L'espérance de vie des médecins

(Didier Modulo 2nde, édition 2004, page 165)

Un journaliste a constaté que, dans son pays, l'espérance de vie d'un médecin est d'environ 75 ans alors que l'espérance de vie d'un homme, en général, est de 68 ans.
Voici un échantillon des durées de vie de 100 hommes :

Durée de vie	0	10	20	30	40	50	60	70	80	90	100
Effectif	2	4	4	2	3	4	11	25	28	15	2

1. L'espérance de vie est la durée de vie moyenne des individus d'une population donnée. Calculer l'espérance de vie pour cet échantillon.
2. Un médecin ne commence à exercer qu'aux alentours de 30 ans. Pour comparer les espérances de vie, on ne considère donc que les hommes dont la durée de vie est supérieure ou égale à 30 ans.
a) Calculer cette nouvelle espérance de vie.
b) Trouvez-vous une différence notable pour cet échantillon par rapport à l'espérance de vie des médecins ?

Cet énoncé est proposé en seconde avec la consigne de rédiger les commentaires qu'il suscite.
Voici la copie d'une lycéenne de 2nde.

Qu'est-ce qu'une espérance de vie ? Ah ! la définition est donnée à la question 1 :
"L'espérance de vie est la durée de vie moyenne des individus d'une population donnée"
Il faut donc attendre que tous les individus de la population soient morts pour calculer cette durée de vie moyenne. Comme il y a de plus en plus de centenaires, il s'agit donc d'une population des années 1880 à 1890. C'est peut-être pour ça que l'on parle d'*hommes*, ce doit être des êtres humains de sexe masculin car à l'époque les femmes ne pouvaient pas être médecins.
Même si l'on prend la population des hommes (masculins) de 1880 à 1890, il faut connaître leur date de naissance et leur date de décès. Les plus âgés ont dû naître dans les années 1800 et les plus jeunes dans la période 1880-1890. Finalement, qu'est-ce qu'une "population donnée" ou comment on se donne une population ? et comment recueillir l'information pour calculer la durée de vie de cette population ?

Passons, ils ont pris 100 hommes décédés et il s'agit de calculer la durée de vie moyenne. C'est étonnant, ils meurent par paquets : 2 à la naissance, 4 à l'âge de 10 ans, 4 à l'âge de 20 ans … non, ce doit être des durées regroupées en classes et les âges 10, 20, …, 100 sont les centres des classes. Il s'agit donc des classes de 5 à 15 ans, de 15 à 25 ans, etc.
Non, cela ne va pas non plus car 0 est le centre de quelle classe ? Au lieu de 0, ce serait bien d'avoir 2.5 ans, centre de la classe 0 à 5.

Bon, le temps passe. Que veut le prof ? Un calcul de moyenne. Allons-y.

1) Durée de vie moyenne des 100 hommes de l'échantillon

$$\frac{1}{100}\left(2\times 0+4\times 10+4\times 20+2\times 30+3\times 40+4\times 50+11\times 60+25\times 70+28\times 80+15\times 90+2\times 100\right)$$

La durée de vie moyenne de cet échantillon de 100 hommes est 67 ans (tout rond !)

2) Durée de vie moyenne des hommes de l'échantillon dont la durée de vie est supérieure ou égale à 30 ans

Combien sont-ils ? On ne devrait pas compter les 2 qui sont morts à 30 ans, car on cherche la durée de vie moyenne qu'il reste à vivre pour ceux qui sont encore en vie à 30 ans. Il y en a 3 qui vivent encore 10 ans, 4 qui vivent encore 20 ans, …, 2 qui vivent encore 70 ans, soit en tout 88 hommes.
Cela donne une durée de vie moyenne pour ces 88 hommes qui vivent plus de 30 ans de :

$$\frac{1}{88}\left(3\times 40+4\times 50+11\times 60+25\times 70+28\times 80+15\times 90+2\times 100\right) = 74.1$$

La durée de vie moyenne des hommes qui vivent plus de 30 ans est d'environ 74 ans (à 30 ans ils ont encore 44 ans en moyenne à vivre).

Quand même, pourquoi est-il écrit que l'on doit prendre les hommes dont la durée de vie est supérieure *ou égale* à 30 ans ? Sans doute faut-il les mettre ces deux qui sont morts à 30 ans.
La durée de vie moyenne des 90 hommes dont la durée de vie est supérieure ou égale à 30 ans est : $\frac{1}{90}\left(2\times 30+3\times 40+4\times 50+11\times 60+25\times 70+28\times 80+15\times 90+2\times 100\right) = 73.1$

Cette durée de vie moyenne est d'environ 73 ans (les 90 hommes qui survivent à 20 ans ont encore 53 ans à vivre en moyenne).

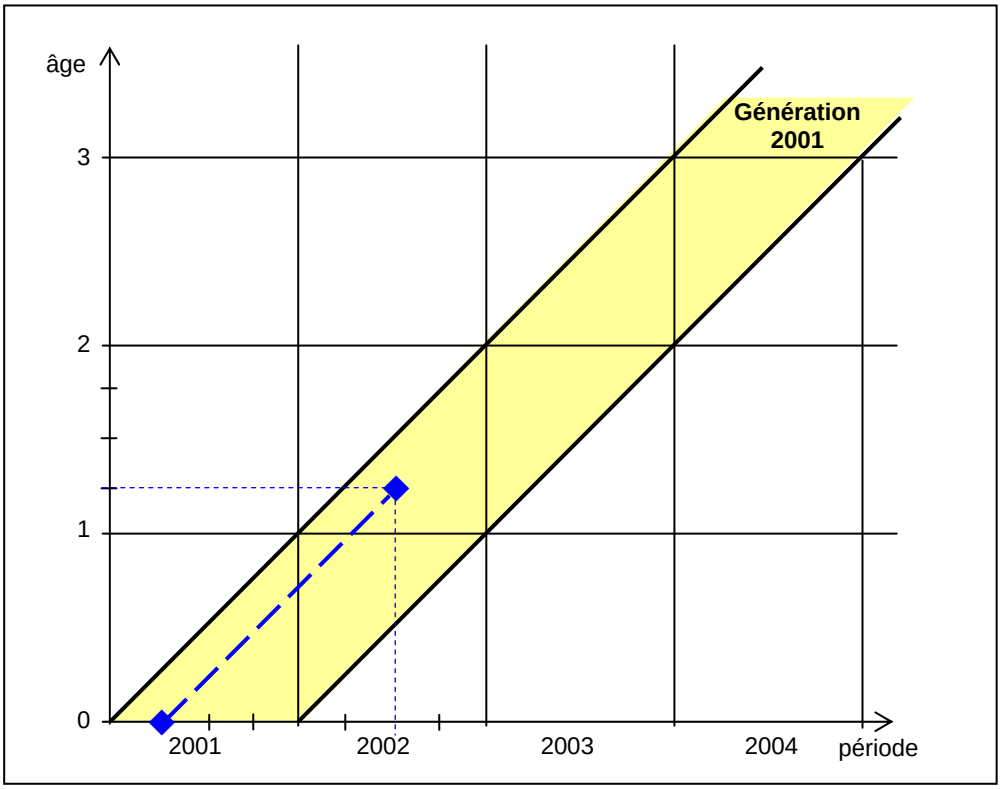
Dernière question : est-ce que je trouve une différence notable (notable ? significative ?) entre 75 ans (espérance de vie des médecins) et 74 ans (espérance de vie des 88 hommes de l'échantillon qui ont vécu plus de 30 ans) ?
J'avais cru comprendre que les moyennes varient d'un échantillon de taille 88 à un autre et qu'il faut étudier cette variabilité pour répondre à ce type de question !

Un peu de démographie

L'espérance de vie recouvre en fait deux notions :
- la durée de vie moyenne d'une génération (ensemble des personnes d'un pays ou une région nées la même année) complètement éteinte (toutes les personnes sont mortes)
- un indice rattaché à une année civile correspondant à la durée de vie moyenne d'une "génération" fictive qui parcourrait tous les taux de mortalité par âge observés cette année là.

Avant de présenter ces deux notions, voici le diagramme incontournable en démographie :

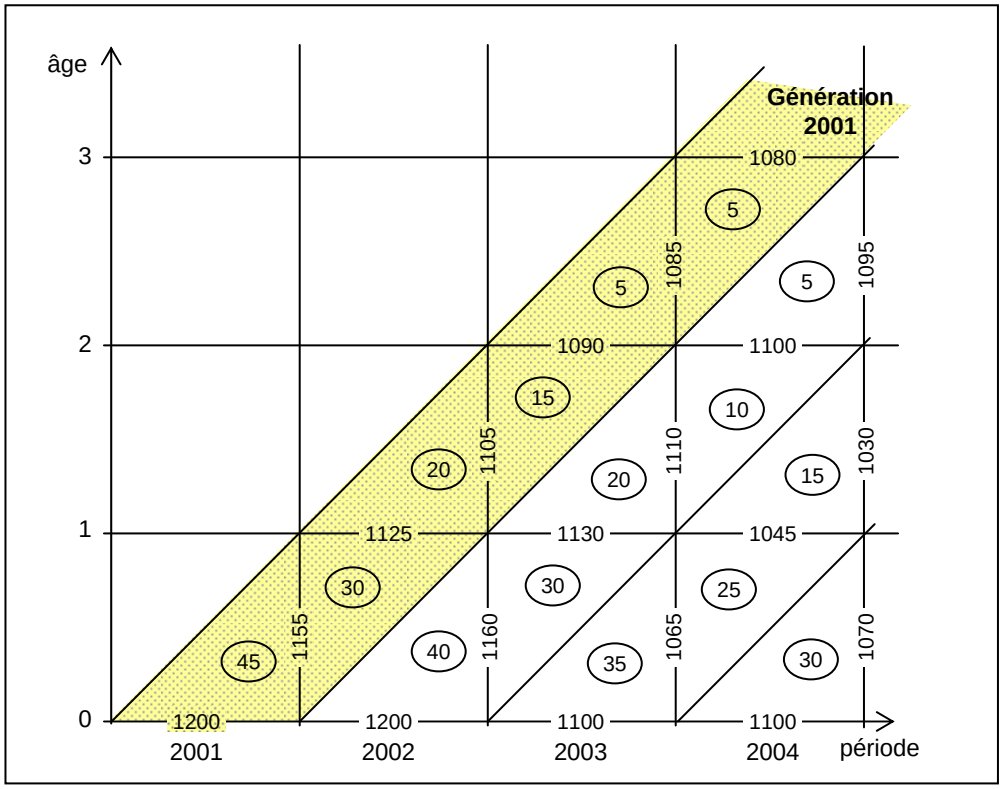
Diagramme de Lexis



Les années civiles ou périodes sont représentées en abscisse, les âges en ordonnées et les générations par des bandes parallèles à la première bissectrice. Le pointillé bleu correspond à la ligne de vie d'un individu né le 1er avril 2001 et décédé le 1er juillet 2002 à l'âge de 1an et 3 mois. Le 1er janvier 2002, cet individu est vivant et il a 9 mois, le 1er avril 2002, il rentre dans sa deuxième année de vie (on dit qu'il a 1 an révolu).
Lorsque l'on considère tous les individus d'une même génération (et d'un même pays ou région, on suppose qu'il n'y a pas de migrations), on place sur les segments les nombres de lignes de vie qui traversent ce segment et on place dans les triangles (entourés dans le graphique ci-dessous) les nombres de décès.

On présente les deux notions sur l'exemple suivant récapitulant les premiers âges de quatre générations (2001 à 2004) d'une population.

Espérance de vie à la naissance (et à deux ans) de la génération 2001



Génération 2001

Pour chaque âge en années révolues x (x entier allant de 0 jusqu'à l'âge au décès du dernier survivant), on note S_x le nombre de survivants, $d(x, x+1)$ le nombre de décès entre les âges x et $x + 1$ et q_x le quotient de mortalité défini par $q_x = \dfrac{d(x, x+1)}{S_x}$. La connaissance de S_0 et de l'une quelconque des trois séries permet de reconstruire l'ensemble des données.

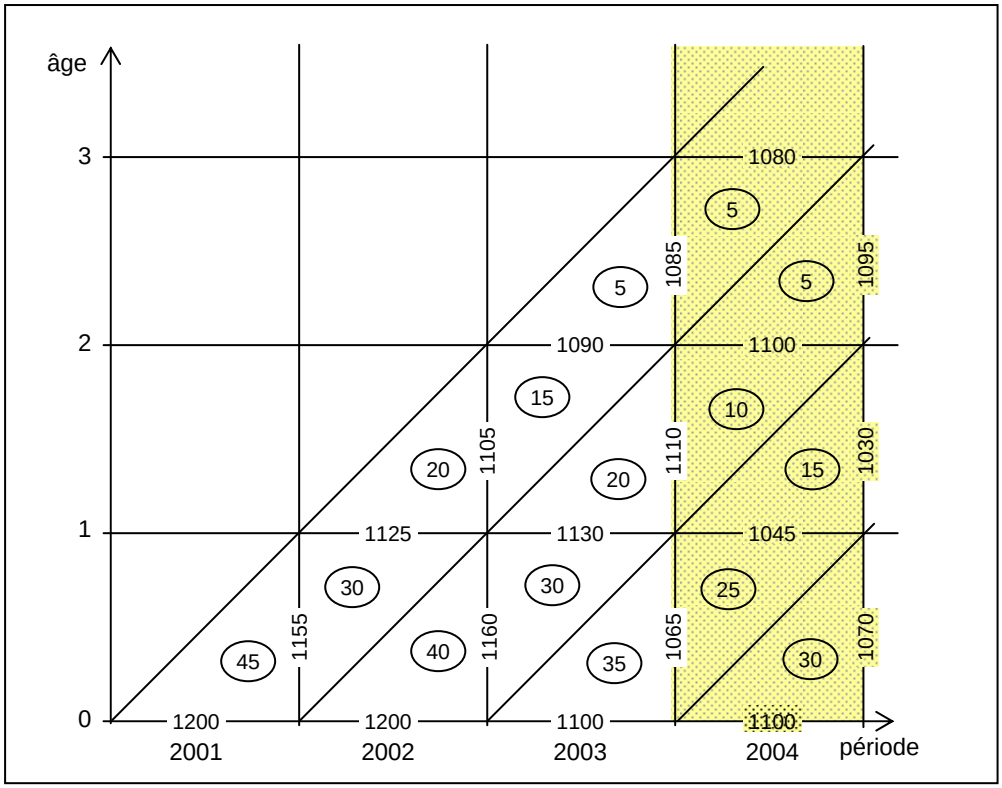
âge x	S_x	$d(x, x+1)$	q_x
0	1200	75	0.0625
1	1125	35	0.0311
2	1090	10	0.0092
3	1080	…	…
…	…	…	…

L'âge en années révolues de 2 ans représente un âge réel compris entre 2 et 3 ans, cet intervalle d'âge est représenté dans les calculs par son centre 2.5 ans.

La durée de vie moyenne de la génération 2001 sera alors calculée (lorsqu'elle sera complètement éteinte) de la façon suivante : $\dfrac{1}{1200}\left[0.5\times 75 + 1.5\times 35 + 2.5\times 10 + ...\right]$

La durée de vie moyenne à 2 ans de la génération 2001 sera alors calculée (lorsqu'elle sera complètement éteinte) de la façon suivante : $\dfrac{1}{1090}\left[2.5\times 10 + ...\right]$. C'est la durée de vie moyenne des 1090 individus de la génération qui ont survécu à l'âge de deux ans.

Espérance de vie à la naissance (et à 2 ans) pour l'année 2004



Année 2004

Pour chaque âge en années révolues *x*, on note $m(x,x+1)$ le nombre de décès à l'âge *x*, N_x la population moyenne d'âge *x* (moyenne entre le 1er janvier 2004 et le 1er janvier 2005, par exemple $N_0 = \frac{1}{2}(1065+1070) = 1067.5$) et t_x le taux de mortalité à l'âge *x* défini par $t_x = \frac{m(x,x+1)}{N_x}$. On construit ensuite la table de mortalité d'une génération fictive (de S_0 = 1000 individus par exemple) en posant $d(x,x+1) = S_x \times t_x$.

âge *x*	*m(x,x+1)*	N_x	t_x
0	55	1067.5	0.0515
1	25	1070	0.0234
2	10	1090	0.0092
3	…	…	…

âge *x*	S_x	*d(x,x+1)*
0	1000	51.5
1	948.5	22.2
2	926.3	8.5
3	917.8	…

L'espérance de vie à la naissance en 2004 est alors disponible dès l'année 2005 et égale à : $\frac{1}{1000}[0.5\times 51.5 + 1.5\times 22.2 + 2.5\times 8.5 + ...]$ et l'espérance de vie à 2 ans en 2004 est : $\frac{1}{926.3}[2.5\times 8.5 + ...]$

Commentaires

1) Finalement, l'espérance de vie moyenne d'une génération correspond bien à une durée de vie moyenne (statistique descriptive, moyenne arithmétique) alors que l'espérance de vie d'une année civile repose sur un modèle (génération fictive qui traverse tous les risques de mortalité) … Il ne s'agit pas pour autant de l'espérance mathématique d'une variable aléatoire, même si l'on peut interpréter le taux de mortalité à l'âge *x* comme la probabilité conditionnellement à la population d'âge *x* de décéder dans l'année.

2) Il est possible de travailler sur des groupes de 5 générations et des groupes d'âges de 5 années.

3) Pour étudier différents modèles de mortalité, il est plus commode d'utiliser une version continue de ces notions, l'âge *x* est un réel, $d(x) = -S'(x)$ et $q(x) = -\frac{S'(x)}{S(x)}$.

4) La première table de mortalité est publiée en 1662 par John Graunt (1620-1674) et construite à partir des bulletins de mortalité tenus à Londres depuis 1603. Il propose même un modèle.

> *Nous avons trouvé que, sur 100 individus conçus et animés, 36 environ meurent avant l'âge de 6 ans et peut-être un seul est survivant à 76 ans. Comme il y a 7 décennies entre 6 et 76, nous avons recherché six nombres moyens proportionnels entre 64, nombre de survivants à 6 ans, et 1, celui qui survit à 76 ans, et nous trouvons que les nombres suivants sont pratiquement assez proches de la vérité, car les hommes ne meurent pas selon des proportions exactes ni en fractions.*
> *Il s'ensuit que, sur ces 100 individus conçus, il en survit 64 au bout de 6 ans, 40 au bout de 16 ans, 25 au bout de 26 ans, 16 au bout de 36 ans, 10 au bout de 46 ans, 6 au bout de 56 ans, 3 au bout de 66 ans, 1 au bout de 76 ans et 0 au bout de 86 ans.*

Ce texte a fait couler beaucoup d'encre : cf. le livre "Histoires de probabilités et de statistiques" de la Commission Inter-IREM d'Épistémologie et d'Histoire des mathématiques, E. Barbin et J.-P. Lamarche coord., Ellipses, 2004. Il faut lire en particulier dans ce livre l'article de Bernard Parzysz sur une correspondance entre les frères Huygens à propos de cette table.

Espérance de vie

Définitions

Il ne faut pas confondre la durée de vie moyenne d'une population réelle complètement éteinte (c'est la moyenne des âges au décès des individus composant cette population) et l'espérance de vie d'une population donnée *une année donnée*, indice démographique dont la définition est donnée ci-après.
De plus, cela n'a pas de sens de considérer une population inter-générationnelle qui regrouperait des nouveaux nés et des centenaires. On s'intéressera à la durée de vie moyenne d'une *génération* (c'est-à-dire, un ensemble de personnes nées la même année civile, "la génération 1960" par exemple). On peut aussi s'intéresser à des sous-populations d'une génération donnée (les femmes de la génération 1960 par exemple).

L'espérance de vie à la naissance de l'année 2005 est la moyenne des durées de vie d'une *génération fictive* soumise aux conditions de *mortalité* de l'année 2005.
L'espérance de vie à 20 ans de l'année 2005 pour les femmes est la moyenne des durées de vie (après 20 ans) d'une génération fictive de femmes ayant 20 ans début 2005 soumise, âge après âge, aux conditions de mortalité des femmes de l'année 2005.
L'âge *x* en démographie signifie âge *x* en *années révolues*, c'est-à-dire âge compris entre *x* (inclus) et *x* + 1 (exclu).
Le *taux de mortalité* des femmes à l'âge *x* pour l'année 2005 est le rapport du nombre de femmes décédées à l'âge *x* en 2005 sur le nombre moyen de femmes d'âge *x* en 2005 (demi-somme du nombre de femmes d'âge *x* au 1er janvier 2005 et au 31 décembre 2005).

Espérance de vie à divers âges (0, 20 et 60 ans) pour les femmes et pour les hommes, en France de 1990 à 2004.

	À la naissance		À 20 ans		À 60 ans	
	Femmes	Hommes	Femmes	Hommes	Femmes	Hommes
1990	81.0	72.7	61.8	53.9	24.2	19.0
1991	81.2	72.9	62.0	54.0	24.4	19.2
1992	81.5	73.2	62.2	54.2	24.7	19.4
1993	81.5	73.3	62.2	54.3	24.6	19.4
1994	81.9	73.7	62.6	54.6	25.0	19.7
1995	81.9	73.9	62.6	54.7	24.9	19.7
1996	82.1	74.1	62.7	54.9	25.0	19.7
1997	82.3	74.6	62.9	55.4	25.2	19.9
1998	82.4	74.8	63.0	55.5	25.3	20.0
1999	82.5	75.0	63.1	55.7	25.3	20.2
2000	82.8	75.3	63.4	56.1	25.6	20.4
2001	82.9	75.5	63.5	56.2	25.7	20.6
2002 (p)	83.0	75.8	63.6	56.5	25.8	20.8
2003 (p)	82.9	75.9	63.4	56.6	25.6	20.8
2004 (p)	83.8	76.7	64.3	57.4	26.5	21.5

(p) : données provisoires.
Champ : France métropolitaine.
Source : Insee, bilan démographique.