Croisement de deux variables catégorielles

I) Proportion et pourcentage (rappel)

1) Définition

Considérons une population de référence E contenant n_E éléments et une sous-population de E, que l'on note A, contenant n_A éléments.

La proportion des éléments de A par rapport à E est :

P = $\frac{n_A}{n_E}$

Si on veut exprimer cette proportion en pourcentage il suffit de multiplier p par 100 : P = $\frac{n_A}{n_E} \times 100$

Pour déterminer le pourcentage des éléments de A par rapport à E, on calcule d'abord la proportion des éléments de A par rapport à E et on ramène l'effectif de référence à 100.

Remarque : Dire que 36% de la population française est du groupe sanguin 0 Rhésus + signifie que la proportion de personne étant du groupe 0+ par rapport à la population française est de $\frac{36}{100}$ (l'effectif total étant ramené à 100).

Exemple 1 : Dans une classe de 25 élèves d'une classe de seconde, 11 ont n'ont pas encore eu 15 ans. Quelle est la proportion d'élèves de moins de 15 ans ?

E représente le nombre d'élèves en classe de seconde et A le nombre d'élèves de seconde ayant moins de 15 ans

P = $\frac{n_A}{n_E}$ = $\frac{11}{25}$ **La proportion d'élèves de moins de 15 ans est** $\frac{11}{25}$

Si on veut se résultat en pourcentage, on fait $\frac{11}{25} \times 100 = 44$

44% des élèves de classe seconde ont moins de15 ans.

Exemple 2 : Dans une entreprise on sait que 30% des salariés partent en vacances en juillet, les autres partant au mois d'août. Ce qui représente un nombre de 150 employés qui sont partis en juillet. Quel est le nombre de salariés dans cette entreprise ?

La proportion de salariés qui part en vacances en juillet est de $\frac{30}{100}$.

Le nombre de personnes parties en vacances en juillet est de 150 donc

n_A =150, on cherche la valeur de n_E.

P = $\frac{150}{n_E} = \frac{30}{100}$ on en déduit que $n_E = \frac{150 \times 100}{30} = 500$

Le nombre de salariés de cette entreprise est de 500.

2) Proportion de proportion et pourcentage de pourcentage

Considérons une population de référence E contenant n_E éléments, une sous-population de E, que l'on note A , contenant n_A éléments et une sous-population de A , que l'on notera B, contenant n_B éléments.

Notons p_1 la proportion de B dans A et p_2 la proportion de A dans E :

La proportion des éléments de B par rapport à E est le produit de p_1 et p_2 :

$$p = p_1 \times p_2$$

Lorsque les proportions sont exprimées en pourcentage, on fait de même : pour calculer le pourcentage d'un pourcentage on multiplie les pourcentages entre eux.

Exemple 1 :

A la rentrée 2020, un lycée compte 35 % d'élèves en seconde parmi lesquelles 40% étudient l'italien en LV2. Parmi les élèves de ce lycée, quelle est la proportion d'élèves de seconde étudiant l'italien en $2^{\text{ème}}$ langue ?

Par rapport à la formule de la leçon : E représente l'ensemble des élèves du lycée.

A représente le nombre d'élèves en classe de seconde de ce lycée et B le nombre d'élèves de secondes qui étudient l'italien en LV2.

P_1 est la proportion d'élèves de seconde qui étudient l'italien : $P_1 = \frac{40}{100}$

P_2 est la proportion d'élèves de seconde par rapport au nombre d'élèves du lycée : $P_2 = \frac{35}{100}$

La proportion d'élèves de seconde étudiant l'italien en $2^{\text{ème}}$ langue est donc :

$P = P_1 \times P_2$ donc $\quad P = \frac{40}{100} \times \frac{35}{100} = \frac{1\,400}{10\,000} = \frac{14}{100} = 0{,}14$

La proportion d'élèves de seconde étudiant l'italien en $2^{\text{ème}}$ langue est donc 0,14 ou $\frac{14}{100}$.

Exemple 2 : 42% de la population française possède le groupe sanguin O, parmi ces personnes, 14% sont de Rhésus -. Quel pourcentage de la population française est du groupe sanguin O- ?

E est la population Française.

A est la population de groupe sanguin O (A est une sous population de E)

B est la population de groupe sanguin O et **Rhésus –** (E est une sous population de A).

Le pourcentage de personnes de groupe O- est :

$P = \frac{42}{100} \times \frac{14}{100} = \frac{588}{10\,000} = \frac{5{,}88}{100} = 0{,}0588$

0,0588% de la population Française est du groupe O-

3) Calcul de TVA

La taxe sur la valeur ajoutée ou TVA est un impôt indirect sur la consommation. C'est un impôt institué en France pour une loi du 10 avril 1954.

En France elle est de 20%, excepté dans quelques secteurs où elle est de 5,5% ou 7%. Ainsi quand vous achetez un produit, le prix se décompose en prix brut (ce qu'on appelle le prix hors taxe) et en prix net (ce qu'on appelle TTC) qui est le prix brut auquel on ajoute la TVA correspondante.

Exemple 1 : Calcul du prix net :

Pour calculer le montant d'une TVA à 20% revient à calculer une proportion de 20% du prix brut, c'est-à-dire multiplier ce prix par 0,2

Un article qui coûte 50 € hors taxe avec une TVA de 20%.

$0{,}2 \times 50 = 10$ Le prix net (prix TTC) est de 50+10 = 60

Le prix TTC est de 60 €

Exemple 2 : Calcul du prix net directement

Pour calculer le prix net directement d'un article dont le prix brut est de 50€ avec une TVA de 20% revient à multiplier 50 par 1,2 .

$50 \times 1{,}2 = 60$

Le prix TTC est de 60 €

Exemple 3 : Calcul du prix brut

Pour retrouver le prix brut directement d'un article dont le prix net est de 150€ avec une TVA de 20% revient à diviser 150 par 1,2 .

$150 \div 1{,}2 = 125$

Le prix brut est de 125 €

II) Tableau croisé d'effectifs de deux variables catégorielles

1) Exemple et définition

Pour la fabrication d'un composant de smartphone, un contrôle de qualité est effectué, mais sa fiabilité n'est pas totale : dans 2 % des cas il accepte un composant défectueux et dans 1 % des cas il rejette un composant conforme. Les résultats sont résumés dans le tableau suivant :

	Conforme	Défectueux	Total
Accepté	1 485	25	1 510
Rejeté	15	1 225	1 240
Total	1 500	1 250	2 750

Marges du tableau

Marges du tableau **Effectif total**

Dans cet exemple deux variables sont étudiées en même temps : des composants défectueux ou pas et des composants rejetés ou pas.

Définition :

De telles variables sont appelées qualitatives ou catégorielles par opposition aux variables quantitatives (poids, tailles, nombre d'enfants par famille…) étudiées les années précédentes qui permettaient d'en calculer les moyennes, médianes… Ces variables représentent une qualité et non une quantité.

Dans ce type de tableau on met dans la colonne de droite et la dernière ligne les totaux appelés marges du tableau et la case en bas à droite donne l'effectif total.

2) Compléter un tableau croisé d'effectifs

Reprenons l'exemple précédent :

Pour la fabrication d'un composant de smartphone, un contrôle de qualité est effectué, mais sa fiabilité n'est pas totale : dans 2 % des cas il accepte un composant défectueux et dans 1 % des cas il rejette un composant conforme. Les résultats sont résumés dans le tableau suivant :

	Conforme	Défectueux	Total
Accepté			1 510
Rejeté			
Total	1 500		

Comment compléter le tableau :

Première colonne : Tout d'abord nous savons qu'il y a 1 500 composants conformes, or 1% sont rejetés.

1% de 1500 : $1500 \times 0{,}01 = 15$. Il y a donc 15 composants rejetés parmi les conformes.

Comme il y a 1500 composants conformes : $1500 - 15 = 1485$, ce résultat nous donne le nombre de composants conformes acceptés. Ainsi la première colonne est remplie.

	Conforme	Défectueux	Total
Accepté	1 485		1 510
Rejeté	15		
Total	1 500		

Première ligne : On sait que le total des composants acceptés est 1510.

Pour avoir le nombre de composants acceptés mais défectueux on obtient :

1 510 – 1485 = 25

On obtient :

	Conforme	Défectueux	Total
Accepté	1 485	25	1 510
Rejeté	15		
Total	1 500		

Deuxième colonne : On sait que le total des composants défectueux acceptés est 25 et qu'ils représentent 2% des composants défectueux.

On obtient le nombre total des composants défectueux : $25 \times \frac{100}{2} = 1250$

Pour avoir les composants défectueux refusés : $1250 - 25 = 1225$ (nombre total des composants défectueux moins ceux qui sont acceptés)

On obtient :

	Conforme	Défectueux	Total
Accepté	1 485	25	1 510
Rejeté	15	1 225	1 240
Total	1 500	1 250	2 750

Pour avoir la dernière colonne il suffit de faire les totaux de chaque ligne :

15 + 1 225 =1 240

1 500 +1 250 = 2 750

L'effectif total est 2750

III) Fréquence conditionnelle. Fréquence marginale

1) Rappel

La fréquence f d'une sous-population A dans une population E est le rapport des effectifs :

$$f = \frac{n_A}{n_E}$$

Exemple : Dans une classe de 17 élèves il y 8 filles, la fréquence des filles est : $\frac{8}{17}$

2) Fréquence marginale

En reprenant le tableau de l'exercice précédent :

	Conforme	Défectueux	Total
Accepté	1 485	25	**1 510**
Rejeté	15	1 225	1 240
Total	**1 500**	1 250	**2 750** **2750**

1) Déterminons la fréquence marginale dont les composants sont acceptés :

Il y a 1510 composants acceptés sur un total de 2750 composants

La fréquence marginale est donc : $f_m = \frac{1510}{2750} \approx$ **0,549**

2) Déterminons la fréquence marginale dont les composants sont conformes :

Il y a 1 500 composants conformes sur un total de 2750 composants

La fréquence marginale est donc : $f_m = \frac{1500}{2750} \approx$ **0,545**

Définition :

Dans un tableau croisé d'effectifs, une sous population dont l'effectif figure dans une marge, ne dépend que d'une seule variable et sa fréquence est appelée de fréquence marginale :

$$f_m(A) = \frac{n_A}{n}$$

3) Fréquence conditionnelle

En reprenant le tableau de l'exercice précédent :

	Conforme	Défectueux	Total
Accepté	1 485	25	1 510
Rejeté	**15** **15**	1 225	**1 240**
Total	**1 500**	1 250	2 750 2750

1) Déterminons la fréquence conditionnelle dont le composant est **rejeté parmi les composants conformes :**

Il y a 1500 composants conformes et 15 composés à la fois conformes et rejetés donc

la fréquence conditionnelle est donc : $f_c = \frac{15}{1500} \approx$ **0,01**

2) Déterminons la fréquence conditionnelle dont les composants **sont conformes parmi les rejetés :**

Il y a 1 240 composants rejetés et 15 conformes parmi les rejetés   donc

la fréquence conditionnelle est donc : $f_c = \frac{15}{1240} \approx$ **0,012**

Définition :

$n_A \neq 0$ **et** $n_B \neq 0$ **la fréquence conditionnelle de B sachant A est :**

$$f_A(B) = \frac{n_{A \cap B}}{n_A}$$

Propriété :

la fréquence conditionnelle de B sachant A est :

$$f_A(B) = \frac{f_{A \cap B}}{f_A}$$

Exemple : Voici ci-dessous les activités préférées d'un groupe de personne :

	Natation	Tennis	Cinéma
Homme	18	25	35
Femme	25	28	55

1. Quelle est la fréquence conditionnelle des personnes préférant la natation parmi les femmes :

$f_A(B) = \frac{f_{A \cap B}}{f_A}$ A représente le nombre de femmes il y en a : 25+28+55= 108

$A \cap B$ représente le nombre de femmes préférant la natation : il en a 25

Donc $f_{\boldsymbol{A}}(B) = \frac{25}{108} \approx 0{,}23$

2. Quelle est la fréquence conditionnelle des femmes parmi les personnes préférant la natation :

$f_{\boldsymbol{B}}(A) = \frac{f_{A \cap B}}{f_B}$ B représente le nombre de personnes préférant la natation : 18+25 = 43

$A \cap B$ représente le nombre de femmes préférant la natation : il en a 25

Donc $f_{\boldsymbol{B}}(A) = \frac{25}{43} \approx 0{,}81$