NORMALISATION DIRECTE

- POURQUOI UNE NORMALISATION ?
- QU'EST-CE QUE LA NORMALISATION ?
- COMMENT FONCTIONNE LA NORMALISATION ?
- TECHNIQUES DE NORMALISATION
- NORMALISATION DIRECTE
- INTERPRÉTATION DES TAUX DE NORMALISATION DIRECTE
- NOTES SUR LA NORMALISATION

À LA FIN DE CE CHAPITRE VOUS DEVEZ ÊTRE CAPABLE DE :

- COMPRENDRE POURQUOI ON NORMALISE DES DONNÉES
- DÉFINIR LA VARIABLE COMPOSITE
- CALCULER L'INDEX DE DISSIMILITUDE
- COMPRENDRE LES TECHNIQUES DE NORMALISATION
- CALCULER DES TAUX STANDARDS AVEC LA MÉTHODE DIRECTE

CHAPITRE 8

NORMALISATION DIRECTE

Pourquoi une normalisation ?

un autre outil de l'analyse de données

Un des buts principaux de l'analyse de données est de placer les données en perspective. On peut consacrer beaucoup de temps et d'énergie à la collecte de données. Souvent les bénéfices des informations collectées ne peuvent pas être exposés en entier à moins qu'on entreprenne un type d'analyse quelconque. Pour la simple raison qu'un listing entier des données peut être très long et encombrant, et peut facilement camoufler les informations qu'il contient. On entreprend alors une analyse des données pour exposer les informations cachées. On compare des chiffres, des proportions ou des taux de population dans le temps pour une même population, ou on compare des populations différentes à un moment donné.

parfois on ne peut pas comparer les vrais chiffres

Si on compare les chiffres réels, très souvent on ne prend pas en compte la taille de la population. Par exemple, si trois fois plus de personnes ont péri dans des accidents de la route en Papouasie-Nouvelle-Guinée qu'à Fidji l'année dernière, on peut dire que la Papouasie a un problème de sécurité du réseau routier. En fait, comme la population de la Papouasie est 5 fois plus élevée que celle de Fiji, c'est le contraire qui est vrai. On fait très souvent cette erreur d'interprétation.

comparer les données entre pays et régions peut induire en erreur

Les données sont également souvent utilisées à mauvais escient en démographie. Dans des domaines comme les statistiques de naissances, décès, mariages et émigration, les statistiques sont souvent utilisées à mauvais escient pour comparer des pays ou des régions.

La probabilité d'évènements démographiques tels que mariages ou décès varie manifestement selon les âges. Les populations qui ont comparativement de plus grands nombres de personnes agées sont susceptibles d'avoir plus de décès et moins de naissances chaque année qu'une communauté de taille égale qui est composée en grande partie de familles jeunes. Il en résulte par exemple que le Danemark, qui a une proportion importante de personnes âgées en comparaison du Kenya, doit avoir plus de décès pour 1000 habitants que le Kenya.

les comparaisons démographiques peuvent être affectées par la structure des âges

Lorsqu'on compare des populations on doit faire attention à ce que la structure des âges de ces populations n'affectent pas sérieusement la comparaison. Les taux bruts de naissances et de décès sont affectés par la proportion de personnes dans les tranches d'âges différentes, et peuvent donner des comparaisons trompeuses (comme pour le Danemark/Kenya). Quand on compare des taux de naissances ou de décès, il faut donc par prudence ajuster les différences dans les compositions par âge.

prenez le Guatemala et la Suède, par exemple

Quand on compare les naissances ou les décès par pays, il est important de voir au-delà des taux bruts, et de regarder la composition des populations. Le taux brut des décès de la Suède est plus important que celui du Guatemala – 11 pour 1000 habitants contre 8 pour 1000, en dépit du fait que l'espérance de vie de la Suède est 78 ans, et 63 ans seulement pour le Guatemala.

Le taux plus important de la Suède est attribué à la différence de composition des âges des deux pays. La "vieille" Suède a 18% de sa population dans la tranche d'âge 65 ans et plus, là où les décès sont le plus susceptible d'arriver, alors que le "jeune" Guatemala a une proportion de personnes âgées de seulement 3% de la population totale. Ainsi, la Suède a une proportion de décès plus élevée chaque année que le Guatemala, bien qu'elle ait de meilleures conditions sanitaires et des taux de décès plus bas par rapport à l'âge. Le succès de la politique de la Suède en matière de santé et de réduction des décès de personnes jeunes a en fait conduit à un taux brut de mortalité plus élevé !

comparer les taux en fonction des âges

Pour être cohérent, on compare les taux en fonction des âges. On peut par exemple comparer le taux annuel de mortalité pour les personnes âgées de 60 à 64 ans au Mexique et aux Ëtats-Unis. Ce taux n'est manifestement pas affecté par le nombre de personnes dans cette tranche d'âge, et montre simplement la probabilité qu'a une personne de cet âge de mourir dans une année donnée.

les variables composites 'segmentent' la population

Si on compare le taux d'un évènement (par exemple les accidents de la route pour 1000 habitants) sur deux populations, on peut calculer les taux bruts et simplement les comparer. Mais ces taux sont souvent affectés par la composition ou la structure de la population (par exemple les accidents de voitures sont plus susceptibles de survenir chez des conducteurs jeunes et inexpérimentés plutôt que chez les conducteurs plus âgés et plus expérimentés).

Lorsqu'on compare deux populations on doit tenir compte de la distribution des âges de chaque population. Dans ce cas on se réfère aux âges en tant que **variable composite**. En voici d'autres exemples :

- ☆ le sexe
- ☆ le statut dans la main d'oeuvre
- ☆ le groupe d'activités
- ☆ le nombre d'employés (pour les enquêtes sur les entreprises)

taux pour chaque catégorie ? …

Donc on conseille de calculer plutôt les taux de chaque catérgorie de la variable composite pour les comparer. Cependant, cette méthode produit en général trop de taux pour donner une image d'ensemble qui puisse être comparable (par exemple, quand le taux de mortalité d'une population A est plus bas ou plus haut qu'une population B).

… plus facile à normaliser que le taux d'ensemble

On peut aussi **normaliser** les taux d'ensemble de deux populations. Ce système applique les taux d'une des populations à la distribution de la variable composite d'une seconde population.

On peut appliquer les taux par âges d'un pays à la structure des âges d'un deuxième pays pour montrer combien de mortalité aurait le premier pays pendant un an s'il avait la structure par âge du second pays.

exemple

Par exemple, le taux brut de mortalité des États-Unis d'Amérique était de 8,5 décès pour 1000 habitants en 1979. Le taux brut de mortalité du Mexique en 1979 était de 6,2. Cependant, si la structure des âges du Mexique avait été la même que celle des États-Unis, pour cette année-là son taux de mortalité normalisé (ajusté par âges) aurait été 11,2, deux fois plus qu'aux États-Unis.

Qu'est-ce que la normalisation ?

définition

La **normalisation** est la technique qui donne la mesure résumée des taux (similaires aux taux bruts), en même temps qu'elle contrôle les différences de composition des populations. Ainsi, la comparaison des taux-types donne une meilleure comparaison des taux étudiés.

égalise les compositions

Les procédures de normalisation rendent identiques les compositions des populations et calculent les taux de manière à ce que leurs différences ne soient pas dues aux variables composites.

peut être utilisé dans de nombreux cas

On peututiliser la normalisation dans de nombreux cas. Le tableau 8.1 donne quelques taux-types souvent comparés, ainsi que les types de populations qui sont comparables, et les variables composites qui peuvent être utilisées pour normaliser les taux.

Tableau 8.1 Taux-types pouvant être normalisés

Taux	**Populations comparées**	**Variables Composites**
Taux de mortalité	Par pays Entre provinces et districts Entre hommes et femmes	Âge Sexe Situation de famille
Fumeurs	Par pays Entre provinces et districts Entre hommes et femmes	Âge Sexe Situation de famille
Revenus moyens	Par pays Entre hommes et femmes	Âge Sexe
Participation à la main d'oeuvre	Par pays Entre provinces et districts Entre hommes et femmes	Âge Sexe Situation de famille
Banqueroute	Par pays Par industries Entre PME et grandes entreprises	Industrie Taille de l'entreprise

variables composites

L'âge est une variable composite qui entraîne la variation de nombreux taux. On trouve aussi le sexe, la situation de famille, les groupes d'activités, la répartition rurale et urbaine, les races ou ethnies, le niveau d'études.

Comment fonctionne la normalisation ?

exemple

Une bonne manière de comprendre comment fonctionne la normalisation est de prendre un exemple. Supposons qu'un chercheur de l'université du Pacifique Sud ait publié les données suivantes :

Tableau 8.2 Pourcentage d'étudiants réussissant les examens de première année d'université

Hommes	60%
Femmes	44%

Source : données fictives

les hommes font mieux que les femmes ?

Le chercheur dit que c'est une preuve suffisante que les hommes font mieux que les femmes à l'université. Vous pensez que c'est un peu bizarre, donc vous décidez de faire un peu plus de recherches et vous demandez les résultats par matière. Afin de simplifier l'exercice, on dit que les étudiants passent seulement leurs examens en Science ou en Anglais :

Tableau 8.3 Pourcentage d'étudiants ayant passé la première année d'université par matière

	Sexe	
Matière	**Hommes**	**Femmes**
Science	70%	80%
Anglais	20%	40%

Source : données fictives

les femmes font mieux que les hommes ?

Si on détaille les résultats on apprend que les femmes font, en fait, mieux que les hommes dans les deux matières. Qu'est-ce qui se passe ici ? Se pourrait-il que la composition des âges de la population (c'est-à-dire le nombre d'étudiants ayant choisi la science et le nombre d'étudiants ayant choisi l'anglais) affecte les résultats ? On regarde les chiffres qui donnent les pourcentages du tableau 8.3 :

Tableau 8.4 Nombre d'étudiants ayant passé leur première année d'université par matière

	Sexe			
	Hommes		**Femmes**	
Matière	**Nombre d'étudiants ayant réussi**	**Nombre total d'étudiants**	**Nombre d'étudiantes ayant réussi**	**Nombre total d'étudiantes**
Science	560	800	80	100
Anglais	40	200	360	900
Total	600	1.000	440	1.000

Source : données fictives

les taux sont bruts

Le taux brut total d'une population est une mesure simple qui résume la situation en un chiffre ($^{600}/_{1.000}$ = 60% pour les hommes et $^{440}/_{1.000}$ = 44% pour les femmes dans l'exemple).

le taux de réussite des hommes est plus élevé

Il n'y a aucun doute que le taux de réussite des hommes est plus élevé que celui des femmes, mais ce fait ne prouve pas nécessairement que les hommes ont de meilleurs résultats que les femmes.

la science a un meilleur taux de réussite

Au tableau 8.3 on voit que la science a un meilleur taux de réussite que l'anglais. 70% des hommes réussissent en science, contre seulement 20% en anglais. 80% des femmes réussissent en science contre 40% en anglais. Au tableau 4.8 on voit que la majorité des hommes est concentrée en science. Le résultat (c'est-à-dire, les hommes ayant un taux de réussite supérieur) est dû au fait que la science est plus facile que l'anglais, et n'a rien à voir avec la faculté plus grande que pourraient avoir les hommes de passer leur première année d'université.

le sexe est la variable composite

Le problème avec les taux bruts est que les chiffres peuvent prêter à confusion à cause des différences de composition des populations comparées. La normalisation peut éliminer de telles conclusions trompeuses. Dans l'exemple des étudiants de l'USP les différences de composition sont dues à la répartition des sexes dans le nombre d'étudiants pour chaque matière.

regarder de quelle manière elles diffèrent

Comment peut-on déterminer si deux populations sont semblables ? On peut utiliser l'**index de dissimilitude** pour mesurer la différence entre deux distributions. L'index de dissimilitude montre dans quelle mesure deux ou plusieurs populations sont "différentes". Pour déterminer l'index de dissimilitude on calcule la moyenne arithmétique des différences absolues entre les taux des populations de chaque catégorie de la variable composite, et ensuite on divise par le nombre de populations. Par exemple, regardez le tableau 8.5 :

Tableau 8.5 Calcul de l'index de dissimilitude

	Pourcentage de la population		Différences de pourcentages	
Matière	**Hommes**	**Femmes**	**Réel**	**Absolu**
Science	80	10	70	70
Anglais	20	90	-70	70
Total	100	100		140

Source : Tableau 8.4

*NOTE: On compare les **pourcentages** dans chaque catégorie de la population, pas les taux.*

formule

$$\textbf{Index de dissimilitude} = \frac{\sum \text{Différence absolue}}{\text{Nombre de populations}}$$

$$= \frac{140}{2} = 70$$

grand index = grande différence

Si les deux distributions sont identiques, la valeur de l'index est zéro. Plus l'index est grand plus les populations sont dissemblables. On doit noter que la valeur de l'index dépend de la manière dont sont répartis les groupes sur la variable composite. Ceci est particulièrement important si on travaille sur les distributions des âges.

besoin de les comparer

On a maintenant établi que les deux populations sont dissemblables, et on doit pouvoir être en mesure d'établir une stratégie pour les comparer. Il y a trois possibilités évidentes :

possibilité 1

1. Déterminer le taux de réussite des hommes et des femmes si le nombre de femmes en science et en anglais est égal au nombre d'hommes.

$$\text{Taux de réussite des hommes} = \frac{600 \ \textit{ayant réussi}}{1.000 \ \textit{au total}} = 60\%$$

Pour calculer le taux de réussite des femmes, on doit d'abord calculer le nombre de femmes qui auraient passé leur examen si on avait le même nombre de femmes en science et en anglais que d'hommes. Si 800 femmes avaient pris la science (le même nombre que les hommes), et que le taux de réussite des femmes est 0,8 (80%), alors 800 × 0,8 = 640 femmes auraient passé leur examen. Si 200 femmes avaient pris l'anglais (le même nombre que les hommes), et que le taux de réussite des femmes est 0,4 (40%), alors 200 × 0,4 = 80 femmes auraient réussi leur examen. Donc, le nombre total de femmes ayant réussi serait :

$$800 \times 0{,}8 + 200 \times 0{,}4 = 720$$

si le même nombre de femmes que d'hommes avaient pris la science et l'anglais

Pour calculer le taux de réussite on divise ce total par le nombre total d'étudiants hommes (1.000). Donc, le taux de réussite des femmes est :

$$= \frac{(0{,}8 \times 800) + (0{,}4 \times 200)}{1.000} = \frac{720}{1.000} = 72\%$$

possibilité 2

2. Déterminer le taux des hommes et des femmes si le nombre d'hommes pour chaque matière est égal au nombre de femmes.

Pour calculer le taux de réussite des hommes on doit d'abord calculer le nombre d'hommes qui auraient réussi si autant d'hommes que de femmes avaient pris chaque matière. Si 100 hommes avaient pris la science (le même nombre que les femmes), et que le taux de réussite des hommes en science est 0,7 (70%), alors 100 × 0,7 = 70 hommes auraient réussi. Si 900 hommes avaient pris l'anglais (le même nombre que les femmes), et que le taux de réussite des hommes en anglais est 0,2 (20%), alors 900 × 0,2 = 180 hommes auraient réussi. Donc, le nombre total d'hommes ayant réussi serait :

$$100 \times 0{,}7 + 900 \times 0{,}2 = 250$$

Pour calculer le taux de réussite des hommes on divise ce total par le nombre total d'étudiantes (1.000) :

$$\text{Donc, le taux de réussite des hommes} = \frac{(0{,}7 \times 100) + (0{,}2 \times 900)}{1.000} = \frac{250}{1.000} = 25\%$$

$$\text{Taux de réussite des femmes} = \frac{440 \ \text{ayant réussi}}{1.000 \ \text{femmes au total}} = 44\%$$

possibilité 3

3. Déterminer le taux des hommes et des femmes si les deux populations ont chacune 500 candidats dans chaque matière.

Notez que le chiffre 500 est arbitraire. On le choisit en présumant que la moitié des hommes a pris la science et que l'autre moitié a pris l'anglais; que la moitié des femmes a pris la science et l'autre moitié l'anglais. On calcule alors les taux de réussite des hommes et des femmes en appliquant ces taux aux chiffres réels.

Taux de réussite des hommes =

$$\frac{(0{,}7 \times 500) + (0{,}2 \times 500)}{1.000}$$

$$= \frac{450}{1.000} = 45\%$$

Taux de réussite des femmes =

$$\frac{(0{,}8 \times 500) + (0{,}4 \times 500)}{1.000}$$

$$= \frac{600}{1.000} = 60\%$$

résumé

Le tableau 8.6 récapitule ces trois stratégies :

Tableau 8.6 Résumé des stratégies de comparaison

Population-type	Pourcentage d'hommes ayant réussi (1)	Pourcentage de femmes ayant réussi (2)	Différence entre les pourcentages (1) - (2)
Hommes	60	72	-12
Femmes	25	44	-19
Pré-définie	45	60	-15

Source : Tableau 8.4.

on ne peut pas comparer les pourcentages 'bruts'

Les différences entre les taux normalisés ne veulent rien dire, mais le sens des différences est montré de manière correcte, et le sera toujours, quelle que soit la population-type utilisée. On peut voir maintenant que rapporter simplement le taux brut du pourcentage d'étudiants qui réussissent leur première année d'université comme l'a fait le chercheur donne une idée fausse des données.

Techniques de normalisation

enlèvent les différences dues à la composition

Comme on l'a déjà dit, les techniques de normalisation rendent les compositions des populations identiques et calculent les taux de manière à ce que leurs différences ne soient pas dues aux variables composites.

utiliser une population-type

La normalisation implique de choisir une **population-type**, ou une **population de référence**, et de substituer de la composition de cette population à celles que l'analyste cherche à comparer.

Dans l'exemple, on a trois populations-types, ou de référence :

1	la population masculine;
2	la population féminine; et
3	une population pré-définie, ou une population fixe.

sélectionnez la poputation-type avec précaution

Le principal défaut de la normalisation en tant que technique est le choix arbitraire de la population. Comme on peut le voir avec l'exemple, on a obtenu des résultats différents selon la population-type utilisée. Cependant, c'est le sens des résultats qui nous intéresse, plutôt que la différence dans les résultats des calculs. Chacun des trois choix de population montre que les femmes ont un taux de réussite supérieur aux hommes, c'est le résultat inverse de celui du chercheur du début de l'exemple.

2 méthodes

On a deux méthodes de standardisation :

1	Directe
2	Indirecte

méthode directe ici

Dans ce chapitre on s'occupe seulement de la **normalisation directe**, qui est la méthode de normalisation favorite.

La **normalisation indirecte** est plus complexe, et on l'utilise quand les taux de chaque catégorie de la variable composite ne sont pas disponibles, ou que les données sont disponibles pour la population-type, mais ne sont pas fiables pour quelques populations. Par exemple, on peut ne pas avoir les taux de mortalité par tranches d'âges pour des Îles éloignées.

Normalisation directe

une population-type

Dans la normalisation directe, on choisit une 'population-type' pour dériver tous les taux normalisés à comparer. Tous les taux normalisés sont comparables du fait que la même population-type est utilisée.

La sélection de la population-type, théoriquement, n'a pas d'importance. Cependant, cette population devra être similaire à celles dont on compare les taux, autrement les résultats seront difficiles à interprêter.

besoins en données

Les données dont on a besoin pour effectuer la normalisation directe sont :

1 Les taux ou les proportions de chaque variable composite (par exemple les études) de toutes les populations à comparer.

2 Les populations des catégories des variables composites dans la population-type (en fait, on a seulement besoin de la distribution des variables composites de la population-type, mais les populations des catégories sont presque toujours disponibles si la distribution est disponible).

formule

Le taux de population normalisé i[1] peut alors être représenté par la formule :

$$i = \frac{\textit{(Chiffre probable s'il s'agissait de la population-type)}}{\textit{(Total de la population-type)}}$$

"chiffre probable s'il s'agissait de la population-type"

Le 'chiffre probable s'il s'agissait de la population-type' est calculé comme suit :

a. pour chaque catégorie de la variable composite, multiplier le taux de la catégorie dans la population i par le chiffre de la population-type de la catégorie; ensuite

b. résumer les résultats en (a) sur la catégorie de la variable composite.

définir R_x^i

Pour dériver une formule qui calcule le taux normalisé, on définit R_x^i comme taux de la catégorie *x* de la variable composite dans la population *i*.

définir P_x

On définit aussi P_x comme chiffre dans la catégorie *x* de la variable composite dans la population-type.

Notez que : $\sum_x P_x$ = total de la population-type.

formule

$$\textbf{Taux normalisé de la population } i = \frac{\sum_x R_x^i . P_x}{\sum_x P_x}$$

appliquer la formule à l'exemple

Prenons l'exemple des étudiants avec la formule.

La variable composite est la matière étudiée. Elle a deux valeurs, la science et l'anglais (c'est-à-dire x = science ou x = anglais). Il y a deux populations – les hommes et les femmes (c'est-à-dire *i* = hommes ou *i* = femmes).

On peut étudier toutes les combinaisons possibles, donc on a quatre taux R_x^i :

R_S^H population = hommes, matière = science

R_S^F population = femmes, matière = science

R_A^H population = hommes, matière = anglais

R_A^F population = femmes, matière = anglais

[1] *Pour la description de l'expression "population i", voir les notes de la section Standardisation à la fin de ce chapitre.*

définir les variables

La **variable composite** est la **matière**, et on a **deux jeux possibles de valeurs** P_X **:** P_S (population-type pour la science) et P_A (population-type pour l'anglais).

Par exemple, si la population-type est la population des hommes, alors P_S = 800 et P_A = 200. De même, si la population-type est celle des femmes, alors P_S = 100 et P_A = 900.

taux normalisé pour les femmes

Si on prend les hommes comme population-type, on peut calculer le taux normalisé pour les femmes. On définit les informations suivantes :

R_S^F	population = femmes, matière = science
R_A^F	population = femmes, matière = anglais
P_S	total de la population-type des hommes en sience (=800)
P_A	total de la population-type des hommes en anglais (=200)

calcul

Le taux normalisé de réussite des femmes $= \dfrac{R_S^F * P_S + R_A^F * P_A}{\sum_x P_x} = \dfrac{0{,}8*800+0{,}4*200}{1.000} = \dfrac{720}{1.000}$

= 72%

taux des hommes

Le taux brut de réussite des hommes = $\dfrac{600 \text{ ayant réussi}}{1.000 \text{ étudiants}}$ = 60%. Le taux normalisé sera le même.

On peut travailler sur le taux normalisé pour les hommes avec les définitions ci-dessus :

R_S^H	population = hommes, matière = science
R_A^H	population = hommes, matière = anglais
P_S	total de la population-type des hommes en science (=800)
P_A	total de la population-type des hommes en anglais (=200)

calcul

Le taux normalisé de réussite des hommes $= \dfrac{R_S^H * P_S + R_A^H * P_A}{\sum_x P_x} = \dfrac{0{,}7*800+0{,}2*200}{1.000} = \dfrac{600}{1.000}$

= 60%

vérification

Comme on a pris les hommes comme population-type, le taux brut est égal au taux normalisé. Si ça n'était pas le cas, on aurait fait une erreur quelque part !

exemple 2

Pour voir à nouveau comment on applique cette formule, regardons les données du tableau 8.7.

Tableau 8.7 Taux de mortalité par tranches d'âges pour l'Australie et le Territoire du Nord

Taux de mortalité par tranches d'âges (pour 1.000 habitants)			
Âge	**Australie**	**Territoire du Nord**	
0	7,494	16,177	R_0^{NT} = 16,177
1–4	0,384	1,091	
5–9	0,178	0,404	
10–14	0,178	0,241	
15–19	0,627	1,258	
20–24	0,867	1,950	
25–29	0,898	2,189	
30–34	0,952	2,151	
35–39	1,136	2,678	
40–44	1,571	3,301	R_{40-44}^{Aust} = 1,571
44–49	2,496	5,153	
50–54	4,198	9,778	
55–59	7,037	16,076	
60–64	11,706	23,706	
65–69	18,892	38,689	
70–74	30,125	44,833	
75–79	49,887	74,872	
80–84	80,637	119,677	
85–89	129,100	129,832	
90–94	202,803	158,299	
95 +	306,127	262,033	
Total	**6,998**	**4,753**	

Source : inconnue

l'âge est une variable composite

Dans cet exemple les populations sont l'Australie et le Territoire du Nord (un territoire de l'Australie). On a deux populations, donc i = l'Australie et i = le Territoire du Nord (TN). La variable composite est l'âge (x = 0, x = 1–4, et ainsi de suite).

variations dans les taux par tranches d'âges

Comme on peut voir au tableau 8.7, le taux brut de mortalité est plus grand pour l'Australie que pour le Territoire du Nord, mais les taux par tranches d'âges sont plus bas pour l'Australie (sauf pour les tranches de 90 à 94 ans et de 95 ans et plus). Ceci devrait nous faire penser que la composition par âges de la population donne une idée fausse des taux réels bruts de mortalité pour la population totale.

une structure des âges plus jeune dans le Territoire du Nord

L'explication de cet anomalie réside dans la proportion de jeunesse plus grande dans le Territoire du Nord, où les taux de mortalité sont bas, par rapport à la population totale de l'Australie où un large pourcentage de la population se trouve dans les tranches d'âges plus élevées, où les taux de mortalité sont élevés. En standardisant le Territoire du Nord on s'apercevrait de ce que serait son taux de mortalité si la structure de la population était la même que pour l'Australie.

la 4ème colonne contient les taux

Dans la quatrième colonne on extrait les chiffres avec la terminologie qu'on vient d'apprendre. D'abord, on dit que R_0^{NT} est 16,177 (le taux de l'âge 0 dans le Territoire du Nord), c'est-à-dire qu'on a 16,177 décès pour 1.000 enfants de moins d'un an. Ensuite, $R_{40-44}^{\mathrm{Aust}}$ = 1,571, le taux de la population australienne âgée de 40 à 44 ans (on a 1,571 décès pour 1.000 habitants de 40 à 44 ans).

taux normalisés

Pour calculer les taux normalisés de mortalité on prend les chiffres du tableau 8.8.

Tableau 8.8 Population de l'Australie

Âge	Population-type (Australie)	P_x
0	259.085	P_0 = 259.085
1–4	1.012.618	
5–9	1.272.208	
10–14	1.241.619	
15–19	1.364.074	
20–24	1.396.764	
25–29	1.399.663	
30–34	1.425.735	
35–39	1.328.387	$P_{35\text{-}39}$ = 1.328.387
40–44	1.294.271	
44–49	1.029.145	
50–54	846.934	
55–59	725.950	
60–64	736.868	
65–69	671.390	
70–74	510.755	
75–79	384.495	
80–84	229.828	
85–89	107.902	
90–94	36.847	
95 +	9.498	
Total	**17.284.036**	$\sum_x \mathbf{P_x}$ **= 17.284.036**

Source : Recensement de la population, 1991

population utilisée comme standard

P_0 représente le nombre d'habitants d'âge 0 dans la population totale (259.085).

$P_{35\text{-}39}$ représente le nombre d'habitants âgés de 35 à 39 ans (1.328.387).

Si on aditionne toutes les valeurs de P_x le total est égal à la population-type totale, 17.284.036 habitants ($\sum_x P_x$ = 17.284.036).

formule

$$\textbf{Taux normalisé de mortalité pour la population } \boldsymbol{i} = \frac{\sum R_x^i . P_x}{\sum_x P_x}$$

appliquer la formule aux données

On applique la formule en construisant et en remplissant un tableau, comme le tableau 8.9.

colonnes 4 et 5

Pour chaque tranche d'âge on multiplie R_x^i (le taux pour chaque tranche d'âge) fois P_x (la population-type totale). Pour simplifier les calculs, on a divisé les taux de chaque tranche d'âge par 1.000 avant de multiplier par le total de la population.

calcul

Pour la tranche d'âge 0 du Territoire du Nord, on multiplie 16,177 par 259.085 et on divise par 1.000, ce qui donne un total de 4.191.

On fait la même chose pour toutes les tranches d'âges et ensuite on aditionne tous ces chiffres de décès probables. On obtient un total de 194.730. On calcule ensuite le taux normalisé en divisant ce chiffre par la somme des valeurs P_x ($\sum_x P_x$), soit le total de la population-type (17.284.036), et en multipliant par 1.000 (comme on a divisé par 1.000 au départ). Ce qui donne un taux normalisé de mortalité pour le Territoire du Nord de 11,266 pour 1.000 habitants.

Tableau 8.9 Calcul des taux normalisés de mortalité

Âge	Taux de mortalité par tranches d'âges (pour 1.000 habitants) Australie R_x^{Aust}	Territoire du Nord R_x^{NT}	Population-type (Australie) P_x	$R_x^i P_x$ - le nombre probable de décès par tranches d'âges* Australie	Territoire du Nord
0	7,494	16,177	259.085	1.942	4.191
1–4	0,384	1,091	1.012.618	389	1.105
5–9	0,178	0,404	1.272.208	226	514
10–4	0,178	0,241	1.241.619	222	299
15–19	0,627	1,258	1.364.074	855	1.716
20–24	0,867	1,950	1.396.764	1.211	2.724
25–29	0,898	2,189	1.399.663	1.257	3.064
30–34	0,952	2,151	1.425.735	1.357	3.067
35–39	1,136	2,678	1.328.387	1.509	3.557
40–44	1,571	3,301	1.294.271	2.033	4.272
44–49	2,496	5,153	1.029.145	2.569	5.303
50–54	4,198	9,778	846.934	3.555	8.281
55–59	7,037	16,076	725.950	5.109	11.670
60–64	11,706	23,706	736.868	8.626	17.468
65–69	18,892	38,689	671.390	12.684	25.975
70–74	30,125	44,833	510.755	15.386	22.899
75–79	49,887	74,872	384.495	19.181	28.788
80–84	80,637	119,677	229.828	18.533	27.505
85–89	129,100	129,832	107.902	13.930	14.009
90–94	202,803	158,299	36.847	7.473	5.833
95 +	306,127	262,033	9.498	2.908	2.489
Total	**6,998**	**4,753**	**17.284.036**	**120.956**	**194.730**

* Le nombre probable de décès par tranches d'âges a été calculé en divisant les taux par 1.000 avant de les multiplier par le total de la population calculée $R_x^i P_x$.

résultats

$$\text{Taux normalisé de mortalité pour l'}\textbf{Australie} = \frac{120.956}{17.284.036} \times 1.000 = 6{,}998$$

$$\text{Taux normalisé de mortalité pour le }\textbf{Territoire du Nord} = \frac{194.730}{17.284.036} \times 1.000 = 11{,}266$$

vérification

Le taux pour l'Australie est 6.998, le même chiffre que pour le taux brut. C'était prévisible, car on a pris la population australienne comme standard pour la comparer avec le Territoire du Nord.

Différence BRUTE de pourcentages

La différence de pourcentages entre les taux bruts de l'Australie et ceux du TN est :

$$= \frac{(4{,}753 - 6{,}998)}{6{,}998}$$

$$= -32{,}08$$

Différence NORMALISÉE de pourcentages

La différence de pourcentages entre les taux normalisés de l'Australie et du TN est :

$$= \frac{(11{,}266 - 6{,}998)}{6{,}998}$$

$$= 60{,}99$$

différences de pourcentages indicatives

Ces différences de pourcentages sont indicatives et ne veulent pas vraiment dire grand chose. Si on avait choisi une autre population-type le pourcentage aurait été différent.

le taux du Territoire du Nord serait plus grand

Cependant, avec cet exemple on peut voir que si le Territoire du Nord avait la même structure des âges que l'Australie, le taux de mortalité serait beaucoup plus élevé. Les planificateurs et les professionnels de santé du Territoire du Nord devront tenir compte de ces incidences.

Interprétation des taux de normalisation directe

situation hypothétique …

Un taux normalisé n'a pas de valeur en soi. La valeur numérique d'un taux de mortalité normalisé par âges, interprété à la lettre, montre le taux brut de mortalité dans une population hypothétique aux taux de mortalité par âges de la population donnée, et à la composition par âges de la population-type.

… pas quand on compare des populations différentes

Le but de la normalisation est de rendre possible la comparaison des taux bruts de populations différentes en contrôlant les différences de composition des âges. C'est-à-dire, ce qu'auraient été les taux de mortalité si les populations étudiées avaient eu la même structure. Les services de la statistique utilisent cette technique quand ils comparent les taux de mortalité de leurs pays avec d'autres pays du Pacifique. La répartition des âges peut varier terriblement d'une île du Pacifique à l'autre, et ainsi fausser les comparaisons.

attention

On rencontre des problèmes quand on fait des comparaisons normalisées. Ils viennent du choix de la population-type, qui peut avoir une influence significative sur les résultats numériques obtenus. C'est particulièrement difficile quand les facteurs contrôlés dans la normalisation concernent des populations avec des différences très marquées dans leur composition. Dans l'exemple de l'université, on obtient des réponses sensiblement différentes selon le choix de la population. Comme on l'a noté, c'est le sens de ces différences qui est important, et pas leur ampleur.

variables démographiques

Il est important de savoir qu'on ne normalise que les variables qu'on ne peut pas changer, ou qui sont difficiles à modifier (par exemple la structure des âges ou des sexes d'une population). On normalise les paramètres qui, à notre avis, peuvent l'être, comme les taux de mortalité, les facteurs de santé, la participation à la main d'oeuvre.

Notes sur la normalisation

POPULATION I

on peut comparer plus de deux populations

Dans l'exemple de l'université, on a comparé seulement deux populations. On peut en comparer plusieurs à la fois (comme 5 pays mélanésiens, ou les 21 pays et territoires membres de la CPS). On doit pouvoir identifier ces populations et on le fait souvent en les codant de 1 à *I*, où I est le nombre total de populations.

exemple

Par exemple, si on veut comparer les taux de mortalité dans les pays mélanésiens, on peut les identifier en utilisant les chiffres suivants :

1 Papouasie-Nouvelle-Guinée

2 Îles Salomon

3 Vanuatu

4 Nouvelle-Calédonie

5 Fidji

La population i sera donc **n'importe laquelle** de ces cinq populations, avec la population 1 qui représente la Papouasie, la population 2 les Îles Salomon, etc..

des lettres pour identifier les populations

On code souvant aussi avec des lettres (ou des groupes de lettres). Dans les pays mélanésiens, par exemple, on peut prendre PNG, IS, Van, NC et FID. Ou peut-être la première lettre de chaque pays : P, S, V, N et F. Toutes ces méthodes se valent, et la meilleure méthode de codage est celle que les utilisateurs comprendront le plus facilement.