# 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](#_Toc244966703)

[Approches de la résolution du problème 3](#_Toc244966704)

[Echantillonage 4](#_Toc244966705)

[a. Schema 4](#_Toc244966706)

[b. Algorithmique 5](#_Toc244966707)

[Statistiques univariées 6](#_Toc244966708)

[a. Structures utilisées : 6](#_Toc244966709)

[b. Données Quantitatives 7](#_Toc244966710)

[a. Schéma de décomposition : 7](#_Toc244966711)

[b. Algorithmes : 8](#_Toc244966712)

[*c.* *Données qualitatives* 12](#_Toc244966713)

[a. Schéma de décomposition : 12](#_Toc244966714)

[b. Algorithmes : 13](#_Toc244966715)

[Statistiques bivariées : croisement qualitatif \* qualitatif 15](#_Toc244966716)

[b. Structures utilisées 16](#_Toc244966717)

[c. Schéma de décomposition 17](#_Toc244966718)

[d. Algorithmes 19](#_Toc244966719)

[Statistiques bivariées : croisement qualitatif \* quantitatif 22](#_Toc244966720)

[a. Structures utilisées : 22](#_Toc244966721)

[b. Schéma de décomposition 24](#_Toc244966722)

[c. Partie Algorithme 26](#_Toc244966723)

# 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 :

1. 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, …

1. 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.

1. 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 traitions 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 :

Fichier de données Analyse des données Résultats statistiques

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.

# Echantillonage

1. Schema

Pré-traitement des données

**Lecture Fichier**

\*1\*

\*2\*

**Echantillonage**

**ANA\_XXX**

NomFichier

i

**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 une échantillonage 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 échatillonage.

1. Algorithmique

##### Echantillonage

**Procédure** Echantillonage (**ES** : monAna : ANA\_XXX, **E**  Nb : Entier) // Pour simplifier la présentation on supposera 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 ← 1

Tant Que ( i <= Nb )

n ← PartieEntiere(alea()\*monAna.NbElements) +1

Present ← Faux

Pour j← 1 à (i-1) pas 1

Si ( n = t2(j) )

Present ← vrai

FinSi

FinPour

Si ( Non Present )

t1(i) ← monAna.Données(n)

t2(i) ← n

i ← i+1

FinSi

Fin Tant Que

**FinProcédure**

# Statistiques univariées

1. Structures utilisées :

Struct ANA\_QUAL

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

Struct ANA\_QUANT

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

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.

1. Données Quantitatives

### a. Schéma de décomposition :

Traitement quantitatif univarié

NomFichier

i

**Lecture Fichier**

\*1\*

\*2\*

**Tri du tableauu**

\*3\*

**Calcul Moyenne**

**Calcul Min**

\*4\*

**Calcul Max**

\*5\*

**Calcul Variance**

\*6\*

**Calcul Ecart-type**

\*7\*

**Calcul quantiles**

\*8\*

**RES\_ANA\_QUANT**

**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

**Procedure** CalculerQuantiles (**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**

**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**

1. Données qualitatives

### a. Schéma de décomposition :

Traitement qualitatif univarié

**Lecture Fichier**

\*1\*

\*2\*

**Calcul Repartition absolue**

**Calcul Repartition relative**

\*3\*

**RES\_ANA\_QUAL**

NomFichier

i

**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(monAnalyseQualtitative, mesResultats.Repartition\_Absolue)

**retourner** mesResultats

**FinFonction**

# Statistiques bivariées : croisement qualitatif \* qualitatif

1. Introduction

Nous avons définit 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 Xi sont les modalités du premier caractère et où Yj les modalités du second:



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



1. 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 structureANA\_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 permettrons d’analyser le lien qui éxiste entre ces deux caractères.

1. Schéma de décomposition

**Fonction \*1\* :**

Traitement bivariée coisement qualitatif \* qualitatif

**Lecture Fichier**

\*1\*

\*2\*

**Ajout des effectifs marginaux**

**Calcul des effectifs théoriques**

\*3\*

**RES\_ANA\_QUAL\_X\_QUAL**

NomFichier

i,j

\*4\*

**Calcul de X²**

\*5\*

**Calcul de CC**

\*6\*

**Calcul de V de Kramer**

Paramètres :

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 :

Tableau(30) : ANA \_QUANT

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

**Fonction \*3\* :**

Paramètres : **E** monAnaQlQl : ANA \_QUAL\_X\_QUAL

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

Description : On calcule les effectifs théoriques de chaque case.

**Fonction \*4\* :**

Paramètres : (monAnaQlQl.QUAL1.NbEtatsPossibles + 1 , monAnaQlQl.QUAL2.NbEtatsPossibles + 1 ) : Reel

Retour : Reel

Description : On calcule .

**Fonction \*5\* :**

Paramètres :

Retour : Reel

Description : On calcule le coefficient de contingence.

**Fonction \*6\* :**

Paramètres :

Retour : reel

Description : On calcule le V de Kramer.

1. Algorithmes

#### Programme d’ajout des effectifs marginaux

**Procedure** 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**

**FinProcedure**

#### Fonction de calcul des effectifs théoriques

**Fonction** EffectifTheorique**(E** monAnalQlQl : ANA\_QUAL\_X\_QUAL ) : Tableau(monANA\_QlQl.Qual1.NbEtatsPossible+1, monANA\_QlQl.Qual2.NbEtatsPossible+1) : Reel

**Variables**

Colonne : Entier

Ligne : Entier

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

**Pour** i 1 **à** Ligne+1 **Faire**

**Pour** j 1 **à** Colonne+1

**Faire**

Monresultat[i][j] (monAnalQlQl.Effectif[i][Colonne]\* monAnalQlQl.Effectif[Ligne][j] ) / monAnalQlQl.Effectif[ligne][colone]

**FinPour**

**FinPour**

**Retourner** Monresultat

**FinFonction**

#### Fonction de calcul de X²

**Fonction** CalculeXcarre**(E** monAnalyseQualQual : ANA\_QUAL\_X\_QUAL ; **E**  Tableau t(monAnaQlQl.QUAL2.NbEtats Possible + 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 (monAnalyseQualQual.Effectif[ i ][ j ] – t[ i ][ j ])² / 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 **Racine**( X² / ( X² + monAnalyseQualQual.Effectif[Ligne][Colonne]))

**Retourner** CC

**FinFonction**

#### Fonction de calcul de V

**Fonction** CalculeV**(E** monAnalQlQl : ANA\_QUAL\_X\_QUAL ; X : Reel) : V : Reel

**Variables**

Colonne  : entier

Ligne  : entier

V  : Reel

Min  : entier

Min0

Colonne monAnalQlQl.QUAL2.NbEtatsPossibles

Ligne monAnalQlQl.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 (X² / monAnalQlQl.Effectif[Ligne][Colonne]\*(min -1))

**Retourner** V

**FinFonction**

# Statistiques bivariées : croisement qualitatif \* quantitatif

1. Structures utilisées :

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 QUALITE

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

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 QUANTITE

* NomQuantité : String
* ValeurMin : Reel
* ValeurMax : 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 DOUBLET\_QUAL\_QUANT

* ValQualité : String
* ValQuantité : 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\_QUANT

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

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.

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\_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

Struct RES\_ANA\_QUAL\_X\_QUANT

* Moyenne\_Quant : Reel
* Qual : Qualité
* Quant : Quantité
* : Reel
* : Reel
* : 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

1. Schéma de décomposition

Analyse des données

RES\_ANA\_QUAL\_X\_QUANT

Fichiers de données I,J

\*5\*

\*4\*

\*3\*

\*7\*

\*6\*

\*2\*

\*8\*

\*1\*

**Lecture Fichier**

**Calcul du lien**

**Calcul Variance Intra**

**Calcul Variance Inter**

**Calcul Variance Totale**

**Calcul Moyenne Totale**

**Création des sous-populations**

**Analyse du tableau de ANA\_QUANT**

**Fonction \*1\* :**

Paramètres :

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 :

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 :

Retour : Reel

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

**Fonction \*5\* :**

Paramètres :

Retour : Reel

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

**Fonction \*6\* :**

Paramètres :

Retour : Reel

Description : On calcule la variance inter.

**Fonction \*7\* :**

Paramètres :

Retour : Reel

Description : On calcule la variance intra.

**Fonction \*8\* :**

Paramètres :

Retour : Reel

Description : On calcule le lien entre les deux caractères.

1. Partie Algorithme

#### Algorithme Découpage en sous-populations

*{Il s’agit d’un tableau de type ANA\_QUANT composé des sous-populations}*

*{Permet de stocker la valeur de la qualité d’une case}*

*{Pour chaque sous population, on s’intéresse à la même quantité}*

*{Pour le moment, chaque sous population est vide}*

*{On va remplir les sous populations}*

*{L’effectif de la sous population est incrémenté d’un. }*

#### Algorithme Analyse univariée sur les sous-populations

*{C’est le tableau qui stocke les résultats des analyses}*

*{Voir la partie une}*

#### Algorithme Calcul Moyenne Totale :

*{On calcule la somme des valeurs}*

#### Algorithme Calcul Variance Totale :

#### Algorithme Calcul Variance Inter

#### Algorithme Calcul Variance Intra

#### Algorithme Calcul du lien