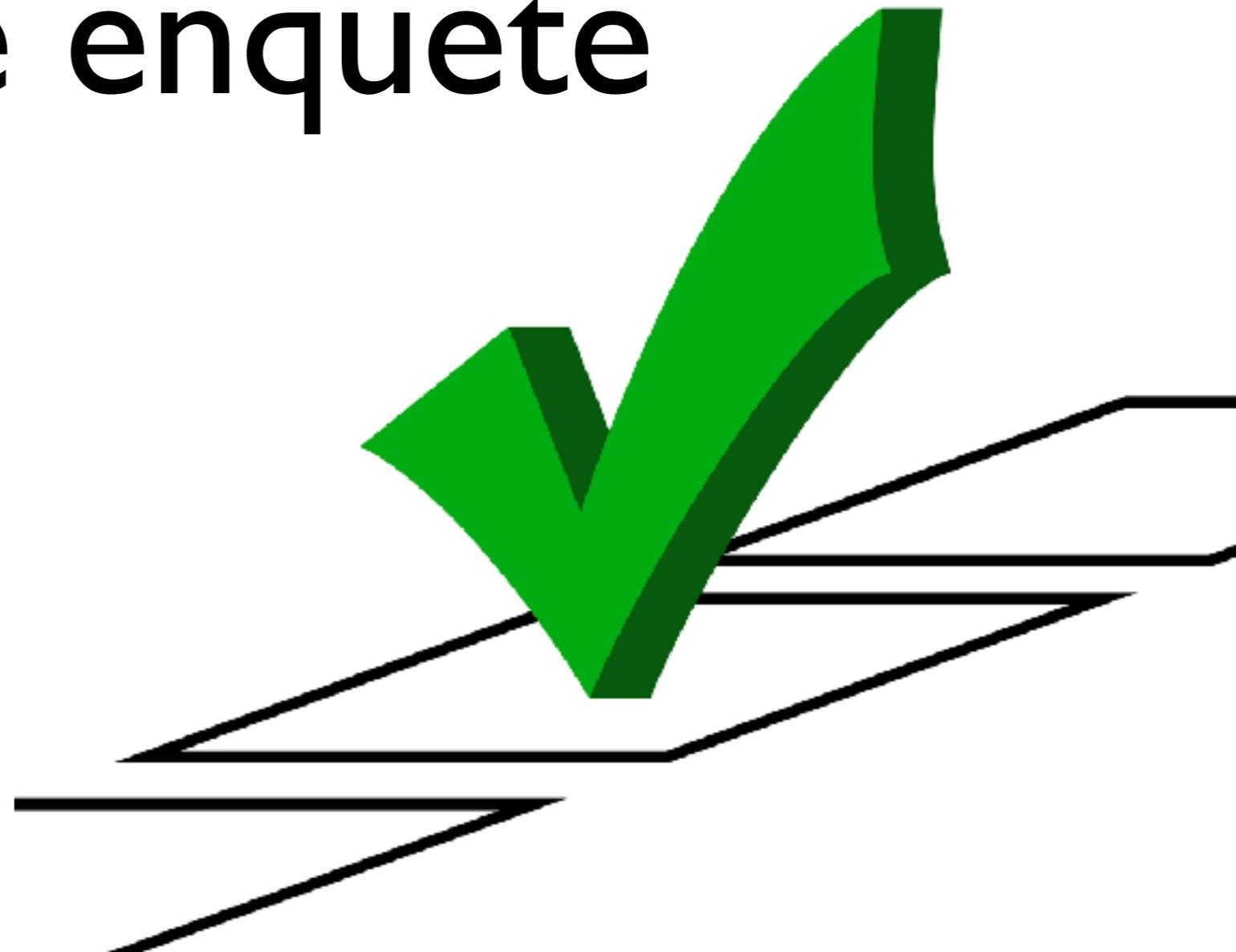


Analyse des Résultats d'une enquête



(Rappels de statistiques)

- **Médiane**

une **médiane** d'un ensemble de valeurs est une valeur m qui permet de couper l'ensemble des valeurs en deux parties égales : mettant d'un côté une moitié des valeurs, qui sont toutes inférieures ou égales à m et de l'autre côté l'autre moitié des valeurs, qui sont toutes supérieures ou égales à m (s'il y a un nombre impair de valeurs, la valeur centrale sera mise des deux côtés)

- **Moyenne**

la **moyenne arithmétique** d'une série statistique est la somme des valeurs divisée par le nombre de valeurs.

Sur une variable quantitative X

- moyenne

$$\bar{X} = \mu_X = \frac{1}{n} \sum_{i=1}^n X_i$$

- variance

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

- écart-type

$$\sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Sur un échantillon de taille n ($n > 1$)

- moyenne

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- variance

$$\sigma_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- écart-type

$$\sigma_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Sur deux variables quantitatives X et Y

- covariance

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y)$$

- coefficient de corrélation

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}$$

$$-1 \leq \rho \leq 1$$

- X et Y indépendantes $\Rightarrow \rho = 0$
- si X et Y corrélées $\Rightarrow \rho > 0$ ou $\rho < 0$
- test du coefficient de corrélation avec un test de Student

- régression linéaire $y = a.x + b$
- estimation des coefficients

$$b = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

$$b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

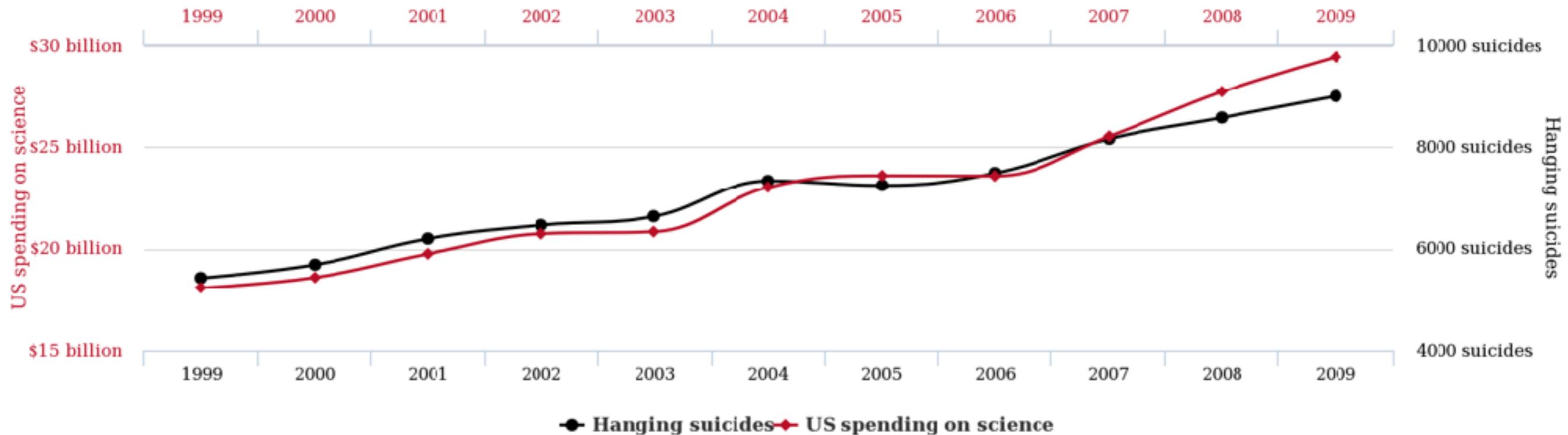
$$a = \bar{Y} - b\bar{X}$$

Les corrélations fallacieuses

- Attention de ne pas corréler des données qui n'ont aucun rapport entre elles !!
- La corrélation n'implique pas la causalité
- Attention donc à l'exploitation du « big data »
- Attention aux échelles qui faussent les comparaisons
- Eviter de comparer des pommes et de oranges

Les corrélations fallacieuses

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation

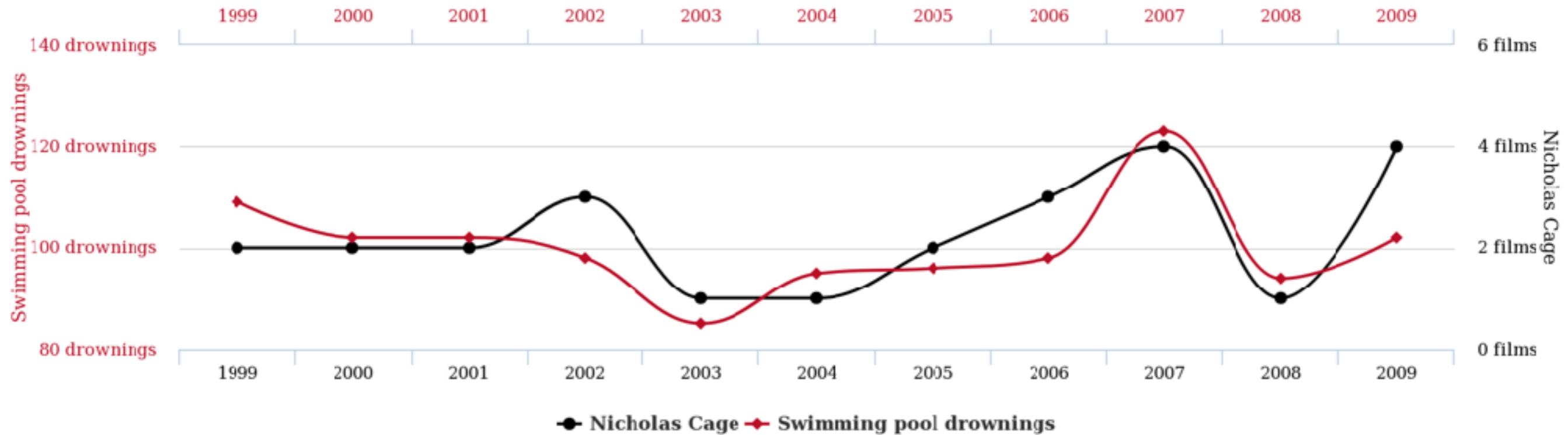


tylervigen.com

99,79 %

Les corrélations fallacieuses

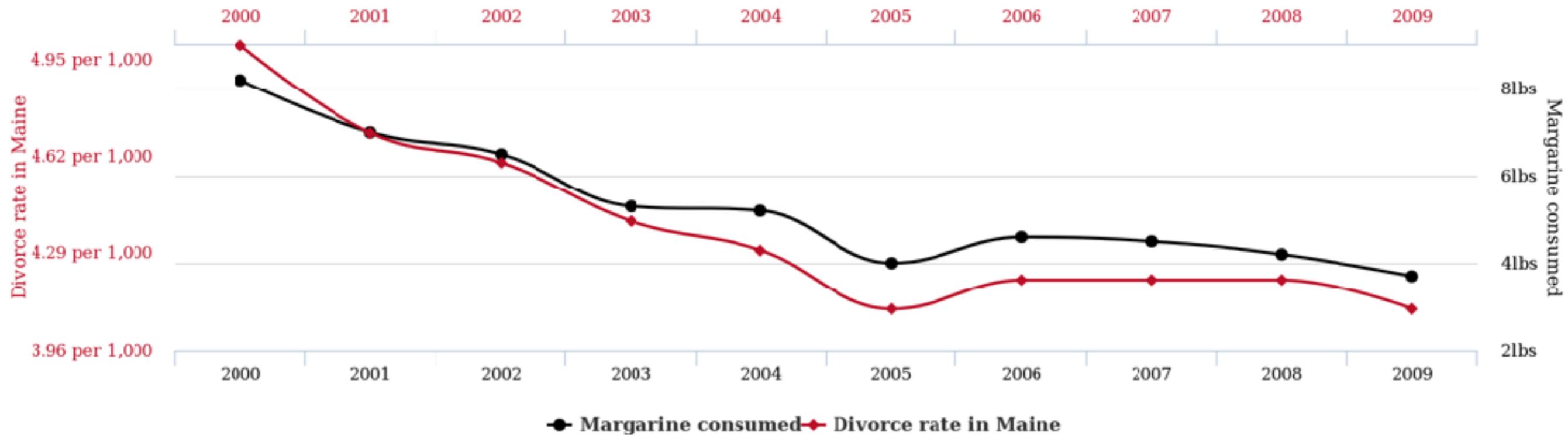
Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in



tylervigen.com

Les corrélations fallacieuses

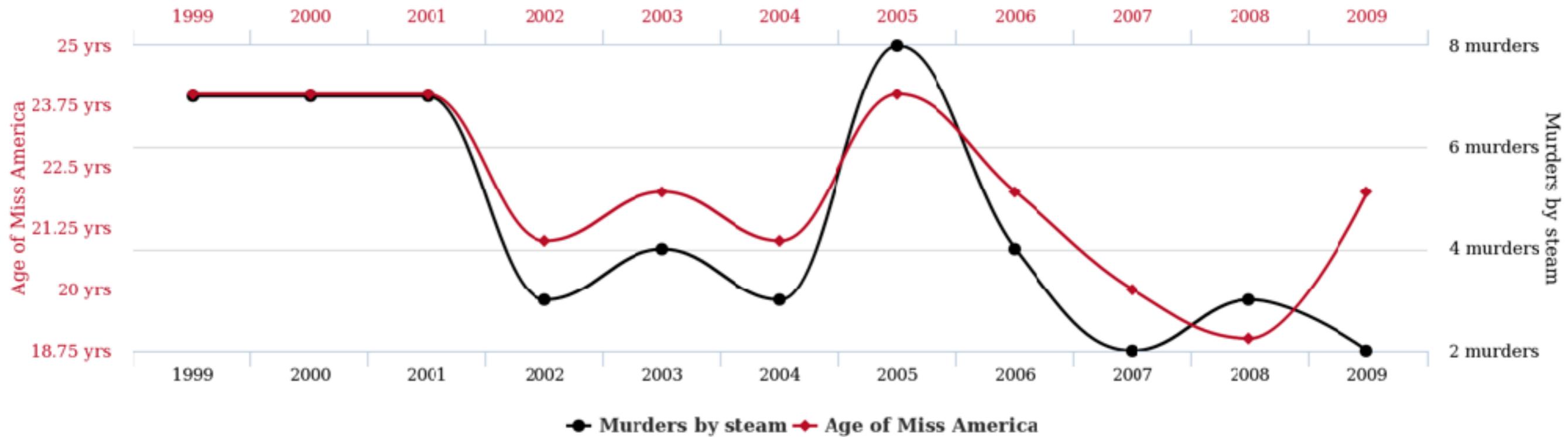
Divorce rate in Maine correlates with Per capita consumption of margarine



tylervigen.com

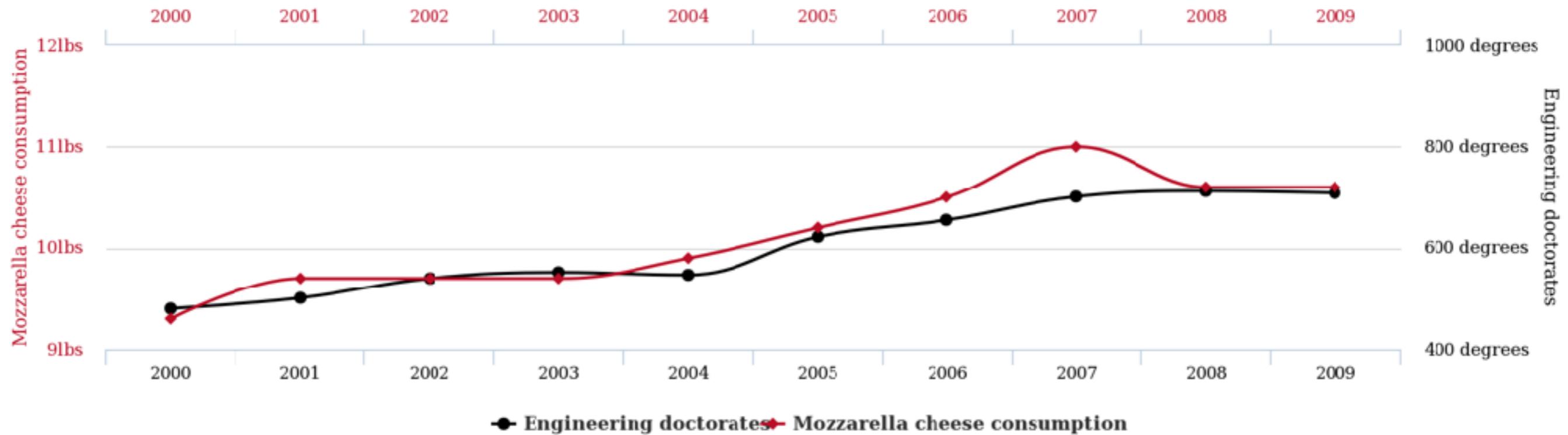
Les corrélations fallacieuses

Age of Miss America correlates with Murders by steam, hot vapours and hot objects



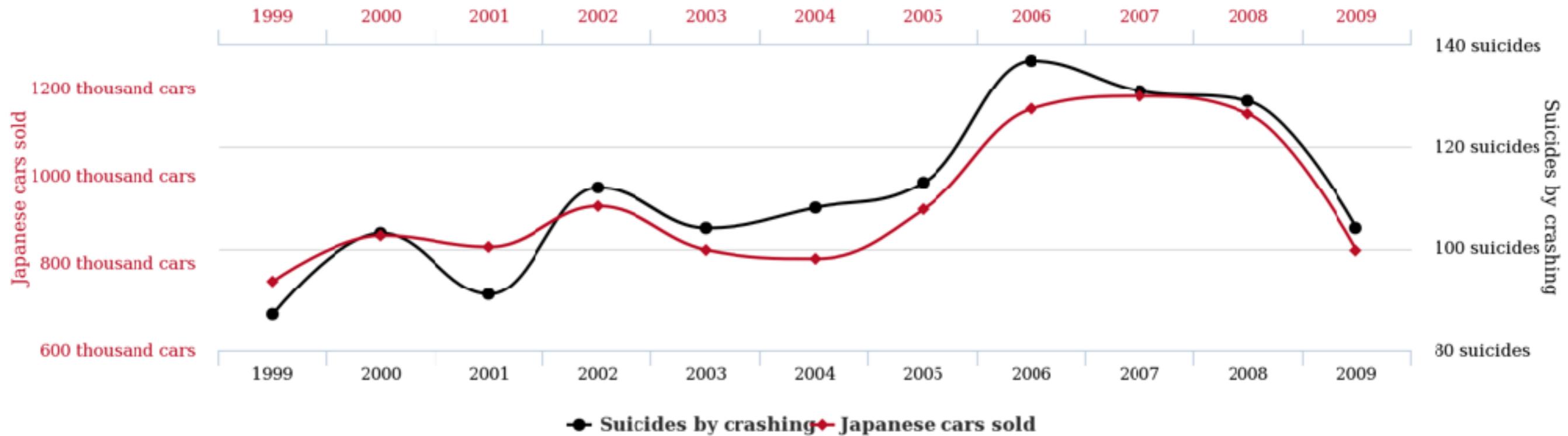
Les corrélations fallacieuses

Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded



Les corrélations fallacieuses

Japanese passenger cars sold in the US
correlates with
Suicides by crashing of motor vehicle



effectif théorique						
		Sociologue	Banquier	Archéologue		
	avec brouette	43,35	33,575	8,075	85	
	sans brouette	58,65	45,425	10,925	115	
		102	79	19	200	
différence						
		Sociologue	Banquier	Archéologue		
	avec brouette	6,35	-2,425	-3,925	0	
	sans brouette	-6,35	2,425	3,925	0	
		0	0	0	0	
X2						
	avec brouette	0,930	0,175	1,908	3,013	
	sans brouette	0,688	0,129	1,410	2,227	
		0	0	0	5,240	0,0728

la probabilité d'obtenir le tableau croisé observé sous l'hypothèse d'indépendance des deux variables est d'environ sept chances sur cent ;

la probabilité que les lignes et les colonnes du tableau sont indépendantes est d'environ sept chances sur cent ;

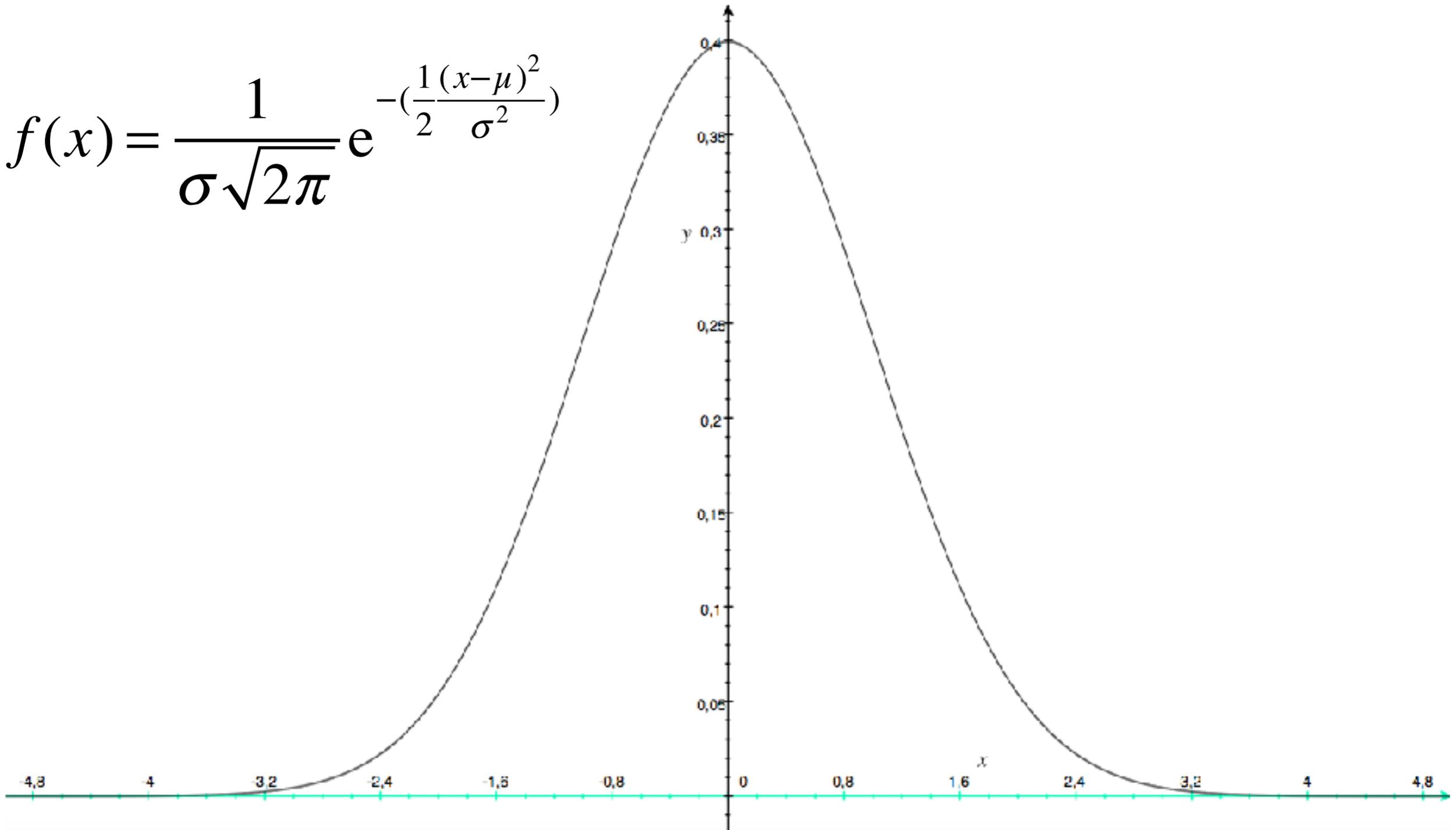
si j'affirme à partir du tableau croisé observé que la profession exercée a une influence sur le fait d'avoir ou non une brouette, j'ai 7 % de chances d'avoir tort, et 93 % de chances d'avoir raison.

Loi normale - Loi de Gauss

$$y = \frac{1}{\sqrt{2\pi}} \exp\left[-x \cdot \frac{x}{2}\right]$$

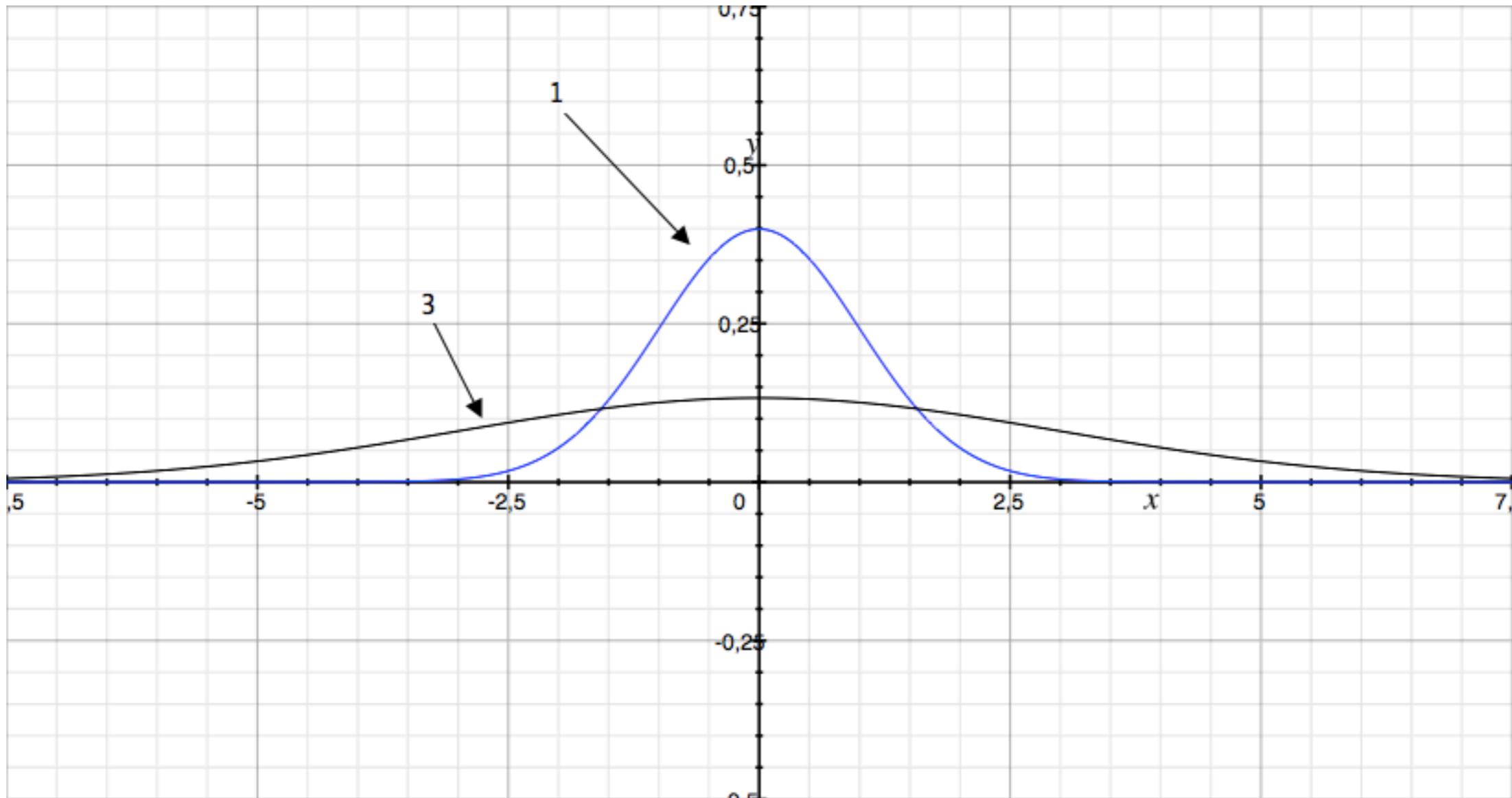
-23-

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right)}$$



Loi normale - Loi de Gauss

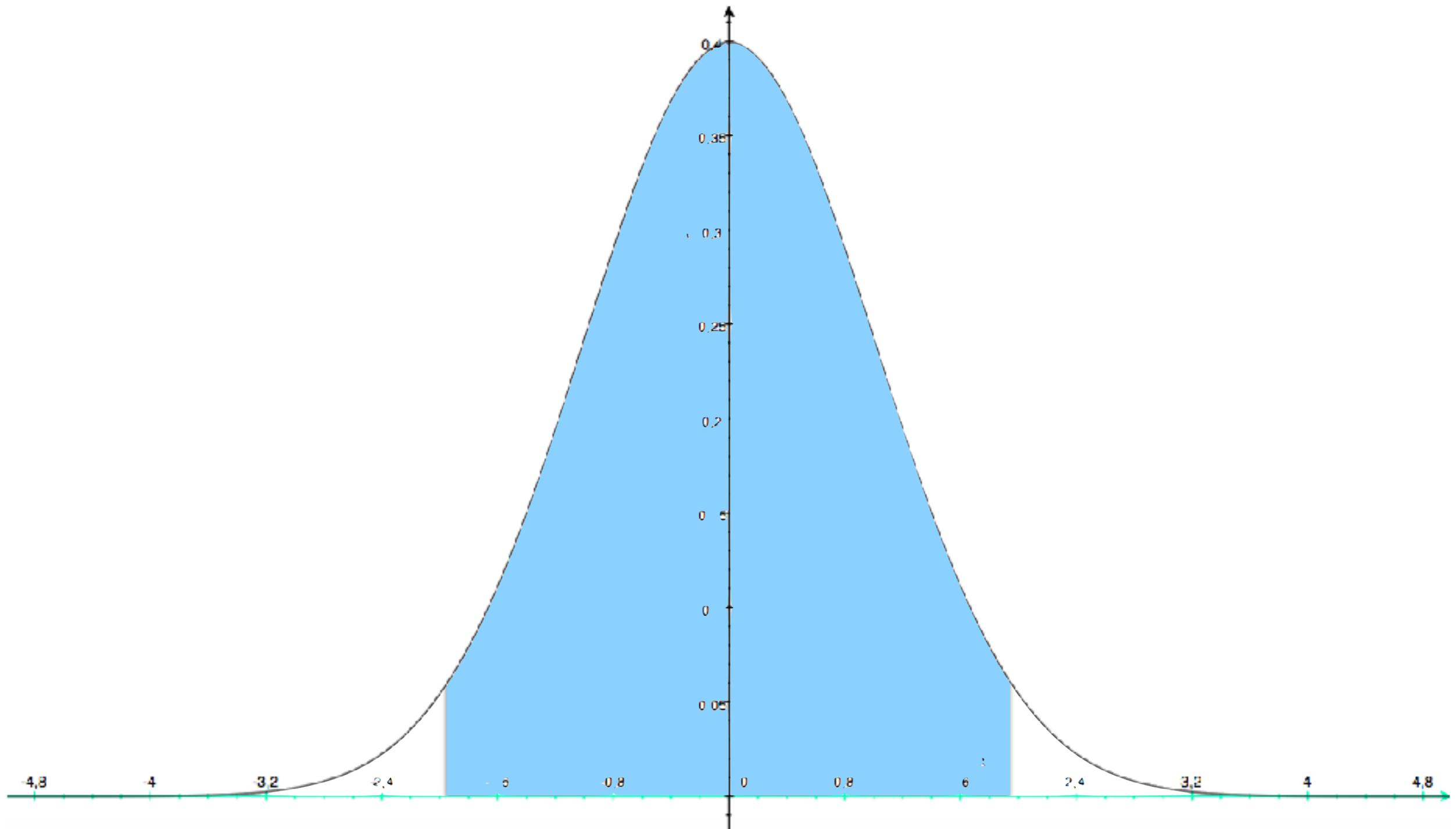
influence de la variance



Loi normale - Loi de Gauss

$$y = \frac{1}{\sqrt{2\pi}} \exp\left(-x \cdot \frac{x}{2}\right)$$

-23



Loi normale - Loi de Gauss

- Intervalles de confiance

$$P(\mu - \sigma \leq Y \leq \mu + \sigma) = 0,6827$$

$$P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) = 0,9545$$

$$P(\mu - 3\sigma \leq Y \leq \mu + 3\sigma) = 0,9973$$

- niveau de confiance à 95 %

$$P(-z \leq Y \leq +z) = 0,95$$

$$\rightarrow z = \pm 1,96$$

Analyse des résultats

Trois phases

- dépouiller
 - calculer
 - conclusions
- ➔ rapport d'enquête

Analyse des résultats

Chaque question devient une variable

mais

- on peut construire plusieurs variables avec une question (certaines questions ouvertes)
- on peut construire une variable à partir de plusieurs questions (variables synthétiques, indices)

Analyse des résultats

- Une variable, c'est ce qui varie. Elle peut prendre plusieurs valeurs possibles (modalités)
- Une variable est définie par l'ensemble de ses modalités

- Différents types de variables:
 - nominale ou catégorie
 - ordonnée ou ordinale
 - quantitative ou numérique

Analyse des résultats

Variable nominale

- valeurs distinctes, peuvent être placées dans n'importe quel ordre
- ex : avec qui êtes vous venus visiter cette exposition ?
 - ▶ seul(e)
 - ▶ en famille
 - ▶ avec des voisins ou amis
 - ▶ avec des collègues de travail
 - ▶ ...
- traitement statistique : fréquence (%), le mode

Analyse des résultats

Variable ordonnée ou ordinale

- il existe un ordre dans les modalités
- ex : classes d'ages, degré d'accord...
- traitement statistique : %, le mode, la médiane, le calcul de % cumulés

Analyse des résultats

Variable quantitative ou numérique

- modalités ordonnées et intervalles constant entre les mesures
- ex : nombre d'enfants, âge, poids, note de 1 à 10,...
- traitement statistique : + moyennes

Analyse des résultats

Numéro	Sexe	Age	Venue	Intérêt	etc
1	1	35	2	1	
2	1	45	3	1	
3	2	28	3	2	
4	1	32	1	1	
etc..					

- **Codification des données**
 - une ligne = 1 questionnaire
 - 1 colonne = 1 variable
 - dans chaque case, 1 réponse, donc une modalité pour un questionnaire donné

Codage des données

Questions fermées à une seule réponse possible

- variables nominales
 - codage variable «sexe»
 - 1 = masculin
 - 2 = féminin
 - 9 = non réponse
 - Codage variable «venue»
 - 1 = seul
 - 2 = voisins, amis
 - 3 = en famille
 - 4 = collègues
 - 9 = non réponse

Codage des données

Variables ordonnées

- codage de la variable «intérêt»
 - 1 = très intéressante
 - 2 = assez intéressante
 - 3 = peu intéressante
 - 4 = pas intéressante du tout
 - 9 = non réponse

Variables quantitatives

on utilise la valeur portée sur le questionnaire

Codage des données

- Questions fermées à choix multiple
 - faire la liste de toutes les modalités de réponse et coder
- Questions fermées à classement
 - faire la liste de toutes les modalités
- Questions ouvertes
 - élaborer une catégorisation des réponses et les traiter comme une question fermée.

Le dictionnaire des codes

Question	Numéro de la variable	Nom de la variable	Description du codage
Numéro du questionnaire	1	numéro	nombre de 1 à 498
Q1. Sexe	2	sexe	1 = masculin 2 = féminin 9 = non réponse
Q2 : Age	3	age	nombre de 17 à 89
Q3 : venu avec qui ?	4	venue	1 = seul 2 = voisins, amis 3 = en famille 4 = collègues,...
Q4 : intérêt de l'exposition	5	intérêt	1 = très intéressante 2 = assez intéressante 3 = peu intéressante,...
Q5...			

Recodage des variables

- faire réduire l'information et avoir des classes à peu près équilibrées,
- trouver les regroupements pertinents.
- ex. variables qualitatives : Rassembler les réponses « tout à fait d'accord » et « plutôt d'accord » sous une nouvelle rubrique « accord ».
- ex. : variables quantitatives : nombre de cigarettes, classes d'âge, nombre de frères et soeurs.

Variables synthétiques

- Variables synthétiques et indices (créés par fusion de réponses à plusieurs questions) :
- Elles sont construites à partir de plusieurs indicateurs destinés à donner une évaluation du même phénomène. On prend donc différentes variables, dont on aligne les modalités (0 – 1) afin d'obtenir des scores : avec cinq indicateurs, on obtient un score qui va de 0 à 5. On crée ensuite une nouvelle variable avec ses modalités spécifiques, qui peut ensuite être croisée avec d'autres.

Nouvelles variables

Ex. : indice de l'« ambiance livresque » : des points attribués en fonction du score obtenus à différentes questions : lecture père, lecture mère, nombre de livres à domicile, etc. On crée une nouvelle variable synthétique « Ambiance livresque » avec les modalités : défavorable, neutre, favorable.

Ex. : indice « homolalie de sexe » : pourcentage d'interlocuteur appartenant au même sexe que l'interviewé avec qui il a discuté sur une semaine.

Ex. : indice typologique pour l'étude de trajectoire ou de parcours typiques . Ex. famille-entourage local : à partir des questions sur la proximité spatiale, subjective, l'existence d'une entraide réelle et l'intensité des contacts.

Analyse des données

- Tris à plat
- Tris croisés
- Analyse multivariée

Analyse des données

Tris à plat

- Décrire le nombre d'individus dans chaque modalité : effectif et pourcentage
- Analyse :
 - faire une description qui utilise des rapports simples et pointer des chiffres remarquables
 - effectuer des recodages : mettre ensemble les «plutôt pas» et les «pas du tout» ou encore distinguer les «tout à fait» et les autres.

Analyse des données

Sexe	Effectif	Pourcentage
Femme	65	56 %
Homme	52	44 %
Total	117	100 %

«La culture, c'est une ouverture sur les autres»

Tout à fait d'accord	527	65 %
Plutôt d'accord	218	27 %
Plutôt pas d'accord	35	4 %
Pas du tout d'accord	6	1 %
Non-réponse	24	3 %
Total	810	100 %

Tris croisés et tableaux de contingence

- Un tableau de contingence permet de représenter deux caractères observés simultanément.
- Ex : Niveau d'études et opinion sur la culture

	Bac +1	Bac+2	Bac+3	Bac+4 et +	Total
Tout à fait d'accord	137	123	131	136	527
Plutôt d'accord	82	65	49	22	218
Plutôt pas d'accord	14	8	8	5	35
Pas du tout d'accord	2	2	1	1	6
Total	235	198	189	164	786

Quels tableaux ?

- vérifier systématiquement les réponses aux questions de l'enquête selon les critères sociodémographiques comme la classique trilogie âge, sexe et niveau social
- réaliser les tableaux qui correspondent aux hypothèses de l'enquête

Variable dépendante et variable indépendante

- La variable indépendante : elle représente dans l'hypothèse un facteur déterminant sur l'activité ou l'opinion considérée
- La variable dépendante : elle est censée subir l'influence de ce facteur (l'objet explicite de l'enquête sur les pratiques et les opinions)
- Ce sont les effectifs des différentes modalités de la variable indépendante (explicative) qui sont pris comme base 100 pour le calcul des pourcentages.

Exemple I

- Si on pense que la variable sexe est un élément déterminant sur la pratique religieuse, on établit un tableau croisé croisant la variable indépendante «sexe» et la variable dépendante «assistance régulière à la messe dominicale»

La tableau suivant infirme cette hypothèse

Aller à la messe			
Sexe	oui	non	Total
Masculin	40	60	100
Féminin	40	60	100

La tableau suivant confirme cette hypothèse

Aller à la messe			
Sexe	oui	non	Total
Masculin	40	60	100
Féminin	60	40	100

Exemple 2

- La variable à expliquer l'opinion sur la culture (variable dépendante) et le facteur explicatif est le niveau d'études (variable indépendante)

	Bac +1	Bac+2	Bac+3	Bac+4	Total
Tout à fait d'accord	58 %	62 %	69 %	83 %	67 %
Plutôt d'accord	35 %	33 %	26 %	13 %	28 %
Plutôt pas d'accord ou pas du tout d'accord	7 %	5 %	5 %	4 %	5 %
Total	100 %	100 %	100 %	100 %	100 %
Effectif	235	198	189	164	786

Analyse des données

Effectuer d'abord un lecture comparative entre les différents groupes (cf sexe, niveau d'études)

- dans les quatre groupes de niveau on a toujours la même hiérarchie de modalités
- on va ensuite essayer de repérer les variations et les extrêmes et pointer les écarts intéressants par rapport à la marge
- Exemple : l'accord total avec l'opinion sur la culture comme ouverture est toujours majoritaire et augmente régulièrement avec le niveau d'étude au détriment de l'accord mitigé. Le niveau de désaccord avec la proposition est assez constant et faible d'un niveau d'étude à l'autre.

Dépendance ou indépendance des variables : le Khi-deux

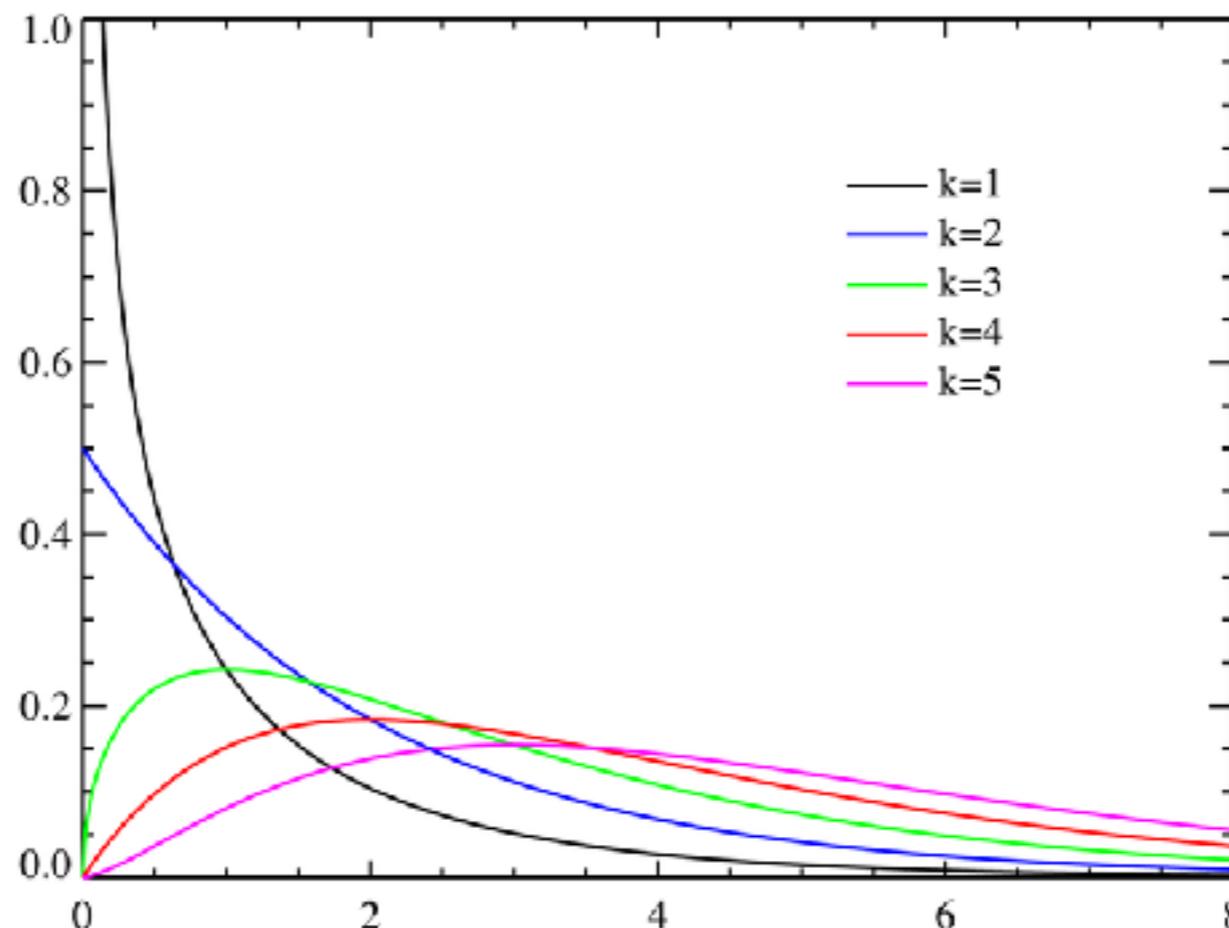
- Si les variables sont indépendantes, le hasard pourrait donner le même tableau
- Si les variables sont dépendantes, on considère qu'elles sont liées
- Une comparaison entre des effectifs théoriques (une répartition des effectifs proportionnelle aux effectifs locaux de la ligne et de la colonne) et les effectifs observés
- Le tableau des effectifs attendus dans l'hypothèse d'indépendance (effectifs théoriques)

Test du Khi-Deux

- test statistique permettant de vérifier l'adéquation d'une série de données à une famille de lois de probabilités, ou de tester l'indépendance de deux variables aléatoires.
- Calcul des écarts observés entre valeur observée (EO) et valeur théorique (ET) = $EO - ET$
- $X^2 = \text{somme } ((EO-ET)^2/ET)$

Test du Khi-deux

Si la valeur du χ^2 est supérieure à une limite déterminée par une table, on rejette l'hypothèse d'indépendance pour admettre que les deux distributions diffèrent significativement. Cela revient à dire que les deux variables sont liées.



Densité du χ^2 en fonction du nombre de degrés de liberté

	Bac +1	Bac+2	Bac+3	Bac+4 et +	Total
Tout à fait d'accord	137	123	131	136	527
Plutôt d'accord	82	65	49	22	218
Plutôt pas d'accord	14	8	8	5	35
Pas du tout d'accord	2	2	1	1	6
Total	235	198	189	164	786

Effectifs Observés	1er cycle	2 et 3ème cycle	Ensemble
Tout à fait d'accord	260	267	527
Plutôt d'accord	147	71	218
Plutôt pas d'accord ou pas du tout d'accord	26	5	41
Ensemble	433	353	786

Calcul du Khi-Deux

Effectifs Théoriques EO	1er cycle	2 et 3ème cycle	Ensemble
Tout à fait d'accord	$527 \cdot 433 / 786 = 290,3$ EO = 260	$527 \cdot 353 / 786 = 236,7$ EO = 267	527
Plutôt d'accord	$218 \cdot 433 / 786 = 120,1$ EO = 147	$218 \cdot 353 / 786 = 97,9$ EO = 71	218
Plutôt pas d'accord ou pas du tout d'accord	$41 \cdot 433 / 786 = 22,6$ EO = 26	$41 \cdot 353 / 786 = 18,4$ EO = 15	41
Ensemble	433	353	786

Test du Khi-Deux

- Calcul des écarts observés entre valeur observée (EO) et valeur théorique (ET) = EO - ET

$$260 - 290,3 = -30,3$$

$$147 - 120,1 = 26,9$$

$$26 - 22,6 = 3,4$$

$$267 - 236,7 = 30,3$$

$$71 - 97,9 = -26,3$$

$$15 - 18,4 = -3,4$$

- $\chi^2 = \text{somme } ((EO-ET)^2/ET)$

$$\chi^2 = (-30,3)^2 / 290,3 + (26,9)^2 / 120,1 + (3,4)^2 / 22,6 + (30,3)^2 / 236,7 + (-26,3)^2 / 97,9 + (-3,4)^2 / 18,4 = 21,62$$

Test du Khi-deux

		Seuil de risque				
		0,1	0,05	0,001	0,0001	0,00001
Nombre de degrés de liberté	1	2,71	3,84	6,64	10,83	15,14
	2	4,61	5,99	9,21	13,82	18,42
	3	6,25	7,82	11,35	16,27	21,11
	4	7,78	9,49	13,28	18,47	23,51
	5	9,24	11,07	15,09	20,52	25,74
	6	10,65	12,59	16,81	22,46	27,86
	7	12,02	14,07	18,48	24,32	29,88

Seuil de risque : risque statistique pris à rejeter l'hypothèse d'indépendance.

Ici on choisit 0,05 soit un risque de 5%

Nombre de degrés de liberté : déterminé par le produit du nombre de modalités de chacune des deux variables moins une = $(3-1)*(2-1) = 2$

Le chiffre indiqué par le tableau est 5,99. Le khi-deux est supérieur à 5,99 : on rejette l'hypothèse d'indépendance

Remarque : ce test n'est valable qu'avec des effectifs théoriques suffisants.

Loi de Khi-deux

Le tableau donne x tel que $P(K > x) = p$

p	0,999	0,995	0,99	0,98	0,95	0,9	0,8	0,2	0,1	0,05	0,02	0,01	0,005	0,001
ddl														
1	0,0000	0,0000	0,0002	0,0006	0,0039	0,0158	0,0642	1,6424	2,7055	3,8415	5,4119	6,6349	7,8794	10,8276
2	0,0020	0,0100	0,0201	0,0404	0,1026	0,2107	0,4463	3,2189	4,6052	5,9915	7,8240	9,2103	10,5966	13,8155
3	0,0243	0,0717	0,1148	0,1848	0,3518	0,5844	1,0052	4,6416	6,2514	7,8147	9,8374	11,3449	12,8382	16,2662
4	0,0908	0,2070	0,2971	0,4294	0,7107	1,0636	1,6488	5,9886	7,7794	9,4877	11,6678	13,2767	14,8603	18,4668
5	0,2102	0,4117	0,5543	0,7519	1,1455	1,6103	2,3425	7,2893	9,2364	11,0705	13,3882	15,0863	16,7496	20,5150
6	0,3811	0,6757	0,8721	1,1344	1,6354	2,2041	3,0701	8,5581	10,6446	12,5916	15,0332	16,8119	18,5476	22,4577
7	0,5985	0,9893	1,2390	1,5643	2,1673	2,8331	3,8223	9,8032	12,0170	14,0671	16,6224	18,4753	20,2777	24,3219
8	0,8571	1,3444	1,6465	2,0325	2,7326	3,4895	4,5936	11,0301	13,3616	15,5073	18,1682	20,0902	21,9550	26,1245
9	1,1519	1,7349	2,0879	2,5324	3,3251	4,1682	5,3801	12,2421	14,6837	16,9190	19,6790	21,6660	23,5894	27,8772
10	1,4787	2,1559	2,5582	3,0591	3,9403	4,8652	6,1791	13,4420	15,9872	18,3070	21,1608	23,2093	25,1882	29,5883
11	1,8339	2,6032	3,0535	3,6087	4,5748	5,5778	6,9887	14,6314	17,2750	19,6751	22,6179	24,7250	26,7568	31,2641
12	2,2142	3,0738	3,5706	4,1783	5,2260	6,3038	7,8073	15,8120	18,5493	21,0261	24,0540	26,2170	28,2995	32,9095
13	2,6172	3,5650	4,1069	4,7654	5,8919	7,0415	8,6339	16,9848	19,8119	22,3620	25,4715	27,6882	29,8195	34,5282
14	3,0407	4,0747	4,6604	5,3682	6,5706	7,7895	9,4673	18,1508	21,0641	23,6848	26,8728	29,1412	31,3193	36,1233
15	3,4827	4,6009	5,2293	5,9849	7,2609	8,5468	10,3070	19,3107	22,3071	24,9958	28,2595	30,5779	32,8013	37,6973
16	3,9416	5,1422	5,8122	6,6142	7,9616	9,3122	11,1521	20,4651	23,5418	26,2962	29,6332	31,9999	34,2672	39,2524
17	4,4161	5,6972	6,4078	7,2550	8,6718	10,0852	12,0023	21,6146	24,7690	27,5871	30,9950	33,4087	35,7185	40,7902
18	4,9048	6,2648	7,0149	7,9062	9,3905	10,8649	12,8570	22,7595	25,9894	28,8693	32,3462	34,8053	37,1565	42,3124
19	5,4068	6,8440	7,6327	8,5670	10,1170	11,6509	13,7158	23,9004	27,2036	30,1435	33,6874	36,1909	38,5823	43,8202
20	5,9210	7,4338	8,2604	9,2367	10,8508	12,4426	14,5784	25,0375	28,4120	31,4104	35,0196	37,5662	39,9968	45,3147
21	6,4467	8,0337	8,8972	9,9146	11,5913	13,2396	15,4446	26,1711	29,6151	32,6706	36,3434	38,9322	41,4011	46,7970
22	6,9830	8,6427	9,5425	10,6000	12,3380	14,0415	16,3140	27,3015	30,8133	33,9244	37,6595	40,2894	42,7957	48,2679
23	7,5292	9,2604	10,1957	11,2926	13,0905	14,8480	17,1865	28,4288	32,0069	35,1725	38,9683	41,6384	44,1813	49,7282
24	8,0849	9,8862	10,8564	11,9918	13,8484	15,6587	18,0618	29,5533	33,1962	36,4150	40,2704	42,9798	45,5585	51,1786
25	8,6493	10,5197	11,5240	12,6973	14,6114	16,4734	18,9398	30,6752	34,3816	37,6525	41,5661	44,3141	46,9279	52,6197
26	9,2221	11,1602	12,1981	13,4086	15,3792	17,2919	19,8202	31,7946	35,5632	38,8851	42,8558	45,6417	48,2899	54,0520
27	9,8028	11,8076	12,8785	14,1254	16,1514	18,1139	20,7030	32,9117	36,7412	40,1133	44,1400	46,9629	49,6449	55,4760
28	10,3909	12,4613	13,5647	14,8475	16,9279	18,9392	21,5880	34,0266	37,9159	41,3371	45,4188	48,2782	50,9934	56,8923
29	10,9861	13,1211	14,2565	15,5745	17,7084	19,7677	22,4751	35,1394	39,0875	42,5570	46,6927	49,5879	52,3356	58,3012
30	11,5880	13,7867	14,9535	16,3062	18,4927	20,5992	23,3641	36,2502	40,2560	43,7730	47,9618	50,8922	53,6720	59,7031
40	17,9164	20,7065	22,1643	23,8376	26,5093	29,0505	32,3450	47,2685	51,8051	55,7585	60,4361	63,6907	66,7660	73,4020
50	24,6739	27,9907	29,7067	31,6639	34,7643	37,6886	41,4492	58,1638	63,1671	67,5048	72,6133	76,1539	79,4900	86,6608
60	31,7383	35,5345	37,4849	39,6994	43,1880	46,4589	50,6406	68,9721	74,3970	79,0819	84,5799	88,3794	91,9517	99,6072
70	39,0364	43,2752	45,4417	47,8934	51,7393	55,3289	59,8978	79,7146	85,5270	90,5312	96,3875	100,4252	104,2149	112,3169
80	46,5199	51,1719	53,5401	56,2128	60,3915	64,2778	69,2069	90,4053	96,5782	101,8795	108,0693	112,3288	116,3211	124,8392
90	54,1552	59,1963	61,7541	64,6347	69,1260	73,2911	78,5584	101,0537	107,5650	113,1453	119,6485	124,1163	128,2989	137,2084
100	61,9179	67,3276	70,0649	73,1422	77,9295	82,3581	87,9453	111,6667	118,4980	124,3421	131,1417	135,8067	140,1695	149,4493
120	77,7551	83,8516	86,9233	90,3667	95,7046	100,6236	106,8056	132,8063	140,2326	146,5674	153,9182	158,9502	163,6482	173,6174
140	93,9256	100,6548	104,0344	107,8149	113,6593	119,0293	125,7581	153,8537	161,8270	168,6130	176,4709	181,8403	186,8468	197,4508
160	110,3603	117,6793	121,3456	125,4400	131,7561	137,5457	144,7834	174,8283	183,3106	190,5165	198,8464	204,5301	209,8239	221,0190
180	127,0111	134,8844	138,8204	143,2096	149,9688	156,1526	163,8682	195,7434	204,7037	212,3039	221,0772	227,0561	232,6198	244,3705
200	143,8428	152,2410	156,4320	161,1003	168,2786	174,8353	183,0028	216,6088	226,0210	233,9943	243,1869	249,4451	255,2642	267,5405
250	186,5541	196,1606	200,9386	206,2490	214,3916	221,8059	231,0128	268,5986	279,0504	287,8815	298,0388	304,9396	311,3462	324,8324
300	229,9634	240,6634	245,9725	251,8637	260,8781	269,0679	279,2143	320,3971	331,7885	341,3951	352,4246	359,9064	366,8444	381,4252
400	318,2596	330,9028	337,1553	344,0781	354,6410	364,2074	376,0218	423,5895	436,6490	447,6325	460,2108	468,7245	476,6064	493,1318
500	407,9470	422,3034	429,3875	437,2194	449,1468	459,9261	473,2099	526,4014	540,9303	553,1268	567,0698	576,4928	585,2066	603,4460
600	498,6229	514,5289	522,3651	531,0191	544,1801	556,0560	570,6680	628,9433	644,8004	658,0936	673,2703	683,5156	692,9816	712,7712
700	590,0480	607,3795	615,9075	625,3175	639,6130	652,4973	668,3308	731,2805	748,3591	762,6607	778,9721	789,9735	800,1314	821,3468
800	682,0665	700,7250	709,8969	720,0107	735,3623	749,1852	766,1555	833,4557	851,6712	866,9114	884,2789	895,9843	906,7862	929,3289
900	774,5698	794,4750	804,2517	815,0267	831,3702	846,0746	864,1125	935,4987	954,7819	970,9036	989,2631	1001,6296	1013,0364	1036,8260

Analyse multivariée

Les tableaux de contingence mettent en évidence les relations entre les variables mais il est imprudent d'interpréter directement une relation comme une relation de cause à effet ou l'indication d'une liaison certaine

- les interprétations d'un lien avéré peuvent être multiples
- il se peut que les deux variables soient toutes les deux influencées par une troisième

ex : taille des enfants / capacité à résoudre un problème mathématique :
contrôle de la variable âge (variable test)

ex : sexe / accident de voiture : contrôle par kilométrage (variable test)

On va donc introduire de nouvelles variables pour contrôler les effets (variables tests)

Analyse multivariée

Emissions	Jusqu'à 40 ans	Plus de 40 ans
Programmes religieux	17 %	26 %
Tribunes politiques	34 %	45 %
Programmes de musique classique	30 %	29 %
Effectif	1000	1300

Relations entre l'âge et l'écoute de trois types d'émissions radiophoniques

Les personnes âgées recherchent d'avantage les programmes religieux et les tribunes politiques mais il n'y pas de différence entre les «jeunes» et les «vieux» pour les programmes de musique classique

Analyse multivariée

- Introduction d'une variable test : le niveau d'instruction
- En pratique, la réalisation des tableaux à trois variables se concrétise par la réalisation d'une série de tableaux croisés à deux dimensions, un par modalité de la variable test.

	Niveau supérieur d'instruction		Niveau inférieur d'instruction	
	Jusque 40 ans	Plus de 40 ans	Jusque 40 ans	Plus de 40 ans
Écoutent	9 %	11 %	29 %	32 %
N'écoutent pas	91 %	89 %	71 %	68 %
Total	100 %	100 %	100 %	100 %
Effectif	600	400	400	900

Écoute des programmes religieux en fonction de l'âge et du niveau d'instruction

- A niveau d'instruction constant, «jeunes» et «vieux» ont à peu près le même niveau d'écoute des programmes religieux. La relation initialement constatée entre âge et écoute des programmes religieux disparaît à l'intérieur de chaque sous-groupe.
- Elle était due au fait que les «jeunes» ont globalement un niveau d'instruction plus élevé que les plus «vieux»
- Par contre, l'écoute des programmes religieux est liée de façon marquée au niveau d'instruction (il y a davantage d'écoute au niveau inférieur)

	Niveau supérieur d'instruction		Niveau inférieur d'instruction	
	Jusque 40 ans	Plus de 40 ans	Jusque 40 ans	Plus de 40 ans
Écoutent	40 %	55 %	25 %	40 %
N'écoutent pas	60 %	45 %	75 %	60 %
Total	100 %	100 %	100 %	100 %
Effectif	600	400	400	900

Écoute des tribunes politiques en fonction de l'âge et du niveau d'instruction

- L'intérêt marqué des plus âgés pour les tribunes politiques est confirmé quel que soit le niveau d'instruction. On observe en outre un plus haut intérêt dans chaque groupe d'âge lorsque le niveau d'instruction est plus élevé.
- La relation initiale est donc conservée. La variable contrôlée a un effet complémentaire de renforcement ou d'atténuation

	Niveau supérieur d'instruction		Niveau inférieur d'instruction	
	Jusque 40 ans	Plus de 40 ans	Jusque 40 ans	Plus de 40 ans
Écoutent	32 %	52 %	28 %	19 %
N'écoutent pas	68 %	48 %	72 %	81 %
Total	100 %	100 %	100 %	100 %
Effectif	600	400	400	900

Écoute des programmes de musique classique en fonction de l'âge et du niveau d'instruction

- Initialement, on n'avait pas constaté de relation entre l'âge et l'écoute de musique classique.
- En intégrant le niveau d'instruction, on constate que, au niveau d'instruction supérieur, l'intérêt est plus élevé chez les plus âgés et, au niveau d'instruction inférieur, c'est le contraire.
- On a donc une interaction entre deux variables explicatives qui doivent être présentes en même temps pour analyser correctement la variable «écoute de la musique classique».

Analyse multivariée

Autres méthodes :

- Régression Linéaire (multilinéaire)
- Analyse en Composantes Principales (ACP)
- Analyse Factorielle des Correspondances (AFC)

