UE ARE «Démarche» 2017-2018

Sciences en société et esprit critique

Autodéfense intellectuelle

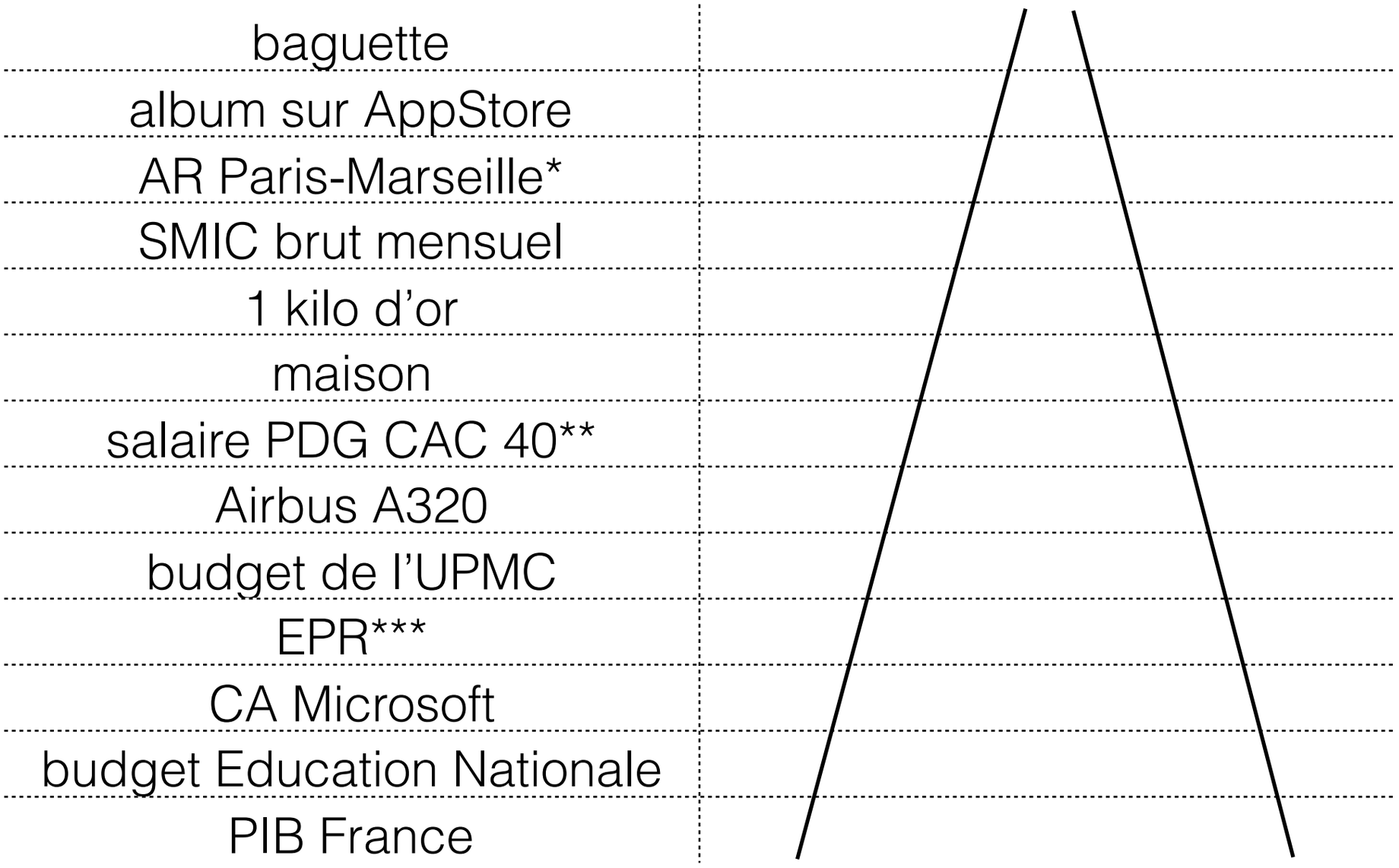
Emmanuel Guigon

Institut des Systèmes Intelligents et de Robotique
Université Pierre et Marie Curie
CNRS / UMR 7222
Paris, France

emmanuel.guigon@sorbonne-universite.fr
e.guigon.free.fr/teaching.html

Les chiffres dans la société

Ordre de grandeur - Quiz (en €)



(*) en train – (**) salaire moyen – (***) Evolutionary Power Reactor

Ordre de grandeur - Quiz (en €)

baguette	1
album sur AppStore	11
AR Paris-Marseille	250
SMIC brut mensuel	1.466
1 kilo d'or	30.000
maison	200.000
salaire PDG CAC 40	4.200.000
Airbus A320	60.000.000
budget de l'UPMC	345.000.000
EPR	8.500.000.000
CA Microsoft	31.000.000.000
budget Education Nationale	146.000.000.000
PIB France	2.200.000.000.000

Ordre de grandeur - Innumérisme

Les chiffres avancés sont-ils plausibles ?

• « 2000 enfants irakiens mouraient à chaque heure depuis dix ans à cause de l'embargo américano-britanique contre ce pays. »
• Calcul !
– 24 heures x 365 jours = 8.760 heures dans un an
– 8.760 x 2.000 enfants = 17.520.000 enfants en un an
– 17.520.000 x 10 ans = 170.520.000 enfants en 10 ans
• et cela dans un pays qui compte 30 millions d'habitants (en 2011) …

Les grands nombres

Attention !
million, milliard, billion, billiard, … en français
million, billion, trillion, quatrillion, … en anglais

Que représente 1 million € ?
- vacances pendant 1000 jours avec 1000€ par jour

Que représente 1 milliard € ?
- vacances pendant 2700 ans avec 1000€ par jour

Les grands nombres

Coût de la guerre en Irak (chiffre de 2004)
113 milliards de $

= inscription de 16.099.088 enfants à un programme d'éducation destiné aux enfants pauvres

= embauche de 2.168.932 enseignants pendant un an dans les écoles publiques

= 2.888.245 bourses universitaires de 4 ans

= 1.626.701 logements

= 404$ par américain

Définitions arbitraires

Compagnie A – Salaire moyen des employés : 22.000 $ – Salaire moyen et profits des propriétaires : 260.000 $

Compagnie B – Salaire moyen des employés : 28.065 $ – Profits moyens des propriétaires : 50.000 $

Compagnie A = Compagnie B

3 propriétaires d'une entreprise de 90 salariés
- 1.980.000 $ de salaires pour 90 salariés
- 110.000 $ de salaire par propriétaire
- 450.000 $ de profits

Définitions arbitraires

Compagnie A – Salaire moyen des employés : 22.000 $ – Salaire moyen et profits des propriétaires : 260.000 $

<u>calcul 1</u>
salaire annuel moyen = 1.980.000/90 = 22.000 $
revenu d'un propriétaire = 110.000 + 450.000/3 = 260.000 $

Compagnie B – Salaire moyen des employés : 28.065 $ – Profits moyens des propriétaires : 50.000 $

<u>calcul 2</u>
profit = 300.000 + 150.000 $
salaire annuel moyen
= 1.980.000 + 3x110.000 + 300.000 = 28.065 $
profit par propriétaire = 150.000/3 = 50.000 $

Données détachées

La publicité dit : « Plus de 80% des personnes testées ont préféré le chocolat Talou »

- Votre goût et non celui de 80% des gens
- Combien de gens ont été testés ?
- Comment l'échantillon a-t-il été réuni ?
- Combien de fois a-t-on testé pour obtenir ce résultat ?
- 800 personnes/1000, 80/100, 8/10, 4/5
- Préféré à quoi ? une autre marque ? quelques unes ? toutes ? lesquelles ?

Changement de définition

• En 1996, un chroniqueur a constaté que des millions d'Américains sont soudain devenus obèses sans avoir pris un seul kilo. L'Organisation de la santé venait de décréter qu'un Indice de Masse Corporelle (IMC) > 25 = obésité, alors qu'aux USA le critère de IMC > 27.6 était précédemment utilisé.

• En 1998, le taux de chômage au UK est passé d'un seul coup de 5% à 7%. À cause d'un changement de définition de «chômeur» comme c'était arrivé déjà 18 fois auparavant. On connait le problème en France actuellement !

Les analyses d'Ivan Illich

Automobile individuelle, la solution au problème de se déplacer de manière efficace d'un point à un autre. Mais des défauts, voire des dangers – pour l'environnement, pour la santé.
Le «système» (bureaucratique et idéologique) est incapable d'envisager de résoudre ces problèmes autrement qu'en augmentant l'offre. La voiture doit permettre de se déplacer rapidement. Chacun en possédant une, s'ensuivent des embouteillages qui ralentissent la vitesse des déplacements ; on réagit en construisant encore plus d'autoroutes, …
Selon Illich, il faut s'efforcer de repenser autrement toute la question. Pour cela, il propose une nouvelle définition de la vitesse, qui demande de considérer le coût social de la voiture (heures d'immobilité, travail pour payer la voiture, son essence, son entretien, ses assurances, les routes, autoroutes, hôpitaux, …).
Illich fait ces calculs et trouve que la vitesse sociale réelle de la voiture n'est pas significativement supérieure à celle … de la calèche.

Exercice : quelle ville est la plus dangereuse ?

• En 2013, il y a eu 50 homicides à **VilleA**, et 50 à **VilleB.** En 2008, il y avait eu 42 meurtres à **VilleA** et 29 à **VilleB.**

• Donc une augmentation de 100x(50- 42)/42=19% à **VilleA**, et 100x(50-29)/29=72% à **VilleB.**

- Mais pourcentage de quoi ?

Exercice : quelle ville est la plus dangereuse ?

VilleA : 550.000 hab. en 2008, 600.000 hab. en 2013.
VilleB : 450.000 hab. en 2008, 800.000 hab. en 2013.

	2008	2013
VilleA	42/550000 = 7,64 pour 100000 hab.	50/600000 = 8,33 pour 100000 hab.
VilleB	29/450000 = 6,44 pour 100000 hab.	50/80000 = 6,25 pour 100000 hab.
Evolution du pourcentage	**?**	**?**

Exercice : quelle ville est la plus dangereuse ?

VilleA : 550.000 hab. en 2008, 600.000 hab. en 2013.
VilleB : 450.000 hab. en 2008, 800.000 hab. en 2013.

	2008	2013
VilleA	42/550000 = 7,64 pour 100000 hab.	50/600000 = 8,33 pour 100000 hab.
VilleB	29/450000 = 6,44 pour 100000 hab.	50/800000 = 6,25 pour 100000 hab.
Evolution du pourcentage	**VilleA** 100x(8,33-7,64)/7,64 = **+9,03% !**	**VilleB** 100x(6,25-6,44)/6,44 = **-2,95% !**

Probabilités

- Au XVIIe, le chevalier de Méré pose une énigme à Blaise Pascal. Méré joue aux dés, et essaie d'en tirer des règles pour gagner.

- Il calcule qu'en tirant 4 fois un dé de six, il aura 4x1/6=2/3 chances de sortir un six.

- Il calcule qu'en tirant 24 fois deux dés de six, il aura 24x1/36=2/3 chances de sortir un double six.

- Son raisonnement est-il juste ?

Probabilités

- P_1 = probabilité de tirer au moins 1 six en 4 jets
- P_2 = probabilité de ne tirer aucun six en 4 jets
- $P_1 = 1 - P_2 = 1 - (5/6)^4 = 1-625/1296 =$ **0,518**

- P_3 = probabilité de tirer au moins 1 double six en 24 jets
- P_4 = probabilité de ne tirer aucun double six en 24 jets
- $P_3 = 1 - P_4 = 1 - (35/36)^{24} = 1 - 0{,}509 =$ **0,491**

Déjouer les fausses prémonitions/superstitions

Une enquête a montré que la plupart des médiums célèbres sont des fils aînés

- Etonnant ?

Déjouer les fausses prémonitions/superstitions

Une enquête a montré que la plupart des médiums célèbres sont des fils aînés

• Pas étonnant : dans une population donnée, surtout lorsque les familles ont assez peu d'enfants (2,3,4), il y a toujours plus de fils aînés. Imaginons une population fictive de familles de 2 enfants chacune. On aura à proportions égales les compositions suivantes :

Fille/Fille – Fille/Garçon – Garçon/Fille – Garçon/Garçon

• Dans 3 cas sur 4, un fils est un fils aîné et une fille est une fille aînée.

Déjouer les fausses prémonitions/superstitions

Prémonition ? Monsieur Paul pensait à une personne X, et dans les 5 minutes qui ont suivi, il reçoit un coup de téléphone lui annonçant que Monsieur X vient de mourir

- Prémonition ?
- Quelle était la probabilité que ça arrive ?

Déjouer les fausses prémonitions/superstitions

- Prémonition ?

- Exercice : Supposons que M. Paul connaisse 1000 personnes (au sens large : ça inclut Zidane par ex) dont il apprendra le décès dans les 30 ans. Ceci est très modeste. Supposons que M. Paul ne pense à chacune de ces personnes qu'une fois en 30 ans (ce qui est très très modeste). Quelle est la probabilité que M. Paul pense à une de ces personnes et que dans les 5 minutes on l'appelle pour lui dire qu'elle vient de mourir ?

Déjouer les fausses prémonitions/superstitions

- Prémonition ?
- Il y a 15.768.000 minutes en 30 ans
- P_1 = probabilité de penser à une personne et qu'on annonce son décès dans les 5 min = 5/15.768.000
- P_2 = probabilité qu'au moins une des fois où on pense à une des 1000 personnes, on annonce son décès dans les 5 minutes = ?

Déjouer les fausses prémonitions/superstitions

- P_1 = probabilité de penser à une personne et qu'on annonce son décès dans les 5 min = 5/15.768.000

- P_2 = probabilité qu'au moins une des fois où on pense à une des 1000 personnes, on annonce son décès dans les 5 minutes = 1 – P_3

- P_3 = probabilité qu'aucun des décès des 1000 personnes ne suivent une pensée de M. Paul = P(ça n'arrive pas pour la personne 1) x P(pas pour 2) x P(pas pour3) x…x P(pas pour 1000) = $(1-P_1)^{1000}$

- Donc P_2 = 1 – P_3 = 1 – 0,999682952 = 0,000317047

- Très faible certes, mais …

Déjouer les fausses prémonitions/superstitions

- $P_2 = 1 - P_3 = 1 - 0{,}999682952 = 0{,}000317047$
- Très faible certes, mais …
- Il y a 50 millions d'habitants dans le pays de M. Paul

• P_4 = probabilité que ça arrive pour au moins un des habitants = $1 - P_5$

• P_5 = probabilité que ça n'arrive à aucun des 50M d'hab = $(1-P_2)^{50000000}$. Donc $P_4 = 0{,}999999999\ldots \approx 1$

• Il y aura en moyenne P_2 x 50M = 15852 cas en 30 ans, soit 528 cas par an. C'est suffisant pour que les gens croient à des prémonitions !

Déjouer les fausses prémonitions/superstitions

- $P_2 = 1 - P_3 = 1 - 0{,}999682952 = 0{,}000317047$
- Très faible certes, mais …
- Même pour un village de 31.232 habitants il sera quasi certain que ça arrive pour au moins un des habitants (P_4’)

- $P_4' = 1 - P_5'$ (probabilité que ça n’arrive à aucun des 31.232 hab) $= 1 - (1-P_2)^{31232} = 0{,}9999 \approx 1$

- Il y aura en moyenne P_2 x 31.232 = 10 cas en 30 ans, soit 1 cas tous les trois ans. C’est suffisant pour que les gens croient à des prémonitions !

Bon anniversaire

Vous connaissez 23 personnes qui sont susceptibles de vous inviter à leur anniversaire. Quelle est la probabilité que deux anniversaires arrivent le même jour ?

La 1ère personne est née n'importe quel jour. Il y a donc 1/365 chance que la 2ème personne soit née ce même jour, soit 364/365 chances qu'elle soit née un autre jour. Pour la 3ème personne, il y a 2/365 chances qu'elle soit née le même jour que l'une ou l'autre des deux premières et 363/365 qu'elle soit née un autre jour. Pour les 23 personnes, la probabilité qu'aucun anniversaire ne coïncide avec un autre est 364/365 x 363/365 x … x 342/365 = 0,46.
Il y a donc plus d'une chance sur deux (54%) que deux anniversaires tombent le même jour parmi un groupe de 23 personnes.

Les faux positifs

Une grave maladie affecte 1 personne sur 1 000 au sein d'une population. Un test existe pour détecter cette maladie : il détecte la maladie, lors qu'elle est présente, dans 99% des cas (et donc ne la détecte pas dans 1% des cas). D'un autre côté, il ne détecte pas de maladie, lorsqu'elle n'est pas présente, dans 98% des cas (et donc la détecte dans 2% des cas, appelé faux positifs).
Le médecin annonce à un patient que son résultat au test est positif. Cette personne doit-elle s'inquiéter ?

A = le patient a la maladie - P(A) = 0.001
B = le test est positif - P(B/A) = 0.99 - P(B/nonA) = 0.02

$$P(A/B) = \frac{P(A)P(B/A)}{P(A)P(B/A) + P(nonA)P(B/nonA)} \rightarrow 4.7\%$$

Statistiques : Mesures de tendance centrale

Définitions

- Moyenne = somme de toutes les mesures divisée par le nombre de mesures

- Médiane = la valeur de la mesure qui se situe au milieu de l'échantillon (qui le coupe en 2 échantillons de même taille)

- Mode = la valeur la plus fréquente parmi l'ensemble des mesures

Statistiques : moyenne, médiane, mode

Prix des matraques Bang chez huit fournisseurs du Service de la Police de Montréal

100$, 129$, 129$, 135$, 139$, 149$, 159$, 179$

- Moyenne = Σ/8 = 141$
- Médiane = (135+139)/2 = 137$
- Mode = 129 $

Les mesures de tendance centrale ne sont pas très différentes mais ce n'est pas toujours le cas.

La compagnie ZZZ

direction : le patron, son frère et six parents
le personnel : 5 contremaîtres et 10 ouvriers

Paul est candidat pour un poste d'employé

Le salaire moyen dans la compagnie est de 6.000 $ par mois.
Durant la période d'essai, Paul touchera 1.500 $ par mois.
Puis, son salaire augmentera vite.

Paul est embauché. Mais après quelques jours, en colère, il demande à voir le patron.
— Vous m'avez menti ! Aucun des ouvriers de ZZZ ne gagne plus de 2.000 $ par mois

La fabrique de
Schpountz
Depuis longtemps

6 000 $
moyenne

4000$
médiane

2 000 $
mode

48 000 $	Le patron
20 000 $	Le frère
5 000 $	les parents
4 000 $	Les contremaîtres
2 000 $	Les employés

Salaires mensuels par personne

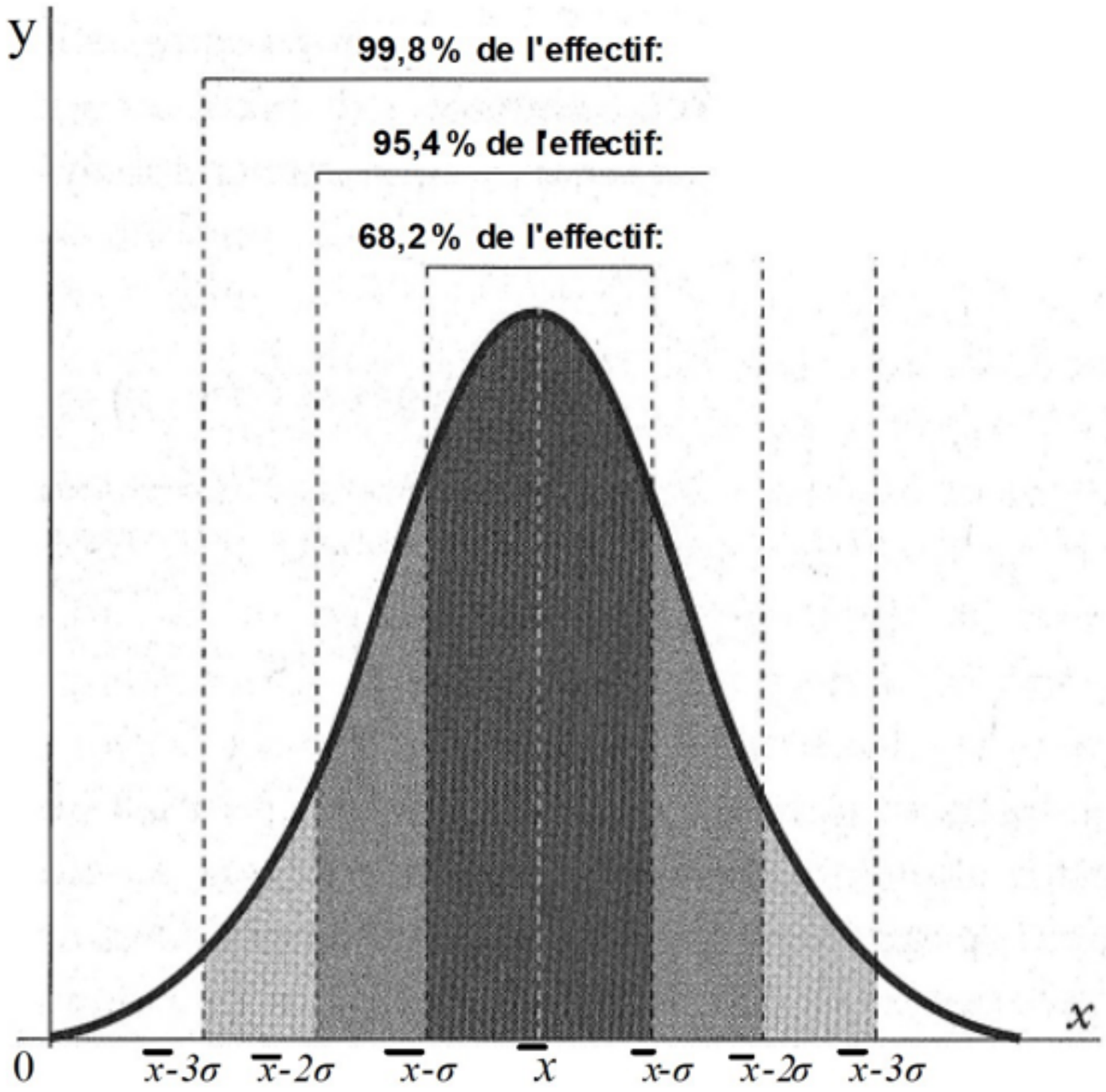
Statistiques : écarts-types

• Ne pas regarder que les mesures de tendance centrale, mais tenir compte aussi des écarts-types

• La dispersion des mesures donne une idée de leur fiabilité, de l'intervalle de confiance

Ex : Vous avez pêché des poissons dans une eau polluée qui rend certains poissons impropres à la consommation. Supposons que la toxicité des poissons se distribue selon une courbe normale. A compter de 7 mg de Cecicela (produit toxique déversé en secret par l'usine de Schpountz), il devient dangereux de manger un poisson. La moyenne des quantités de Cecicela retrouvées dans les poissons de ce cours d'eau est de 4 mg. Allez-vous en manger ?

Loi normale



$\bar{x}$ moyenne σ écart-type

Statistiques : écarts-types

Ex : Vous avez pêché des poissons …

- Moyenne de l'échantillon $\bar{x} = 4$ mg

– écart-type $\sigma = 1$ mg

99,8% de l'effectif a une toxicité comprise entre $\bar{x}-3\sigma$ et $\bar{x}+3\sigma$, donc entre 1 et 7
prob(toxicité > 7) = **0,001**

– écart-type $\sigma = 4$ mg

68,2% de l'effectif a une toxicité comprise entre $\bar{x}-\sigma$ et $\bar{x}+\sigma$, donc entre 0 et 8
prob(toxicité > 7) > **0,159**

Les sondages

Inférer les propriétés d'une population à partir de l'examen d'une petite partie de cette population (échantillon)

- un sondage n'est pas un recensement

- comment construire un échantillon ?
- une cuillerée de soupe pour juger une soupe
- quelques modèles de matraque pour évaluer les matraques

- échantillon représentatif : suffisamment grand et non biaisé

Les sondages du Literary Digest

Revue américaine très lue en son temps, qui menait depuis les années 1920 des sondages lors de la tenue d'élections présidentielles

Année	Envoi (en M)	Votants (en M)	Ecart (%)*
1920	11	?	6
1924	16,5	?	5,1
1928	18	?	44
1932	20	3	0,9
1936	10	2,3	FAUX

(*) par rapport aux résultats effectifs

Les sondages du Literary Digest

En 1936, sur la base des 2,3 millions de votes retournés sur les 10 millions envoyés, la revue avait annoncé l'élection de Alfred Mossznan Landon, l'adversaire républicain du démocrate Franklin Delano Roosevelt.

Un jeune psychologue, George Gallup, avait pour sa part interrogé 4500 personnes et, sur cette base, prédit l'élection de F. D. Roosevelt. C'est ce dernier qui l'avait effectivement emporté, avec 60,8 % des voix, contre seulement 36,6 % pour son adversaire, soit l'une des plus importantes majorités de toutes les présidentielles américaines.

Les sondages du Literary Digest

Explication : échantillon grand mais biaisé

La revue choisissait les gens à qui elle envoyait un bulletin parmi ses abonnés et au hasard dans le bottin de téléphone. Par ces deux manières, elle sur-sélectionnait des gens plus fortunés et plus enclins à voter pour le candidat républicain (puisqu'ils avaient choisi de s'abonner à cette revue plutôt conservatrice ou qu'ils avaient, en 1936, les moyens de se payer le téléphone).

Les sondages

- Comment choisir la taille de l'échantillon ?

Pas de réponse simple. Cela dépend de la population étudiée, du degré de précision souhaité, des moyens financiers disponibles, de la question posée, …

La plupart des sondages portent sur des échantillons de 1000 à 2000 personnes. Au delà, la précision obtenue ne vaut pas la dépense.

- Comment sélectionner les individus ?

- Au hasard
- Par stratification, par grappes, par quotas

Les sondages

- Comment poser la question ?

Une bonne question n'est ni ambiguë, ni biaisée ; posée de la même manière à tous les sondés, elle est comprise par chacun de la même façon.

Ex : 86 % des policiers interrogés sont favorables à l'achat de nouvelles matraques Bing suite à la question

« Compte tenu de l'augmentation du nombre de dangereux anarchistes et de l'efficacité démontrée des matraques Bing pour les ramener à la raison d'État, approuvez-vous le remplacement des matraques désuètes de la Police par les économiques et ergonomiques matraques Bing ? »

Les sondages pré-électoraux

Pierre Bourdieu

Les présupposés de ces sondages sont contestables, puisqu'ils présument que chacun peut avoir une opinion ; que toutes les opinions se valent ; et qu'il y a « consensus sur les problèmes, autrement dit qu'il y a accord sur les questions qui méritent d'être posées ».

L'opinion publique que dévoilent les sondages est « un artefact pur et simple, dont la fonction est de dissimuler que l'état de l'opinion, à un moment donné du temps, est un système de forces, de tensions et qu'il n'est rien de plus inadéquat pour représenter l'état de l'opinion qu'un pourcentage »

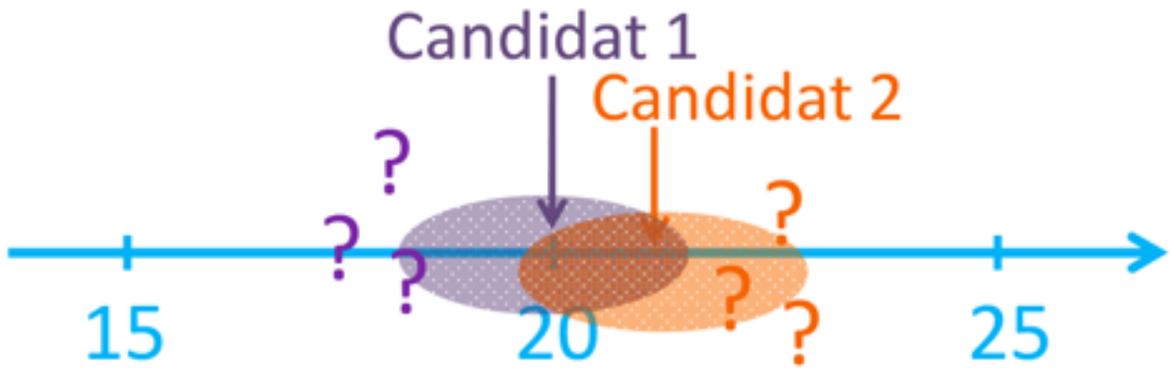
Les chiffres dans les sondages

% intentions de vote	
candidat 1	candidat 2
20	22

Le candidat 2 en tête !

➔ Réelle différence ?

Les chiffres dans les sondages



Intervalle de fluctuation (approximation)*

Prop. intentions de vote Nb de personnes qui se prononcent

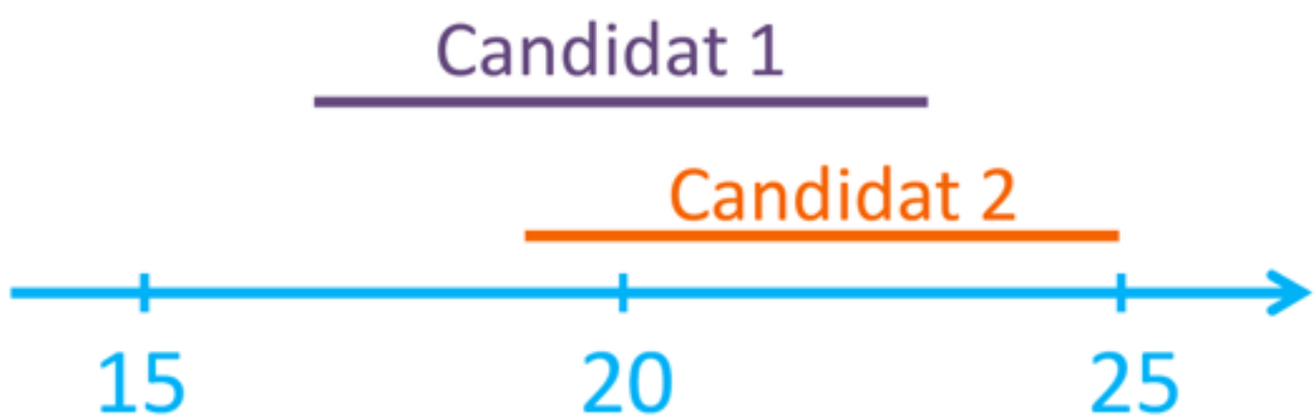
$$I = \left[p - \frac{1}{\sqrt{n}}; \; p + \frac{1}{\sqrt{n}} \right]$$

* formule simplifiée de l'intervalle fluctuation quand $n \geq 25$ et $0{,}2 \leq p \leq 0{,}8$

Les chiffres dans les sondages

Échantillon de **1000** personnes

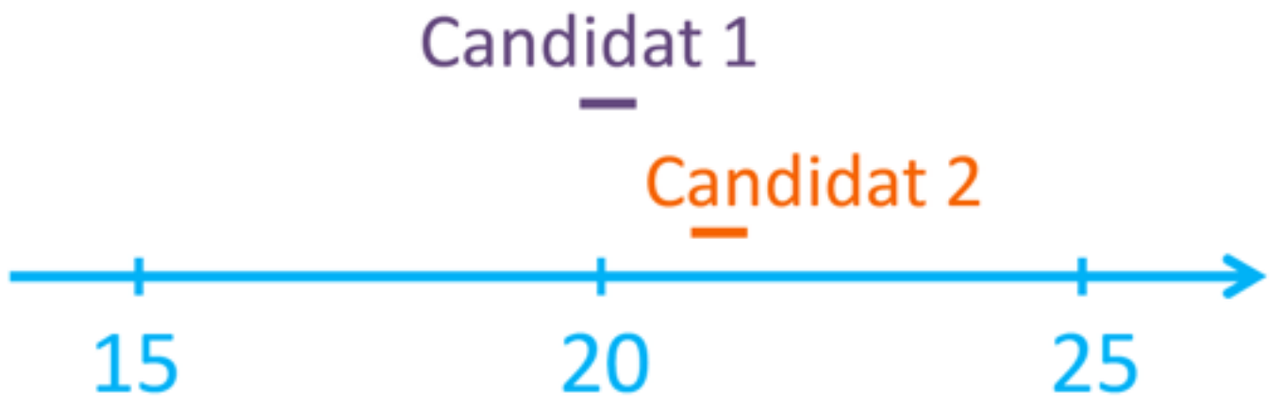
Prop. intentions de vote	
candidat 1	candidat 2
20 **+/- 3**	22 **+/- 3**
17-23	**19-25**



Les chiffres dans les sondages

Échantillon de **100000** personnes

Prop. intentions de vote	
candidat 1	candidat 2
20 **+/- 0.3**	22 **+/- 0.3**
19.7-20.3	**21.7-22.3**



Les chiffres dans les sondages

~~Échantillon de **1000** personnes~~

Prop. intentions de vote	
candidat 1	candidat 2
20 **+/- 3**	22 **+/- 3**
17-23	**19-25**

40 % ne se prononcent pas

Échantillon de **600** personnes

Prop. intentions de vote	
candidat 1	candidat 2
20 **+/- 4**	22 **+/- 4**
16-24	**18-26**

Corrélation et causalité

Post hoc ergo procter hoc : le fait d'avoir constaté et établi une corrélation ne signifie pas qu'on ait trouvé une relation de causalité

Les variables A et B sont corrélées. Cela peut signifier que :

- A cause B
- B cause A
- A et B sont accidentellement liés sans avoir entre eux de lien de causalité
- A et B dépendent d'un troisième facteur C

Régression vers la moyenne

Francis Galton (1822–1911) : relation entre la taille des pères et celle des fils

– des pères de grande taille tendent à avoir des fils de grande taille et des pères de petite taille tendent à avoir des fils de petite taille

– des pères particulièrement grands tendent à avoir des fils moins grands qu'eux et, inversement, des pères de très petite taille tendent à avoir des rejetons moins petits qu'eux.

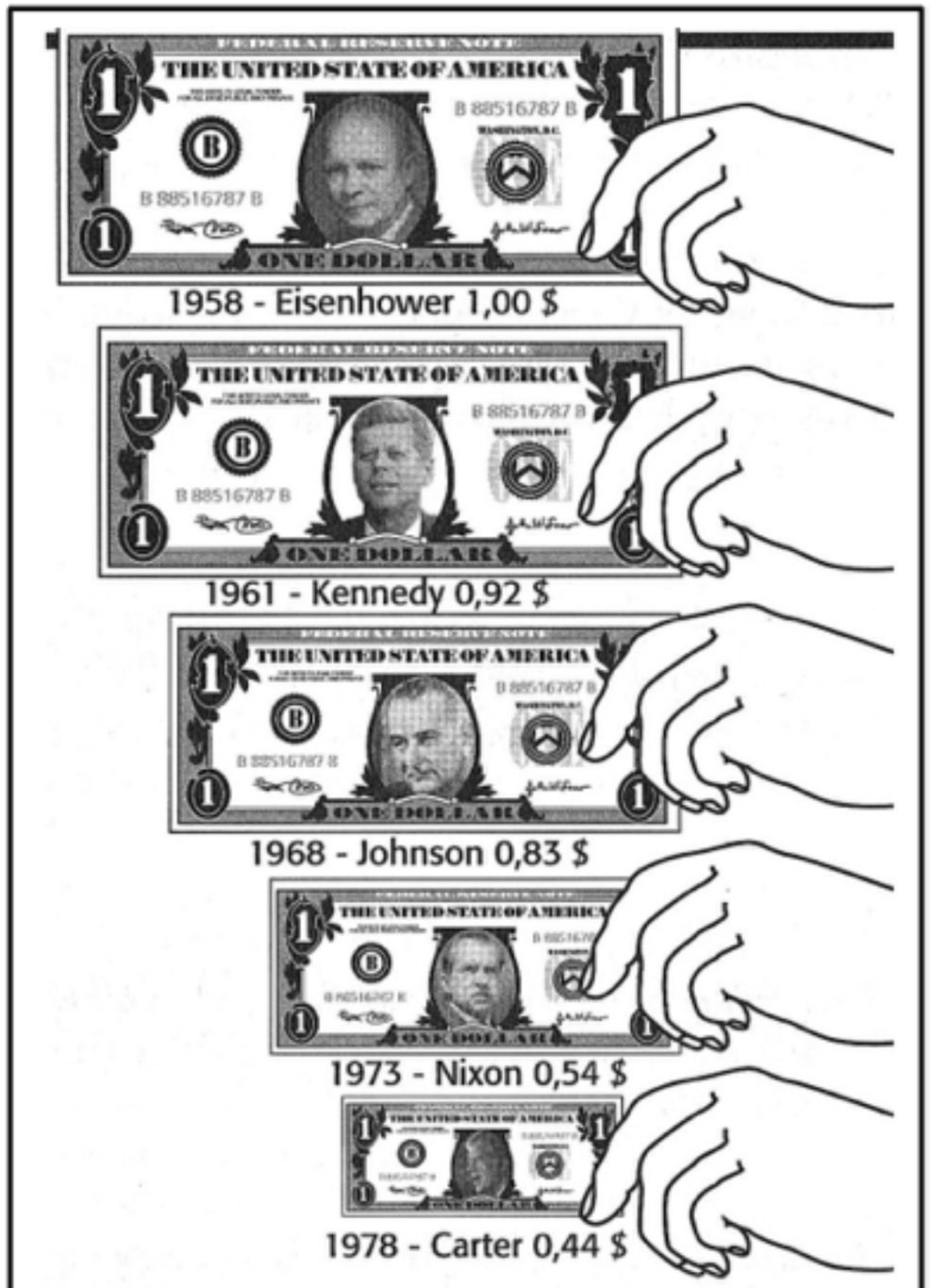
Régression vers la moyenne

– des pères particulièrement grands tendent à avoir des fils moins grands qu'eux et, inversement, des pères de très petite taille tendent à avoir des rejetons moins petits qu'eux.

Explication : Corrélations imparfaites entre deux variables

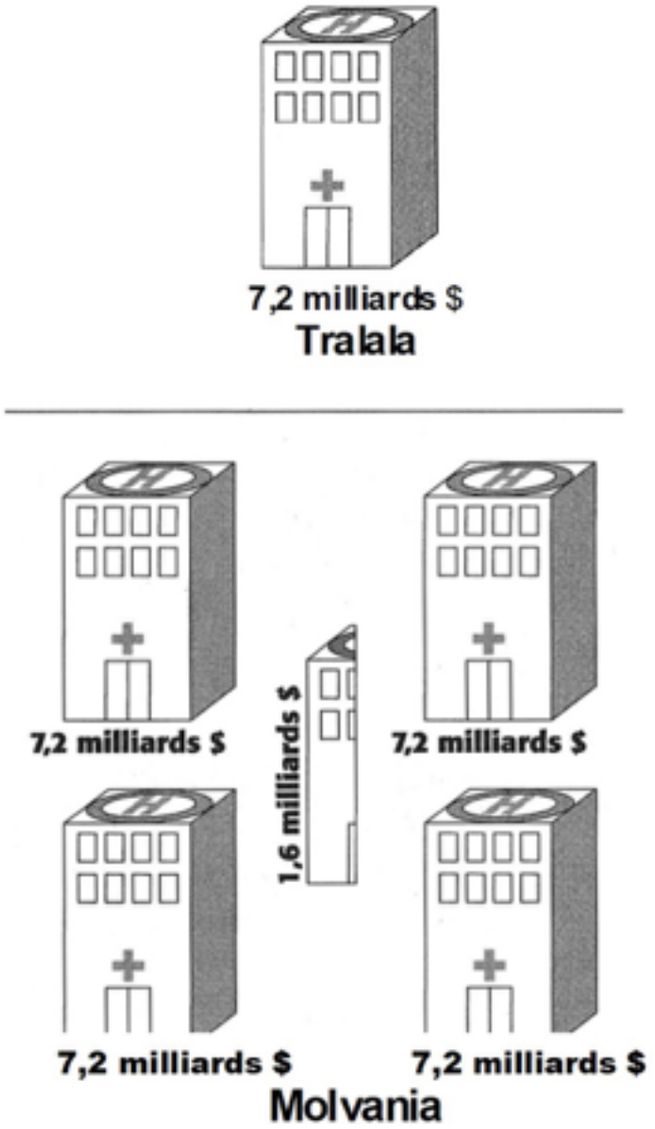
De nombreux facteurs définissent la taille d'une personne : la taille de son père, de sa mère, son environnement, … Une taille extrême est due à un concours de circonstance exceptionnel. Cet événement exceptionnel tend à être corrélé avec un événement moins exceptionnel.

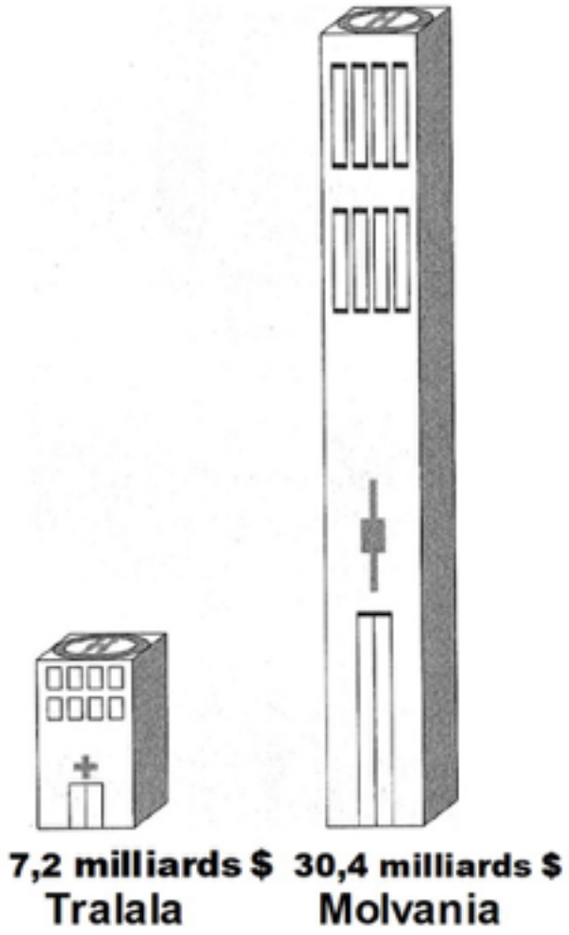
Illustrations et graphiques

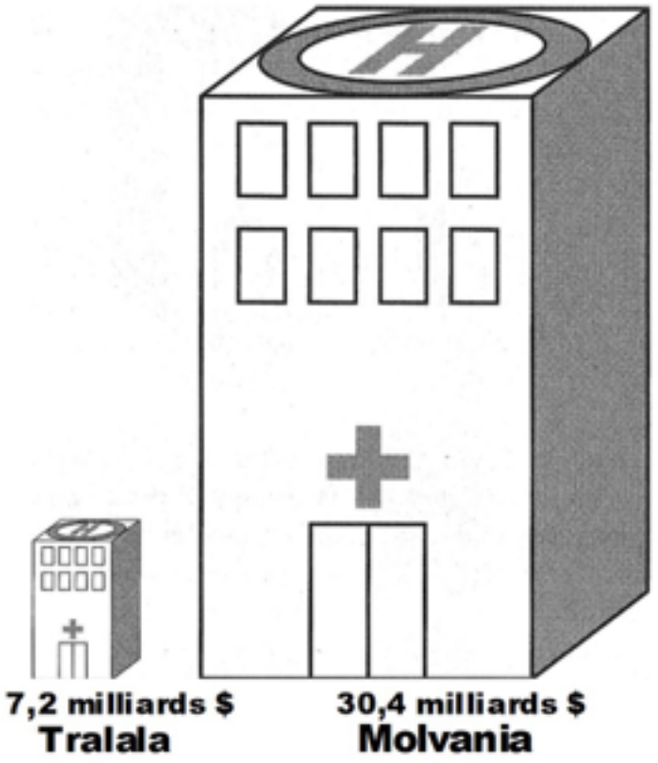


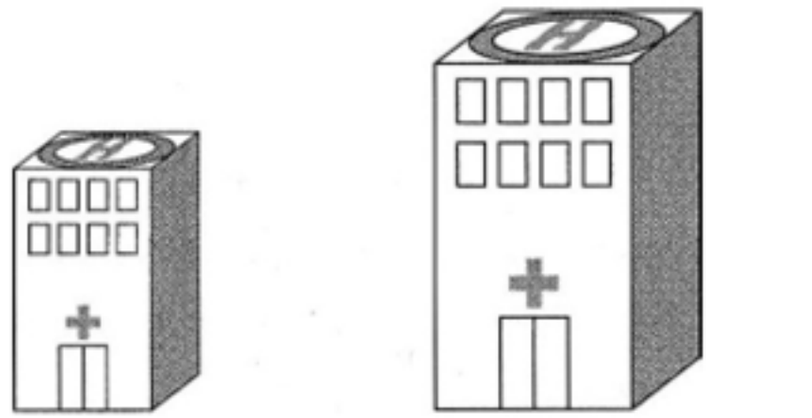
La dévaluation du dollar

Illustrations et graphiques

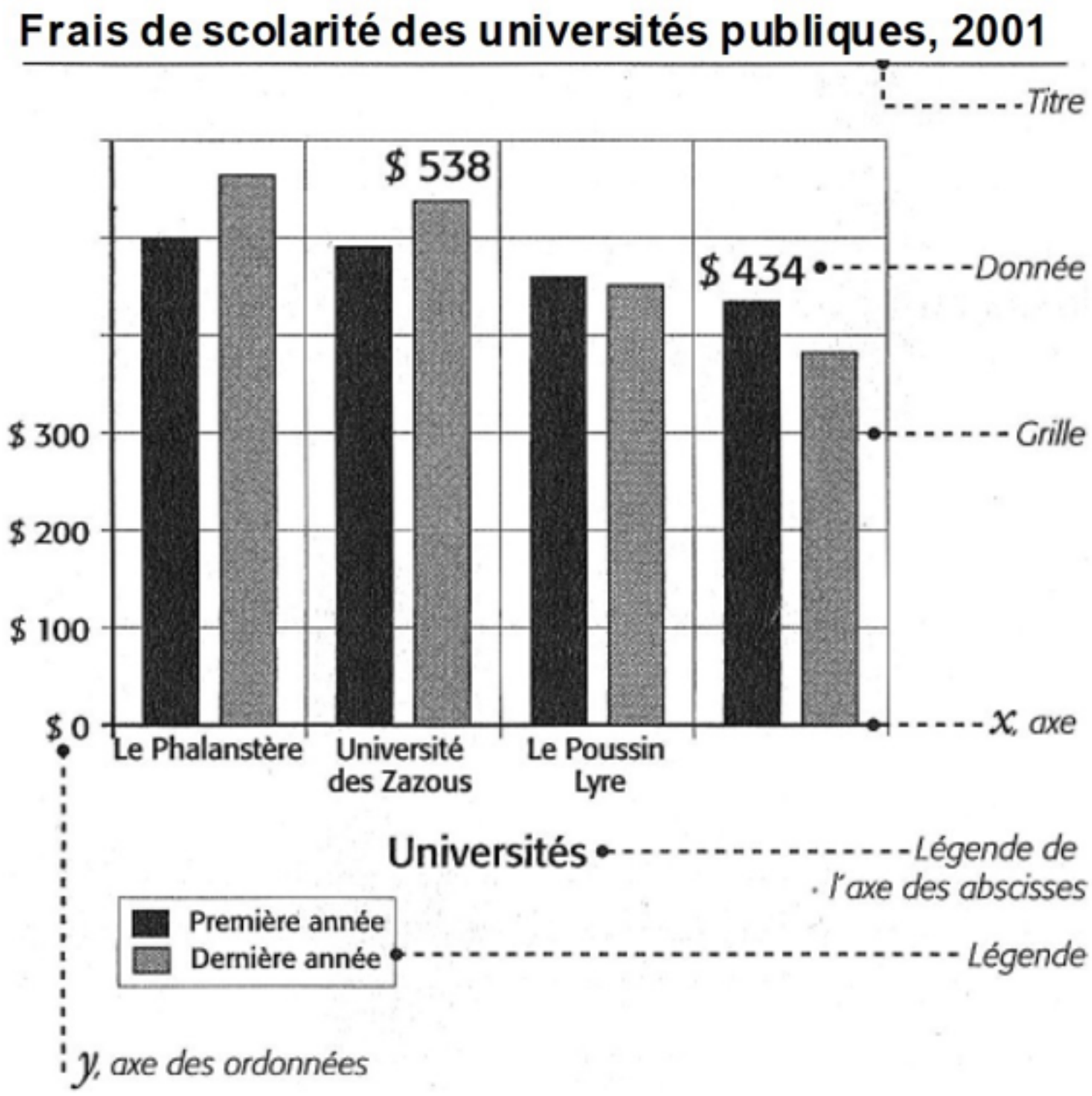




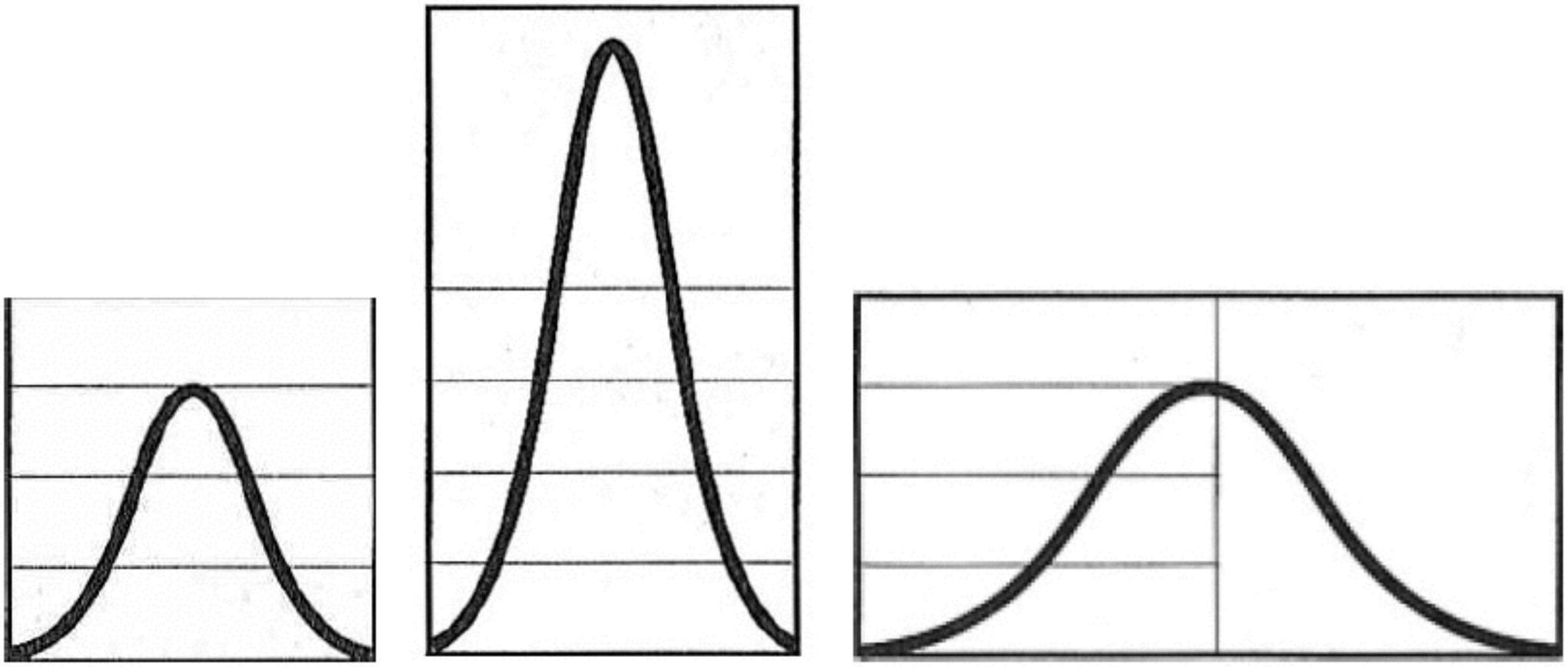




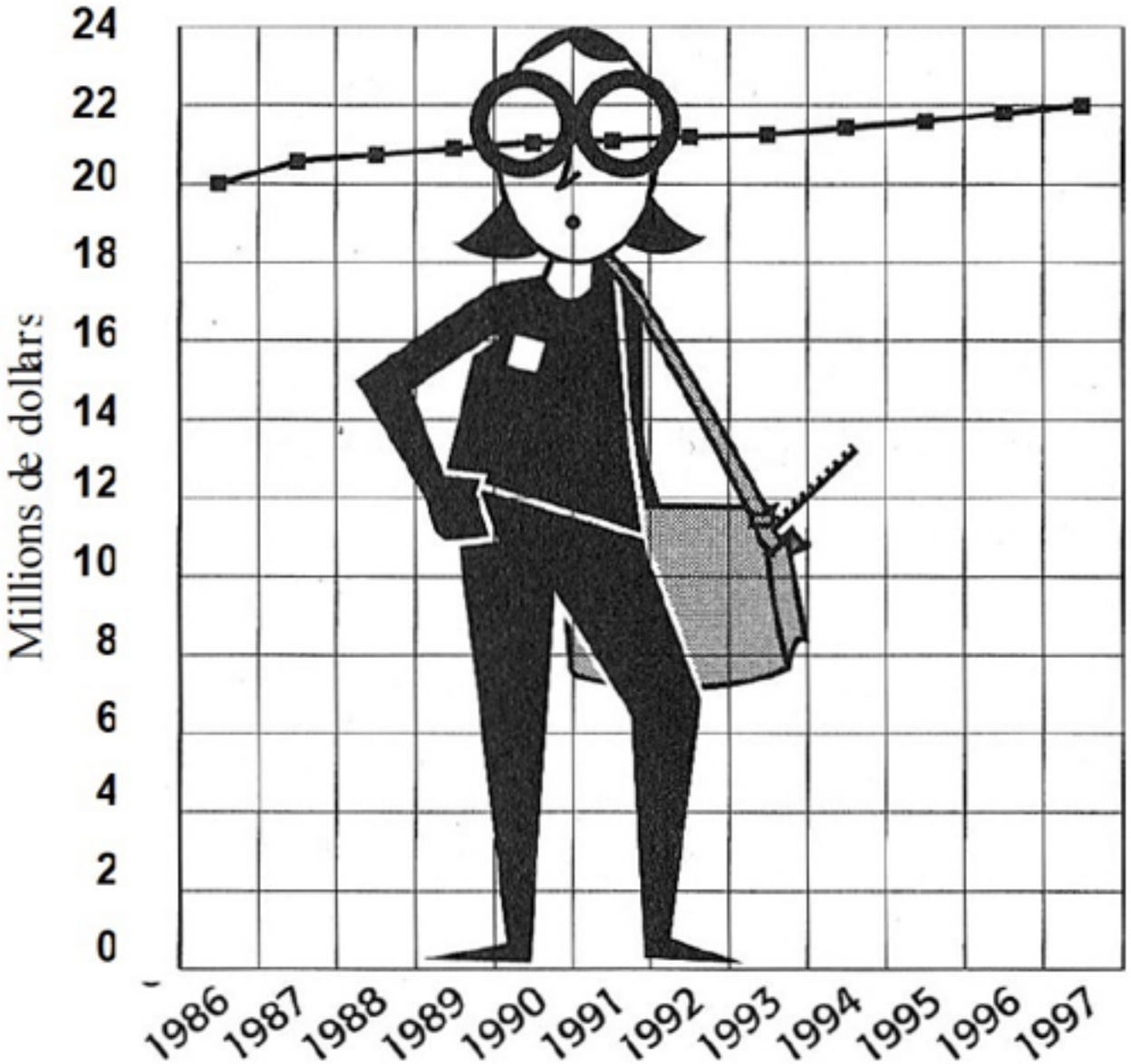
Illustrations et graphiques

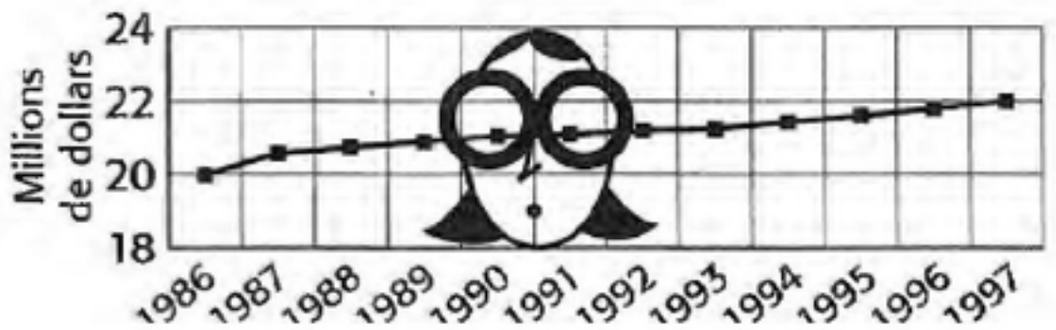


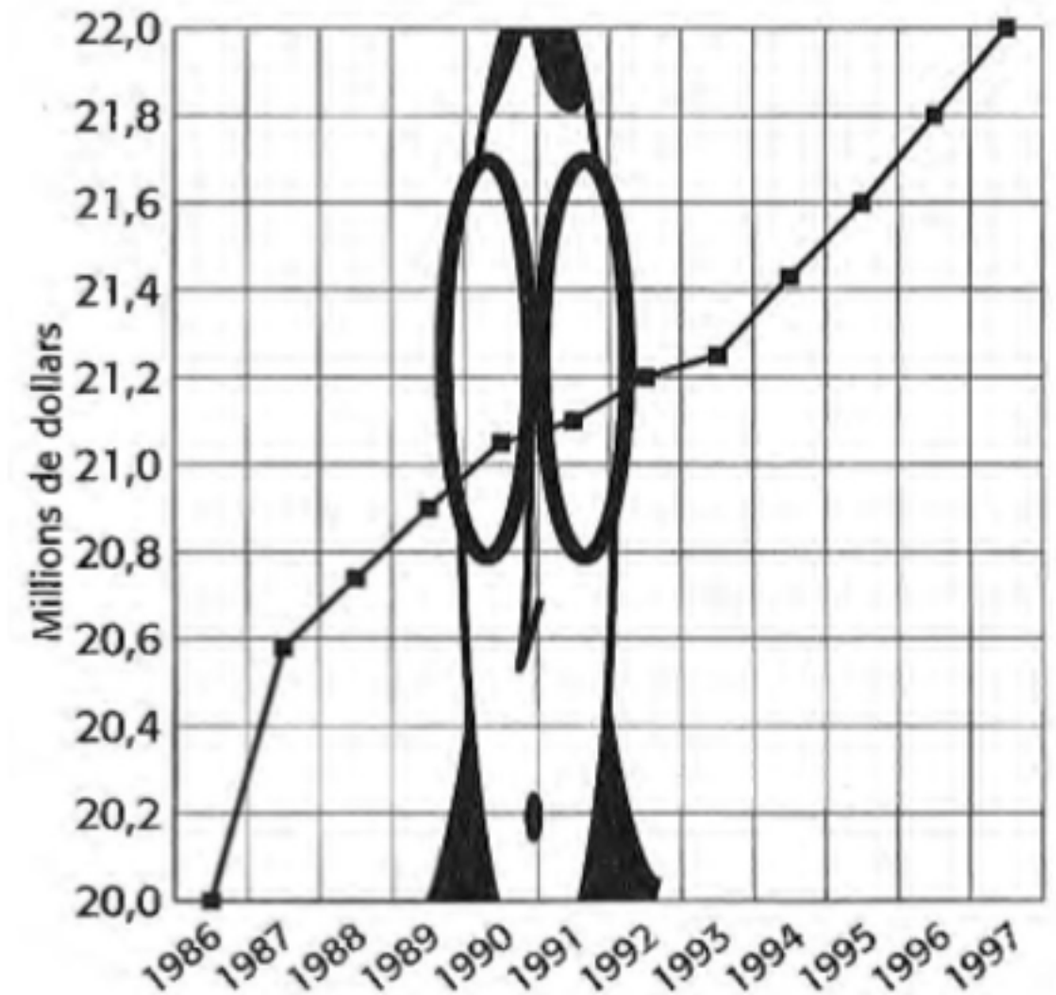
Illustrations et graphiques



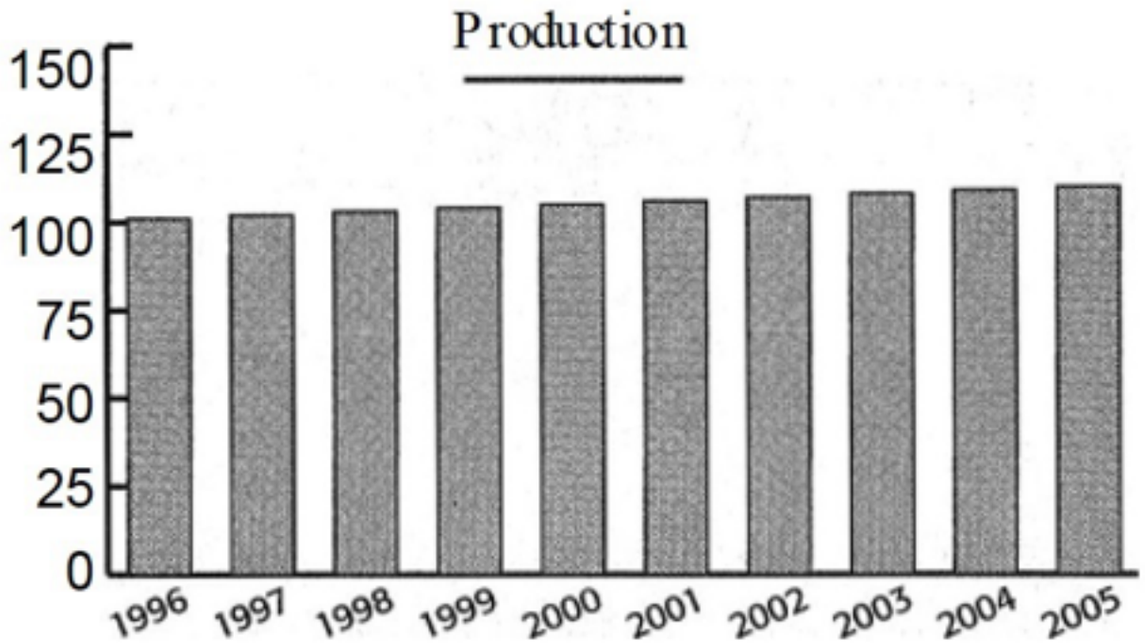
Illustrations et graphiques

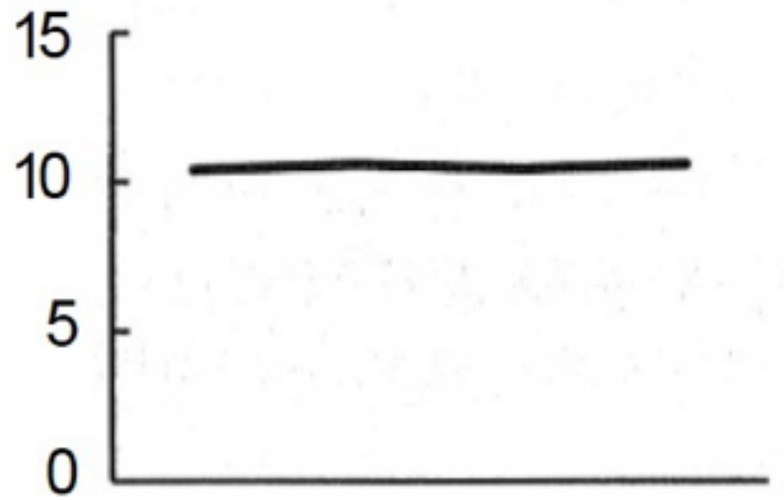


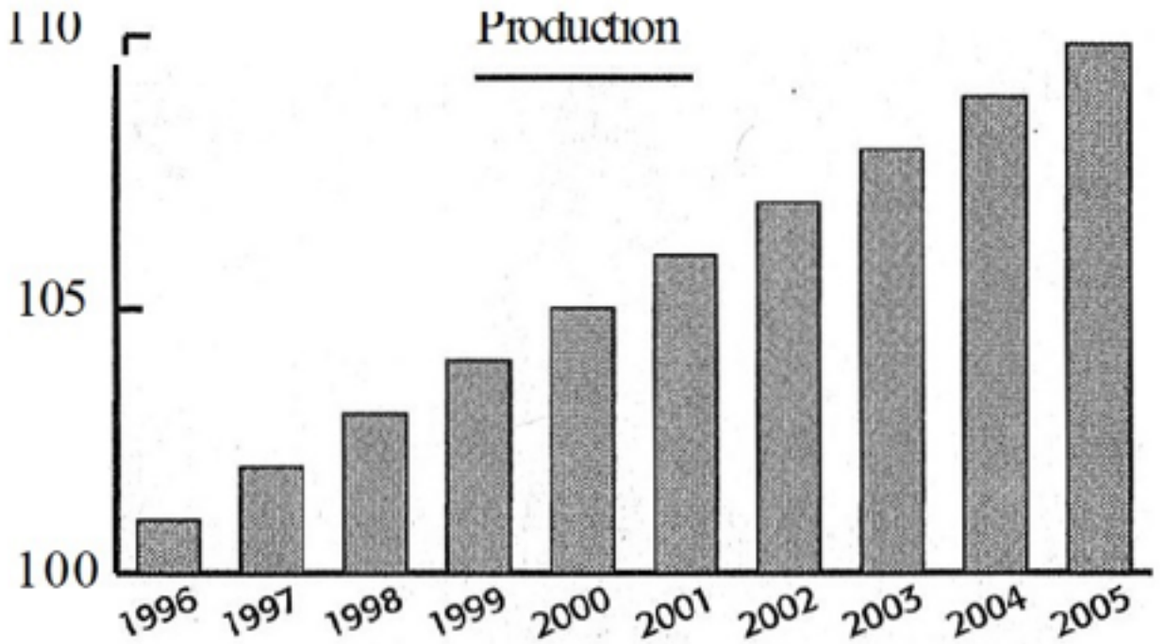


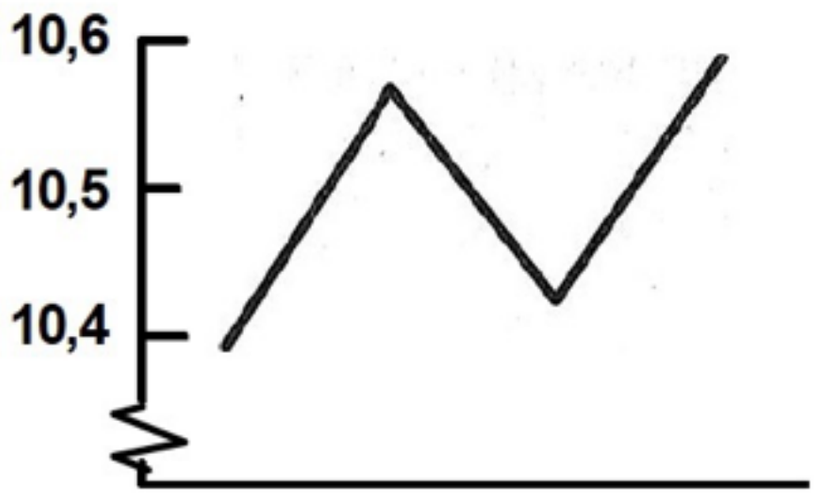


Illustrations et graphiques









L'épreuve de sélection de Wason

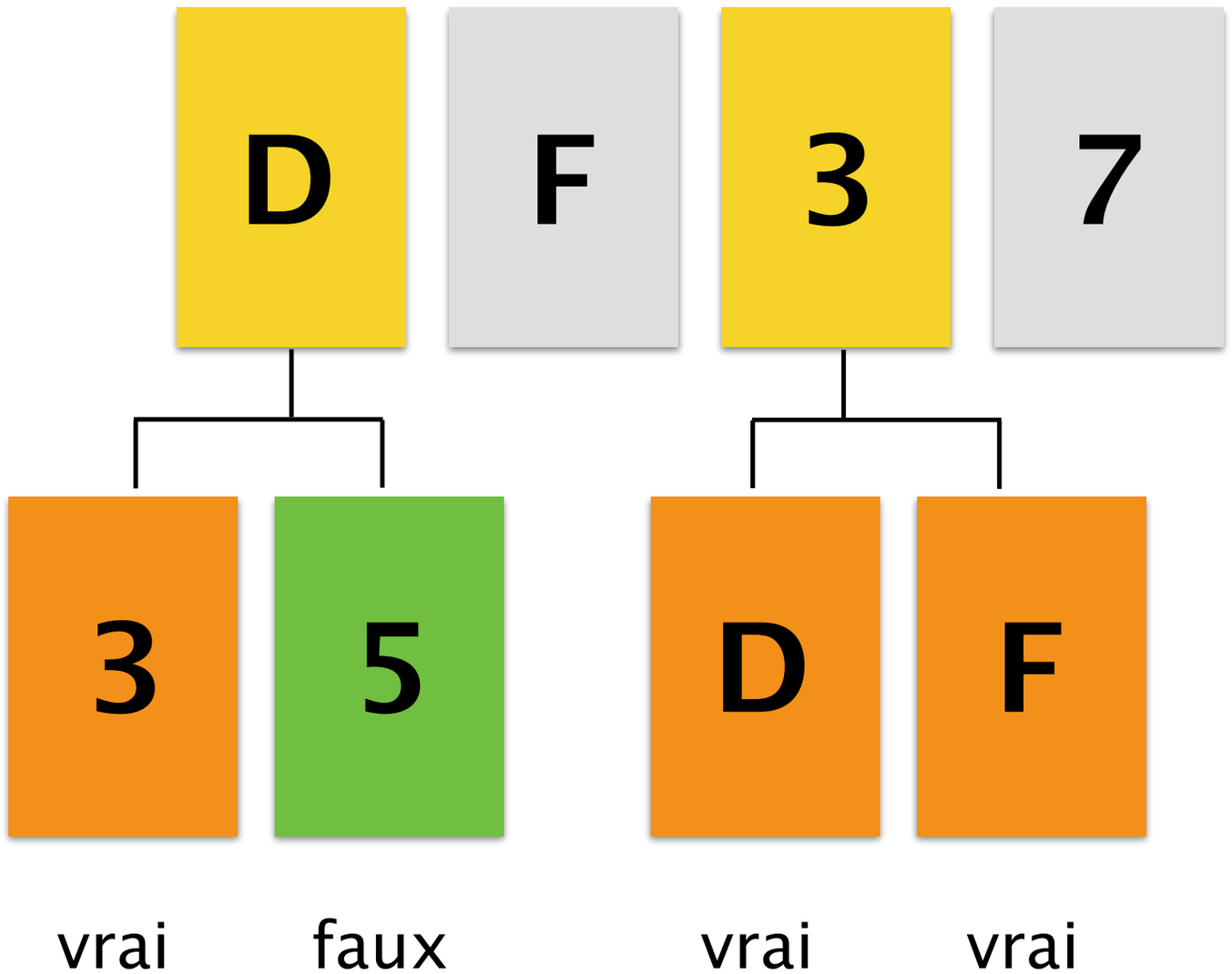
Quatre cartes

Chaque carte présente sur une face une lettre et sur l'autre un chiffre.

Quelles cartes devrez-vous retourner pour vérifier que la règle suivante a été respectée : **si une carte présente un D sur une face, alors elle doit avoir un 3 sur son autre face** ?

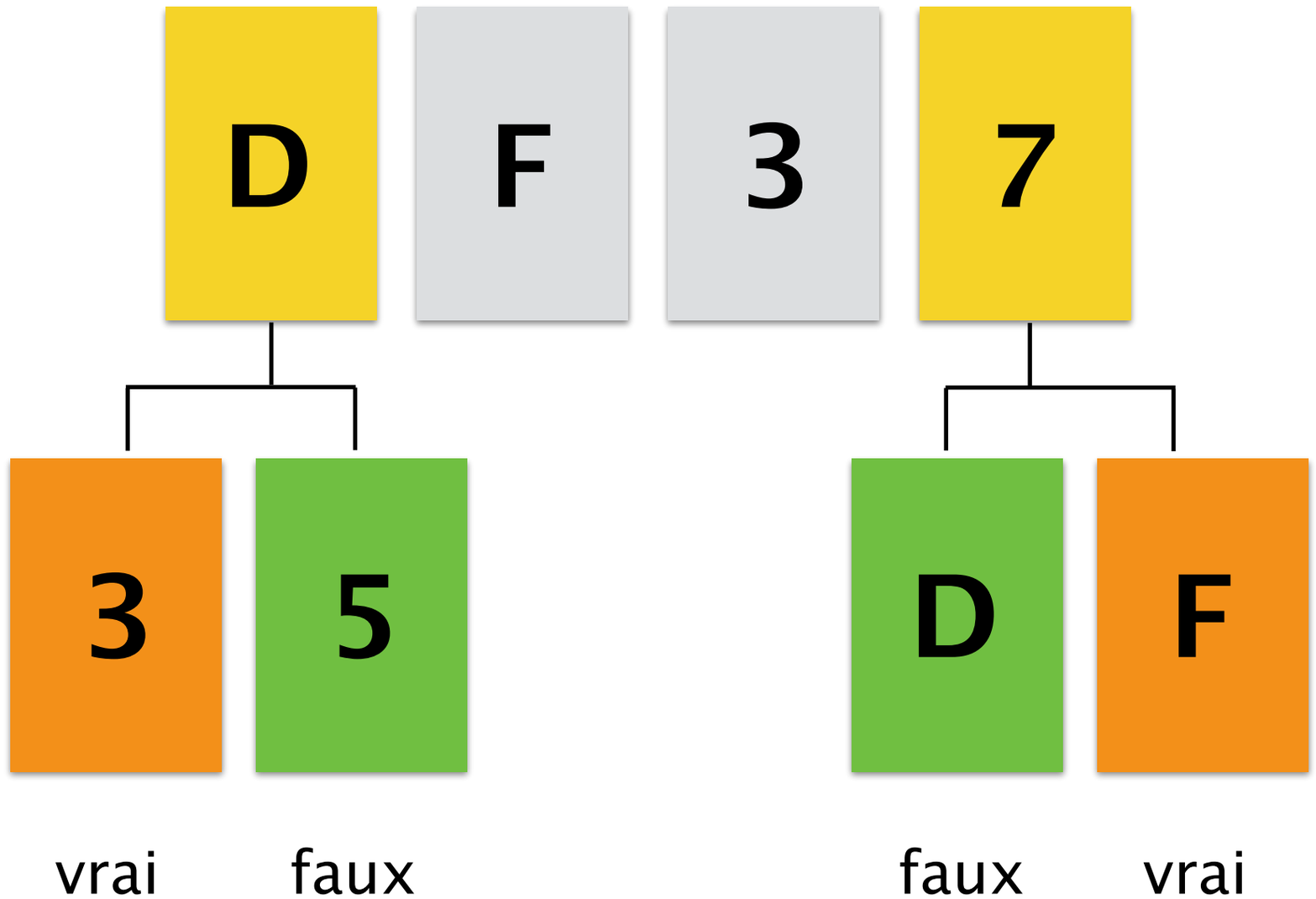
L'épreuve de sélection de Wason

Réponse majoritaire



L'épreuve de sélection de Wason

Réponse correcte



L'épreuve de sélection de Wason

Illustre la tendance à rechercher des exemples qui confirment et à négliger ceux qui infirment une hypothèse.