

# INTRODUCTION

Vous avez décidé d'être partie prenante d'une étude, la présente formation est destinée à vous donner les informations nécessaires et susceptibles de vous aider à la réaliser.

Chaque étude est placée sous la responsabilité d'un membre d'une équipe de travail qui est alors un "chef de projet".

# Réflexions préalables à l'étude

- FORMULATION DU PROBLEME
- ETUDES ET ANALYSES PRELIMINAIRES
- PROJET D'ETUDE
- LE CHEF DE PROJET
- DEROULEMENT DE L'ETUDE

# FORMULATION DU PROBLEME

- Analyse du problème à résoudre
- Définition des objectifs de l'étude
- Formulation des questions d'études

# Analyse du problème à résoudre

- Lancement d'un nouveau produit :  
chances de réussite ;
- Quel prix retenir pour le lancement ?
- Etude de notoriété ;
- Etude de satisfaction ;
- Etude des attitudes et de comportement ;
- étude de l'efficacité d'une campagne publicitaire ;

# Définition des objectifs de l'étude

- Caractéristiques du produit ;
- Prix ;
- Concurrence ;
- Critères de satisfaction ;
- Fréquences d'achat ;
- Comportement d'achat ;

# Formulation des questions d'études

- Mesurer les intentions d'achats ;
- Estimer les volumes des ventes ;
- Comparer les prix des concurrents ;
- Mesurer le degré de satisfaction ;
- Mesurer la régularité des actes d'achat ;

# **ETUDES ET ANALYSES PRELIMINAIRES**

**Identifier les informations nécessaires**

**Rechercher les informations  
disponibles**

# Identifier les informations nécessaires

- Environnement micro économique
- Environnement macro économique

# Rechercher les informations disponibles

- Etude documentaire

  - sources internes / externes

  - Sources publiques

  - Sources professionnelles et associatives

  - Sources commerciales privées

- Rencontres et discussions

  - Entreprises

  - Distributeurs

  - clients

# PROJET D'ETUDE

- Objectifs de l'étude ;
- Choix méthodologiques ;
- Délai de réalisation de l'étude ;
- Budget de l'étude ;
- Membres de l'équipe d'étude.

# CHEF DU PROJET

- \* Le chef de projet élabore l'avant projet, soit seul soit avec le commanditaire de l'étude.
- \* Il négocie le budget de l'étude avec le commanditaire.
- \* Il vérifie la validation des échantillons et autres paramètres envisagés par l'entreprise.
- \* Il vérifie le respect des délais.
- \* Il vérifie aussi l'adéquation entre le projet et la réalisation.
- \* Il est responsable des problèmes qui peuvent survenir vis à vis du commanditaire.

# déroulement de l'étude

# Briefing avant l'étude

**Le chef de projet a normalement déjà pris connaissance du contenu de l'étude au travers de l'avant-projet, il réunit l'équipe de l'étude pour lui redonner l'ensemble des tenants et aboutissants de cette étude: but, délais, délimitation géographique, phases de l'étude, établissement des questionnaires(où, comment, exigences particulières,...)**

# Etude proprement dite

- Echantillonnage
- Questionnaire à établir, tester puis poser à l'échantillon visé
- Saisie des données
- Analyse des données sur ordinateur

# Rapport

- clair ;
- circonstancié ;
- soigné ;
- correspondre aux exigences du commanditaire ;
- détaillé ;
- complet ;
- ordonné. (l'analyse doit être séparée des graphiques et du tri statistique)

# Soutenance orale devant le commanditaire

- claire ;
- détaillée ;
- il faut montrer que l'on possède le sujet ;
- il faut être préparé à répondre aux questions possibles du commanditaire ;
- il faut se servir des outils audio-visuels.

# ECHANTILLONNAGE

# Enquête

Ensemble des opérations de collecte et de traitement de données relatives à quelques domaines que ce soit.

# Population cible

Rassemblement de tous les cas qui répondent à un ensemble de caractères spécifiques. Appelée aussi univers ou ensemble statistique, c'est l'ensemble des éléments auxquels on s'intéresse.

# Sondage

Enquête incomplète, enquête partielle ou enquête par échantillonnage, c'est une enquête au cours de laquelle seulement une partie des unités de base de la population sont observée.

# Echantillon

Ensemble des unités de base sélectionnées et réellement observées au cours d'un sondage.

# Echantillonnage

Ensemble des opérations qui permettent de sélectionner de façon organisée les éléments de l'échantillon.

# Base de sondage

Enumération ou présentation ordonnée de toutes les unités de base constituant la population.

# Erreur d'échantillonnage

Ecart entre les résultats obtenus auprès d'un échantillon et ce que nous apprendrait un recensement comparable de la population. Plus la taille de l'échantillon est grande plus l'erreur d'échantillonnage diminue.

# Détermination de la taille de l'échantillon

**la taille d'échantillon dépend essentiellement de deux facteurs :**

**La précision souhaitée : plus on souhaite des résultats précis, plus l'échantillon nécessaire est important.**

**Le budget disponible : plus on augmente la taille, plus le coût de l'enquête s'accroît.**

# Cas d'une population de loi inconnue

Taille d'échantillon pour estimer une moyenne

$$n = \frac{\sigma^2}{\varepsilon^2 \times \alpha}$$

**Exemple :** Un parc de loisirs souhaite estimer à 10dh près le montant moyen d'achats effectués par chaque visiteur. une étude pilote menée sur 50 visiteurs choisis au hasard a montré que l'écart- type des achats est :  $\sigma = 100$  dh

Si on se fixe un seuil de confiance  $(1-\alpha) = 95\%$ , La taille de l'échantillon est donc :

$$n = \frac{100^2}{10^2 \times 0,05} = 2000$$

# Cas d'une population de loi inconnue

## Taille d'échantillon pour estimer une proportion

$$n = \frac{pq}{\varepsilon^2 \times \alpha}$$

Si on se fixe un seuil de confiance  $(1-\alpha) = 95\%$ , La taille de l'échantillon est donc :

$$n = \frac{0,65 \times 0,35}{0,05^2 \times 0,05} = 1820$$

# Cas d'une population normale

## Taille d'échantillon pour estimer une moyenne

Cas des prélèvements dans une population finie avec remise ou dans une population infinie sans remise

$$n = Z^2_{1-\frac{\alpha}{2}} \frac{\sigma^2}{\varepsilon^2}$$

**Exemple : Un parc de loisirs souhaite estimer à 10dh près le montant moyen d'achats effectués par chaque visiteur. une étude pilote menée sur 50 visiteurs choisis au hasard a montré que l'écart- type des achats est :  $\sigma = 100$  dh**

Si on se fixe un seuil de confiance  $(1-\alpha) = 95\%$ , La taille de l'échantillon est donc :

$$n = 1,96^2 \frac{100^2}{10^2} = 384,16 = 385$$

# Cas d'une population normale

## Taille d'échantillon pour estimer une proportion

Cas des prélèvements dans une population finie avec remise ou dans une population infinie sans remise

$$n = Z^2_{1-\frac{\alpha}{2}} \frac{pq}{\varepsilon^2}$$

**Exemple :** Le parc souhaite estimer la proportion des visiteurs qui font des achats à cinq points prés.

L'enquête pilote a estimé cette proportion à 65%, c'est à dire  $p = 0,65$

Si on se fixe un seuil de confiance  $(1-\alpha) = 95\%$ , La taille de l'échantillon est donc :

$$n = 1,96^2 \frac{0,65 \times 0,35}{0,05^2} = 349,58 = 350$$

# Cas d'une population normale

## Taille d'échantillon pour estimer une moyenne

Cas des prélèvements dans une population finie sans remise

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} \sigma^2 N}{\varepsilon^2 N + Z^2_{1-\frac{\alpha}{2}} \sigma^2}$$

# Cas d'une population normale

## Taille d'échantillon pour estimer une proportion

Cas des prélèvements dans une population finie sans remise.

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} p q N}{\varepsilon^2 N + Z^2_{1-\frac{\alpha}{2}} p q}$$

# Méthodes d'échantillonnage

- Méthodes d'échantillonnage probabilistes
- Méthodes d'échantillonnage empiriques

# Méthodes d'échantillonnage probabilistes

- Echantillonnage aléatoire et simple
- Echantillonnage stratifié
- Echantillonnage par degrés
- Echantillonnage systématique

# Echantillonnage aléatoire et simple

Un échantillonnage est aléatoire si tous les individus de la population ont la même chance de faire partie de l'échantillon; il est simple si les prélèvements des individus sont réalisés indépendamment les uns des autres.

# Echantillonnage aléatoire et simple

Si la population est finie, l'échantillonnage aléatoire et simple correspond au tirage aléatoire avec remise, qui permet de traiter les populations finies comme des populations infinies.

# Echantillonnage aléatoire et simple

- Constituer la base de sondage qui correspond à la liste complète et sans répétition des éléments de la population ;
- Numéroté ces éléments de 1 à N ;
- Procéder, à l'aide d'une table de nombres aléatoires ou d'un générateur de nombres pseudo aléatoires à la sélection des unités différentes qui constitueront l'échantillon.

# Echantillonnage aléatoire et simple

## Exemple :

On souhaite avoir un échantillon aléatoire et simple de 5 entreprises parmi une population de 22 entreprises. On dispose de la base de sondage c'est à dire la liste complète et sans répétition des 22 entreprises numérotées de 1 à 22. On prend un extrait d'une table de nombre aléatoire par exemple :

# Echantillonnage aléatoire et simple

On choisit au hasard un nombre de la table, supposons ce nombre 06121. Comme  $N=22$ , on va retenir le premier groupe de 2 chiffres, ce qui donne les  $N^\circ$  : 06, ensuite 12 ; 19 ; 17 ; les nombres (82,77 et 92) sont inutilisables. La cinquième entreprise sera le  $N^\circ 10$ .

10480 15011 01536 02011 81647 91646  
22368 46573 25595 85393 30995 89198  
24130 48390 22527 97265 76393 64809  
42167 93093 06243 61680 07856 16376  
37570 39975 81837 16656 **06121** 91782  
77921 06907 11008 42751 27756 53498

# Echantillonnage stratifié

L'échantillonnage stratifié est une technique qui consiste à subdiviser une population hétérogène, d'effectif  $N$ , en  $P$  sous populations ou « strates » plus homogènes d'effectif  $N_i$  de telle sorte que  $N = N_1 + N_2 + \dots + N_p$ . Un échantillon, d'effectif  $n_i$ , est par la suite, prélevé indépendamment au sein de chacune des strates en appliquant un plan d'échantillonnage au choix de l'utilisateur. Le plus souvent, on procédera par un échantillonnage aléatoire et simple à l'intérieur de chaque strate.

# Echantillonnage stratifié

La stratification peut entraîner des gains de précision appréciables, elle facilite en outre les opérations de collecte des données et fournit des informations pour différentes parties de la population.

# Echantillonnage stratifié

Pour la répartition de l'effectif total,  $n$ , de l'échantillon dans les différentes strates, La première solution, dite proportionnelle, consiste à conserver la même fraction d'échantillonnage dans chaque strate. Une seconde solution, dite optimale, tient compte du budget de l'enquête.

# Répartition proportionnelle

La répartition proportionnelle consiste à répartir la taille de l'échantillon  $n$  en utilisant la même fraction de sondage  $f$  dans chacune des strates. Cette solution tient compte d'un seul facteur qui est le poids de chaque strate.

# Répartition proportionnelle

Désignons par  $w_i$  le poids de la strate et par  $f$  la fraction de sondage constant :

$$f = \frac{n}{N} \quad \text{et} \quad w_i = \frac{N_i}{N}$$

le nombre d'unités à choisir dans chacune des strates est donc :

$$n_i = w_i \times n = f \times N_i$$

# Répartition proportionnelle

## Exemple :

Dans une population de 10000 entreprises, réparties en 5000 petites entreprises, 3000 moyennes entreprises et 2000 grandes entreprises, on souhaite avoir un échantillon de 500 entreprises.

Fraction de sondage constante :

$$f = 500 / 10000 = 5 \%$$

# Répartition proportionnelle

Strate	Effectif de la strate	Taille de l'échantillon
Petite	5000	$5000 * 0,05 = 250$
Moyenne	3000	$3000 * 0,05 = 150$
Grande	2000	$2000 * 0,05 = 100$
Total	10000	500

# Répartition proportionnelle

Strate	Effectif de la strate	Poids de la strate	Taille de l'échantillon
Petite	5000	50%	$500 * 0,5 = 250$
Moyenne	3000	30%	$500 * 0,3 = 150$
Grande	2000	20%	$500 * 0,2 = 100$
Total	10000		500

# Répartition optimale

- Budget total de l'enquête,  $G$
- Poids de la strate,  $w_i$
- Coût de la collecte de l'information dans la strate,  $c_i$
- Dispersion à l'intérieur de la strate, mesurée par l'écart type  $\sigma_i$ .

# Répartition optimale

le nombre d'unités à choisir dans chacune des strates est donné par :

$$n_i = k \frac{w_i \sigma_i}{\sqrt{c_i}}$$

$$k = \frac{G}{\sum w_i \sigma_i \sqrt{c_i}}$$

# Répartition optimale

## Exemple :

Dans la population des 10000 entreprises, on a pu avoir les informations suivantes :

Strate	$w_i$	$c_i$	$\sigma_i$
Petite	0,5	50	0,8
Moyenne	0,3	75	1,5
Grande	0,2	100	2,2

# Répartition optimale

le nombre d'entreprises à choisir dans chacune des strates est donné par :

$$k = \frac{G}{\sum w_i \sigma_i \sqrt{c_i}} = \frac{5000}{0,5 \times 0,8 \times \sqrt{50} + 0,3 \times 1,5 \times \sqrt{75} + 0,2 \times 2,2 \times \sqrt{100}} = 449,42$$

$$n_1 = 449,42 \times \frac{0,5 \times 0,8}{\sqrt{50}} = 26 \text{ petites entreprises}$$

$$n_2 = 449,42 \times \frac{0,3 \times 1,5}{\sqrt{75}} = 24 \text{ moyennes entreprises}$$

$$n_3 = 449,42 \times \frac{0,2 \times 2,2}{\sqrt{100}} = 20 \text{ grandes entreprises}$$

# Echantillonnage par degrés

L'échantillonnage par degrés regroupe toute une série de plans d'échantillonnage caractérisés par un système ramifié et hiérarchisé d'unités.

# Echantillonnage par degrés

Dans le cas de deux degrés, par exemple, on considère que la population est constituée d'un certain nombre d'unités de sondage du premier degré (unités primaires), chacune de ces unités étant constituée d'un certain nombre d'unités du second degré. (unités secondaires)

On réalise d'abord un échantillonnage d'unités du premier degré. Ensuite, dans chaque unité sélectionnée au premier degré, on prélève un échantillon d'unités du second degré. Le mode de sélection pouvant varier d'un degré à l'autre.

# Echantillonnage par degrés

L'échantillonnage par degrés s'impose lorsqu'il est impossible d'inventorier les éléments de toute la population et qu'il est possible d'énumérer les unités prélevées au premier degré. Il permet une concentration du travail sur le terrain et donc une réduction des coûts.

Pour un même nombre total d'observations, il faut citer sa plus faible efficacité que l'échantillonnage aléatoire et simple.

# Echantillonnage par degrés

## Exemple :

Pour étudier le niveau de consommation des ménages d'une ville, on a tiré aléatoirement 5 quartiers. Dans chaque quartier sélectionné, on retient une rue sur 5, dans chaque rue retenue, on retient un immeuble sur 3, et dans chaque immeuble, un ménage par étage sera questionné.

# Echantillonnage systématique

L'échantillonnage systématique est une technique qui consiste à prélever des unités d'échantillonnage situées à intervalles égaux. Le choix du premier individu détermine la composition de tout l'échantillon.

# Echantillonnage systématique

Si on connaît l'effectif total de la population  $N$  et qu'on souhaite prélever un échantillon d'effectif  $n$ , l'intervalle entre deux unités successives à sélectionner est donné par :

$$k = \frac{N}{n}$$

# Echantillonnage systématique

Connaissant  $k$ , on choisit le plus souvent, pour débiter, un nombre aléatoire,  $i$ , compris entre 1 et  $k$ . le rang des unités sélectionnées est alors  $i, i+k, i+2k, \dots$

# Echantillonnage systématique

L'échantillonnage systématique est facile à préparer et, en général facile à exécuter, il réduit le temps consacré à la localisation des unités sélectionnées.

# Echantillonnage systématique

Si les éléments de la population se présentent dans un ordre aléatoire (pas de tendance) l'échantillonnage systématique est équivalent à l'échantillonnage aléatoire et simple. Par contre si les éléments de la population présentent une tendance, l'échantillonnage systématique est plus précis que l'échantillonnage aléatoire.

# Echantillonnage systématique

## Exemple :

On veut sélectionner un échantillon de 30 entreprises au sein d'une population de 1800 entreprises.

$$k = \frac{1800}{30} = 60$$

Ainsi on va tirer une entreprise toutes les 60 en partant d'un nombre tiré aléatoirement entre 1 et 60.

Supposons ce nombre est le 15. On va donc sélectionner la 15<sup>ème</sup> entreprise puis la 75<sup>ème</sup>, la 135<sup>ème</sup>. jusqu'à la 1755<sup>ème</sup> ce qui nous donnera l'échantillon de 30 entreprises.

# Méthodes d'échantillonnage empiriques

- **Echantillonnage accidentel (De convenance)**
- Echantillonnage à priori
- Echantillonnage « Boule de neige »
- **Echantillonnage par Quotas**

# Echantillonnage accidentel

Il s'agit d'un échantillon constitué d'individus qui se trouvaient accidentellement à l'endroit et au moment où l'information a été collectée.

Enquêtes réalisées dans la rue, les lieux publics, en sortie de super marché ...

Questionnaires figurant dans les magazines et renvoyés spontanément.

Les échantillons accidentels ne peuvent être considérés représentatifs d'aucune population. Il est risqué de généraliser à une population donnée des résultats obtenus par un échantillon accidentel.

# Echantillonnage à priori

C'est un échantillonnage par jugement à priori. Il consiste à sélectionner des individus dont on pense, avant de les interroger, qu'ils peuvent détenir l'information.

Le risque de ce type d'échantillonnage est de considérer des individus, apparemment représentatifs de la population étudiée.

# Echantillonnage « Boule de neige »

Cette méthode est réservée aux populations composées d'individus dont l'identification est difficile ou qui possèdent des caractéristiques rares.

La méthode consiste à faire construire l'échantillon par les individus eux-mêmes. Il suffit d'en identifier un petit nombre initial et de leur demander de faire appel à d'autres individus possédant les mêmes caractéristiques.

# Echantillonnage par Quotas

L'échantillonnage par quotas est l'échantillonnage non probabiliste le plus connu, et finalement le mieux accepté comme substitut aux méthodes probabilistes dans le cas où ces dernières rencontreraient des contraintes de base de sondage. Mais la représentativité de la population étudiée reste douteuse.

# Echantillonnage par Quotas

L'échantillonnage par quotas consiste à étudier la structure de la population selon des critères choisis (quotas) empiriquement. L'échantillon est ensuite construit de manière à constituer une reproduction en miniature de la population sur ces critères.

# Echantillonnage par Quotas

L'échantillonnage par quotas est une forme simplifiée de l'échantillonnage stratifié à fraction de sondage constante. Les quotas représentent les variables de stratification.

Une fois les quotas sont fixés, les individus sont sélectionnés à la convenance de l'enquêteur.

# Echantillonnage par Quotas

Les critères servant de base à la définition des quotas ne doivent pas être nombreux. Au-delà de 3 critères, la démarche devient complexe. Les quotas doivent être construits sur une base de données fiables ( statistiques disponibles ) indiquant la répartition de la population sur les critères choisis. Les critères les plus utilisés dans les études de marché sont économiques et socio-démographiques en particulier l'âge, le sexe, la catégorie socioprofessionnelle, ...

# Echantillonnage par Quotas :

## Exemple

On souhaite avoir un échantillon de 1000 individus. La structure de la population selon trois critères est la suivante :

Age	Structure de la population	Répartition de l'échantillon
20 à 29 ans	40 %	400
30 à 49 ans	35 %	350
50 à 60 ans	25 %	250
Total	100 %	1000

# Echantillonnage par Quotas :

## Exemple

### Structure de la population

Age	Sexe	Masculin	Féminin	Total
20 à 29 ans		48 %	52 %	100 %
30 à 49 ans		49 %	51 %	100 %
50 à 60 ans		45 %	55 %	100 %

# Echantillonnage par Quotas :

## Exemple

Répartition de l'échantillon

Age	Sexe	Masculin	Féminin	Total
20 à 29 ans		192	208	400
30 à 49 ans		172	178	350
50 à 60 ans		113	137	250

AGE	Sexe	Sans	Etudiant	Agric	Artisans	Prof libérales	Employés	Ouvriers	Total
	CSP								
20-29	M	10%	30%	5%	6%	9%	25%	15%	100%
	F	15%	25%	2%	10%	8%	30%	10%	100%
30-49	M	8%	5%	15%	22%	15%	15%	20%	100%
	F	20%	4%	10%	16%	14%	24%	12%	100%
50-60	M	6%	2%	25%	22%	18%	17%	10%	100%
	F	35%	1%	20%	20%	6%	13%	5%	100%

AGE	Sexe	Sans	Etudiant	Agric	Artisans	Prof libérales	Employés	Ouvriers	Total
	CSP								
20-29	M	19	58	10	12	17	48	28	192
	F	31	52	4	21	17	62	21	208
30-49	M	14	9	26	38	26	26	33	172
	F	36	7	18	28	25	43	21	178
50-60	M	7	2	28	25	20	19	12	113
	F	48	1	27	27	8	18	8	137

# Test de représentativité de l'échantillon

Pour valider les résultats de l'étude, et pour minimiser la marge d'erreur, il est indispensable de vérifier la représentativité de l'échantillon.

# Test de représentativité de l'échantillon

- Mise en tableau des résultats
- Test de représentativité

# Mise en tableau des résultats

Il faut dans un premier temps élaborer un tableau dans lequel seront notées les données théoriques concernant la population totale, ainsi que les données relatives à l'étude, ce tableau est appelé tableau de test de représentativité.

# Test de représentativité

Appelé test d'ajustement ou test Khi-deux, ce test est destiné à tester si l'échantillon observé peut être considéré comme représentatif de la population étudiée. Pour cela on a besoin de connaître les caractéristiques de la population. Le test d'ajustement consiste à calculer la valeur de la statistique khi-deux qui permet d'évaluer les écarts entre les fréquences observées dans l'échantillon et les fréquences théoriques correspondantes à la population étudiée. Lorsque les écarts seront petits, on acceptera la représentativité, sinon on la rejettera.

# Test de représentativité

- Formuler une hypothèse nulle
- Calculer la valeur de Khi-deux observé
- Déterminer la valeur de Khi-deux théorique
- Résultat du test de représentativité

# Formuler une hypothèse nulle

---

l'échantillon observé est représentatif de la population étudiée.

# Calculer la valeur de Khi-deux observé

$$\text{Khi deux} = \sum \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}}$$

# Déterminer la valeur de Khi-deux théorique

On se fixe un seuil de confiance  $\alpha$ , on se reporte à la table de distribution de Khi-deux, et on cherche la valeur correspondante à une probabilité égale à  $(1-\alpha)$  et un nombre de "degrés de liberté (ddl)" égale au nombre de caractéristiques de la population moins 1.

# Résultat du test de représentativité

la valeur de Khi-deux observé est nulle lorsque les fréquences observées sont toutes égales aux fréquences théoriques, c'est à dire lorsqu'il y a une concordance parfaite entre l'échantillon observé et la population étudiée. Cette valeur est d'autant plus grande que les écarts entre les fréquences observées et théoriques sont plus grands. Il découle de la comparaison du Khi-deux observé et du Khi-deux théorique :

# Résultat du test de représentativité

- Si Khi-deux théorique est strictement supérieur au Khi-deux observé, l'hypothèse nulle est acceptée. Donc **l'échantillon est représentatif de la population étudiée.**
- Si Khi-deux théorique est inférieur ou égal au Khi-deux observé, l'hypothèse nulle est rejetée. En d'autres termes, **l'échantillon n'est pas représentatif.** Il faut donc réaliser un redressement d'échantillon.

# Test de représentativité : exemple

Une enquête a été réalisée au près d'un échantillon de 400 individus prélevé au sein d'une population cible de 4 millions d'individus. Les données que l'on possède sur cette population sont les suivantes :

**Hommes 48% soit 1,92 millions d'hommes**

**Femmes 52% soit 2,08 millions de femmes**

# Test de représentativité : exemple

Sexe	Hommes		Femmes		Total
	%	Effectifs	%	Effectifs	
Aucun	35	672000	50	1040000	1712000
Primaire	30	576000	25	520000	1096000
Secondaire	15	288000	10	208000	496000
Formation professionnelle	13	249600	10	208000	457600
Supérieur	7	134400	5	104000	238400
Total	100	1920000	100	2080000	4000000

# Test de représentativité : exemple

## Caractéristiques de l'échantillon :

Sexe	Hommes		Femmes		Total
	%	Effectifs	%	Effectifs	
Aucun	32	61	54	112	173
Primaire	28	54	23	48	102
Secondaire	18	35	12	25	60
Formation professionnelle	14	27	8	17	44
Supérieur	8	15	3	6	21
Total	100	192	100	208	400

# Test de représentativité : exemple

## Mise en tableau des résultats

Sexe	Hommes		Femmes	
	Effectifs théoriques	Effectifs observés	Effectifs théoriques	Effectifs observés
Aucun	67,2	61	104	112
Primaire	57,6	54	52	48
Secondaire	28,8	35	20,8	25
Formation professionnelle	24,96	27	20,8	17
Supérieur	13,44	15	10,4	6
Total	192	192	208	208

# Test de représentativité : exemple

**Hypothèse nulle :**

l'échantillon observé est représentatif  
de la population étudiée.

# Test de représentativité : exemple

$$\text{Khi-deux observé} = [(61-67,2)^2/67,2 + (54-57,6)^2/57,6 + (35-28,8)^2/28,8 + (27-24,96)^2/24,96 + (15-13,44)^2/13,44 + (112-104)^2/104 + (48-52)^2/52 + (25-20,8)^2/20,8 + (17-20,8)^2/20,8 + (6-10,4)^2/10,4]$$

$$\text{Khi-deux observé} = 6,80$$

# Test de représentativité : exemple

**Khi-deux théorique** : si on se fixe un seuil de confiance  $\alpha=5\%$ , on se reporte à la table de distribution de Khi-deux, et on cherche la valeur correspondante à une probabilité égale à  $(1-\alpha)=0,95$  et un nombre de "degrés de liberté (ddl)" égale  $(10-1)=9$ , on trouve :

Khi-deux théorique = 16,92

# Test de représentativité : exemple

## Résultat du test de représentativité

Khi-deux théorique est strictement supérieur au Khi-deux observé, l'hypothèse nulle est acceptée. Donc **l'échantillon est représentatif de la population étudiée.**

# Redressement de l'échantillon

Si l'échantillon sélectionné n'est pas représentatif de la population étudiée, après le test, il est nécessaire de le redresser. On redressera l'échantillon si par exemple certaines caractéristiques de la population sont sur-représentées ou sous-représentées. On peut ainsi tirer de l'échantillon original un échantillon représentatif.

# Redressement de l'échantillon

Pour redresser l'échantillon, il existe deux méthodes distinctes :

- Redressement par extraction
- Redressement par substitution

# Redressement par extraction

La méthode dite par extraction élimine tous les individus au-delà du pourcentage à obtenir. L'effectif de l'échantillon est donc d'autant plus réduit que les pourcentages théoriques sont différents des pourcentages observés. Cette méthode permet d'aboutir à un échantillon plus représentatif mais au risque d'obtenir un échantillon ridiculement petit ou même à une impossibilité de traitement (le nombre minimum de questionnaires nécessaire pour traiter une étude est de 60. D'autre part, pour valider les résultats et pour effectuer des calculs statistiques, il faut au minimum 5 effectifs par classe.).

# Redressement par extraction :

## Exemple

1200 personnes interrogées pour une étude concernant l'attitude de la population face à un produit suivent la distribution suivante :

HOMMES: 55% soit 660 interrogés      FEMMES: 45%  
soit 540 interrogés

Or, la population totale a une distribution :

HOMMES:45% soit **540** pers. à interroger.      FEMMES:  
55% soit **660** personnes à interroger.

# Redressement par extraction : Exemple

Après redressement, on aura :

HOMMES: 660-120=540 individus.

FEMMES: 540 individus. (résultat inchangé)

D'où une meilleure répartition :

HOMMES: 50% soit 540 individus

FEMMES: 50% soit 540 individus

# Redressement par extraction

Cette méthode a le très gros inconvénient de faire fondre l'échantillon original. Dans l'exemple, nous avons perdu 120 individus sondés. La précision des estimations des résultats s'en trouve fatalement affectée : par ce procédé, tout se passe comme si on avait interrogé un plus petit nombre de personnes. Mais, on obtient une meilleure représentativité de la population.

La méthode par extraction est intéressante si la taille de l'échantillon est supérieure à 1000.

# Redressement par substitution

A l'opposé, la méthode par substitution remplace comme son nom l'indique, tout individu en surnombre dans une modalité sur représentée par un individu choisi au hasard dans une modalité sous-représentée. Ce dernier se trouve donc dupliqué. L'effectif de l'échantillon est maintenu, contrairement à la méthode par extraction ( on a vu que l'on était passé dans l'exemple d'un échantillon de 1200 individus à un échantillon de 1080), mais au prix de nombreuses duplications qui tendent à uniformiser l'échantillon. En fait, tout se passe comme si les individus des catégories sous-représentées avaient été interrogées plusieurs fois.

# Redressement par substitution

Les imperfections de l'échantillon sont éliminées sans risque apparent puisque la taille de l'échantillon ne se trouve pas affectée.

La méthode par substitution est valable si le différentiel de redressement est inférieur à 30%.

# Redressement par substitution : Exemple

La population totale compte :

HOMMES : 45%

FEMMES : 55%

L'échantillon de 2000 individus étudiés compte :

HOMMES : 740 individus (37%) ;

FEMMES : 1220 individus (61%) ;

NREP : 40 individus (2%)

# Redressement par substitution : Exemple

On pose le tableau suivant :

	OBTENUS		SOUHAITES		DIFFERENCE
	% Nbre	%	Nbre	S-O	
NREP 2		40	0	0	-2%
FEMMES		61	1220	55	1100 -6%
HOMMES		37	740	45	900 +8%

DIFFERENCIEL DE REDRESSEMENT=

$$2\%+6\%+8\%=16\%.$$

# Redressement par substitution :

## Exemple

L'échantillon redressé correspond ici au nombre souhaité :

<b>HOMMES : 45%</b>	<b>900 individus.</b>
<b>FEMMES : 55%</b>	<b>1100 individus.</b>

La répartition par sexe correspond alors à celle de la population totale.

On remarque ici que le différentiel de redressement est inférieur à 16%, ce qui est un bon résultat et qui ne remet pas en questions la validation des observations.

# Redressement de l'échantillon

- la méthode par extraction est intéressante si on sonde plus de 1000 personnes
- la méthode par substitution est valable si le différentiel de redressement est inférieur à 30%.
- Si l'échantillon n'est pas représentatif, et qu'on ne peut utiliser aucune des méthodes, il nous faut administrer de nouveaux questionnaires. C'est pourquoi il faut vérifier le plus tôt possible la représentativité de l'échantillon et tout de suite le redresser.

# ELABORATION DU QUESTIONNAIRE

Un questionnaire est un ensemble de questions construit dans le but de générer l'information nécessaire à l'accomplissement des objectifs d'une étude.

La caractéristique essentielle d'un questionnaire est d'être un instrument de standardisation de l'information collectée dans une enquête. En effet si chaque répondant voit ou entendra les mêmes questions, les informations recueillies, c'est à dire les réponses, seront donc comparables pour l'ensemble des répondants.

# ELABORATION DU QUESTIONNAIRE

La construction d'un questionnaire, dans son fond comme dans sa forme, doit inciter le répondant à coopérer, à se sentir impliqué et à donner des réponses complètes et exactes aux questions qui lui sont posées.

# Règles liées au fond du questionnaire

Le contenu du questionnaire dépend des objectifs de l'étude, et des hypothèses de travail que l'on aura formulées.

Le questionnaire sera construit à partir d'une liste des informations à rechercher. Cette démarche permet de n'oublier aucune information importante tout en évitant les questions inutiles.

# Les types de questions

- Les questions ouvertes
- Les questions fermées
- Echelles à catégories spécifiques

# Les questions ouvertes

Une question ouverte laisse l'individu interrogé totalement libre du choix de sa réponse dans sa forme et dans sa longueur.

On distingue 2 types de questions ouvertes :

# Les questions ouvertes

Questions à réponses numériques.

- Quel âge avez-vous ?
- Combien d'enfant à charge avez-vous ?
- Combien de Km par mois effectuez-vous avec votre voiture ?

Ces questions ne posent pas de problèmes.

# Les questions ouvertes

## Questions à réponse littéraire

- Quel est votre profession ?
- Quels marques d'ordinateurs connaissez-vous ?

Il faut s'efforcer de se limiter à des questions pour les quelles la réponse tient en un seul mot ou groupe de mots.

Si la réponse prend la forme d'un discours, la mise en forme de l'information en vue d'un traitement devient compliquée.

# Les questions fermées

Les questions fermées sont des questions pour les quelles on impose au répondant une forme précise de réponse et un nombre limité de choix de réponses.

On distingue trois types de questions fermées :

# Les questions fermées

## Questions dichotomiques

Ces questions offrent 2 choix de réponse, celle ci doit être unique, c'est l'un ou l'autre.

Sexe ?

1. Masculin

2. Féminin

Avez –vous un compte bancaire ?

1. oui

2. non

Ces questions se prêtent bien aux traitements statistiques.

# Les questions fermées

## Questions au choix multiples et réponse unique

Le répondant dispose d'une liste de choix mais sa réponse doit être unique. Les modalités de la liste doivent être exhaustives et mutuellement exclusives.

# Les questions fermées

Quelle est votre situation familiale ?

1. Célibataire
2. Marié
3. Divorcé
4. Veuf
5. Séparé

# Les questions fermées

## Questions au choix multiples et réponses multiples

Le répondant est autorisé à sélectionner plusieurs des possibilités proposées dans une liste.

Pour traiter ce type de questions, il est préférable d'envisager chaque modalité comme une question dichotomique avec réponse oui/non.

# Les questions fermées

Parmi les raisons suivantes indiquez celles qui vous guident dans le choix d'une destination touristique :

1. environnement naturel
2. dépaysement
3. culture
4. réputation culinaire
5. sites archéologiques
- 6 .autres (préciser)

# Echelles à catégories spécifiques

- Echelle de Likert
- Différentiel sémantique
- Echelles à supports sémantiques
- Echelles d'intention

# Echelle de Likert

Cette échelle a pour but de mesurer l'attitude d'un individu celui-ci attribut à une liste de propositions relatives à l'objet étudié une note indiquant son degré d'accord ou de désaccord, avec chacune des propositions énoncées.

# Echelle de Likert

Pour vous, protéger l'environnement ça consiste à :

	Pas d'accord		ni en accord ni en désaccord			tout à fait d'accord	
Utiliser des lessives sans Phosphates.	1	2	3	4	5	6	7

Traiter les Déchets Ménagers.	1	2	3	4	5	6	7
-------------------------------	---	---	---	---	---	---	---

Acheter des produits emballés.	1	2	3	4	5	6	7
--------------------------------	---	---	---	---	---	---	---

# Différentiel sémantique

Cette échelle consiste à demander au répondant de se situer quelque part entre deux objectifs de sens parfaitement opposés.

# Différentiel sémantique

Pour vous, l'administration X est :

	Extrêmement	Très	Assez	Sans opinion	Assez	Très	Extrêmement	
Agréable	1	2	3	4	5	6	7	Désagréable
Active	1	2	3	4	5	6	7	Inactive
Moderne	1	2	3	4	5	6	7	Dépassé
Proche du public	1	2	3	4	5	6	7	Eloignée du public

# Echelles à supports sémantiques

Le principe est de bâtir, pour une population donnée, une échelle dont les libellés sont à des distances, psychologiquement égales l'un de l'autre.

# Echelles à supports sémantiques

Comment jugez-vous le confort de ces différentes voitures ?

Marque	Extrêmement			Faible				moyen		Bon			Extrêmement	
	Mauvais												Bon	
BMW	0	10	20	30	40	50	60	70	80	90	100			
Ford	0	10	20	30	40	50	60	70	80	90	100			

# Echelles d'intention

Ces échelles ont pour but de recueillir une intention déclarée de comportement de la part du répondant.

# Echelles d'intention

Si Mc Donald ouvrait près de ton lieu de travail :

Je ne prendrais						je prendrais mon
Certainement pas	1	2	3	4	5	déjeuner à Mc Donald
Mon déjeuner						c'est sûr
A Mc Donald						

# Les niveaux de mesure

On distingue trois niveaux :

- **Echelles nominales (ou catégorielles)**
- **Echelles Ordinales**
- **Echelles continues : mesure métrique**

# Echelles nominales (ou catégorielles)

Quand une mesure est de nature nominale, les nombres jouent uniquement un rôle d'étiquettes permettant d'identifier des objets, des propriétés ou des événements. Les chiffres utilisées n'ont aucune relation entre eux : ni ordre, ni distance, ni, origine.

# Echelles nominales (ou catégorielles)

Quelle est votre situation familiale ?

1. Célibataire
2. Marié
3. Divorcé
4. Veuf
5. Séparé

# Echelles Ordinales

Le niveau de mesure ordinal identifie une relation d'ordre entre les propriétés des objets étudiés. Les chiffres utilisés ont une relation d'ordre mais aucune relation de distance, ils ne permettent pas de mesurer le degré de différence entre les objets.

# Echelles Ordinales

Classer les 4 types de produit suivants  
chaque caractéristique :

Produit caractéristique	Huile D'olive	Beurre	Huile de Tournesol	Margarine
Facilité de gestion	2	1	4	3
Goût	1	2	4	3
Facilité d'utilisation	3	4	1	2
Qualité nutritionnelle	1	3	2	4

# Echelles continues : mesure métrique

Ce niveau de mesure est le plus complet, en plus des relations d'appartenance et d'ordre, il conserve les distances et possède une origine.

# Echelles continues : mesure métrique

Indiquez la population de votre ville :

.....

# Règles liées à la forme du questionnaire

## Rédaction du questionnaire

Le vocabulaire et le style utilisés pour la rédaction d'un questionnaire doivent être ceux employés par les répondants et non ceux du rédacteur.

# Règles liées à la forme du questionnaire

## Rédaction du questionnaire

Tout vocabulaire technique est à éviter, sauf si l'étude est menée auprès de spécialistes.

Tout vocabulaire ambigu est à éviter, c'est à dire tout mot ou expression dont la compréhension prête à confusion soit parce qu'il possède plusieurs sens, soit parce qu'il est subjectif.

L'ambiguïté peut également provenir d'utilisations différentes d'un même mot selon la région, la classe sociale ou la classe d'âge.

Le vocabulaire employé doit être le plus simple le moins ambigu et adapté à la population concernée par l'étude.

# Règles liées à la forme du questionnaire

## Style de rédaction

Il est fréquent que le style de rédaction soit source de biais, soit dans la compréhension des questions, soit dans les réponses.

Il faut éviter les questions contenant une négation ou double négation. En plus de sa complexité, il y a le risque d'orienter la réponse dans un sens particulier.

Il faut aussi éviter les questions tendancieuses, par exemple la tendance naturelle à dire oui.

Une question ne doit être longue, elle ne doit pas dépasser une ligne à une ligne et demi.

# Règles liées à la forme du questionnaire

## Structure du questionnaire

La structure du questionnaire concerne l'ordre d'apparition des questions, la longueur et la présence de questions spéciales.

Il se structure autour de six points :

# Règles liées à la forme du questionnaire

## Structure du questionnaire

1. Présentation de l'enquêteur et de l'enquête
2. Questions d'introduction
3. Questions qualifiantes
4. Questions de mise en route
5. Questions spécifiques
6. Questions d'identification

# 1. Présentation de l'enquêteur et de l'enquête

- Organisme chargé de l'étude
- Commanditaire de l'étude
- Intérêt de l'étude
- Objectifs de l'étude

## 2. Questions d'introduction

---

leur but est d'initier la conversation et gagner la sympathie du répondant et l'intéresser au questionnaire qui va suivre. Il s'agit de questions générales souvent ouvertes.

# 3. Questions qualifiantes

Elles sont utiles si on a besoin de savoir si l'individu possède l'information recherchée. Elles sont souvent de forme dichotomique. Elles permettent d'orienter le répondant à des parties spécifiques du questionnaire.

# 4. Questions de mise en route

---

Elles servent à centrer progressivement le thème de l'étude. On commence par des questions simples, petit à petit les questions deviennent plus complexes.

# 5. Questions spécifiques

C'est le cœur du questionnaire, les questions doivent amener les informations les plus importantes pour l'étude.

A ce stade, le répondant s'est imprégné de l'esprit de l'étude et a noué une relation de sympathie avec l'enquêteur, on peut donc lui demander des efforts de réflexion ou des informations personnelles.

# 6. Questions d'identification

---

Elles décrivent les répondants. Cette partie fournit les informations qui serviront à classer les répondants en groupes distincts. Il s'agit de variables de segmentation.

# Mise en page du questionnaire

- Aérer les questions ;
- Utiliser de temps à autre des symboles;
- Soigner la typographie du texte : gras – italique – souligné;
- Imprimer sur du papier de qualité, avec en tête de l'organisme responsable de l'étude;
- Utiliser, si nécessaire, de la couleur.

# Codage du questionnaire

Le codage d'un questionnaire a pour but de faciliter la transformation des informations d'un état brut à un état qui va permettre leur traitement statistique. Normalement, le codage doit être fait dès la rédaction des questions.

La codification vise à faciliter la saisie, et l'analyse des données. Elle consiste à donner à chaque réponse un code. (un chiffre, une lettre ou les deux)

# Codage du questionnaire

- Identification des variables relatives à chaque question.
- Prévoir un code pour les non-réponses

Ce code indique que l'information recherchée n'a pas pu être recueillie. Le répondant refuse ou oublie de répondre, l'enquêteur oublie de reporter la réponse.

Eviter d'utiliser le code zéro. On doit utiliser un code qui ne peut pas être utilisé comme réponse. De préférence retenir le même code pour tout le questionnaire. Par exemple 99.

# Codage du questionnaire

- Prévoir l'identification individuelle du questionnaire et de l'enquêteur

Cette identification permet un contrôle à posteriori de la qualité du travail effectué.

L'identification se fait en général par un code à 4 chiffres. Les deux premiers chiffres indiquent qui était l'enquêteur et les deux suivants le Numéro du questionnaire pour cet enquêteur.

# Codage du questionnaire

Les questions fermées sont codifiées au moment de l'élaboration du questionnaire. Les questions ouvertes sont codifiées après l'administration de tous les questionnaires. Il est préférable de lire un échantillon ou la totalité des questionnaires avant de décider la codification en fonction des gammes de réponses.

# Codage du questionnaire

Le code doit avoir les caractéristiques suivantes :

- La fiabilité : une même réponse doit avoir le même code;
- L'exhaustivité : le code doit être complet;
- L'exclusivité : une réponse donnée ne doit pas avoir plus qu'un code.

# Codage du questionnaire

Le codage doit se faire directement sur le support de l'enquête.

Un contrôle est nécessaire pour détecter et corriger les erreurs qui peuvent se produire lors de la codification. Pour cela, un double codage est parfois intéressant, si non un contrôle d'un échantillon aléatoire de questionnaires est nécessaire.

# ANALYSE DES RESULTATS

---

**- SAISIE**

**- CONTROLE**

**- ANALYSE DES DONNEES**

# SAISIE

La saisie consiste à reporter les codes sur un support informatique sous forme d'une matrice de données (variables/individus) en utilisant un programme adéquat. Cette opération nécessite une concentration extrême et un contrôle sévère.

# CONTROLE

Le contrôle consiste à détecter parmi les données saisies ceux jugés erronés pour les corriger ou les éliminer. Cette opération est appelée épuration des données, Elle comporte les deux étapes suivantes :

# CONTROLE

---

- **Contrôle de validité**
- **Contrôle de cohérence**

# Contrôle de validité

Il consiste à vérifier pour chaque variable si les codes ou les données saisis correspondent bien aux codes prévus lors de la codification ou au domaine de définition de la variable.

Le contrôle de validité peut se faire grâce à l'élaboration de tableaux de fréquences.

# Contrôle de cohérence

Il consiste à vérifier la cohérence des renvois prévus dans le questionnaire, c'est à dire vérifier pour certaines variables si les codes ou les données saisis concordent bien avec les codes ou les données saisis d'autres variables.

Le contrôle de validité peut se faire grâce à l'élaboration de tableaux de fréquences conditionnelles.

# ANALYSE DES DONNEES

- Présenter sous forme exploitable (tableaux statistiques, graphiques, diagrammes, etc.....) C'est ce qu'on appelle la statistique descriptive.
- Substituer à cet ensemble de renseignements quelques données simples, comparer cette représentation simplifiée à des répartitions théoriques, et tenter de donner une explication du phénomène observé et en vérifier le bien fondé.
- Interpréter, conclure et définir avec précision le degré de confiance à accorder aux conclusions générales induites à partir des faits observés.

# ANALYSE DES DONNEES

L'analyse des données peut faire l'objet d'une variable, et on parlera d'analyse unidimensionnelle. Elle peut concerner deux variables, on parle alors d'analyse bidimensionnelle. Elle peut concerner plusieurs variables, on parle alors d'analyse multidimensionnelle.

# **ANALYSE UNIDIMENSIONNELLE**

## **(ou uni-variée)**

- TRIS A PLAT**
- REPRESENTATIONS GRAPHIQUES**
- LA REDUCTION DES DONNEES**

# TRIS A PLAT

Le tri à plat consiste à condenser les données sous forme d'un tableau statistique appelé distribution de fréquences.

# Cas d'un caractère qualitatif

Une distribution de fréquences d'un caractère qualitatif se présente sous forme d'un tableau très simple. La première colonne du tableau indique la nomenclature, c'est à dire la liste des modalités du caractère étudié. La deuxième colonne comporte une succession de nombres entiers appelés effectifs ou fréquences absolues. Une troisième colonne comporte une succession de nombres décimaux appelés pourcentages ou fréquences relatives.

La somme des fréquences absolues est évidemment égale à l'effectif total.

La somme des fréquences relatives est toujours égale à 1.

# Cas d'un caractère qualitatif

nationalité d'un échantillon de 500 touristes visitant le Maroc.

Nationalité	Nombre de touristes (fréquences absolues)	Pourcentage des touristes (fréquences relatives)
Française	85	17 %
Allemande	106	21,2 %
Italienne	62	12,4 %
Hollandaise	44	8,8 %
Belge	40	8 %
américaine	70	14,0 %
Autres nationalités	93	18,6 %
Total	500	100 %

# Cas d'une variable statistique discrète

Lorsque l'effectif des données brutes d'une variable statistique discrète est très grand, il est presque impossible de présenter les données sous forme d'une série statistique. Lorsque le nombre d'observations qui se répètent est grand, il est plus facile de condenser les données dans une distribution de fréquences.

# nombre d'employés chez 1000 commerçants .

Nombre d'employés	Nombre de commerçants	proportion des commerçants	Fréquences absolues cumulées croissantes	Fréquences absolues cumulées décroissantes	Fréquences absolues cumulées décroissantes	Fréquences absolues cumulées décroissantes
0	50	5 %	50	1000	5 %	100 %
1	100	10 %	150	950	15 %	95 %
2	200	20 %	350	850	35 %	85 %
3	150	15 %	500	650	50 %	65 %
4	120	12 %	620	500	62 %	50 %
5	160	16 %	780	380	78 %	38 %
6	130	13 %	910	220	91 %	22 %
7	90	9 %	1000	90	100 %	9 %

# Cas d'une variable statistique continue

Lorsqu'on a une grande quantité de données brutes d'une variable statistique continue, il est presque impossible de présenter les données sous forme d'une série statistique. Lorsque le nombre d'observations distinctes est grand, il est plus facile de condenser les données dans une distribution de fréquences groupée en classes et de déterminer le nombre d'observations appartenant à chaque classe, bien que le regroupement de données détruise une partie de l'information initiale.

# Cas d'une variable statistique continue

On détermine la plus grande et la plus petite valeur des données brutes et on calcule l'étendue.

# Cas d'une variable statistique continue

On partage l'étendue en classes suffisamment nombreuses de même amplitude. Lorsque cela est impossible, on considère des classes d'amplitudes variables ou des classes ouvertes. Le nombre de classes est habituellement compris entre 5 et 20, cela dépend des données. Il est préférable de choisir les classes de telle sorte que les points centraux coïncident avec des données réellement observées et éviter que les bornes de classes coïncident avec des données réellement observées. Ceci tend à diminuer l'erreur de groupement.

# Cas d'une variable statistique continue

Dans chaque classe, on détermine le nombre d'observations, pour cela il suffit de compter, à partir des données brutes, le nombre d'observations appartenant à la classe.

# poids en kilogramme de 80 personnes

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

# poids en kilogramme de 80 personnes

Poids	Nombre de personnes	Nombre de personnes cumulé		Pourcentage	Pourcentage cumulé	
		croissant	décroissant		croissant	décroissant
50-54	1	1	80	1,25 %	1,25 %	100 %
55-59	2	3	79	2,5 %	3,75 %	98,75 %
60-64	11	14	77	13,75 %	17,5 %	96,25 %
65-69	10	24	66	12,5 %	30 %	82,5 %
70-74	12	36	56	15 %	45 %	70 %
75-79	21	57	44	26,25 %	71,25 %	55 %
80-84	6	63	23	7,5 %	78,75 %	28,75 %
85-89	9	72	17	11,25 %	90 %	21,25 %
90-94	4	76	8	5 %	95 %	10 %
95-99	4	80	4	5 %	100 %	5 %
TOTAL	80			100 %		

# REPRESENTATIONS GRAPHIQUES

Bien qu'un tableau statistique renferme toute l'information rassemblée, il est très utile de le traduire par un graphique. La représentation graphique d'une distribution de fréquences permet de visualiser et de déceler ses principales caractéristiques.

Selon la nature du caractère étudié on utilise différents modes de représentations graphiques. Le principe des différents graphiques est le respect de la proportionnalité entre les fréquences et les surfaces.

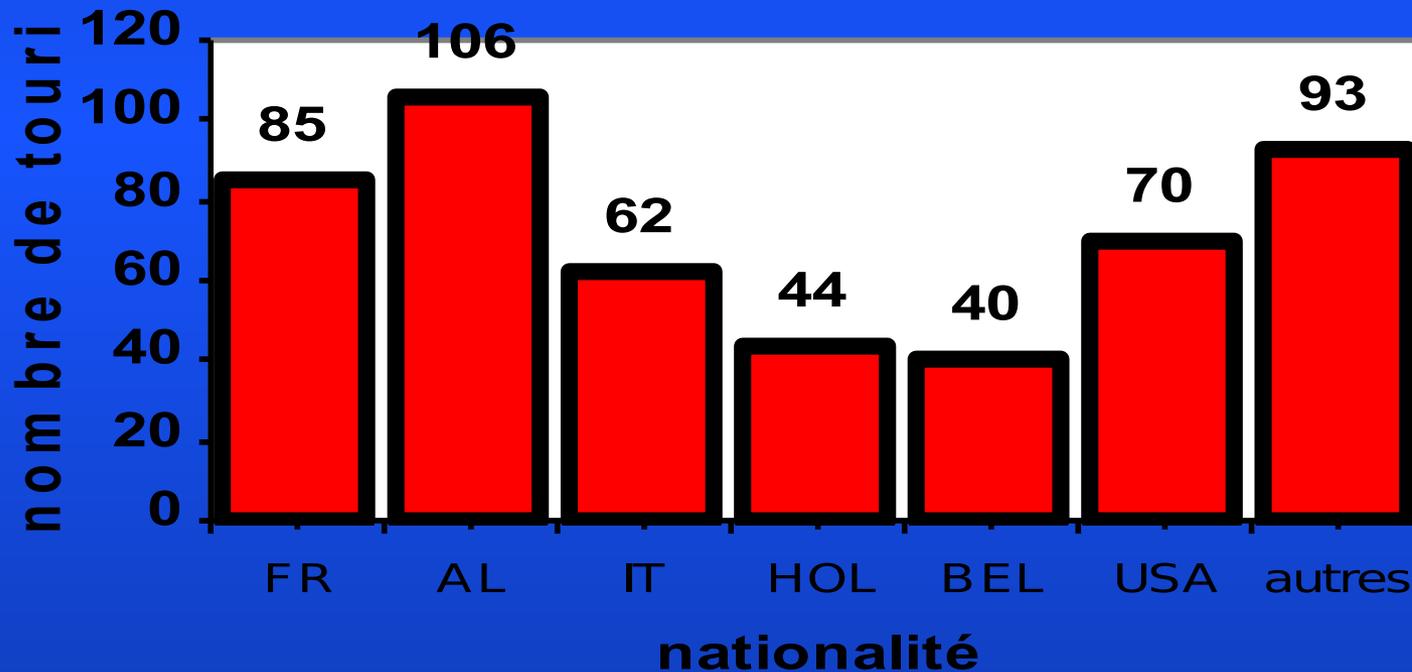
# Cas d'un caractère qualitatif

---

- Diagramme en tuyaux d'orgue
- Diagramme circulaire ou graphique à secteurs

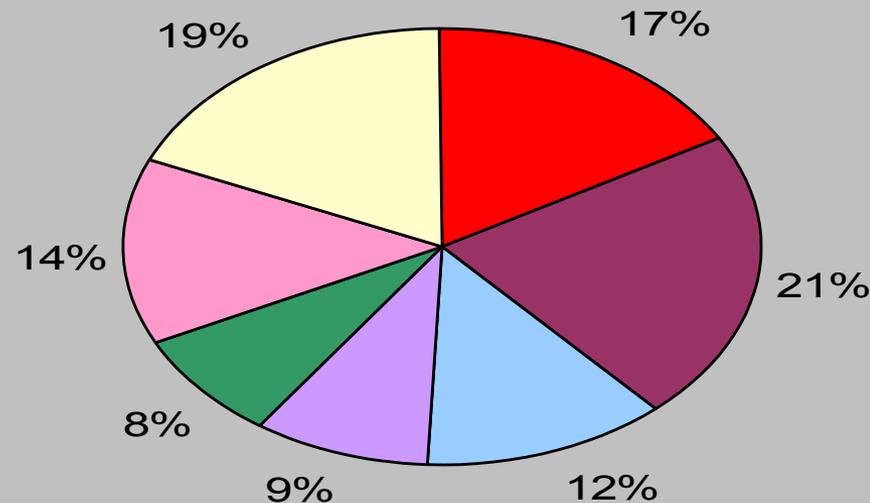
# Diagramme en tuyaux d'orgue

nationalité des touristes visitant le Maroc

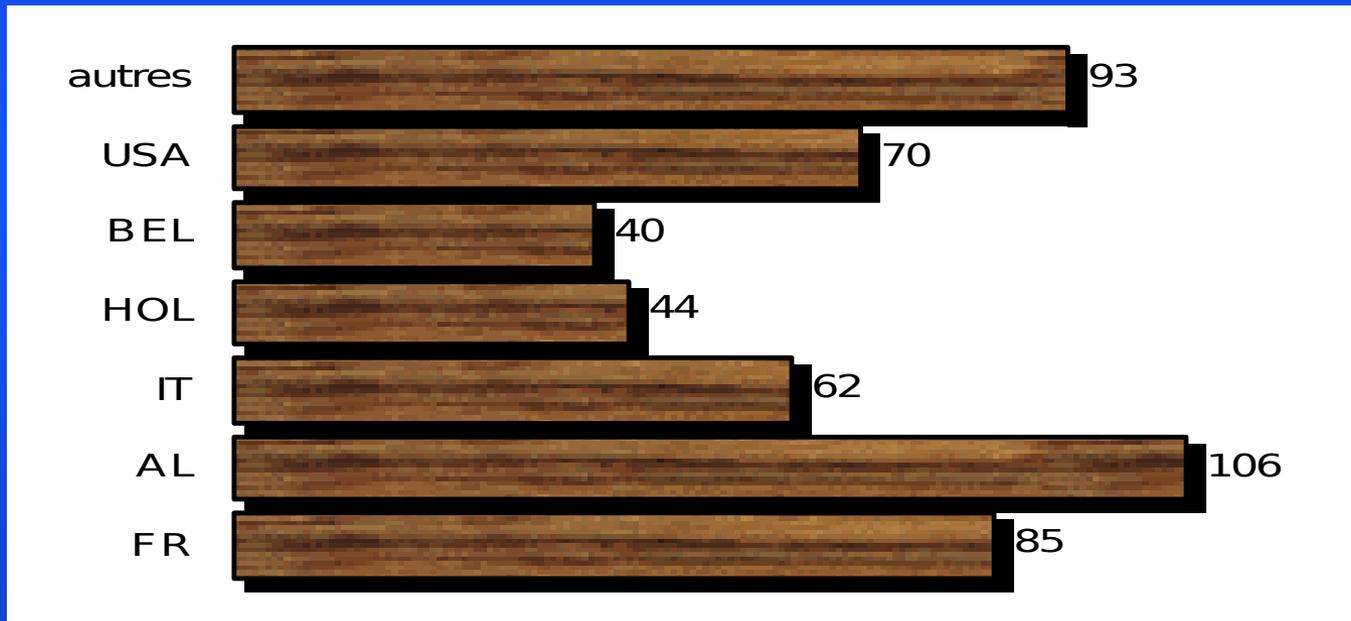


# Diagramme circulaire ou graphique à secteurs

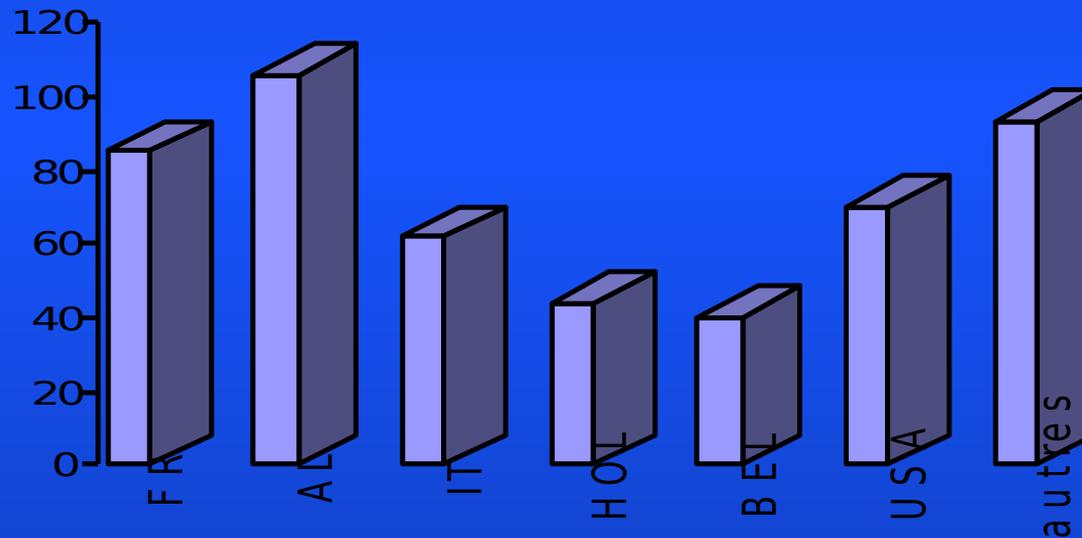
nationalité des touristes visitant le Maroc



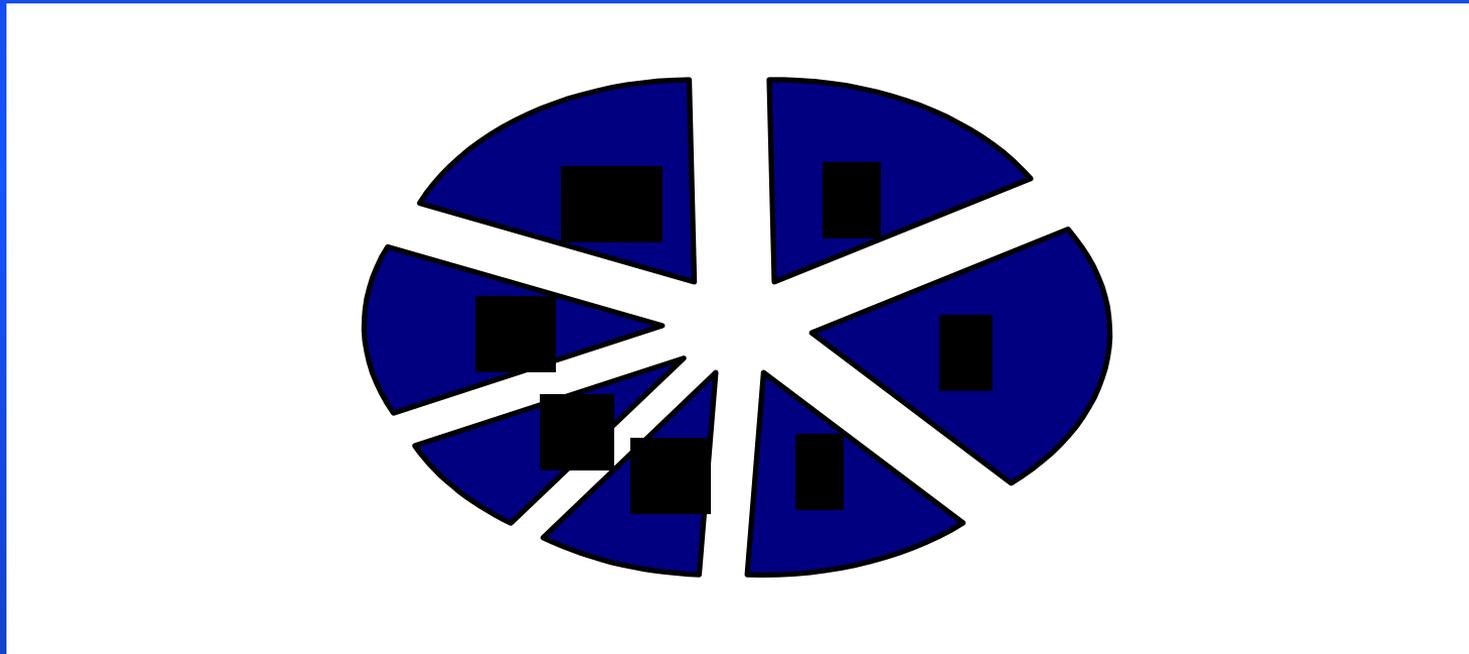
# Autres modes de représentations graphiques



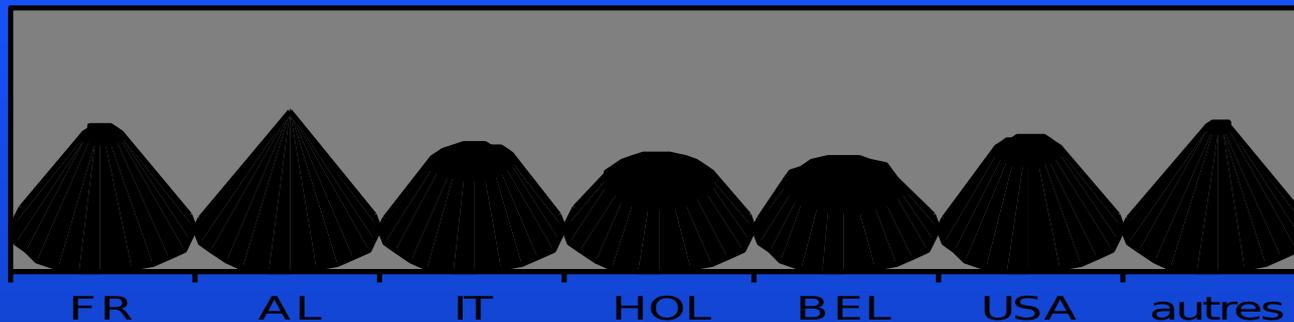
# Autres modes de représentations graphiques



# Autres modes de représentations graphiques



# Autres modes de représentations graphiques

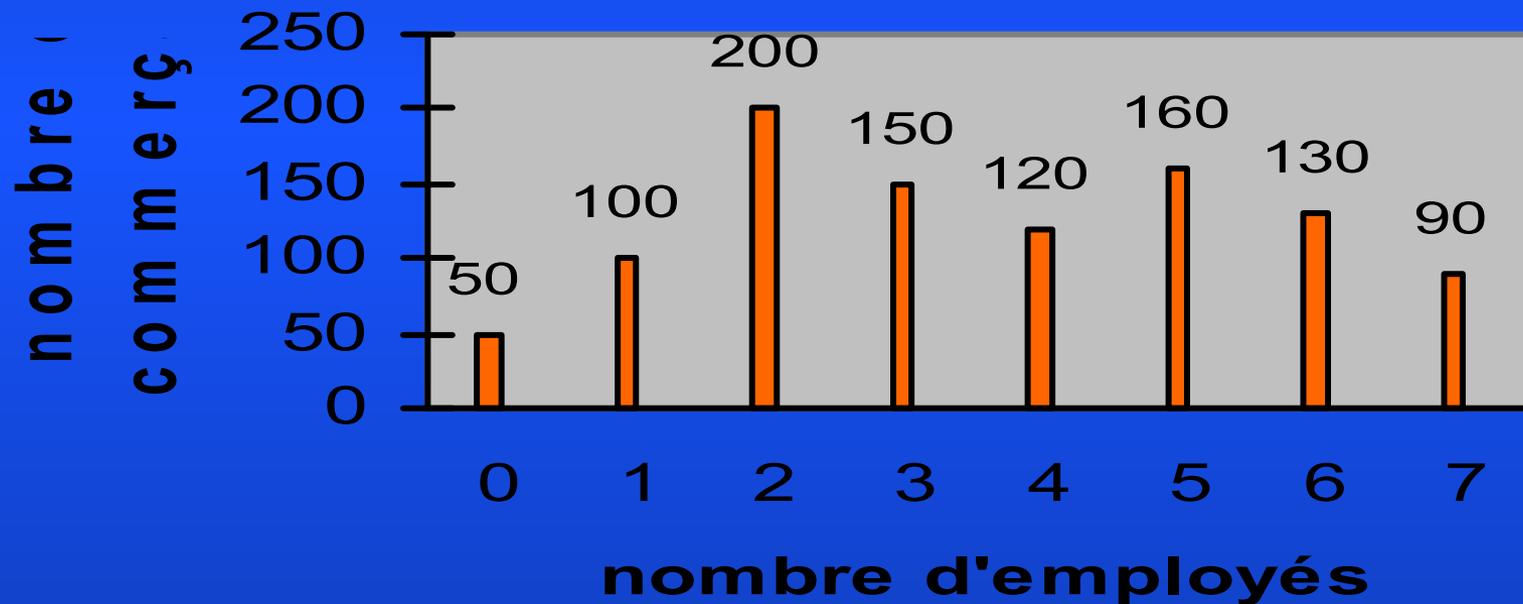


# Cas d'une variable statistique discrète

- Diagramme en bâtons
- Polygone de fréquences
- Polygone de fréquences cumulées ou diagramme en escaliers

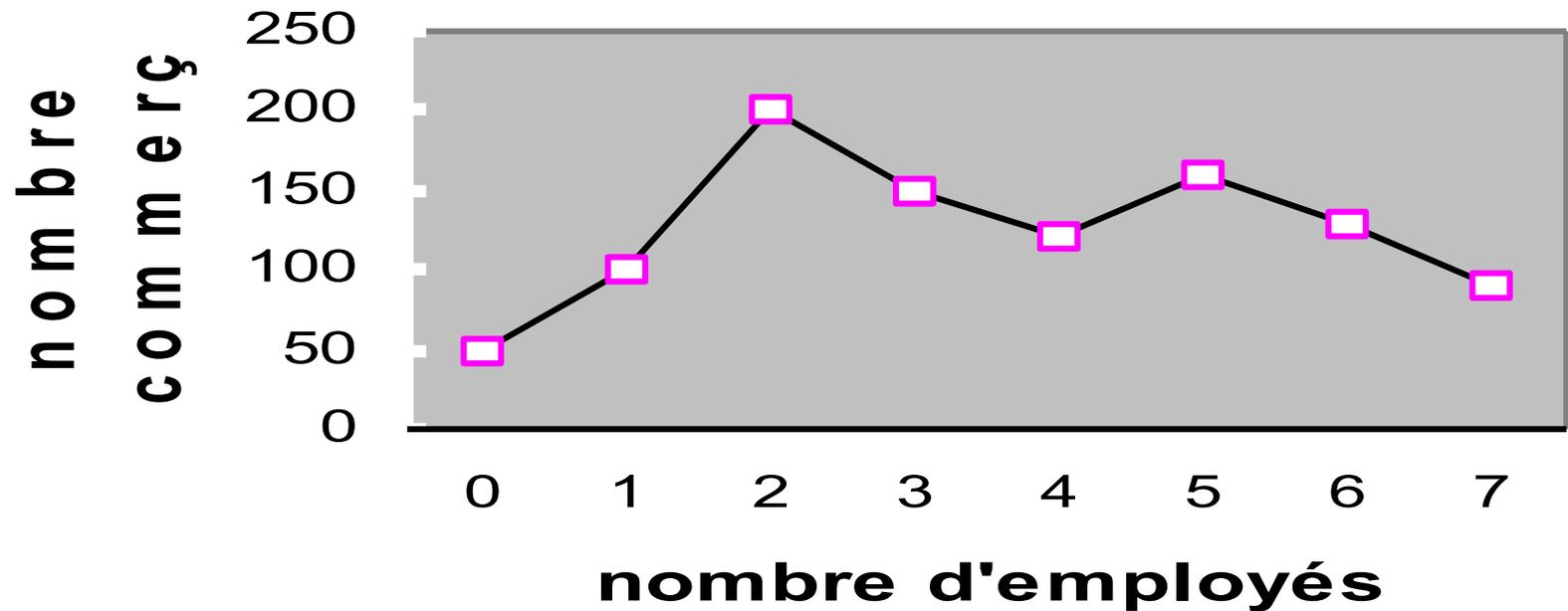
# Diagramme en bâtons

nombre d'employés chez 1000 commerçants



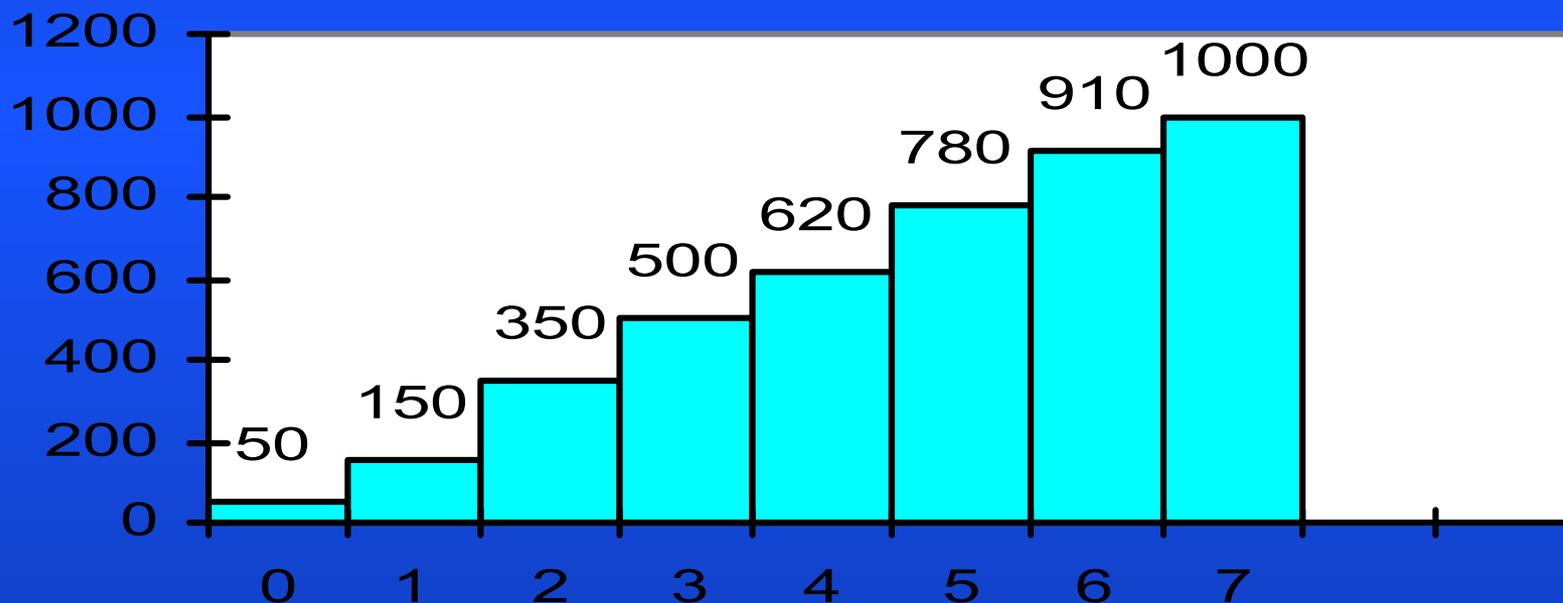
# Polygone de fréquences

nombre d'employés chez 1000 commerçants



# Polygone de fréquences cumulées ou diagramme en escaliers

nombre d'employés chez 1000 commerçants



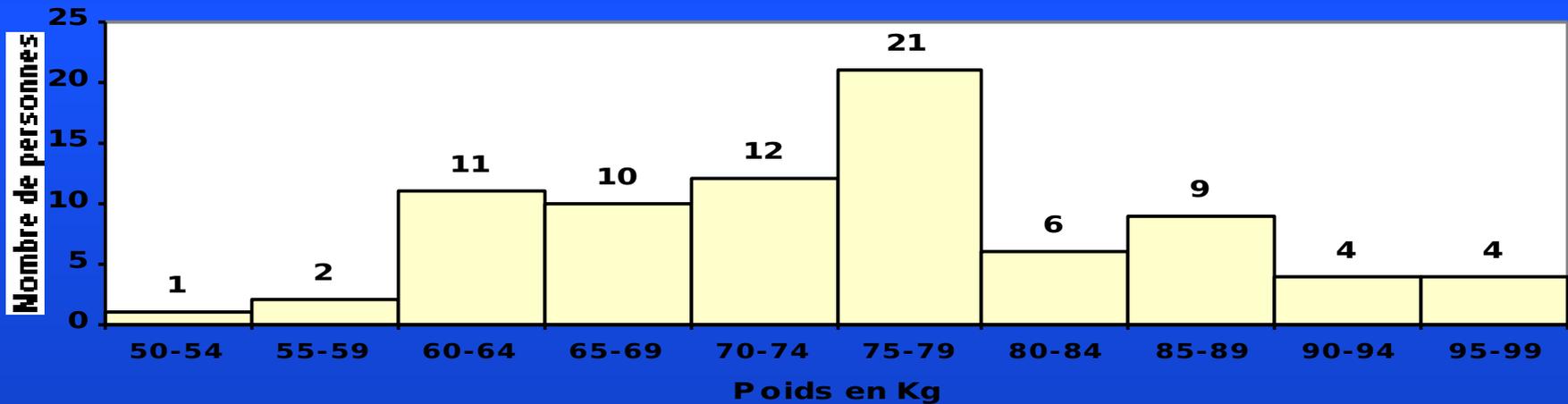
# Cas d'une variable statistique continue

---

Pour représenter une variable statistique continue, on utilise le plus souvent ce qu'on appelle histogramme.

# HISTOGRAMME

poids en kilogramme de 80 personnes



# LA REDUCTION DES DONNEES

On a vu qu'on peut décrire les données brutes à l'aide des tableaux statistiques et des graphiques. La réduction des données est une troisième méthode d'analyse des données, elle a pour objet le calcul de paramètres statistiques qui permettent de caractériser de façon simple les séries statistiques et les distributions de fréquences. On distingue plusieurs types de paramètres statistiques :

# LA REDUCTION DES DONNEES

---

- Les paramètres de position
- Les paramètres de dispersion
- Les paramètres de forme

# Les paramètres de position

Appelés aussi valeurs de tendance centrale, ils servent à caractériser l'ordre de grandeur des observations. Les principaux paramètres de position sont : la moyenne, le mode, et la médiane.

# Les paramètres de dispersion

Ces paramètres permettent de chiffrer la variabilité des valeurs observées autour d'un paramètre de position. Les principaux paramètres de dispersion sont : la variance, l'écart type, le coefficient de variation.

# Les paramètres de forme

Ces paramètres permettent de caractériser la forme et la normalité des valeurs observées. Les principaux paramètres de forme sont : le coefficient d'asymétrie et le coefficient d'aplatissement.

# Moyenne

*On peut réduire un ensemble d'observations en une seule observation constante appelée moyenne. La moyenne est donc une valeur telle que, c'est comme si toutes les observations étaient égales à cette valeur. La moyenne peut être simple si toutes les observations ont le même poids, ce qui n'est pas toujours le cas. La moyenne pondérée intervient dans le cas où les observations n'auraient pas la même importance. Il s'agit d'associer à chaque observation un coefficient de pondération indiquant son poids parmi les autres observations.*

# Moyenne arithmétique

qu'on appelle couramment moyenne, d'une série de plusieurs observations est égale à la somme de toutes les observations divisée par le nombre de ces observations.

# Moyenne géométrique

*elle est calculée pour des observations positives. Elle est égale à la racine  $n^{\text{ème}}$  du produit de toutes les observations. Elle est utilisée principalement lorsqu'on raisonne en taux (taux de croissance).*

# Moyenne harmonique

elle est égale à l'inverse de la moyenne arithmétique des inverses des observations. Son usage s'impose lorsque la variable statistique est un quotient (coût moyen, vitesse moyenne etc. ...).

# Le mode

Le mode d'une série statistique est l'observation que l'on rencontre le plus fréquemment. Le mode peut ne pas exister, et s'il existe, il peut ne pas être unique.

# La médiane

La médiane d'une variable statistique est une valeur pour laquelle, la moitié des observations lui sont inférieure ou égales et la moitié supérieure ou égales. La médiane partage donc le nombre total d'observations en deux parties égales. La médiane est un paramètre statistique qui ne dépend que du nombre d'observations. pour déterminer la médiane, il faut raisonner en terme de fréquences cumulées, la médiane est alors la valeur de la variable qui correspond à la moitié de l'effectif total.