

Kafémath Virtuel du 30 avril 2020

Statistiques, probabilités, logiques:

Attention!



Moruroa – Mur de la légion 5^e RE.

Incontournables statistiques

Statistiques omniprésentes:

Sondages, élections, sociologie, politique, sports, ...

Grand nombre de données:

A recueillir, classer, traiter : → utilisation d'outils mathématiques

Interprétation:

Problèmes de mauvaises utilisations: biais, surinterprétation,...

La partie est-elle représentative du tout?

Nous ne serons pas exhaustif

Présentation partielle et partielle

Pas big data

Stimuler le sens critique

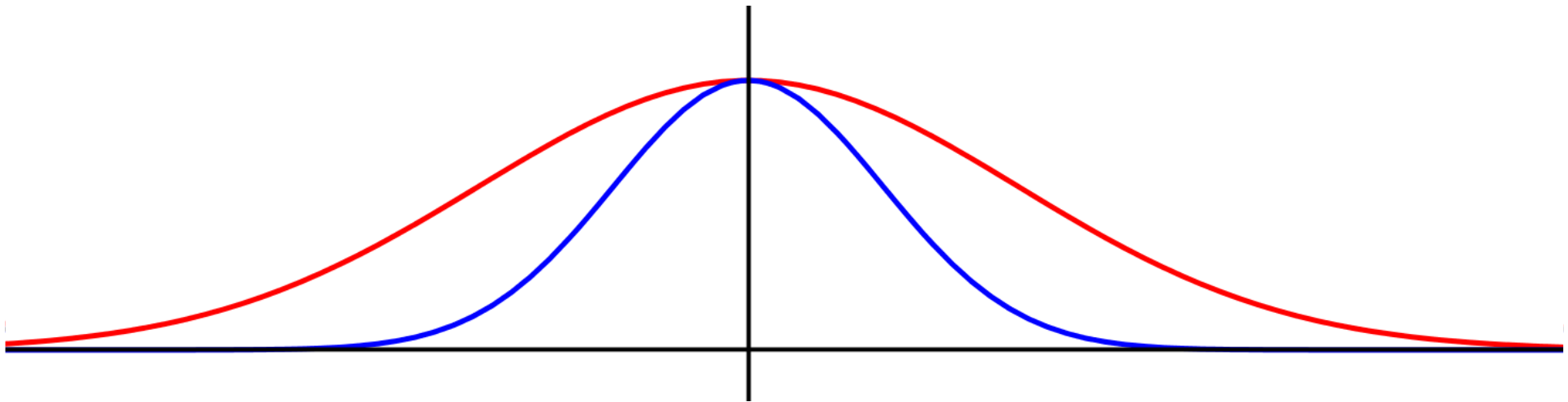
STATISTIQUES (*status*)

La partie est-elle représentative du tout?

- **Statistiques descriptives:** « décrire » des données
(= *réduction d'information*)
- **Inférence statistique :** « induire » les caractéristiques d'une population à partir d'un échantillon.
- **Statistique mathématique:** estimateurs non biaisés (notion de précision).
Repose sur la théorie des probabilités

Notions de bases

Moyenne, écart-type, variance médiane,...



Seuil de pauvreté

Domaines associés aux Statistiques

Probabilités = probs habilité

2 enfants dont Anne (F): probabilité d'avoir une famille *paritaire*?

Anne

F	F
F	G
G	F
G	G

Benoît
2/3

Les Logiques = l'éloge Hic

Vu (et entendu) à la télé:

Masque efficace 100% \Rightarrow utile

Masque non efficace 100% \Rightarrow inutile

Domaines associés aux Statistiques

Probabilités = probs habilité

2 enfants dont Anne (F): probabilité d'avoir une famille *paritaire*?

Anne			Aînée		
F	F	Benoît 2/3	F	F	Benoît 1/2
F	G		F	G	
G	F		G	F	
G	G		G	G	

Les Logiques = l'éloge Hic

Vu (et entendu) à la télé:

Masque efficace 100% \Rightarrow utile

Masque non efficace 100% \Rightarrow inutile

Domaines associés aux Statistiques

Probabilités = probs habilité

2 enfants dont Anne (F): probabilité d'avoir une famille *paritaire*?

Anne	Aînée	Ma fille																								
<table border="1"><tr><td>F</td><td>F</td></tr><tr><td>F</td><td>G</td></tr><tr><td>G</td><td>F</td></tr><tr><td>G</td><td>G</td></tr></table> Benoît 2/3	F	F	F	G	G	F	G	G	<table border="1"><tr><td>F</td><td>F</td></tr><tr><td>F</td><td>G</td></tr><tr><td>G</td><td>F</td></tr><tr><td>G</td><td>G</td></tr></table> Benoît 1/2	F	F	F	G	G	F	G	G	<table border="1"><tr><td>F</td><td>F</td></tr><tr><td>F</td><td>G</td></tr><tr><td>G</td><td>F</td></tr><tr><td>G</td><td>G</td></tr></table> Benoît 1	F	F	F	G	G	F	G	G
F	F																									
F	G																									
G	F																									
G	G																									
F	F																									
F	G																									
G	F																									
G	G																									
F	F																									
F	G																									
G	F																									
G	G																									

Les Logiques = l'éloge Hic

Vu (et entendu) à la télé:

Masque efficace 100% \Rightarrow utile

Masque non efficace 100% \Rightarrow inutile

Vocabulaire

Déduction

Induction

Biais:

cognitifs

d'autocomplaisance

de mesure

corrélation-causalité

Erreur écologique

(population \nRightarrow individu)

Erreur atomiste

(individu \nRightarrow population)

Les convictions sont des ennemis de la vérité plus dangereux que les mensonges.

Humain, trop humain, FN

Chose n'est icy plus commune :

Le bien nous le faisons, le mal c'est la fortune,
On a toujours raison, le destin toujours tort.

L'ingratitude &

l'injustice des hommes envers la Fortune

Jean de la Fontaine

Paradoxes : Simpson, ...



Où sont-elles?

Etude effectuée en juillet 2007 (Europe + U.S.):

Combien avez-vous eu de partenaires sexuels au cours de votre vie, même s'il s'agit de partenaires d'un soir?

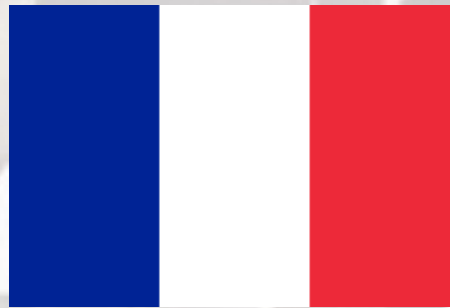
Où sont-elles?

Etude effectuée en juillet 2007 (Europe + U.S.):

Combien avez-vous eu de partenaires sexuels au cours de votre vie, même s'il s'agit de partenaires d'un soir?

N_H = nb de partenaires/H

N_F = nb de partenaires/F



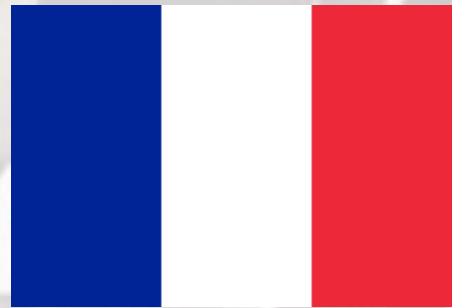
Où sont-elles?

Etude effectuée en juillet 2007 (Europe + U.S.):

Combien avez-vous eu de partenaires sexuels au cours de votre vie, même s'il s'agit de partenaires d'un soir?

N_H = nb de partenaires/H

N_F = nb de partenaires/F



$N_H / N_F = 2,5$

1,6

1,3

> 1

$N_H = 11,5$

13,6

13,8

Où sont-elles?

Poncifs:

Hommes vantards

Femmes mémoire sélective

Biais potentiels:

Femme = Homme

Population (16 < âge < 64),

Flux de populations:

gigolos, gérontophiles,

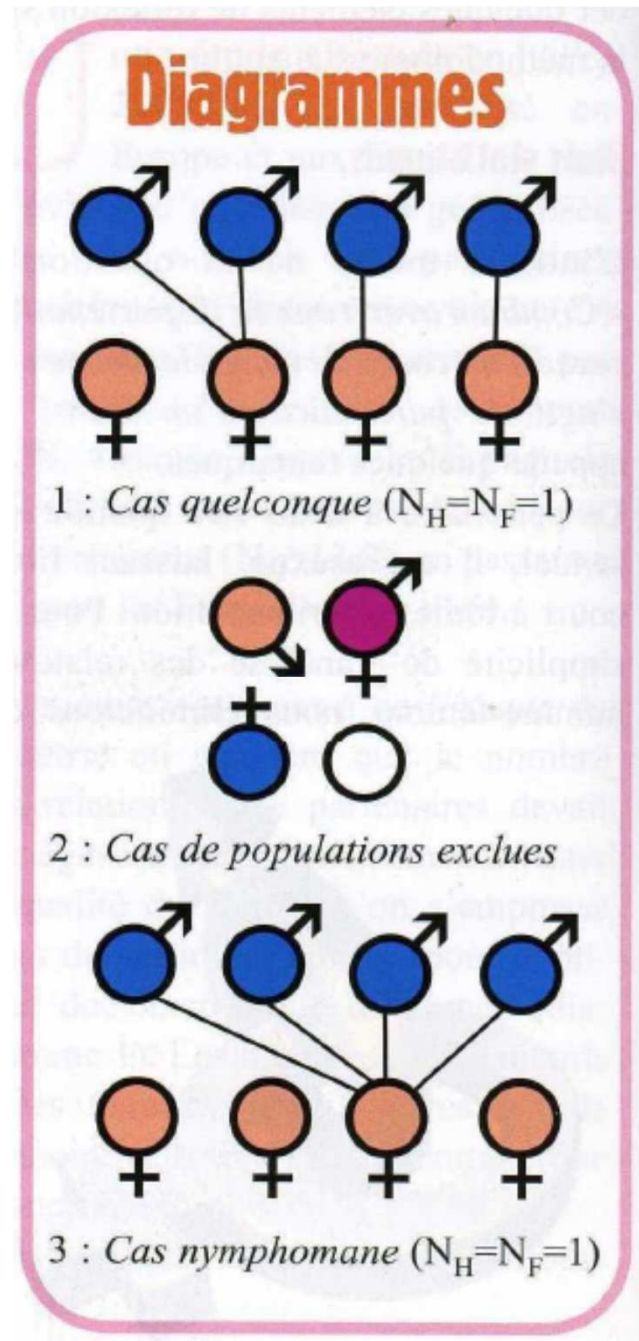
veuves argentées, donc joyeuses,

étudiantes nordiques,

Péripatéticiennes,

Nymphomanes, ...

*Nous sommes souvent abusés par les mythes,
et il est difficile de le constater.*



Quelques paradoxes statistiques

Paradoxe de l'amitié (Scott L. Feld, 1991)

Une majorité d'individus ont, *en moyenne*, moins d'amis que leurs amis.

Quelques paradoxes statistiques

Paradoxe de l'amitié (Scott L. Feld, 1991)

Une majorité d'individus ont, *en moyenne*, moins d'amis que leurs amis.

Paradoxe de Rogers : En déplaçant un élément a de l'ensemble A dans l'ensemble B, on peut faire augmenter les moyennes des deux ensembles. Cd. : $\bar{B} < a < \bar{A}$

Quelques paradoxes statistiques

Paradoxe de l'amitié (Scott L. Feld, 1991)

Une majorité d'individus ont, *en moyenne*, moins d'amis que leurs amis.

Paradoxe de Rogers : En déplaçant un élément a de l'ensemble A dans l'ensemble B, on peut faire augmenter les moyennes des deux ensembles. Cd. : $\bar{B} < a < \bar{A}$

Paradoxe de Braess (1968) : Une extension du réseau routier peut entraîner des temps de trajet plus longs.

Quelques paradoxes statistiques

Paradoxe de l'amitié (Scott L. Feld, 1991)

Une majorité d'individus ont, *en moyenne*, moins d'amis que leurs amis.

Paradoxe de Rogers : En déplaçant un élément a de l'ensemble A dans l'ensemble B, on peut faire augmenter les moyennes des deux ensembles. Cd. : $\bar{B} < a < \bar{A}$

Paradoxe de Braess (1968) : Une extension du réseau routier peut entraîner des temps de trajet plus longs.

Paradoxe de Simpson (1951) – **Yule** (1903):

Agglomération: Union des caractéristiques différente des caractéristiques de l'union.

Paradoxe de Simpson

Agglomération de résultats:

1 classe de 32 filles

1 classe de 32 garçons

3 options: Philo, Maths, Physique

% de réussite	Maths	Philo	Physique	Global
Filles	67	31	86	
Garçons	56	20	72	

Paradoxe de Simpson

Agglomération de résultats:

1 classe de 32 filles

1 classe de 32 garçons

3 options: Philo, Maths, Physique

% de réussite	Maths	Philo	Physique	Global
Filles	67	31	86	56
Garçons	56	20	72	59

Paradoxe de Simpson

Agglomération de résultats:

1 classe de 32 filles

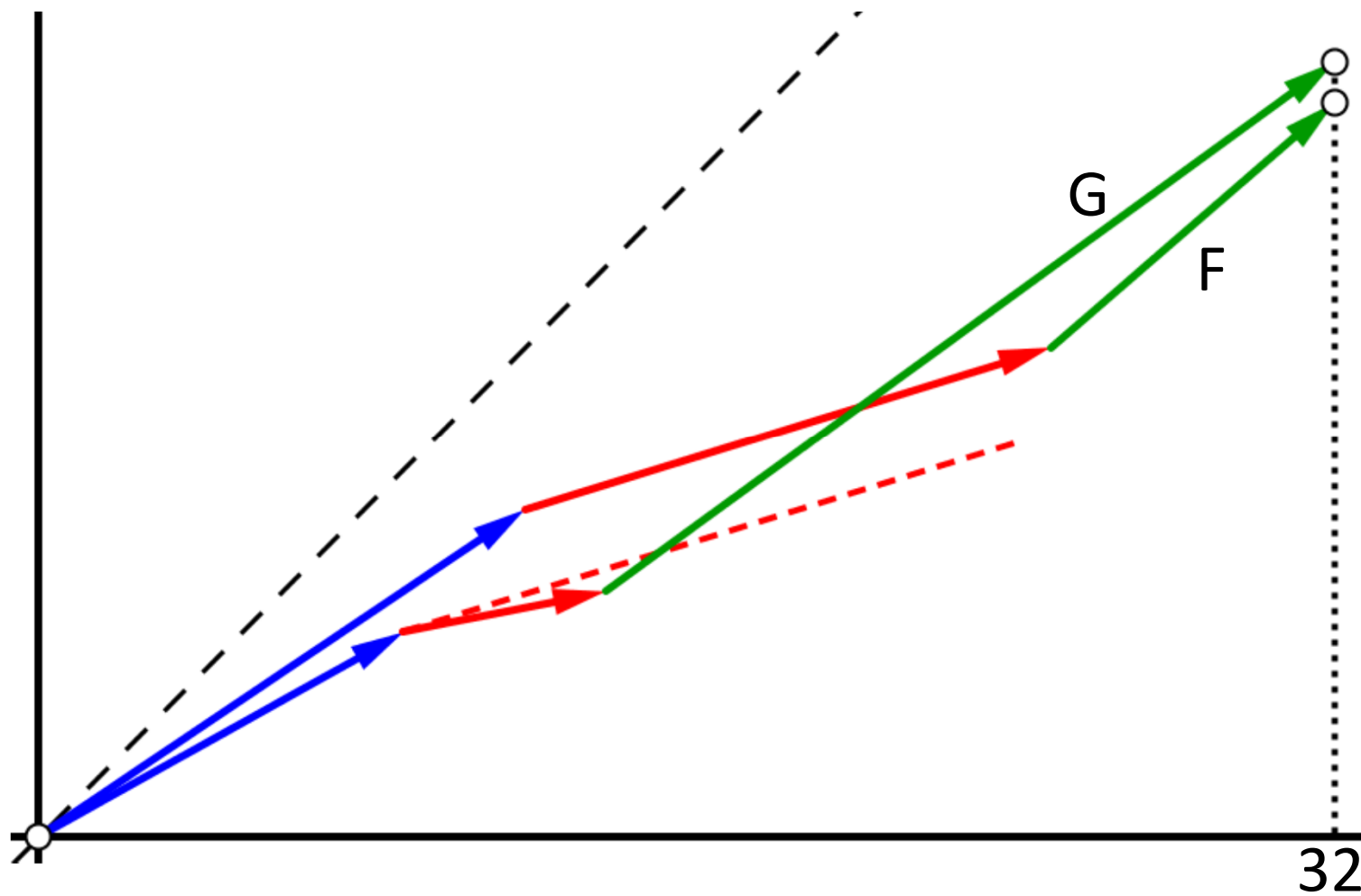
1 classe de 32 garçons

3 options: Philo, Maths, Physique

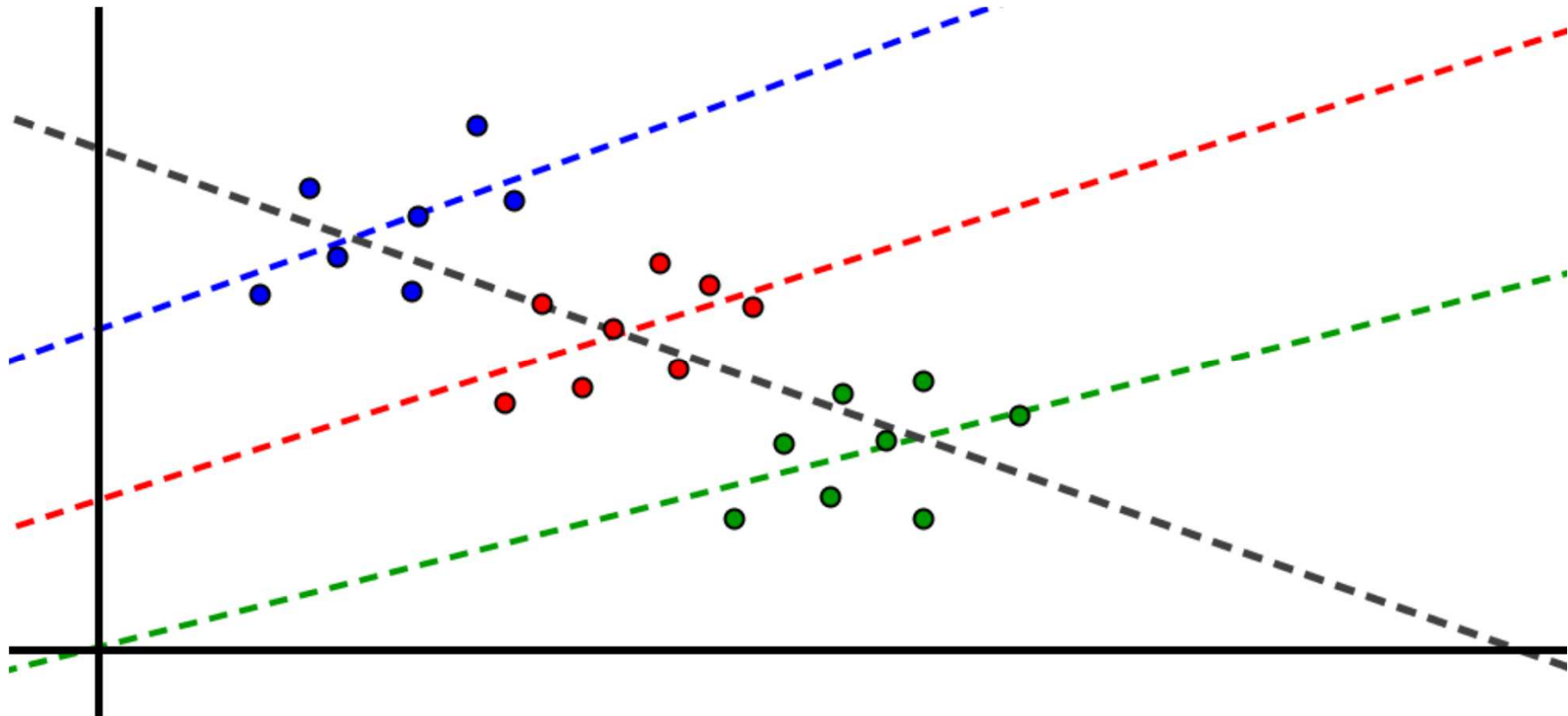
% de réussite	Maths	Philo	Physique	Global
Filles	67	31	86	56
Garçons	56	20	72	59

		Maths	Philo	Physique	Total
Effectif	Filles	12	13	7	32
	Garçons	9	5	18	32
Nb. de réussite	Filles	8	4	6	18
	Garçons	5	1	13	19

Paradoxe de Simpson



Paradoxe de Simpson



2 classes de 100 élèves: 100 G, 100 F

2 matières: Maths (\diamond), Physique (\circ)

Paradoxe de Simpson

Les élèves choisissent une seule matière.

Réussite:

%	F	G
\diamond	20	30
\circ	60	90

Effectifs:

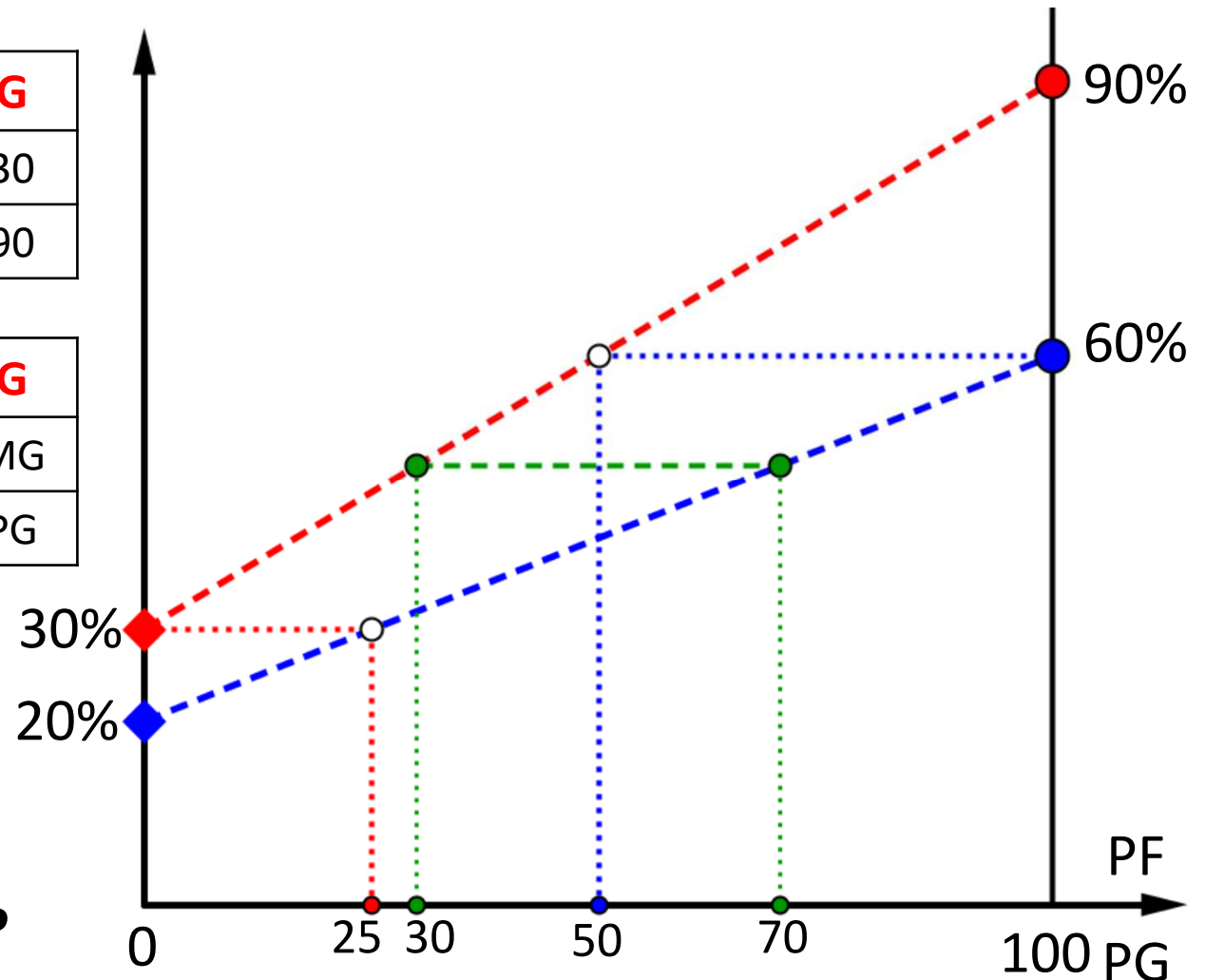
Nb	F	G
\diamond	MF	MG
\circ	PF	PG

Moyenne par classe:

$$30 < G < 90$$

$$20 < F < 60$$

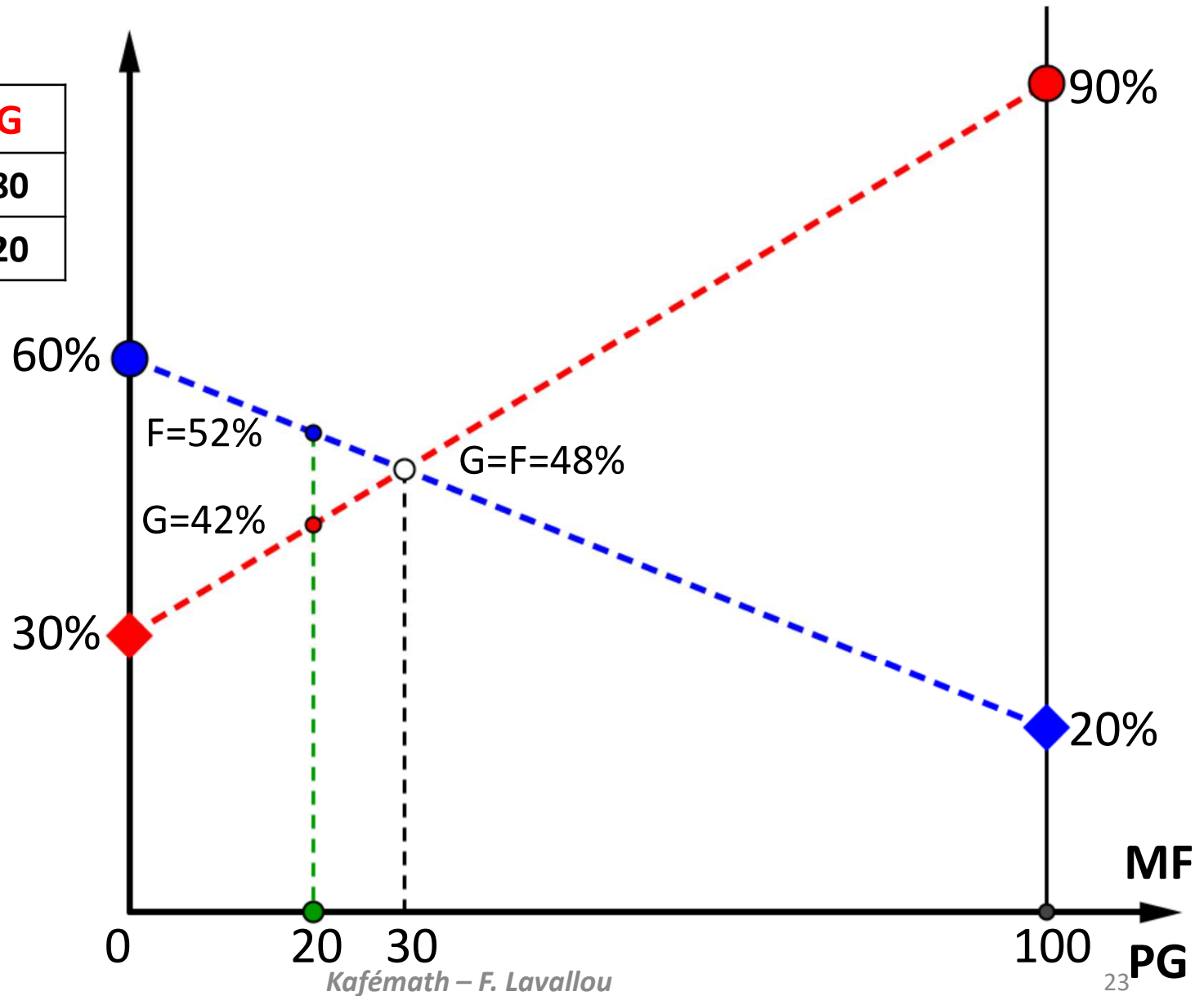
Peut-on avoir $F > G$?



Paradoxe de Simpson

Exemple:

Nb	F	G
◇	20	80
○	80	20



Prises de tests, prise de têtes

Une maladie mortelle touche 1 personne sur 100.

Traitement: 20% en meurent.

Test: fiabilité = 0,95.

Question: ***Faut-il appliquer le traitement à une personne déclarée positive par le test?***

Plupart des médecins: **OUI** (*sous influence du taux de fiabilité*).

Problème:

Probabilité qu'une personne positive soit infectée \neq 95% (fiabilité)

Prises de tests, prise de têtes

Une maladie mortelle touche 1 personne sur 100.

Traitement: 20% en meurent.

Test: fiabilité = 0,95.

Question: ***Faut-il appliquer le traitement à une personne déclarée positive par le test?***

Forte dépendance du taux de base (1%): $95/590 \approx 16\% \ll 95\%$

Personnes:	Malades	Saines	Total:
Test > 0	95	495	590
Test < 0	5	9405	9410
Total:	100	9900	10000

On sauve 76 personnes (/95) et en condamne 99 (/495)

Bilan = - 23 personnes (Traitement : 123 décès, sinon: 100 décès!)

Que les pros se tâtent avant de répondre!

Erreurs judiciaires

Probabilité très faible \Rightarrow tricherie?

En 1999, Sally Clark condamnée pour le meurtre de ses deux fils, à un an d'intervalle.

Les statistiques donnent la probabilité $p = 1/8543$ (!) pour la mort du nourrisson dans un couple aisé non fumeur.

Le pédiatre « expert » donne alors la probabilité $p^2 = 1/73000000$.

Conclusion: 700000 naissances/an \Rightarrow 1 cas/siècle \Rightarrow coupable!

Donc, tout vainqueur du loto ($p = 1/14000000$) a triché!?

Prise en compte du milieu, proba garçon double, proba d'un second décès = 10 fois plus $\Rightarrow 1/170000$

Effectivement: un ou deux cas de mort double chaque année!

Citations:

Mettez un pied dans le four et l'autre dans un seau d'eau glacée. D'après les statisticiens, vous devriez vous sentir très à l'aise en moyenne.

Les statistiques sont vraies quant à la maladie et fausses quant au malade; elles sont vraies quant aux populations et fausses quant à l'individu.

Il y a trois sortes de mensonges: les mensonges, les sacrés mensonges et les statistiques.

Les statistiques montrent que les personnes qui fêtent le plus d'anniversaire deviennent les plus vieilles.

L'être humain moyen a un testicule et un sein.

Le loto est un impôt sur les gens qui ne comprennent pas les statistiques.

Les statistiques, c'est comme le bikini. Ce qu'elles révèlent est suggestif. Ce qu'elles dissimulent est essentiel.

Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales.

“La démocratie, ce curieux abus de la statistique.” Jose Luis Borges