

2. VARIABLES STATISTIQUES QUALITATIVES

Définition : une variable est une caractéristique pouvant prendre plusieurs des valeurs d'un ensemble d'observations possibles ,auquel une mesure ou une qualité peut être appliquée. Par exemple, la variable « Sexe » peut avoir 2 observations différentes possibles : « Masculin » et « Féminin ». D'autres variables possibles sont l'âge, le poids, la couleur, etc.

La distinction entre les types de variables décide des analyses statistiques utilisables, des représentations graphiques utilisables ou encore des tests statistiques. On doit pouvoir choisir les bonnes analyses pour que le travail soit valide.

Variables qualitatives :

Définition : comme le nom l'indique, les variables qualitatives contiennent des valeurs qui expriment une qualité comme le *sexe*, la *couleur* ou le *nom*. Elles peuvent être :

Nominales, comme par exemple le nom des journaux, le signe astrologique, le nom des personnes, la couleur;

Ordinales, désigne le rang ou la préférence comme : un peu, moyen, beaucoup,...

Variables qualitatives nominales

Définition

Une variable est qualitative nominale quand ses valeurs sont des éléments d'une catégorie non hiérarchique. C'est-à-dire que ses éléments ne peuvent pas se ranger dans une gradation logique.

Exemple de gradation logique: pas d'accord, plutôt d'accord, d'accord, fortement d'accord.

Exemples

Variables qualitatives nominales

Variables

Sexe

Profession

Catégories

Masculin

Féminin

Ingénieur

Médecin

Pharmacien

Utilisation

Les variables qualitatives nominales ne se prêtent qu'au calcul du mode.

Contre-indication

Pour être utilisables avec les logiciels statistiques, les variables qualitatives nominales sont encodées numériquement. Par exemple, on peut définir que la variable « Sexe » à la valeur « 1 » pour masculin et « 2 » pour féminin.

Ces nombres n'ont pas de signification et ne peuvent faire l'objet d'opérations arithmétiques comme par exemple le calcul d'une moyenne. En réalité, ce sont des « *numéros* ». Le problème est que les logiciels vont quand même faire les calculs, même si les résultats n'ont aucun sens. Que veut dire la moyenne *Ingénieur + Médecin + Pharmacien*?

Variables qualitatives ordinales

Définition

Une variable est qualitative ordinale quand ses valeurs sont des éléments d'une catégorie hiérarchique. C'est-à-dire que ses éléments peuvent être rangés dans une gradation logique.

Exemple de gradation logique :

- Niveau de scolarité : primaire, secondaire, universitaire;
- Jugement, appréciation : Jamais, rarement, à l'occasion, fréquemment, toujours.

On a ainsi une gradation des valeurs, une *relation d'ordre* entre chacune des appréciations.

Utilisation

Permet d'avoir une appréciation d'éléments difficilement quantifiables : préférences, goûts, appréciations.

Contre-indication

La gradation hiérarchique n'est pas nécessairement régulière. Ainsi, la gradation dans la scolarité de l'exemple ci-dessus n'est pas la même : le niveau *primaire* représente 6 années d'études tandis que le niveau *universitaire* représente 4 années. Avec des variables qualitatives ordinales, on calcule la médiane ou le mode plutôt que la moyenne.

Exemples de variables qualitatives ordinales

Variables qualitatives ordinales

Variables

Catégories

Préférences

Tout a fait en désaccord
Plutôt en désaccord
D'accord
Plutôt d'accord
Tout a fait d'accord

Éducation

Analphabète
Primaire
Secondaire
Collégial
Universitaire

Les modalités d'un caractère qualitatif peuvent être nominales ou ordinales.

En termes nominales les modalités sont exprimées par des noms et ne sont pas hiérarchisées ex couleur des yeux, groupes sanguins. Un caractère nominal peut être dichotomique s'il ne peut prendre que deux modalités ex sexe.

Mesurées dans une échelle ordinale, les modalités traduisent le degré d'un état caractérisant un individu sans que ce degré ne puisse être défini par un nombre qui résulte d'une mesure. Les modalités sont alors hiérarchisées ex stade d'une maladie (très satisfaisant ; satisfaisant ; moyen).

2.1. constitution d'une série qualitative

Une série statistique correspond aux différentes modalités d'un caractère sur un échantillon d'individus appartenant à une population donnée. Le nombre d'individus qui constituent l'échantillon étudié s'appelle la taille de l'échantillon.

Exemple. Afin d'étudier la structure de la population des stagiaires de 1 ère année, une étude peut être entreprise sur un échantillon de 50 stagiaires portant sur leur degré de satisfaction des résultats de l'actuel gouvernement. (voir corrigé)

2.2. Tableaux statistiques

Le tableau statistique d'effectifs ou de fréquences est un mode synthétique de présentation des données. Sa constitution est immédiate dans le cas des caractères qualitatifs et discrets mais nécessite en revanche une transformation des données dans le cas continu.

Le tableau est constitué par les modalités du caractère des effectifs et des fréquences relatives et cumulées (voir corrigé)

2.3. REPRESENTATION GRAPHIQUE

La représentation graphique des données relatives à un caractère unique repose sur la proportionnalité des longueurs, ou des aires, des graphiques, aux effectifs, ou aux fréquences, des différentes modalités du caractère. Pour un caractère qualitatif, on utilise principalement trois types de représentation graphique : le diagramme en bâtons, la représentation par tuyaux d'orgue et la représentation par secteurs.

Lorsque le caractère étudié est la répartition géographique d'une population, la représentation graphique est un cartogramme.

a) Diagramme en bâtons.

Nous portons en abscisse les modalités, de façon arbitraire.

Nous portons en ordonnée des segments dont la **longueur** est proportionnelle aux effectifs (ou aux fréquences) de chaque modalité.

Nous appelons **polygone statistique**, ou **diagramme polygonal**, la ligne obtenue en joignant les sommets des bâtons.

b) Tuyaux d'orgue.

Nous portons en abscisses les modalités, de façon arbitraire.

Nous portons en ordonnées des rectangles dont la **longueur** est proportionnelle aux effectifs, ou aux fréquences, de chaque modalité.

c) Secteurs.

Les diagrammes circulaires, ou semi-circulaires, consistent à partager un disque ou un demi-disque, en tranches, ou secteurs, correspondant aux modalités observées et dont la **surface** est proportionnelle à l'effectif, ou à la fréquence, de la modalité.

Ces diagrammes conviennent très bien pour des données politiques ou socio-économiques.

Les diagrammes circulaires s'obtiennent en calculant l'angle ($a^\circ = f_i \times 360^\circ$ ou 180°) c'est-à-dire la fréquence relative multipliée par 360° cercle ou 180° semi cercle

d) Exemple.

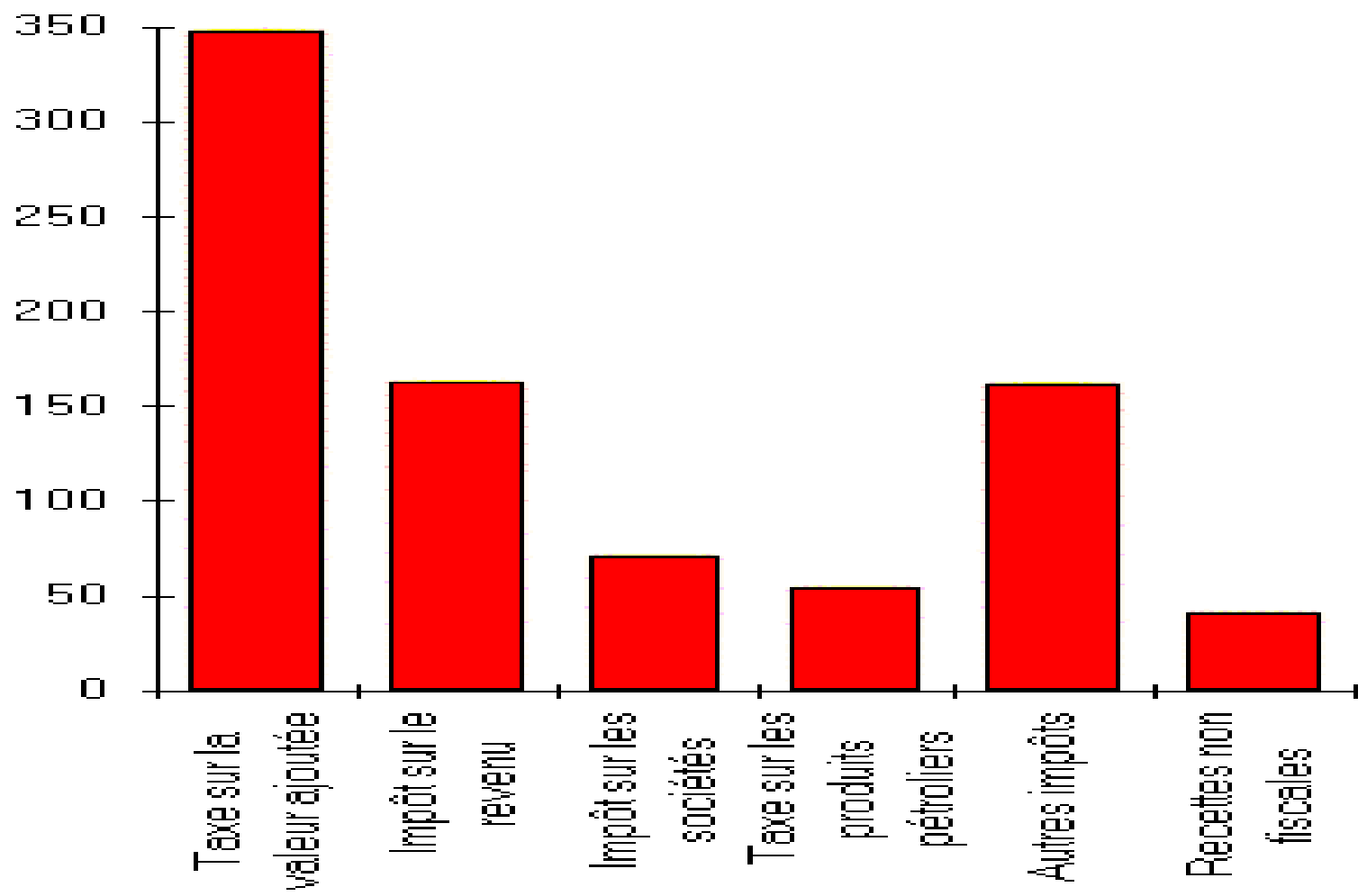
En 2005, les recettes du budget de l'Etat se présentaient de la façon suivante (en milliards) :

Le caractère étudié, la nature des recettes du budget de l'Etat, est un caractère qualitatif.

Taxe sur la valeur ajoutée	348
Impôt sur le revenu	163
Impôt sur les sociétés	71
Taxe sur les produits pétroliers	54
Autres impôts	161
Recettes non fiscales	41
TOTAL	838

Dans la représentation en **tuyaux d'orgue**, les différentes modalités du caractère (les diverses sources de recettes du budget de l'Etat) sont représentées par des segments sur l'axe des ordonnées.

Pour chaque abscisse on porte un rectangle dont la longueur est proportionnelle au montant correspondant de la recette (effectif).

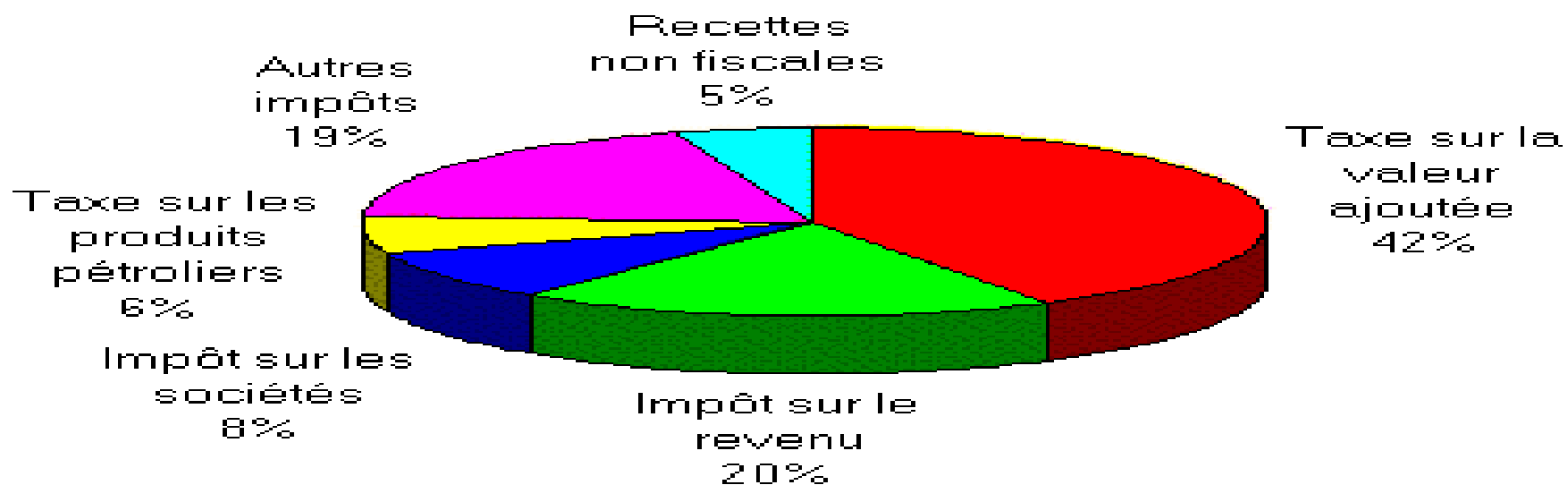


Histogramme des effectifs

Dans la représentation par **diagramme en bâtons**, les différentes modalités du caractère (les diverses sources de recettes du budget de l'Etat) sont représentées par des points sur l'axe des ordonnées. Pour chaque abscisse, on porte un segment vertical dont la longueur est proportionnelle au montant correspondant de la recette (rectangle de largeur nulle).

Dans le **diagramme circulaire**, chaque secteur a une surface proportionnelle à l'importance de la recette dans le budget. L'angle au centre représentant une modalité est donc proportionnelle à l'importance de la recette dans le budget.

Source de recettes	Recette		Angle (degrés)	Angle cumulé	
	MF	%		\widehat{xOy}	$\widehat{xOy} - 180^\circ$
Taxe sur la valeur ajoutée	348	41,5	149,5	149,5	
Impôt sur le revenu	163	19,5	70,0	219,5	39,5
Impôt sur les sociétés	71	8,5	30,5	250,0	70,0
Taxe sur les produits pétroliers	54	6,4	23,2	273,2	93,2
Autres impôts	161	19,2	69,2	342,4	162,4
Recettes non fiscales	41	4,9	17,6	360,0	180,0
TOTAL	838	100			



Recettes 1982

Diagramme circulaire des fréquences

e) Cartogrammes.

Un cartogramme est une carte géographique dont les secteurs **géographiques** sont coloriés avec une couleur différente suivant l'effectif ou suivant la fréquence du caractère étudié.

2.4. EXERCICES

Un opérateur téléphonique veut savoir si ses abonnés au téléphone portable sont satisfaits de leur forfait. Il fait une enquête auprès de 300 abonnés au téléphone portable. Le tableau 1 résume les réponses obtenues :

	Forfait Ultra-Prime	Forfait Super-Plus	Total
Satisfait	150	50	200
Non satisfait	50	50	100
TOTAL	200	100	300

1. Quelle est la population ? Qu'est-ce qu'un individu ? Quel est l'échantillon ? Quelles sont les variables en jeu ? leur type ?
2. Quel est l'effectif de la modalité "Satisfait" dans cet échantillon ?
3. Calculer la fréquence correspondante.
4. Calculer la fréquence de la modalité "Forfait Ultra-Prime" dans cet échantillon.
5. Quel est le pourcentage de non-satisfaits dans l'échantillon.
6. Calculer la proportion de non-satisfaits parmi les abonnés au forfait Ultra-Prime.
7. Calculer la proportion de non-satisfaits parmi les abonnés au forfait Super-Plus.
8. Calculer la proportion d'abonnés à chacun des deux forfaits parmi les clients satisfaits.
9. Calculer la proportion d'abonnés à chacun des deux forfaits parmi les clients non-satisfaits.
10. Parmi les 300 abonnés interrogés, quel pourcentage représentent les abonnés satisfaits au forfait Ultra-Prime ?
11. Selon vous, les deux forfaits donnent ils également satisfaction à leurs abonnés ?