

# Peut-on (vraiment) faire des statistiques sur les données AT-MP ?

Journée d'information du CISME  
consacrée aux Thésaurus

11 avril 2013 – Institut Léonard de Vinci – Paris La Défense

# Que faire de données a priori non discriminantes ?

PRO COD	LIBELLE	045A	046A	047A	049A	049B	051A	052A	053A	055A	057A	059A
7245	MONTEURS ET REPARATEURS DE LIGNES	0	0	1	0	0	0	0	0	0	156	0
7311	MECANICIENS REPARATEURS D'INSTRUM	0	0	0	0	0	0	0	0	0	96	0
7312	FACTEURS ET ACCORDEURS D'INSTRUME	0	0	0	0	0	0	0	0	0	40	0
7313	JOAILLIERS ET ORFEVRES	0	0	0	0	1	0	0	0	0	74	0
7321	POTIERS ET ASSIMILES (PRODUITS CERA	0	0	0	0	0	0	0	0	0	58	0
7322	SOUFFLEURS MOULEURS TAILLEURS ME	0	0	0	0	0	0	0	0	0	72	0
7323	GRAVEURS SUR VERRE	0	0	0	0	0	0	0	0	0	20	0
7324	PEINTRES DECORATEURS SUR VERRE E	0	0	0	0	0	0	0	0	0	56	0
7331	OUVRIERS DES METIERS D'ARTISANAT SI	0	0	7	0	0	0	0	0	0	70	0
7332	OUVRIERS DES METIERS D'ARTISANAT TE	0	0	0	0	0	0	0	0	0	82	0
7341	COMPOSITEURS TYPOGRAPHES ET ASSI	0	0	0	0	0	0	0	0	0	128	0
7342	STEREOTYPEURS ET CLICHEURS GALVA	0	0	0	0	0	0	0	0	0	7	0
7343	GRAVEURS D'IMPRIMERIE ET PHOTOGRA	0	0	0	0	0	0	0	0	0	39	0
7344	PERSONNEL DE LA PHOTOGRAPHIE ET A	0	0	0	0	0	0	0	0	0	22	0
7345	RELIEURS ET ASSIMILES	0	0	0	0	0	1	0	0	0	95	0
7346	IMPRIMEURS SERIGRAPHES	0	0	0	0	0	0	0	0	0	110	0
7411	BOUCHERS POISSONNIERS ET ASSIMILES	0	4	0	1	0	0	0	0	0	5381	0
7412	BOULANGERS PATISSIERS ET CONFISEU	0	0	0	1	0	0	0	0	0	699	0
7413	OUVRIERS DE LA FABRICATION DES PRO	0	0	0	0	0	0	0	0	0	199	0
7414	OUVRIERS DE LA CONSERVERIE DE FRUI	0	0	0	0	0	0	0	0	0	85	0
7415	DEGUSTATEURS,CLASSEURS DE DENREI	0	0	0	0	0	0	0	1	0	21	0
7416	OUVRIERS PREPA.TABAC,ET,FABRICATION	0	0	0	0	0	0	0	0	0	4	0
7421	OUVRIERS DU TRAITEMENT DU BOIS	0	0	4	0	0	0	0	0	0	82	0
7422	EBENISTES MENUISIERS ET ASSIMILES	0	0	76	1	0	1	0	0	0	420	0
7423	REGLEURS ET REGLEURS-CONDUCTEURS	0	0	1	0	0	0	0	0	0	82	0
7424	VANNIERS BROSSIERS ET ASSIMILES	0	0	0	0	0	0	0	0	0	26	0

Peut-on faire des statistiques sur les données AT-MP

2013-088-CNAMTS-DRP-MSTAT-PJ

# Plan de la présentation

Peut-on faire des statistiques sur les données AT-MP ?

Autrement-dit, cela vaut-il la peine d'enregistrer des données en fonction des nomenclatures assises sur des thésaurus détaillés ?

1. ☹️

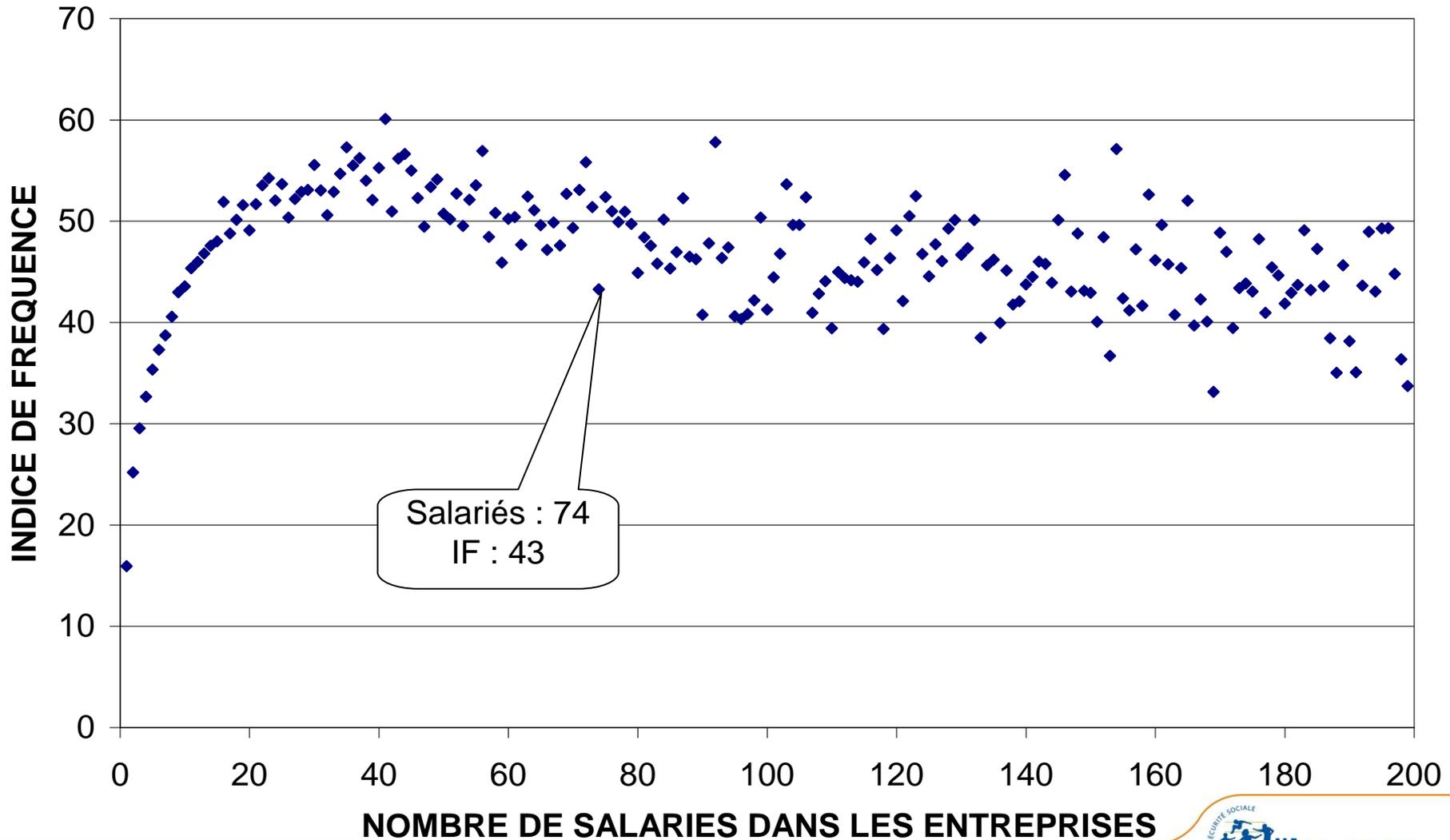
2. 😊

3. **Conclusions**

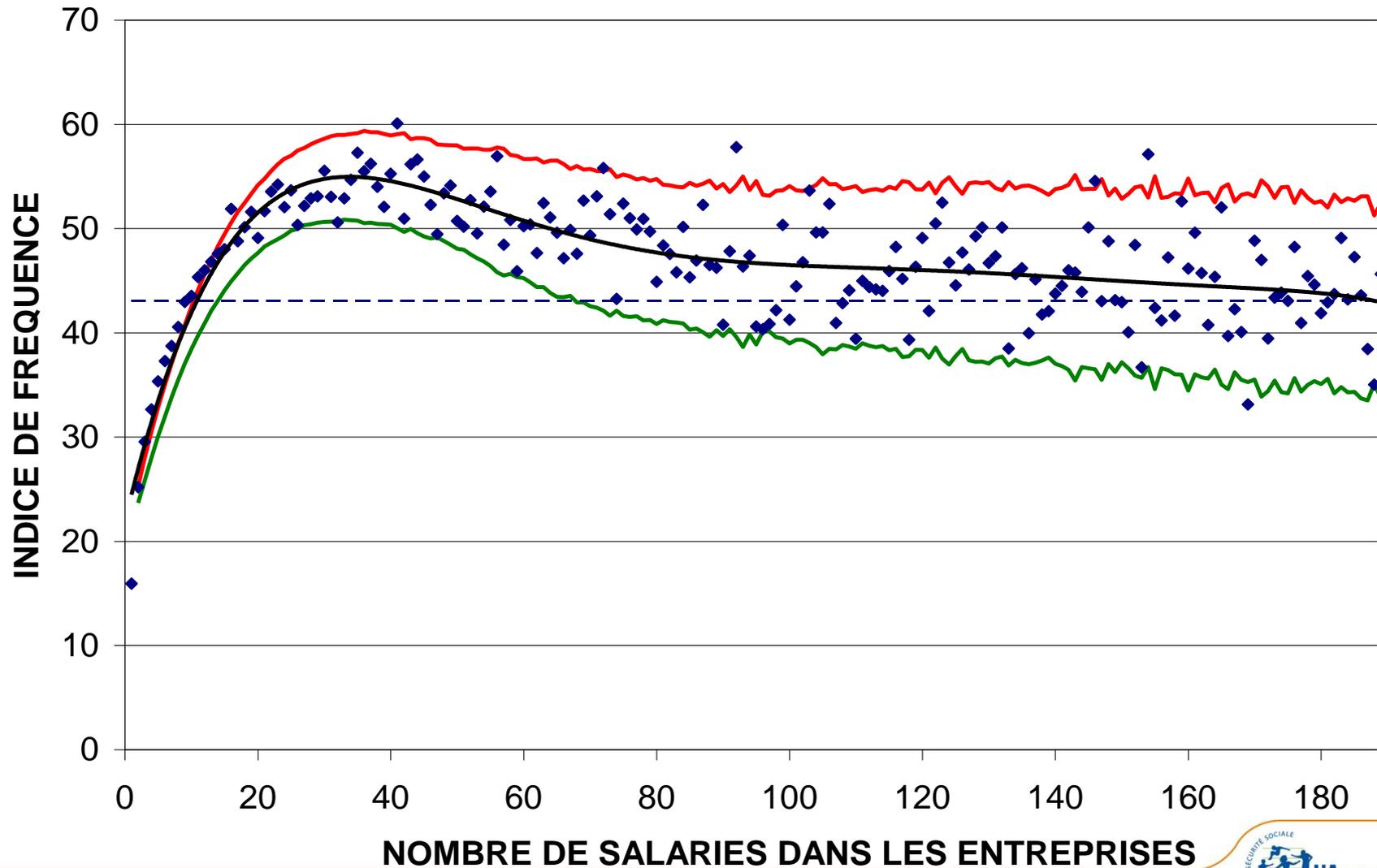


1 - ☹️

# L'exemple des indices de fréquence (IF) des accidents du travail en fonction des effectifs



# Au-delà de 20 salariés, le constat d'une certaine dispersion autour d'une valeur moyenne

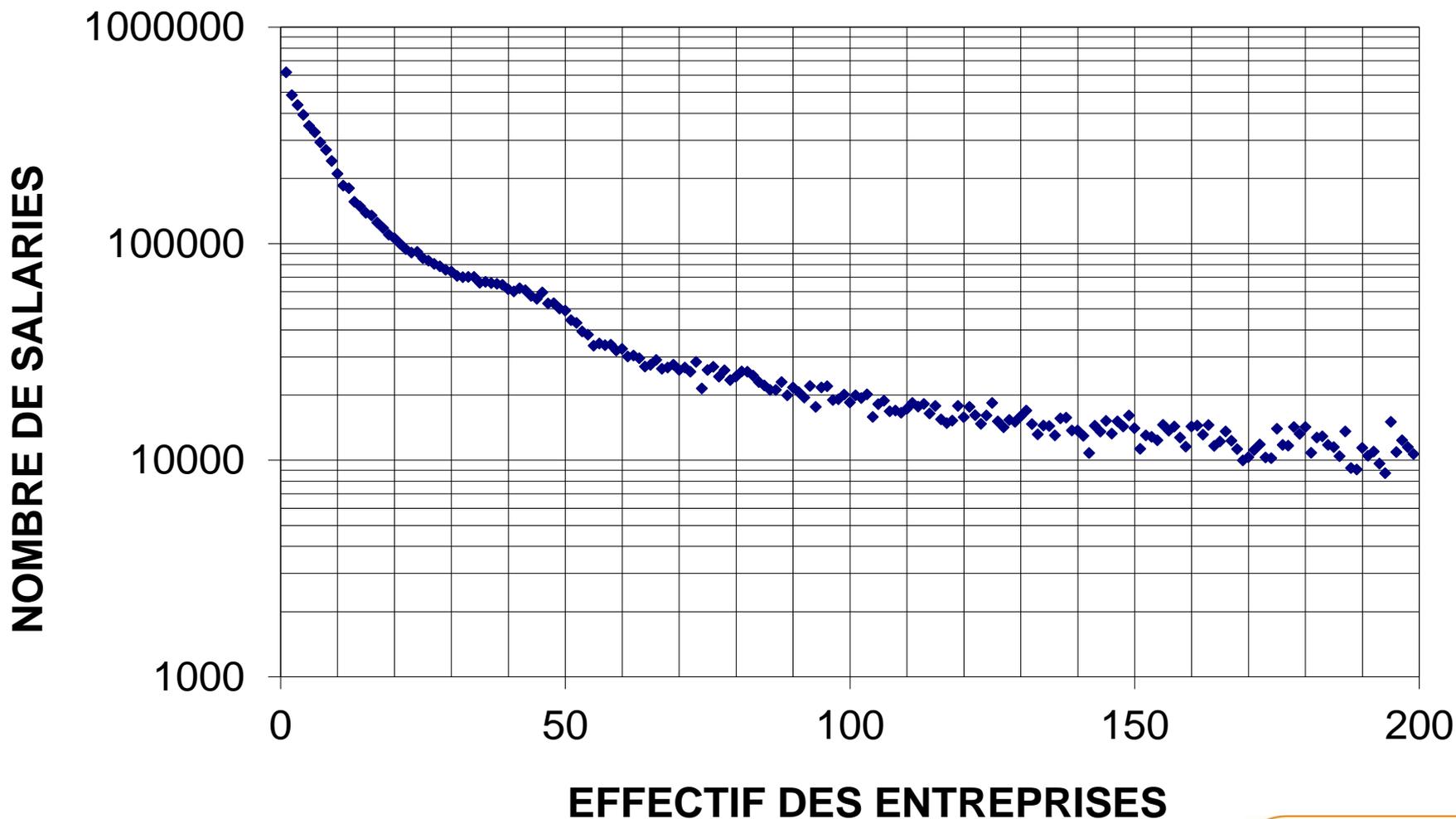


NOMBRE DE SALARIÉS DANS LES ENTREPRISES

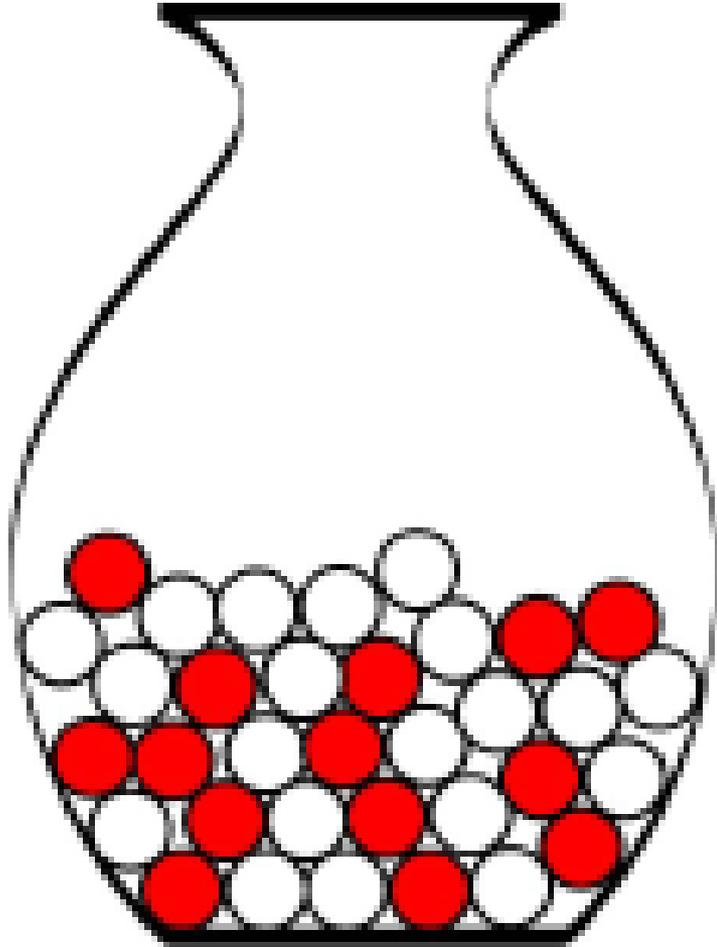
Peut-on faire des statistiques sur les données AT-MP

2013-088-CNAMTS-DRP-MSTAT-PJ

# L'explication de la dispersion par le nombre de salariés sur lequel est calculé l'indice de fréquence

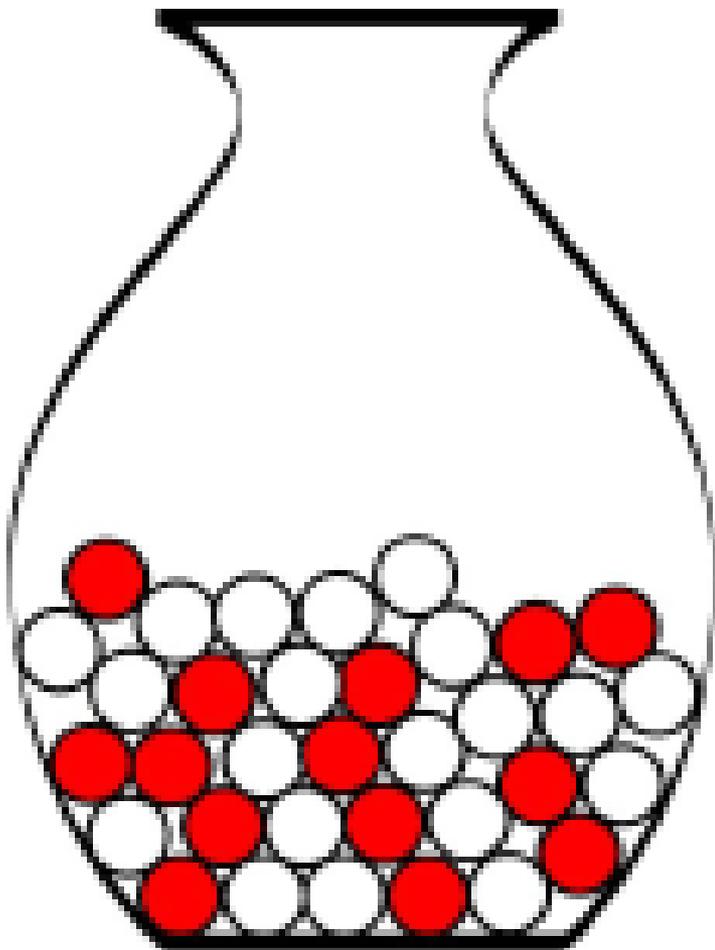


# Pour comprendre, un petit rappel du cours de probabilité



- On place dans une urne 1000 boules :
  - 960 blanches,
  - et 40 rouges (c-à-d 4% ou 40‰).
- Si on tire une boule 1000 fois de suite (en remettant et en mélangeant),
  - on ne va pas récupérer 40 boules rouges exactement,
  - mais (probablement) entre 30 et 54 boules rouges,
- ... alors que la main qui tire est bien « exposée » à 40 boules rouges.

# Suite du rappel du cours de probabilité



## ❑ Toujours pour 40 boules parmi 1000 (soit 4%), on récupère :

- ❑ **si on tire 10 boules :**  
**de 0 à 3 boules rouges,**  
c'est-à-dire entre 0 et 30% du tirage  
ou encore entre 4%-4% et 4%+26% ;
- ❑ **si on tire 100 boules :**  
**entre 2 et 10 boules rouges,**  
c'est -à-dire entre 2% et 10% du tirage  
ou encore entre 4%-2% et 4%+6% ;
- ❑ **si on tire 1000 boules :**  
**entre 30 et 54 boules rouges**  
c'est-à-dire entre 3% et 5.4% du tirage  
ou encore entre 4%-1% et 4%+1.4% ;
- ❑ **si on tire 10 000 boules :**  
**entre 364 et 440 boules rouges,**  
c'est-à-dire entre 3.6% et 4.4% du tirage  
ou encore entre 4%-0.4% et 4%+0.4% ;
- ❑ **si on tire 100 000 boules :**  
**entre 3880 et 4123 boules rouges,**  
c'est-à-dire entre 3.8% et 4.1% du tirage  
ou encore entre 4-0.2% et 4+0.2%.

# Fin du rappel du cours de probabilité

Les intervalles « de confiance » du transparent précédent peuvent être calculés par la formule de l'intervalle de confiance pour une proportion  $p$  inconnue

$$f - u_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} < p < f + u_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}$$

**f**  
fréquence  
d'apparition du  
phénomène  
dans  
l'échantillon

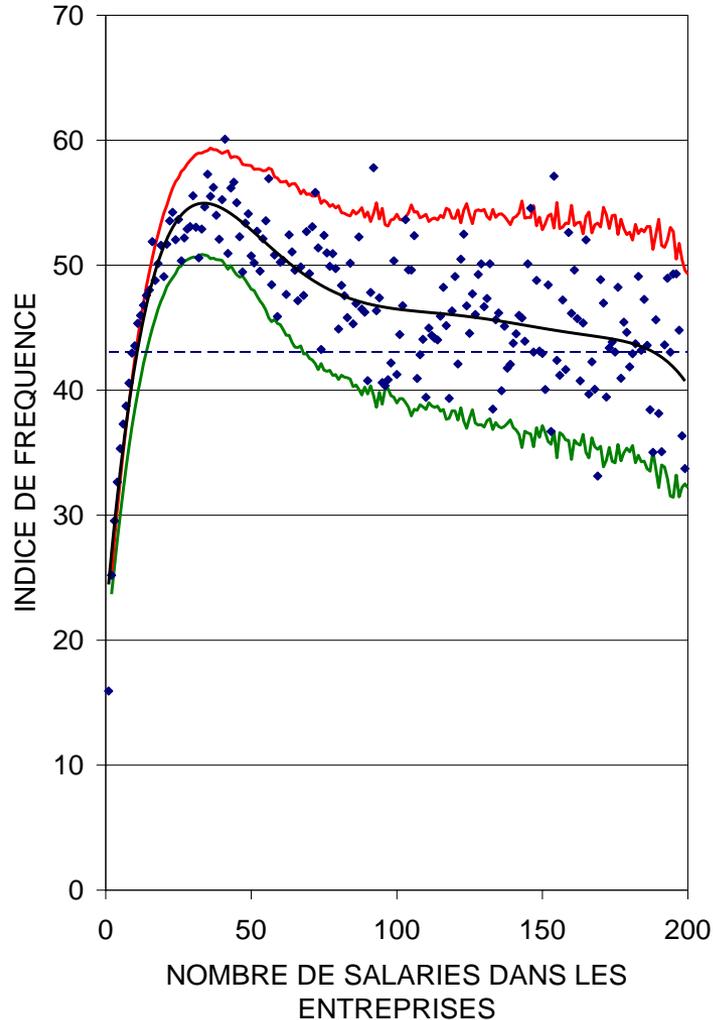
**n**  
taille  
de  
l'échantillon

**p**  
probabilité  
du  
phénomène  
(inconnue)

**u**  
paramètre  
dépendant  
du niveau de  
confiance  
souhaité

Confiance	u
95%	1.96
97.5%	2.24
99%	2.58
99.9%	3.29

# Le concept de probabilité d'accident



- **Quand on applique LA formule en remplaçant :**
  - la proportion de boules rouges dans l'urne par l'indice de fréquence moyen
  - et le nombre de boules tirées par le nombre de salariés sur lequel l'indice de fréquence est calculé

**on obtient les courbes (vertes et rouges) qui encadrent le nuage de points.**

- **Tout se passe comme si les salariés sur lesquels l'indice de fréquence est calculé n'étaient qu'un échantillon qui permet d'approcher un risque mesurable par une probabilité d'accident.**

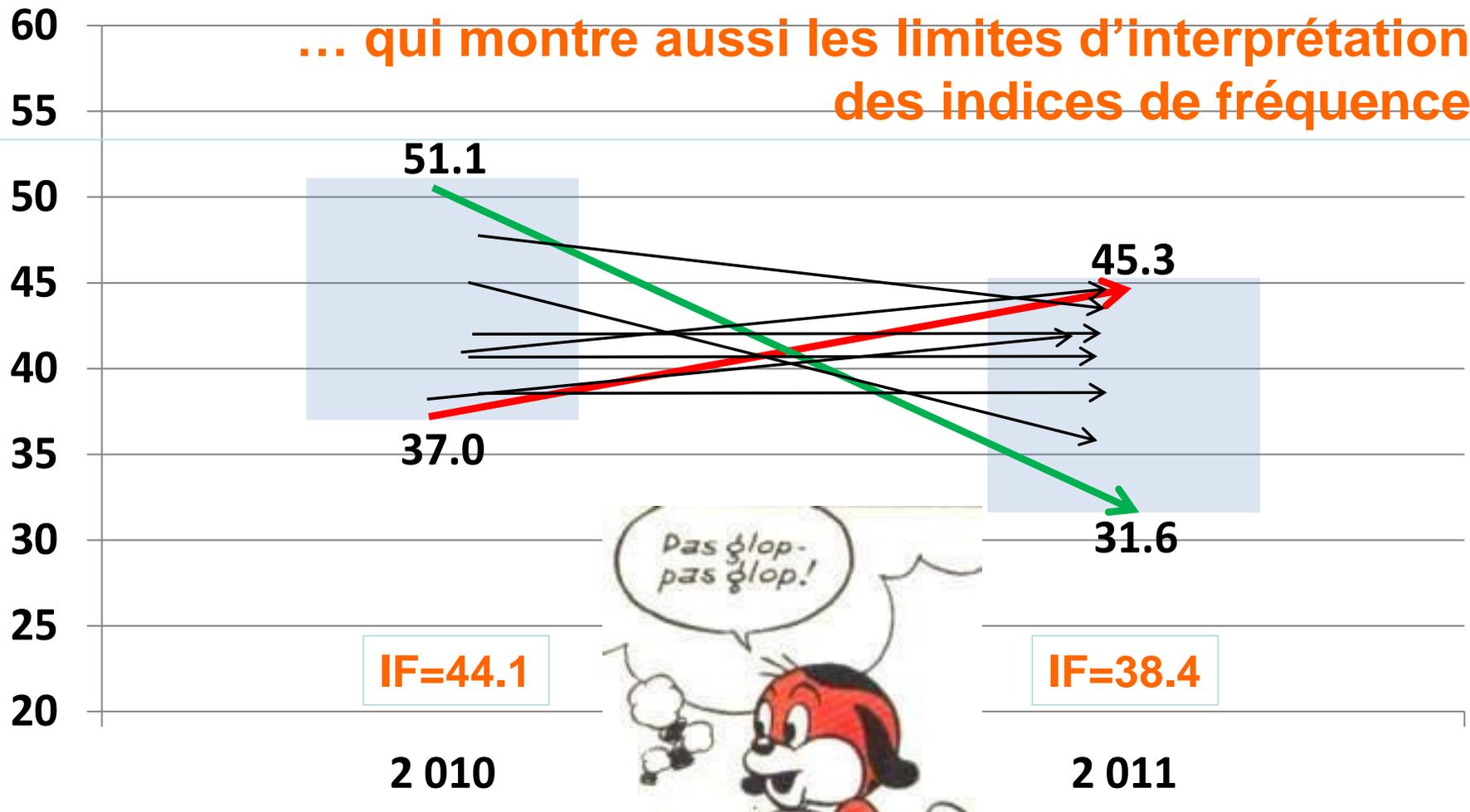
# Conséquence :

## pour une probabilité d'accident de 40 ‰

Nombre de salariés de l'activité	Nombre d'AT annuel min	Indice de fréquence min	Nombre d'AT annuel max	Indice de fréquence max
10	0	0	3	300
100	2	20	10	100
1 000	30	30	54	54
10 000	364	36	440	44
100 000	3880	39	4123	41

# La probabilité d'accident quasiment inaccessible une année donnée...

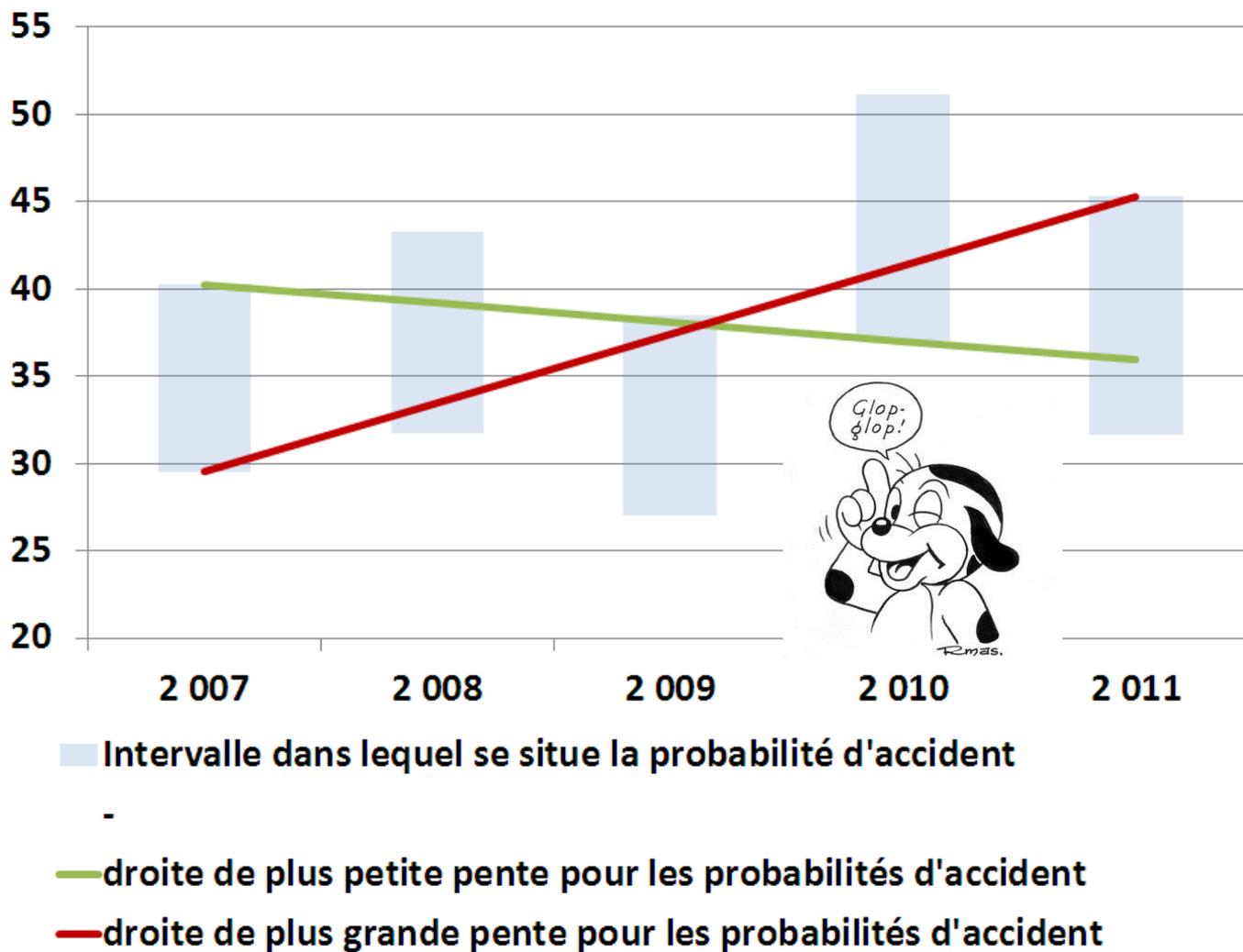
... qui montre aussi les limites d'interprétation des indices de fréquence





2 - ☺

# ... mais qui peut devenir statistiquement accessible sur une plus longue période



- Dans ce cas, les IF des deux dernières années sont en trompe-l'œil.
- 90% des droites sont à la hausse (contre 10% à la baisse)
- La moyenne des pentes des droites à la hausse est 5.1% / an.



# 3 - Conclusions

# Conclusion n°1



Source : the new-yorker  
(Jean-Loup Chiflet, edition les Arènes, l'humour au bureau)

*Et je persiste à dire que ce n'est qu'une théorie.*

# Conclusion n°2

**Dès lors qu'on dispose de données sur la durée,  
calées sur des nomenclatures bien faites,  
tout reste statistiquement possible.**



**D'où l'intérêt d'accumuler les données  
sur le nomenclatures / Thésaurus Harmonisés  
et de se donner les moyens et la peine  
de constituer les bases correspondantes.**