

Cours « Les principes de l'Analyse démographique » Master 1e année par Alexandre Avdeev,

Chapitre 1 Introduction : population, mouvement de la population et démographiques, notions de base

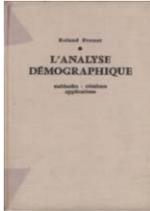
- Introduction, la bibliographie
- Notion (définition) de la population dans l'analyse démographique
- Indices (indicateurs) démographiques
 - *de composition de la population* → Indices de structure : rapports, proportions
 - *de dynamique de la population* → Indices d'occurrence : taux (hasard), quotient (probabilité, risque)
- Mouvement de la population : équation du bilan démographique (**équation fondamentale** du mouvement de la population)
 - Équation en termes de la masse (volume) et en termes des taux (valeurs réduites annualisées)
 - Approches des calculs des taux bruts (population exposée, hypothèses linéaires et exponentielle)
- Composition de la population : structures par âge et par sexe
 - Représentation graphique et l'analyse visuelle (pyramide démographique, rapport de masculinité, ratios de dépendance)
 - Indicateurs relatifs aux structures par sexe et par âge (âges moyen et médian, indices comparatifs, indice de vieillissement de Billeter)
 - Indicateurs associés à la qualité des données ou à la déclaration de l'âge (Indice de précision des Nation Unies, Indices de concentration (de préférence dans la déclaration des âges) des UNASAI, de Whipple, de Bachi et de Myers)

Tous les matériaux sont disponibles en libre d'accès du septembre au juillet sur les EPI de l'Université Paris 1 Panthéon Sorbonne

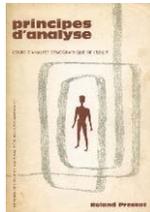
https://cours.univ-paris1.fr/fixe/IDUP_M1-Analyse-Demographique

et toute l'année sur le site <http://dmo.econ.msu.ru/Teaching/demo/>

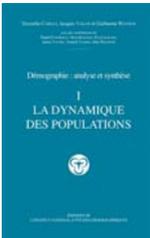
Manuels d'analyse démographique (en français) :



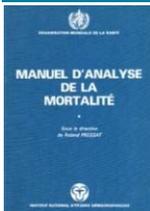
*** **Roland Pressat** – *L'analyse démographique : concepts, méthodes, résultats.*, 2e édition entièrement refondue. Paris: INED, 1969, 322 p. (première édition en 1961)



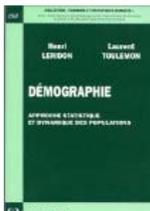
*** **Roland Pressat** – *Principes d'analyse : cours d'analyse démographique de l'I.D.U.P.* Paris: INED, 1966, 322 p. ([version PDF à télécharger](#))



*** **Louis Henry** – *Démographie. Analyse et modèles*, Paris: Édition de l'INED, 1984, 342 p.



** **Roland Pressat (dir.)** - *Manuel d'analyse de la mortalité*. Paris : Édition de l'INED, 1985, 164 p.

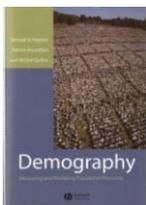


** **Henri Leridon et Laurent Toulemon**, *Démographie. – Approche statistique et dynamique des populations.*, Paris: Economica, 1997, 440 p.



** **Graziella Caselli, Jacques Vallin, Guillaume Wunsch (dir.)** - *Démographie. Analyse et synthèse. Volume I : « La dynamique des populations. »* Paris: INED, 2001, 552 p.

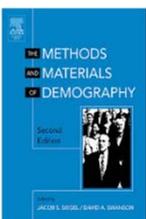
Manuels d'analyse démographique (en anglais) :



*** **Samuel H. Preston, Patrick Heuveline and Michel Guillot** – *Demography. Measuring and Modeling Population Processes*. Blackwell Publishing, 2000, 291 p.

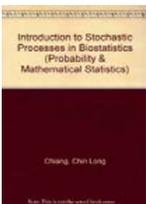
C'est un manuel de base pour ce cours, il est vivement recommandé pour l'acquisition (10-15 € sur Amazon)

« ... l'ouvrage... présente avec clarté et précision, en douze chapitres, les concepts et les méthodes permettant d'analyser la vie des populations » (*Population*, 2002, 57-3 pp. 591-592)



*** **Jacob Siegel David Swanson (eds)** - *The Methods and Materials of Demography*, 2nd Edition, January 2004, Elsevier Academic Press, 802 p. (ISBN: 0-12-641955-8)

Un manuel fondamental et pratique comportant les exemples d'analyse démographique appliquée aux populations des Etats Unis, disponible en PDF sur Web, vivement recommandé



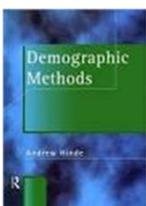
*** **Chin Long Chiang** - *Introduction to Stochastic Processes in Biostatistics*. John Wiley & Sons, 1968, 332 p. (Collection : Probability & Mathematical Statistics)

Un manuel fondamental d'analyse démographique (très) mathématisée, comportant notamment les principes d'estimations des statistiques des indicateurs de survie (variance, écart type) indispensable pour les comparaisons et la confirmation des hypothèses



** **Cox, Peter R.** *Demography* Cambridge, The University press, 1950, 326 p. (1st edition); 1970, 477 p. (4th edition)

« Manuel de démographie publié pour l'Institut des actuaires et la faculté des actuaires, il s'agit surtout des méthodes de démographie pure, mais aussi d'une évocation très sommaire de quelques problèmes d'économie démographique, tels que *population optimum*, charge des vieillards, etc.. La technique est fort bien exposée, mais date malheureusement quelque peu sur divers points. » (selon *Population*, 1953, 8-1 pp. 166-167)



** **Hinde, Andrew** - *Demographic methods*, London, UK. Hodder Arnold Publication, 1998, 320 p.

« Ce livre appartient à une catégorie d'ouvrages, assez rares, qui entendent faire le tour des principes et des méthodes permettant l'analyse statistique des populations, d'autant, nous dit l'auteur, qu'il se situe entre les ouvrages qui esquivent les vraies difficultés méthodologiques et ceux faisant appel à des connaissances mathématiques assez poussées (essentiellement, écrit-il, l'algèbre et le calcul matriciels). » (selon *Population*, 2001 56-3 pp. 478-479)



*** Manuel en ligne : *Population Analysis for Policies & Programmes*. Paris, Paris: International Union for the Scientific Study of Population (IUSSP). Disponible sur <https://papp.iussp.org/index.html>

Démographie : objet d'étude et méthodes

étudie l'évolution et la reproduction de l'effectif et des structures des *populations humaines* (et en occurrence celles non humaines, biologiques) dans le temps (et dans l'espace).

Population (*définition statistique*) : un ensemble des unités d'observation possédantes d'un même caractère (tous ce qu'on peut compter).

Caractères:

- physique (humains, sexe, âge),
- géographique (lieu de résidence p.ex. : population de la Europe),
- politique/administratifs (p.ex. : population de France/celle française),
- culturel (p.ex. : population francophone, chrétienne-protestante),
- social (p.ex. : population paysanne, agricole) etc...

« Ensemble » → une méta notion
(n'a pas de définition)

population démographique ≠ une population statistique

Population (*définition démographique*) : un ensemble des individus formant un système qui se reproduit (natalité versus mortalité; émigration versus immigration)

Analyse démographique : l'application *des méthodes quantitatives* (mathématiques, statistiques) aux études de la dynamique des populations humaines (et non humaines) et de ses composants (phénomènes démographiques).

≠ statistique(s)
de la population

Exercice de réflexion : présenter les populations statistiques et celles démographiques sur le diagramme de Venn

Vocation et objectifs de l'analyse démographique

- Observer, quantifier et comprendre les tendances des phénomènes démographiques (dynamiques) et de la composition des populations :
 - Statistique démographique/de la population : recensements, enquêtes, données de l'enregistrement d'état civil, registres etc.
 - Évènement démographique : **naissances** → Phénomène démographique : **natalité, fécondité** ;
 - Évènement démographique : **décès** → Phénomène démographique : **mortalité**, la durée moyenne de vie ;
 - Évènements : **mariage**, divorces (formation et la dissolution des couples) → Phénomènes : **nuptialité, divortialité** ;
 - Évènements : **changement du lieu de résidence** → Phénomènes : **migration**
- Construire les indicateurs adéquats aux phénomènes démographiques (statistiques et modèles)
- Distinguer les tendances magistrales des fluctuations conjoncturelles, l'essentiel du passager.
- Alimenter les sciences sociales (économie, sociologie, histoire, géographie etc.) et les politiques publiques pour appréhender et gérer les dynamiques sociétales
- Vocation majeur « instrumentale », utilitaire → **la prévision**: les perspectives et la projection de la population pour l'administration, pour la prise des décisions politiques, pour la gestion des biens, pour la protection sociale etc...
 - Produire des hypothèses sur la mortalité, la fécondité, la nuptialité et les migrations (qu'est-ce c'est ?)
 - Estimer les effectifs et des structures de la population par âge, par sexe, par l'état matrimonial
 - Estimer la distribution de la population par territoires
 - Estimer des structures et la taille des ménages etc.

Population dans l'analyse démographique (objet d'études)

- **Evolution historique** (une abstraction, un modèle, un nombre):
 - Avant XVI-XVII siècle : le nombre appliqué aux phénomènes sociaux n'était qu'une expression de la quantité des objets (des unités dénombrables)
 - **N** que l'on peut dénombrer (recensement), diviser, combiner, e.g., 5040 ménages de la population dans la cité idéale de Platon;
 - Après, – le nombre est une fonction → effectif de la population est une fonction de temps **P(t)** qui suit une loi mathématique e.g., la croissance géométrique de la population malthusienne
 - analyse mathématique appliquée aux phénomènes démographiques (croissance de l'effectif, mortalité, fécondité etc.)
- **Quatre pierres angulaires d'analyse démographique :**
 - La (les) loi(s) mathématique(s) de croissance de la population
 - Equation du bilan démographique, ou équation fondamentale du mouvement de la population
 - Cycle de reproduction (remplacement des générations)
 - Durée (analyse de durée ou de survie) : la durée de vie ; la durée des intervalles entre des événements démographiques (lois de la fécondité, de la mortalité, de la nuptialité et de migration) qui se résume en tables démographiques (tables de mortalité, tables des mariages etc.)
- Instruments (outils) méthodologiques indispensables fournis par la théorie de probabilité.

Au sommet de l'Analyse Démographique = la théorie de la population stable et l'équation intégrale de reproduction de la population (équation de Lotka)

F.R. Sharpe Ph.D. & A.J. Lotka M.A. "A problem in age-distribution" (L), *The Philosophical Magazine, Series 6, Volume 21, 1911, No 124, p.435-438*

- **Aujourd'hui: population – un système dynamique complexe**
 - On y ajoute des variables environnementaux, comportementaux, génétiques etc. : limites de la croissance, population logistique, l'idée de la saturation de l'environnement

1. La loi de croissance de la population : effectif varie avec le temps

Soit t – le temps discret,

$P(t)$ – effectif de la population au moment t ,

$k = \frac{P(t)}{P(t-1)}$ – taux (quotient) de croissance

$$\left. \begin{array}{l} P(t) = k \cdot P(t - 1) \\ \text{si } k \text{ est constant, pour } t \text{ intervalles } \rightarrow \end{array} \right\}$$

$$P(t) = k^t \cdot P(0) \quad (1)$$

$k < 1 \rightarrow$ diminution d'effectif

$k = 1 \rightarrow$ effectif stationnaire

$k > 1 \rightarrow$ croissance d'effectif

$$\frac{1}{t} \ln \frac{P(t)}{P(0)} = \ln k \rightarrow k = e^{t \ln \frac{P(t)}{P(0)}} = \exp\left(\frac{1}{t} \ln \frac{P(t)}{P(0)}\right)$$

Soit t – le temps continu, dP – la variation de l'effectif de la population avec dt – la variation infinitésimale de temps, on en déduit r – le taux (quotient) d'accroissement

$$r = \frac{dP}{dt} \cdot \frac{1}{P}$$

qui est une dérivée logarithmique $\frac{f'}{f} = (\ln f)'$ \rightarrow

$$P(t) = e^{rt} \cdot P(0) \quad (2)$$

$r < 0 \rightarrow$ diminution d'effectif

$k = 0 \rightarrow$ effectif stationnaire

$r > 1 \rightarrow$ croissance d'effectif

puisque (1) \equiv (2) $\rightarrow k = e^r \equiv k = \exp(r)$

soit b – taux de naissances

d – taux de décès

$$\left. \begin{array}{l} \text{soit } b \text{ – taux de naissances} \\ \text{soit } d \text{ – taux de décès} \end{array} \right\} \rightarrow \text{alors } r = b - d, \text{ si } b=0 \rightarrow P(t) = P(0) \cdot e^{-d \cdot t}$$

fonction de survie de la table de mortalité :
variation de l'effectif avec l'âge

Idee de base : la croissance est une propriété de la population \rightarrow population (effectif) est une fonction de temps (toutefois le problème de Fibonacci nous amène à une idée sur la reproduction biologique)

Exercice: l'an 2020 AC = l'an 7581 après la Création, la population mondiale = 7 795 millions, $r = ?$

2. Equation du bilan démographique (naissances, décès, migration)

Les populations (démographique) sont des ensembles (dans le sens statistique) des individus (humains ou non-humains) qui se reproduisent biologiquement : les individus produisent leurs semblables et disparaissent avec l'âge.

Ce sont les naissances qui alimentent l'augmentation de l'effectif d'une population démographique et les décès qui le réduisent (on parle de l'accroissement naturel).

On distingue toutefois:

les populations fermées, dont la variation de l'effectif dépend uniquement des naissances et des décès

$$P(t + \Delta t) = P(t) + B(\Delta t) - D(\Delta t)$$

les populations ouvertes dont les effectifs varient à causes de la migration: augmentées par l'immigration (in-migration) et diminuées de l'émigration (out-migration)

$$P(t + \Delta t) = P(t) + B(\Delta t) - D(\Delta t) + I(\Delta t) - E(\Delta t)$$

Symboles utilisés dans ces équations:

$P(t)$ et $P(t+\Delta t)$ – effectif au moment t et au moment $t + \Delta t$

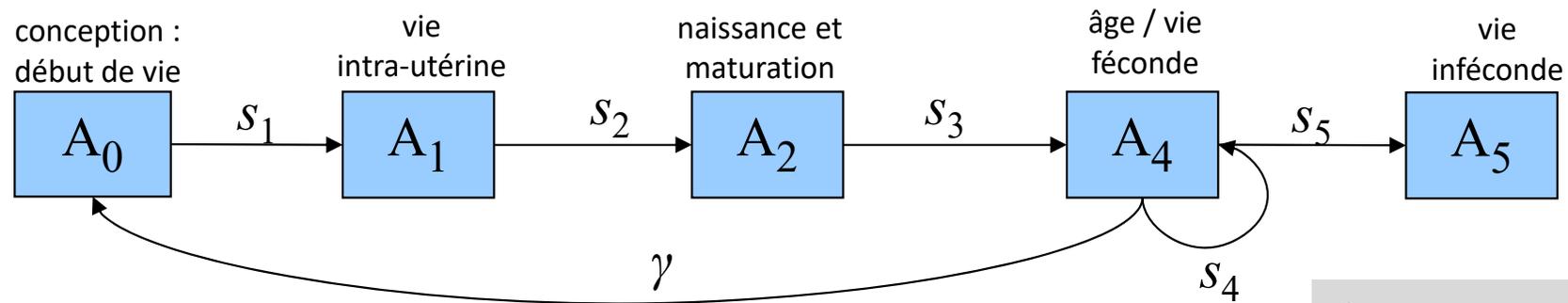
$B(\Delta t)$ – nombre de naissances pendant une période Δt

$D(\Delta t)$ – nombre de décès pendant une période Δt

$I(\Delta t)$ – nombre d'immigrés arrivant pendant une période Δt

$E(\Delta t)$ – nombre d'émigrants pendant une période Δt

3. Cycle de vie et la reproduction : processus démographiques comme transitions entre les états



$$F = \delta \gamma \varphi s_f$$

F ← fécondité est une fonction des arguments $\delta, \gamma, \varphi, s_f$

S_4 Question qui s'impose :
quelle est la durée de la
vie féconde ?

s_i – probabilité de survie dans l'intervalle (a_i, a_{i+1}) ;

s_f – probabilité de survie de la conception à l'âge de la maternité ;

φ – fertilité (*fécondité naturelle*) ;

γ – probabilité d'accouplement ;

δ – le rapport des sexes à la naissance (*à la conception*) ;

Nota : dans les populations non-humaines $s_5 \rightarrow 0$

Soit

a_m – âge moyen à la maternité ;

F – le nombre (moyen) de filles survivantes à l'âge a_m par femme ;

r – le taux d'accroissement de la population

$$r = \frac{1}{a_m} \cdot \ln F$$

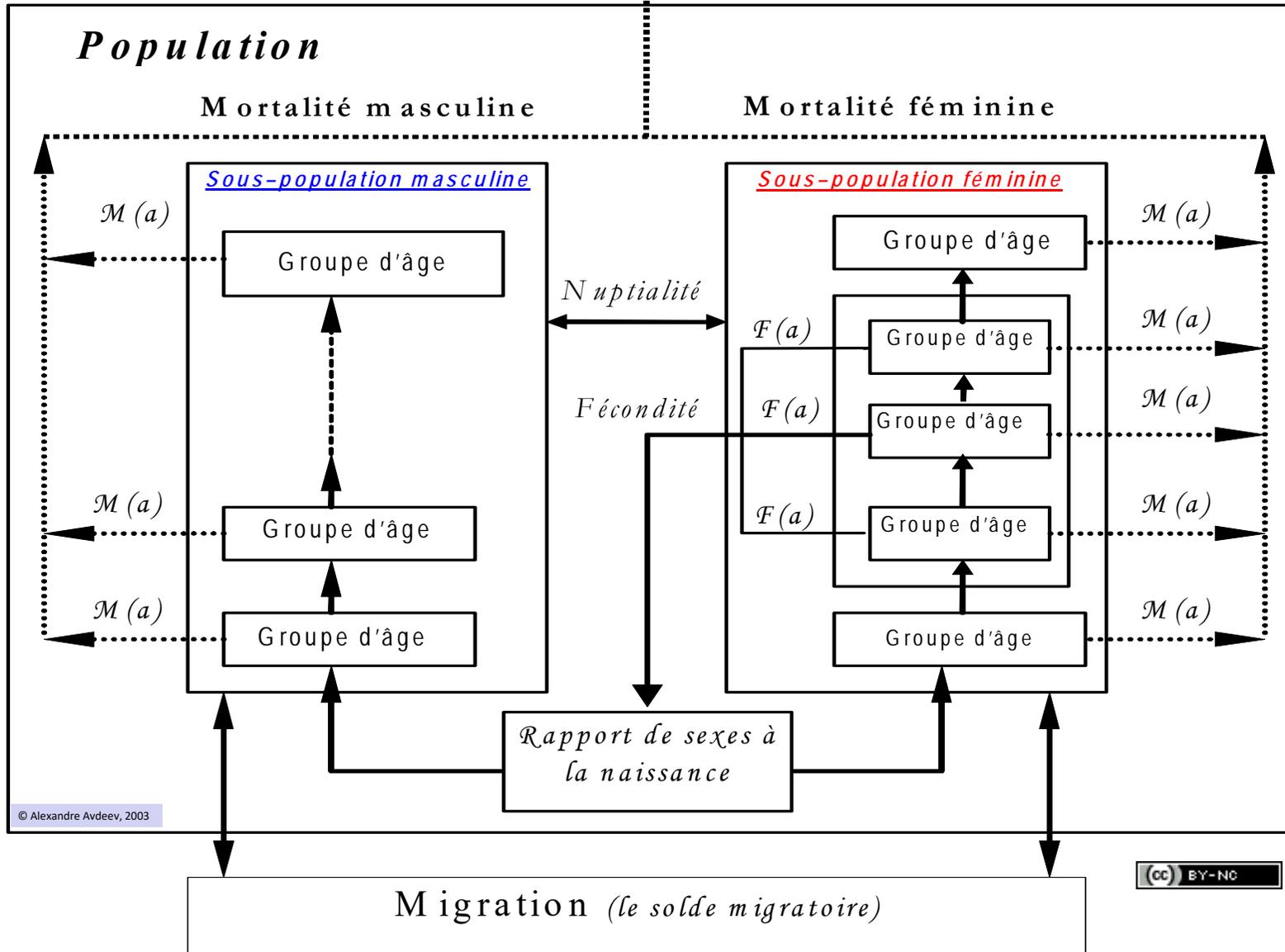
On peut donc déduire la loi de la croissance exponentielle du modèle des cycles de reproduction

Idée de base : la croissance est une propriété de l'espèce (de sa fécondité et de sa survie)
→ croissance est une mesure du remplacement des générations

4. La structure dynamique de la population :

les flux (passage) et les stocks (durée), la population comme un système d'autoreproduction

$$1 = \int_0^t e^{-\rho \cdot a} \cdot p(a) \cdot f(a) da$$



Sources d'information en démographie

1. Les recensements et les registres de la population → information sur l'effectif et structures :

- lieu et type de résidence (urbain, rural) ;
- âge, sexe, lieux de naissances, état matrimonial et nationalité des personnes ;
- langue maternelle et maîtrise de langues ;
- la durée de résidence sur le lieu du recensement et lieu de résidence lors du recensement précédent (il y 5 ans, etc.) des personnes ;
- composition des ménages (liens de parenté et les relations entre les membres de ménage) ;
- la fécondité accomplie (nombre de propres enfants nés et vivants) ;
- caractéristiques socio-économiques des personnes ;
- condition des logements ;
- état de santé (infirmité).

2. La statistique d'état civil → enregistrement des événements vitaux :

- naissance (date, lieu, nom et prénoms, sexe, l'ordre (biologique) de naissance, âge, noms, prénoms, lieu de naissances et nationalité des parents, leur statut matrimonial, leurs âges, activité et profession des parents, ...)
- décès (date, lieu, nom et prénoms, sexe, âge et/ou date de naissance, lieu de naissance, nationalité, cause(s) de décès, lieu de résidence, lieu de décès, statut matrimonial, activité, profession, etc...)
- mariage (date, lieu, noms, prénoms, sexes, dates et lieux de naissances, nationalité, états matrimoniaux antérieurs, résidence des époux, les enfants communs etc...)
- divorce / séparation (caractéristiques des conjoints, caractéristiques du mariage, enfants communs)

3. La statistique de migration internationale entrant:

- âge, sexe, lieu de naissance, nationalité, état matrimonial et familial, pays de départ, pays de naissances etc.

Quantifier la population et les phénomènes démographiques

- **Mesurer les masses** ou les volumes (effectif d'une population à un moment donné et le nombre d'événements démographiques repéré sur un intervalle de temps) :
 - Naissances (par rang) et décès (par cause)
 - Mariages (premiers et remariages) et divorces
 - Immigration – émigration; etc.
- **Mesurer la variation**, la composition ou la structure (effet de masse éliminé, effectifs réduits) :
 - Croissance : rapport numérique des effectifs de la même population aux différentes périodes historiques
 - Structure : rapport numérique entre les composantes d'une population ou des sous-populations
 - Par âge et par sexe (% des hommes et des femmes, % des groupes d'âge)
 - États matrimoniaux et familiaux, types de ménages
 - Complémentaires: nationalité, éducation, type du lieu d'habitation, CSP etc.
- **Mesurer les fréquences** des phénomènes démographiques (estimer la probabilité d'un événement démographique)
 - Quotients ou **la fréquence** des événements sur un intervalle de temps **réduite à l'effectif** de la population
 - Taux bruts ou **la fréquence réduite à l'effectif** de la population **et à la durée de la période** d'observation
 - Taux ajustés (comparatifs) ou **la fréquence réduite à l'effectif, à la durée de la période et à la composition des populations**
 - Indicateurs dérivés : la durée d'attente d'un événement (naissance, décès, mariage, divorce etc.)
 - Un problème de calculs (propre à la démographie) : comment accorder les mesures prises sur l'intervalle (événements) avec celles du moment (populations productrices) sachant qu'elles varient avec le temps

Les indices démographiques relatifs à la structure

1. Les rapports/ratios : $\frac{A}{B} \quad A \notin B \quad 0 \leq \frac{A}{B} < +\infty$

les ratios en démographie sont définis sur l'ensemble des nombres rationnels non négatifs

Exemples: - le rapport de masculinité

$$RM = \frac{\text{nombre d'hommes}}{\text{nombre de femmes}}$$

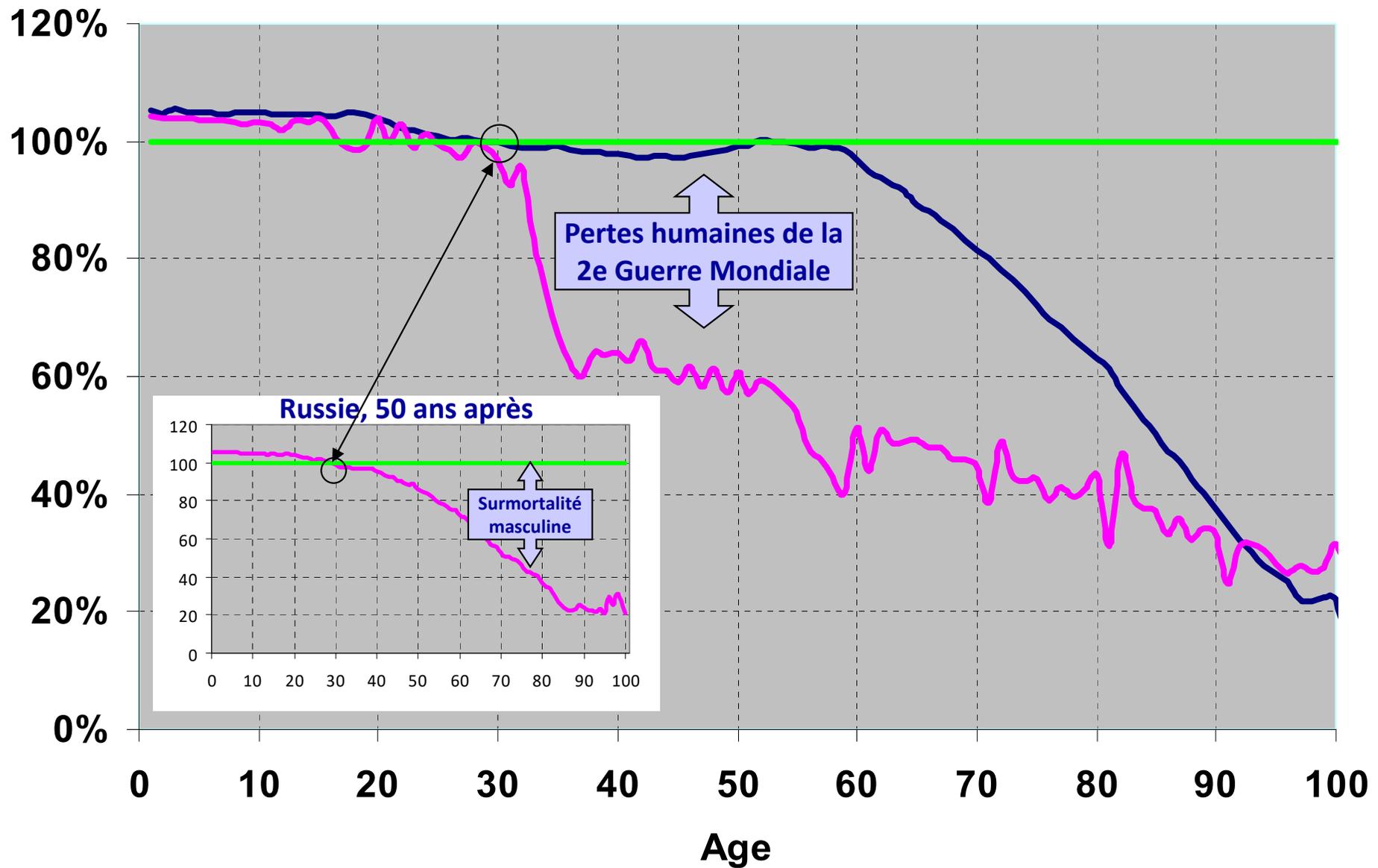
- la descendance des générations féminines dans une enquête rétrospective

$$DF_a = \frac{\text{nombre d'enfants nés vers l'âge "a"}}{\text{nombre de femmes à l'âge exacte "a"}}$$

- la « charge » démographique

$$CD = \frac{\text{Population de moins 15 et de plus de 65 ans}}{\text{Population active (15 - 65 ans)}}$$

Rapport de sexes en France en 2000 et en Russie en 1959



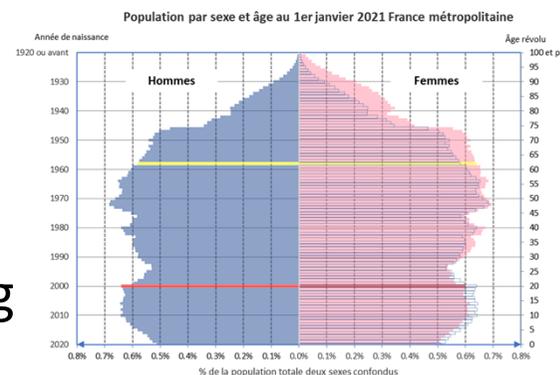
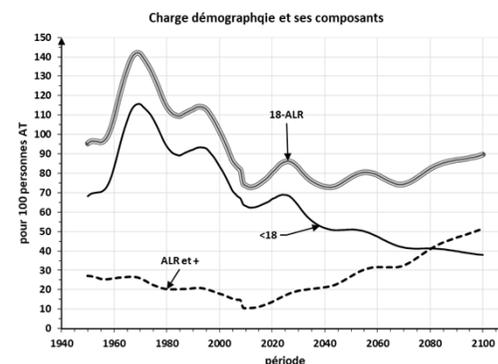
— France, 1.1.2000 — Russie, 1959 — Ligne d'égalité des sexes

Les indices démographiques relatifs à la structure (suite)

2. Proportions (parts) $0 \leq \frac{A}{A+B} \leq 1; A \text{ et } B \in C = A+B$

Exemples:

- ✓ % de la population active (que veut dire « active » ?)
- ✓ Structure de la population par groupes d'âge (pyramides démographiques)
- ✓ Proportion des survivants à l'âge « x » par rapport à leur effectif initial à l'âge « $x-1$ »
- ✓ Calendrier de la fécondité (nombre d'enfant nés vers l'âge « x » (à l'âge « x » par rapport au nombre d'enfant nés vers l'âge de 50 ans)



Les indices démographiques relatifs à la dynamique de la population

1. Indicateur de la variation : quotient (probabilité, risque = eng.: risk)

Les quotients montrent (quantifient) dans quelle mesure l'effectif à la fin d'une période est différent de celui au début de cette période → mesure d'avancement.

C'est un indicateur standardisé par rapport à l'effectif initial, mais il varie en fonction de la durée de l'intervalle de temps (période d'observation) → la variation divisée par l'effectif initial :

$$Q = \frac{P_{t+\Delta} - P_t}{P_t} = \frac{\Delta P_{\Delta t}}{P_t} = \frac{P_{t+\Delta}}{P_t} - 1 \quad \text{Il est toujours défini sur l'intervalle de } -1 \text{ à } +\infty$$

2. Indicateur de la vitesse de variation : taux (hasard, taux instantané = eng.: hazard rate)

Les taux montrent (quantifient) la variation de l'effectif (ou la production des événements démographiques) par rapport de l'unité de temps et de l'effectif de la population durant une période → mesure de la vitesse (moyenne).

C'est un indicateur standardisé par rapport à la variation de l'effectif et de la durée d'un intervalle de temps (période d'observation), il est donc stable par rapport à la durée d'observation (tendance centrale).

Le plus souvent il est estimé à partir de l'hypothèse sur l'accroissement uniforme (linéaire) sur l'intervalle considéré :

$$T = \frac{P_{t+\Delta} - P_t}{0,5 \cdot (P_t + P_{t+\Delta}) \cdot \Delta t} = \frac{\Delta P_{\Delta t}}{\overline{P}_{(t,t+\Delta t)} \cdot \Delta t}$$

Exercice de réflexion :

Développer le concept de quotient et celui de taux aux phénomènes démographiques (natalité-fécondité, mortalité)

Analyse du mouvement de la population (dynamique de l'effectif)

- Les flux et les stocks: équation fondamentale du mouvement de la population
- Taux bruts relative à la dynamique de la population

Flux et stocks

Équation fondamentale du bilan démographique (en termes de masse)

$$P_t = P_0 + N_{0,t} - D_{0,t} + I_{0,t} - E_{0,t}$$

P_t – nombre de personnes survivantes au moment t (*stock*)

P_0 – nombre de personnes survivantes au moment 0 (*stock*)

$N_{0,t}$ – nombre de naissances durant la période entre 0 et t (*flux*)

$D_{0,t}$ – nombre de décès durant la période entre 0 et t (*flux*)

$I_{0,t}$ – nombre migrants arrivés durant la période entre 0 et t (*flux*)

$E_{0,t}$ – nombre migrants partis durant la période entre 0 et t (*flux*)

$N_{0,t} - D_{0,t} \rightarrow$ excédent (déficit) naturel

$I_{0,t} - E_{0,t} \rightarrow$ excédent (déficit), ou solde migratoire = migration nette

$I_{0,t} + E_{0,t} \rightarrow$ migration brute

4. Équation du bilan démographique en masse et normalisée (en termes des taux dits « bruts »)

volume $P_t = P_0 + N_{0,t} - D_{0,t} + I_{0,t} - E_{0,t} \rightarrow$

L'idée : **décomposer** la variation (accroissement)

accroissement $P_t - P_0 = N_{0,t} - D_{0,t} + I_{0,t} - E_{0,t} \rightarrow$

vitesse de change $\frac{P_t - P_0}{\bar{P} \cdot t} = \frac{N_{0,t} - D_{0,t} + I_{0,t} - E_{0,t}}{\bar{P} \cdot t} \rightarrow$

L'information sur la masse/volume est perdue
(principe de réduction de l'information)

$$TBA_{0,t} = (TBN_{0,t} - TBM_{0,t}) + (TBI_{0,t} - TBE_{0,t}) \rightarrow$$

$$TBA_{0,t} = TAN_{0,t} + TBAM_{0,t}$$

Les taux bruts principaux

TBN – taux brut de natalité	}	Mouvement naturel	}	Deux types de flux
TBM – taux brut de mortalité				
TBI – taux brut d’immigration	}	Mouvement spatial (migration)		
TBE – taux brut d’émigration				

Ces taux bruts correspondent aux quatre composants de l'équation du bilan démographique

Principe de correspondance dans les calculs des taux :

Il faut toujours rapporter les événements (le nombre d'événement) à *une population étant effectivement exposée au risque des ces événements* = « population exposée »

Comparaison de deux populations : France et Danemark, 2000 et 2001

Tableau avec les valeurs absolus

Année	Pays	Ajustement Statistique	Décès	Population au 1er janvier	Population au 31 décembre	Population exposée	Naissances vivantes
2000	Danemark	596	57 986	5 330 020	5 348 616	5 339 318	67 084
	France	94 456	530 864	58 858 198	59 240 516	59 049 357	774 782
2001	Danemark	18	58 338	5 349 212	5 368 336	5 358 774	65 458
	France	94 455	531 073	59 266 572	59 642 350	59 454 461	770 945

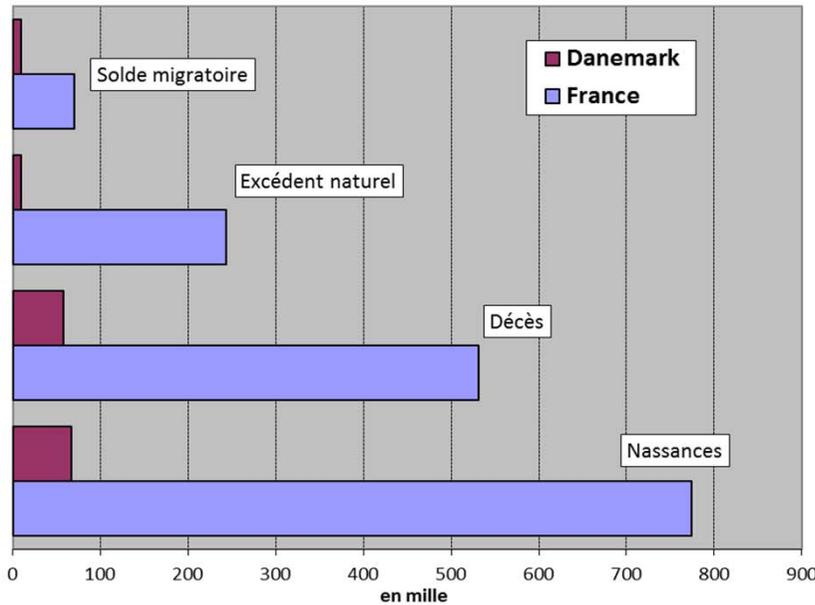
Tableau avec les taux ‰

Année	Pays	Taux de accroissement	Taux brut de mortalité	Taux brut de natalité	Taux d'accroissement naturel	Taux d'accroissement migratoire
2000	Danemark	3.59	10.86	12.56	1.7	1.78
	France	6.92	8.99	13.12	4.13	1.19
2001	Danemark	3.57	10.89	12.22	1.33	2.24
	France	7.05	8.93	12.97	4.03	1.43

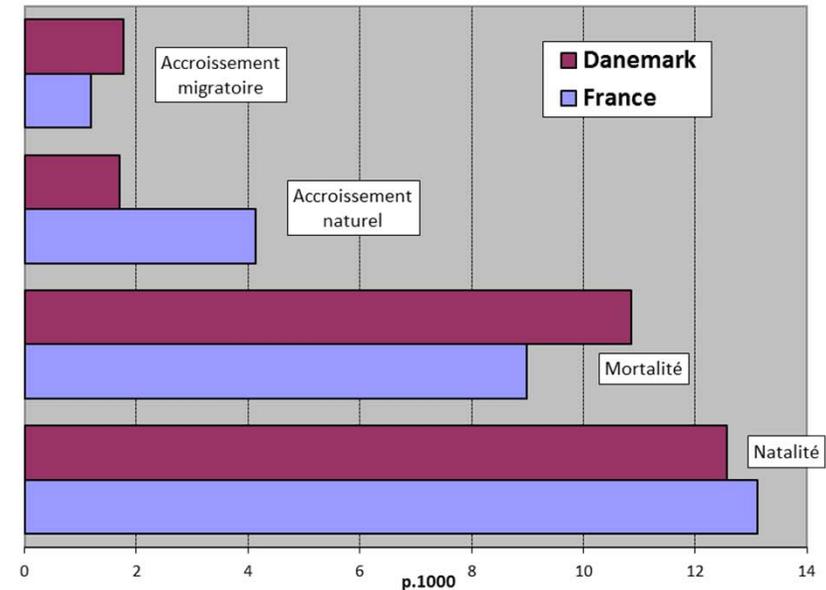
Comparaison (analyse) visuelle (sur un graphique)

Comparaison des valeurs absolus

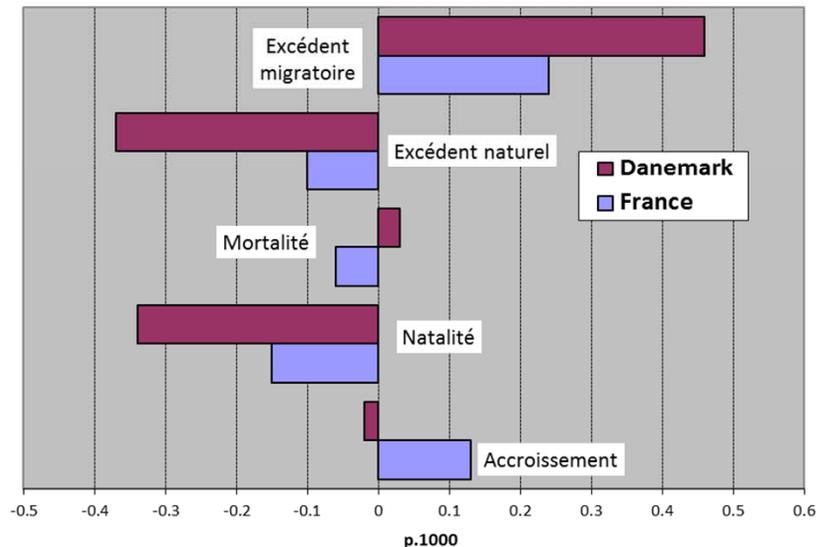
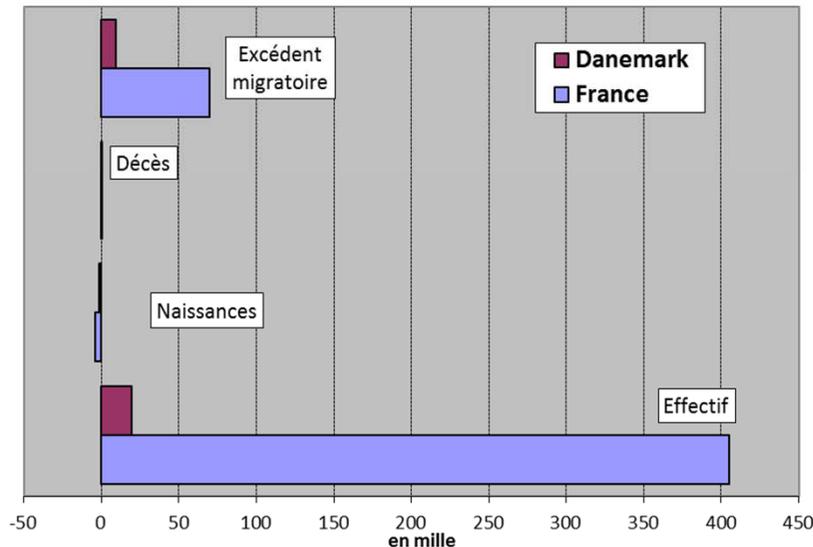
Situation statique
(deux populations, même période)



Comparaison des taux (ou valeurs standardisées)



Dynamique de la situation
(variation des indicateurs dans chaque population)



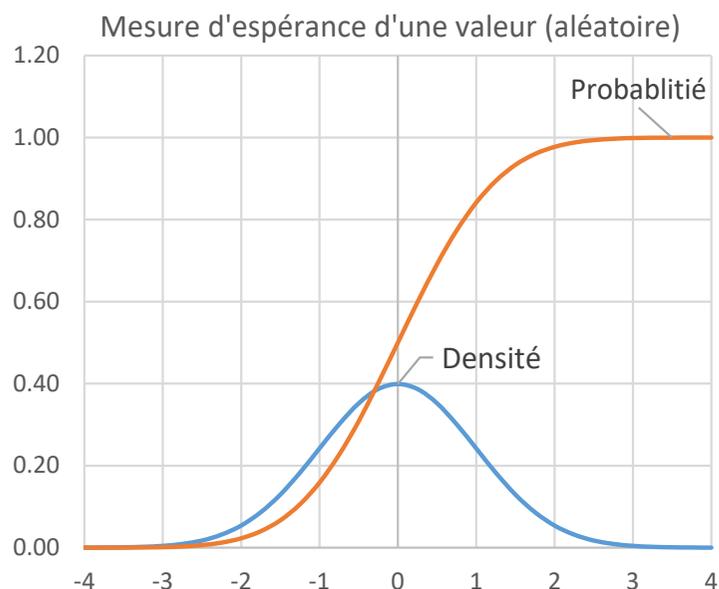
Indices démographiques relatifs à l'intensité des flux ou les indices d'occurrence

$$\text{Quotient} = \frac{\text{Nombre d'événements enregistrés (nombre d'occurrences)}}{\text{Population (exposée au risque) au début de la période d'observation}}$$

- convient pour les processus et les populations homogènes
- une proportion, une mesure de la variation (sous l'influence des événements non renouvelables), sinon l'interprétation devient difficile;
- son score varie en fonction de la durée d'une période d'observation
- statistiquement correspond à la distribution de risque ou de probabilité (risque relatif / probabilité tout court)

$$\text{Taux} = \frac{\text{Nombre d'événements enregistrés (nombre d'occurrences)}}{\text{Population (exposée au risque) pendant toute la période d'observation}}$$

Relation entre le taux et le quotient



- statistiquement correspond à la densité de risque relatif (ou la densité de risque, on dit aussi « hasard » ou « le taux d'échec »)
- application universelle (pour processus homogènes et hétérogènes)
- mesuré comme un risque moyen (ou instantané) de l'occurrence durant une période d'observation → une statistique non biaisée
- représente la tendance centrale = plus la période d'observation est longue → moins d'influence des aléas (ex. g. épidémies)
- population exposée = nombre de personnes-années vécues durant la période d'exposition →
taux démographique est toujours annualisé = réduit à une année d'exposition
(année – une unité conventionnelle, ni universelle, ni obligatoire)
- mathématiquement, le quotient sur l'intervalle $[x, x+n)$ ${}_n q_x = \int_{i=x}^{x+n} \tau(i)$
où $\tau(i)$ – la fonction de la densité sur cet intervalle

Calculs des taux : trois méthodes d'estimation de l'effectif de population exposée (nombre d'années vécues = N.A.V.)

$$\text{Taux} = \frac{\text{Nombre d'événements enregistrés (nombre d'occurrences)}}{\text{Effectif de la population exposée au risque}}$$

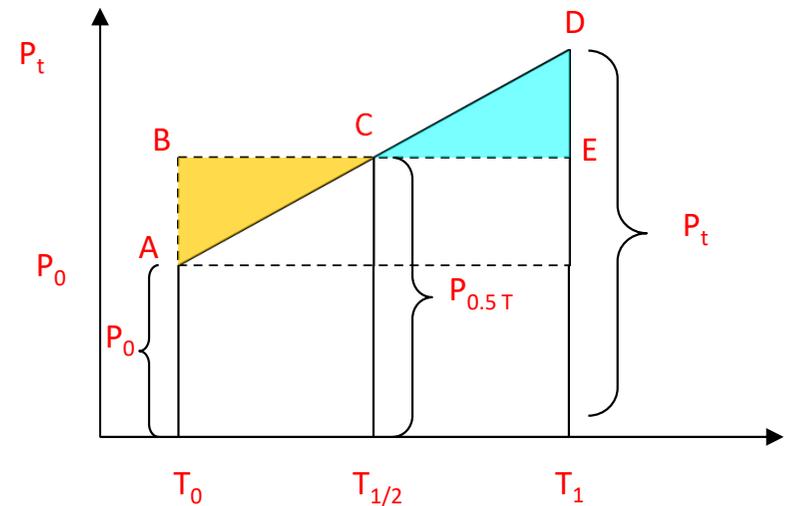
Problème : Comment peut-on estimer (trouver un estimateur pour) le dénominateur en absence des données individuelles sachant qu'il varie durant la période d'observation ?

Méthode 1 (le plus souvent utilisée)

$$\text{Calcul usuel des taux} \rightarrow \text{Taux} = \frac{E_{(t_0, t_T)}}{T \cdot \bar{P}_{(t_0, t_T)}}$$

$\bar{P}_{(t_0, t_1)}$ – la population dite « moyenne », ou dans ce cas la population au milieu de l'intervalle de temps entre t_0 et t_1 sur lequel on observe les événements E , estimée à partir d'une hypothèse de la croissance linéaire (progression arithmétique)
 (on dit « le taux centré », puisque il représente une probabilité instantanée juste au milieu de l'intervalle, en anglais « central rate »)

T – la durée de l'intervalle entre t_0 et $t_1 \rightarrow T = t_1 - t_0$

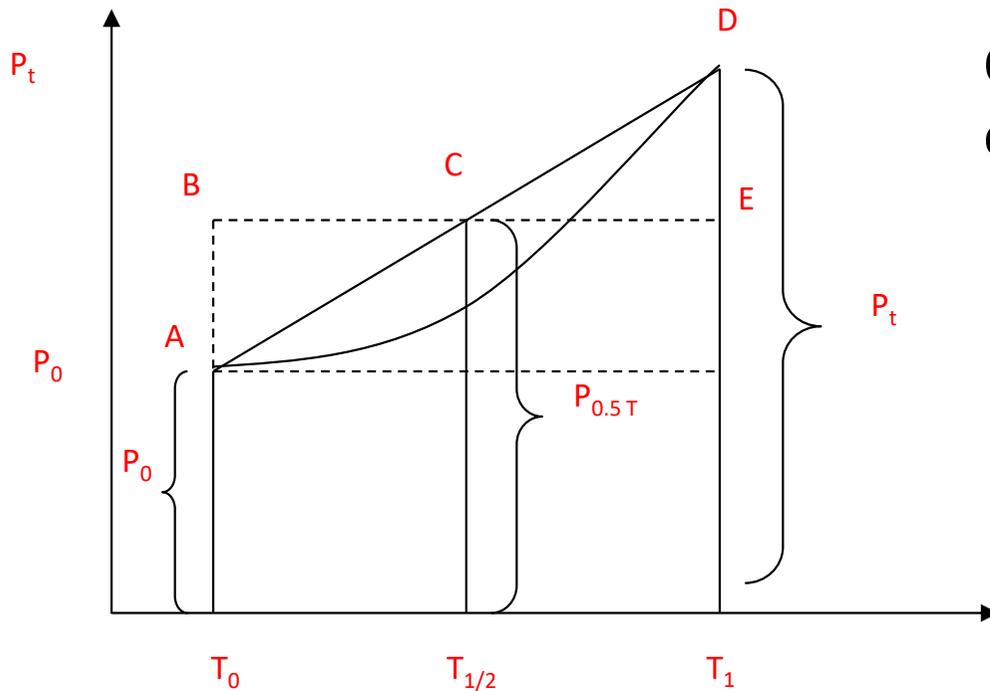


Par ailleurs, on voit que le score du NAV est égal à la surface du trapèze $A D T_1 T_0 = \frac{1}{2} (P_0 + P_t) \times (T_1 - T_0)$

Question d'autocontrôle : pourquoi considère-t-on le produit de la période d'observation et de la population à son milieu comme une population exposée au risque durant cette période ? Prouvez qu'une telle procédure est correcte et sous quelles conditions (hypothèses).

Estimation du NAV pour une population croissante exponentielle

Méthode 2 (si on ne dispose que la population à deux extrémités) :



On suppose (une hypothèse) que la croissance soit exponentielle

$$\text{Soit } P_t = P_0 e^{rt} \rightarrow r = \frac{\ln P_t - \ln P_0}{t} \quad (1)$$

$$\text{or par définition } r = \frac{P_t - P_0}{NAV} \quad (2)$$

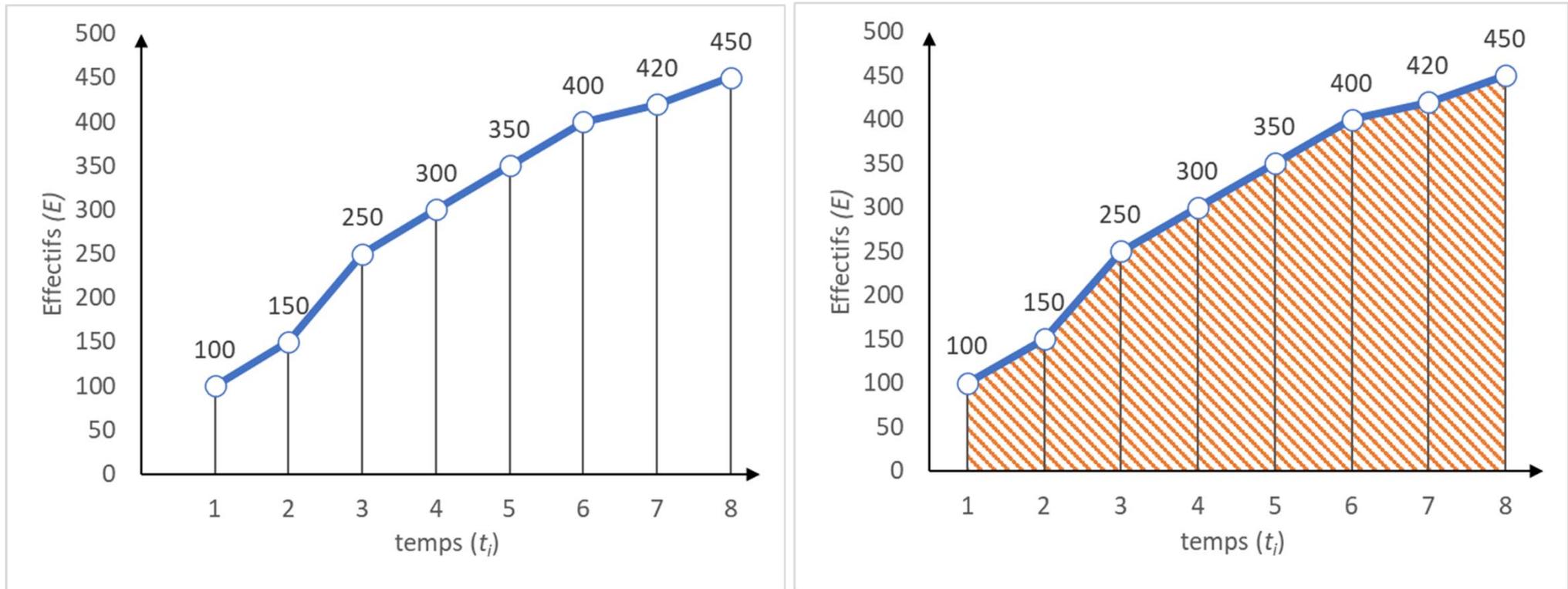
$$\text{alors (1) } \equiv \text{(2)} \rightarrow \frac{\ln P_t - \ln P_0}{t} = \frac{P_t - P_0}{NAV} \rightarrow$$

$$NAV = \frac{P_t - P_0}{\ln P_t - \ln P_0} \cdot t$$

autocontrôle : répondez à la question « Combien d'humains sont nés depuis la création du monde ? »

Estimation du NAV pour une population en croissance irrégulière

Méthode 3 (si on dispose des effectifs intermédiaires sur plusieurs intervalles) :



On peut prendre l'intégral par la méthode de trapèze

Soit $S_t = NAV$ dans un intervalle $t=T/n$, alors

$$S_x = \frac{P_t + P_{t+1}}{2} \cdot t \rightarrow NAV = \left(\frac{P_0 + P_n}{2} + \sum_{t=1}^{n-1} P_t \right) \cdot \frac{T}{n}$$

Un exercice pour comprendre l'importance du choix d'une méthode d'estimation du NAV

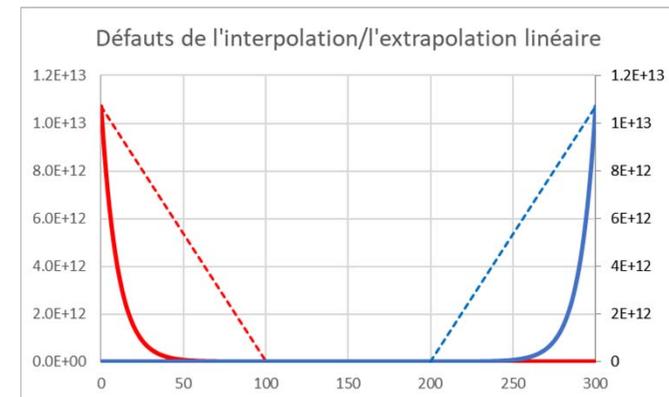
Vous travaillez dans une société d'assurance et le patron vous a demandé d'estimer le nombre d'annuités à toucher durant 10 ans, sachant l'effectif de la population assurée P_i au début de l'année i était :

$$P_1 = 100\ 000$$

$$P_5 = 86\ 071$$

$$P_{10} = 74\ 082$$

1. Calculez le taux d'accroissement sur l'intervalle entre t_0 et t_{10} (sur les autres intervalles).
2. Estimez le NAV selon toutes les méthodes et comparez les résultats (écart absolu et relatif).
3. Quelle approximation est meilleure ?



Méthodes de calcul de l'effectif de la population exposée

(du nombre d'années vécues sous le risque, ou nombre de personne – années \equiv NAV)

à partir des données individuelles

Méthode I

(observation directe en continu)

On observe **7 individus** (7 lignes de vie) durant la période entre le 1 janvier 1981 et le 1 janvier 1982 (période = 1 an)

On enregistre le moment d'entrée en observation et le moment de sortie de l'observation de chaque personne et on compte le nombre de trimestre d'observation sachant que 1 trimestre = $\frac{1}{4}$ d'année.

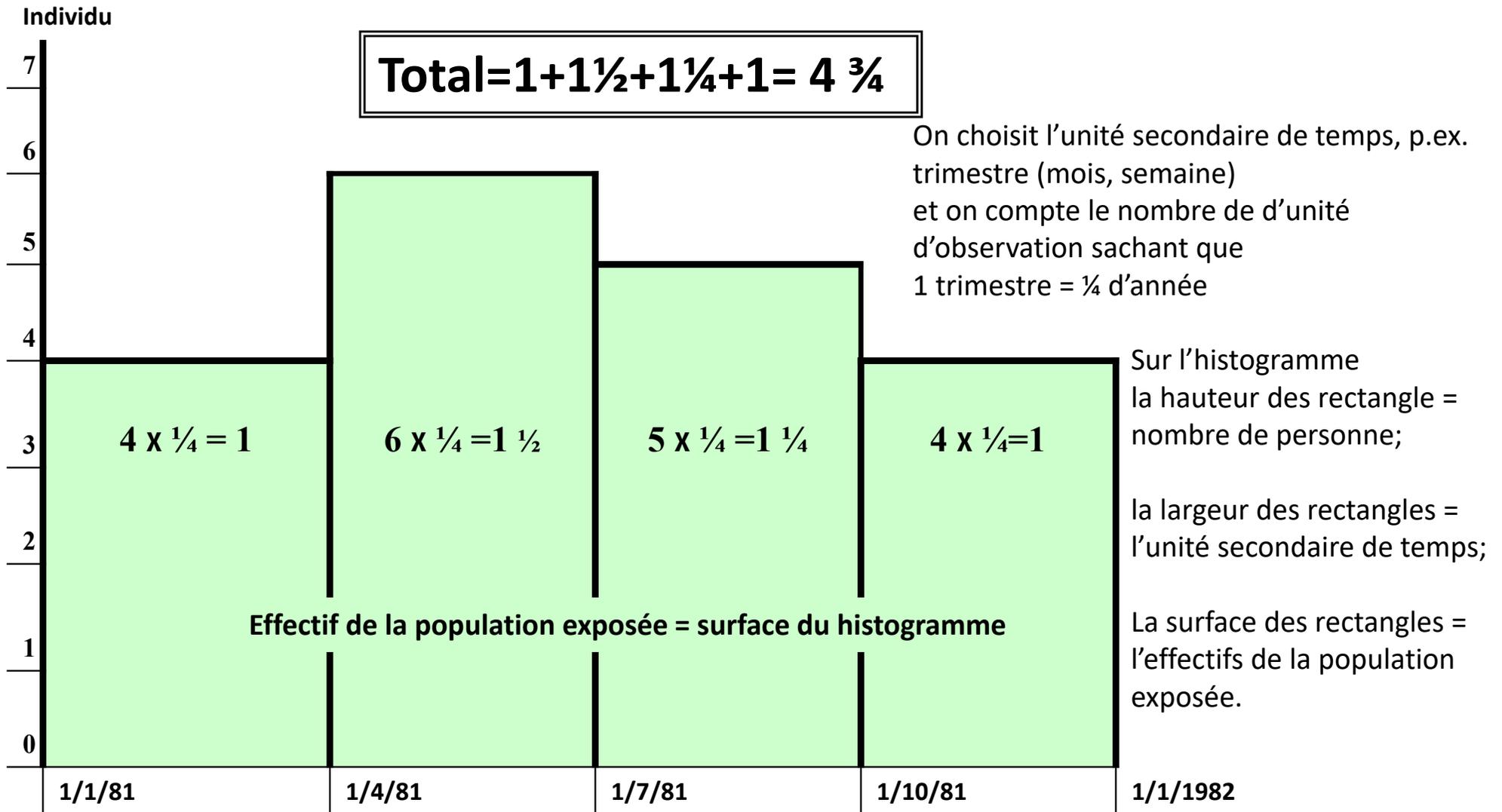
Individu	Personne-années vécues				
	I trimestre	II trimestre	III trimestre	IV trimestre	
1	→				$4 \times \frac{1}{4} = 1$
2	→				$\frac{3}{4}$
3		→			$\frac{3}{4}$
4	→				1
5	→				$\frac{1}{2}$
6		→			$\frac{1}{2}$
7				→	$\frac{1}{4}$
	I trimestre	II trimestre	III trimestre	IV trimestre	$4 \frac{3}{4}$ Total

Effectif de la population exposée = $4, \frac{3}{4}$ personnes-années

Méthode II

(observation directe
en temps discret)

Les lignes de vie sont converties en nombre de personnes qui sont vivant au chaque moment entre le 1 janvier 1981 et le 1 janvier 1982. On enregistre le nombre d'individu présent durant chaque intervalle.



Les deux méthodes donnent le même résultat, mais il ne sont pas tout à fait équivalentes. Dans cet exemple les sorties d'observation ont eu lieu toujours à la fin de trimestre, donc il s'agit d'une coïncidence.

Quantification et l'analyse de la structure d'une population par âge et par sexe

- Les indices relatifs à la structure (ratios et proportions)
- Présentation graphique: pyramides démographiques
- Indicateurs usuels : rapports de dépendance
- Évaluation de qualité des données sur la structure par âge et par sexe :
 - UN ASAI (United Nation Age-Sex Accuracy Index)
 - Indice de Whipple (préférences pour les âges avec le 0 et le 5)
 - Indice de Bachi (préférences pour n'importe quel chiffre)
 - Indice synthétique (mêlé) de Myers

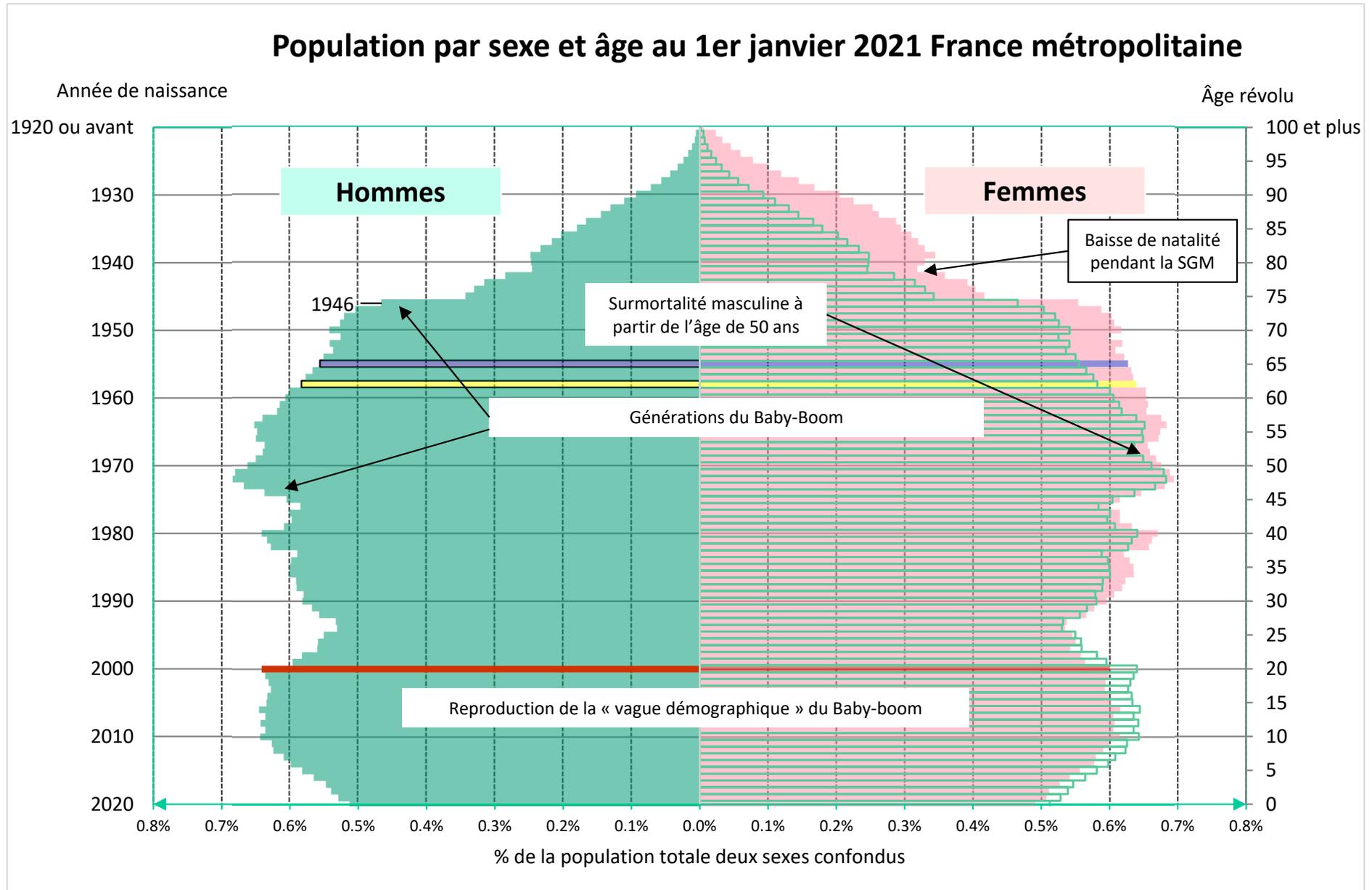
Analyse de la structure de population par âge et par sexe

- donne une image de la population dans son ensemble
- raisonnablement prévisible, elle ne dépend que des naissances, de la mortalité et des migrations historiques
- un écart entre la structure observée et celle attendue permet d'évaluer la qualité des données (e.g. celles d'un recensement)

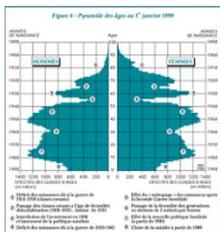
Méthodes d'analyse et approches:

- Présentations graphiques (pyramides, survie des générations : les histogrammes superposés)
- Rapport des effectifs des générations (ratios des âges « x ») = $\frac{N_{x+n}}{N_x}$
- Rapport des effectifs des sexes (ratio de sexes, dit « rapport de masculinité ») = $\frac{N_m}{N_f}$
- Excédent (déficit) of males/femelles dans une population = $\frac{N_m - N_f}{N_m + N_f}$
- Indices intégraux (associés à la qualité des données)
 - Indice de précision âge-sexe de l'Organisation des Nations Unies (*UN Age-Sex Accuracy Index*)
 - Indices de préférences (rebondissement) des âges de Whipple, de Bachi et de Myers
 - Modèle de populations stables (avec ses variantes)
 - Cohérence interne (états matrimoniaux etc.)

Présentation de la structure par sexe et par âge d'une population: la pyramide démographique



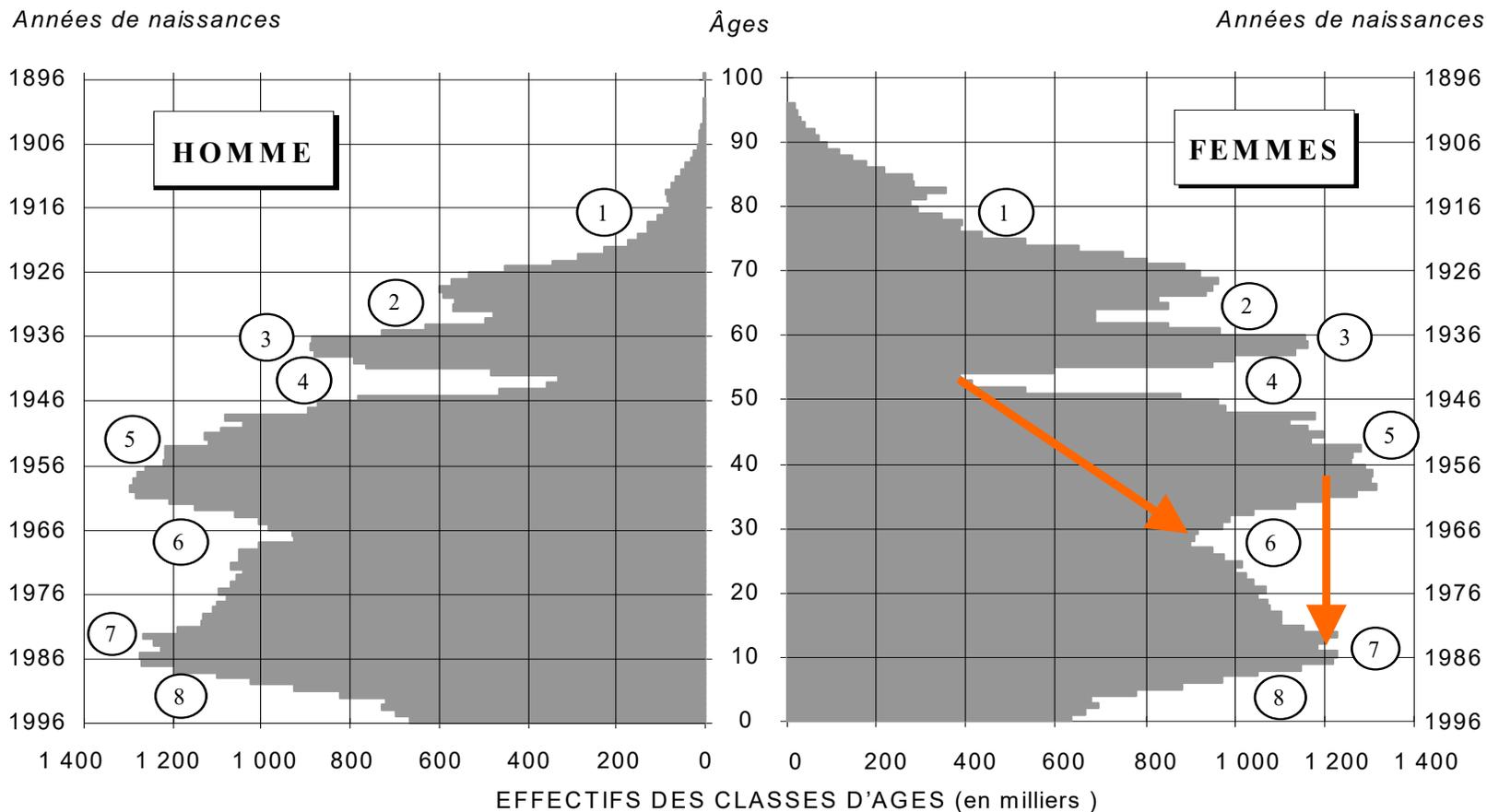
Source des données: INSEE



Les empreintes historiques sur la pyramide des âges et leurs conséquences démographiques.

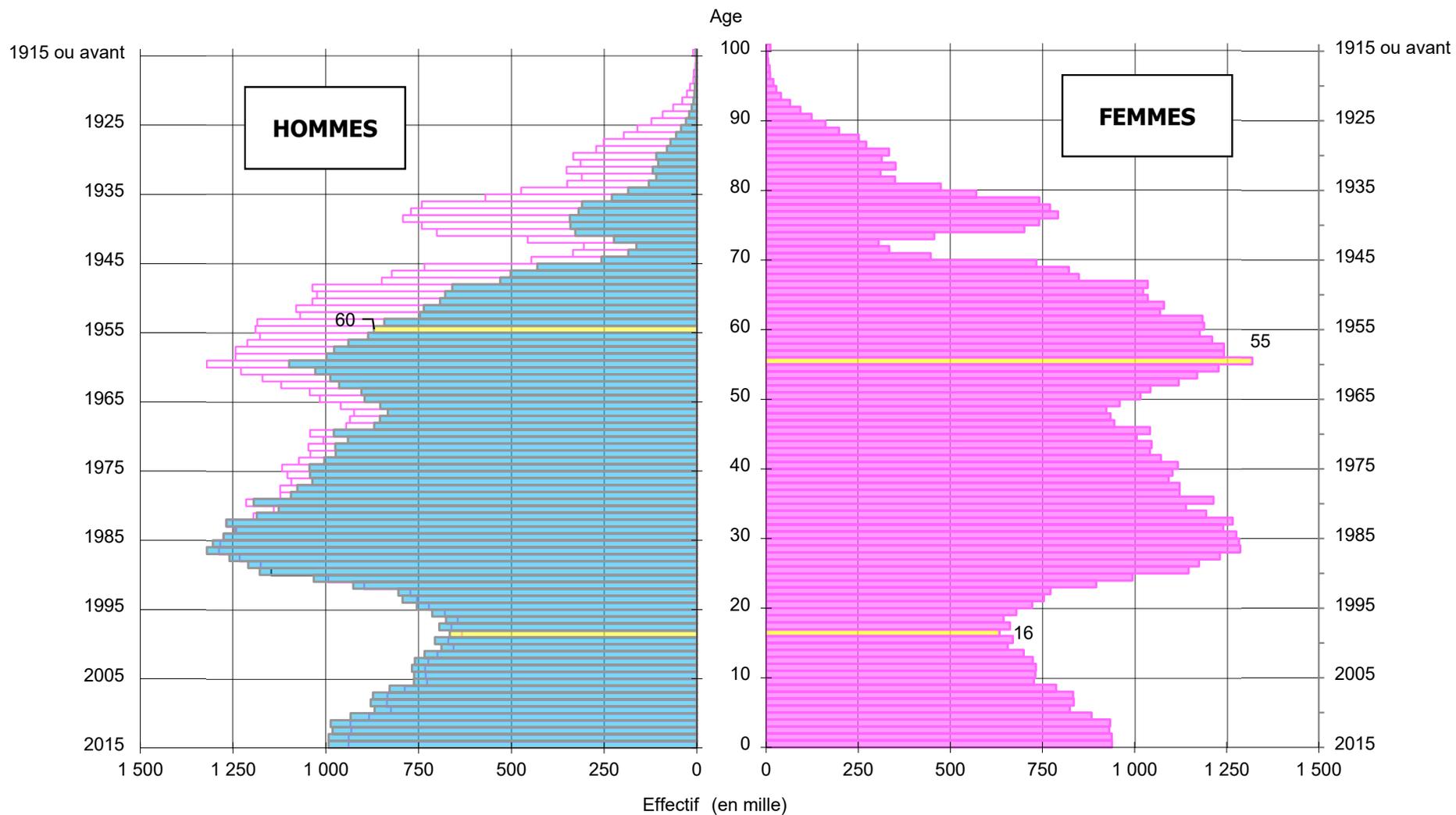
Population de la Russie au 1 janvier 1997

Pour plus de détails voir :
[A. Avdeev, A. Blum \(1999\)](#)
 « La population russe : des raisons d'espérer ? »
Population et Sociétés
 n° 351, novembre 1999

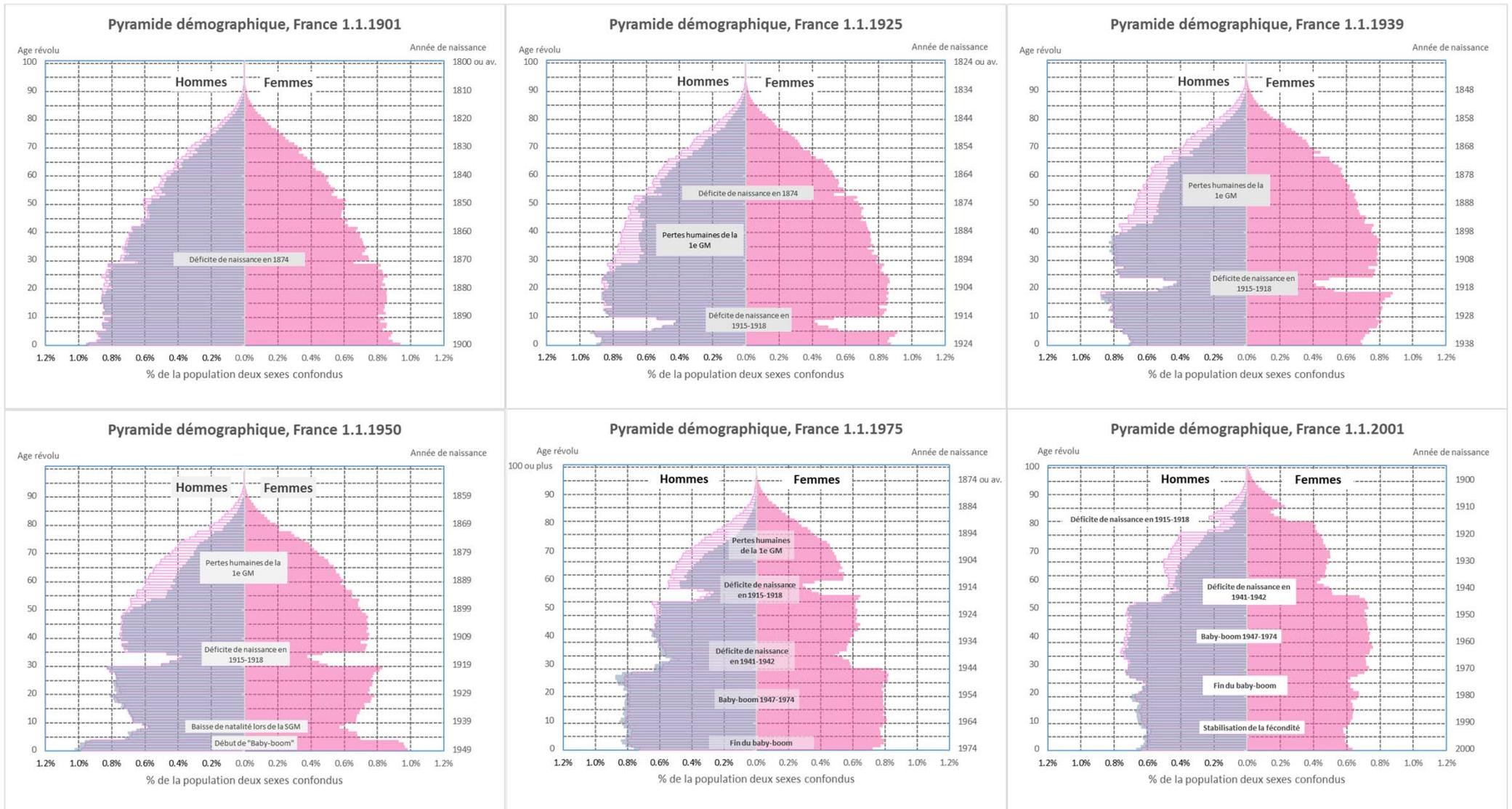


- | | |
|--|--|
| ① Déficit des naissances dû à la guerre 1914-1918 (classe creuses) | ⑤ Effet du « rattrapage » des naissances après la guerre |
| ② Passage des classe creuses à l'âge de fécondité, la collectivisation et la « dékoulakisation » 1928-1932, la famine de 1933. | ⑥ Passage de la fécondité des générations en dessous de 2 enfant par femme |
| ③ Interdiction de l'avortement en 1936 et lancement de la politique nataliste. | ⑦ Effet de nouvelle politique familiale (à partir de 1983) |
| ④ Déficit des naissances dû à la guerre 1941-1945 | ⑧ Chute de la natalité à partir de 1989 |

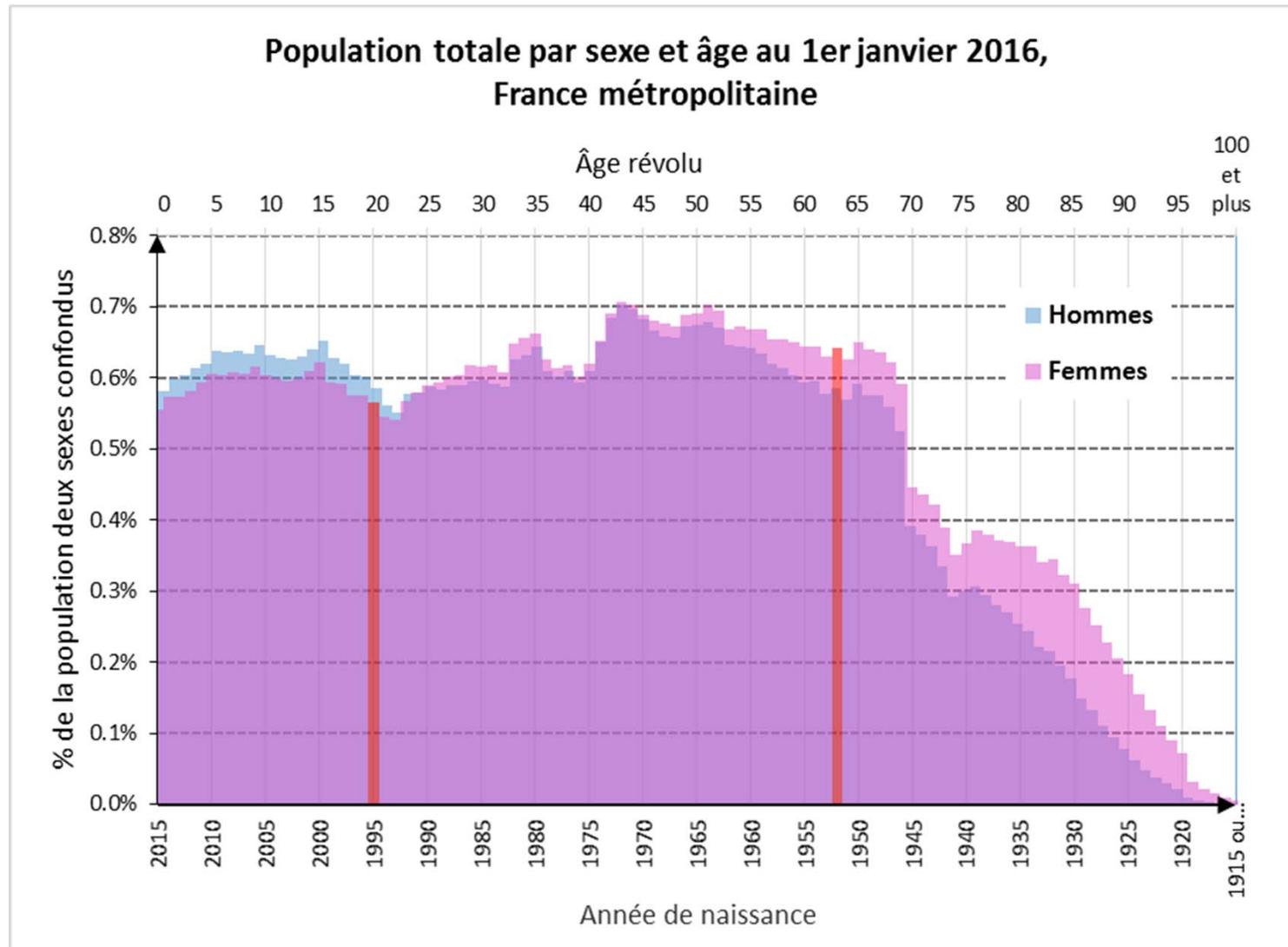
Structure de la population de Russie au 1 janvier 2016: les empreintes de l'histoire ne s'effacent pas vite



Un siècle d'histoire démographique de la France en six images



Une autre façon de présenter graphiquement la structure par âge et sexe

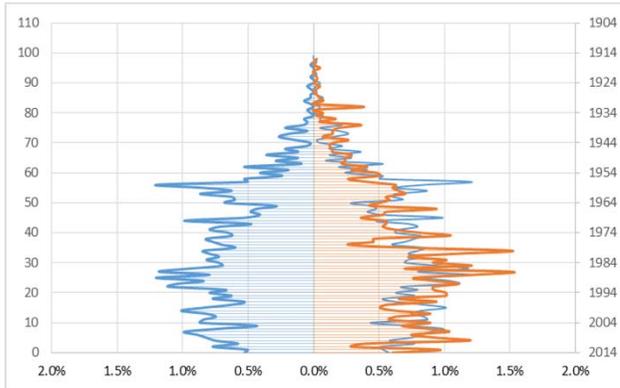


Comment peut-on dessiner des pyramides d'âge comparables?

Une pyramide des âges présente deux (ou plus) histogrammes superposés

(graphique en barres (rectangles) représentant la densité d'une distribution)

Soit N_x^S - effectif de la population de sexe **S** et de l'âge **x**



- variables : 1) f_x^m proportion de l'effectif d'un groupe d'âge de sexe masculin dans le total de la population

$$f_x^s = \frac{N_x^s}{\sum_s \sum_x N_x^s}$$

2) f_x^m proportion de l'effectif d'un groupe d'âge de sexe féminin dans le total de la population

- base d'un rectangle : = l'amplitude (a_i) d'une classe d'âge (peut varier d'une classe à l'autre)

- hauteur d'un rectangle : = la densité (h) de la distribution par âge (la fréquence divisée par l'amplitude de l'intervalle)

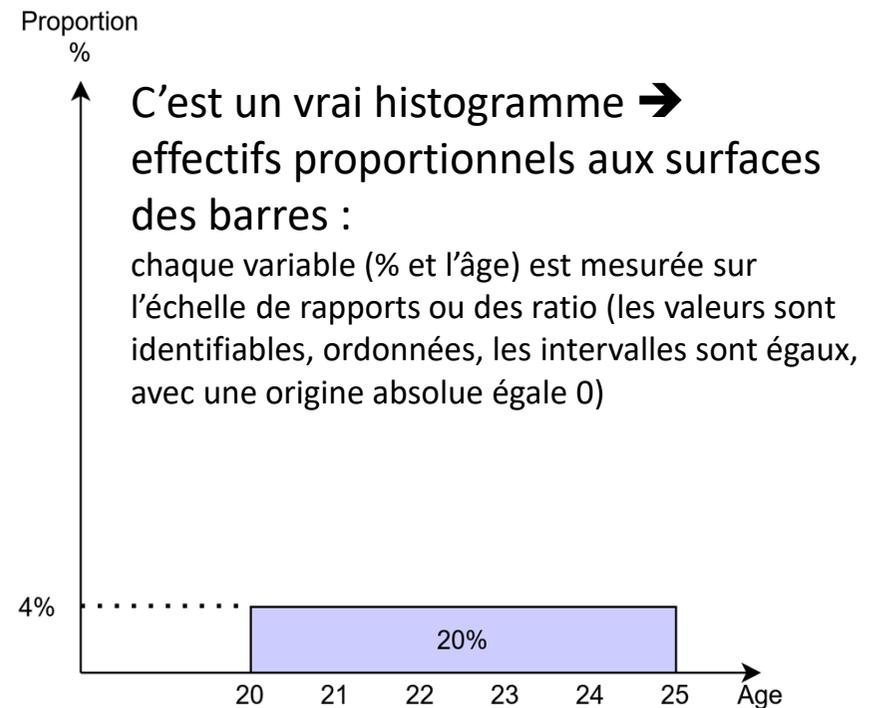
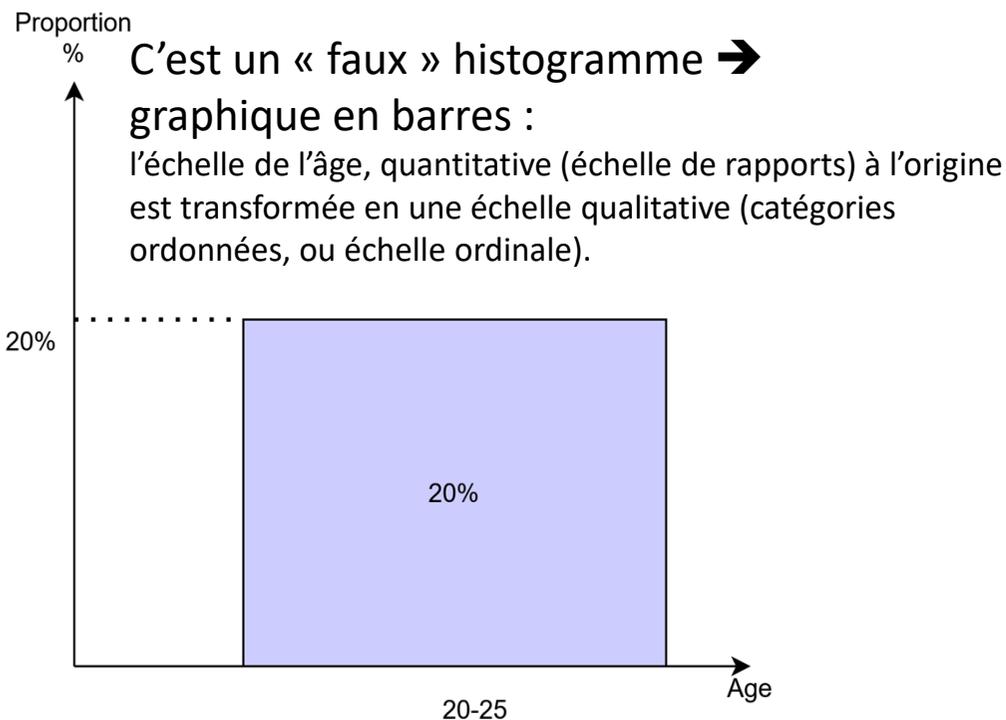
$$h = \frac{f_x^s}{a_x}$$

- surface d'un rectangle : = proportionnelle donc à l'effectif d'un groupe d'âge (à la part d'un groupe d'âge dans l'effectif total de la population)

Question d'autocontrôle : pourquoi rapporte-t-on l'effectif d'un group « sexe-âge » au total de la population et non au total d'un sexe?

Choisir correctement la hauteur d'un rectangle dans une histogramme

- L'âge est une variable catégorielle, qualitative et ordonnée :
- L'intervalle 20-25 ans est UNE UNITÉ d'échelle, quoi que son amplitude soit de 5 ans
- La hauteur du rectangle = 20% = proportion de l'effectif
- La surface du rectangle = 20% x 1
- L'âge est une variable quantitative et continue :
- L'unité de l'échelle = UNE ANNÉE D'ÂGE
- L'intervalle 20-25 ans composé de 5 unités d'âge révolu (20, 21, 22, 23 et 24)
- La hauteur du rectangle = 4% (20% : 5)
- La surface du rectangle = 20% (4 % x 5)



Le mauvais choix peut fausser la position du ou des mode(s) sur le graphique. C'est une grave erreur !!!

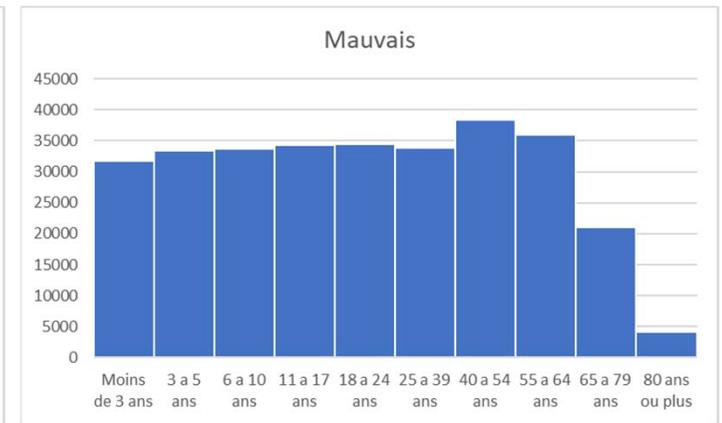
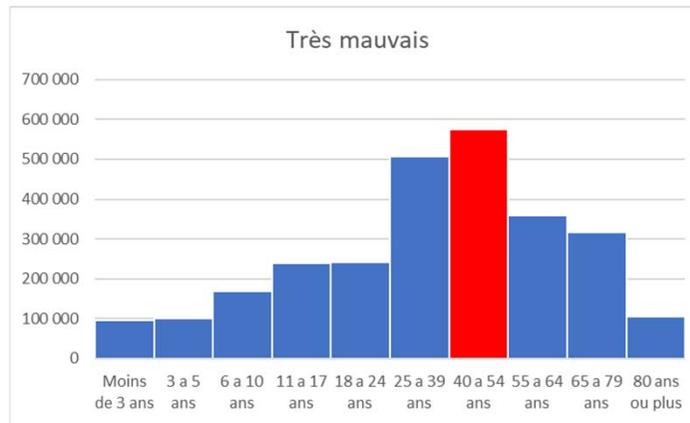
Exemple d'une pyramide aux intervalles d'âge inégaux

Données:

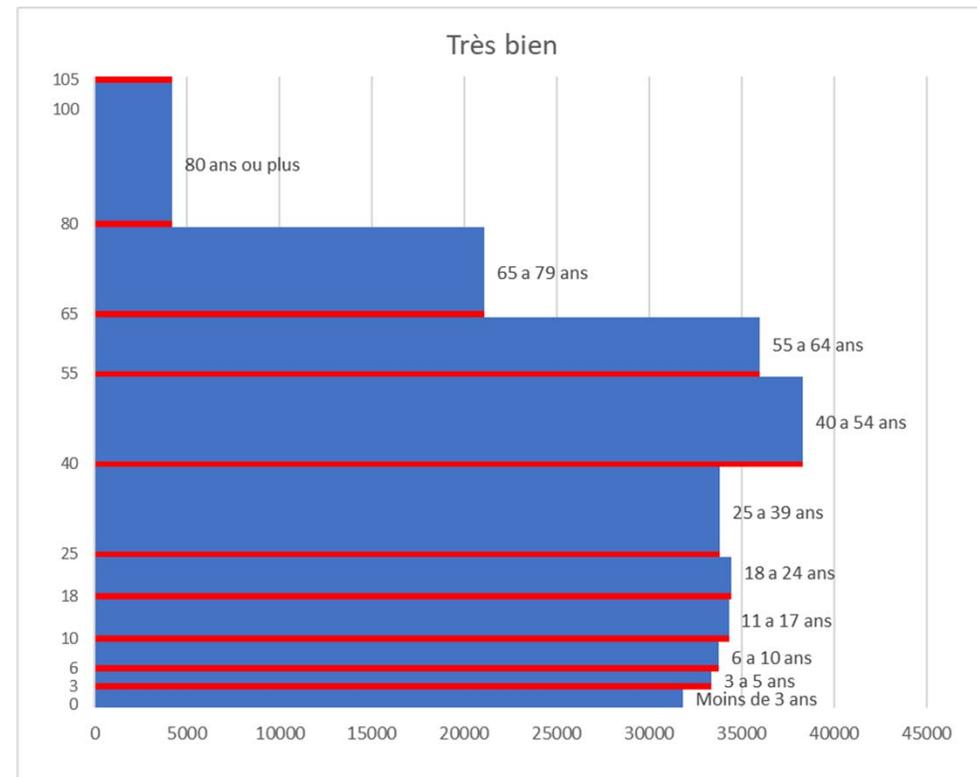
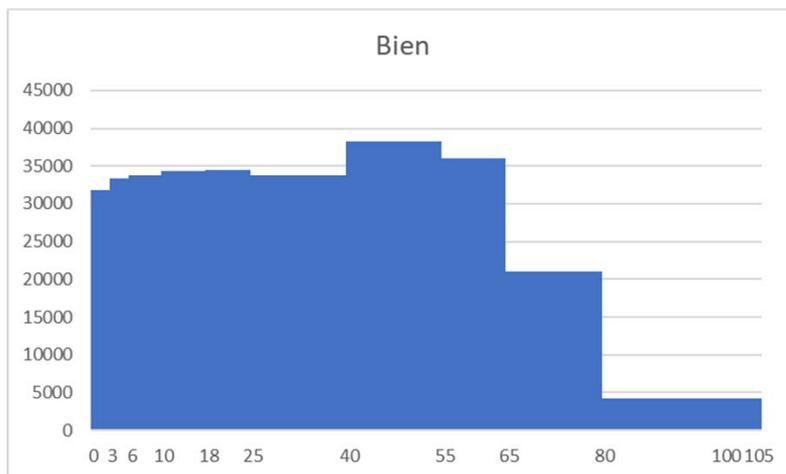
POP1A - Population par sexe
et âge regroupé en 2014
Région du Grand Est (44)

Group d'âge	Hommes	Femmes
Moins de 3 ans	95 453	91 395
3 a 5 ans	100 113	96 338
6 a 10 ans	168 844	161 120
11 a 17 ans	240 243	226 906
18 a 24 ans	241 194	228 300
25 a 39 ans	507 209	506 950
40 a 54 ans	574 710	581 980
55 a 64 ans	359 855	376 984
65 a 79 ans	316 241	369 919
80 ans ou plus	104 696	206 194
Ensemble	2 708 558	2 846 087

Source : Insee, RP2014 exploitation principale,
géographie au 01/01/2016.

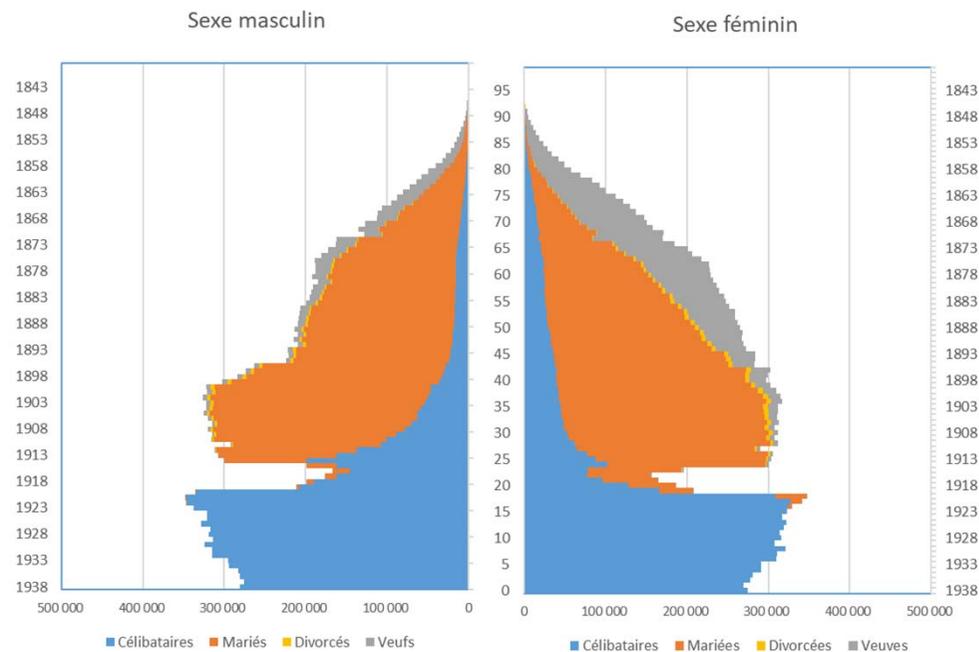
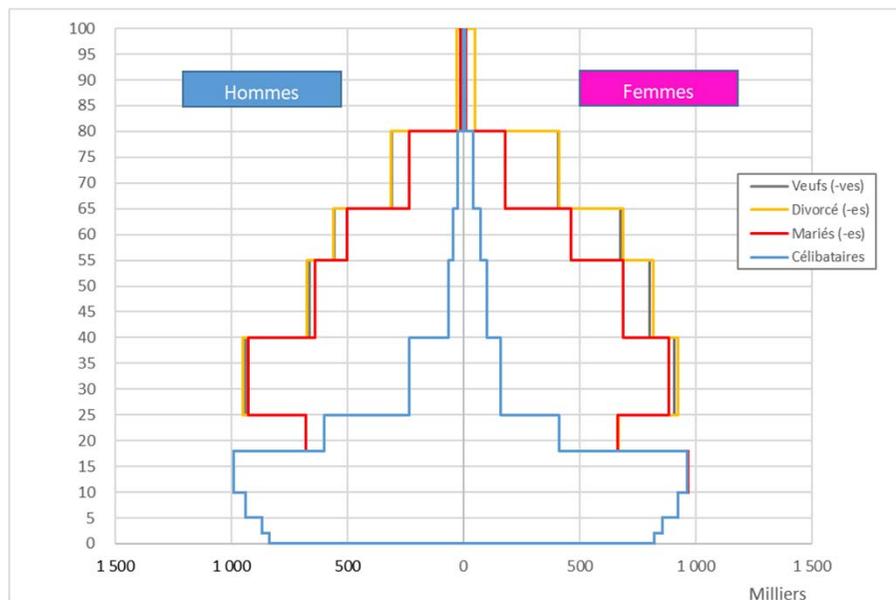


Le faux mode (à gauche) et les intervalles d'âge inadéquats (les deux)

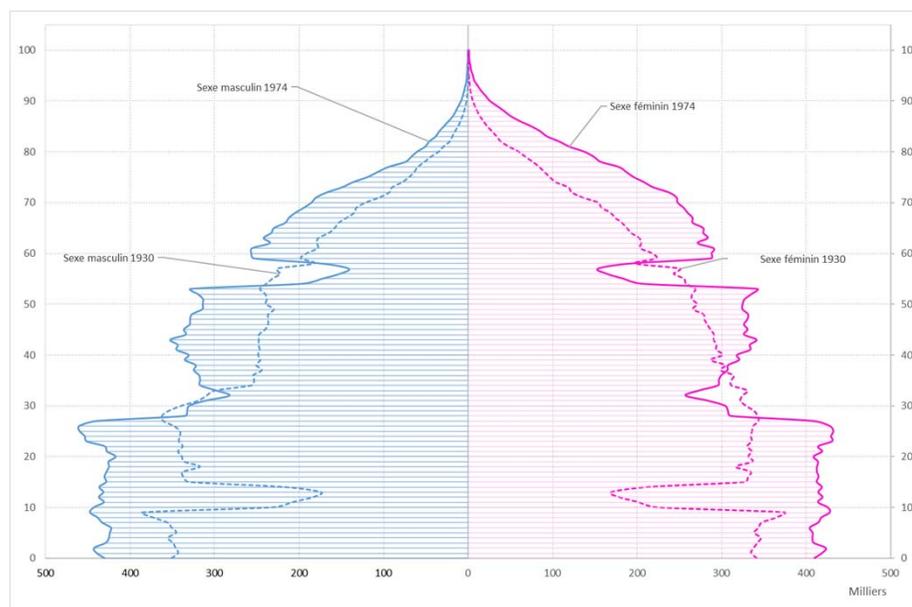


Plusieurs histogrammes superposés : analyse visuelle comparative

Population de la France par âge, sexe et état matrimonial en 1930



Population de la France par âge et par en 1930 et 1974



Comment analyser la structure par âge et par sexe ?

Ratios âge/sexe spécifiques à la structure par âge et par sexe:

- ratio hommes/femmes par âge (**rapport de masculinité** - ${}_nRM_x$)

soit ${}_nN_x^S$ - effectif de la population dans l'intervalle d'âge entre x (compris) et $x+n$ (non compris) de sexe $S = F$ (féminin), ou $=M$ (masculin)

$${}_nRM_x = \frac{{}_nN_x^M}{{}_nN_x^F} \cdot 100$$

On dit: autant d'hommes pour 100 femmes

- proportion d'hommes ${}_nPM_x = \frac{{}_nN_x^M}{{}_nN_x^F + {}_nN_x^M} \cdot 100$

- on peut par ailleurs démontrer que ${}_nRM_x = \frac{{}_nPM_x}{1 - {}_nPM_x} \cdot 100$ et inversement ${}_nPM_x = \frac{{}_nRM_x}{1 + {}_nRM_x} \cdot 100$

Indicateurs statistiques descripteurs :

$$\bar{x}^S = \frac{\sum_{x=0}^{\infty} {}_nN_x^S \cdot \left(x + \frac{n}{2}\right)}{\sum_{x=0}^{\infty} {}_nN_x^S} \rightarrow \text{âge moyen de la population (total ou deux sexes confondus, de sexe masculin, de sexe féminin)}$$

$$\sum_{x=0}^{Me} {}_nN_x^S = \sum_{x=Me}^{\infty} {}_nN_x^S \rightarrow \text{âge médian de la population (total ou deux sexes confondus, de sexe masculin, de sexe féminin)}$$

Indicateurs comparatifs :

Soit $r_{x,i}$ - proportion de catégorie x dans une population i

- indice dissimilarité linéaire (de deux population) : $ID = 0.5 \cdot \sum_{x=1}^n |r_{x,2} - r_{x,1}|$

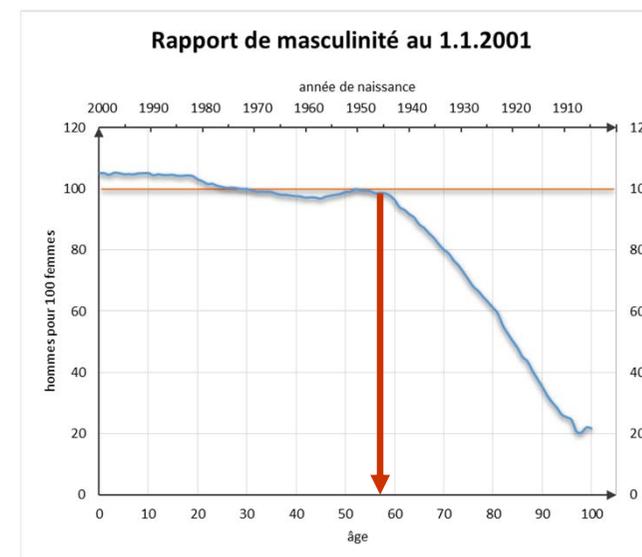
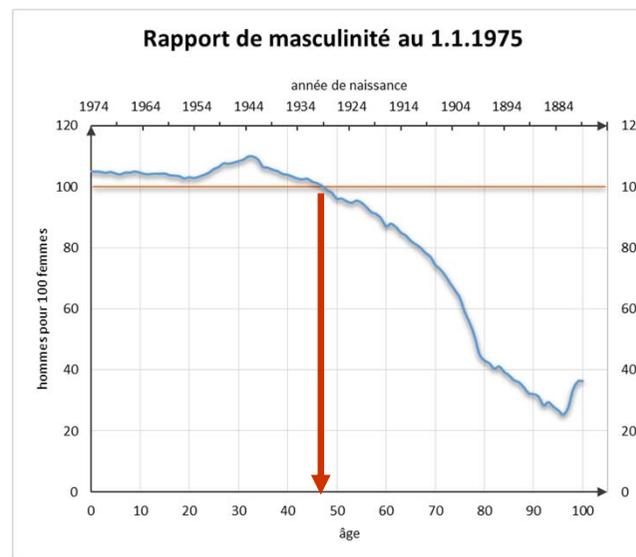
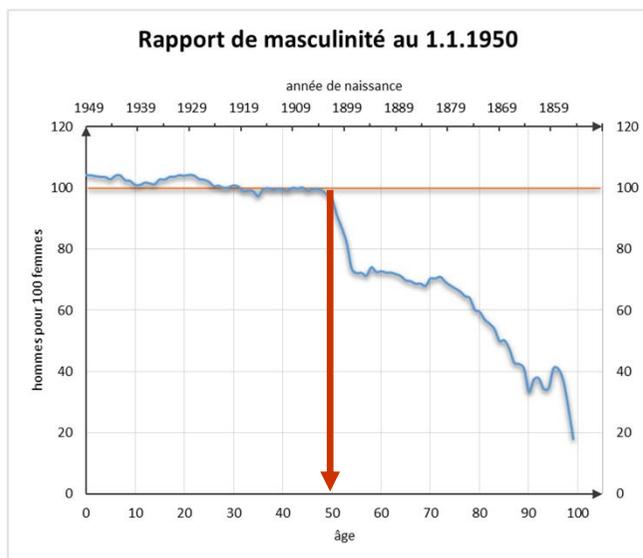
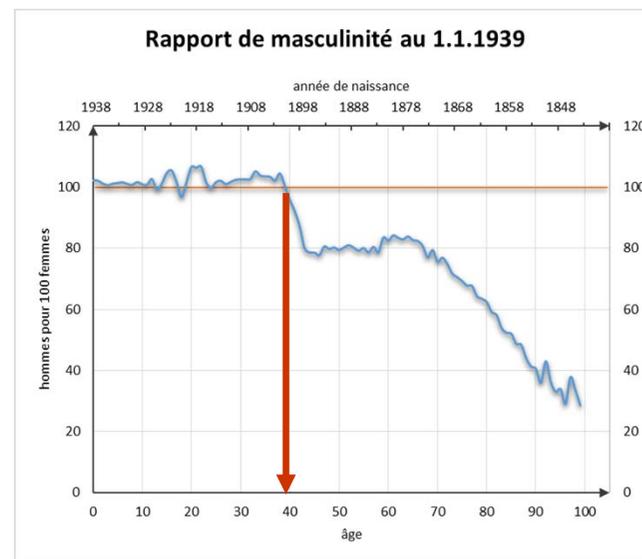
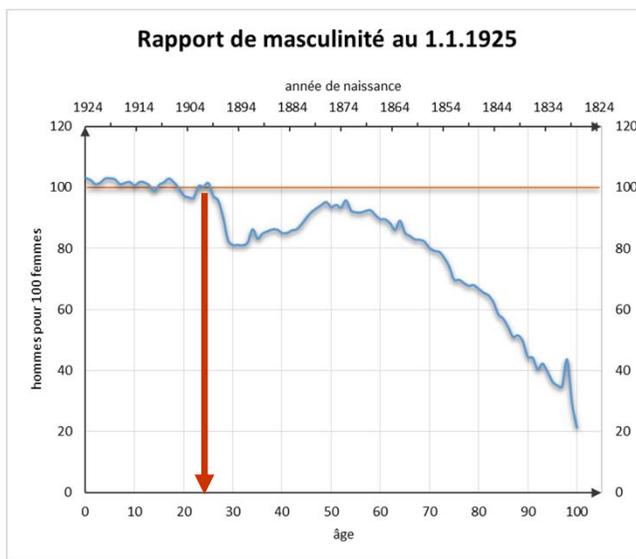
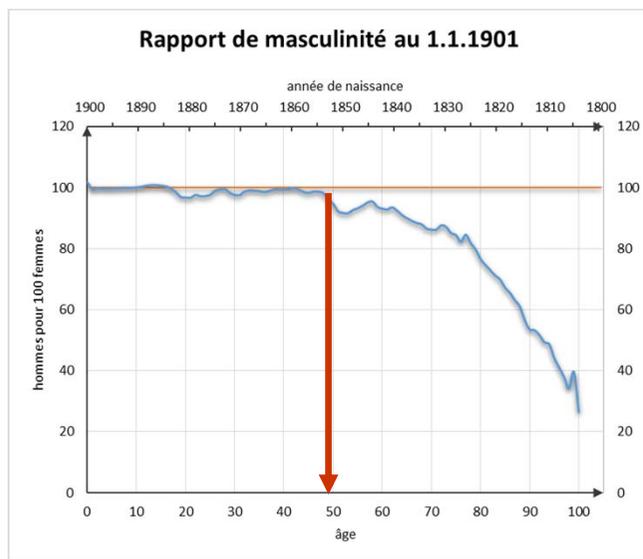
- indice de la différence relative (de deux population) : $IDR = 0.5 \cdot \frac{1}{n} \cdot \sum_{x=1}^n \left| \left(\frac{r_{x,2}}{r_{x,1}} \cdot 100 \right) - 100 \right|$

- indice généralisé de la dissimilarité (de Szalay)

$$I_{sz} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \left(\frac{s_{i,1} - s_{i,2}}{s_{i,1} + s_{i,2}} \right)^2}, \quad 0 < I_{sz} < 1$$

e.g. pour évaluer l'homogénéité territoriale (comparer les régions)

Le rapport de sexe ou le rapport de masculinité en France



L'âge de début et les causes du déséquilibre ?

Autres indicateurs caractérisant la structure la structure par âge

- **ratio enfants/femmes** (enfants/mères) « *REM* » caractérisant le remplacement des générations

soit ${}_5N_0$ - effectif de la population dans l'intervalle d'âge entre l'âge 0 (compris) et l'âge de 5 ans (non compris) = l'âge révolu 0-4 ans

${}_{35}N_{15}^F$ - effectif de la population de sexe féminin dans l'intervalle d'âge entre 15 (compris) et 50 (non compris) = l'âge féconde 15-49 ans révolus

$$REM = \frac{{}_5N_0}{{}_{35}N_{15}^F} \cdot 100 \quad \dots \text{enfants pour 100 femmes,}$$

REM peut servir pour une estimation indirecte de la fécondité (moyenne sur 5 ans)

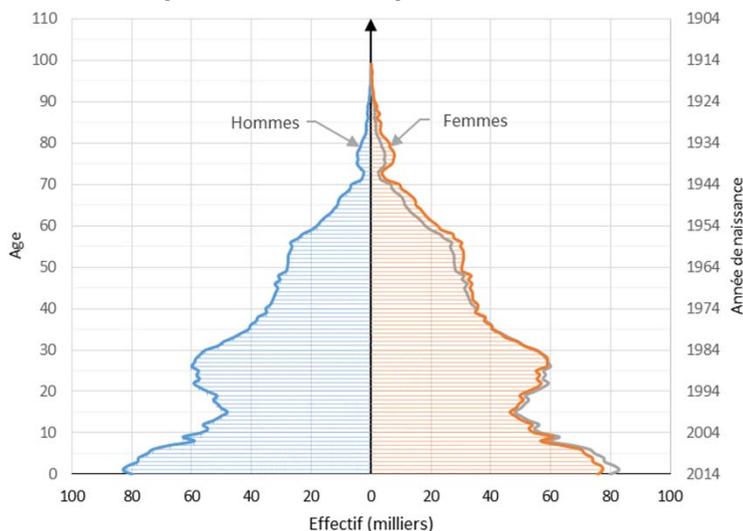
- **ratio de dépendance *RD*** ou la charge (*le dividende*) « démographique », on distingue en plus *RDj* de jeunes et *RDv* des vieux

$$\left. \begin{aligned} RDj &= \frac{{}_nN_0}{{}_{65-n}N_n} \cdot 100 \\ RDv &= \frac{{}_\infty N_{65}}{{}_{65-n}N_n} \cdot 100 \end{aligned} \right\} \Rightarrow RD = \frac{{}_nN_0 + {}_\infty N_{65}}{{}_{65-n}N_n} \cdot 100 \quad \underline{\text{Nota:}} \text{ le choix de } n \text{ est arbitraire}$$

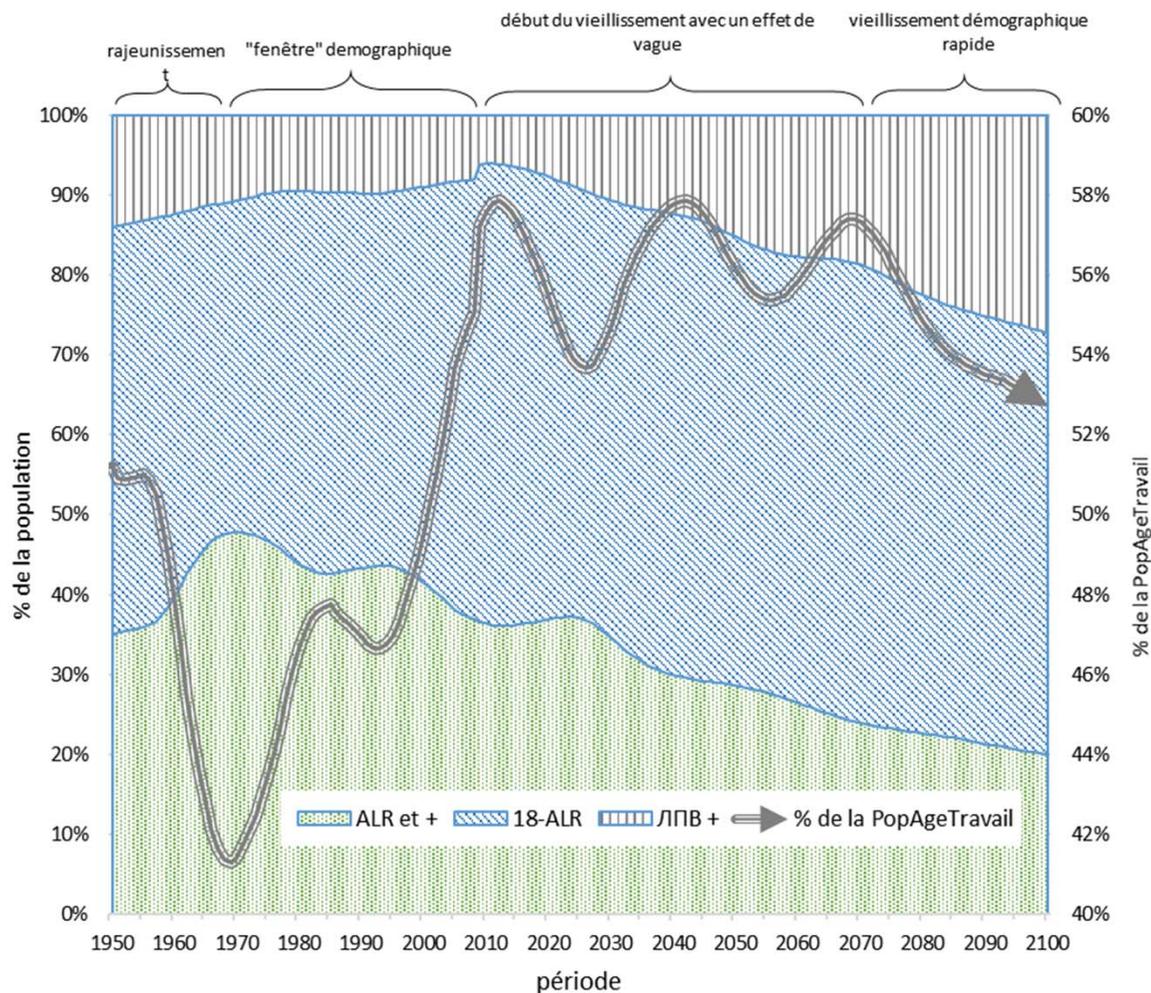
RD – rapport entre (l'effectif de) la population hors l'âge de travail et (l'effectif de) la population à l'âge de travail

Transition démographique et la dynamique des trois grands groupes d'âge dans la République Kirghize

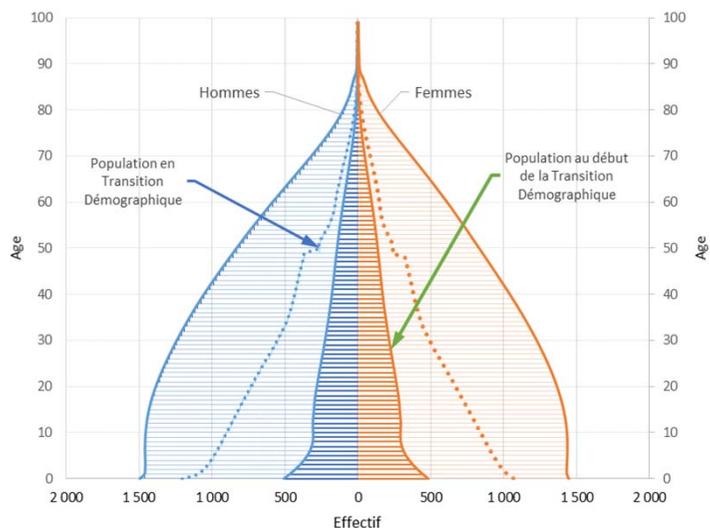
Population au 1 janvier 2017



Dynamique historique et perspectives des rapports entre trois grands groupes d'âge 1950-2100

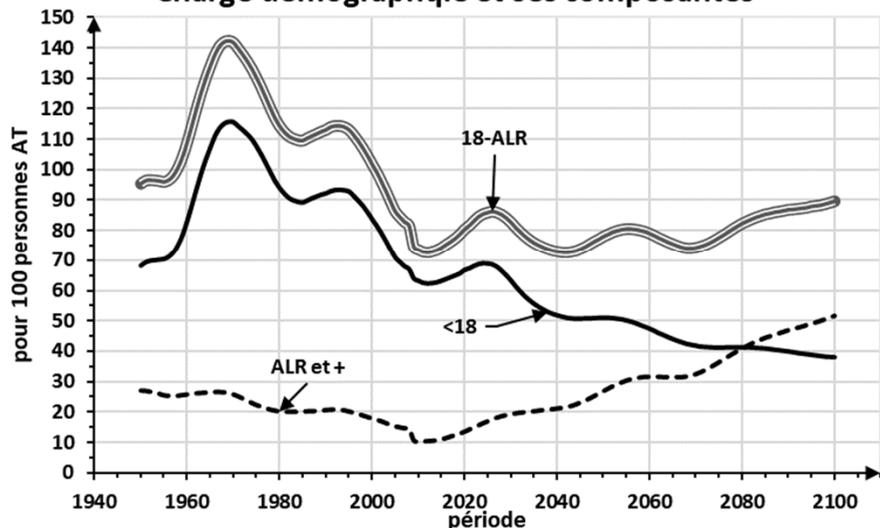


Modèle de l'évolution de la structure par âge et par sexe

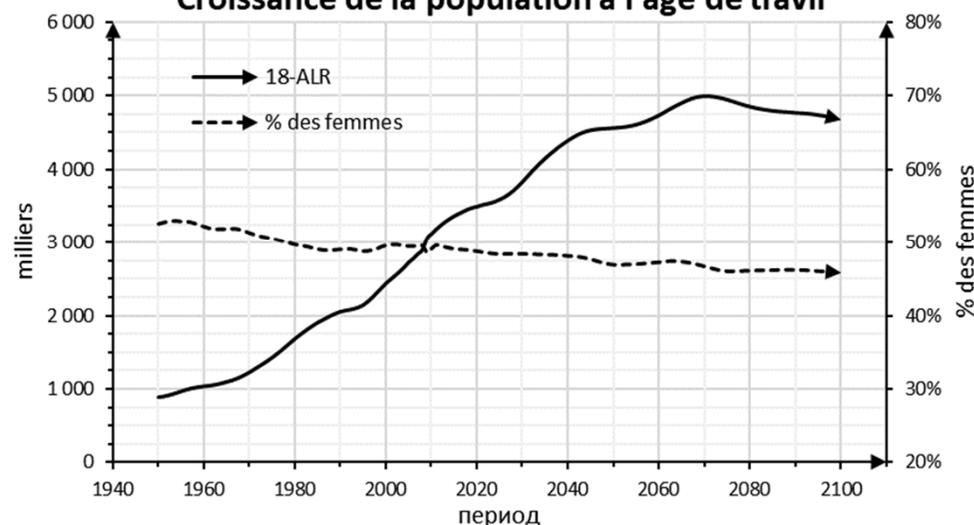


Analyse descriptive des perspectives de la structure par âge et par sexe de la population de la République Kirghize

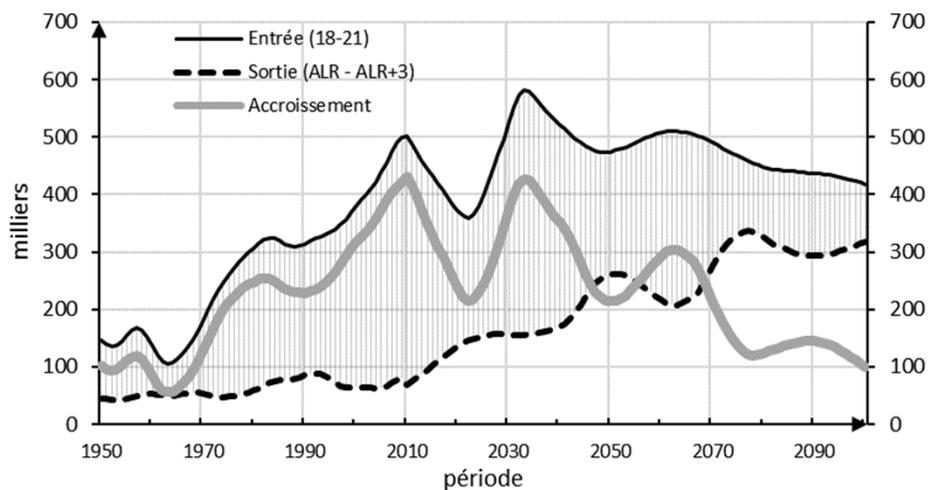
Charge démographique et ses composantes



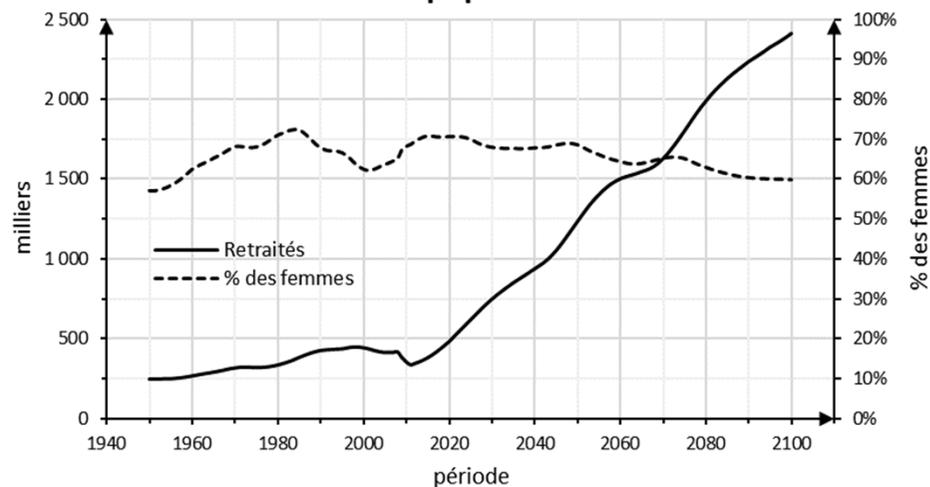
Croissance de la population à l'âge de travail



Remplacement des générations au marché de travail



Croissance de la population à la retraite



Indice de Billeter (IB) : quantification du vieillissement de la population

Ernst P. Billeter "Eine Maßzahl zur Beurteilung der Altersverteilung einer Bevölkerung", *Swiss Journal of Economics and Statistics (SJES)*, 1954, vol. 90, issue IV, 496-505

Soit $N(0-14)$ – l'effectif de la population à l'âge révolu de 0-14 ans
 $N(50+)$ – l'effectif de la population à l'âge révolu 50 ans et plus
 $N(15-49)$ – l'effectif de la population à l'âge révolu de 15-49 ans

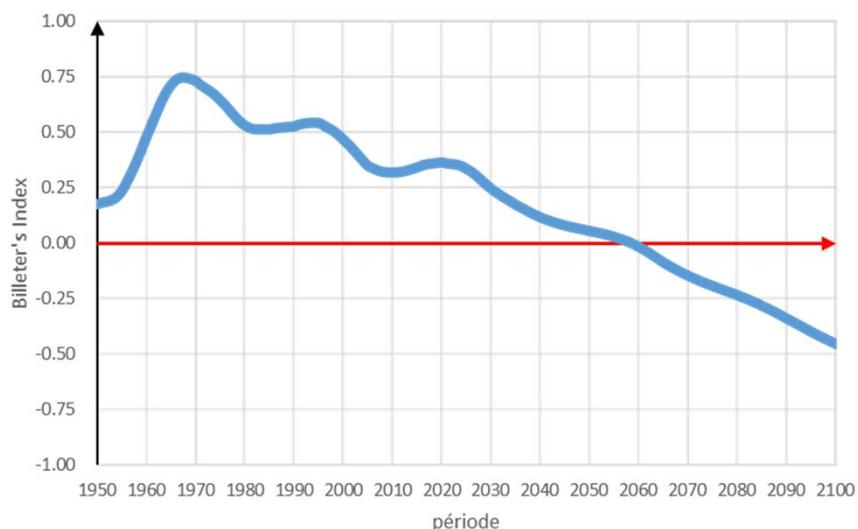
$$IB = \frac{N_{0-14} - N_{50+}}{N_{15-49}}$$

$IB > 0$ pour les populations plutôt « jeunes »
 $IB < 0$ pour les populations plutôt « vieilles »

Les « + » : facile à calculer et facile à interpréter dans les termes du binôme « vieille – jeune » .

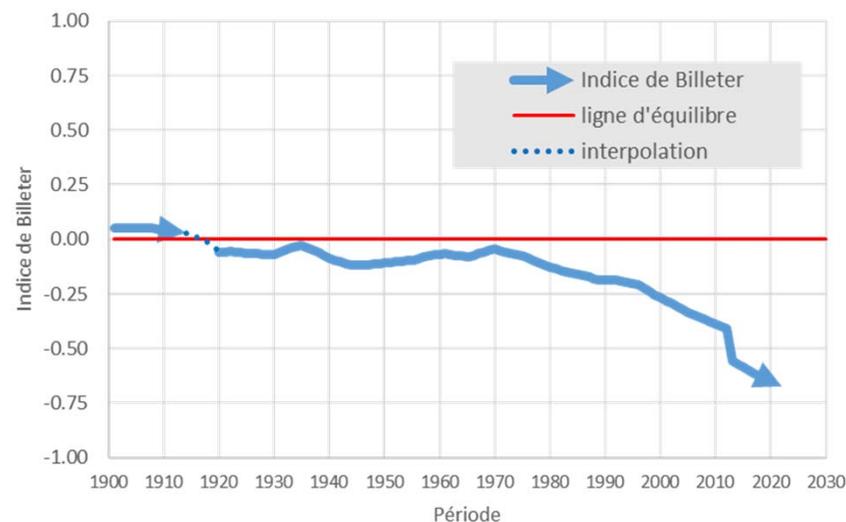
Les « - » : ne tient pas compte à des empreintes historiques, convient peu pour les comparaisons internationales

Dynamique de l'indice de Billeter en Kirghizie de 1950 à 2020 avec les perspectives jusqu'à 2100



Calculs à partir des données UN World Population Prospects, 2019 revision

Dynamique de l'indice de Billeter en France de 1901 à 2022



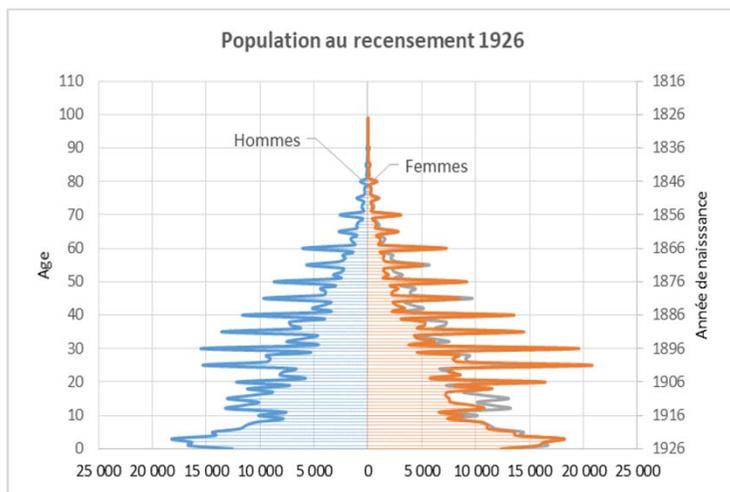
Calculs à partir des données de l'INSEE (fichier fm_t6.xls, version 2023), avec une interpolation pour la période 1914-1919

Qualité des données sur la structure par âge et par sexe

Lecture: Franc Hobbs « Age and Sex Composition » in J.S.Siegel and D.A.Swanson (eds), *The Methodes and Materials in Demography*, second edition, 2008, p.125-173 (136-140 notamment)

+ **notre petit outil d'analyse « Pyramides »** sur l'EPI ou sur <http://dmo.econ.msu.ru/teaching/demo/index.htm>

UN Age-Sex Accuracy Index – un estimateur simple de la cohérence des données



On attend que les rapports entre les effectifs des générations successives et des nombres des hommes et des femmes respectent certaines règles:

- Effectif d'une génération se situe numériquement entre les deux générations voisines (celle d'avant et celle d'après)
- Le rapport des effectifs des hommes et des femmes de mêmes générations est stable

ONU Indice de précision : se calcule pour les groupes d'âge quinquennaux inférieurs 70 ans à partir des écarts moyens des ratios par sexe et par âges (s'il n'y avait pas des événements perturbateurs dans le passé ces ratios doivent être près de 100 dans tous les groupes d'âges).

Écart moyen de 100 quantifie la distorsion: <20 (précis); de 20 à < 40 (imprécis); 40 et + (très imprécis)

Trois composants:

1. Écart moyen des rapports des effectifs des sexes par âge (ERS)
2. Écart moyen des rapports des effectifs des hommes des groupes d'âges successifs (RA^m)
3. Écart moyen des rapports des effectifs des femmes groupes d'âges successifs (RA^f)

ONU IP = (3 x E.R.S.) + R.A.^m + R.A.^f → voir les calculs sur la diapositive suivante

Calculs de l'IP ONU (Kirghizie , 1926)

Age	Hommes	Femmes	${}_nRS_x$	$abs({}_nRS_x - {}_nRS_{x-n})$	$abs(100 - {}_nRA_x^m)$	$abs(100 - {}_nRA_x^f)$
0-4	77 840	76 340	101.96			
5-9	55 823	53 756	103.85	1.88	14.398	10.253
10-14	52 585	43 454	121.01	17.17	1.616	9.005
15-19	51 074	41 753	122.32	1.31	9.344	6.989
20-24	40 834	46 327	88.14	34.18	18.048	0.597
25-29	48 579	50 351	96.48	8.34	22.248	18.374
30-34	38 642	38 744	99.74	3.26	11.118	6.860
35-39	38 372	32 844	116.83	17.09	15.262	4.134
40-44	27 940	24 336	114.81	2.02	12.166	4.176
45-49	25 248	17 949	140.67	25.86	7.158	10.028
50-54	19 183	15 563	123.26	17.40	2.868	7.848
55-59	14 251	10 912	130.60	7.34	6.658	18.972
60-64	11 352	11 371	99.83	30.77	11.469	38.993
65-69	6 117	5 450	112.24	12.41	22.619	31.334
70-74	4 458	4 503	99.00			
Téléchargez ce fichier et faites cet exercice			Somme =	179.02	154.97	167.56
			Nombre =	13	13	13
			Moyenne =	13.77	11.92	12.89

$${}_5RS_x = \frac{{}_5N_x^m}{{}_5N_x^f} \cdot 100$$

$${}_5RA_x^s = abs \left[100 - \frac{{}_5N_x^s}{0.5 \cdot ({}_5N_{x-5}^s + {}_5N_{x+5}^s)} \cdot 100 \right]$$

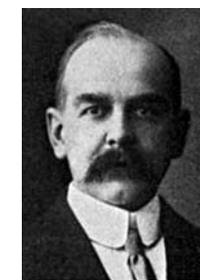
$$IP_ONU = 3 \cdot \overline{ERS} + \overline{RA}^m + \overline{RA}^f$$

$$(13.8 \times 3) + 11.9 + 12.9 = 66.1$$

Écart moyen de 100 quantifie la distorsion: <20 (précis); de 20 à < 40 (imprécis); 40 et + (très imprécis)

Qualité des déclarations d'âge au recensement

Indice de préférence d'âge de Whipple



George Chandler Whipple
(02.03.1866 – 27.11.1924)

Whipple, George C. (1919). *Vital Statistics*. New York: John Wiley & Sons. 1st edition, 640 p.

(voir p.180 édition de 1924: "The U. S. Census Bureau in studying the error due to the abnormal use of round numbers has made use of a measure termed the " Index of Concentration." This was taken to be the "per cent which the number reported as multiples of 5 forms of one-fifth of the total number between ages 23 to 62 years, inclusive.") <https://archive.org/details/vitalstatisticsi00geor>

+ **notre petit outil d'analyse « Pyramides »** sur l'EPI ou sur <http://dmo.econ.msu.ru/teaching/demo/index.htm>

+ **Pour les amateurs de R** : Riffe T, Aburto JM, Alexander M, Fennell S, Kashnitsky I, Pascariu M and Gerland P. (2019) DemoTools: An R package of tools for aggregate demographic analysis URL: <https://github.com/timriffe/DemoTools/> .

L'idée génératrice $\rightarrow \bar{x} = \frac{\sum_i^n x_i}{n} \Rightarrow \frac{\bar{x} \cdot n}{\sum_i^n x_i} = 1$

Indice de Whipple

- attirance des âges qui se terminent par 0, ou par 5, ou par 0 et 5
- se base sur l'hypothèse de la linéarité (rectangularité) sur les intervalles d'âge entre 23 et 62 ans : l'effectif de la population à l'âge que se termine par 0 et/ou par 5 doit être égale à peu près à la moyenne des effectifs aux deux âges précédents et deux âges suivants...
- au-delà de ces limites l'hypothèse de rectangularité ne tient pas (la décroissance était trop rapide à l'époque)
- facile à calculer
- varie de 100 (pas de préférence pour les 0 et les 5 à la déclaration) à 500 (il n'y a que les âges avec 0 et 5)

$$W_{0;5} = \frac{(P_{25} + P_{30} + P_{35} + \dots + P_{50} + P_{55} + P_{60}) \cdot 5}{\sum_{x=23}^{62} P_x} \cdot 100 \quad \text{ou bien à part pour les 0 et les 5} \quad W_0 = \frac{(P_{30} + P_{40} + P_{50} + P_{60}) \cdot 10}{\sum_{x=23}^{62} P_x} \cdot 100$$

Ensuite on peut calculer la moyenne de W_0 et W_5 pour obtenir le score de l'indice de Whipple général

$$\rightarrow W_{0;5} = \frac{W_0 + W_5}{2} \leftarrow W_5 = \frac{(P_{25} + P_{35} + P_{45} + P_{55}) \cdot 10}{\sum_{x=23}^{62} P_x} \cdot 100$$

Recommandation de l'ONU pour $W < 105$ (très précises); 105-110 (assez précises); 110-125 (à peu près); 125-175 (mauvaise); 175 > (très mauvaises)

Il existe les autres modifications de cet indice.

Cf. Spoorenberg Thomas, 2007, « La qualité des déclarations par âge : extension et application de l'indice de Whipple modifié », *Population*, 2007/4 (Vol. 62), p. 847-859. DOI : 10.3917/popu.704.0847. URL : <https://www.cairn.info/revue-population-2007-4-page-847.htm>

Scores de l'indice de Whipple et leurs interprétations

$$W_{0,5} = \frac{(P_{25} + P_{30} + P_{35} + P_{40} + \dots + P_{50} + P_{55} + P_{60}) \cdot 5}{\sum_{x=23}^{62} P_x} \cdot 100$$

Théoriquement l'IW varie de 0 à 500.

	Score	Interprétation
	0	aucun '0' ni '5' ne sont pas déclarés (situation théorique, données sont mauvaise)
	100	pas de préférence pour le '0' ou '5'
	<105	données sont bonnes (effet est aléatoire)
	105-110	données sont suffisamment bonne (l'effet est plutôt aléatoire)
	110-125	données sont approximatives
	125-175	forte préférence pour les '0' et les '5' (l'effet n'est pas aléatoire)
	≥ 175	données sont très mauvaises
	500	que des '0' et les '5' sont déclarés (situation théorique).



Indice de Bachi

Bachi, Roberto, 1954. "Measurement of the Tendency to Round off Age Returns." *Bulletin de l'Institut International de Statistique* (Proceedings of the 28th Session, Rome), 34(3): 129–138.

Bachi, Roberto,
16.01.1909 (Rome) –
26.11.1995 (Jérusalem)

+ notre petit outil d'analyse « Pyramides » sur l'EPI ou sur <http://dmo.econ.msu.ru/teaching/demo/index.htm>

- Cette méthode d'évaluation se base sur les mêmes principes que celle de Whipple : linéarité.
- En revanche cet indicateur évalue tous les chiffres (de 0 à 9)
- Les calculs s'appliquent aux âges de 23 à 77 ans
- L'hypothèse : l'effectif de la population à l'âge x égale à la moyenne des effectifs aux âges $x-5$ et $x+5$

$$I_0 = \left| 100 \cdot \frac{P_{30} + P_{40} + P_{50} + P_{60} + P_{70}}{0.5 \cdot P_{25} + \sum_{x=26}^{74} P_x + 0.5 \cdot P_{75}} - 10 \right| \quad I_1 = \left| 100 \cdot \frac{P_{31} + P_{41} + P_{51} + P_{61} + P_{71}}{0.5 \cdot P_{26} + \sum_{x=27}^{75} P_x + 0.5 \cdot P_{76}} - 10 \right| \quad I_2 = \left| 100 \cdot \frac{P_{32} + P_{42} + P_{52} + P_{62} + P_{72}}{0.5 \cdot P_{27} + \sum_{x=28}^{76} P_x + 0.5 \cdot P_{77}} - 10 \right|$$

$$I_3 = \left| 100 \cdot \frac{0.5 \cdot P_{23} + P_{33} + P_{43} + P_{53} + P_{63} + 0.5 \cdot P_{73}}{0.5 \cdot P_{23} + \sum_{x=24}^{72} P_x + 0.5 \cdot P_{73}} - 10 \right| \quad I_4 = \left| 100 \cdot \frac{0.5 \cdot P_{24} + P_{34} + P_{44} + P_{54} + P_{64} + 0.5 \cdot P_{74}}{0.5 \cdot P_{25} + \sum_{x=26}^{73} P_x + 0.5 \cdot P_{74}} - 10 \right| \quad \text{Idem pour 5,6 et 7}$$

$$I_8 = \left| 100 \cdot \frac{P_{28} + P_{38} + P_{48} + P_{58} + P_{68}}{0.5 \cdot P_{23} + \sum_{x=24}^{72} P_x + 0.5 \cdot P_{73}} - 10 \right| \quad I_9 = \left| 100 \cdot \frac{P_{29} + P_{39} + P_{49} + P_{59} + P_{69}}{0.5 \cdot P_{24} + \sum_{x=25}^{73} P_x + 0.5 \cdot P_{74}} - 10 \right|$$

Il est supposé que dans l'idéal (une distribution uniforme) le score des indices soit proche à 0, S'il est = 10 → le chiffre en question est absent

$$I_{Bachi} = 0.5 \cdot \sum_{n=0}^9 I_n$$

L'indice de Bachi se fabrique comme **une demi-somme** des indices partiels
Son score peut varier de 0 (la distribution uniforme) à 90 (tous les âges se terminent par le même chiffre)

Il existe la pratique de ne pas diviser la somme de indices partiels pas 2:
dans ce cas le score de l'indice de Bachi varie de 0 à 180



Robert Julius Myers,
31.10.1913 – 13.02.2010

Indice synthétique de préférence d'âge de Myers

- estime les préférences pour chaque des chiffres de 0 à 9
- peut être utilisé à toute sorte de déclaration des durées (allaitement, etc.) ou des valeurs numériques
- varie de 0 (distribution uniforme=pas de préférence) à 180 (tous les âges terminent par le même chiffre) ou à 90, si on divise par deux la somme pour que le score soit comparable à celui de l'indice de Bachi

Chiffre terminal de l'âge déclaré	Effectifs de la population		Poids pour		Population mélangée		Écart absolu du pourcentage de 10%
	commençant par l'âge 10 + a	commençant par l'âge 20 + a	col.1	col.2	Nombre (1)x(3)+(2)x(4)	pourcentage %	
a	(1)	(2)	(3)	(4)	(5)	(6)	(7)
0	1 879 327	1 544 435	1	9	15 779 242	12.18	2.18
1	1 391 999	1 081 801	2	8	11 438 406	8.83	1.17
2	1 550 738	1 220 036	3	7	13 192 466	10.18	0.18
3	1 383 402	1 080 815	4	6	12 018 498	9.27	0.73
4	1 426 235	1 102 274	5	5	12 642 545	9.75	0.25
5	1 603 138	1 268 426	6	4	14 692 532	11.34	1.34
6	1 390 982	1 073 492	7	3	12 957 350	10.00	0.00
7	1 297 957	976 244	8	2	12 336 144	9.52	0.48
8	1 346 442	1 019 407	9	1	13 137 385	10.14	0.14
9	1 140 819	847 826	10	0	11 408 190	8.80	1.20
Somme	X	X	X	X	129 602 758	100.0	7.65
Indice synthétique de préférence d'âge = 7.65 ou bien (somme ÷ 2) =							3.83

Calculs:

1. les sommes des populations dont l'âge termine par 0, 1, 2...9 entre 10 et 79 ans (ex. pour $a=1 \rightarrow P_{11}+P_{21}+P_{31}+\dots+P_{71}$)
2. le même calcul pour âges entre 20 et 89 (ex. pour $a=3 \rightarrow P_{23}+P_{33}+P_{43}+\dots+P_{73}$)
3. pondération des sommes pour obtenir les populations mélangées par le nombre de combinaisons (étapes 1 et 2)
4. transformation des populations mélangées en pourcentage du total
5. calculer l'écart du pourcentage obtenu de 10 pour chaque population et la somme des écarts

Myers, R. J. 1954. "Accuracy of Age Reporting in the 1950 United States Census." *Journal of the American Statistical Association* 49(268): 826–831.

Myers, R. J. 1940. "Errors and Bias in the Reporting of Ages in Census Data." *Transactions of the Actuarial Society of America* 41: Part II (104): 395–415.

(Reproduced in U.S. Bureau of the Census. 1951. *Handbook of Statistical Methods for Demographers*, pp. 115–125.)

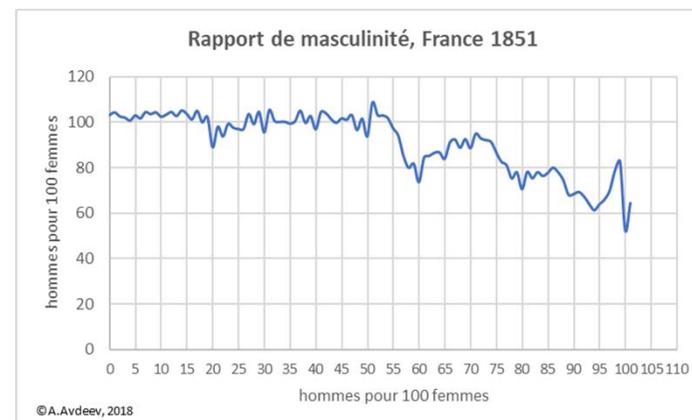
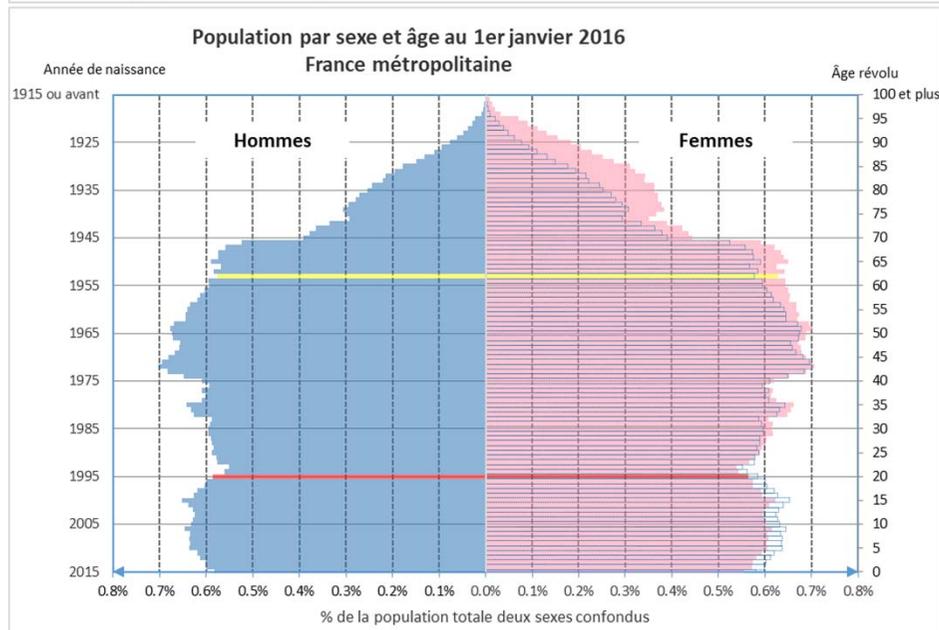
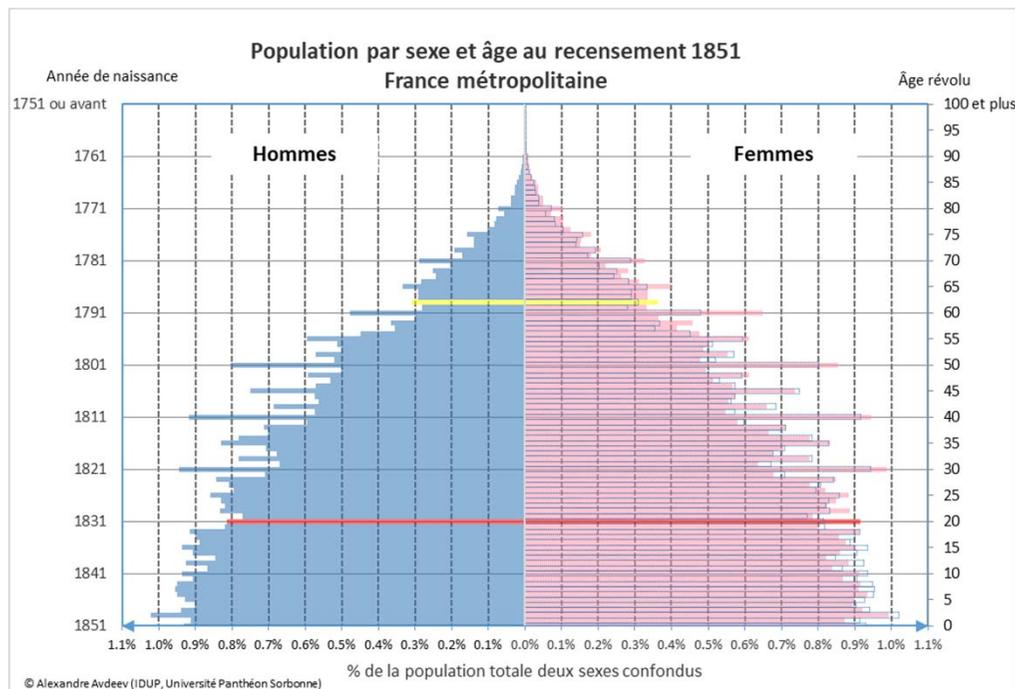
Interprétation de l'indice mélangé de Myers

- Si la somme (colonne 6) pour un chiffre **est supérieure à 10** (à 10% de la population mélangée), cela signifie sa **surreprésentation** dans les déclarations des âges (préférence pour ce chiffre)
- Si, en revanche, la somme (colonne 6) **est inférieure à 10**, cela signifie que ce chiffre est **sous-représenté** dans les déclarations des âges (on l'évite)
- S'il n'y a aucune préférence pour les chiffres, cet indice pourrait être très proche au zéro
- Il existe la pratique de ne pas diviser par 2 la somme des écarts de 10: dans ce cas le score de *l'indice de Myers varie de 0 à 180* (voir ex.c. E.Arriaga, 1994)

Pour avoir plus d'information sur la correction des défauts des données sur la distribution de la population par âge et par sexe consultez E.Arriaga et al. (1994), *Population Analysis with Microcomputer*, al, vol.1, « Presentation of techniques », documentation of the US Census Bureau, available on <https://www2.census.gov/software/pas/documentation/pamvi-archive.pdf>

Exemples des calculs : 2-TD\TD-1_Pyramide\Population de France 1851 Whipple-Myers.xlsx

Population de la France en 1851 et en 2016



	Hommes	Femmes	Ensemble
Population totale	17 777 012	17 976 515	35 753 527
Moins de 20 ans	6 555 559	6 356 119	12 911 678
de 20 à 64 ans	10 143 115	10 382 160	20 525 275
65 ans ou plus	1 078 338	1 238 236	2 316 574

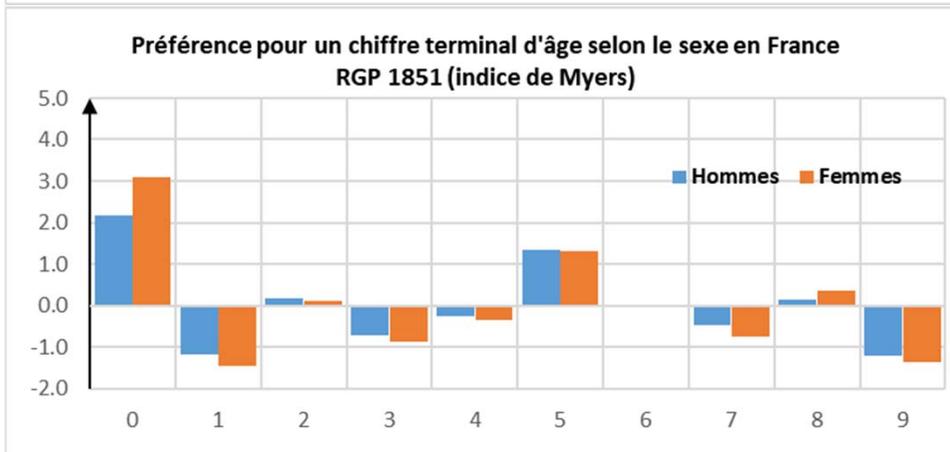
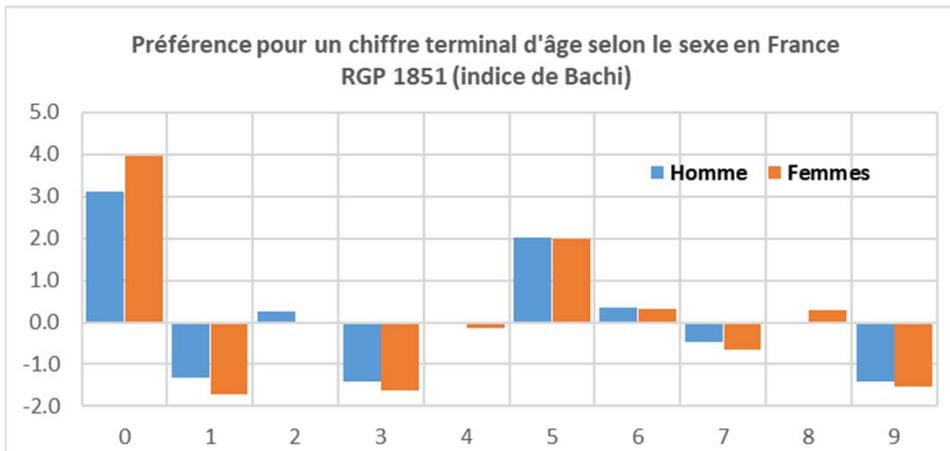
	Hommes	Femmes	Deux sexes
Age moyen	30.36	31.27	30.82
Age médian	27.00	27.73	27.365
Charge démographique p.100 personnes de 20 à 65 ans			
par <20 et 65+	75.3	73.1	74.2
par <20	64.6	61.2	62.9
par 65 et plus	10.6	11.9	11.3
par 80 et plus	1.0	1.3	1.2

	Hommes	Femmes	Deux sexes
Age moyen	39.75	42.57	41.22
Age médian	38.80	41.82	40.02
Charge démographique p.100 personnes de 20 à 65 ans			
par <20 et 65+	73.8	79.2	76.5
par <20	44.7	41.3	43.0
par 65 et plus	29.1	37.9	33.6
par 80 et plus	7.5	13.4	10.5

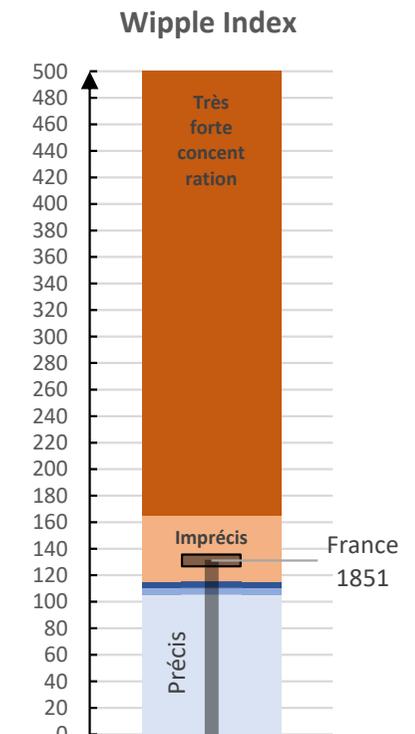
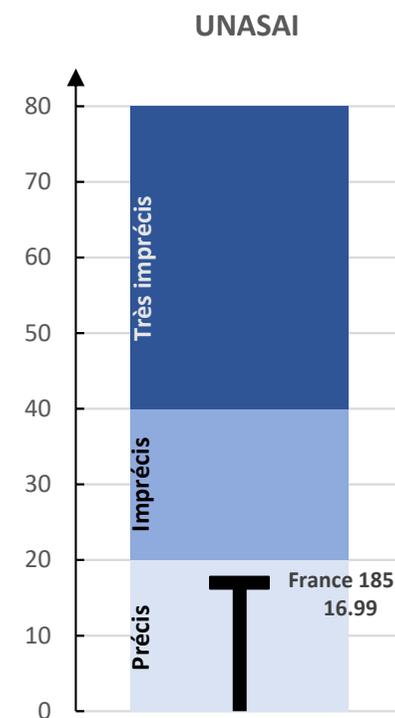
Indicateurs de la qualité des données sur la déclaration d'âge au RGP 1851 en France

Indicateur (h+f)	score	Interprétation
UNASAI	16.99	Données satisfaisantes (sur les intervalles quinquennaux)
Indice de Whipple	123.9	Données imprécises
Indice de Bachi	5.38	Données imprécises (voir graphique de TD)
Indice mélangé de Myers	8.71	Forte préférence pour le 0 et pour le 5

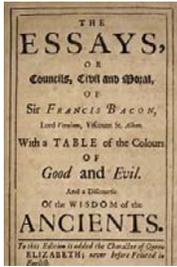
UNASAI atteste les données par groupes d'âge quinquennaux comme satisfaisantes
 WI les place dans la zone d'imprécision
 IB et IM montrent une préférence pour les 0 et le 5 dans la déclaration d'âge (les femmes déclarent 0 plus souvent que les hommes)



Conclusion: il vaut mieux utiliser les données groupées



Exercice pour autocontrôle : faire le même pour la France, 2016



L'apparition du mot et de la notion « population »

1625 – invention du mot « population » par Sir **Francis Bacon (22.01.1561–09.04.1626)**, dans **“Essayes or Counsels, Civill and Morall”** (avec 58 essais) publié en **1625**

(c'est la troisième édition, la première *“Essayes: Religious Meditations. Places of Perswasion and Disswasion. Seene and Allowed”* a été publiée en 1597 avec 10 essais, la 2^e en 1612 avec 38 essais),

Essai XV : « Of Seditious and Troubles » [« Sur excitation à la rébellion et troubles »] :



“Generally, it is to be foreseen that **the population** of a kingdom (especially if it be not mown down by wars) do not exceed the stock of the kingdom which should maintain them. Neither is the population to be reckoned only by number; for a smaller number that spend more and earn less do wear out an estate sooner than a greater number that live lower and gather more. Therefore the multiplying of nobility and other degrees of quality in an over proportion to the common people doth speedily bring a state to necessity; and so doth likewise an overgrown clergy; for they bring nothing to the stock; and in like manner, when more are bred scholars than preferments can take off.”

Texte complet est accessible sur : <http://www.authorama.com/essays-of-francis-bacon-16.html>

« Généralement on doit veiller que **la population** d'un Royaume, (spécialement si elle n'est pas fauchée par les guerres) n'excède pas les ressources du royaume nécessaires à leur *entretien*. Aucune population ne doit être évaluée uniquement par son nombre, puisque celle moins nombreuse qui dépense plus et gagne moins épuise l'Etat plus rapidement que celle nombreuse qui vive plus modestement et thésaurise davantage. Par conséquent, la multiplication de la noblesse et d'autres états de qualité dans une proportion élevée par rapports aux gens communs doit amener un Etat dans le besoin; de même pour le surcroit du clergé qui n'apporte rien, et aussi quand le nombre des gens lettrés dépasse le nombre de places que le service peut leur offrir ».

Cependant, dans les premières éditions françaises ce mot a été traduit en *peuple* ou *monde*.

Les contemporains de F.Bacon: *Galileo Galilei* (1564-1642), Italie, *René Descartes* (1596-1650), France, *Tommaso Campanella* (1568-1639) Italie-France



Denis Diderot,
né le 5 octobre 1713 à
Langres et mort le 31
juillet 1784

L'Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers est une encyclopédie française,

éditée de 1751 à 1772 sous la direction de Denis Diderot et, partiellement, de Jean Le Rond d'Alembert



Jean Le Rond d'Alembert
né le 16 novembre 1717
à Paris où il est mort le
29 octobre 1783.

Tom 13, 1765-12, Pomacies – Reggio

POPULATION, s. f. (Phys. Polit. Morale.) ce mot est abstrait, pris dans l'acception la plus étendue, il exprime le produit de tous les êtres multipliés par la génération ; car la terre est peuplée non-seulement d'hommes, mais aussi des animaux de toutes espèces qui l'habitent avec eux. La reproduction de son semblable est dans chaque individu le fruit de la puissance d'engendrer ; la population en est le résultat. Mais cette expression s'applique plus particulièrement à l'espèce humaine ; & dans ce sens particulier, elle désigne le rapport des hommes au terrain qu'ils occupent, en raison directe de leur nombre & inverse de l'espace.

Étienne Noël Damilaville,
né à Bordeaux le 21 novembre 1723 et
mort le 13 décembre 1768,



Annexe 2

La stabilité de l'effectif est-elle une règle, la croissance est-elle une anomalie ?

Dans l'Antiquité: un nombre est une quantité fixe

Platon (428-348 av.J.-C.) dans *La République* et *Les Lois* imagine une population stationnaire (5040 familles, ~20 000 citoyens libres) et une politique qui maintient cette stationnarité



Gravure de Ambrosius Holbein pour une édition de 1518. Dans le coin en bas à gauche le voyageur Raphael Hythlodæus décrivant l'île.

Pour Platon, la population de la cité était un ensemble des citoyens libres, les esclaves et les métèques en étaient exclus. De fait, la société de la cité était composée de trois populations (libres, esclaves et métèques) qui ne se mélangeaient pas, mais chacune avait une fonction sociale bien définie.

Les mêmes idées ont été retenues et développées par

- Aristote (384-322 av.J.-C.), *La Politique*
- Sir Thomas More (1478-1535)
Utopia, 1518, Londres (en latin), traduction française en 1550 à Paris: *l'Utopie ou le traité de la meilleure forme de gouvernement*
- Tommaso Campanella (1568-1639)
Civitas solis, 1623, Francfort, (appendice à la *Philosophia realis*). Traduction française en 1841

Croissance est une qualité *intrinsèque* de la population

Problème de Fibonacci (XIII s.) : multiplication des lapins

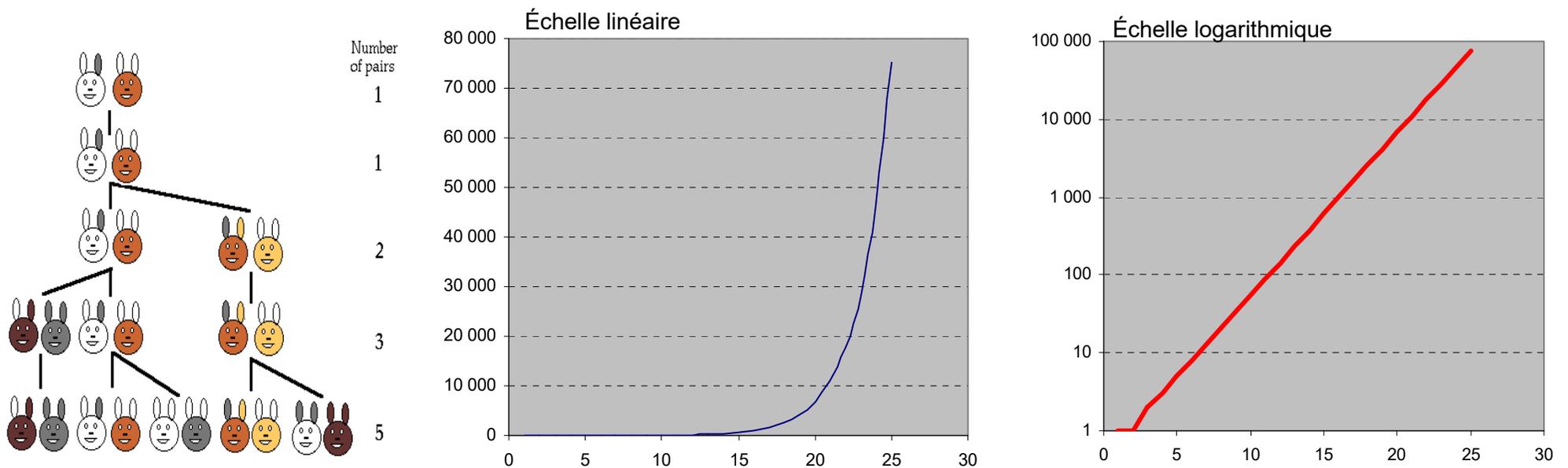
Leonardo Pisano, Fibonacci (*en italien : Figlio Buono Nato Ci*) ~1170 – ~1250 : *Liber abaci*, (Livre des calculs) rédigé en 1202, on ne dispose qu'une édition de 1228...

La croissance de lapins (des arbres) :

« Possédant initialement un couple de lapins, combien de couples obtient-on *dans douze mois*, si chaque couple engendre tous les mois un nouveau couple à compter du second mois de son existence ? » **Réponse : 144 couples**

1, 1, (1+1=2), (1+2=3), (2+3=5), = 1,1,2,3,5,8,13,21, etc... ($n_i = n_{i-1} + n_{i-2}$, $i > 2$)

La croissance de lapins durant 25 mois (75 025 couples)



Annexe 4

La Survie des humains est gérée par une loi

DIGESTORUM SEU PANDECTARUM / LIBER
TRIGESIMUSQUINTUS / TITULUS II.

AD LEGEM FALCIDIAM: 68. *Aemilius-Macer au liv.2 sur la Loi du vingtième des successions*

« Ulpien prescrit la méthode suivante pour calculer les alimens faits à quelqu'un. Les alimens laissés à quelqu'un depuis le bas âge jusqu'à vingt ans sont réputés devoir durer trente ans, et on retient sur ces alimens la Falcidie en conséquence de ce calcul. Etc. »

Domitius Ulpianus,
juriste romain, 170—228

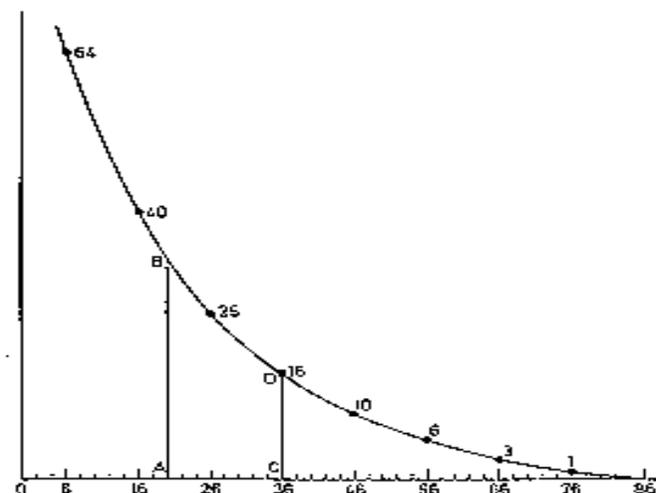
Âge du bénéficiaire	Durée de l'usufruit
0–19	30
20–24	28
25–29	25
30–34	22
35–39	20
40–49	(60-x-1)
50–54	9
55–59	7
60–	5

1662 –**John Graunt**, citoyen de Londres publie

Natural and Political Observations Mentioned in a following Index and made upon the Bills of Mortality

Viz. of 100 there dies

Within the first six years	36	The fourth	6
The next ten years, or Decad	24	The next	4
The second Decad	15	The next	2
The thrid Decad	9	The next	1



1669 Christiaan Huygens (1629-1695), Netherlands

La première représentation graphique de la fonction de distribution continue: la table de mortalité de John Graunt avec la démonstration comment peut-on trouver la durée médiane de vie après avoir atteint un âge donné

Rapport des sexes: une découverte de première importance



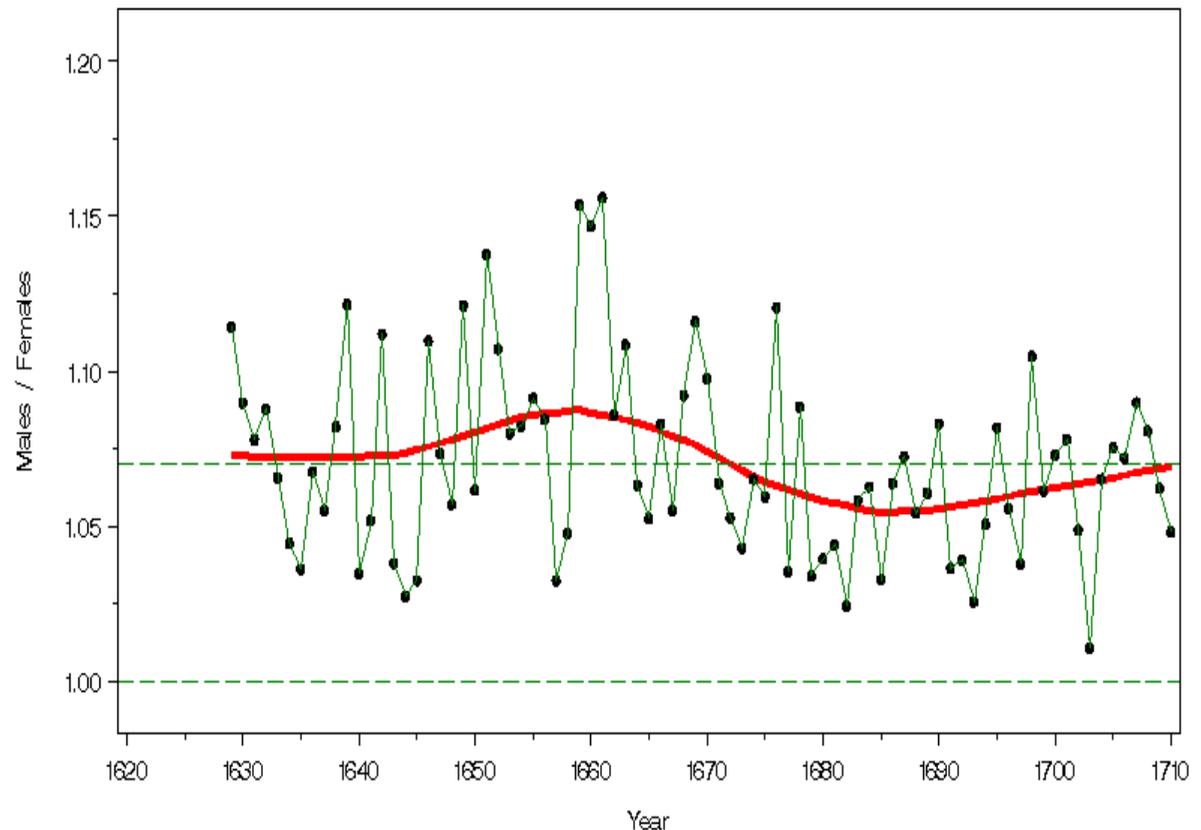
1711 - John Arbuthnot (1667-1735), Ecosse.

Il a réalisé le premier test statistique de signification (la différence entre les observations et une hypothèse « nulle ») pour démontrer « the **guiding hand of a devine** » qui maintenait un rapport des sexes à la naissance presque constant à Londres en 1629-1710

Cette priorité est cependant contestée au profit de J.P.Süssmilch, auteur de « L'ordre divine... » paru 40 ans après:

«Le pasteur Süssmilch a été le premier à tenter de traiter systématiquement la question du taux de masculinité, et il a introduit à ce sujet le constat que «pour 1000 fillettes nées, il vient 1050 garçons», une formule promise au succès parmi les démographes malgré ses problèmes évidents»

Source: Wikipédia avec une référence à « Le sexisme de la première heure. Hasard et sociologie », Éric Brian et Marie Jaisson, *Raisons d'agir*, 2007, page 22



Encore quelques éléments d'histoire :

1. Les sources :



Alessandro Farnese
1468-1549
élu pape sous le nom
Paulus III en 1534

→ Registres des naissances, des décès et des mariages :
Ordonnance de Villers-Cotterêts, 1539, François I (baptêmes et sépultures);

Concile de Trente, 1545-1563, Paul III;

Ordonnance de Blois, 1579 (mariages);

L'acte de naissance de l'état civil + registre de Naissances, de Mariages et de Décès (NMD), le 20 septembre 1792 (avant c'était BMC=baptêmes, mariages et sépultures)

→ Recensement (général de la population) :

1846, Belgique, Adolphe Quetelet ; 1851 Angleterre, William Farr ;
1851, France



A. Quetelet
1796-1874

2. Le mot : Démographie → 1855, France, **Achille Guillard** : *Éléments de statistique humaine ou **démographie** comparé*. Paris, Guillaumin et Cie (titre d'origine « Éléments de la démographie comparée »)