



Cours de zététique

Pièges statistiques

Sébastien Paumier



Qu'est-ce que la zététique ?

- du grec *zêtêin* = chercher
- Littré: "méthode dont on se sert pour pénétrer la raison des choses"
- Larousse: "se dit des méthodes de recherches scientifiques: méthode zététique"
- Henri Broch: "art du doute"



Objectif du cours

- éveil de l'esprit critique:
 - jauger la fiabilité d'une information
 - percevoir les argumentations biaisées
 - viser à l'objectivité en connaissant ses limites (mauvaise intuition, illusions, etc)
- étude des mécanismes d'influence
- compréhension de la méthodologie scientifique



Thèmes étudiés

- pièges statistiques, probabilités et hasard
- paradoxes, pièges logiques, rhétorique
- illusions, manipulation, arnaques
- mesures, ordres de grandeurs
- fiabilité des informations
- méthodologie scientifique
- pseudo-sciences et charlatans



Organisation

- 12 cours+12 TD
- contrôle terminal: 1 examen sur table
- contrôle continu: 1 dossier à préparer



Le pouvoir des statistiques

- peu de gens les maîtrisent, mais tout le monde les respecte
- en imposent, même si elles sont creuses:
 - "la moto est 32% plus dangereuse que la voiture"
- aberrations non intuitives:
 - "ce prix a baissé de 250% en 3 ans"



Pourcentages

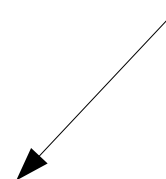
- indication d'une proportion:
 - exemple: 93% des enfants aiment les frites
- servent à exprimer une variation:
 - augmentation: "j'avais 100 billes, j'en ai maintenant 107" = +7%
 - diminution: "ce chemisier était à 400€, je l'ai payé 300€" = -25%



Pourcentages

A=quantité de départ
x=variation en % de A

$$\text{quantité d'arrivée} = A + x \times A = A \times (1 + x)$$



multiplication et non pas addition!

Conséquence: 2 augmentations de 5% \neq +10%



Pourcentages en chaîne

- si la quantité A augmente 2 fois de 5%:
 - 1^{ère} augmentation: $A + A \times 5/100 = A \times 1,05$
 - la deuxième augmentation porte sur A déjà augmenté:
 - 2^{ème} fois: $(A \times 1,05) + (A \times 1,05) \times 5/100$
 $= (A \times 1,05) \times 1,05$
 $= A \times 1,1025$
- 10,25% d'augmentation
et non pas 10%!



Pourcentages en chaîne

- +5% et -5% ne se compensent pas:

$$A \times 1,05 \times 0,95 = A \times 0,9975$$

- exemple: "après une augmentation hier de 20% du cours de l'action BINIOU, elle a subi 2 reculs successifs de 9%"

$$A \times 1,20 \times 0,91 \times 0,91 = A \times 0,99372$$

apparence: **+2%**

réalité: **-0,628%**



Importance de la formulation

- importance de la formulation des énoncés
- "ça a augmenté de moitié":
 - $100 \Rightarrow 150$?
 - $100 \Rightarrow 200$?
- "ça a augmenté de 150%":
 - $100 \Rightarrow 150$?
 - $100 \Rightarrow 250$?



Importance de la formulation

		ouvriers	cadres	
2008	salaire effectifs	200 € 1000	2000 € 100	→ 363,64 € en moyenne
2009	salaire effectifs	180 € 600	1800 € 500	→ 916,34 € en moyenne

↓ ↓

-10% -10%

Les syndicats: "Les salaires ont baissé de 10% en moyenne"

Le patron: "Le salaire moyen a augmenté de 152%"

Tout le monde dit vrai!



Importance de la formulation

2 BINIOUS achetés = 1 remboursé!

(en bon d'achat sur le même produit)

50% d'économies ?

- on en achète 2
- on a un bon pour en acheter un 3^{ème}

conclusion: 3 produits pour 2 payés
= 33% d'économies



Importance de la formulation

- "les ventes de yaourts BINIOU ont baissé de 5%"
- 5% de quoi ?
 - du nombre de yaourts BINIOU vendus ?
 - de l'argent encaissé pour les yaourts BINIOU vendus (peut différer du cas précédent si promotions) ?
 - de la proportion de yaourts BINIOU parmi tous les yaourts vendus ?



Importance de la formulation

- "la voiture est la 1^{ère} cause de décès"
- tout dépend des catégories:
 - vrai si on a (40% voiture, 20% glisser dans la salle de bain, 20% avaler de la soude, 20% s'étouffer avec un bretzel)
 - faux si on compte: (40% voiture, 60% accidents domestiques)



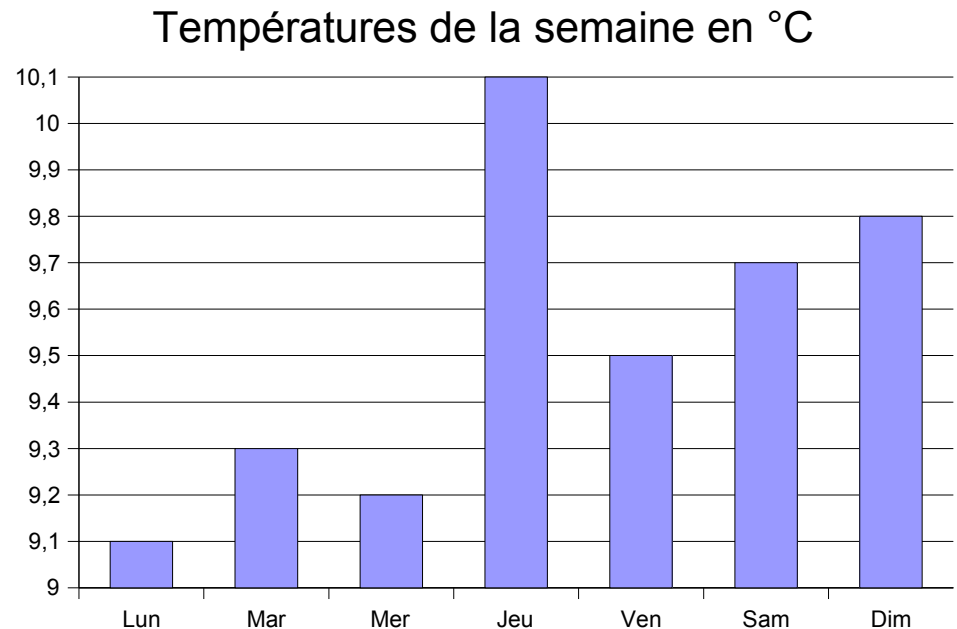
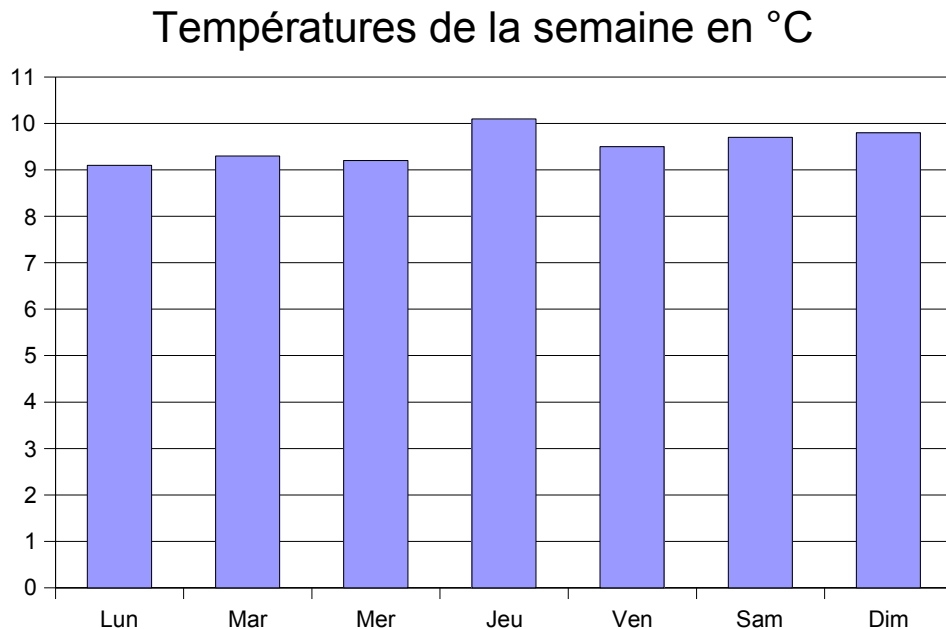
Moyenne

- moyenne habituelle:
 - somme/nombre d'éléments
- si dans une entreprise:
 - les 47 employés touchent 1000 €/mois
 - le patron touche 1 000 000 €/mois
 - la moyenne de $(47 \times 1000 + 1000000)/48$
 $= 21\,812,5$ €/mois est-elle représentative de
quoi que ce soit ?
- ne rend pas compte de la répartition des
valeurs



Courbes et histogrammes

- attention au point de départ des échelles
- ces 2 courbes de températures donnent-elles la même impression ?

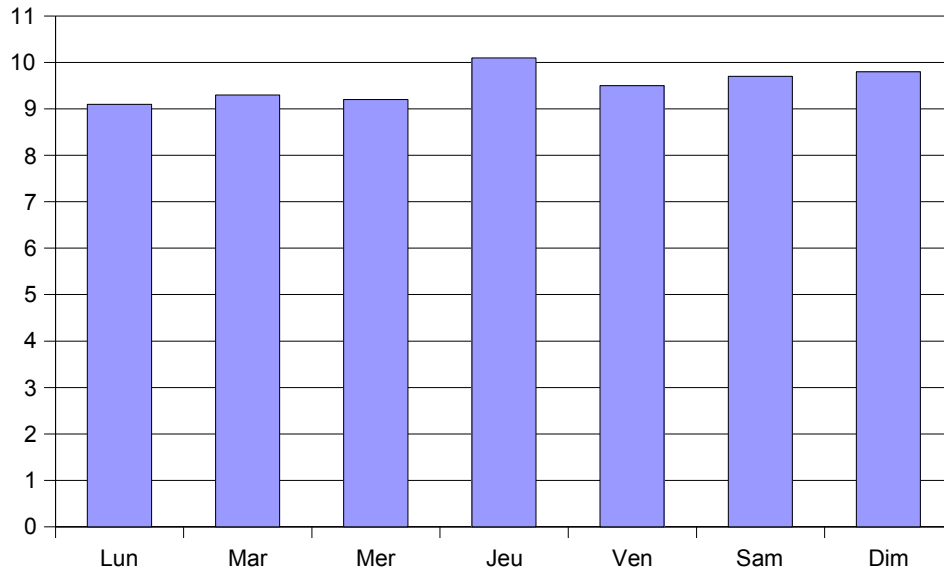




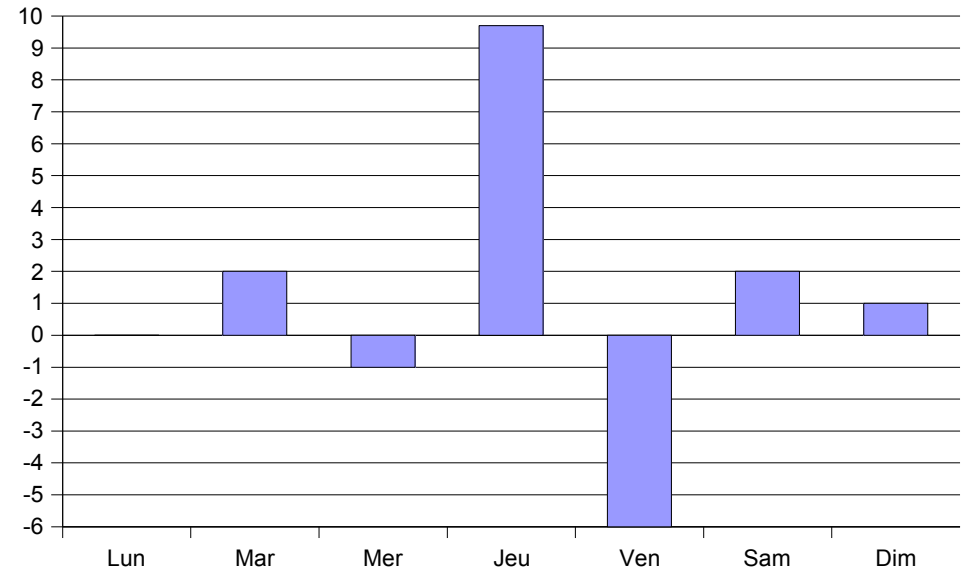
Courbes et histogrammes

- attention aux unités
- ces 2 courbes de températures donnent-elles la même impression ?

Températures de la semaine en °C



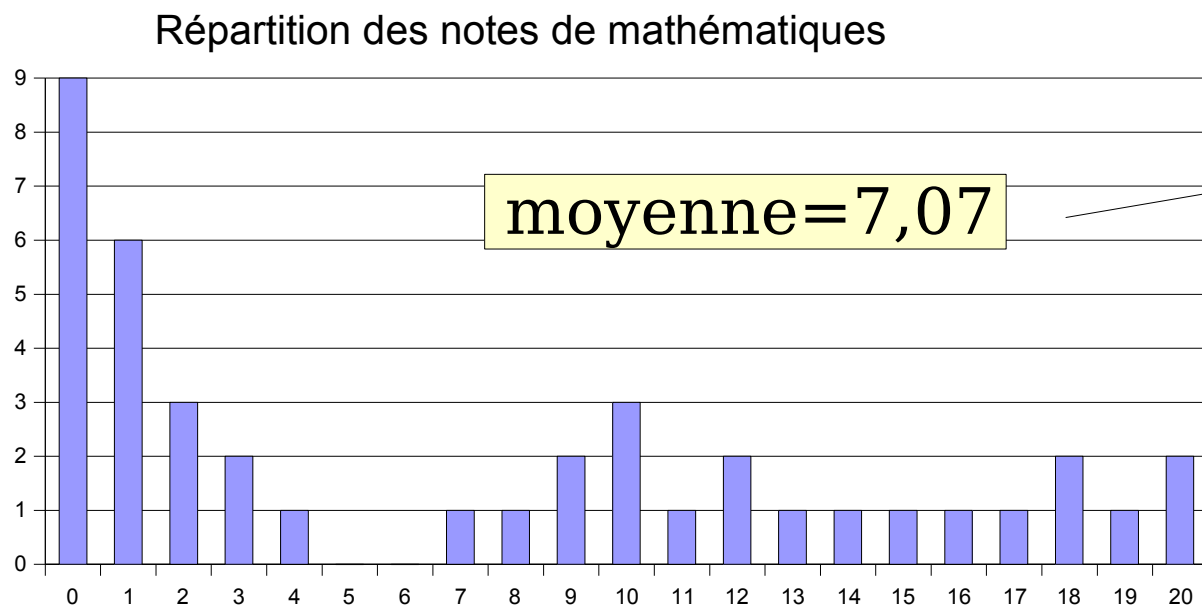
Progression des températures de la semaine en %





Courbes de distribution

- on ne représente pas une valeur, mais le nombre de sujets pour cette valeur
- permet de relativiser une moyenne



moyenne=7,07

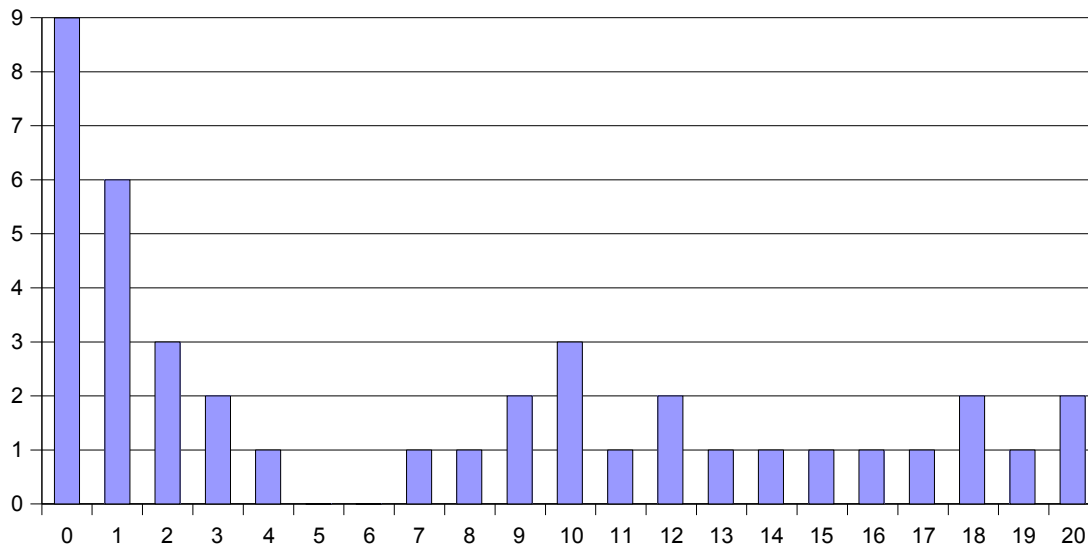
ne rend pas compte
du pic d'échec



Courbes de distribution

- médiane=valeur telle que la moitié des sujets est au-dessus et l'autre en-dessous

Répartition des notes de mathématiques



moyenne = 6,8
médiane = 4

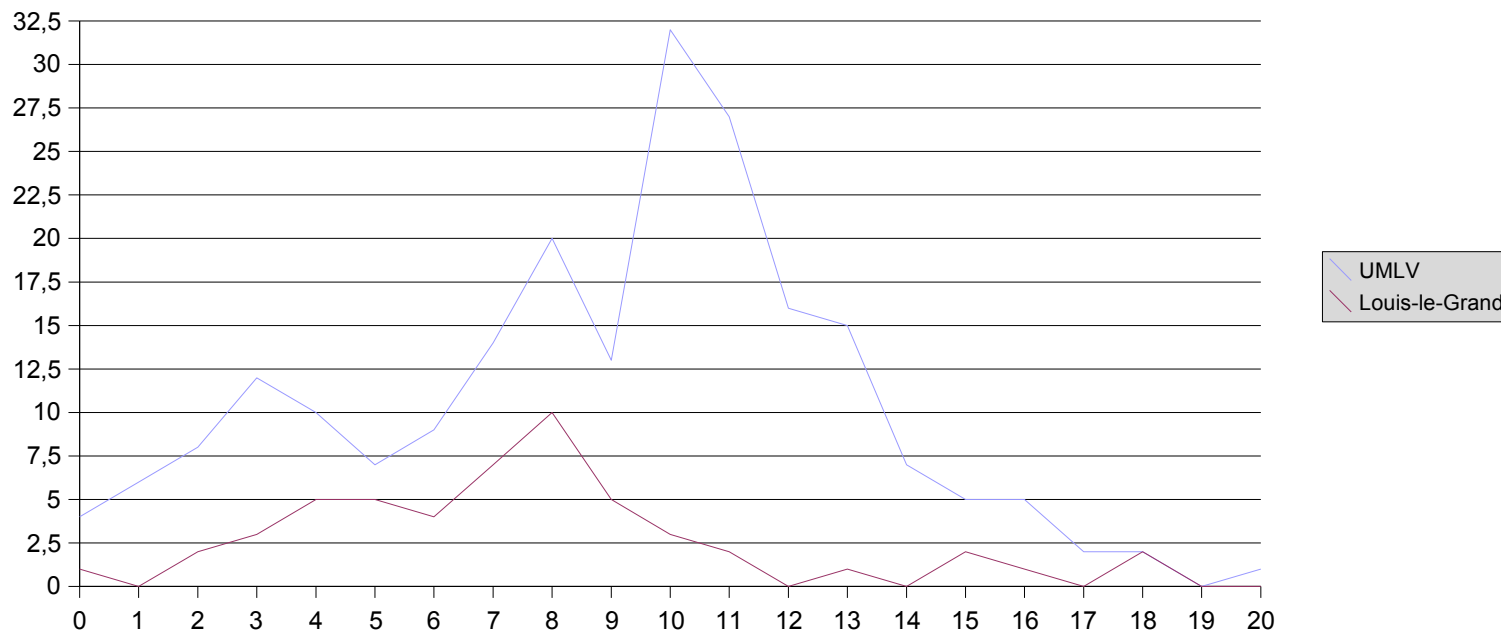




Courbes de distribution

- attention aux fausses impressions!
- les L1 de Marne-la-Vallée meilleurs en maths qu'une prépa de Louis-le-Grand ?

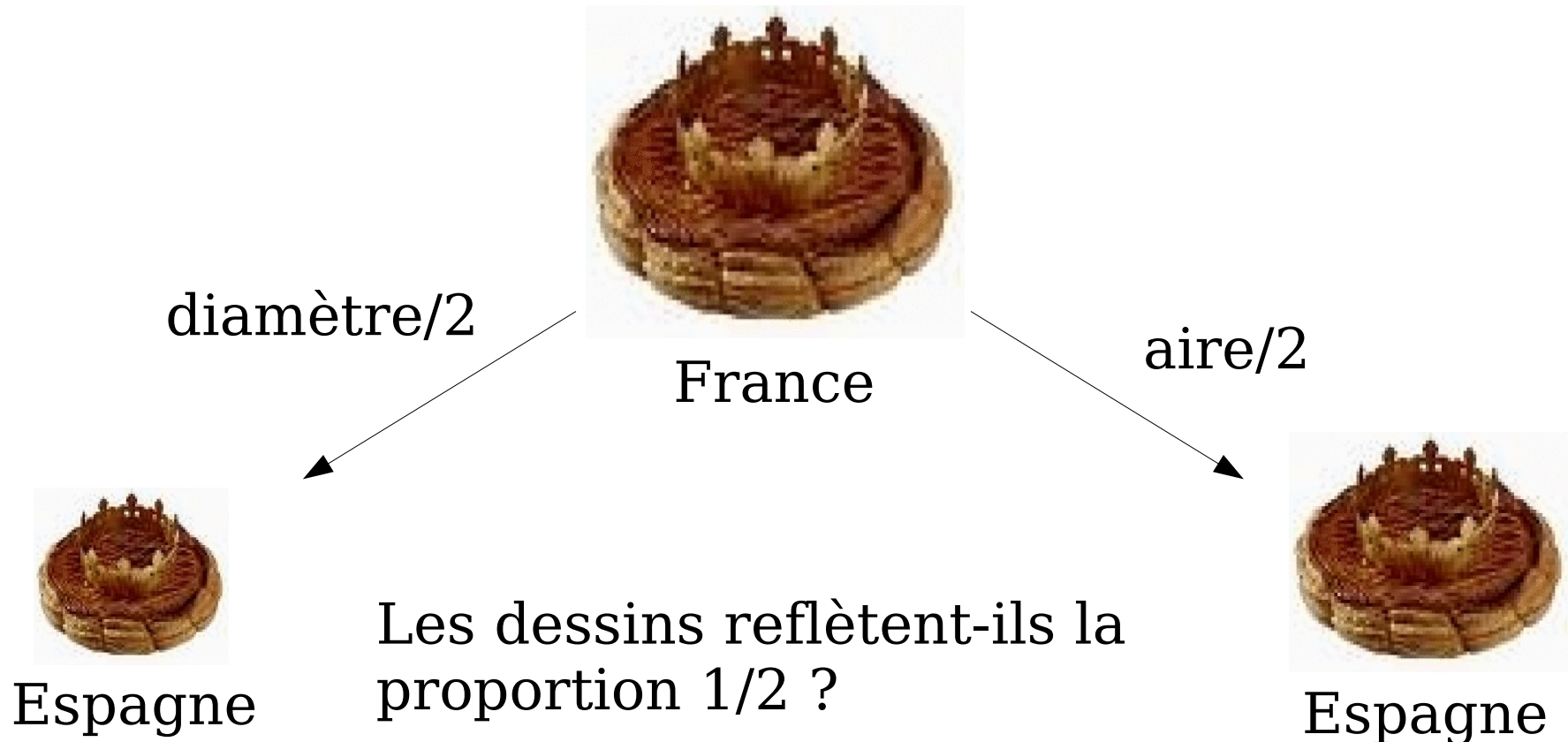
Répartition des notes de mathématiques





Proportions des dessins

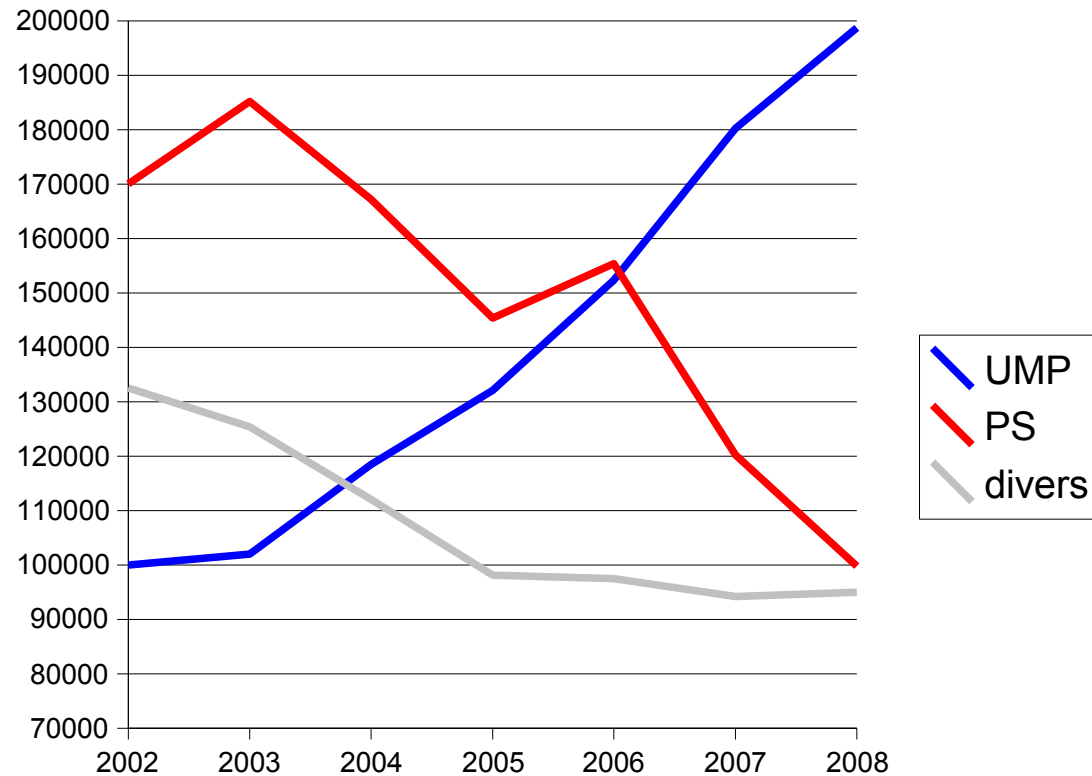
- "La France consomme 2 fois plus de galettes que l'Espagne"





Heuristiques de lecture

- Que penser du graphique suivant ?



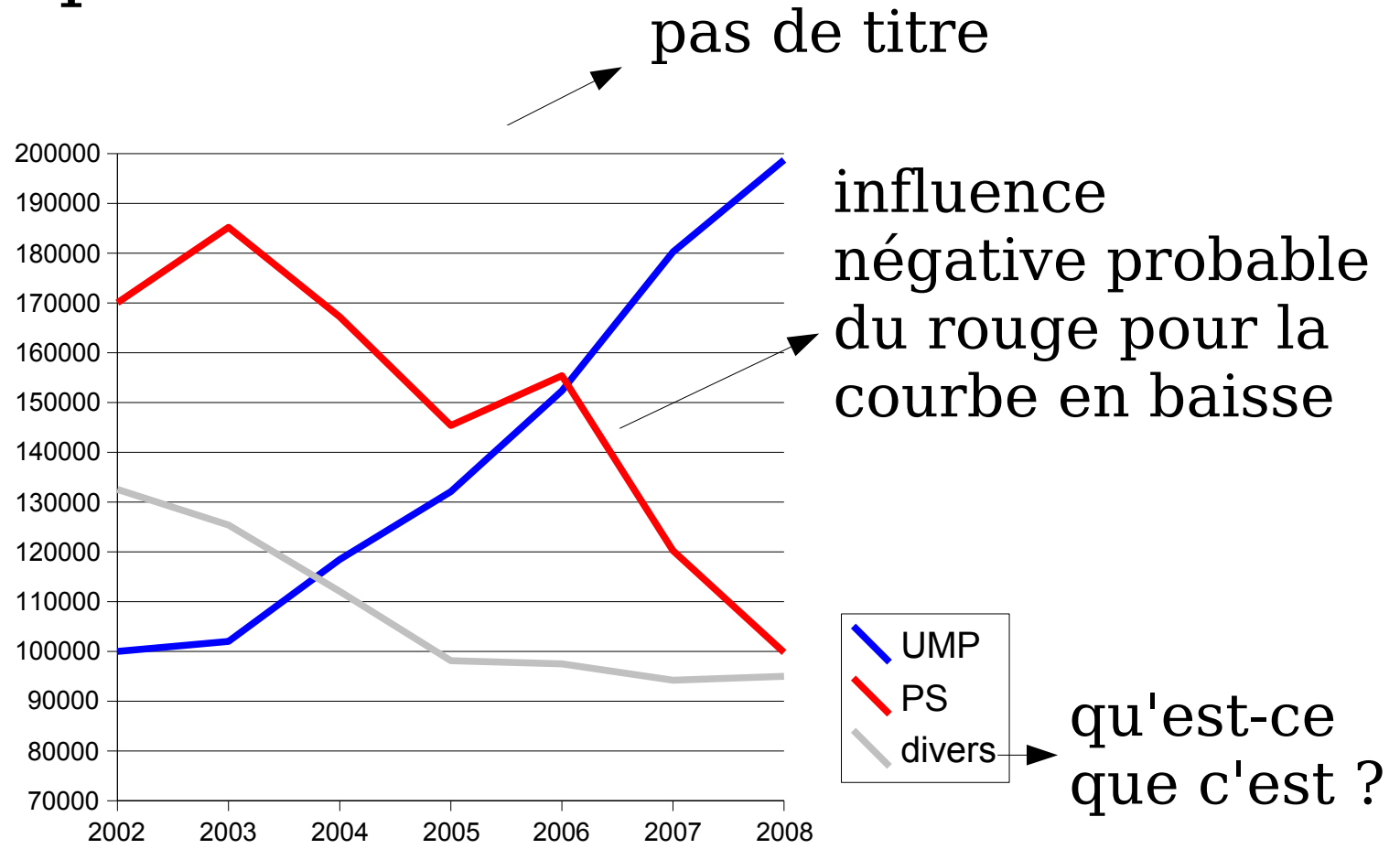


Heuristiques de lecture

- bonne réponse: rien

pas d'unité:

- dépenses ?
- militants ?
- voix ?





Heuristiques de lecture

- dangers de l'interprétation abusive:
 - ajouts souvent faux,
 - inconscients,
 - résistant à l'argumentation logique une fois ancrés dans le cerveau



Régression vers la moyenne

- les sportifs craignent de faire la une du magazine *Sports Illustrated*, car ils croient que cela leur porte la poisse
- les chiffres leur donnent raison



Quel est ce mystère ?



Régression vers la moyenne

- un sportif qui est à la une du magazine *Sports Illustrated* y est en général pour ses très bonnes performances
- il y a donc de fortes chances pour qu'il fasse moins bien ensuite
- CQFD



Et le beau temps après la pluie, c'est pareil alors ?



Les faux infimes

- "l'anorexie nerveuse touche environ 2% de jeunes femmes et entraîne la mort d'un cinquième de ses victimes"

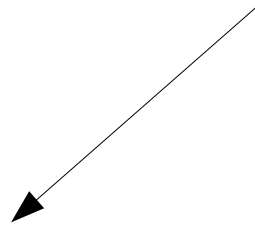
The Times, 31 mai 2000, exemple cité par J. Whyte

- ça n'a pas l'air d'être très important
- mais...



Les faux infimes

- les faits: il y avait 3,5 millions de femmes entre 15 et 25 en Grande-Bretagne en 1999
- $3,5 \text{ millions} \times 2\% \times 20\% = 14000$



- 14000 mortes par an à cause de l'anorexie: est-ce plausible ?
- en 1999: 855 décès de 15-25 ans, dont 13 cas d'anorexie



Négligence des taux de base

- une maladie mortelle touche 1 personne sur 1000
- l'opération est mortelle 1 fois sur 2
- un test fiable à 90% permet de savoir si une personne est atteinte:
 - 90% des malades sont positifs
 - 90% des sains sont négatifs
- si un patient est positif au test, faut-il l'opérer ?



Négligence des taux de base

- pour une population de 10000, on a donc:
 - 10 personnes vraiment malades et 9990 saines
- sur les 10 malades:
 - 90% seront positives au test, soit 9 déclarées malades
- sur les 9990 saines:
 - 10% seront positives au test, soit 999 déclarées malades



Négligence des taux de base

- sur ces 1008 (999+9) personnes déclarées malades, seules 9 le sont réellement, soit environ 1%
- si on opère ces 1008 personnes, on va en tuer la moitié *alors que 99% sont saines!*
- $1/1000$ =faux infime
- principe de précaution ?



Croissance exponentielle

- "D'après une étude menée par le bureau réputé Keith, Mye & Hath, le nombre de suicides par bretzel a doublé tous les 3 ans depuis 1963. Dès que ces conclusions ont été rendues publiques, les autorités ont blablabla..."
- problème de santé publique ou propagande du lobby des cacahuètes ?



Croissance exponentielle

- $2008 - 1963 = 45$ ans
- le nombre de suicides par bretzel a donc doublé $45/3 = 15$ fois
- "doublé 15 fois" = multiplié par $2^{15} = 32768$
- pour 10 morts en 1963, on en aurait donc 320000 cette année
- est-ce bien raisonnable ?



Significativité

- degré de signification d'un résultat statistique = probabilité que l'écart entre:
 - l'observation
 - la prédiction théoriquesoit dû au hasard
- exemple: tester si un dé est truqué



Significativité

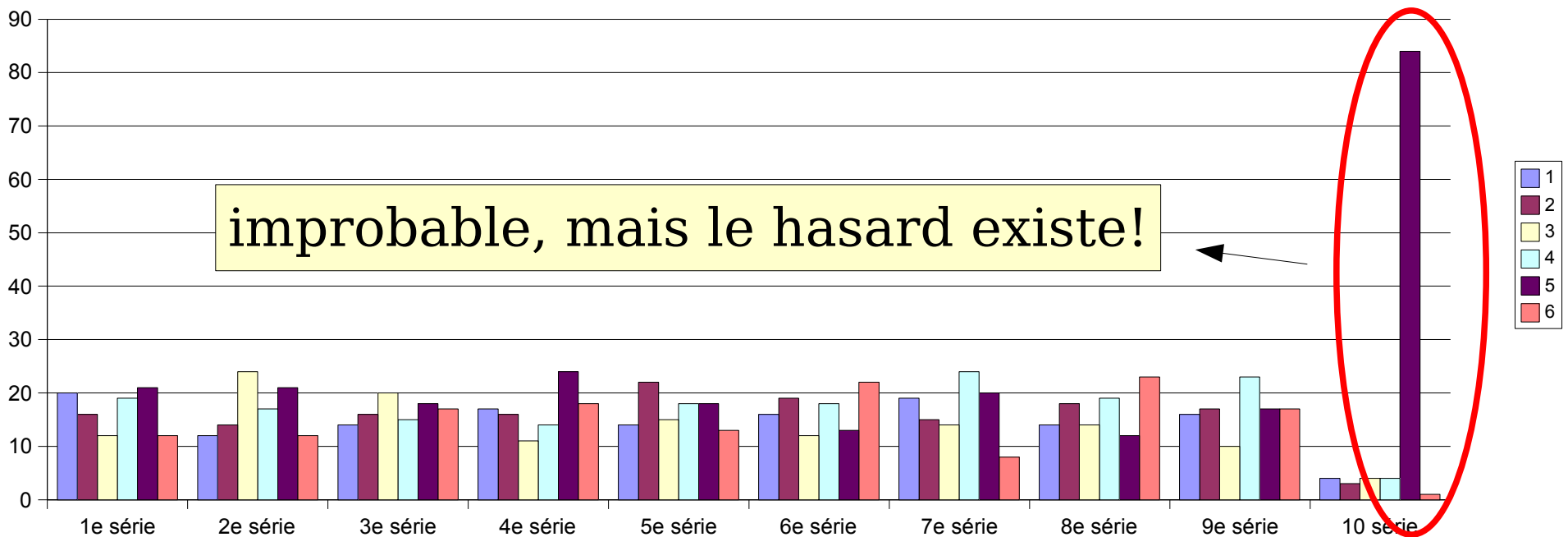
- sur 6 lancers: 3 fois le 5
 - résultat non significatif car l'échantillon est trop petit
- sur 1 million de lancers: 900000 fois le 5
 - résultat probant, mais très fastidieux à obtenir
- sur 1000 lancers: 248 fois le 5
 - que conclure ?



Échantillonnage

- il faut regarder la distribution:

échantillonnage en 10 séries de 100 lancers



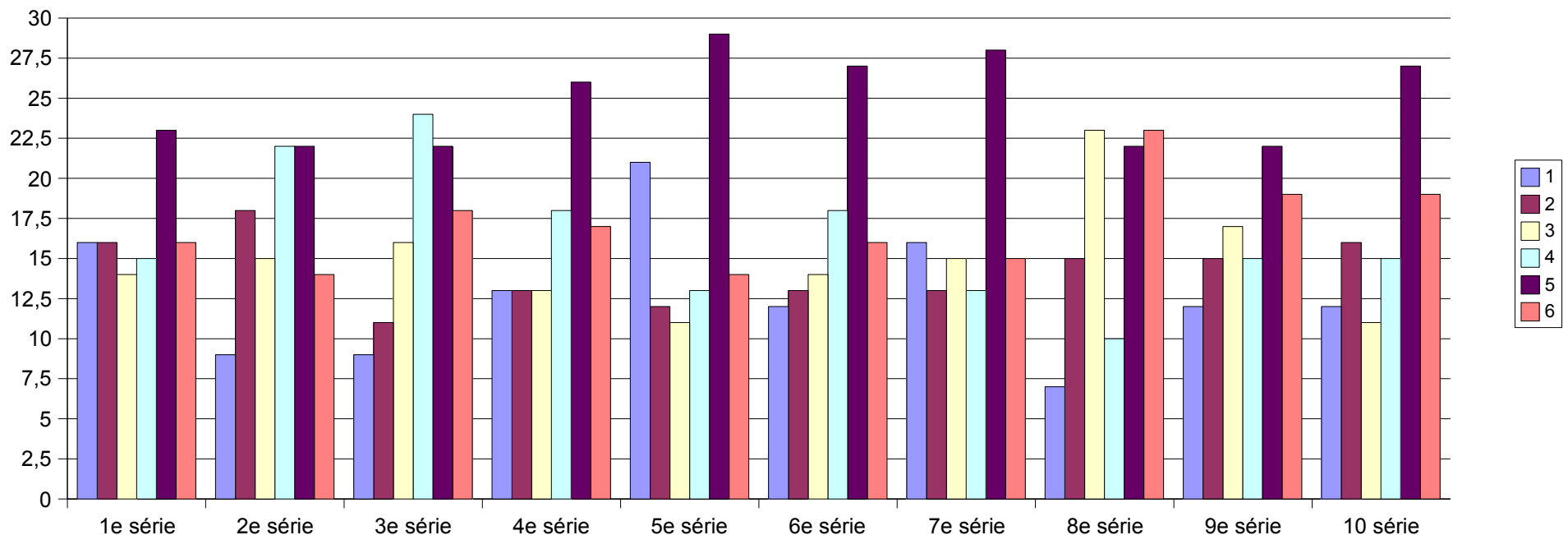
- si le dé était pipé, le 5 ressortirait plus dans chaque série



Échantillonnage

- on conclurait au dé pipé si on avait:

échantillonnage en 10 séries de 100 lancers



- cette fois, il aurait fallu 10 coups de chance, ce qui est peu raisonnable



Significativité

- le seuil de significativité dépend du test:
 - pour les dés, on compte des faits clairs
 - pour un sondage, on doit tenir compte de biais (ne pas oser avouer qu'on vote X, etc)
 - pour une mesure de similarité entre copies d'examen, il faut tenir compte du fait qu'il n'y a pas 50 façons de répondre à certaines questions
 - etc.



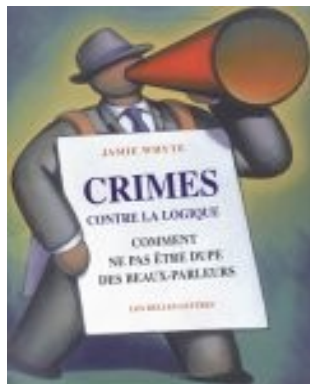
Bibliographie



Méfiez-vous des statistiques
Nicolas Gauvrit



Petit cours d'autodéfense intellectuelle
Normand Baillargeon



Crimes contre la logique
Jamie Whyte