Jugement critique et statistique
$\rightarrow$
La statistique au service de l'émission de jugements critiques

Gilles Stoltz

Université Paris Sud — CNRS — HEC Paris

Quelques exemples d'intervalles de confiance

$\rightarrow$ Les chiffres ne parlent pas d'eux-mêmes !

TwitterAudit

- étudie 5,000 followers
- objectif = proportion totale de vrais comptes

Capture effectuée le 4 octobre 2017 à 9h 20

Followers : 39,915,316 au total

Score : 59.8% en réalité (mais arrondi)

Nombres 23,869,359 et 16,045,957 trop beaux pour être vrais...

TwitterAudit

« Score » de 59.8% : proportion d'échantillon
(proportion de vrais comptes parmi les 5,000 followers étudiés)

Pas directement reportable sur la population de 39,915,316 followers !

Intervalle de confiance : $59.8\% \pm 1.4\%$

Nombre total de vrais followers : 23.3M – 24.5M

Morale : Détailler la généralisation échantillon $\rightarrow$ population

Mes étudiants ont écrit 23,310,544 – 24,428,174
et ont été outrés que je ne leur accorde pas tous les points...

Les journalistes et les sondages...

Les instituts fournissent la marge d'incertitude des sondages :

$\pm 3\%$ si environ 1,000 sondés

Signifie que la proportion de population est égale à celle d'échantillon à $\pm 3\%$ près, avec grande confiance

Les journalistes français ne la comprennent pas, donc ne la reprennent pas et disent à la place :

« Nous vous rappelons que les sondages ne sont qu'un instantané de l'opinion et ne préjugent pas de l'élection à venir »

Aux Etats-Unis (la peur des procès ?) ils indiquent ces marges.

Les journalistes et les sondages... en France

Méthodologie

- Étude réalisée auprès d'un **échantillon de 801 personnes**, représentatif de la **population d'Alsace, âgées de 18 ans et plus et inscrites sur les listes électorales.**
- L'échantillon a été constitué selon la méthode des quotas, au regard des critères de sexe, d'âge, de catégorie socioprofessionnelle, de catégorie d'agglomération et de département de résidence.
- **Mode d'interrogation:** L'échantillon a été interrogé par téléphone au domicile des personnes.
- **Dates de terrain:** les interviews ont été réalisées **les 16 et 17 mars 2010**.
- OpinionWay rappelle par ailleurs que les résultats de ce sondage doivent être lus en tenant compte des marges d'incertitude : 3 à 4 points au plus pour un échantillon de 800 répondants.
- La notice de cette enquête est consultable à la commission des sondages.

Le Figaro-LCI - Baromètre OpinionWay - Fiducial des élections régionales – Région Alsace / 18 Mars 2010 page 3

Pour le second tour des élections régionales en Alsace, mars 2010

Les journalistes et les sondages... en France

minimare

- **Noté**, ce titre du « Monde » (22/8) : « Niches fiscales : M. Sarkozy annonce 10 milliards d'euros de hausses d'impôts ». Même avec les niches, il faut appeler un chat un chat.
- **Ce titre**, dans « Les Echos » (20/8) : « Voté il y a trois ans, le paquet fiscal est resté un talon d'Achille pour la majorité ». Un talon menacé de rupture.
- **Le « JDD »** (22/8) consacre un quart de page à expliquer la hausse de 2 points de la cote de Sarko, selon l'Ifop. Manque de bol, un sondage Viavoice pour « Libé » (23/8), réalisé aux mêmes dates, le donne en baisse de 1 point. Il en faudrait, des tiers de page, pour expliquer les contradictions entre sondeurs !
- **Cours de fiscalité** donné par le secrétaire général de l'UMP, Xavier Bertrand (« Le Parisien », 21/8) :
- **Ségolène Royal** rope 1, 21/8) les Sarko sur la sécuri fie d'« agitation fré cace ». Seulement
- **La même Ségolè** Figaro », 18/8) : « J sur moi et mon pè les hommes, donc voir, donc moi fille forcément castratr que « je » parle de
- **De Vincent Peillo** péen PS (« Le Pa « On a laissé se cor velle oligarchie, mé fonctionnaires pass nets ministériels, lié litique, faisant des f dans le privé. » Pe parler aux anciens cabinets Fabius ou aujourd'hui salariés

mbe !

Source : Canard enchaîné, août 2010 : Pan sur le bec ! (variations non significatives)

Les journalistes et les sondages... aux Etats-Unis

Hillary Clinton Holds Double-Digit Lead Over Donald Trump, Poll Finds

Melissa Chan @melissalchan | June 26, 2016

Clinton leads Trump by 51% to 39% in a new national poll

Democratic presidential candidate Hillary Clinton holds a double-digit lead over presumptive GOP nominee Donald Trump, according to a new national poll released Sunday.

The former Secretary of State leads Trump by 51% to 39% among registered voters nationwide, a new Washington *Post*-ABC News poll shows. The poll also found that 56% of the public at large say Trump stands against their beliefs and 64% say he does not have the necessary credentials to be president. Fifty-six percent feel strongly that he is unqualified.

John Locher—AP

The poll, which randomly sampled 1,001 adults between June 20 and June 23, also found that half of Americans are anxious about Clinton possibly leading the country should she win in November. (The margin of sampling error for overall results is plus or minus 3.5 percentage points.)

Source : Time.com, 26 juin 2016

Conclusions sur les intervalles de confiance

Les chiffres ne parlent pas d'eux-mêmes, il faut les faire parler !

→ Expliquer quelles généralisations sont possibles
de l'échantillon à la population

P.ex. procurer des marges d'incertitudes (et arrondir ses nombres) :

> Dans toute statistique, l'inexactitude du nombre est
> compensée par la précision des décimales.
> Georges Elgozy (économiste français, 1909–1989)

Tests d'hypothèses

Principe et premiers exemples

Objectif

Mettre à l'épreuve une assertion / en prouver une autre

The Lady tasting tea (Fisher, 1935)

4 tasses lait puis thé
4 tasses thé puis lait
Présentées simultanément

Hypothèses

- H_0 [raisonnable] : Ne peut distinguer le goût
- H_1 [à prouver] : Distingue le goût

Sous H_0, reconnaissance aléatoire
Probabilité $1/70 \approx 1.4\%$ de bien classer toutes les tasses

Si c'est le cas, on rejette H_0 pour H_1 ; sinon, on conserve H_0

Objectif

Mettre à l'épreuve une assertion / en prouver une autre

Hypothèses

- H_0 [raisonnable] : Opinion commune ou vérité actuelle
- H_1 [à prouver] : Une vérité nouvelle

On considère une certaine quantité (« statistique de test »)

- Loi connue sous H_0
- Comportement différent sous H_1

Si valeur obtenue sur les données est peu probable sous H_0 :
Rejet de H_0 (i.e., H_0 est peu plausible) $\rightarrow$ progrès négatif

Sinon :
Conservation de H_0 (i.e., H_0 n'est pas impossible) $\rightarrow$ zone grise

Drague tactile

Source : Nicolas Guéguen, Social influence, 2(2):81–97, 2017

3 JH touchant ou non des JF avant demande de numéro
120 JF non touchées : 12 num. / 120 JF touchées : 23 num.

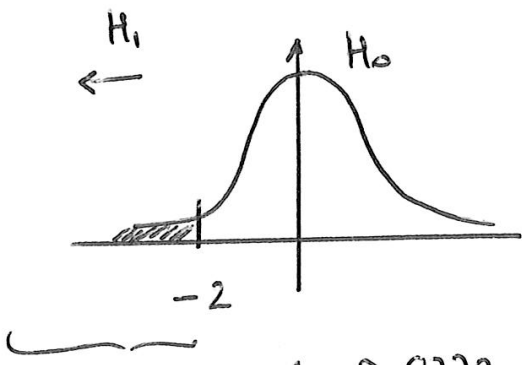
Hypothèses (cf. toucher très léger, non ressenti)

- H_0 [raisonnable] : Pas d'impact du toucher
- H_1 [à prouver] : Plus grand succès avec toucher

Statistique de test :

$$S = \frac{\overline{x}_{120} - \overline{y}_{120}}{\sqrt{(2/120)\,\overline{xy}_{240}(1 - \overline{xy}_{240})}}$$

Sous H_0 : loi normale standard
Sous H_1 : grandes valeurs positives
Valeur de S sur les données : ≈ -2.0



P-value = 1 - 0.9772
≃ 2.3% < 5%

Drague tactile

Source : Nicolas Guéguen, Social influence, 2(2)81–97, 2017

3 JH touchant ou non des JF avant demande de numéro
120 JF non touchées : 12 num. / 120 JF touchées : 23 num.

Hypothèses (cf. toucher très léger, non ressenti)

- H_0 [raisonnable] : Pas d'impact du toucher
- H_1 [à prouver] : Plus grand succès avec toucher

Question : le taux de $23/120 \approx 19.2\%$ est-il significativement (H_1) supérieur au taux de $12/120 = 10\%$, ou non (H_0) ?

I.e., la différence entre les deux proportions pourrait-elle être uniquement imputée au hasard (H_0) ou non (H_1) ?

Réponse :
L'indice de crédibilité (P–valeur) de H_0 est 2.3%, on rejette H_0

Tests d'hypothèses

Tests du χ^2 d'ajustement
$\rightarrow$ pour détecter des manipulations

Elections présidentielles en Iran, 2009

The Washington Post

The Devil Is in the Digits

By Bernd Beber and Alexandra Scacco
Saturday, June 20, 2009 12:02 AM

Since the declaration of Mahmoud Ahmadinejad's landslide victory in Iran's presidential election, accusations of fraud have swelled. Against expectations from pollsters and pundits alike, Ahmadinejad did surprisingly well in urban areas, including Tehran -- where he is thought to be highly unpopular -- and even Tabriz, the capital city of opposition candidate Mir Hussein Mousavi's native East Azarbaijan province.

Others have pointed to the surprisingly poor performance of Mehdi Karroubi, another reform candidate, and particularly in his home province of Lorestan, where conservative candidates fared poorly in 2005, but where Ahmadinejad allegedly captured 71 percent of the vote. [...]

Elections présidentielles en Iran, 2009

29 provinces
4 candidats
$\rightarrow$ 116 nombres

Chiffres des unités :
Aucune information

Donc uniformément distribués (H_0)...

Sauf si inventés (H_1)

Case	Province	Candidat	Voix	Chiffre
20	Kurdistan	Ahmadinejad	315689	9
21	Markazi	Ahmadinejad	572988	8
22	Mazandaran	Ahmadinejad	1289257	7
23	Qazvin	Ahmadinejad	498061	1
24	Qom	Ahmadinejad	422457	7
25	Semnan	Ahmadinejad	295177	7
26	Sistan and Baluchistan	Ahmadinejad	450269	9
27	Tehran	Ahmadinejad	3371523	3
28	Yazd	Ahmadinejad	337178	8
29	Zanjan	Ahmadinejad	444480	0
30	Ardabil	Karroubi	2319	9
31	Azarbaijan, East	Karroubi	7246	6
32	Azarbaijan, West	Karroubi	21609	9
33	Bushehr	Karroubi	3563	3
34	Chahar Mahaal and Bakhtiari	Karroubi	4127	7
35	Fars	Karroubi	16277	7
36	Gilan	Karroubi	7183	3
37	Golestan	Karroubi	10097	7
38	Hamadan	Karroubi	12032	2
39	Hormozgan	Karroubi	5126	6
40	Ilam	Karroubi	7471	1

Affichage des données | Affichage des variables

Elections présidentielles en Iran, 2009

H_0 [raisonnable] : Derniers chiffres uniformément distribués
H_1 [à prouver] : Derniers chiffres non uniformément distribués

$H_0 \leftrightarrow$ données authentiques / $H_1 \leftrightarrow$ données manipulées

Chiffre

	Effectif observé	Effectif théorique	Résidu
0	9	11,6	-2,6
1	11	11,6	-,6
2	8	11,6	-3,6
3	9	11,6	-2,6
4	10	11,6	-1,6
5	5	11,6	-6,6
6	14	11,6	2,4
7	20	11,6	8,4
8	17	11,6	5,4
9	13	11,6	1,4
Total	116		

Test

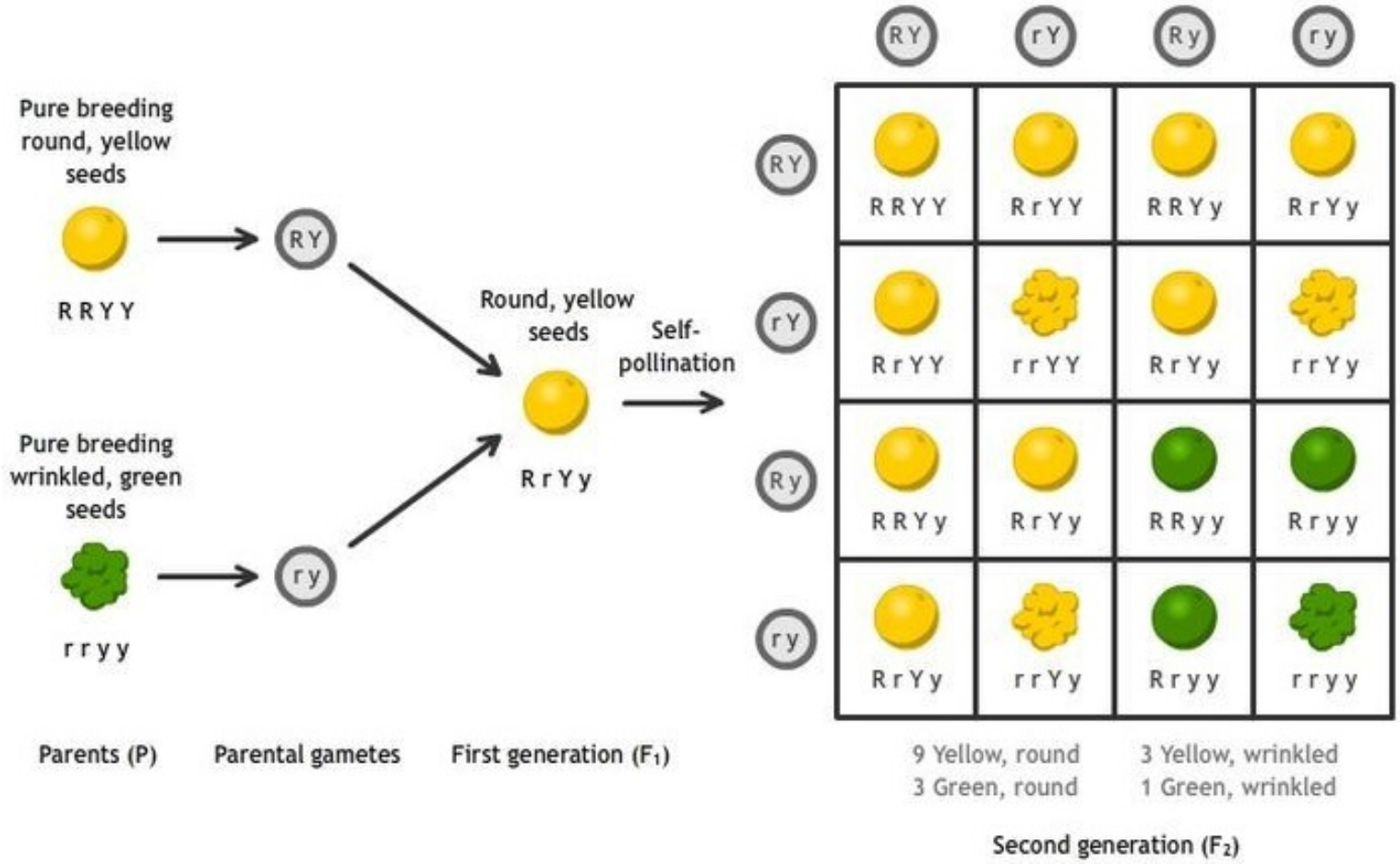
	Chiffre
Khi-deux	15,552
ddl	9
Signification asymptotique	,077

P–valeur : 7.7%
(indice de crédibilité de H_0)

Conservation de H_0 :
On ne peut pas exclure
l'authenticité de ces données

Théorie génétique de Mendel mais données revisitées par Fisher !

Voici une des nombreuses expériences de croisement de Mendel



→ Répartitions attendues 9/16, 3/16, 3/16 et 1/16

Théorie génétique de Mendel mais données revisitées par Fisher !

Hypothèses

- H_0 [conformité] :
  Phénotypes répartis selon la loi (9/16, 3/16, 3/16, 1/16)
- H_1 [non-conformité] : Phénotypes répartis selon une autre loi

$H_0 \leftrightarrow$ hypothèse de hasard génétique
$H_1 \leftrightarrow$ les voies du Seigneur sont impénétrables

Données recueillies après culture dans le jardin du monastère

Phénotype

	Effectif observé	Effectif théorique	Résidu
Lisses et jaunes	315	312,8	2,3
Lisses et verts	108	104,3	3,8
Ridés et jaunes	101	104,3	-3,3
Ridés et verts	32	34,8	-2,8
Total	556		

Test

	Phénotype
Khi-deux	,470
ddl	3
Signification asymptotique	,925

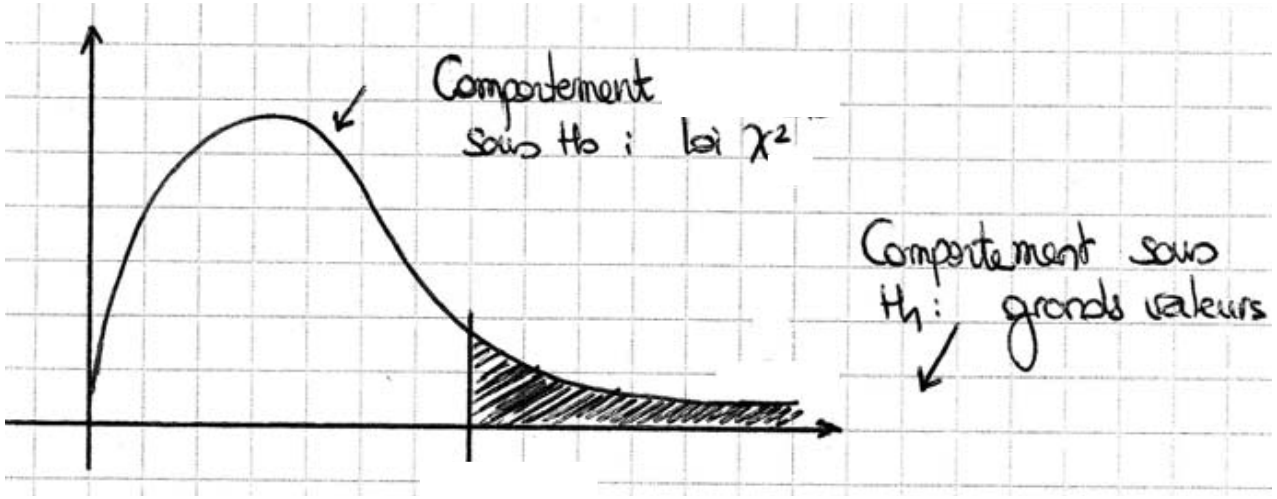
Excellente conformité : P–valeur de 92.5%

Théorie génétique de Mendel mais données revisitées par Fisher !

Les données de Mendel sont trop belles pour être vraies !

Car les P–valeurs des données de Mendel (calculées rétrospectivement) le sont : systématiquement autour de 90%.

Elles auraient dû être uniformément distribuées entre 0% et 100% :



→ Multiplication artificielle de la taille d'échantillon

Ex : 56 plants → 556 plants / mais c'était de bonne guerre !

Tricheries comptables et loi de Benford

Loi des premiers chiffres significatifs qui nous entourent (et en particulier, des nombres issus de l'économie)

Années 90, à NYC, application par le Parquet financier :
Détection dans les comptes de grandes entreprises

Tomasz Michalski et moi-même :
Détection dans les balances des paiements des Etats, dans certaines conditions macro-économiques

Références :

The Review of Economics and Statistics, 95(2):591-616, 2013

http://www.breves-de-maths.fr/quels-nombres-lhomme-produit-il/

Evocations d'autres sujets

Une méthodologie statistique avec des limites claires

Association $\neq$ causalité

Ex : Conso. de chocolat et Prix Nobel (liaison linéaire significative)

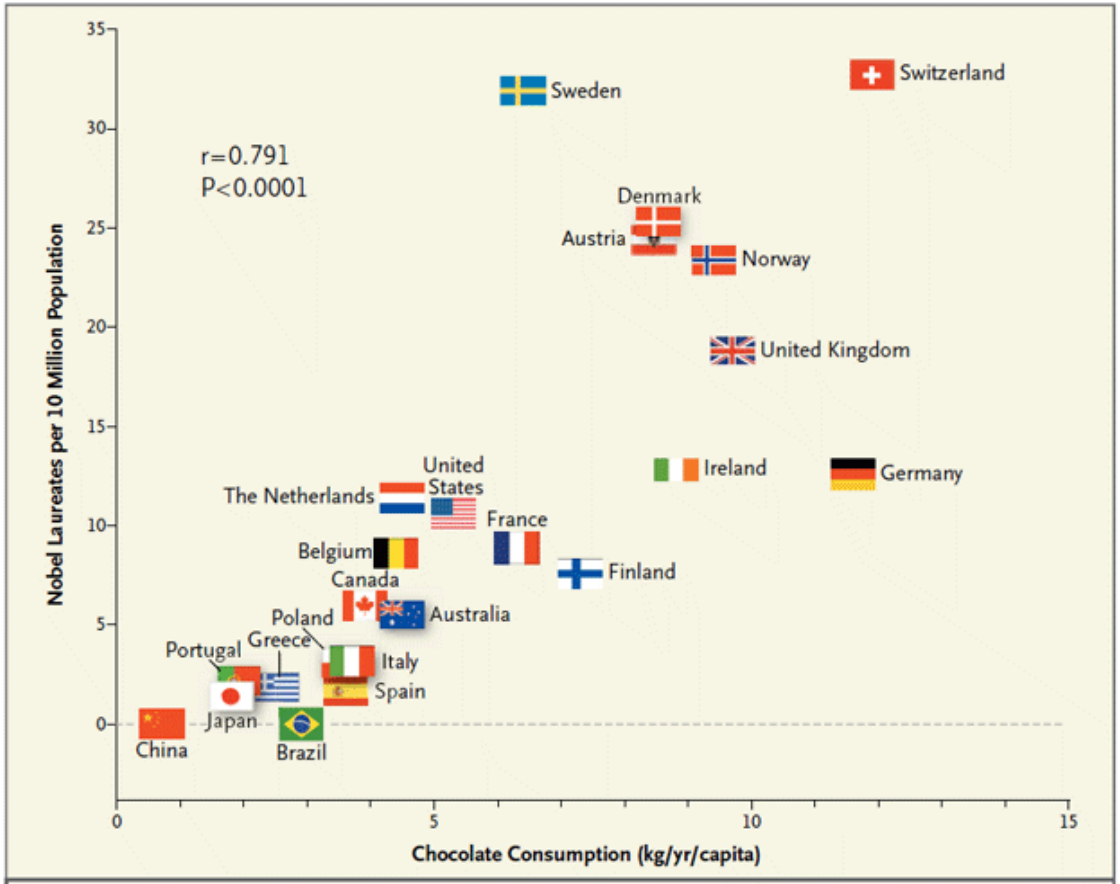


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NEW ENGLAND JOURNAL OF MEDICINE

Source : Franz Messerli, The New England Journal of Medicine, 2012

Regarde-t-on la bonne population ?

Consultations citoyennes volontaires
Biais de motivation → sous-populations

- Statut de La Poste (extrême-gauche, 2009)
- Indépendance de la Catalogne (automne 2017)
- Mariage gay en Australie (automne 2017)

Débats sur le pouvoir d'achat (traitement) des fonctionnaires, 2008

- Syndicats : point d'indice
- Gouvernement : évolutions individuelles sur quelques années

Ensemble des fonctionnaires en 2008 vs.
Fonctionnaires en poste sur toute la durée considérée

Conclusions et réflexions

Méthodologie statistique précise

Les chiffres ne parlent pas d'eux-mêmes, il faut les faire parler !

Intervalles de confiance :

- discuter la précision de nos données d'échantillon
- en termes de généralisation à la population

Tests d'hypothèses :
Permettent de voir si un fait est établi ou non à partir des données

- Oui, il l'est (avec grande confiance / faible risque d'erreur)
- On ne peut pas exclure ni qu'il le soit, ni qu'il ne le soit pas

Cf. H_0 est tenue pour vraie jusqu'à preuve contraire

La P-valeur : indique la force de la conclusion (rejet H_0) ou de l'absence de conclusion possible (H_0 conservée)

Missions du statisticien

La statistique est un bikini / un shorty.
Ce qu'elle révèle est suggestif, ce qu'elle cache est vital.
Arthur Koestler (romancier et essayiste hongrois, 1905–1983)

→ Eclairer la décision, mais pas la prendre !

A cet effet, quantifier le risque et les marges d'incertitude

C'est au décisionnaire de les peser et d'endosser la responsabilité de l'action.