

Introduction à la démarche sociologique



LECTURE DE DOCUMENTS PRESENTANT DES DONNEES STATISTIQUES

Jean-François GUILLAUME, Professeur

Année académique 2017-2018

Précision initiale

Les indications qui figurent dans les pages qui suivent sont inspirées d'une note rédigée en 1986 par M. Paul Minon, Professeur à l'Université de Liège et Membre du conseil supérieur de statistique, et intitulée *Suggestions pour aider à lire un tableau statistique*.

Il entendait dans cette note « s'inspirer largement de préoccupations auxquelles les experts en méthode statistique ne sont pas toujours sensibles, peut-être parce qu'ils les jugent trop triviales ». Et il poursuivait : « si le développement de la méthode statistique et le raffinement des calculs entrepris en conséquence permettent d'aller bien au-delà de ces préoccupations, ils ne paraissent néanmoins pas autoriser à les négliger complètement ».

Partons d'un exemple. Ce document est extrait du site « Statistics Belgium », de la Direction générale Statistique et Information économique qui a pour mission de produire les statistiques nationales officielles de la Belgique.

Accidents de la route 2008													
Evolution du nombre de victimes													
Année	Nombre total de victimes	*	Nombre de victimes										
			Tués	*	Mortellement blessés	*	Tués 30 jours	*	Blessés graves	*	Blessés légers	*	
2008	65 381	-2,3	835	-12,8	109	-3,5	944	-11,9	6 782	-3,1	57 654	-2,0	
2007	66 915	0,9	958	1,3	113	-12,4	1 071	-0,4	6 997	0,0	58 847	1,0	
2006	66 344	-0,2	946	-1,0	129	-3,0	1 075	-1,3	6 999	-3,8	58 270	0,3	
2005	66 476	-	956	-	133	-	1 089	-	7 272	-	58 114	-	
2004	-	-	-	-	-	-	-	-	-	-	-	-	
2003	-	-	-	-	-	-	-	-	-	-	-	-	
2002	66 299	-0,7	1 236	-11,3	119	29,3	1 355	-8,8	8 223	-8,1	56 721	0,7	
2001	66 779	-3,8	1 394	2,2	92	-13,2	1 486	1,1	8 948	-9,1	56 345	-3,0	
2000	69 430	-	1 364	-	106	-	1 470	-	9 846	-	58 114	-	
1995	71 754	-	1 338	-	111	-	1 449	-	12 717	-	57 588	-	
1990	88 160	-	1 739	-	237	-	1 976	-	17 479	-	68 705	-	
1985	76 315	-	1 431	-	370	-	1 801	-	16 732	-	57 782	-	
1980	84 700	-	1 790	-	606	-	2 396	-	19 929	-	62 375	-	
1975	84 478	-	1 589	-	757	-	2 346	-	19 389	-	62 743	-	
1970	107 777	-	1 544	-	**	-	-	-	24 585	-	81 648	-	
1965	100 786	-	1 392	-	**	-	-	-	17 847	-	81 547	-	
1960	79 747	-	1 097	-	**	-	-	-	13 894	-	64 756	-	
1955	51 185	-	828	-	**	-	-	-	9 687	-	40 670	-	

* Augmentation ou diminution en % par rapport à l'année précédente.

** Jusqu'en 1972, les chiffres concernant les mortellement blessés sont compris dans ceux des blessés graves ou blessés légers

Tués 30 jours = tués sur place + mortellement blessés

Attention: total pas exactement égal à la somme des parties. Voir annexe 3.

A ce tableau s'ajoutent des indications reprises dans les deux annexes suivantes.

Annexe 1 – Définitions utilisées

Accident : un accident entre deux usagers de la route ou plus est considéré comme un seul accident. Seuls les accidents ayant eu lieu sur la voie publique et ayant fait des morts ou des blessés sont comptabilisés dans cette statistique. Ne sont donc pas pris en compte : les collisions et les accidents sur terrain privé ou survenus lors de compétitions sportives. Les accidents n'ayant entraîné que des dommages matériels ont cessé d'être comptabilisés depuis 1973.

Nature de l'accident : la nature de l'accident se réfère à la première collision (p.ex. une voiture heurte d'abord une autre voiture et ensuite un arbre : il s'agit alors d'une collision entre deux usagers de la route).

Tué 30 jours : Toute personne décédée sur place ou endéans les 30 jours suivant la date de l'accident.

Blessé grave : toute personne blessée dans un accident de la circulation et dont l'état nécessite une hospitalisation de plus de 24 heures.

Blessé léger : toute personne blessée dans un accident de la circulation et à laquelle ne s'appliquent pas les qualificatifs de blessé grave ou de blessé mortellement.

Accident du week-end : accidents survenant entre le vendredi 22 h et le lundi 5 h 59.

Annexe 2 – Qualité des chiffres

Les données des tués sont les données les plus fiables et les plus stables. La probabilité qu'un accident fasse l'objet de l'intervention de la police ou des parquets y est en effet la plus grande. Les données des blessés légers sont très probablement sous-estimées, notamment pour les usagers faibles (piétons, cyclistes). Sur base d'études belges et internationales, le taux d'enregistrement par la police est en effet évalué à 90% pour les accidents mortels (les données des parquets nous permettent cependant d'améliorer les résultats). Il se situe aux alentours de 50% pour les victimes hospitalisées et est de moins de 20% pour les personnes très légèrement blessées (non hospitalisées).

Sources : http://statbel.fgov.be/fr/binaries/accidents_dossier_2008_fr_tcm326-33655_tcm326-33655.xls

Avant d'interpréter les données chiffrées reprises dans le document ou d'en tirer des enseignements, il faut d'abord apprendre à « le lire ». Et pour ce faire, il faut procéder par étape.

Celui qui est parvenu à bien lire le document évitera de dire qu'en 2008, 65.381 belges ont été victimes d'un accident de la route. Pourquoi cette affirmation n'est-elle pas valide ?

Première étape :

IDENTIFIER LA SOURCE DES DONNEES QUI FIGURENT DANS LE DOCUMENT

Comment identifier la source des données reprises dans un tableau ?

Tout d'abord, en regardant où ce tableau a été publié.

Dans notre exemple, le tableau est publié sur le site de Statbel.

Statbel, on l'a dit plus haut, est l'acronyme de « Statistics Belgium », service attaché à la Direction Générale Statistique et Information économique du Service Public Fédéral « Economie, Petites et Moyennes Entreprises, Classes moyennes et Energie ». Cette terminologie, généralement abrégée par les trois lettres SPF, désigne ce que l'on appelait jadis « Ministère », soit l'ensemble de l'administration mise au service d'un ou de plusieurs ministres pour les matières qui les concernent.

Attention ! L'organisme qui édite ou publie un document reprenant des données statistiques n'a pas nécessairement collecté ces données !

Les données présentées dans cette publication de Statbel sur les accidents de la route n'émane pas de cette administration. Elles ont été collectées par d'autres services que ceux du SPF Economie, PME, Classes moyennes et Energie.

Il faut alors identifier ces autres sources. Ce qui n'est pas simple : les informations ne sont pas toujours données de façon explicite.

En identifiant la façon dont les données ont été collectées.

Il existe plusieurs façons de récolter des données statistiques.

Premièrement, par des enquêtes d'un type particulier, mais qui ne sont actuellement plus organisées en Belgique : **les recensements généraux de la population et des logements**. Les recensements doivent logiquement permettre d'atteindre chacun des éléments de la population visée.

Depuis le premier recensement à caractère scientifique initié en 1846 par Adolphe Quetelet, on a organisé à intervalles relativement réguliers (tous les dix ans), 16 recensements généraux de la population et des logements en Belgique. Dès l'origine, et particulièrement à partir de 1856, le recensement général de la population a couvert différents thèmes de la vie en société. Il ne s'est pas limité à un simple dénombrement des habitants à but administratif. Il a étendu sa portée pour constituer un utile relevé d'ordre social, économique et démographique. Avec l'amélioration des outils mis à disposition des administrations publiques, on n'a plus eu recours au recensement général pour dénombrer la population : depuis la loi du 19 juillet 1991 relative aux registres de population, le Registre national des personnes physiques (RNPP) est devenue la source unique du chiffre officiel de la population. Le recensement a conservé un important rôle dans la collecte

d'informations socio-économiques. Ce glissement dans les fonctions dévolues au recensement général de la population et des logements explique son changement d'intitulé intervenu en 2001 : « Enquête socio-économique générale 2001 ».

Plus récemment, en 2011, l'amélioration des outils techniques et informatiques a permis de recourir aux bases de données existantes et d'abandonner l'enquête menée auprès de tous les citoyens belges. C'est un important changement de méthode.

D'après <http://statbel.fgov.be/fr/statistiques/chiffres/population/recensement/>

Deuxièmement, par des enquêtes spécialement organisées **auprès d'un échantillon de probabilité**. Pour les modes de constitution d'un échantillon de probabilité, on se reportera au cours de Méthodologie de l'enquête.

Troisièmement, par des enquêtes réalisées sur **un échantillon qui n'est plus tout à fait un échantillon de probabilité**.

Exemples :

L'enquête EU-HBS (European Union – Household Budget Survey) porte sur les dépenses et les revenus des ménages. Elle vise à dresser un état des habitudes de consommation de la population sur une année. Cette enquête est aussi la source principale pour composer le panier de l'indice des prix à la consommation¹.

Dans l'enquête EU-HBS, la collaboration active des ménages participants est sollicitée : ils doivent compléter un questionnaire ménage avec l'aide de l'enquêteur, des questionnaires individuels pour chacun des membres âgés de plus de 12 ans et un carnet de recettes et dépenses dans lequel ils enregistrent tous leurs revenus et leurs dépenses pendant un mois. Ce qui ne va pas sans soulever certaines difficultés : il faut ainsi que les ménages soient effectivement capables de comptabiliser leurs recettes et leurs dépenses durant une année entière.

Par ailleurs, la démarche est assez lourde : à partir de 2012, l'enquête se déroulera tous les deux ans plutôt que chaque année, le questionnaire a été raccourci, les enquêteurs travaillent sur des tablettes et les ménages peuvent noter leurs dépenses dans une application internet plutôt que sur papier. Ces évolutions ne sont pas non plus sans susciter une série de biais. Enfin, il est prévu que les ménages qui participent de manière « satisfaisante » à l'enquête (tous les documents sont bien complétés) reçoivent une compensation financière.

Parallèlement à ce changement de méthode dans la collecte des données, l'échantillon a été élargi (5.000 ménages participants par an, au lieu de 300 ménages interrogés chaque mois jusqu'en 2010) mais il est intégré à un autre dispositif d'enquête, l'enquête sur les forces de travail

Sources : http://statbel.fgov.be/fr/statistiques/collecte_donnees/enquetes/budget_des_menages/

¹ Qu'est-ce que l'indice des prix à la consommation ? Il consiste en un instrument de mesure de l'évaluation du coût de la vie (généralement sa hausse) et de l'inflation. Le coût de la vie est appréhendé sur la base des prix d'un panier de biens et services acquis par les ménages et représentatifs de leurs habitudes de consommation. Ce panier de produits et services est composé pour une période de deux années. On peut alors mesurer le niveau des prix pour une année, mais aussi les fluctuations entre deux périodes : entre l'année de base (par exemple, 2004) (base=100) et la période de référence (par exemple, août 2012) (indice=121,36). Dans l'exemple repris, on peut considérer qu'entre fin 2004 et août 2012, le coût de la vie a augmenté de 21,36%. On notera que depuis 2008, la composition du panier de l'indice est évaluée tous les 2 ans en fonction des changements d'habitudes de consommation des ménages. L'indice tient ainsi davantage compte des nouveaux produits sur le marché de la consommation.

Les enquêtes sur l'emploi du temps réalisées en Belgique en 1999 et 2005 visaient à donner une image aussi fidèle que possible des occupations quotidiennes des personnes. Ici aussi, les participants devaient noter dans un journal toutes leurs activités, le moment auquel elles sont effectuées, la personne avec laquelle cette activité est effectuée, l'endroit où elle a eu lieu,... Ces deux enquêtes ont été associées à l'enquête sur le budget des ménages.

Les modalités prises par la collecte des données (durée du dispositif, longueur des questionnaires, utilisation d'outils informatiques, etc.) et la nature de ces données doivent inciter à une certaine prudence. On doit également considérer le fait que les réponses fournies ne sont étayées d'aucune justification et que les répondants ont pu craindre que leurs déclarations aient des répercussions sur leur sort.

Quatrièmement, dans des documents officiels émanant de pouvoirs publics.

Ce qui est alors saisi, ce ne sont pas des réalités sociales mais des réalités administratives, officielles. Ces réalités ont souvent des implications juridiques.

Reprenons ici l'exemple de départ : « les accidents de la route survenus en 2008 » et « l'évolution du nombre de victimes ».

Pour identifier la source des données, il faut relire attentivement les annexes, surtout la deuxième de ces annexes.

Il nous faudra également poser des questions complémentaires. Par exemple, que se passe-t-il lorsqu'un accident de la route survient ?

Une brochure éditée par le SPF Justice nous éclaire à ce propos.

Dès que l'accident implique des morts ou des blessés, il est obligatoire d'en aviser sur-le-champ un service de police.

(...) En cas de blessures graves le service de police n'est pas habilité pour soigner ou transporter les blessés. Si cela n'a pas encore été fait, il veillera lui-même à appeler les services médicaux nécessaires. Si les blessés graves sont déjà pris en charge par un service de secours et transportés dans un établissement hospitalier, le service de police présent sur les lieux de l'accident établira les données relatives aux personnes transportées : quelle est leur identité ? Où étaient-elles au moment de l'accident ? Comment ont-elles été impliquées dans l'accident ? Quel est leur état ? Quel est le service qui les a évacuées, à destination de quel hôpital ?

(...) Le décès sera constaté par un médecin. Si une ambulance a été appelée et qu'à son arrivée sur les lieux de l'accident la victime est décédée, celle-ci ne pourra être transportée par l'ambulance.

(...) Dans la pratique actuellement suivie en Belgique, le corps de la victime décédée doit rester sur place afin de préserver tous les éléments de preuve et de permettre ainsi aux autorités et aux spécialistes désignés par le parquet de relever ces éléments. Ces constatations doivent être faites afin de faciliter ultérieurement le déroulement de l'instruction.

(...) Par la délivrance du permis d'inhumer, le procureur du Roi autorise la famille à disposer du corps de la victime décédée.

(...) Tout décès sur la voie publique doit être communiqué immédiatement à l'administration communale de l'endroit où s'est produit l'accident. Cette déclaration sera faite par le service de police.

(Information et conseils aux familles des victimes d'accidents de la route, Service public fédéral Justice, Bruxelles, 2009)

Considérons également les précisions apportées dans une brochure éditée par l'IBSR (Casteels Y. et M. Scheers, 2008, « Evolution de la sécurité routière en Belgique 2000-2006 », IBSR, Observatoire pour la sécurité routière, Bruxelles, pp. 7-9).

En Belgique comme dans la plupart des pays européens, les informations concernant les accidents corporels sont collectées par les services de police. Pour chaque accident corporel, les forces de police remplissent un procès-verbal et, normalement, un *Formulaire d'analyse des accidents de la route avec tués ou blessés* (FAC). Ce questionnaire, reprenant un grand nombre d'informations sur les accidents, permet la création de la base de données des accidents. Pour ce faire, les données récoltées via les FAC sont envoyées au SPF Economie DG SIE où elles sont corrigées (notamment en y ajoutant les décédés 30 jours émanant des PV des parquets non repris dans la base de la police) et validées. Ensuite, la base de données est transmise à divers instituts, dont l'IBSR et les régions, pour exploitation.

En Belgique comme dans tous les autres pays possédant des statistiques sur les accidents de la circulation, une partie des accidents corporels ne sont pas enregistrés par les services de police. Les raisons de ce sous-enregistrement sont multiples et d'ailleurs partiellement inconnues en Belgique : on sait qu'une partie des accidents corporels ne sont pas enregistrés tout simplement parce que les usagers impliqués, malgré l'obligation légale, n'appellent pas les forces de l'ordre ; il arrive aussi que les policiers n'arrivent jamais sur place, ou que le FAC ne soit pas rempli, ou se perde, etc. Il n'existe pas de recherches approfondies récentes sur le sujet en Belgique. On estime toutefois que le taux d'accidents non enregistrés dépend de la gravité des accidents (plus un accident est grave, plus les chances d'enregistrement sont importantes), du type d'utilisateur impliqué (les accidents impliquant un cycliste sont moins souvent enregistrés que les accidents impliquant une voiture), de l'âge des usagers impliqués, etc.

D'où proviennent alors les données présentées dans le document de départ ?

.....

.....

.....

.....

.....

.....

Comment ont-elles été rassemblées ?

.....

.....

.....

.....

.....

Il faut alors témoigner d'une certaine prudence ou d'une certaine réserve à l'égard des données présentées. Quelles sont les limites qui peuvent nous inciter à cette prudence ou à cette réserve ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Les données présentées dans le document intitulé « Accidents de la route 2008 » ne reflètent que très imparfaitement les réalités vécues qui leur sont liées. Il faut donc les manipuler avec prudence. Les choses ne sont pas toujours aussi simples qu'elles ne paraissent de premier abord.

Cette prudence et cette méfiance systématiques obligent à d'abord à s'interroger sur la définition des termes utilisés dans le document. Dans le cas présent, qu'est-ce qu'un « accident de la route » ?

D'ultimes appels à la précaution peuvent encore être adressés lorsque l'on envisage les sources des données.

Primo, parce que les données collectées constituent la partie émergée des faits originels. C'est par exemple le cas lorsqu'il s'agit d'étudier la délinquance à travers des données relatives aux condamnations prononcées par les tribunaux et même à travers des données relatives aux infractions constatées ; de connaître les revenus sur base des déclarations à l'impôt sur les personnes physiques.

Secundo, parce qu'un même document officiel peut fournir des données fiables et des données imprécises. C'est par exemple le cas des actes de déclaration de mariage qui fournissent des informations fiables sur l'âge, la nationalité, la situation matrimoniale antérieure des conjoints, mais sont beaucoup plus imprécis sur leur situation professionnelle.

Tertio, parce que ces données ont pu être ajustées en fonction de données antérieures. L'organisme qui les publie procède donc à une estimation.

Deuxième étape :

IDENTIFIER SUR QUOI PORTENT LES DONNEES PRESENTEES

Il faut ici procéder de façon systématique, et en prolongeant les réflexions qui ont été entamées sur la source des données. Quatre questions doivent être posées :

1. Quelle est la population concernée ?
2. A quel(s) moment(s) du temps cette population est-elle étudiée ?
3. Cette population est-elle caractérisée par des critère(s) de différenciation ?
4. Quelle(s) sorte(s) d'expression statistique retrouve-t-on ?

Nous l'avons vu précédemment, il ne faut pas se limiter à la lecture du titre, mais chercher des informations complémentaires dans la présentation des données, dans les notes qui l'accompagnent ou dans le contexte où il s'insère.

1.- La population concernée

Pour identifier la « population » concernée, il faut tout d'abord définir ce qu'est une population.

Une population constitue une collection d'éléments suffisamment homogènes pour s'insérer dans une même série statistique.

Puis définir ce qu'est une série statistique.

Une série statistique est constituée de la suite des valeurs attribuées à chaque élément constitutif d'une population pour un caractère donné.

Cette étape est déterminante pour la suite du travail de lecture et d'analyse d'un document reprenant des données statistiques. Parce que si la population concernée n'a pas été identifiée correctement, le risque est grand de tirer des conclusions erronées !

Dans notre situation de départ, qui envisage « les accidents de la route en 2008 », il est erroné de conclure qu'en 2008, 65.381 belges ont été victimes d'un accident de la route. Et si cette erreur est commise, c'est parce que la population concernée n'a pas été identifiée correctement.

Comment faire alors pour ne pas se tromper ?

Premièrement, en faisant preuve de méfiance sur le sens des mots. Les termes et les mots utilisés pour désigner cette population doivent être considérés avec prudence.

Exemples :

1. Dans l'enquête européenne sur le budget des ménages, une personne seule peut constituer un ménage alors que le sens commun y verrait plutôt une unité de vie collective... Ainsi, pour Statbel, « un ménage est constitué soit par une personne vivant habituellement seule, soit

par deux ou plusieurs personnes qui, unie ou non par des liens de parenté, occupent habituellement un même logement et y vivent en commun ».

2. Lorsqu'il s'agit de comptabiliser les déclarations de « cohabitation légale », il n'est pas question de recenser le nombre de personnes qui sont en couple sans être mariées. Cette notion de « cohabitation légale » désigne une situation de vie commune de deux personnes qui ont fait une déclaration, conformément à l'article 1476 du Code civil. Pour faire une déclaration de cohabitation légale, les deux personnes qui le désirent ne doivent pas être liées par un mariage ou par une autre cohabitation légale. C'est un acte par lequel les personnes qui vivent sous un même toit fixent les contributions respectives aux charges de la vie commune. Ainsi, Paul, 75 ans, et sa sœur Virginie, 72 ans, qui vivent dans la maison familiale depuis le décès de leurs parents, peuvent faire une déclaration de cohabitation légale. Ce que ne pourraient pas faire Casimir, 40 ans, et Philomène, 38 ans, mariés et parents de deux enfants.
3. C'est encore le cas lorsqu'il s'agit de déterminer le chiffre de la population « active », par exemple en Région wallonne. La population active n'est pas composée des personnes qui exercent une activité professionnelle sur le territoire de la Région wallonne, mais elle est composée des personnes résidant sur ce territoire, qui y travaillent ou qui travaillent ailleurs (population active occupée), ou qui sont sans emploi et à la recherche d'un emploi (population active au chômage) (source : <https://www.iweps.be/indicateur-statistique/population-active-taux-dactivite-taux-demploi-taux-de-chomage-administratifs-commune/>).
4. Dans la situation de départ, la notion d'accident de la route est définie de façon précise.

Accident : un accident entre deux usagers de la route ou plus est considéré comme un seul accident. Seuls les accidents ayant eu lieu sur la voie publique et ayant fait des morts ou des blessés sont comptabilisés dans cette statistique. Ne sont donc pas pris en compte : les collisions et les accidents sur terrain privé ou survenus lors de compétitions sportives. Les accidents n'ayant entraîné que des dommages matériels ont cessé d'être comptabilisés depuis 1973.

Deuxièmement, en considérant qu'une population n'est pas nécessairement un ensemble composé d'êtres humains. Le mot a une portée plus générale en statistique. On peut être confronté à :

- une population faite de collectivités groupant des quantités variables d'êtres humains (ménages, communes, entreprises, écoles, hôpitaux,...) ;
- une population composée d'autres êtres que d'êtres humains (animaux, végétaux, marchandises,...) ;
- une population composée d'unités physiques ou monétaires (kilomètres de voies, tonnes de marchandises, montants de recettes ou de dépenses, etc.) ;
- une population constituée d'événements. C'est le cas dans l'exemple suivant où les auteurs d'une brochure portant sur les mariages et les divorces célébrés en 2006, et éditée par la Direction Générale Statistique et Information Economique du SPF Economie, Classes Moyennes et Energie, apportent quelques précisions importantes sur les données présentées.

- La statistique des mariages est établie à partir des bulletins de Modèle II (tel que prévu par l'Arrêté royal du 26 août 1966) qui sont établis par les administrations des communes dans lesquelles ont été célébrés ces mariages. Elle reprend donc tous les mariages célébrés devant les officiers d'état civil des communes belges ; mais elle ne reprend pas ceux célébrés à l'étranger, dans les consulats ou dans les corps belges casernés à l'étranger, ni ceux célébrés sur le territoire belge devant des officiers d'état civil jouissant du statut d'extraterritorialité ⁽¹⁾. Leur répartition géographique peut se faire selon la commune de mariage. Il en résulte une statistique de fait, comme dans les tableaux 1 et 4. Elle peut également se faire selon la résidence habituelle de l'un ou de l'autre conjoint avant le mariage, comme dans les tableaux 2 et 3, où apparaissent de ce fait des mariages de personnes domiciliées à l'étranger.
- Depuis le 1er juin 2003, la loi autorise le mariage entre personnes de même sexe (Loi du 13 février 2003 ouvrant le mariage à des personnes de même sexe et modifiant certaines dispositions du Code civil) et depuis lors, un peu plus de 5.000 mariages de ce type ont été célébrés en Belgique. Le bulletin de mariage cependant ne mentionne pas explicitement le sexe de chacun des conjoints, en considérant toujours qu'ils sont de sexes différents (à gauche les renseignements relatifs à l'époux – aujourd'hui le premier conjoint – et à droite, les renseignements relatifs à l'épouse – aujourd'hui le second conjoint).
- La présente statistique ne permet donc pas de distinguer les mariages selon qu'ils concernent des personnes de sexes différents ou des personnes de même sexe. Et les caractéristiques de chacun des époux ne doivent plus être interprétées globalement comme étant respectivement les caractéristiques de l'époux et les caractéristiques de l'épouse.
- **Il faut encore préciser que cette statistique est une statistique d'événements (les mariages) et non une statistique de personnes.** Si chaque événement-mariage implique bien 2 personnes-mariées, la comparaison entre les 2 statistiques n'est pas toujours aisée lorsque les données utilisées proviennent de sources différentes. Ainsi, une statistique des mariages, et particulièrement des mariages de personnes de même sexe, est extraite du Registre national. Il s'agit d'une statistique de personnes, qui n'est pas directement transposable en statistique d'événements, du fait d'erreurs ou d'oublis éventuels de transcription, mais surtout du fait que toutes les personnes qui se marient en Belgique ne sont pas nécessairement inscrites au Registre national. Le rapprochement des 2 sources ne peut donc se faire sans précautions.

- Source : http://statbel.fgov.be/fr/binaries/Mariages%20et%20divorces%20en%202006_tcm326-119496.pdf

- ⁽¹⁾ Il s'agit ici des mariages célébrés dans des ambassades ou consulats étrangers ayant leur siège en Belgique.

Question :

Dans les données portant sur les accidents de la route 2008, identifiez la population de référence. Prenez en considération la source des données, la façon dont les données ont été produites, la distinction entre statistique d'événements et statistique de personnes.

.....

.....

.....

.....

.....

.....

2.- Le moment du temps

Règle générale :

D'ordinaire, il s'agit du moment où les données relatives à cette population ont été collectées.

Le tableau saisit, photographie l'état d'une population en un moment fort bref, souvent en une seule journée.

Exception :

Quand la population est composée d'événements, le tableau reprend des événements qui se sont déroulés tout au long d'une période de temps assez étendue (souvent une année entière).

Attention :

Il faut toujours bien déterminer en quoi consiste exactement la population étudiée.

Des données portant sur les chômeurs relevés dans le pays en mars 2009 n'auront pas le même sens si elles ont été obtenues un jour déterminé de ce mois ou si elles additionnent toutes les personnes qui ont été mises en chômage durant le mois.

Les données présentées dans un tableau statistique ne sont pleinement valables que pour le moment du temps auquel elles se rapportent.

Ce moment peut avoir été exceptionnel.

Il appartient en tout cas au passé, avant qu'un tableau soit publié et qu'un lecteur en prenne connaissance.

Méfiance donc si on veut appréhender les réalités actuelles au départ d'un tableau.

Les difficultés sont minimales si les données sont extrêmement stables ou ne se modifient que faiblement et lentement.

Les difficultés sont plus importantes si les données sont susceptibles de se modifier fortement et vite.

Question :

Dans les données portant sur les accidents de la route 2008, identifiez le ou les moments du temps.

.....

.....

.....

.....

.....

.....

.....

3.- Les critères de différenciation

Si les éléments d'une population se différencient (se caractérisent différemment sous tel ou tel aspect), on peut alors regrouper les éléments de cette population en catégories ou en classes.

Quels critères de différenciation peuvent être utilisés ?

Un attribut qui ne peut pas s'exprimer par une valeur numérique :

- Caractériser des individus selon le sexe, l'état civil, la nationalité,...
- Caractériser des entreprises selon leur régime juridique (SA, SPRL,...) ou d'après la nature de leur production.
- Caractériser des décès d'après leur cause initiale.
- Caractériser des sommes d'argent inscrites au budget de l'Etat selon leur affectation.

Un attribut qui s'exprime à travers une valeur numérique et qui est une « variable » au sens strict du terme :

- Caractériser des travailleurs selon leur âge, le montant de leur rémunération.
- Caractériser des communes selon leur superficie, le nombre de leurs habitants.
- Caractériser des déclarations fiscales selon le montant des sommes déclarées.

Attention :

Il arrive qu'une population soit étudiée à différents moments du temps : par exemple, on recense le nombre de mariages célébrés chaque année au cours d'une période donnée, ou durant la même année par mois ou par trimestre.

Nous ne considérerons pas ici qu'il s'agit d'un critère de différenciation à proprement parler.

Pourquoi ? Parce que nous partons de l'idée qu'une population constitue une collection d'éléments suffisamment homogènes pour s'insérer dans une même série statistique.

Or la distribution des événements dans une perspective temporelle (comme par exemple, le nombre de mariages célébrés depuis 1831 en Belgique) n'a pas le même statut que la distribution des événements selon certains critères au cours d'une période de référence (comme par exemple, la distribution des mariages selon la différence d'âge des conjoints en 2012 en Belgique) : la première série s'inscrit dans un intervalle ouvert alors que l'autre série s'inscrit dans un intervalle clôturé.

Question :

Dans les données portant sur les accidents de la route 2008 et sur la cohabitation légale, identifiez les critères de différenciation.

.....
.....
.....
.....
.....
.....
.....
.....
.....

4.- Les expressions statistiques

On peut retrouver différentes formes, de la plus simple à la plus élaborée :

- des fréquences simples, des chiffres absolus déterminés sur base d'un seul critère de différenciation ;
- des fréquences relatives, souvent des pourcentages (%), mais pas uniquement (par exemple, ‰) ;
- la combinaison de fréquences simples et de fréquences relatives, sur base de deux ou plusieurs critères de différenciation ;
- des expressions statistiques raffinées (moyennes, médiane, coefficient, taux, indice, ratio, instrument de mesure de la dispersion, instrument de mesure ou d'appréciation de l'association entre des critères de différenciation).

Question :

Dans les données portant sur les accidents de la route 2008, identifiez les expressions statistiques.

Accidents de la route 2008 :

.....
.....
.....
.....

Troisième étape :

PRECISER LE NOMBRE DE TABLEAUX FIGURANT DANS LE DOCUMENT DE REFERENCE

Nous partirons une fois encore de l'idée qu'un tableau porte sur la distribution des éléments constitutifs d'une population à un moment du temps, sur la base d'un ou de plusieurs critères de différenciation.

L'homogénéité de la population autorisera certains calculs au départ des données présentées dans un tableau.

Pour préciser si nous sommes face à un tableau ou à plusieurs tableaux, il faut apprendre à identifier les « faux » critères de différenciation de la population et l'absence d'homogénéité de la collection des éléments présentés.

Reprenons l'exemple des accidents de la route 2008.

Sommes-nous face à un tableau ou à plusieurs tableaux ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Exercice individuel

Considérez avec attention le document figurant sur la page suivante.

Les données ont été extraites du site web de l'INSEE (Institut National de la Statistique et des Etudes Economiques, France, http://www.insee.fr/fr/themes/tableau.asp?reg_id=98&ref_id=CMPTF05446)

Répondez ensuite aux questions suivantes.

D'où proviennent les données ?

.....

.....

.....

.....

Quelle est la population concernée ?

.....

.....

.....

.....

A quel moment du temps est-elle saisie ?

.....

.....

.....

.....

Quels sont les critères de différenciation ?

.....

.....

.....

.....

La forme des expressions statistiques ?

.....

.....

.....

.....

Avons-nous à faire à un tableau ou à plusieurs tableaux ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Fréquentation du cinéma dans le monde

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007 (r)	2008
Nombre d'entrées (en millions de spectateurs)											
Allemagne	148,9	149,0	152,5	177,9	163,9	149,0	156,7	127,3	136,7	125,4	129,4
Espagne	112,1	131,3	135,3	146,8	140,7	137,5	143,9	127,6	121,6	116,9	107,8
France métropolitaine	170,6	153,6	165,8	187,5	184,4	173,5	(r) 195,7	175,5	188,8	177,9	189,7
Italie	118,5	103,5	104,2	113,3	115,6	110,5	116,3	105,6	106,1	116,4	111,6
Royaume-Uni	135,2	139,1	142,5	155,9	175,9	167,3	171,3	164,7	156,6	162,4	164,2
Union européenne (1)	828,0	810,0	844,0	935,0	938,0	890,0	1 006,0	894,0	926,0	917,0	(e) 920,0
Etats-Unis (r)	1 438,0	1 440,0	1 385,0	1 437,0	1 597,0	1 520,0	1 484,0	1 378,0	1 395,0	1 400,0	1 364,0
Japon	153,1	144,7	135,4	163,3	160,8	162,3	170,1	160,5	164,3	163,2	160,5
Australie	79,8	88,0	82,2	92,5	92,5	89,8	91,5	82,2	83,6	84,7	84,6
Nombre d'entrées moyen par habitant											
Allemagne	1,8	1,8	1,9	2,2	2,0	1,8	1,9	1,5	1,7	1,5	1,6
Espagne	2,9	3,3	3,4	3,7	3,4	3,3	3,5	2,9	2,8	2,6	2,4
France métropolitaine (r)	2,9	2,6	2,8	3,2	3,2	3,0	3,3	3,0	3,2	3,0	3,2
Italie	2,1	1,8	1,8	1,9	1,9	1,9	2,0	1,9	1,8	1,9	1,9
Royaume-Uni	2,3	2,4	2,4	2,6	2,9	2,8	2,8	2,7	2,6	2,7	2,7
Union européenne (1)	2,2	2,2	2,3	2,5	2,5	2,3	2,2	1,9	2,0	1,9	(e) 1,9
États-Unis	5,5	5,4	5,2	5,3	5,7	5,4	5,2	4,7	4,8	4,6	4,5
Japon	1,2	1,1	1,1	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3
Australie	4,4	4,8	4,5	5,1	4,7	4,5	4,6	4,0	4,0	4,0	4,0

e : données estimées

r : données révisées

(1) : Union européenne à 25 pays de 2004 à 2006 et à 27 pays à compter de 2007.

Source : Centre national de la cinématographie.

Quatrième étape :

DECRYPTER LE TABLEAU

Nous considérons que nous avons affaire à un tableau lorsque les données qui y sont présentées sont relatives à une population, c'est-à-dire qu'elles constituent une collection d'éléments suffisamment homogènes pour s'insérer dans une même série statistique.

Notre utilisation du terme « tableau » est donc plus restrictive que celle du sens commun.

Pour « entrer » dans le tableau (ou dans les tableaux) repris sur le document analysé, il convient d'abord d'en préciser la forme.

1.- Les tableaux de fréquences à une seule entrée

Un tableau statistique est un tableau de fréquences lorsque les éléments d'une population sont répartis d'après le nombre d'éléments correspondant à chacune des modalités distinguées sur base d'un critère de différenciation.

Un tableau de fréquences reste à un tableau à une seule entrée tant qu'il n'y est tenu compte que d'un seul critère de différenciation.

Un exemple :

Prénom des garçons nés en 2014 en Belgique

Louis	622
Lucas	619
Arthur	610
Adam	575
Noah	562
Liam	513
Mohamed	468
Nathan	455
Jules	442
Mathis	428
Victor	420
Gabriel	393
<i>Etc.</i>	

1.1.- Leur présentation

Un tableau de fréquences à une seule entrée est le plus souvent disposé en lignes. Une ligne y est prévue pour chacune des modalités distinguées sur base du critère de différenciation retenu.

Parfois, des tableaux de fréquences à une seule entrée sont disposés en colonnes et sur une seule ligne, en prévoyant alors une colonne pour chacune des modalités distinguées sur base du critère de différenciation envisagé.

Le plus souvent, le tableau mentionne en dernière ligne le total des éléments qui y sont dénombrés. Parfois, cette mention figure en première ligne. Dans le cas d'un tableau disposé en colonnes, le total est mentionné dans la première colonne ou dans la dernière colonne.

Parfois aussi, on ne trouve aucune mention du total.

On peut également retrouver des totaux partiels additionnant les éléments repris sur plusieurs modalités apparentées du critère de différenciation envisagé.

Précisez pour les données sur les accidents de la route, pour la cohabitation légale et pour la fréquentation du cinéma dans le monde, s'il s'agit bien d'un tableau de fréquences à une entrée. Identifiez ensuite la disposition des données, le total des éléments dénombrés et les éventuels totaux partiels.

Accidents de la route 2008 :

.....

.....

.....

.....

Cohabitation légale :

.....

.....

.....

.....

Fréquentation du cinéma dans le monde :

.....
.....
.....
.....

1.2.- Les fréquences absolues

Un tableau de fréquences à une seule entrée présente des fréquences absolues quand il dénombre et répartit les éléments d'une population qui ont pu être effectivement repérés dans la réalité.

Les éléments ne sont pas toujours dénombrés par unités. On peut les dénombrer en multiples d'unités (milliers, millions, milliards,...). De ce fait, des virgules et des décimales peuvent apparaître dans ce tableau qui reste cependant un tableau de fréquences absolues.

Attention : ne pas perdre de vue les multiples d'unités quand on procède à des opérations de calcul statistique au départ de ces tableaux.

Précisez pour les données sur les accidents de la route, pour la cohabitation légale et pour la fréquentation du cinéma dans le monde, la référence utilisée pour comptabiliser les observations.

Accidents de la route 2008 :

.....
.....

Cohabitation légale :

.....
.....

Fréquentation du cinéma dans le monde :

.....
.....

1.3.- Les fréquences relatives

Les fréquences relatives introduites dans un tableau de fréquences à une seule entrée montrent comment une population se répartit sur base d'un critère de différenciation. Mais on n'utilise plus les « contingents » d'éléments, on montre comment la population se répartit proportionnellement, par référence à un nombre conventionnellement fixé.

Il est alors plus facile de comparer des données portant sur des populations qui n'ont pas un volume identique.

Quel est ce nombre conventionnellement fixé ?

On pourrait utiliser l'unité : les fréquences relatives consisteraient alors en fractions ordinaires ou en fractions décimales.

Mais d'habitude, les fréquences relatives sont calculées en référence à cent unités et s'expriment donc en pourcentages qui sont souvent assortis d'une ou de deux décimales.

Parfois, elles sont calculées en référence à mille ou à dix mille unités, ce qui revient à calculer des pourcentages accompagnés respectivement d'une ou de deux décimales.

Les fréquences relatives accompagnent généralement les fréquences absolues, dans une colonne juxtaposée à la colonne réservée aux fréquences absolues.

Si les fréquences absolues sont disposées sur une seule ligne, les fréquences relatives occupent la ligne suivante.

Parfois, les fréquences relatives sont reprises dans un tableau distinct du tableau consacré aux fréquences absolues.

Les fréquences relatives n'ont pas beaucoup de sens quand elles concernent une population dont le volume est inconnu ou dont le volume reste insuffisant.

Des pourcentages sont trompeurs quand la population effectivement dénombrée ne compte pas plus d'une centaine d'unités.

Des décimales accolées à des pourcentages ne se justifient que si la population est de loin plus nombreuse.

En définitive, les fréquences absolues offrent autant d'intérêt. Elles sont déjà fort instructives par elles-mêmes et peuvent être facilement transformées en fréquences relatives si cela paraît nécessaire.

Précisez pour les cas des accidents de la route et de la fréquentation du cinéma dans le monde, si l'on a fait usage de fréquences relatives pour présenter les données et si oui, dans lequel des tableaux.

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.4.- Quelles informations peut-on extraire d'un tableau de fréquences à une seule entrée ?

1°) Comparer les fréquences absolues et les fréquences relatives qui y figurent pour :

- établir à quel point la distribution des fréquences est équilibrée ou non ;
- identifier la modalité du critère de différenciation qui correspond à la plus haute des fréquences (la catégorie « modale » ou le « mode » de la série constituée par le tableau) ;
- le cas échéant, constater que la série comporte deux ou plusieurs modes ;
- si la série ne comporte qu'un seul mode, établir si ce mode rassemble la majorité absolue des éléments de la population ou une majorité relative ;
- identifier des modalités aux plus faibles effectifs ;
- identifier des modalités auxquelles aucune fréquence ne correspond au sein de la population dénombrée ;
- ordonner les différentes modalités d'après l'importance des fréquences qui y correspondent ;
- calculer des rapports entre des fréquences qui correspondent aux différentes modalités.

2°) Se montrer critique à l'égard des catégories établies dans le tableau : la façon dont on a opéré le découpage en modalités a pu gonfler artificiellement l'un ou l'autre des fréquences.

3°) Examiner de près les catégories « résiduelles » (appellation du type : « autres », etc.) pour voir ce qu'elles peuvent recouvrir.

4°) Eventuellement regrouper plusieurs des modalités pour créer de nouvelles catégories.

5°) Lorsque le critère de différenciation mis en œuvre dans le tableau se traduit par des valeurs numériques, et qu'il devient donc une « variable » au sens strict du terme, on a d'autres possibilités. On peut calculer différentes sortes d'expressions statistiques.

- On peut ordonner les valeurs d'une variable et les additionner de manière à calculer des fréquences « cumulées ».

Exemple : des individus répartis par âge. On peut établir combien d'entre eux ont dépassé ou n'ont pas dépassé tel ou tel âge.

- On peut calculer la valeur « médiane », soit la valeur de la variable qui sépare en deux contingents d'importance égale les éléments d'une série ordonnée.
- On peut calculer les écarts interquartiles ou interdéciles.

On peut calculer par référence à la moyenne, la variance ou l'écart-type, pour préciser la signification d'une mesure de la tendance centrale et discerner à quel point une population est hétérogène.

- On peut procéder à une représentation graphique des données (histogramme, système de coordonnées,...) afin de percevoir quelle forme revêt la distribution de ces données.

6°) On peut extraire d'autres informations en ne considérant plus uniquement les données contenues dans le tableau.

- Comparer les données contenues dans le tableau à d'autres données établies en fonction du critère de différenciation du tableau et qui sont souvent présentes dans la publication où le tableau se trouve.
- Comparer la population du tableau soit à une population plus vaste dans laquelle elle s'insère, soit à une population différente.
- Comparer la population du tableau à une population analogue mesurée à un autre moment du temps.

Attention : les ressemblances sont tout aussi instructives que les différences.

- Confronter la distribution des données à un modèle théorique (courbe de Gauss, courbe en J, courbe en U ou pour la distribution d'une population, pyramide des âges).

Remarque :

C'est quand on part d'une question précise qu'il est le plus aisé d'extraire des informations. Par exemple, si l'on veut identifier la quantité d'électricité produite par les différentes sources existantes (nucléaire, éolien, hydraulique, centrales thermiques, autres sources éventuelles), on pourra plus facilement aller chercher l'information dans le tableau adéquat.

Application : Pour chacun des cas proposés, et en deans les limites qui ont été précisées, procédez à une première lecture critique des données présentées.

Accidents de la route 2008 :

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Fréquentation du cinéma dans le monde

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Dans le cas des données relatives à la fréquentation du cinéma, on est face à un « faux » tableau de fréquences à une entrée. Pourquoi ?

.....
.....
.....
.....
.....
.....

Pourrait-on établir, au départ des données présentées dans le document, que :

- en 2000, 60% des entrées de cinéma dans le monde ont été vendues aux Etats-Unis ?

.....
.....
.....
.....
.....
.....

- en 2000, 18% des entrées de cinéma dans l'Union européenne ont été vendues en Allemagne ?

.....
.....
.....
.....
.....
.....

- en 2000, on a vendu dix fois plus d'entrées de cinéma aux Etats-Unis qu'en Espagne ?

.....
.....
.....
.....
.....
.....

En conclusion, face à un « faux » tableau de fréquences à une entrée :

- peut-on transformer les fréquences absolues en fréquences relatives ?

.....

.....

.....

.....

.....

.....

- peut-on comparer les données entre elles ?

.....

.....

.....

.....

.....

.....

Ne pourrait-on pas aménager les données de façon à obtenir un « vrai » tableau de fréquences à une entrée ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

3.- Les tableaux de fréquences à double entrée

Il s'agit ici de tableaux qui montrent comment une population se répartit en tenant compte simultanément de deux critères de différenciation.

Un exemple :

Prénoms des filles nées en 2014 par région

	Belgique	Bruxelles	Flandre	Wallonie
Emma	653	64	381	208
Louise	584	40	344	200
Elise	480	19	351	110
Olivia	459	53	215	191
Lina	383	119	155	109
Marie	376	17	309	50
Lucie	359	37	78	244
Ella	348	10	329	9
<i>Etc.</i>				

Partons à présent du tableau suivant. Il est extrait d'une brochure réalisée en 2006 par des chercheurs d'un centre d'études de l'ULB (TEF, Travail Emploi Formation) pour le SPP Politique Scientifique Fédérale (service public de programmation politique scientifique), sur base du programme Atlas de valorisation des résultats de l'Enquête socio-économique 2001. Cette brochure a été publiée par le SPF Economie, PME, Classes moyennes et Energie et est disponible à l'adresse suivante :

http://statbel.fgov.be/fr/binaries/p009n018_fr%5B1%5D_tcm326-34058.pdf

On trouve à la page 29 de ce rapport le tableau suivant.

Tableau 2 : Niveau de diplôme le plus élevé obtenu selon le sexe (Belgique, 2001)

	Jamais suivi d'enseignement	Primaire	Sec. inférieur	Sec. supérieur	Supérieur (type court, type long, université)	Inconnu	Effectifs
Hommes	3.7%	14.1%	23.3%	30.0%	23.1%	5.8%	3.797.922
Femmes	4.3%	17.4%	21.9%	27.1%	22.6%	6.6%	4.088.287
TOTAL H+F	4.0%	15.8%	22.6%	28.5%	22.8%	6.2%	7.886.209

Source : INS-ESE 2001. Calculs : Centre TEF, ULB

Répondez à présent aux questions suivantes.

Quels sont les critères de différenciation présents dans ce document ?

Premier critère : le sexe.

Nombre de modalités : deux (homme-femme)

Deuxième critère :

Nombre de modalités :

On peut considérer ici qu'on est face à deux tableaux distincts : lesquels ?

.....

.....

.....

.....

.....

.....

Exprimez par une phrase complète ce que signifie la mention « 23,3% » (case grisée)

.....

.....

.....

.....

.....

Exprimez par une phrase complète ce que signifie la mention « 22,8% » (case grisée)

.....
.....
.....
.....

Y aurait-il une raison à la disposition des données dans le tableau ? Partons ici de l'idée que les modalités distinguées sur base d'un critère de différenciation sont disposées en lignes parce que le tableau privilégie ce critère et fournit en ce qui le concerne les données les plus intéressantes. Le critère de différenciation repris en colonnes devient alors accessoire. Les auteurs de la recherche privilégieraient-ils une perspective bien précise ?

.....
.....
.....
.....
.....

Remarque : la disposition des données est parfois liée à des raisons très pratiques de mise en page. La plupart des tableaux sont plus hauts que larges et les colonnes sont donc réservées au critère de différenciation offrant le moins de modalités.

Exprimez par une phrase complète ce que signifie la mention « 7.886.209 » (coin inférieur droit) ?

.....
.....
.....
.....
.....

Dans un tableau de fréquences à double entrée, des totaux partiels y mentionnent souvent au bout de chaque ligne le nombre d'éléments dénombrés sur cette ligne et au bas de chaque colonne, le nombre d'éléments dénombrés dans cette colonne. Est-ce le cas ici ?

.....
.....
.....
.....

Si un tableau à double entrée peut indiquer uniquement des fréquences absolues ou des fréquences relatives, il peut indiquer conjointement des fréquences absolues et des fréquences relatives.

Est-il possible au départ des données contenues dans le tableau de retrouver les effectifs absolus pour chacune des modalités distinguées ? Par exemple, combien de femmes âgées de plus de 65 ans étaient-elles diplômées de l'enseignement supérieur en 2001 ?

.....

.....

.....

.....

.....

.....

.....

.....

Au départ de cette opération mathématique, on peut souligner certaines limites ou difficultés liées à la présentation des données en termes de fréquences relatives.

- Les fréquences relatives sont habituellement exprimées en pourcentages et ces pourcentages sont souvent assortis d'une ou deux décimales dont l'existence ne se justifie vraiment que si le volume de la population est suffisant.
- Quand des fréquences relatives sont introduites dans un tableau à double entrée, elles y sont placées en regard des fréquences absolues qu'elles expriment dans une colonne adjacente ou plus rarement, sur une ligne immédiatement inférieure. Un tableau peut alors gravement s'alourdir. On retrouve alors parfois les données relatives dans un tableau apparemment distinct du tableau contenant les fréquences absolues.
- De toute façon, les fréquences relatives ne prennent vraiment d'intérêt que s'il est possible de déceler les fréquences absolues qu'elles traduisent.

4.- Quels enseignements tirer d'un tableau à double entrée ?

Premièrement, prendre une vue globale des données qui y figurent.

Examiner comment tous les éléments de la population envisagée se répartissent d'après un des critères de différenciation qui y sont mis en œuvre et/ou selon les différents critères de différenciation présents. Ces données figurent soit en fin de ligne, soit en fin de colonne. Ces données seront plus éclairantes si elles sont exprimées en fréquences absolues et en fréquences relatives.

Deuxièmement, comparer des sous-populations distinguées au sein de la population totale que le tableau recouvre.

Ces sous-populations peuvent être constituées à partir de l'un et l'autre des critères de différenciation que le tableau met en œuvre. Le plus souvent, on prêtera surtout attention à l'un des deux critères.

Notons que la comparaison est facilitée quand elle porte sur des séries de fréquences relatives (que le tableau contient ou qu'il faut calculer). La comparaison s'enrichit encore quand elle porte en outre sur des expressions statistiques résumant des séries de fréquences, particulièrement lorsque ces séries sont établies en fonction des valeurs d'une variable (au sens strict du terme), comme par exemple : calcul de la moyenne, de la médiane,...

Attention !

Il ne faut pas confondre des différences d'intensité et des différences de fréquences. Ainsi, l'âge moyen d'une sous-population n'est pas nécessairement plus élevé parce que cette sous-population comprend proportionnellement plus d'individus ayant dépassé un âge déterminé. On pourra nuancer ces constats en recourant à différentes sortes d'expressions statistiques (voir à ce propos le cours de Statistique descriptive).

Pour enrichir le travail de comparaison

- La comparaison restera très frustrée si on se limite à constater l'existence de différences. Il faut encore préciser en quoi ces différences se manifestent, quelle est leur importance et si ces différences sont statistiquement significatives compte tenu du volume des sous-populations.
- Des ressemblances peuvent être tout aussi remarquables et parfois plus remarquables que des différences.
- Enfin, il est tout aussi instructif d'utiliser successivement l'un des deux critères puis l'autre pour composer les sous-populations. Par exemple : examiner la répartition par niveau de formation des populations masculines et féminines d'un pays ; comparer la répartition par sexe des différents niveaux de formation.
- On pourra également représenter graphiquement la répartition des données (bâtonnets ; fromages ; etc.).

Troisièmement, recourir aux procédés de calcul statistique qui permettent de mieux se prononcer sur la signification et la force de l'association entre les deux critères : à quel point la répartition des éléments d'une population sur base d'un de ces critères est-elle associée à leur répartition sur base de l'autre critère ? C'est par exemple la finalité du Calcul du Chi².

Attention !

Il s'agit également de dépasser le stade de la corrélation statistique, pour aller vers une explication de cette association entre deux variables. Et donc de ne pas trop vite inférer qu'il y a un lien de cau-

salité entre les deux variables : il ne s'agit pas de soutenir qu'un phénomène reflété par un critère de différenciation (comme le sexe) serait à la cause d'un phénomène reflété par l'autre critère (comme la mortalité).

Il ne faut pas perdre de vue qu'il peut y avoir d'autres critères de différenciation qui peuvent entrer en ligne de compte.

5.- Les tableaux de fréquences à plus de deux entrées

Par exemple : répartition des habitants d'un pays selon le sexe, l'âge, la nationalité et la province où ils habitent. Ces tableaux s'avèrent extrêmement lourds et complexes.

Comment extraire une information de ces tableaux ?

En procédant comme détaillé ci-dessus : reconstituer différents tableaux de fréquences à double entrée et étudier chacun d'eux séparément.

Tableau 3 : Niveau de diplôme le plus élevé obtenu selon l'âge et le sexe (Belgique, 2001)

	Jamais suivi d'enseignement	Primaire	Sec. inférieur	Sec. supérieur	Supérieur (type court, type long, université)	Inconnu	Effectifs
Hommes							
18-24 ans	0.9%	2.7%	19.9%	58.1%	16.1%	2.3%	428.735
25-34 ans	1.9%	3.1%	17.6%	40.9%	33.9%	2.6%	676.379
35-64 ans	3.9%	13.2%	26.7%	26.2%	25.0%	4.9%	1.990.203
65 ans et +	6.6%	33.8%	21.2%	13.2%	11.5%	13.6%	702.605
Total	3.7%	14.1%	23.3%	30.0%	23.1%	5.8%	3.797.922
Femmes							
18-24 ans	0.8%	1.8%	12.9%	58.1%	24.6%	1.8%	419.888
25-34 ans	1.9%	2.7%	13.8%	36.9%	42.6%	2.1%	667.831
35-64 ans	4.4%	14.6%	26.6%	25.8%	23.8%	4.8%	1.997.712
65 ans et +	7.0%	39.2%	21.9%	10.4%	6.1%	15.4%	1.002.856
Total	4.3%	17.4%	21.9%	27.1%	22.6%	6.6%	4.088.287
TOTAL H+F	4.0%	15.8%	22.6%	28.5%	22.8%	6.2%	7.886.209

Source : INS-ESE 2001. Calculs : Centre TEF, ULB

6.- Les tableaux présentant des expressions statistiques qui ne sont plus des fréquences

Les indices

Des valeurs numériques fort similaires sont parfois comparées en calculant ce que l'écart séparant deux de ces valeurs représente en proportion de l'une de ces valeurs.

Partons d'un exemple tiré du rapport d'activités 2016 de la Fédération des Brasseurs Belges.

Ce rapport est disponible à l'adresse suivante :

(<https://publish.folders.eu/fixed/1058842?token=0d4d6c961c12740ef91be39450f534a8&pageMode=single>)

On y trouve notamment des données sur l'évolution de la consommation de bière en Belgique (calculée en hectolitres), entre 2008 et 2016, selon le circuit de distribution de ce produit (« FOOD », c'est-à-dire les commerces, et « HORECA », c'est-à-dire hôtellerie, restaurants et cafés).

Années	FOOD	HORECA
2008	4.404.301	4.389.934
2009	4.451.417	4.228.181
2010	4.394.240	4.042.911
2011	4.593.411	3.980.505
2012	4.444.750	3.783.088
2013	4.454.953	3.637.617
2014	4.459.451	3.625.801
2015	4.432.187	3.518.779
2016	4.326.573	3.362.575

Le calcul d'indices nous permettra de mieux représenter l'évolution de la consommation de bière pour chacun des deux circuits de distribution et de comparer l'évolution respective de cette consommation.

Comment élabore-t-on un indice ?

Premièrement, en choisissant une année de référence et en lui attribuant une valeur de base de 100.

Deuxièmement, en divisant le chiffre de la consommation correspondant à une autre année par le chiffre de l'année de référence, puis en multipliant le résultat de cette division par 100.

Si l'on constate que ce résultat est supérieur à 100, c'est donc qu'entre l'année retenue et l'année de référence, le chiffre de la consommation a augmenté.

Si l'on constate que ce résultat est inférieur à 100, c'est donc qu'entre l'année retenue et l'année de référence, le chiffre de la consommation a diminué.

Exemple :

Années	FOOD	INDICE	CALCUL
2008	4.404.301	100	
2009	4.451.417	101,06	$(4.451.417/4.404.301) \times 100$
2010	4.394.240	99,77	
2011	4.593.411		
2012	4.444.750		
2013	4.454.953		
2014	4.459.451		
2015	4.432.187		
2016	4.326.573	98,23	

Années	HORECA	INDICE	CALCUL
2008	4.389.934	100	
2009	4.228.181	96,31	$(4.228.181/4.389.934) \times 100$
2010	4.042.911	92,09	
2011	3.980.505		
2012	3.783.088		
2013	3.637.617		
2014	3.625.801		
2015	3.518.779		
2016	3.362.575	76,60	

En utilisant les deux indices, on peut mieux appréhender l'évolution respective de la consommation de bière selon les deux circuits de distribution du produit.

Ces opérations ne sont pas toujours simples, et parfois même très complexes, comme quand on utilise des indices synthétiques élaborés en combinant les résultats de plusieurs rapports entre des valeurs d'indice et les valeurs qui leur servent de base. C'est le cas par exemple de l'indice des prix à la consommation.

Les taux

Les indices mettent en relation des valeurs numériques fort similaires. D'autres expressions, comme les taux ou les ratios, mettent en relation des valeurs numériques plus disparates.

Exemples :

- taux général de natalité ou taux général de mortalité : calculé en divisant le nombre de naissances ou de décès survenus pendant un an dans le pays ou dans une de ses divisions administratives, puis en multipliant le quotient par 1.000
- taux de chômage : calculé en divisant le nombre de chômeurs par le chiffre de la population active (personnes occupées + chômeurs) âgée de 15 à 64 ans

- densité d'une population : est aussi une espèce de taux, puisque le nombre d'habitants d'un pays ou d'une division administrative de ce pays, est divisé par la superficie (en km² ou en ha) du pays ou de la division administrative.

Ces valeurs générales sont discutables.

Un taux général de natalité et un taux général de mortalité sont affectés par la structure d'âge de la population intervenant en dénominateur.

C'est pourquoi un taux de fécondité et un taux de mortalité par âge sont plus éclairants.

Un taux de criminalité établi au niveau d'une division administrative du pays en fonction de nombre d'habitants de la division administrative est contestable du fait que les condamnés dans cette division administrative n'habitent pas nécessairement dans cette division administrative et que des habitants de la division administrative sont en revanche condamnés ailleurs.

La définition d'un taux de chômage varie suivant la définition du terme chômage qui y est mise en œuvre et suivant que la population de référence comprend la totalité de la population active ou exclut certaines de ses composantes.

Considérons ici un dernier exemple : le taux de réussite à l'examen théorique pour le permis de conduire catégorie B, selon l'âge et par région.

En 2015, en Belgique, on a répertorié 296.696 examens théoriques pour le permis de conduire catégorie B.

Ces données sont tirées du rapport d'activités du GOCA (Groupement des entreprises agréées de contrôle automobile et du permis de conduire ASBL). Ce rapport est disponible à l'adresse suivante :

http://www.goca.be/upload/rapports_annuels/GOCA_RapportAnnuel2015_FR.pdf

On peut y retrouver ce qui suit :

	< 20	20-30	>30
2015 Bruxelles	52,67%	47,76%	35,53%
2014 Bruxelles	49,75%	46,05%	33,82%
2015 Flandre	68,82%	57,98%	41,54%
2014 Flandre	63,24%	52,61%	36,63%
2015 Wallonie	56,66%	50,82%	36,90%
2014 Wallonie	55,02%	48,49%	35,88%
2015 Belgique	63,48%	53,21%	38,41%

Il s'agit bien de taux. Et pour vous assurer que les choses ont été bien comprises, formulez par une phrase complète ce que signifie le chiffre de « 68,82% ».

.....

Conclusions

Dans la lecture d'un tableau statistique, Paul Minon conseillait de procéder avec rigueur !

1. Déterminer clairement l'objet du tableau.
2. Déterminer clairement la source des données qu'il contient.
3. Discerner les enseignements globaux que le tableau livre avant de s'attacher à l'un ou l'autre détail.

Bien avant les calculs statistiques sophistiqués, il rappelait l'importance des calculs les plus simples : ils restent les plus indispensables et leurs résultats nous apprennent déjà beaucoup. Si l'on procède à des calculs et à des représentations graphiques, il faut réfléchir constamment afin de décider lesquelles de ces opérations sont les plus pertinentes. Enfin, pour extraire une information pertinente d'un tableau, il convient de ne pas oublier de comparer les données qu'il contient et les données contenues par ailleurs.

QUESTION D'EXAMEN – 2^e SESSION 2016-2017

Le document qui suit est extrait d'une publication de la Fédération Wallonie-Bruxelles, datant de 2016 et intitulée « La Fédération Wallonie-Bruxelles en chiffres ». A la page 243 de cette publication, dans le chapitre consacré au sport, il est notamment question de la fréquentation des activités sportives subsidiées par les Centres de Conseil du Sport. On répertorie 7 Centres du Conseil du Sport en région de langue française et en région bilingue de Bruxelles-Capitale : Bruxelles ; Mont-Saint-Guibert (province du Brabant wallon) ; Liège (province de Liège), Mons et Charleroi (province du Hainaut) ; Namur (province de Namur) et Arlon (province de Luxembourg).

Les Centres de Conseil du Sport (CCS) sont au service des particuliers, des communes, des provinces, des écoles et des associations. Ils informent et offrent un accompagnement sportif, administratif et de gestion en matière de subventions, de prêt de matériel, de prix et de trophées et d'organisation d'activités sportives. Ils travaillent également avec les clubs sportifs affiliés à une fédération sportive, les centres sportifs locaux (et locaux intégrés) et les fédérations sportives reconnues.

Note(s) :

- « **Eté sport** » : organisation d'un stage sportif pendant la période estivale pour encourager les nouvelles initiatives de découverte et/ou des pratiques sportives et en permettre l'accès au plus grand nombre (sans limite d'âge).
- « **Eté jeunesse** » : partenariat entre un club affilié à une fédération sportive et une association au service des jeunes.
Objectif : encourager l'initiation sportive de jeunes de 5 à 17 ans et favoriser l'insertion sociale et culturelle en partageant des expériences positives par le biais de la découverte de la pratique sportive et de ses valeurs intrinsèques.
- **Cross** : développement des capacités d'endurance chez l'enfant du degré moyen ou supérieur de l'enseignement fondamental (classes de la 3^e à la 6^e primaire tous réseaux confondus).
- « **Mon club, mon école** » : partenariat entre un club affilié à une fédération sportive et un établissement scolaire (tous réseaux confondus). Objectif : sensibilisation à la pratique sportive pour les élèves de 8 à 14 ans et initiation à l'arbitrage pour ceux de 15 à 18 ans, dans le but d'ouvrir les infrastructures scolaires en dehors des heures de cours d'éducation physique et de prolonger la découverte sportive en pratique régulière en s'affiliant à un club.

Evolution de la fréquentation des activités subsidiées ou organisées par les Centres de Conseil du sport de 2012 à 2014

Nombre de participants			
Activités	2012	2013	2014
Eté sport	50.202	57.420	51.646
Eté jeunesse	10.590	8.460	9.850
Cross	17.452	27.142	<u>28.402</u>
Mon club, mon école	4.505	4.649	5.083

Répondez à présent aux questions suivantes.

- Quel est l'objet du document ? **6 pts**
- Ce document contient-il un ou plusieurs tableaux ? Précisez le type du ou des tableau(x) que vous avez identifié(s). **4 pts**
- Que signifie le chiffre de 28.402 ? **2 pts**
- Combien de jeunes belges francophones ont participé à un stage sportif en 2013 ? **2 pts**
- En 2014, on peut établir que 94.981 personnes âgées de plus de 5 ans ont pris part à une activité subsidiée ou organisée par les Centres de Conseil du Sport ?
Vrai ou faux ? Argumentez. **2 pts**
- Entre 2012 et 2014, le nombre de cross organisés dans les différentes provinces francophones a augmenté de 63%.
Vrai ou faux ? Argumentez. **2 pts**
- Plus de la moitié des subsides distribués en 2013 par les Centres de Conseil du Sport ont été affectés à l'opération « Eté sport ».
Vrai ou faux ? Argumentez. **2 pts**