

Projet de pôle : Analyse de l'information

COUDERT T. - LOISEAU K. - MACHIZAUD A. - PRAYER T. - TYGREAT T.

Livrable n°2 :

Découpage du problème & Approche algorithmique

Sommaire

Récapitulatif du sujet.....	3
Approches de la résolution du problème.....	3
Echantillonnage	4
a. Schema	4
b. Algorithmique.....	5
Statistiques univariées.....	6
a. Structures utilisées :.....	6
b. Données Quantitatives.....	7
a. Schéma de décomposition :.....	7
b. Algorithmes :	8
c. <i>Données qualitatives</i>	12
a. Schéma de décomposition :.....	12
b. Algorithmes :	13
Statistiques bivariées : croisement qualitatif * qualitatif.....	15
b. Structures utilisées	16
c. Schéma de décomposition	17
d. Algorithmes	19
Statistiques bivariées : croisement qualitatif * quantitatif.....	22
a. Structures utilisées :.....	22
b. Schéma de décomposition	24
c. Partie Algorithme	26

Récapitulatif du sujet

Nous cherchons à mener une étude statistique sur une population d'eistiens qui ont fini leur cursus. Les données en notre possession sont diverses : sexe, origine scolaire, origine géographique, mention au bac, etc.

Notre étude se concentrera sur l'influence des différents paramètres sur son choix de l'option. A partir de ce sujet, nous pouvons effectuer trois types d'analyses pour ce livrable :

- a. Une analyse statistique univariée
Le but ici est de décrire le profil eistien dans notre population totale : pourcentage de IFI, DSI, etc., pourcentage de filles, ...
- b. Une analyse statistique bivariée : croisement qualitatif * qualitatif
On croisera le choix de l'option avec d'autres caractères qualitatifs tels que l'origine géographique ou l'origine scolaire par exemple.
- c. Une analyse statistique bivariée : croisement qualitatif * quantitatif
Dans celle-ci, on pourra combiner le choix d'option en fonction des notes obtenues, âge, etc.

Le fait que nous ne traitons pas le cas de « l'analyse statistique : croisement quantitatif * quantitatif » provient de notre choix du sujet. Nous nous intéressons à l'influence de différents caractères sur le choix de l'option, or le choix de l'option est un caractère qualitatif donc on ne peut que faire des analyse croisées de type « qualitatif*qualitatif » et « qualitatif*quantitatif ».

Cependant lors de prochaines analyses statistiques multivariées, on pourra s'intéresser à un niveau supérieur de croisement, avec plusieurs caractères quantitatifs croisés.

Approches de la résolution du problème

Afin de mener à bien notre projet nous avons découpé notre problème selon le schéma suivant :

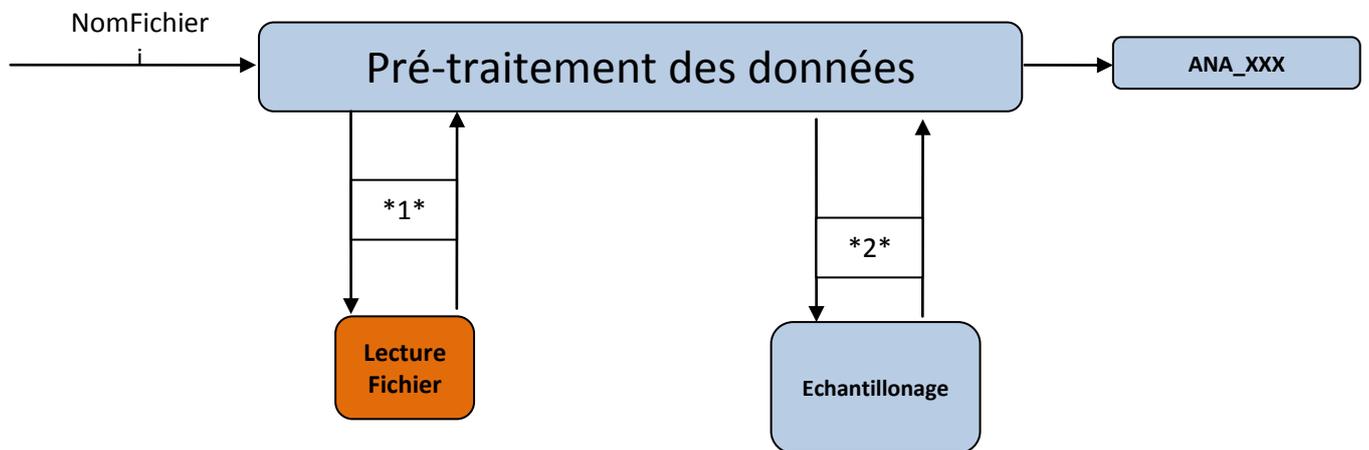


Après avoir reçu notre fichier en entrée, nous garderons uniquement les données qui nous sont utiles en sélectionnant les colonnes correspondantes, puis nous effectuerons diverses opérations pour traiter les données selon les statistiques que l'on veut en tirer.

Ces étapes seront décrites dans chaque sous-partie qui correspondent à une analyse différente car l'approche n'est pas identique selon la statistique voulue.

Echantillonnage

a. Schema



Fonction *1* :

Paramètre : E nom_Fichier : String ; E i : Entier ; E j : Entier

Retour : ANA_XXX

Description : à partir du fichier, on va récupérer les données des colonnes i et j et créer une structure d'analyse quelconque qui contient tous les individus. ANA_XXX désigne n'importe quelle structure qui commence par ANA.

Procédure *2* :

Paramètres : ES monAna : ANA_XXX, E Nb : Entier

Description : on réalise un échantillonnage sur la population de départ. On va donc modifier le contenu du paramètre d'entrée. L'entier Nb passé en paramètre permet de préciser le nombre d'individus que l'on veut garder après échantillonnage.

b. Algorithmique

Echantillonnage

Procédure Echantillonnage (ES : monAna : ANA_XXX, E Nb : Entier) // Pour simplifier la présentation on suppose ici que l'on traite une analyse univariée.

Variables

Tableau t1(1000) : Reel

Tableau t2(Nb) : Entier // Permet de stocker les indices des cases déjà sélectionnées

i,j,n : entier

Present : booléen

$i \leftarrow 1$

Tant Que ($i \leq Nb$)

$n \leftarrow \text{PartieEntiere}(\text{alea()} * \text{monAna.NbElements}) + 1$

 Present \leftarrow Faux

 Pour $j \leftarrow 1$ à (i-1) pas 1

 Si ($n = t2(j)$)

 Present \leftarrow vrai

 FinSi

 FinPour

 Si (Non Present)

$t1(i) \leftarrow \text{monAna.Données}(n)$

$t2(i) \leftarrow n$

$i \leftarrow i + 1$

 FinSi

Fin Tant Que

FinProcédure

Statistiques univariées

a. Structures utilisées :

Struct ANA_QUANT

- Quant : Quantité
- NbElements : reel
- Données : Tableau(1000) : reel

Struct ANA_QUAL

- Qual : QUALITE
- NbElements : Entier
- Données : Tableau(1000) : String

Les structures ANA_QUANT et ANA_QUAL servent à stocker les données nécessaires à l'analyse univariée (respectivement quantitative et qualitative). Elles comportent chacune un tableau : des réels pour un caractère quantitatif, des chaînes de caractères pour une analyse qualitative. On y trouve également le nombre d'éléments, utile pour pouvoir faire des boucles sur le tableau. Enfin, un élément Quantité (ou Qualité) donne des informations sur le caractère étudié.

Struct QUANTITE

- NomQuantité : String
- ValeurMin : reel
- ValeurMax : reel

Struct QUALITE

- Nom_Qualité : String
- NbEtatsPossibles : Entier
- EtatsPossibles : Tableau(30) : String

Ces structures donnent des informations sur le caractère étudié : un label permettant d'afficher son nom (option, sexe, âge, promo, moyenne, etc.). Une valeur min et une valeur max permettent de borner le caractère quantitatif (exemple : une note se situe entre 0 et 20). Dans le cas d'un caractère qualitatif, un entier donne le nombre d'état que peut prendre le caractère, et un tableau de chaînes de caractères les récapitule. Par exemple dans le cas des options, le nombre d'états possibles serait 9, et on trouverait dans le tableau : « IFI » ; « ICOM » ; « GL » ; etc.

Struct RES_ANA_QUANT

- Quant : Quantité
- Moyenne : Reel
- Mediane : Reel
- Variance : Reel
- Ecart_Type : Reel
- Min : Reel
- Max : Reel
- Premier_Quartile : Reel
- Troisième_Quartile : Reel
- NbElements : Entier

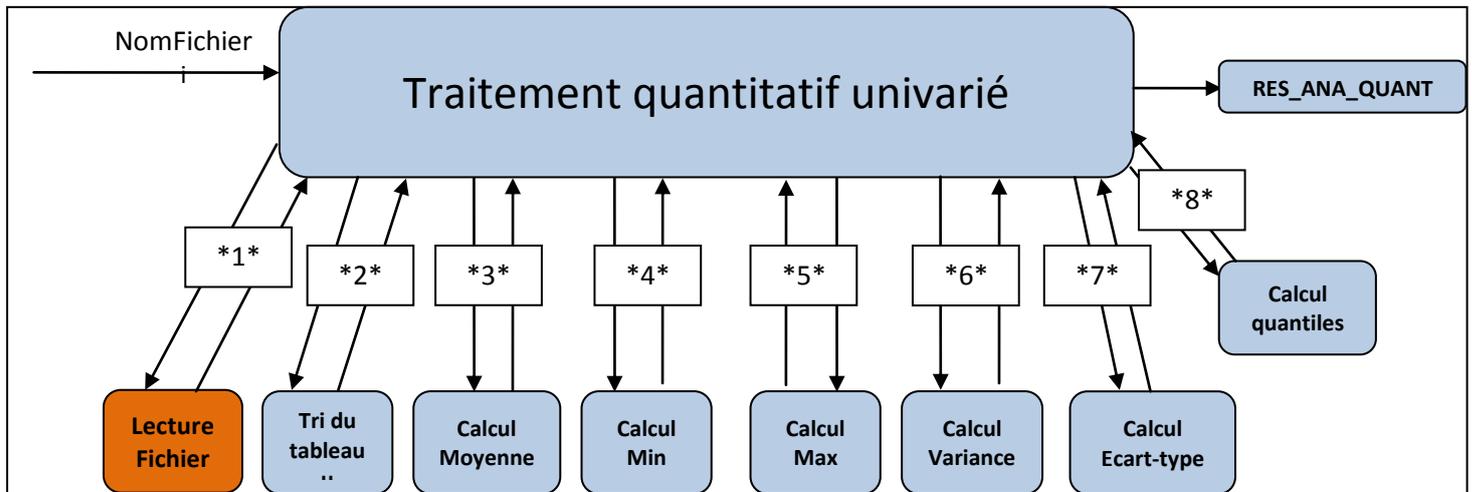
Struct RES_ANA_QUAL

- Qual : QUALITE
- Repartition_Absolue : Tableau (100) : Entier
- Repartition_Relative : Tableau (100) : reel
- NbElementsTot : Entier

Les structures RES_ANA_QUANT et RES_ANA_QUAL permettent de rassembler tous les résultats de l'analyse. Ainsi, on peut faire une fonction qui appelle toutes les autres et retourne une structure de ce type qui comporte la moyenne, l'écart-type, etc.

b. Données Quantitatives

a. Schéma de décomposition :



Fonction *1* :

Paramètre : E nom_Fichier : String ; E i : Entier

Retour : ANA_QUANT

Description : partir du fichier, on va récupérer les données de la colonne i et créer une structure du type ANA_QUANT.

Procédure *2* :

Paramètres : ES maQuantité : ANA_QUANT

Description : trie le tableau de données d'ANA_QUANT dans l'ordre croissant, pour pouvoir calculer les quantiles plus facilement

Fonction *3* :

Paramètres : E maQuantité : ANA_QUANT

Retour : réel

Description : calcul la moyenne de la série statistique

Fonction *4* :

Paramètres : E maQuantité : ANA_QUANT

Retour : réel

Description : Calcul la valeur minimum de la série statistique

Fonction *5* :

Paramètres : E maQuantité : ANA_QUANT

Retour : réel

Description : Calcul la valeur maximum de la série statistique

Fonction *6* :

Paramètres : E maQuantité : ANA_QUANT ; E moyenne : reel

Retour : reel

Description : fonction qui calcule la variance de la série statistique

Fonction *7* :

Paramètres : E variance : reel

Retour : reel

Description : fonction qui calcule l'écart-type à partir de la variance de la série statistique

Procédure *8* :

Paramètres : E maQuantité : ANA_QUANT ; ES mediane : reel ; ES quantile1 : reel ; ES quantile 2 : reel

Description : procédure qui calcule les quantiles de la série statistiques

b. Algorithmes :

Tri croissant d'un tableau

```
Procédure trierTableauCroissant (ES : maQuantite : ANA_QUANT) // On implémente ici le tri à bulles
  Variables
  desordre : booleen
  tmp : reel

  desordre ← vrai
  Tantque desordre = vrai
    Faire
    desordre ← faux //par défaut, on considère que le tableau est rangé
    Pour i ← 1 à (maQuantite.nbElement - 1) pas 1
      Faire
      Si (maQuantite.Données(i) > maQuantite.Données(i+1))
        Faire //l'ordre est mauvais, on inverse les deux
        tmp ← maQuantite.Données(i)
        maQuantite.Données(i) ← maQuantite.Données(i+1)
        maQuantite.Données(i+1) ← tmp
        //on a trouvé deux cases mal ordonnées, le tableau n'est donc pas encore trié
        desordre ← vrai
      Finsi
    FinPour
  FinTantQue
FinProcédure
```

Calcul d'une moyenne

```
Fonction CalculerMoyenne (E : maQuantité : ANA_QUANT) : reel
  Variables
  somme : reel

  somme ← 0
  Pour i ← 1 à maQuantité.nbElements
    Faire
    somme ← somme + maQuantité.Données(i)
  FinPour
  Retourner somme/maQuantité.nbElements
FinFonction
```

Recherche d'un minimum

Fonction calculerMin (E : maQuantité : ANA_QUANT) : reel

Variables

min : reel

min ← maQuantité.Données(1) // le min est initialisé à la première valeur

Pour i ← 2 à maQuantité.nbElements

Faire

Si maQuantité.Données(i) < min

Faire

min ← maQuantité.Données(i)

Finsi

FinPour

Retourner min

FinFonction

Recherche d'un maximum

Fonction calculerMax (E : maQuantité : ANA_QUANT) : reel

Variables

max : reel

max ← maQuantité.Données(1) // le max est initialisé à la première valeur

Pour i ← 2 à maQuantité.nbElements

Faire

Si maQuantité.Données(i) > max

Faire

max ← maQuantité.Données(i)

Finsi

FinPour

Retourner max

FinFonction

Calcul de la variance

Fonction calculerVariance(E : maQuantité : ANA_QUANT ; E : moyenne : reel) : reel

Variables

variance : reel

variance ← 0

Pour i ← 1 à maQuantité.nbElements

Faire

variance ← variance + (maQuantité.Données(i) - moyenne)*(maQuantité.Données(i) - moyenne)

FinPour

Variance ← Variance / maQuantité.nbElements

Retourner Variance

FinFonction

Calcul de l'écart type

Fonction calculerEcartType(E : variance : reel) : reel

Retourner racineCarre(variance)

FinFonction

Calculs des différents quartiles

Procédure CalculerQuartiles (E : maQuantité : ANA_QUANT ; ES : mediane : reel ; ES : quartile1 : reel ; ES : quartile3 : reel)

```
Si (maQuantité.nbElements modulo 2) = 1 // si le nombre d'éléments est impair
    Faire
        mediane ← maQuantité.Données( maQuantité.nbElements/2 + 1) // Le
// tableau étant trié, on peut directement prendre la case du milieu
    Sinon // si le nombre d'éléments est pair
        Faire
            // on fait la moyenne des deux cases du milieu
mediane ← ( maQuantité.Données( maQuantité.nbElements/2 ) + maQuantité.Données(
maQuantité.nbElements/2 + 1 ) ) / 2
    FinSi

// on calcule maintenant les quartiles
Si (maQuantité.nbElements modulo 4) = 0
    Faire
        quartile1 ← maQuantité.Données( maQuantité.nbElements/4 )
        quartile3 ← maQuantité.Données( 0,75*maQuantité.nbElements )
    SinonSi (maQuantité.nbElements modulo 4) = 1
        Faire
            quartile1 ← maQuantité.Données( (maQuantité.nbElements/4) + 1 )
quartile3 ← maQuantité.Données( 0,75*maQuantité.nbElements + 1 )
        SinonSi (maQuantité.nbElements modulo 4) = 2
            Faire
                quartile1 ← maQuantité.Données( (maQuantité.nbElements/4) + 2 )
                quartile3 ← maQuantité.Données( 0,75*maQuantité.nbElements + 2 )
            Sinon // (nbElements modulo 4) = 3
                Faire
                    quartile1 ← maQuantité.Données( (maQuantité.nbElements/4) + 3 )
                    quartile3 ← maQuantité.Données( 0,75*maQuantité.nbElements + 3 )
            FinSi
    FinSi

FinProcédure
```

Fonction qui effectue les différents traitements sur la quantité et renvoie les résultats

Fonction TraitementQuantitatifUnivarié (E : monAnalyseQuantitative : ANA_QUANT) : RES_ANA_QUANT

Variables

mesResultats : RES_ANA_QUANT

mesResultats.Quantité ← monAnalyseQuantitative.Quantité

mesResultats.nbElements ← monAnalyseQuantitative.nbElements

trierTableauCroissant (monAnalyseQuantitative)

mesResultats.Moyenne ← CalculerMoyenne (monAnalyseQuantitative)

mesResultats.Min ← CalculerMin (monAnalyseQuantitative)

mesResultats.Max ← CalculerMax (monAnalyseQuantitative)

mesResultats.variance ← calculerVariance (monAnalyseQuantitative, mesResultats.Moyenne)

mesResultats.ecart_type ← calculerEcartType(mesResultats.variance)

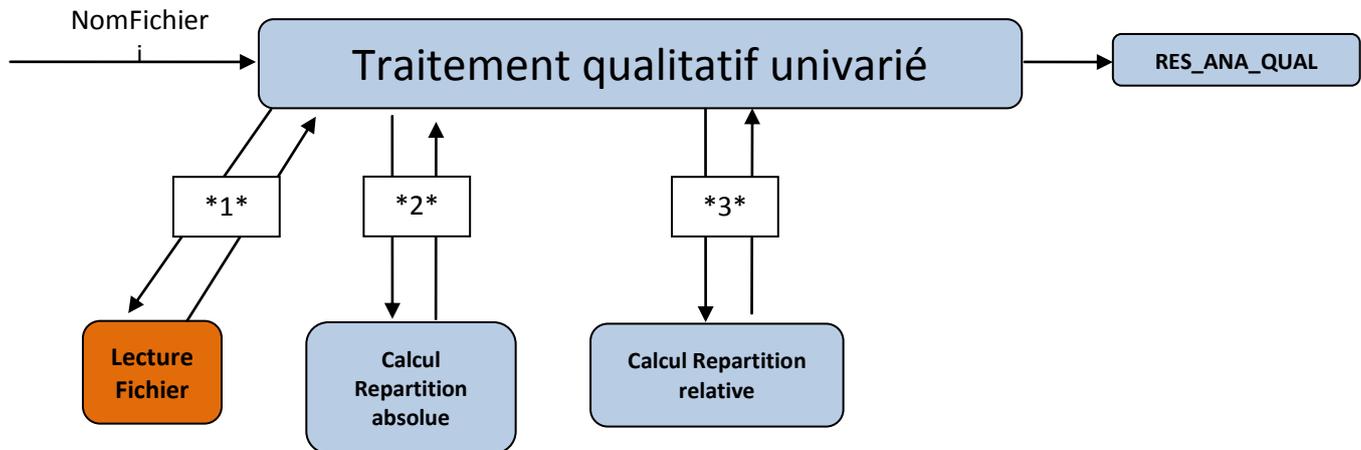
CalculerQuantiles (monAnalyseQuantitative, mesResultats.mediane, mesResultats.Premier_Quartile,
mesResultats.Troisième_Quartile)

Retourner mesResultats

FinFonction

c. Données qualitatives

a. Schéma de décomposition :



Fonction *1* :

Paramètre : E nom_Fichier : String ; E i : Entier ; E j : Entier

Retour : ANA_QUANT

Description : partir du fichier, on va récupérer les données de la colonne i et créer une structure du type ANA_QUANT.

Fonction *2* :

Paramètre : E monAnalyseQualitative : ANA_QUAL

Retour : Tableau (100) : Entier

Description : retourne le tableau de répartition absolue (en nombre) de la série statistique

Fonction *3* :

Paramètre : E monAnalyseQualitative : ANA_QUAL ; E tab : Tableau(100) : Entier

Retour : Tableau (100) : reel

Description : retourne le tableau de répartition relative (en fréquence) de la série statistique

b. Algorithmes :

Calcul de répartition absolue

Fonction CalculRepartitionAbsolue (E monAnalyseQualitative : ANA_QUAL) : tableau(100) : Entier

variables

tab : Tableau(100) : Entier

QualitéTMP : String

Position : Entier

Pour i ← 1 à monAnalyseQuantitative.Qualité.nbEtatsPossibles

Faire

Tab (i) ← 0 // on initialise le tableau

FinPour

Pour i ← 1 à monAnalyseQuantitative.nbElements

Faire

QualitéTMP ← monAnalyseQualitative.Données(i)

// on récupère l'indice de la case dont la valeur doit être incrémentée

Pour j ← 1 à monAnalyseQualitative.Qualité.nbEtatsPossibles

Faire

Si QualitéTMP = monAnalyseQualitative. Qualité.EtatsPossibles(j)

Faire

Position ← j

FinSi

FinPour

Tab(j) ← Tab(j)+1

FinPour

retourner tab

FinFonction

Calcul de la répartition relative

Fonction CalculRepartitionRelative(E monAnalyseQualitative : ANA_QUAL ; E tab : tableau (100) : Entier) :

Tableau (100) : reel

Variables

TabRes : Tableau(100) : reel

Pour i ← 1 à monAnalyseQuantitative.Qualité.nbEtatsPossibles

Faire

TabRes (i) ← tab(i) / monAnalyseQualitative.nbElements

FinPour

Retourner tabRes

FinFonction

Fonction qui effectue les différents traitements sur la qualité et renvoie les résultats

Fonction TraitementQualitatifUnivarié (E monAnalyseQualitative : ANA_QUAL) : RES_ANA_QUAL

Variables

mesResultats : RES_ANA_QUAL

mesResultats.Qualité ← monAnalyseQualitative.Qualité

mesResultats.nbElementsTot ← monAnalyseQualitative.nbElements

mesResultats.Repartition_Absolue ← CalculerRepartitionAbsolue(monAnalyseQualitative)

mesResultats.Repartition_Relative ← CalculerRepartitionRelative(monAnalyseQualitative,
mesResultats.Repartition_Absolue)

retourner mesResultats

FinFonction

Statistiques bivariées : croisement qualitatif * qualitatif

a. Introduction

Nous avons défini une fonction qui rajoute les effectifs marginaux de nos deux caractères au bord du tableau d'effectif.

Concrètement lors de la lecture du Fichier nous allons récupérer un tableau d'effectif comme celui-ci où X_i sont les modalités du premier caractère et où Y_j les modalités du second:

	Y_1		Y_j		Y_l
X_1	$n_{1,1}$				$n_{1,l}$
X_i	$n_{i,1}$		$n_{i,j}$		$n_{i,l}$
X_k	$n_{k,1}$		$n_{k,j}$		$n_{k,l}$

Et lors de l'étape 1, nous allons rajouter une ligne et une colonne dans laquelle figureront les effectifs marginaux comme ci-dessous :

	Y_1		Y_j		Y_l	
X_1	$n_{1,1}$				$n_{1,l}$	$n_{1,\cdot}$
X_i	$n_{i,1}$		$n_{i,j}$		$n_{i,l}$	$n_{i,\cdot}$
X_k	$n_{k,1}$		$n_{k,j}$		$n_{k,l}$	$n_{k,\cdot}$
	$n_{\cdot,1}$		$n_{\cdot,j}$		$n_{\cdot,l}$	card \leftarrow

b. Structures utilisées

Struct ANA_QUAL_X_QUAL

- Qual1 : qualité
- Qual2 : qualité
- NbElement : Integer
- Effectif [Qual1.NbEtatsPossible + 1] x [Qual2.NbEtatsPossible + 1]

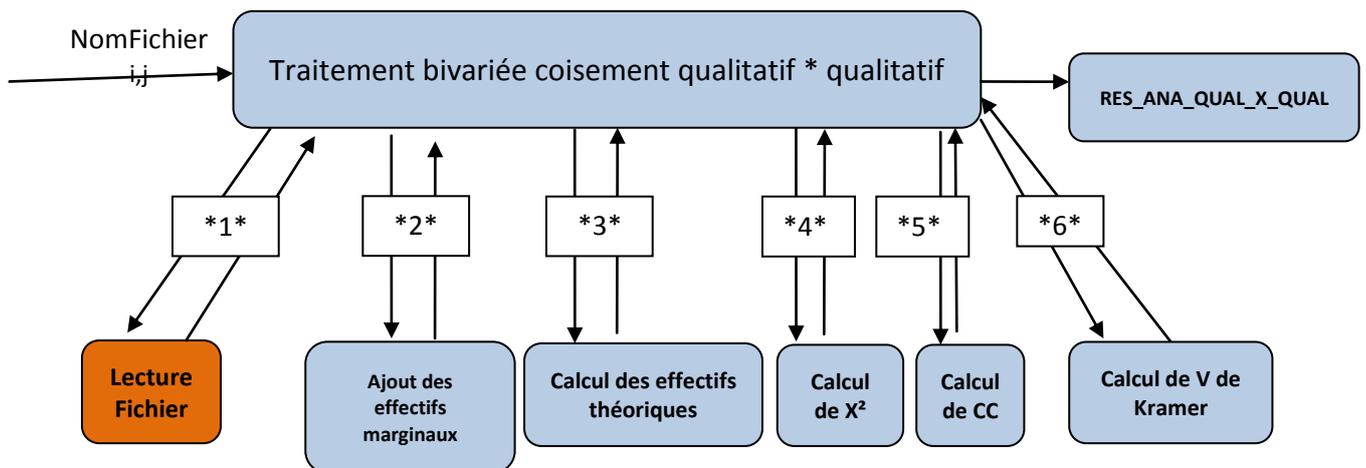
Struct RES_ANA_QUAL_X_QUAL

- Qual1 : qualité
- Qual2 : qualité
- EffectifTheorique [Qual1.NbEtatsPossible+1] x [Qual2.NbEtatsPossible+1]
- X² : float
- CC : float
- V : float

Les deux structures ci-dessus sont construites sur le même principe que les présentées dans les autres sections. Ainsi on retrouve dans la structure ANA_QUAL_X_QUAL regroupe toutes les données nécessaires au traitement bivarié de deux caractères qualitatifs.

RES_ANA_QUAL_X_QUAL regroupe dans une seule entité tous les chiffres qui permettront d'analyser le lien qui existe entre ces deux caractères.

c. Schéma de décomposition



Fonction *1* :

Paramètres : E nomFichier : String, E i : entier, E j : entier

Retour : ANA_QUAL_X_QUAL

Description : A partir du fichier, on va récupérer les données des colonnes i et j et créer une structure du type ANA_QUAL_X_QUAL.

Procédure *2* :

Paramètres : ES monAnaQIQI : ANA_QUAL_X_QUAL

Retour : Tableau(30) : ANA_QUANT

Description : On ajoute une ligne et une colonne qui correspondent aux effectifs marginaux.

Fonction *3* :

Paramètres : E monAnaQIQI : ANA_QUAL_X_QUAL

Retour : Tableau(monAnaQIQI.QUAL1.NbEtatsPossibles + 1, monAnaQIQI.QUAL2.NbEtatsPossibles + 1) : Reel

Description : On calcule les effectifs théoriques de chaque case.

Fonction *4* :

Paramètres : E AnaQIQI : ANA_QUAL_X_QUAL, E tableau t(monAnaQIQI.QUAL1.NbEtatsPossibles + 1, monAnaQIQI.QUAL2.NbEtatsPossibles + 1) : Reel

Retour : Reel

Description : On calcule χ^2 .

Fonction *5* :

Paramètres : E AnaQlQt : ANA_QUAL_X_QUANT,

E Xsi: reel

Retour : Reel

Description : On calcule le coefficient de contingence.

Fonction *6* :

Paramètres : E AnaQlQt : ANA_QUAL_X_QUANT , E Xsi : reel

Retour : reel

Description : On calcule le V de Kramer.

d. Algorithmes

Programme d'ajout des effectifs marginaux

Procédure AjoutEffectifMarginaux(**ES** monAnalyseQualQual : ANA_QUAL_X_QUAL)

Variables

Colonne : Entier

Ligne : Entier

tempMarginal : Reel

Colonne ← monAnalyseQualQual.QUAL2.NbEtatsPossibles

Ligne ← monAnalyseQualQual.QUAL1.NbEtatsPossibles

tempMarginal ← 0

Pour i ← 1 à ligne+1

Faire

tempMarginal ← 0

Pour j ← 1 à colonne+1

Faire

tempMarginal ← tempMarginal + monAnalyseQualQual.Effectif[i][j]

FinPour

monAnalyseQualQual.Effectif[i][colonne+1] = tempMarginal ;

FinPour

tempMarginal ← 0

Pour i ← 1 à colonne+1

Faire

tempMarginal ← 0

Pour i ← 1 à ligne+1

Faire

tempMarginal ← tempMarginal + monAnalyseQualQual.Effectif[i][j]

FinPour

monAnalyseQualQual.Effectif[ligne+1][j] = tempMarginal ;

FinPour

FinProcédure

Fonction de calcul des effectifs théoriques

Fonction EffectifTheorique(**E** monAnalQIQI : ANA_QUAL_X_QUAL) :

Tableau(monANA_QIQI.Qual1.NbEtatsPossible+1, monANA_QIQI.Qual2.NbEtatsPossible+1) : Reel

Variables

Colonne : Entier

Ligne : Entier

Tableau Monresultat (monAnalQIQI.Qual1.NbEtatsPossible+1, monAnalQIQI.Qual2.NbEtatsPossible+1): reel

Pour i ← 1 à Ligne+1

Faire

Pour j ← 1 à Colonne+1

Faire

Monresultat[i][j] ← (monAnalQIQI.Effectif[i][Colonne] * monAnalQIQI.Effectif[Ligne][j]) / monAnalQIQI.Effectif[ligne][colone]

FinPour

FinPour

Retourner Monresultat

FinFonction

Fonction de calcul de X^2

Fonction CalculeXcarre(**E** monAnalyseQualQual : ANA_QUAL_X_QUAL ; **E** Tableau t(monAnaQIQI.QUAL2.NbEtatsPossible + 1):Reel

Variables

Colonne : Entier

Ligne : Entier

X : Reel

Colonne ← monAnalyseQualQual.QUAL2.NbEtatsPossibles

Ligne ← monAnalyseQualQual.QUAL1.NbEtatsPossibles

Pour i ← 1 à Ligne+1

Faire

Pour j ← 1 à Colonne+1

Faire

$X \leftarrow (\text{monAnalyseQualQual.Effectif}[i][j] - t[i][j])^2 / \text{monAnalyseQualQual.Effectif}[i][j]$

FinPour

FinPour

Retourner X

FinFonction

Fonction de calcul de CC

Fonction CalculeCC(**E** monAnalyseQualQual : ANA_QUAL_X_QUAL ; **E** X :Reel) :Reel

Variables

Colonne : Entier

Ligne : Entier

Colonne ← monAnalyseQualQual.QUAL2.NbEtatsPossibles

Ligne ← monAnalyseQualQual.QUAL1.NbEtatsPossibles

$CC \leftarrow \text{Racine}(X^2 / (X^2 + \text{monAnalyseQualQual.Effectif}[Ligne][Colonne]))$

Retourner CC

FinFonction

Fonction de calcul de V

Fonction CalculeV(E monAnalQIQI : ANA_QUAL_X_QUAL ; X : Reel) : V : Reel

Variables

Colonne : entier
Ligne : entier
V : Reel
Min : entier

Min ← 0

Colonne ← monAnalQIQI.QUAL2.NbEtatsPossibles

Ligne ← monAnalQIQI.QUAL1.NbEtatsPossibles

Si (Colonne +1 < Ligne+1)

Min ← Colonne+1

FinSi

Si(Colonne+1 = Ligne+1)

Min ← Colonne+1

FinSi

Si (Ligne+1 < Colonne+1)

Min ← Ligne+1

FinSi

V ← Racine ($X^2 / \text{monAnalQIQI.Effectif[Ligne][Colonne]} * (\text{min} - 1)$)

Retourner V

FinFonction

Statistiques bivariées : croisement qualitatif * quantitatif

a. Structures utilisées :

Struct QUALITE

- NomQualité : String
- NbEtatsPossibles : Entier
- EtatsPossibles : Tableau(30) : String

La structure QUALITE permet de définir un caractère de type qualitatif. Elle possède trois champs : le nom du caractère en question, le nombre d'états qu'il peut prendre, et enfin un tableau qui contient les états qu'il peut prendre.

Struct QUANTITE

- NomQuantité : String
- ValeurMin : Reel
- ValeurMax : Reel

La structure QUANTITE permet de définir un caractère de type quantitatif. Elle possède trois champs : le nom du caractère en question, et les valeurs extrêmes qu'il peut prendre.

Struct DOUBLET_QUAL_QUANT

- ValQualité : String
- ValQuantité : Reel

La structure DOUBLET_QUAL_QUANT permet de réunir ensembles les valeurs de deux caractères différentes d'une seule et même personne. Cela permet de les stocker dans un tableau à une dimension. Donc ici, il s'agit d'un caractère qualitatif et d'un caractère quantitatif

Struct ANA_QUANT

- Quant : Quantité
- NbElements : Entier
- Données : Tableau(1000) : Reel

La structure ANA_QUANT permet de regrouper dans une même entité l'ensemble des informations nécessaires pour effectuer une analyse statistique d'un caractère quantitatif. Elle possède trois champs : la structure quantité, qui précise ce que l'on étudie, le nombre d'individus, et enfin un tableau qui contient les valeurs.

Struct ANA_QUAL_X_QUANT

- Qual : Qualite
- Quant : Quantite
- NbElements : Entier
- Données : Tableau(1000) :
Doublet_Qual_Quant

La structure ANA_QUANT_QUAL_X_QUANT regroupe les informations nécessaires à une analyse bivariée entre un caractère qualitatif et un autre quantitatif. On retrouve 4 champs : la qualité, la quantité, le nombre d'individus étudiés, et en fin un tableau qui contient les valeurs. Pour plus de commodité, les valeurs sont regroupées dans un doublet.

Struct RES_ANA_QUANT

- Quant : Quantité
- Moyenne : Reel
- Mediane : Reel
- Variance : Reel
- Ecart_Type : Reel
- Ecart_Median : Reel
- Min : Reel
- Max : Reel
- Premier_Quartile : Reel
- Troisième_Quartile : Reel
- NbElements : Entier

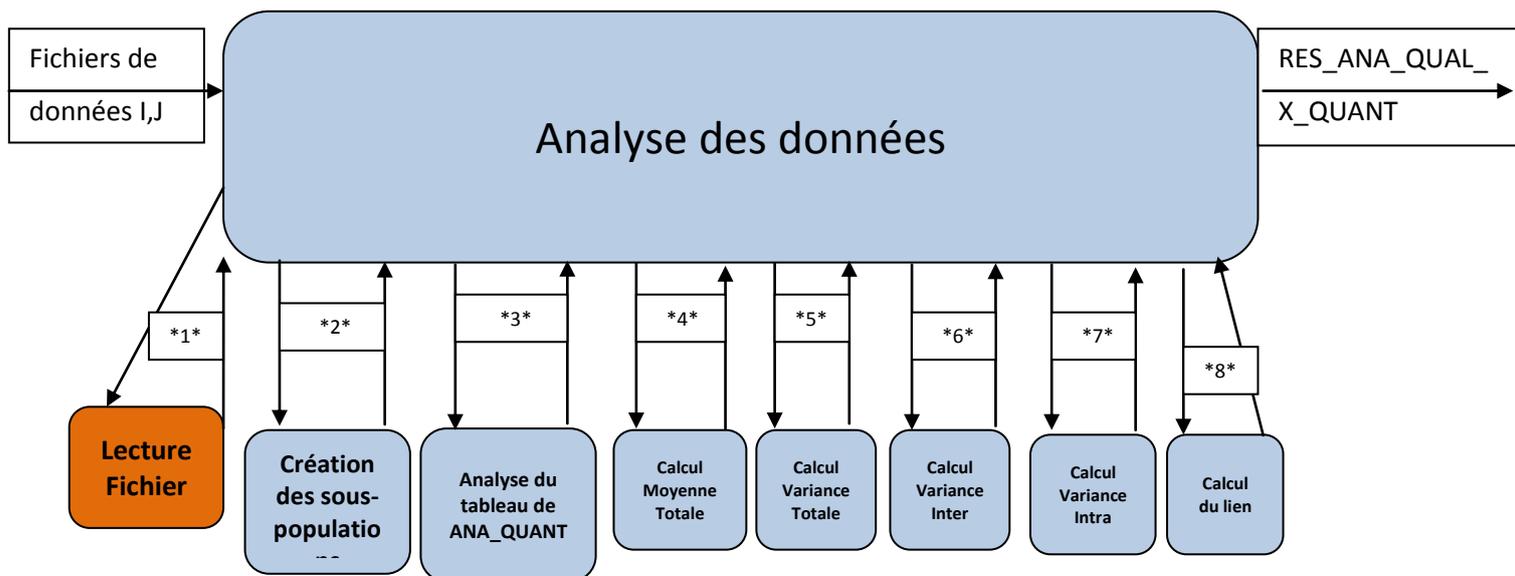
La structure RES_ANA_QUANT permet de stocker dans une même entité l'ensemble des résultats d'une analyse quantitative. On y retrouve la quantité étudiée, le nombre d'individus, puis les outils statistiques : moyenne, médiane, variance, écart type, écart médian, le min, le max, le premier quartile et le troisième quartile.

Struct RES_ANA_QUAL_X_QUANT

- Moyenne_Quant : Reel
- Qual : Qualité
- Quant : Quantité
- Var_{TOT} : Reel
- Var_{INTRA} : Reel
- Var_{INTER} : Reel
- Lien : Reel
- Res_Sous_Pop : Tableau(30) :
RES_ANA_QUANT
- NbElements : Entier

La structure RES_ANA_QUAL_X_QUANT permet de stocker les résultats d'une analyse bivariée entre un caractère qualitatif et un caractère quantitatif. On retrouve en plus des informations sur les caractères, les trois variances qui existent, la moyenne de la quantité, les analyses quantitatives sur les sous populations et bien sur le lien qui existe entre ces deux caractères

b. Schéma de décomposition



Fonction *1* :

Paramètres : *E nomFichier* : String, *E i* : entier, *E j* : entier

Retour : ANA_QUAL_X_QUANT

Description : A partir du fichier, on va récupérer les données des colonnes i et j et créer une structure du type ANA_QUAL_X_QUANT.

Fonction *2* :

Paramètres : *E AnaQlQt* : ANA_QUAL_X_QUANT,

Retour : Tableau(30) : ANA_QUANT

Description : A partir de ANA_QUAL_X_QUANT, on découpe en sous population de même caractère qualitatif et on crée pour chaque sous population une structure ANA_QUANT. On retourne l'ensemble des ANA_QUANT obtenus dans un tableau.

Fonction *3* :

Paramètres : *E Tableau(30)* : ANA_QUANT

Retour : Tableau(30) : RES_ANA_QUANT

Description : A partir du tableau d'ANA_QUANT, on effectue les calculs d'analyse univariée pour un caractère quantitatif pour chaque sous population du tableau. On récupère donc un ensemble de RES_ANA_QUANT que l'on retourne dans un tableau.

Fonction *4* :

Paramètres : *E AnaQlQt* : ANA_QUAL_X_QUANT

Retour : Reel

Description : On calcule la moyenne du caractère quantitatif sur l'ensemble de la population.

Fonction *5* :

Paramètres : E AnaQlQt : ANA_QUAL_X_QUANT,

E Moyenne_{TOTALE} : reel

Retour : Reel

Description : On calcule la variance du caractère quantitatif sur l'ensemble de la population.

Fonction *6* :

Paramètres :

E tableau t1(30): RES_{ANAQUANT}, E NbElementsDuTableauRes : entier, E Moyenne_{totale} : reel, E card : entier

Retour : Reel

Description : On calcule la variance inter.

Fonction *7* :

Paramètres : E tableau t1(30): RES_ANA_QUANT,

E NbElementsDuTableauRes : entier,

E card : entier

Retour : Reel

Description : On calcule la variance intra.

Fonction *8* :

Paramètres : E Variance_{totale} : reel, E Variance_{inter} : reel

Retour : Reel

Description : On calcule le lien entre les deux caractères.

c. Partie Algorithme

Algorithme Découpage en sous-populations

```
Fonction DécoupageSousPopulation (E AnaQlQt : ANA_QUAL_X_QUANT) : tableau(30)
    : ANA_QUANT
variables
    Tableau t(30) : ANA_QUANT {Il s'agit d'un tableau de type ANA_QUANT composé
    des sous-populations}
    i, pos : entier
    QualiteTMP : string {Permet de stocker la valeur de la qualité d'une case}
pour i ← 1 à AnaQlQt.Qual.NbEtatsPossibles pas 1
    t(i).Quant ← AnaQlQt.Quant {Pour chaque sous population, on s'intéresse à la même quantité}
    t(i).NbElements ← 0 {Pour le moment, chaque sous population est vide}
fin pour
pour i ← 1 à ANA_QUAL_X_QUANT.NbElements pas 1 {On va remplir les sous
    populations}
    QualiteTMP ←
    ANA_QUAL_QUANT.Données(i).ValQualité {On récupère la valeur de la qualité}
pour j ← 1 à ANA_QUAL_X_QUANT.Qual.NbEtatsPossibles pas 1
    Si (QualiteTMP = ANA_QUAL_X_QUANT.Qual.EtatsPossibles(j)
        pos =
        j {On récupère la position de la case dans laquelle il faut l'ajouter}
fin pour
    t(pos).NbElements ← t(pos).NbElements + 1 {L'effectif de la sous population
    est incrémenté d'un. }
    t(pos).Données(t(pos).NbElements) ←
    ANA_QUAL_QUANT.Données(i).ValQuantité
fin pour
    retourner t
fin fonction
```

Algorithme Analyse univariée sur les sous-populations

```
Fonction AnalyseUnivariéeSousPop(E tableau t1(30) : ANA_QUANT, E NbEtatsPossibles : entier) : tableau(30) : RES_ANA_QUANT
    variables
        Tableau t(30) : RES_ANA_QUANT {C'est le tableau qui stocke les résultats des analyses}
        i : entier
    pour i ← 1 à NbEtatsPossibles pas 1
        t(i) ← TraitementUnivariéQuantitatif(t1(i)) {Voir la partie une}
    fin pour
    retourner t
fin fonction
```

Algorithme Calcul Moyenne Totale :

```
Fonction CalculMoyenneTotale(E AnaQlQt : ANA_QUAL_X_QUANT) : reel
    variables
        Moyenne : reel
        i : entier
    Moyenne ← 0
    pour i ← 1 à AnaQlQt.NbElements pas 1 {On calcule la somme des valeurs}
        Moyenne ← Moyenne + AnaQlQt.Données(i).ValQuantité
    fin pour
    retourner Moyenne/AnaQlQt.NbElements
fin fonction
```

Algorithme Calcul Variance Totale :

```
Fonction CalculVarianceTotale(E AnaQlQt : ANA_QUAL_X_QUANT, E Moyennetotale : reel) : reel
    variables
        Variance : reel
        i : entier
    Variance ← 0
    pour i ← 1 à AnaQlQt.NbElements pas 1
        Variance ← Variance + (AnaQlQt.Données(i).ValQuantité - Moyennetotale)2
    fin pour
    retourner Variance/AnaQlQt.NbElements
fin fonction
```

Algorithme Calcul Variance Inter

Fonction CalculVarianceInter(**E** tableau t1(30): RES_ANA_QUANT, **E** NbElementsDuTableau : entier, **E** Moyenne_{totale} : reel, **E** card : entier) : reel

variables

Variance : reel

i : entier

Variance ← 0

pour i ← 1 à NbElementsDuTableauRes pas 1

Variance

← Variance + (t1(i).Moyenne – Moyenne_{totale})²
* t1(i).NbElements

fin pour

retourner $\frac{\text{Variance}}{\text{card}}$

fin fonction

Algorithme Calcul Variance Intra

Fonction CalculVarianceInter(**E** tableau t1(30): RES_ANA_QUANT, **E** NbElementsDuTableau : entier, **E** card : entier) : reel

variables

Variance : reel

i : entier

Variance ← 0

pour i ← 1 à NbElementsDuTableauRes pas 1

Variance ← Variance + (t1(i).Variance)² * t1(i).NbElements

fin pour

retourner $\frac{\text{Variance}}{\text{card}}$

fin fonction

Algorithme Calcul du lien

Fonction CalculLien(**E** Variance_{totale} : reel, **E** Variance_{inter} : reel) : reel

variables

retourner $\frac{\text{Variance}_{\text{inter}}}{\text{Variance}_{\text{totale}}}$

fin fonction