LA STATISTIQUE POUR CHERCHER, MODÉLISER, RAISONNER, COMMUNIQUER…

Atelier SFdS - APMEP P1-14 – Toulouse 19 octobre 2014

Philippe DUTARTE
Groupe enseignement de la statistique - SFdS

Un des objectifs de cet atelier est de montrer, par des exemples d'activités de classe, en quoi la statistique permet, dans le cadre des programmes de collège et de lycée, de développer les compétences mathématiques.

Les **fichiers** Excel correspondant à cet atelier sont accessibles à l'adresse :
http://www.sfds.asso.fr/367-Ressources_pour_le_secondaire

Chercher

Analyser un problème.
Extraire, organiser et traiter l'information utile.

Activité 1 : un indice de bonheur ?

« Le PIB mesure tout, sauf ce qui fait que la vie vaut la peine d'être vécue. »

Robert KENNEDY.

L'indice de développement humain (IDH) est calculé par le Programme de développement humain des Nations Unies à partir de l'espérance de vie, du niveau d'éducation et du niveau de vie. Le fichier « idh.xls » ou « idh.ods » fournit les données 2013 (rapport UNDP 2014 source : http://hdr.undp.org/fr/data).
Voir fichiers idh.xls et idh_activite_3eme.xls pour une expérimentation en classe de troisième par Fabienne GLEBA, professeure au collège V. Duruy de Fontenay-sous-Bois (Val-de-Marne).

1. Proposer un regroupement des pays en trois classes selon leur niveau d'IDH : « très élevé », « moyen à élevé », « faible à moyen ».

Coup de pouce
On pourra représenter la répartition des IDH.

2. Utiliser des arguments statistiques pour situer les IDH de la France et de la Zambie (0,561).

Coup de pouce
Comment la répartition des IDH explique-t-elle que la moyenne est inférieure à la médiane ? Quel indicateur vous semble préférable pour situer un pays ?

3. La même organisation des Nations Unies fournit un **indice d'inégalité de genre** (IIG), prenant notamment en compte le taux d'éducation des filles et la représentation des femmes au parlement. Reprendre l'étude statistique précédente à propos de l'IIG afin de situer deux pays par rapport aux autres.

Modéliser

Traduire en langage mathématique une situation réelle.
Utiliser, comprendre, élaborer une simulation numérique prenant appui sur la modélisation.

Activité 2 : Noël au balcon, Pâques au tison ?

Fatah CHALBI
Collège Elsa Triolet 94 Champigny-sur-Marne.

Problème : le dicton « Noël au balcon, Pâques au tison » est-il vrai ?

Coup de pouce
On propose les températures (à Paris) des deux dernières années à Noël et à Pâques.

Date	Noël 2012	Pâques 2013
Température	12°C	6°C

Date	Noël 2013	Pâques 2014
Température	10°C	16°C

Activité en classe de troisième

On considère les températures à Paris de 1900 à 2004. Les températures brutes ont été extraites à partir du site Néerlandais *http://www.ecad.eu/dailydata/customquery.php*
Ouvrir le classeur Etude_dicton_Eleve.xls

A - Premier cycle de modélisation

1. Dans la feuille **Résumé 1er tour**, compléter les colonnes de températures de Noël et de Pâques.
Indication : les données sont disponibles dans les onglets **Noël 1er tour** *et* **Pâques 1er tour.**
2. Dans les cellules C108 (et D108), calculer la moyenne des températures de Noël (respectivement de Pâques).
3. Dans la colonne E :
$T°_{\text{moy Noël}}$ $T°_{\text{moy Pâques}}$
Lorsque $T°_{\text{Noël}} \geq T°_{\text{moy Noël}}$ et que $T°_{\text{Pâques}} \leq T°_{\text{moy Pâques}}$ alors le tableur renvoie OUI.
Lorsque $T°_{\text{Noël}} \geq T°_{\text{moy Noël}}$ et que $T°_{\text{Pâques}} > T°_{\text{moy Pâques}}$ alors le tableur renvoie NON.
Lorsque $T°_{\text{Noël}} < T°_{\text{moy Noël}}$ alors le tableur ne renvoie rien.
Calculer les fréquences d'apparition du « Oui » et du « Non ».
Indication : la fonction NB.SI() *permet de dénombrer.*
4. Que dire de la validité du dicton ? Expliquer.

B – Deuxième cycle de modélisation (avec les vraies dates de Pâques)

En utilisant les données des pages Résumé 2ème tour ; Noël 2ème tour et Pâques 2ème tour, reprendre les questions du A.

C - Troisième cycle de modélisation (en prenant en compte les températures de la veille et du lendemain de Noël et de Pâques)

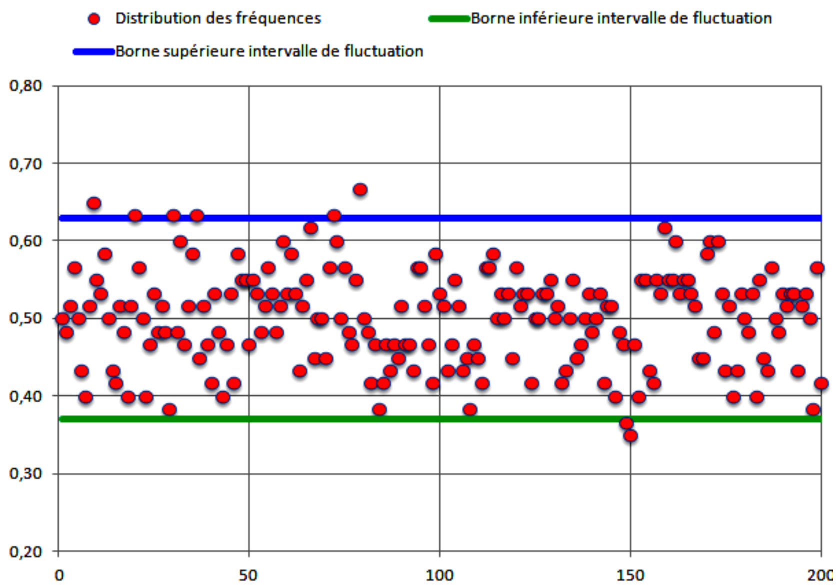
En utilisant les données des pages **Résumé 3ème tour ; Noël 3ème tour** et **Pâques 3ème tour,** reprendre les questions du A.
Fichier « réponses » : Etude_dicton_Professeur.xls

Activité en classe de seconde

On reprend l'activité de troisième, à laquelle s'ajoute un travail autour de la prise de décision pour valider ou invalider le résultat obtenu grâce à l'utilisation de l'intervalle de fluctuation au programme de la classe de seconde générale et technologique (en première, on pourrait utiliser la loi binomiale…).

Les résultats obtenus dans la première partie laissent penser que notre étude pourrait s'apparenter au lancer d'une pièce de monnaie équilibrée sur un échantillon de taille 60 (il y a 59 « oui/non » pour les tours 1 et 2 et 62 « oui/non » pour le tour 3). On assimile l'issue pile à la véracité du dicton.

1. On a simulé sur tableur 200 échantillons de taille 60 de lancers d'une pièce supposée équilibrée. Les points correspondent à la fréquence de pile sur chaque échantillon.



a) Calculer les bornes de l'intervalle de l'intervalle de fluctuation :

$I = \left[0{,}5 - \dfrac{1}{\sqrt{60}}\ ;\ 0{,}5 + \dfrac{1}{\sqrt{60}}\right]$ et le visualiser sur le graphique ci-dessus.

b) Combien de fréquences se trouvent en dehors de l'intervalle I ? Peut-on dire, dans ce cas précis, que pour 95% au moins de ces échantillons de taille 60, la fréquence de pile appartient à l'intervalle I ?

2. Ouvrir le fichier Simulation_piece_de_monnaie.xls.

Taper un grand nombre de fois sur la touche F9. Que constatez-vous ?

3. Au seuil de 95 %, peut-on valider ou invalider les résultats de l'étude du dicton ?

Modéliser

Traduire en langage mathématique une situation réelle.

Activité 3 : Jefferson a-t-il rencontré Laplace ? Le cas étonnant de la Constitution américaine

En 1998, un tribunal américain a mené l'enquête pour savoir si Thomas Jefferson, l'un des pères de la Constitution américaine de 1787, avait pu rencontrer le mathématicien français Pierre-Simon Laplace lors de son ambassade à Paris à la fin du règne de Louis XVI (rapporté par Alain Desrosières dans *Prouver et gouverner* 2014). Ce procès opposait la Chambre des représentants, à majorité républicaine, au département du Commerce, relevant de l'administration démocrate de Bill Clinton. La Constitution prévoit un poids relatif des États américains au niveau fédéral selon leurs populations respectives, fondé sur un « dénombrement réel » (*actual enumeration*). Or les méthodes de recensement exhaustif ont tendance à sous-estimer certaines catégories de population, notamment dans les quartiers pauvres, comme les Noirs ou les Hispaniques, traditionnellement plus proches des démocrates. L'administration responsable du recensement souhaitait procéder par sondages aléatoires, ce que les républicains considéraient comme non constitutionnel. La question était de savoir si Jefferson avait pu rencontrer Laplace, précurseur de la méthode d'échantillonnage aléatoire, dont il avait calculé les risques d'erreur (le tribunal américain a conclu que la réponse était « probablement non »).

Le 30 novembre 1785, Laplace lut à l'Académie des sciences un mémoire « Sur les naissances, les mariages et les morts... » dont voici des extraits :
(source http://gallica.bnf.fr/ark:/12148/bpt6k77599c/f40)

« La population est un des plus sûrs moyens de juger de la prospérité d'un empire, et les variations qu'elle éprouve, comparées aux événements qui les précèdent, sont la plus juste mesure de l'influence des causes physiques et morales sur le bonheur ou sur le malheur de l'espèce humaine. Il est donc intéressant, à tous égards, de connaître la population de la France, d'en suivre les progrès et d'avoir la loi suivant laquelle les hommes sont répandus sur la surface de ce grand royaume. [...]
Quoique les naissances soient la source de la population, elles ne suffisent pas cependant pour la déterminer : […] voyons comment on peut y parvenir. […]
On choisira donc un grand nombre de paroisses dans toutes les provinces du royaume pour avoir un milieu entre les petites différences que les causes locales peuvent apporter dans les résultats ; on fera ensuite un dénombrement exact de leurs habitants à une époque donnée, et, par le relevé des naissances durant les dix années qui précèdent cette époque, on déterminera le nombre correspondant des naissances annuelles. En divisant par ce nombre celui des habitants, on aura le rapport de la population aux naissances, d'une manière d'autant plus précise que le dénombrement sera plus considérable. […] Le rapport de la population aux naissances, déterminé par la méthode précédente, ne peut jamais être rigoureusement exact ; en lui supposant même une précision rigoureuse, il resterait encore sur la population de la France l'incertitude qui nait de l'action de causes variables. La population de la France, tirée des naissances annuelles, n'est donc qu'un résultat probable, et par conséquent susceptible d'erreurs. C'est à l'analyse des hasards à déterminer la probabilité de ces erreurs et jusqu'à quel point on doit porter le dénombrement pour qu'il soit très probable qu'elles seront renfermées dans d'étroites limites. Ces recherches dépendent d'une théorie nouvelle et encore peu connue, celle de la probabilité des événements futurs prise des événements observés […].
Les dénombrements déjà faits en France et comparés aux naissances donnent à peu près 26 pour le rapport de la population aux naissances annuelles ; or, si l'on prend un milieu entre les naissances des années 1781 et 1782, on a 973 054,5 pour le nombre de naissances annuelles dans toute l'étendue de ce royaume, en y comprenant la Corse ; en multipliant donc ce nombre par 26, la population de la France entière sera de 25 299 417 habitants. Maintenant je trouve par mon analyse que, pour avoir une probabilité de 1 000 contre 1, de ne pas se tromper d'un demi million dans cette évaluation de la population de la France, il faudrait que le dénombrement qui a servi à déterminer le facteur 26 eût été de 771 469 habitants. […] Voici l'analyse qui m'a conduit à ce résultat.
Considérons une urne qui renferme une infinité de boules blanches et noires dans un rapport inconnu […]. Il est facile d'appliquer ces résultats [longs calculs précédents] à la théorie de la population déduite des naissances, car on peut considérer chaque naissance annuelle comme étant représentée par une boule noire, et chaque individu existant comme étant représenté par une boule blanche… »

Reprenons la modélisation de Laplace, celle d'une urne de Bernoulli contenant une infinité de boules noires (naissances) et blanches (individus) dans un rapport donné, et utilisons les outils de calcul actuels.

1. a) Plutôt que de considérer le « rapport de la population aux naissances », estimé à 26 pour la France de 1781-1782, prenons en compte le taux de natalité c'est-à-dire le rapport des naissances à la population. Montrer que pour la France de cette époque, on peut supposer que le taux de natalité p vaut 0,038 5 (c'est-à-dire 38,5 naissances pour 1 000 habitants).

b) À titre de comparaison, rechercher sur Internet le taux actuel de natalité en France ainsi qu'un pays dont le taux de natalité actuel est proche de 38,5 pour 1 000.

2. On suppose que l'urne contient un nombre infini de boules noires et blanches avec une proportion de boules noires valant $p = 0{,}038\,5$. On prélève au hasard 771 469 boules dans l'urne et on désigne par X la variable aléatoire correspondant au nombre de boules noires dans l'échantillon ainsi obtenu.

a) Déterminer l'espérance de X.

Coup de pouce

Quelle loi suit X ?

b) Laplace admet une fluctuation de plus ou moins 2 % (500 000 sur 25 299 417) autour de la valeur attendue (l'espérance) et évoque une probabilité de 1 pour 1 000 de « se tromper ». Peut-on justifier cette affirmation ?

Coup de pouce

Calculer $P(29\,107 \leq X \leq 30\,296)$.

Représenter

Choisir un cadre adapté pour traiter un problème ou pour représenter un objet mathématique.
Passer d'un mode de représentation à un autre.
Changer de registre.

Activité 4 : Mesurer l'inégalité des revenus

La courbe de Lorenz et le coefficient de Gini sont des outils qui permettent de visualiser et de quantifier les inégalités de revenus.

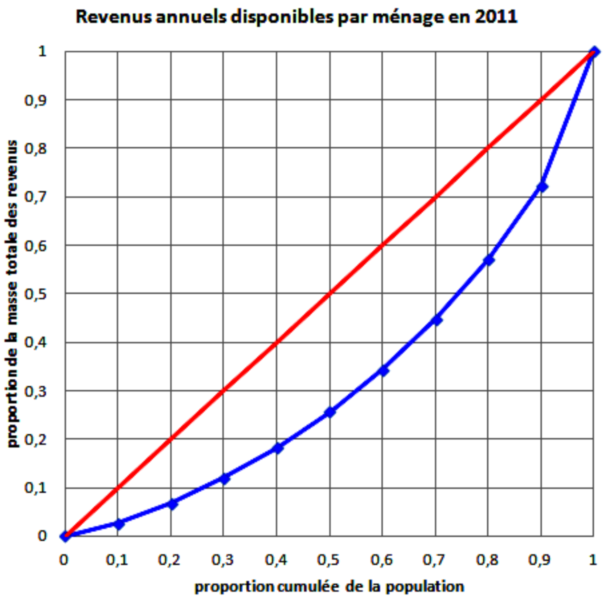
On s'intéresse à la répartition des revenus disponibles des ménages en France en 2011 (source disponible en ligne : *Le revenu et le patrimoine des ménages*, INSEE édition 2014).
Reproduire sur une feuille de calcul le tableau de données ci-contre.

Lecture : pour les 10% des ménages les plus modestes, le revenu annuel disponible en 2011 est inférieur à 13 070 euros et le revenu moyen est de 9 480 euros.

D4 f_x

	A	B	C	D	E
1	Distribution des revenus annuels disponibles par ménage en 2011				
2	proportion cumulée de la population	limite supérieure du revenu annuel disponible	revenu annuel disponible moyen de la tranche	proportion de la masse totale des revenus	proportion cumulée de la masse totale des revenus
3	0	0	0	0	0
4	0,1	13 070,00 €	9 480,00 €		
5	0,2	16 830,00 €	15 060,00 €		
6	0,3	20 380,00 €	18 570,00 €		
7	0,4	24 470,00 €	22 400,00 €		
8	0,5	29 010,00 €	26 730,00 €		
9	0,6	34 210,00 €	31 510,00 €		
10	0,7	40 490,00 €	37 250,00 €		
11	0,8	48 680,00 €	44 280,00 €		
12	0,9	62 980,00 €	54 810,00 €		
13	1	***	99 750,00 €		
14					

1. a) Lire le revenu médian annuel disponible en 2011.
b) Calculer le revenu moyen disponible en 2011 (on remarquera que chaque revenu moyen par tranche correspond à 10% de la population).
2. Indiquer les formules à entrer en D4 et en E4 qui, recopiées vers le bas, permettent de compléter les colonnes D et E.
3. La courbe ci-contre, nommée courbe de Lorenz, est obtenue avec le tableur en sélectionnant les points dont les coordonnées figurent en colonnes A et E. La diagonale rouge correspond à une répartition totalement équitable des revenus, où, par exemple 60% de la masse totale des revenus est distribuée à 60% de la population.



a) Lire le pourcentage de la population aux plus faibles revenus à qui est distribuée 60% de la masse totale des revenus.
b) L'inégalité de la distribution des revenus est visualisée par l'écart entre la diagonale rouge et la courbe de Lorenz située en-dessous. Le coefficient de Gini, défini comme le double de l'aire (en unités d'aires) comprise entre la diagonale rouge et la courbe de Lorenz, est compris entre 0 (égalité complète) et 1 (inégalité totale) et mesure l'inégalité de la distribution. Calculer le coefficient de Gini pour 2011.

Coup de pouce
Entrer en F4 la formule =0,1*(E3+E4)/2 fournissant l'aire du trapèze situé entre le premier segment de la courbe de Lorenz et l'axe des abscisses. Recopier vers le bas jusqu'en F13 puis faire la somme pour obtenir l'aire située entre la courbe de Lorenz et l'axe des abscisses.
4. Les données pour 2004 sont fournies ci-dessous (source : INSEE). Les inégalités se sont-elles accrues entre 2004 et 2011 ?

Distribution des revenus annuels disponibles par ménage en 2004		
proportion cumulée de la population	limite supérieure du revenu annuel disponible	revenu annuel disponible moyen de la tranche
0,1	11 480,00 €	8 790,00 €
0,2	14 410,00 €	13 010,00 €
0,3	17 580,00 €	15 940,00 €
0,4	20 940,00 €	19 250,00 €
0,5	24 600,00 €	22 780,00 €
0,6	28 620,00 €	26 570,00 €
0,7	33 170,00 €	30 840,00 €
0,8	39 360,00 €	36 080,00 €
0,9	49 550,00 €	43 930,00 €
1	***	72 170,00 €

Voir le fichier : inegalite_revenus.xls

Calculer

Effectuer un calcul automatisable à la main ou à l'aide d'un instrument.
Exercer l'intelligence du calcul. Contrôler les calculs.

Activité 5 : Risque, coût d'opportunité et aide à la décision

En économie, dans une situation à risque où l'on est confronté à plusieurs choix, le coût d'opportunité d'un choix donné est le meilleur gain que l'on peut obtenir en choisissant l'un des autres choix. Examinons un cas pratique.

Une compagnie aérienne assure chaque jour la liaison Paris-Toulouse avec un avion comportant 75 places. On considère que chaque jour les 75 places sont réservées mais que la probabilité pour qu'un client ayant réservé se présente à l'embarquement est 0,92, indépendamment d'un client à l'autre. La compagnie s'interroge sur l'opportunité, ou non, d'effectuer une surréservation.

1. a) La compagnie prévoit de surréserver en effectuant 80 réservations. On considère la variable aléatoire X correspondant au nombre de passagers se présentant à l'embarquement. Déterminer la loi de X puis la probabilité qu'il y ait au moins un client en surnombre.

b) Les pertes d'opportunité sont ainsi définies : chaque place vide coûte 200 € en recette perdue et chaque client en surnombre coûte 400 €, la compagnie s'engageant à rembourser le double de la valeur du billet.

En construisant une feuille de calcul, vérifier que l'espérance des pertes d'opportunité est 510,46 €.

Coup de pouce

E2 f_x =LOI.BINOMIALE(A2;80;0,92;FAUX)

	A	B	C	D	E	F
1	k	nombre de places vides	nombre de clients en surnombre	coût	probabilité	espérance
2	0	75	0	15 000,00 €	1,7668E-88	0,00 €
3	1	74	0	14 800,00 €	1,6255E-85	0,00 €
4	2	73	0	14 600,00 €	7,3838E-83	0,00 €
5	3	72	0	14 400,00 €	2,2078E-80	0,00 €
6	4	71	0	14 200,00 €	4,8874E-78	0,00 €
75	73	2	0	400,00 €	0,15139786	60,56 €
76	74	1	0	200,00 €	0,16469633	32,94 €
77	75	0	0	- €	0,15152062	- €
78	76	0	1	400,00 €	0,11463731	45,85 €
79	77	0	2	800,00 €	0,06848463	54,79 €
80	78	0	3	1 200,00 €	0,03029128	36,35 €
81	79	0	4	1 600,00 €	0,00881898	14,11 €
82	80	0	5	2 000,00 €	0,00126773	2,54 €
83					total	510,46 €

2. Déterminer de façon analogue l'espérance des pertes d'opportunité lorsque la compagnie choisit de ne pas faire de surréservation et effectue 75 réservations.

3. Que faut-il choisir : effectuer 75 réservations ou 80 ? Justifier que le coût d'opportunité correspondant au choix de 75 réservations au lieu de 80 réservations est, en moyenne, de 689,54 € par vol.

4. En utilisant le tableur, vérifier que le choix retenu à la question précédente est le meilleur.

Le coût d'opportunité correspondant au choix de ne pas effectuer de surréservation (plutôt que tout autre choix) est donc, en moyenne, de 689,54 € par vol.

Voir le fichier : risque_cout_opportunite.xls

Raisonner

Utiliser les notions de la logique élémentaire pour bâtir un raisonnement.
Utiliser différents types de raisonnement.
Effectuer des inférences (inductives, déductives) pour obtenir de nouveaux résultats, conduire une démonstration, confirmer ou infirmer une conjecture, prendre une décision.

Activité 6 : Discrimination ou effet de structure ?

Des effets de structure dans les données ont été mentionnés par les statisticiens Karl Pearson (1899) et Udny Yule (1903) mais le phénomène a été pour la première fois décrit par Edward Simpson dans un article de 1951 et est souvent connu sous le nom de « paradoxe de Simpson ».
Un des exemples réels les plus connus se produisit lorsque l'Université de Californie à Berkeley fut poursuivie pour discrimination sexiste dans les admissions, la différence entre les taux d'admission des filles et des garçons étant jugée trop importante pour être due au hasard.
Le tableau suivant fournit les données des admissions à Berkeley des six principaux départements de l'Université pour la rentrée 1973.

Garçons		Filles	
Nombre de candidats	Taux d'admission	Nombre de candidats	Taux d'admission
2 691	0,45	1 835	0,3

1. a) Calculer le taux d'admission moyen pour ces six départements.
b) Si l'on suppose que les $n = 1\,835$ candidates constituent un échantillon aléatoire issu d'une population où le taux d'admission est $p = 0{,}39$, le programme de seconde prévoit une fluctuation de la fréquence observée comprise entre $p - \frac{1}{\sqrt{n}}$ et $p + \frac{1}{\sqrt{n}}$ dans plus de 95 % des cas. Si la fréquence observée n'appartient pas à cet intervalle, nous dirons qu'elle présence un écart « significatif ». Le taux d'admission des filles présente-t-il un écart significatif par rapport au taux moyen ?
2. Si l'on considère les taux d'admission pour chacun des six départements, on observe les résultats suivants.

Département	Garçons		Filles	
	Nombre de candidats	Taux d'admission	Nombre de candidats	Taux d'admission
A	825	0,62	108	0,82
B	560	0,63	25	0,68
C	325	0,37	593	0,34
D	417	0,33	375	0,35
E	191	0,28	393	0,24
F	373	0,06	341	0,07

a) Vérifier que les taux d'admission du premier tableau sont cohérents avec ceux du second tableau.
b) Pourquoi la situation peut-elle être qualifiée de paradoxale ? Les écarts observés sont-ils significatifs ?
Coup de pouce
Le taux d'admission des filles dans le département A présente-t-il un écart significatif (en leur faveur) par rapport à ce taux moyen d'admission dans ce département ?
Le taux d'admission des filles dans le département C présente-t-il un écart significatif (en leur défaveur) par rapport à au taux moyen d'admission dans ce département ?

Communiquer

Développer une argumentation mathématique correcte à l'écrit ou à l'oral.
Critiquer une démarche ou un résultat.
S'exprimer avec clarté et précision à l'oral et à l'écrit.

Activité 7 : Docteur Spock et l'affaire des jurées disparues

Extrait de Hans Zeisel et Harry Kalven « *Parking tickets and missing women : Statistics and the Law*. ».

« Cet exemple est issu du procès en 1968 du pédiatre et auteur Benjamin Spock à la cour du district de Boston pour conspiration contre le service militaire en encourageant à la résistance contre la guerre au Vietnam. Dans ce procès, la défense contesta la légalité de la méthode de sélection du jury. Bien que plus de la moitié des jurés éligibles à Boston furent des femmes, il n'y avait aucune femme dans le jury du docteur Spock. Or ce dernier, plus qu'aucun autre accusé, en souhaitait la présence parce que tant de femmes avaient élevé leurs enfants « selon le docteur Spock » ; de plus, les sondages d'opinion montraient que les femmes, en général, étaient davantage opposées à la guerre du Vietnam.
La question était de savoir si l'absence totale de femmes était accidentelle dans ce jury ou si elle résultait d'une discrimination systématique. Le raisonnement statistique devait fournir une réponse.
À la cour du district de Boston, les jurés sont sélectionnés en trois étapes. L'annuaire de la ville est utilisé pour la première étape ; à partir de celui-ci, le greffier est supposé sélectionner 300 noms au hasard, c'est-à-dire comme dans une loterie. La deuxième étape de sélection intervient lorsqu'un procès est sur le point de débuter. À partir des 300 noms dans la boîte, les noms de 30, ou davantage, jurés potentiels sont choisis. Ces personnes sont convoquées à la cour le matin du procès. Dans la troisième étape, douze jurés sont véritablement sélectionnés après interrogation, à la fois par le procureur et par la défense.
Le tableau suivant indique le pourcentage de femmes dans quelques 46 procès après la deuxième étape pour les sept juges de la cour fédérale du district de Boston.

Juge du procès Spock	Autres juges du district de Boston					
	A	B	C	D	E	F
6,4% 8,7%						
13,3% 13,6% 15% 15,2% 17,7% 18,6%	16,8%				17,7% 19,7%	16,5%
23,1%		27% 28,9%	21% 23,4% 27,5% 27,5%	24,3% 29,7%	21,5% 27,9%	20,7% 23,5% 26,4% 26,7% 29,5% 29,8%
	30,8% 33,6%	32% 32,7% 35,5%	30,5% 31,9% 32,5% 33,8% 33,8%		34,8%	31,9% 36,2%
	40,5% 48,9%	45,6%			40,2%	
Moyenne : 14,6%	Moyenne : 34,1%	Moyenne : 33,6%	Moyenne : 29,1%	Moyenne : 27%	Moyenne : 27%	Moyenne : 26,8%

L'effectif des jurés potentiels après l'étape 2 (minimum 30) est variable, les moyennes calculées en tiennent compte.

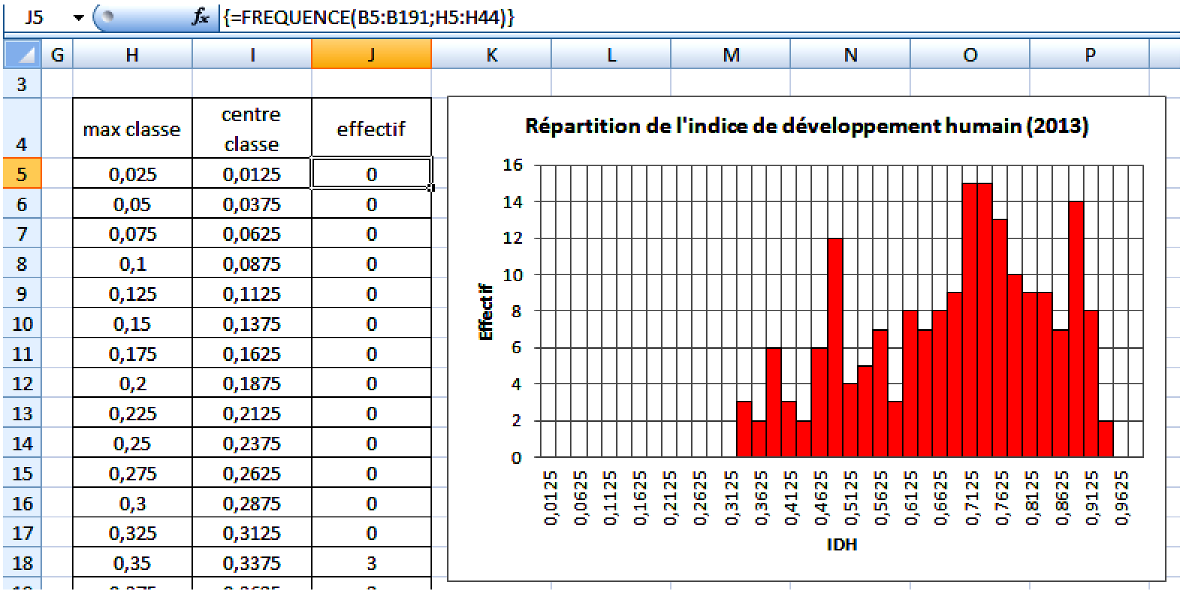
La proportion moyenne de femmes choisies par les six collègues du juge du procès Spock était 29%, et de plus, les moyennes de ces six juges étaient proches de la moyenne du groupe. Cela suggère que la proportion de femmes parmi la liste de 300 noms de la boîte à jurys était proche de 29%. Mais le tableau montre aussi que le juge du procès Spock avait constamment de petits pourcentages de femmes avec une moyenne de seulement 14,6%, pratiquement moitié moins que ses collègues.
Il est possible, bien sûr, que la méthode de sélection du juge du procès Spock soit la même que celle de ses six collègues. Mais quelle est la probabilité qu'une différence aussi large (ou plus large encore) qu'entre 14,6% et 29% puisse se produire par hasard ? Un calcul statistique révèle que la probabilité est d'environ 10^{-18} pour que « la chance au tirage » conduise à une distribution de jurées femmes égale à celle du juge de ce procès ou plus extrême. La conclusion est donc sans appel : les jurés potentiels pour le juge du procès ont été choisis dans la liste centrale d'une façon qui, systématiquement, réduisait la proportion des femmes.
Ainsi la proportion des femmes parmi les jurés potentiels souffrait d'une double réduction impropre, tout d'abord lorsque le greffier de la cour réduit leur profil de la majorité dans l'annuaire de la ville à 29% dans les listes pour les jurys, puis lorsque le juge conduit à abaisser de 29% à 14,6% sa moyenne personnelle. Dans le procès Spock seule une femme jurée potentielle vint devant la cour, et elle fut aisément éliminée à l'étape 3 par le procureur dans son quota de choix (pour lequel il n'a pas à donner de raison). »

1. a) L'effectif des jurés potentiels après l'étape 2 dans l'ensemble des sélections du juge du procès Spock est d'environ 700 personnes. Quel est, approximativement, le nombre total de jurés potentiels femmes sélectionné par ce juge ?
b) En modélisant la situation décrite à l'aide d'une loi binomiale de paramètres $n = 700$ et $p = 0,29$, retrouver l'ordre de grandeur de la probabilité citée dans le texte.
2. Peut-on considérer qu'il existe une différence significative des fréquences de jurées potentielles obtenues, d'une part, par le juge du procès Spock, et d'autre part par le juge C et par le juge F (dont les effectifs de jurés potentiels sont approximativement 700) ?
Coup de pouce
Déterminer l'intervalle de fluctuation au seuil de 95% de la fréquence d'un caractère sur un échantillon aléatoire de taille 700 prélevé dans une population où la proportion du caractère est 0,29.

Éléments de réponse

1. Un indice de bonheur ?

1. Répartition :



J5 {=FREQUENCE(B5:B191;H5:H44)}

	max classe	centre classe	effectif
5	0,025	0,0125	0
6	0,05	0,0375	0
7	0,075	0,0625	0
8	0,1	0,0875	0
9	0,125	0,1125	0
10	0,15	0,1375	0
11	0,175	0,1625	0
12	0,2	0,1875	0
13	0,225	0,2125	0
14	0,25	0,2375	0
15	0,275	0,2625	0
16	0,3	0,2875	0
17	0,325	0,3125	0
18	0,35	0,3375	3

Les limites des trois groupes peuvent être fixées à environ 0,59 et 0,86.

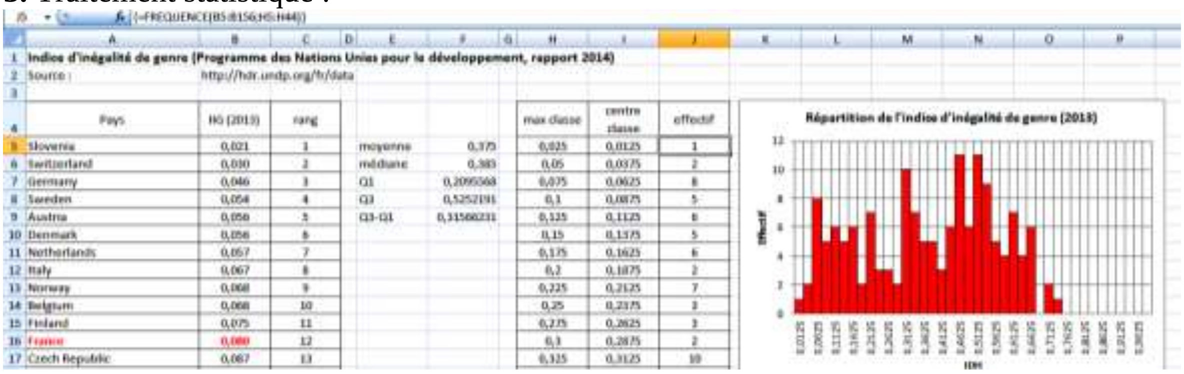
2. La dissymétrie de la répartition explique que la moyenne est inférieure à la médiane : l'histogramme présente une queue importante de faibles IDH.

Pour situer un pays, la médiane est préférable à la moyenne. Par exemple, la République Dominicaine avec un IDH d'environ 0,7 obtient un résultat supérieur à la moyenne mais la moitié des pays font mieux qu'elle.

L'IDH de la France se situe en 2013 au 20[e] rang sur 187 pays, mais, dans le groupe des pays possédant un IDH très élevé, elle se situe plutôt à la fin.

L'IDH de la Zambie en 2013 est proche du troisième quartile : 75% des pays font mieux qu'elle, 25% font moins bien.

3. Traitement statistique :



{=FREQUENCE(B5:B156;H5:H44)}

Indice d'inégalité de genre (Programme des Nations Unies pour le développement, rapport 2014)

Source : http://hdr.undp.org/fr/data

	Pays	IIG (2013)	rang			max classe	centre classe	effectif
5	Slovenia	0,021	1	moyenne	0,375	0,025	0,0125	1
6	Switzerland	0,030	2	médiane	0,383	0,05	0,0375	2
7	Germany	0,046	3	Q1	0,2095568	0,075	0,0625	8
8	Sweden	0,054	4	Q3	0,5252191	0,1	0,0875	5
9	Austria	0,056	5	Q3-Q1	0,31566231	0,125	0,1125	6
10	Denmark	0,056	6			0,15	0,1375	5
11	Netherlands	0,057	7			0,175	0,1625	6
12	Italy	0,067	8			0,2	0,1875	2
13	Norway	0,068	9			0,225	0,2125	7
14	Belgium	0,068	10			0,25	0,2375	3
15	Finland	0,075	11			0,275	0,2625	3
16	France	0,080	12			0,3	0,2875	2
17	Czech Republic	0,087	13			0,325	0,3125	10

2. Noël au balcon, Pâques au tison ?

Compte-rendu de l'activité en classe de troisième

Cette activité s'est déroulée en salle informatique durant une séance d'une heure. À la fin de la séance, les élèves sont invités à présenter leur réponse à la classe en utilisant leurs feuilles de classeur (tableur).

Déroulement de la séance

1ère étape : Le professeur projette à l'aide d'un dispositif de vidéo-projection les deux images présentant un père Noël au soleil et un lapin de Pâques sous la neige.
La consigne n'est pas donnée de suite, le professeur demande aux élèves « quels questionnements ces images suscitent-elles ? ». Les élèves ont été surpris par l'absence, a priori, de consigne.
2ème étape : L'enseignant dévoile le dicton accompagnant les images et énonce la consigne. Les élèves discutent et communiquent leurs compréhensions (et incompréhension) du dicton. Les images aidant à donner du sens au dicton, le professeur puise des éléments de réponses d'élèves pour en fixer le sens.
3ème étape : Les élèves discutent de la validité du dicton, les échanges sont fructueux. Le professeur demande aux élèves de prouver leurs dires. Un seul binôme a eu l'idée de s'appuyer sur des données réelles de températures pour « tester le dicton ». Le professeur projette deux tableaux regroupant les températures à Noël et à Pâques pour les deux dernières années afin de prolonger les débats.
4ème étape : La séance est bien engagée et les élèves adhèrent à l'étude. Le résultat du test sur les deux dernières années ne suffisant pas à prendre position sur la validité du dicton, une étude statistique à l'aide du tableur est proposée aux élèves. (cf : fichiers et fiche d'activité élève joints).
5ème étape : Le professeur distribue les fiches d'activité, les élèves sont invités à ouvrir le fichier « *Etude_dicton_Eleve.xlsx* » pour réaliser le travail.

Le professeur commente le choix de prendre 3 modèles simples et réduits à un seul paramètre à savoir la température.
1er modèle (1er tour de modélisation) :
Dans ce modèle nous travaillons sur les températures du 25 décembre et le 08 avril (Cette date correspond au milieu du segment allant du 22 mars au 25 avril (dates extrêmes de Pâques).
2ème modèle (2ème tour de modélisation) :
On prend les températures du 25 décembre et les dates réelles de Pâques (données par un algorithme de Gauss).
3ème modèle (3ème tour de modélisation) :
On prend les températures sur un segment autour de Noël et Pâques (segment : J-1 ; J ; J+1).

On définit alors des critères afin de s'entendre sur ce que nous appelons un «Noël au balcon » et des « Pâques au tison ».
Pour Noël :
Le critère retenu pour notre étude est le suivant :
Un « Noël au balcon » est un Noël à température supérieure à la température moyenne à Noël ou « normale ».
Pour Pâques **:**
De la même façon nous avons retenu le critère suivant :

Un « Pâques au tison » est des Pâques à température inférieure à la température moyenne à Pâques ou « normale ».
Le professeur passe dans les groupes, et répond aux questions d'ordres technique (copier les données d'une colonne, etc.).

6ème étape : Exposition des résultats trouvés par des binômes volontaires (vidéo-projection) et observation de fréquences proches de 0,5.
Résumé du premier tour :

E112 =NB.SI(E2:E106;"OUI")

	B	C	D	E	F
103	2001	2,4	10,8		
104	2002	12,6	9,8	Non	
105	2003	7,2	5,4	Oui	
106	2004	3,3	7		
107					
108		4,4	9,8		
109					
110					
111				Nombre	Fréquence
112			OUI	31	0,53
113			NON	28	0,47
114			TOTAL	59	1,00
115					

Résumé des deuxième et troisième tour :

E111 =NB.SI(E2:E106;"oui")

	C	D	E	F
109				
110			Nombre	Fréquence
111		OUI	28	0,47
112		NON	31	0,53
113		TOTAL	59	1

E113 =NB.SI(E2:E106;"oui")

	C	D	E	F
111				
112			Nombre	Fréquence
113		OUI	32	0,53
114		NON	30	0,47
115		TOTAL	62	1

Conclusion : le dicton n'est pas statistiquement justifié.

3. Jefferson a-t-il rencontré Laplace ? Le cas étonnant de la Constitution américaine

1. a) $\dfrac{973054{,}5}{25299417} \approx \dfrac{1}{26} \approx 0{,}038\ 5$.

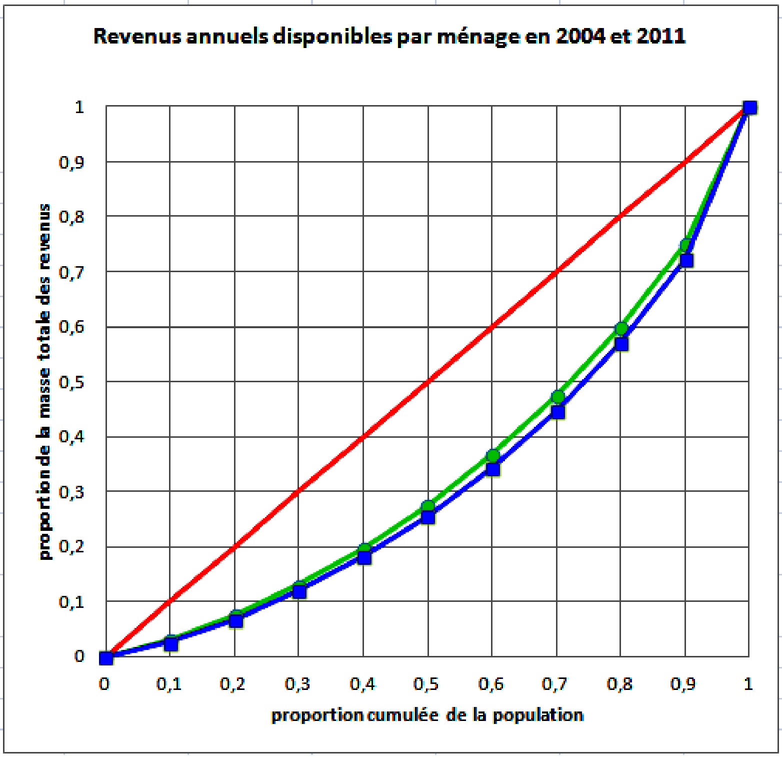
b) Le taux de natalité en France pour 2013 est de 12,2 naissances pour 1 000 habitants. Le taux de natalité au Mozambique, par exemple, est estimé à 38,5 naissances pour 1 000 habitants (2007).
2. a) X suit la loi binomiale de paramètres $n = 771\ 469$ et $p = 0{,}038\ 5$.
Son espérance est $np \approx 29\ 702$.
b) Sur un tableur, la formule :
=LOI.BINOMIALE(30296;771469;0,0385;VRAI) – LOI.BINOMIALE(29106;771469;0,0385;VRAI)
donne une probabilité d'environ 0,999 6.
La probabilité d'obtenir un résultat situé en-dehors de l'intervalle précédent est donc d'environ 0,4 pour 1 000. C'est l'ordre de grandeur annoncé par Laplace.

4. Mesurer l'inégalité des revenus

1.a) Me = 29 010 €.
b) $\bar{x}$ = 35 984 €.
2. On peut entrer en D4 la formule =C4/SOMME(C$4:C$13).
On peut entrer en E4 la formule =E3+D4.
3. a) Environ 82% de la population aux plus faibles revenus se partage 60% des revenus.
b) Le coefficient de Gini pour 2011 vaut environ 0,35.
4. On peut copier la feuille de calcul précédente et substituer les valeurs de 2004 à celles de 2011. La superposition des courbes de Lorenz montre que les inégalités se sont aggravées. Ce que quantifie le coefficient de Gini, qui vaut environ 0,32 pour 2004.

F15 =2*(0,5-F14)

	F
1	
2	aire du trapèze
3	
4	0,00151887
5	0,0052858
6	0,01028822
7	0,01636888
8	0,02363146
9	0,0321589
10	0,04207907
11	0,05364252
12	0,06746786
13	0,08752938
14	0,33997097
15	0,32005806
16	
17	
18	
19	



5. Risque, coût d'opportunité et aide à la décision

1. a) X suit la loi binomiale de paramètres n = 80 et p = 0,92.
$P(X > 75) \approx 0{,}223$.
b) Voir le fichier risque_cout_opportunite.xls.
2. L'espérance des pertes d'opportunité lorsque la compagnie choisit de ne pas faire de surréservation et effectue 75 réservations est 1 200 €.
3. 1 200 – 510,46 = 689,54 €.
4. On obtient les résultats suivants.

n	75	76	77	78	79	80	81	82
Espérance de perte	1 200	1 017,06	840,49	683,29	565,86	510,46	533,47	640,12

6. Discrimination ou effet de structure ?

1. a) Taux d'admission moyen environ 0,39.
b) 0,35 $\notin$ [0,36 ; 0,42].
2. a) On calcule, à partir du tableau 2, les taux d'admission moyens pour les garçons et pour les filles pour vérifier si l'on retrouve ceux du tableau 1. On trouve environ 0,445 et 0,303 ce qui est cohérent.
b) Pour 4 départements sur 6, le taux d'admission des filles est supérieur à celui des garçons, pourtant, globalement, on observe la situation contraire.
Taux d'admission moyen pour le département A : environ 0,64.
0,82 $\notin$ [0,54 ; 0,74]. La différence est significative.
Taux d'admission moyen pour le département C : environ 0,35.
0,34 $\in$ [0,31 ; 0,39]. La différence n'est pas significative.

7. Docteur Spock et l'affaire des jurées disparues

1.a) Il y a environ 102 femmes parmi les jurés potentiels du juge du procès Spock.
b) On suppose que le choix des 700 jurés potentiels s'effectue au hasard, et avec remise, parmi la liste comportant une proportion p = 0,29 de femmes. On considère la variable aléatoire X correspondant au nombre de femmes obtenu. Cette variable aléatoire suite la loi binomiale de paramètres 700 et 0,29.
On obtient $\mathrm{P}(X \leq 102) \approx 2 \times 10^{-19}$. C'est bien l'ordre cde grandeur annoncé dans l'article.
2. On observe que $P(X \leq 179) \approx 0{,}024$; $P(X \leq 180) \approx 0{,}029$; $P(X \leq 210) \approx 0{,}735$; $P(X \leq 211) \approx 0{,}761$. On en déduit que l'intervalle de fluctuation de $\dfrac{X}{700}$ à 95% est $\left[\dfrac{180}{700}, \dfrac{211}{700}\right]$ c'est-à-dire environ [0,257 ; 0,302].
La valeur observée pour le juge du procès Spock, 0,146, n'appartient pas à l'intervalle précédent. La différence est donc significative au seuil de 95%.
Les valeurs observées pour le juge C, 0,291, et pour le juge F, 0,268, appartiennent à l'intervalle précédant. La différence avec 0,29 n'est pas jugée significative.