique à l'IUT Lumière Lyon II DUT Statistique et Informatique Décisionnelle (STID)
- ▶ Concrètement :
 - ▶ 1/3 de temps enseignant (bac+1/+2/+3) en statistique dans la formation STID à l'IUT
 - ▶ 1/3 de temps de responsable pédagogique STID
 - ▶ 1/3 de temps recherche en statistique
- ▶ Thème de recherche : le choix en présence de plusieurs critères / votants ; l'analyse de classements

<http://eric.univ-lyon2.fr/~arolland/>

LA RECHERCHE EN MATH

- ▶ Réfléchir
- ▶ Ecrire des articles corrigés par les pairs / Aller parler devant d'autres chercheurs en conférence
Exemple : 10ème colloque francophone sur les sondages à Lyon en octobre 2018
- ▶ Vulgariser nos recherches (vers le grand public pour les SHS, vers la technologie pour les sciences)
- ▶ Et pour les maths/stat : “inventer” ou “découvrir” de nouveaux objets mathématiques et essayer de comprendre leurs propriétés

QUELQUES EFFETS DONT IL FAUT SE MÉFIER QUAND ON FAIT DES SONDAGES...

- ▶ constitution de l'échantillon
- ▶ sincérité des votants
- ▶ effet de halo
- ▶ biais de confirmation
- ▶ présentation des données
- ▶ corrélation et causalité
- ▶ confusion

(Merci à JB Aubin et Thibaut Espinasse)

PRINCIPE DU SONDAGE

- ▶ on ne peut pas interroger toute la population
- ▶ alors on interroge une petite partie **prise au hasard**
- ▶ la plupart du temps l'échantillon recueilli ressemble assez à la population : on peut obtenir une estimation de la variable recherchée

Exemples :

- ▶ contrôle qualité par échantillonnage (industrie)
- ▶ étude de marché (marketing)
- ▶ recensement de la population et enquêtes emploi, santé, handicap...(INSEE)
- ▶ opinion politique

CONSTITUTION DE L'ÉCHANTILLON

- ▶ la précision d'un sondage par échantillon ne dépend que de la (racine carrée de la) **taille** de l'échantillon, et pas de la taille de la population !
ordre de grandeur : 1000 individus $\leftrightarrow \pm 3\%$

CONSTITUTION DE L'ÉCHANTILLON

- ▶ la précision d'un sondage par échantillon ne dépend que de la (racine carrée de la) **taille** de l'échantillon, et pas de la taille de la population !
ordre de grandeur : 1000 individus $\leftrightarrow \pm 3\%$
- ▶ il faut un tirage **vraiment aléatoire** sur toute la population
 - ▶ sondage par téléphone ?
 - ▶ sondage sur twitter ?

CONSTITUTION DE L'ÉCHANTILLON

- ▶ la précision d'un sondage par échantillon ne dépend que de la (racine carrée de la) **taille** de l'échantillon, et pas de la taille de la population !
ordre de grandeur : 1000 individus $\leftrightarrow \pm 3\%$
- ▶ il faut un tirage **vraiment aléatoire** sur toute la population
 - ▶ sondage par téléphone ?
 - ▶ sondage sur twitter ?
- ▶ on peut redresser (ou procéder par quota) si on connaît exactement **toutes** les variables qui influent sur la réponse
 - ▶ sexe ? domicile ? métier ?

SINCÉRITÉ DES VOTANTS

Qui fume régulièrement du cannabis avant de venir au lycée ?

SINCÉRITÉ DES VOTANTS

Qui fume régulièrement du cannabis avant de venir au lycée ?

Questions délicates : violences conjugales, comportements amoraux, actes illégaux....

EFFET DE HALO

QUESTION

EFFET DE HALO

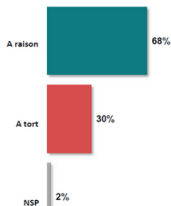
QUESTION

Pensez-vous que la morale d'un homme peut être jugée à la façon dont il traite les animaux ?



L'opinion sur l'abandon du projet d'aéroport de Notre-Dame-des-Landes

Q. Diriez-vous que le gouvernement a raison ou a tort d'abandonner le projet d'aéroport de Notre-Dame-des-Landes ?



Vote au 1 ^{er} Tour – Présidentielle 2017					Vote au 2 nd Tour	
J.-L. Mélenchon	B. Hamon	E. Macron	F. Fillon	M. Le Pen	E. Macron	M. Le Pen

72%	84%	76%	59%	63%	72%	63%
28%	15%	24%	38%	35%	26%	35%

1.7k



SONDAGE EXCLUSIF/ INFOGRAPHIES - Selon notre enquête Odoxa et Dentsu Consulting pour *Le Figaro* et France Info, les trois quarts des Français soutiennent la décision du gouvernement d'abandonner le projet d'aéroport et de reporter au printemps l'évacuation de la ZAD.

C'est un large plébiscite. La décision de l'exécutif d'abandonner la construction de l'aéroport de Notre-Dame-des-Landes et de reporter au printemps l'évacuation de la ZAD recueille un large soutien dans l'opinion. Selon notre sondage Odoxa pour *Le Figaro* et France info*, réalisé mercredi et jeudi matin, les trois quarts (76%) des Français saluent le choix annoncé par le premier ministre Édouard Philippe mercredi. Une majorité de sympathisants de droite (63%) l'approuvent aussi. L'enquête révèle que 74 % des Français soutiennent l'abandon du projet et 72% le fait de différer l'évacuation de la ZAD.

EDOUARD PHILIPPE A ANNONCÉ QU'IL RENONÇAIT FINALEMENT À LA CONSTRUCTION DE L'AÉROPORT DE NOTRE DAME DES LANDES, ET QU'IL LAISSAIT JUSQU'AU PRINTEMPS AUX OCCUPANTS ILLÉGAUX POUR ÉVACUER LA ZONE. VOUS PERSONNELLEMENT, ESTIMEZ-VOUS QUE C'EST UNE BONNE OU UNE MAUVAISE DÉCISION ?

Très mauvaise décision

1 % NSP



BIAIS DE CONFIRMATION



PRÉSENTATION DES DONNÉES TENDANCIEUSES...

La participation des voyageurs au coût de fonctionnement des transports en Île-de-France

Le coût de fonctionnement des trains, métros, trams et bus en Île-de-France chaque année est d'environ

10 milliards €*
* estimation 2016

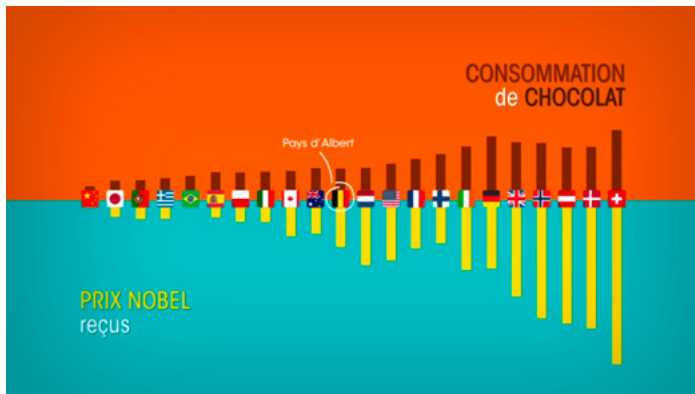
dont
72%
sont financés
par les employeurs,
les collectivités
et d'autres recettes

et
28%

sont financés par
les voyageurs via l'achat
de forfaits *À Navigo* et tickets



CORRÉLATION ET CAUSALITÉ



Quelques sources : Chaine Youtube "La statistique expliquée à mon chat"

Site web "spurious correlations"

CORRÉLATION ET CAUSALITÉ

Exemple (au lycée) : la longueur moyenne des cheveux est inversement proportionnelle à la taille.

EFFET DE CONFUSION / PARADOXE DE SIMPSON

Exemple

Lycée A

	succès	échec
Section ES	90	10
Section S	240	160

Lycée A

	succès	échec
Section ES	90%	10%
Section S	60%	40%

Lycée B

	succès	échec
Section ES	240	60
Section S	50	50

Lycée B

	succès	échec
Section ES	80%	20%
Section S	50%	50%

EFFET DE CONFUSION / PARADOXE DE SIMPSON

Exemple

Lycée A			Lycée B		
	succès	échec		succès	échec
Section 1	90	10	Section 1	240	60
Section 2	240	160	Section 2	50	50
Total	330	170	Total	290	110

Lycée A			Lycée B		
	succès	échec		succès	échec
Section 1	90%	10%	Section 1	80%	20%
Section 2	60%	40%	Section 2	50%	50%
Total	66%	34%	Total	72,5%	26,5%

Les élèves de la section 1 réussissent mieux au lycée A ; les élèves de la section 2 réussissent mieux au lycée A ; Mais le lycée B a de meilleurs résultats globaux !

EFFET DE CONFUSION / PARADOXE DE SIMPSON

