

Jour de naissance

Document présenté par:

Pascale Balland
108 avenue du Général de Gaulle
94 700 Maisons Alfort

Professeur de mathématiques
au
Collège Pierre Brossolette
154 bis avenue Pierre Brossolette
94 170 Le Perreux Sur Marne
tel: 01 48 71 13 63
fax 01 48 71 00 63
RNE 0940788W

avec tous mes remerciements à Mme Trinquet pour son aide
précieuse.

Jour de naissance

Présentation et objectifs :

Ce travail est un complément sur la notion de hasard et de traitement de données statistiques. Il s'inscrit dans le programme de 4ème de mathématiques paragraphe 1.3 « traitements de données » et 1.1 « utilisation de la proportionnalité », ainsi que du B.2.I .

Problématique :

Le jour de naissance (lundi, mardi, mercredi, jeudi, vendredi, samedi ou dimanche) dans la semaine est-il dû au hasard?

Compétences travaillées :

Cette étude a permis de travailler sur les savoir-faire suivants :

1. notions de statistiques
 - Etudier le caractère d'une population réelle : effectuer un sondage, récolter, ordonner des données en vue de leur traitement ; s'interroger sur la taille de l'échantillon ;
 - Utiliser des outils statistiques : fréquence, construire un tableau de données, un graphique, intérêt des différents types de graphiques.
2. notion de probabilités : le hasard et l'ordre qui en découle ; évolution de la fréquence d'une série de données aléatoires en fonction du nombre.
3. B2I :
 - Utilisation et intérêt du tableur (*openoffice. Org Calc*) pour compiler des données ; créer, modifier et insérer une formule dans une feuille de calcul ; créer un graphique à partir des données d'une feuille de calcul.
 - Envoyer et recevoir un fichier joint sur messagerie.

Modalité de mise en œuvre :

Cette activité s'est déroulée durant des séances d'atelier, c'est à dire avec des groupes d'une quinzaine d'élèves, ce qui a permis de favoriser la prise en main d'un tableur en salle informatique ainsi que l'échange entre élèves.

Déroulement :

J'ai demandé aux élèves d'interroger leur entourage sur leur date de naissance. Suite à une remarque de certains, il m'a fallu insister sur le fait que nous voulions faire une véritable étude, et que des données inventées fausseraient sûrement le résultat et ne nous permettraient pas de tirer des conclusions intéressantes.

Nous avons discuté de « l'échantillon » à sonder, et nous avons convenu qu'il devait comporter des personnes de tous les âges. Ils ont donc interrogé chacun 10 personnes (parents, oncles et tantes, grands-parents, frères et sœurs, cousins et cousines).

1ère partie : traitement des données (annexe 1, première et deuxième parties)

Le premier travail a été de déterminer le jour de naissance dans la semaine à partir de la date de naissance.

Après l'inévitable « c'est facile madame » et des ébauches de calculs compliqués, j'ai présenté aux élèves une formule fournie par *openoffice.org calc* : « =JOURSEM() »; cette dernière renvoie un chiffre en fonction de la date de naissance insérée (1 pour dimanche, 2 pour lundi ...).

Ils ont alors élaboré un tableau à partir de leur série de donnée (voir annexe 1).

Remarque : lorsque nous avons voulu trier par ordre décroissant la colonne B, nous nous sommes aperçus que rien ne se passait. Cela m'a permis de leur expliquer que le tri se faisait sur la formule insérée et non sur le chiffre (le logiciel « voit » les formules).

Lors de la mise en commun de leurs observations, ils se sont rendu compte que chacune de leur série était d'effectif trop faible pour en tirer des conclusions.

Ils ont alors mis leurs données en commun, en vérifiant que celles-ci ne se recoupaient pas (j'avais pris soin pour cela de ne pas leur demander les dates de naissance parmi les amis du collège).

Lors de la lecture du diagramme circulaire, les élèves se sont aperçus que la répartition dans la semaine n'était pas régulière.

Deux avis sont alors apparus:

1. Puisque la répartition n'est pas régulière, c'est que les jours de naissance ne sont pas le fait du hasard. En effet, ils attendaient une répartition à peu près identique tous les jours de la semaine.
2. C'est justement parce qu'on ne choisit pas le jour de sa naissance que cette répartition n'est pas régulière dans la semaine.

La suite du travail a donc consisté en deux points:

1. D'une part, interpréter les résultats obtenus (émettre des hypothèses pouvant les expliquer, élaborer une stratégie pour vérifier ces hypothèses).
2. D'autre part, étudier le comportement d'une série de données aléatoires modélisant le problème.

2ème partie: interprétation des résultats à l'aide d'un questionnaire (annexe 2 et 3)

Pour expliquer le déséquilibre dans la répartition des données, les élèves ont été amenés à émettre l'hypothèse que les naissances devaient être parfois « programmées ».

Je leur ai alors donné à lire un extrait d'un compte rendu d'une conférence du *Collège national des gynécologues et obstétriciens français* du 30 novembre 1995 publié sur internet (voir annexe 2).

Suite à cette lecture, nous nous sommes rendu compte que notre échantillon n'était pas bien choisi: en effet, on ne peut visualiser l'incidence du déclenchement des accouchements sur le jour de naissance qu'à partir de 1970, puisque avant cette date il ne se pratiquait pas ou très peu.

D'après l'article, le déclenchement des naissances se faisant de façon plus régulière à partir de 1995, nous avons décidé de trier nos données en considérant d'une part toutes les dates de naissance avant 1970 et d'autre part toutes celles après 1995.

A cette occasion, je leur ai montré l'avantage de comparer les données des différents tableaux sous la forme de fréquence, l'effectif total n'étant pas le même ; on peut ainsi répondre plus facilement aux questions du type « dans quel tableau la « part » d'un jour précis est-elle la plus importante ».

L'étude du document en annexe 3, résultat du tri effectué à partir de toutes les données récoltées dans trois classes, nous a montré que la répartition des naissances dans la semaine est plus régulière avant 1970 avec une nette diminution de la fréquence des naissances le dimanche après 1995.

Cela semble confirmer l'hypothèse émise.

3ème partie: modélisation du problème (annexe 4)

En utilisant la fonction « ALEA.ENTRE.BORNES » du tableur, nous avons construit une série de données aléatoires modélisant le problème : une suite de nombres entiers compris entre 1 et 7, le 1 désignant le dimanche, 2 le lundi etc...

Dans un premier temps, notre série était de même effectif que la série de données récoltées par sondage. Mais à la lecture des différents résultats (tableau des effectifs et des fréquences ainsi que les diagrammes) nous nous sommes aperçus que la répartition n'était pas plus régulière que celle observée sur les données sondées.

Je leur ai alors fourni un fichier légèrement modifié pour qu'ils puissent générer rapidement une nouvelle série aléatoire (touche F9) et observer les différents diagrammes se modifier en conséquence.

Nous nous sommes aperçus que la série aléatoire finissait par « bien » se comporter au bout d'un très grand nombre de données, et qu'elle semblait se stabiliser autour d'une fréquence de 0,14 ou $1/7$.

Ce travail nous a permis de dégager les deux conclusions suivantes:

1. le « paradoxe » du hasard : on observe une certaine régularité dans les données choisies au hasard.
2. celle-ci apparaît lorsqu'on traite un grand nombre de données.

Conclusions : (extraits de travaux d'élèves en annexe 5)

Suite à ce travail, j'ai fait remarqué aux élèves qu'il faut être prudent dans l'interprétation des résultats d'une étude statistique.

En effet, nous nous sommes rendu compte que la taille de l'échantillon étudié joue un rôle très important : avec un échantillon trop faible, rien ne nous permet d'affirmer, quelque soit la répartition des valeurs observées, que celle-ci n'est pas le fait du hasard.

Etant donné que l'étude s'est faite sur moins de 200 valeurs pour chaque période, on ne peut exclure l'effet du hasard sur la régularité observée avant 1970 tout comme sur l'irrégularité après 1995. On ne peut donc pas affirmer que notre échantillon comporte des naissances programmées malgré l'apparente cohérence des résultats de l'étude avec l'hypothèse émise.

Critiques :

Le travail de compilation des données a été très fastidieux ; cette phase nous a pris beaucoup de temps et a fait perdre de vue le fil directeur de l'étude à certains enfants. Il faudra réfléchir à une méthode plus efficace pour regrouper toutes les données.

J'ai mené ce travail de façon linéaire, comme présenté dans ce document, mais il pourrait être intéressant suite aux conclusions de la 1ère partie de le scinder en deux:

- un premier groupe d'élève travaillant sur l'interprétation des résultats.
- un deuxième sur la série aléatoire.

Chaque groupe pourrait alors présenter les résultats de ses recherches sous la forme souhaitée à l'autre groupe. Objectifs : favoriser les échanges et mettre en place une démarche scientifique de recherche et de remise en question.

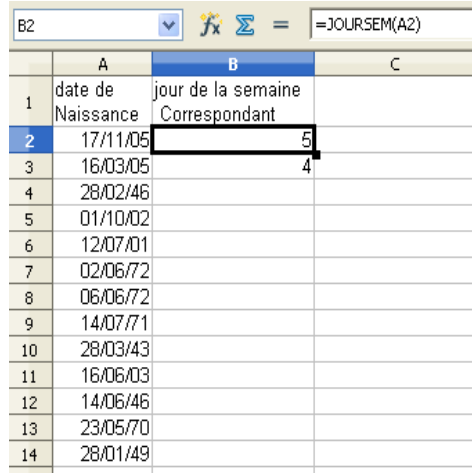

Annexe 1

ETUDE STATISTIQUE

Nous avons récolté les dates de naissance de deux populations différentes : des collégiens et des personnes faisant partie de la génération de leur parents ou grands-parents.

L'étude portera sur le jour de la semaine de leur naissance (est-ce un mardi, un dimanche...).

PREMIERE PARTIE

<i>Exercice préliminaire : trouver le jour de la semaine correspondant à une date précise.</i>	Ouvre une feuille de calcul open.office.org calc . Enregistre-la dans le dossier de la classe sous le nom jour de naissance . Dans la case S1 tape 4/01/2010 . Dans S2 =JOURSEM(S1).
A quel jour de la semaine correspond chaque chiffre?	
<i>Construire une série statistique à partir des données récoltées en trouvant le jour de la semaine correspondant à chaque date de naissance.</i>	Recopie le tableau suivant dans ta feuille de calcul. Etends la formule B2 jusqu'en B14. Complète le tableau avec tes propres données récoltées. 
<i>Ranger les jours de la semaine par ordre décroissant</i>	Sélectionne la colonne B et clique dans la barre d'outils sur l'icône 
<i>Que constates-tu? Compare avec les données d'un camarade. Le nombre de données récoltées est-il suffisant pour en tirer une conclusion générale?</i>	

DEUXIEME PARTIE

Regroupe dans un même tableau les données de tous tes camarades.

Que constates-tu?

Le nombre de données récoltées est-il suffisant maintenant pour en tirer une conclusion générale?

Annexe 2

Questionnaire :

I-Etude des données traitées.

1. Observe le tableau et le diagramme de ta classe.

Quelles constatations peux-tu faire sur la répartition des jours dans la semaine ?

Ceci correspond-il à ce que tu attendais ? Pourquoi ?

Qu'est-ce qui pourrait expliquer cela ?

2. Voici l'extrait d'un compte rendu d'une conférence du *Collège national des gynécologues et obstétriciens français* du 30 novembre 1995

QUELLE EST LA PLACE DU DÉCLENCHEMENT [DES ACCOUCHEMENTS] EN FRANCE ET DANS LE MONDE?

En France, de 1972 à 1995, la fréquence des déclenchements a varié de 8,5 à 20%, mais il existe une disparité importante suivant les régions et suivant les établissements.

La fréquence des déclenchements pour indications médicales est stable de 1988 à 1994: 11,2%; mais ceux sans indication médicale ont augmenté progressivement de 4,2 à 7,3 %.

A l'étranger, dans les années 80, la fréquence était pour: le Danemark, de 12%, la Pologne, de 15 %, la Norvège, de 17,5 %, le Pays de Galles, de 36%, la Finlande, de 17 à 29%, alors qu'elle était de 10,4% en France

Sources : <http://www.gyneweb.fr/sources/obstetrique/consens.htm>

Quelle information supplémentaire cet article nous fournit-il ?

A partir de quelle date les déclenchements d'accouchement sont-ils devenus plus fréquents ?

Comment modifier en conséquence la façon dont nous avons traité les données ?

3. Dans la cellule B4 du tableau, modifie la formule pour ne prendre en compte que les dates supérieures à 1995 : $\$N\2 devient $\$N\63 (seul le n° de la ligne à partir de laquelle on compte les valeurs est modifié).

Qu'observes-tu ? _____

4. Observe les tableaux qui regroupent les données de trois classes.

Quelles remarques peux-tu faire ?

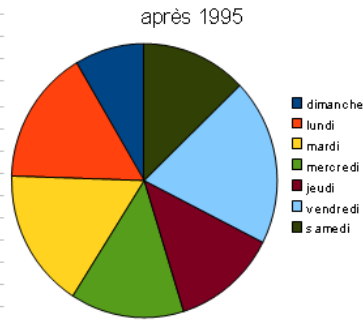
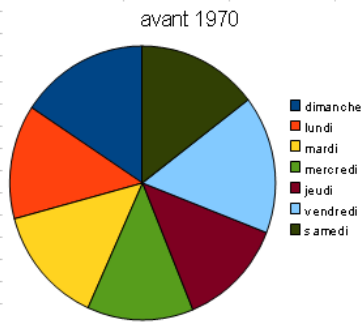
Annexe 3

Tableau de répartition des naissances dans la semaine avant 1970

jour de la semaine	dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi	total
	1	2	3	4	5	6	7	
effectif	28	24	25	23	23	29	26	178
fréquence	0,16	0,13	0,14	0,13	0,13	0,16	0,15	1

Tableau de répartition des naissances dans la semaine après 1995

jour de la semaine	dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi	total
	1	2	3	4	5	6	7	
effectif	16	30	31	26	24	37	24	188
fréquence	0,09	0,16	0,16	0,14	0,13	0,2	0,13	1



Date de Naissance sem
 01/08/12
 21/09/14
 11/11/18
 01/06/20
 18/03/23
 03/05/25
 25/05/25
 03/03/28
 20/04/28
 16/04/30
 20/05/30
 20/08/30
 22/01/31
 17/04/31
 29/05/32
 08/08/32
 24/12/32
 04/02/33
 17/10/33
 10/12/33
 24/03/34
 24/04/34
 22/10/34
 08/03/35
 28/06/35
 12/11/35
 13/03/36
 17/07/36
 27/02/37
 07/03/37
 14/12/37
 12/10/38
 13/04/39
 31/07/39
 26/12/39
 03/03/40

Annexe 4

Etude de données aléatoires.

I- Construire une série aléatoire modélisant l'étude

<i>Choisir au hasard 100 nombres compris entre 1 et 7.</i>	Dans le même fichier, ouvre une nouvelle feuille. Dans la cellule A1 tape =ALEA.ENTRE.BORNES(1;7) Etends cette formule verticalement jusqu'en A100
<i>Compter le nombre de fois où apparaissent les différentes valeurs de 1 à 7, et regrouper ces données dans un tableau.</i>	Copie le tableau ci-dessous. Dans la cellule D3 tape la formule =NB.SI(\$A\$1:\$A\$100;D2) Etends cette formule horizontalement. Dans la cellule K3 calcule la somme de D3 à J3.

II – Etude des résultats obtenus:

Dans le fichier tableur que ton professeur t'a donné, observe les tableaux situés dans la feuille « données aléatoires ».

Si tu appuies sur la touche « F9 » de ton clavier (à utiliser avec modération !), tu pourras modifier les données.

1) Comment se comportent les données dans les tableaux 1 et 2 ?

Appuie sur la touche « F9 » ; les données sont alors recalculées. Qu' observes-tu ?

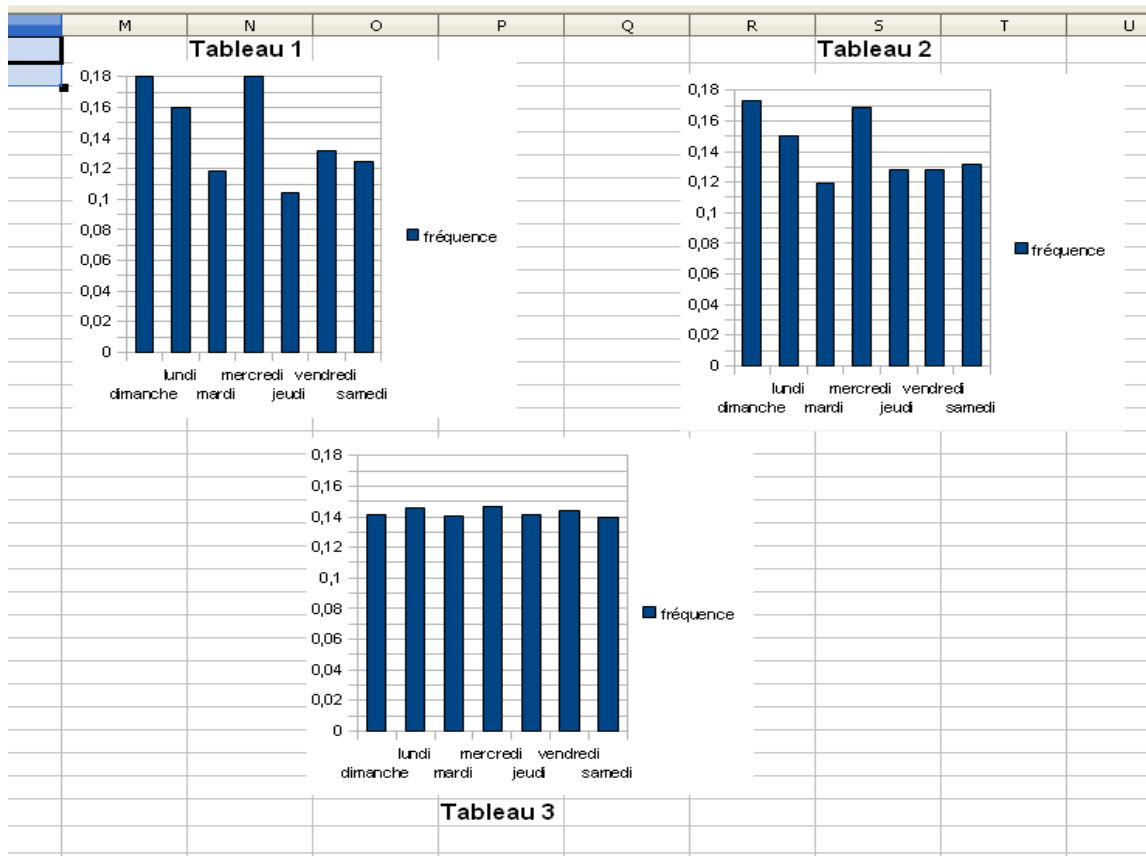
Comment devraient-elles se comporter, sachant que le choix des dates se fait entièrement au hasard ? _____

2) Observe maintenant le tableau 3 et les diagrammes associés ; appuie sur « F9 ». Quelles remarques peux-tu faire ? _____

3) Le nombre de données récoltées lors du sondage est-il suffisant pour tirer des conclusions de cette étude?

K1 Σ = =ENT(ALEA()*7+1)

	A	B	C	D	E	F	G	H	I	J	K
1	jour de la s*	dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi			4
2		1	2	3	4	5	6	7			1
3	effectif	26	23	17	26	15	19	18	144		4
4	fréquence	0,18	0,16	0,12	0,18	0,1	0,13	0,13	1		5
5											6
6											2
7											3
8											1
9											7
10	jour de la s*	dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi			5
11		1	2	3	4	5	6	7			1
12	effectif	84	73	58	82	62	62	64	485		4
13	fréquence	0,17	0,15	0,12	0,17	0,13	0,13	0,13	1		3
14											3
15											4
16											5
17											7
18											7
19	jour de la s*	dimanche	lundi	mardi	mercredi	jeudi	vendredi	samedi			6
20		1	2	3	4	5	6	7			5
21	effectif	562	577	558	583	562	570	555	3967		7
22	fréquence	0,14	0,15	0,14	0,15	0,14	0,14	0,14	1		7
23											7
24											3
25											7
26											7
27											4
28											1



Annexe 5

Extrait 1

Etudes statistique sur les naissances

Nous avons effectué l'étude suivante :

Dans les tableaux 1 et 2, les données sont peu nombreuses et donc les fréquences sont inégales alors que si l'on choisi environ 4 000 données (aléatoirement) les fréquences sont quasiment les mêmes.

Si l'on recalcule toutes les données, moins il y en a et plus les fréquences sont désordonnées, plus il y en a et plus les fréquences sont proches. Donc il faut au moins 4 000 données pour pouvoir conclure une étude statistique sur les jours de naissance.

Extrait 2

I- A partir des tableaux et des diagrammes de 4C, on constate que la répartition des jours de naissances est très irrégulière.

Nous nous attendions a des résultats plus réguliers.

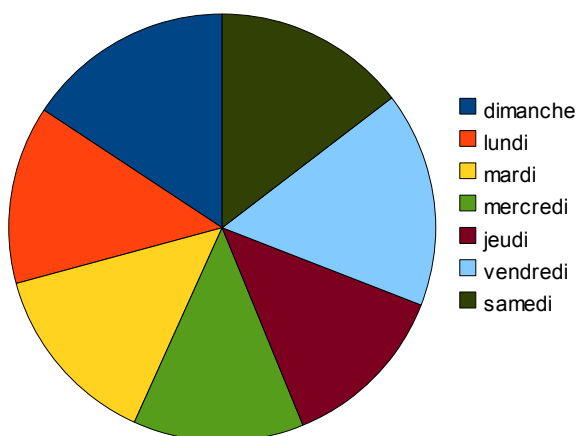
Il faudrait récolté plus de date de naissances.

Extrait 3

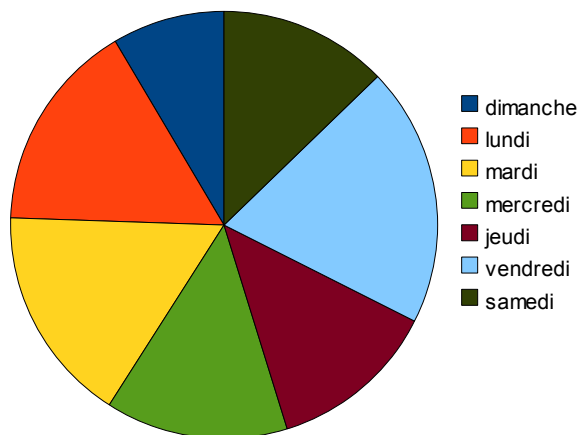
Résumé des maths atelier

D'abord, on a cherché des dates à inscrire puis on les a toutes rangées dans un tableau et on a cherché dans quelle jour il ya avait le plus de naissances, on a fait un graphique et on l'a visualisé. On a regroupé toutes les données des 4eB, des 4eC et des 4eE puis on a fait 3 tableaux pour chercher d'abord la fréquence des naissances avant 1970 puis ensuite après 1995 et pour finir avec toutes les valeurs

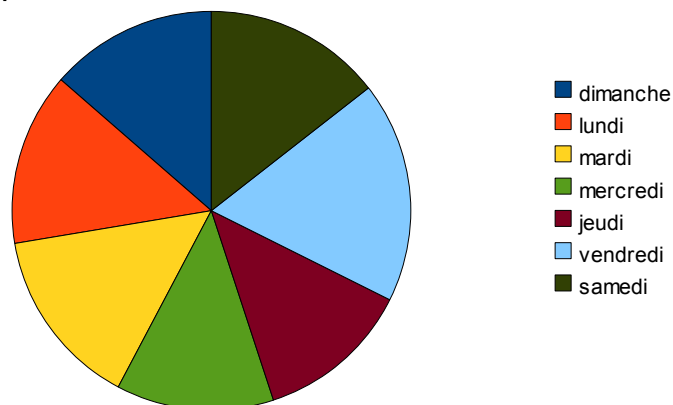
Répartition avant 1970



Répartition après 1995



répartition des naissances dans la semaine



.On remarque que la part du dimanche est plus faible dans le tableau après 1995 que avant 1970 donc cela confirme l'hypothèse que le déclenchement a une influence sur le jour de naissance.