Université Sidi Mohamed Ben Abdellah
Factulté de Sciences et Techniques-Fès
Département de Mathématique

Master Sciences et Techniques MACS
Mathématique et Application au Calcul Scientifique

MEMOIRE DE FIN D'ETUDES
Pour l'Obtention du Diplôme de Master Sciences et Techniques

Modélisation et traitement statistique des données du procédé de séchage et d'emballage de la levure sèche à Lesaffre Maroc

Réalisé par :

- **Salma BRIHI**

Encadré par :

- **Mme. EZZAKI Fatima** (FST-Fès)
- **Mr. EL YAGHMOURI Fouad** (Lesaffre Maroc)

Soutenu Le 15 juin 2016 devant le jury composé de :
- Mme. F. EZZAKI
- Mr. M. ETTAOUIL
- Mme. O. AMMOR
- Mme. A.RAHMOUNI HASSANI

Stage effectué à : « **Lesaffre-Maroc** »

Année Universitaire 2015/2016

Faculté des Sciences et Techniques - Fès
✉B.P. 2202 - Route d'Imouzzer - FES
☎212 (0) 35 60 29 53 FAX : 212 (0) 35 60 82 14
Site web : http ://www.fst-usmba.ac.ma

Remerciements

Avant de commencer la présentation de ce travail, je profite de l'occasion pour remercier toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.

Je tiens à exprimer mes vifs remerciements à Mme. EZZAKI Fatima, d'avoir accepeté d'encadrer mon projet de fin d'études, ainsi que pour son soutien, ses remarques pertinentes et son encouragement.

Je tiens à exprimer ma profonde reconnaissance et toutes mes pensées de gratitude à Mr. Fouad EL YAGHMOURI, qui m'a accompagné de près durant tout ce travail, pour sa disponibilité, pour la confiance qu'il a su m'accorder et les conseils précieux qu'il m'a prodigués tout au long de la réalisation de ce projet. Mes remerciements vont aussi à Mr. Saad EL FIRDAOUSSI et Mr. Ahmed TIMMOUCHE pour leurs soutiens et leurs remarques pendant toute la période du stage, et à toute l'équipe de la société « Lesaffre-Maroc » qui m'a accueilli au début de ce stage.

Je tiens à remercier aussi les membres du jury ; Mr. ETTAOUIL, Mme AMMOR et Mme RAHMOUNI HASSANI de m'avoir honoré en acceptant de juger mon modeste travail.Veuillez trouver ici le témoignage de mon respect le plus profond.

Mes remerciements vont aussi à tous mes professeurs, enseignants et toutes les personnes qui m'ont soutenus jusqu'au bout, et qui n'ont pas cessé de me donner des conseils très importants en signe de reconnaissance.

Dédicaces

Que ce travail témoigne de mes respects :

⇝ A ***mes parents*** :
Grâce à leurs tendres encouragements et leurd grands sacrifices, ils ont pu crée le climat affectueux et propice à la poursuite de mes études.
Aucune dédicace ne pourrait exprimer mon respect, ma considération et mes profonds sentiments envers eux.
Je prie le bon Dieu de les bénir, de veiller sur eux, en espérant qu'ils seront toujours fiers de moi.

⇝ A ***mes frères***.

⇝ A La famille ***BRIHI*** et ***ZELLOU*** :
Ils vont trouver ici l'expression de mes sentiments de respect et de reconnaissance pour le soutien qu'ils n'ont cessé de me porter.

⇝ A tous ***mes professeurs*** :
Leurs générosité et leur soutien m'oblige de leurs témoigner mon profond respect et ma loyale considération.

⇝ A tous ***mes amis et mes collègues*** :
Ils vont trouver ici le témoignage d'une fidélité et d'une amitié infinie.

Salma BRIHI

Table des matières

Remerciements **2**

Dédicaces **3**

Introduction **6**

Présentation de l'entreprise d'accueil « Lesaffre-Maroc » 7

I Historique du groupe Lesaffre **8**

II Lesaffre Maroc **17**
II.1 Organigramme de la société : . . . 19
II.2 Carte d'identité de la société : . . . 19
II.3 La chaîne de production de la levure de boulangerie : . . . 21

Partie théorique 22

Régression linéaire **23**

III Régression linéaire simple : **23**
III.1 Modèle linéaire . . . 23
III.1.1 Définition : . . . 23
III.1.2 Estimation des paramétres β_0 et β_1 . . . 24
III.1.3 Estimation du paramétre σ^2 . . . 24
III.1.4 Inférence concernant la moyenne de la distribution conditionnelle Y à $X = X_h$ 25
III.2 Analyse de la variance . . . 25

IV Régression linéaire multiple **27**
IV.1 Modèle linéaire . . . 27
IV.1.1 Définition : . . . 27
IV.1.2 Estimation des paramétres $\beta_i \quad \forall i = 1, ..., n$ . . . 27
IV.2 Analyse de la variance en régression multiple . . . 28
IV.2.1 Estimation de σ^2 : . . . 28
IV.2.2 Test de signification de la régression dans son ensemble . . . 29
IV.2.3 Contribution marginale et estimation par intervalle . . . 29
IV.2.4 Estimation de $E(Y_h)$ par intervalle de confiance . . . 30

Analyse des données **31**

V Analyse en composantes principales **31**
V.1 Introduction . . . 31
V.2 Histoire . . . 32
V.3 La méthode : . . . 32
V.4 Les données : . . . 32
V.5 Nuage de points : . . . 34
V.6 Centrage et réduction des données : . . . 34
V.7 Détermination du premier axe principal Δ_1 : . . . 35
V.7.1 Première composante principale Y_1 . . . 36
V.7.2 Propriétés de la première composante principale Y_1 : . . . 36
V.8 Deuxième axe principal Δ_2 : . . . 37

V.9 Représentation des variables : 37

VI L'analyse factorielle des correspondances : 37
VI.1 Le tableau de données : 38
VI.1.1 Données : 38
VI.2 Liaison et indépendance entre deux variables qualitatives : 40
VI.3 Comment l'AFC appréhende-t-elle l'écart à l'indépendance ? 40
VI.4 Nuage des lignes et nuage des colonnes 41
VI.4.1 Le nuage des profils lignes N_I 41
VI.4.2 Le nuage des profils colonnes N_J 42
VI.5 Représentation des nuages des lignes et des colonnes 42
VI.6 Interprétation simultanée des lignes et des colonnes 43

Partie pratique 45

VII Position du problème : 46
VII.1 Séchage 46
VII.2 Emballage sous vide : 48
VII.2.1 Le procédé d'emballage sous vide 49

Annex 60

Application de l'AFC 60

Introduction

Après avoir prodigué de nombreuses connaissances théoriques et pratiques, la formation à la FST intègre un stage pour que l'étudiant met en pratique ses connaissances, se rende compte de l'efficacité de son savoir-faire et ait une idée sur le monde du travail pour pouvoir s'intégrer dans la vie active, et s'engager à prendre des responsabilités. Dans cet objectif, nous avons effectué un stage de fin d'études à la société « Lesaffre-Maroc ».

Durant cette période de stage mon travail est porté sur le suivi du procédé de séchage et d'emballage de la levure sèche et puis la modélisation et le traitement statistique des données de ce dernier. Le présent rapport explique la démarche adoptée afin de répondre à l'objectif de ce sujet. Il est structuré en trois parties.

Dans la première partie nous donnons une présentation générale du groupe Lesaffre et de la société d'accueil et du procédé de fabrication.

La seconde partie est consacrée à un rappel mathématique sur la régression linéaire et l'analyse des données.

Enfin, la dernière partie représente l'application de la régression linéaire et l'ACP sur les données du procédé de séchage et d'emballage et nous donnerons un exemple sur l'AFC.

Présentation de l'entreprise d'accueil « Lesaffre-Maroc »

I Historique du groupe Lesaffre

⋆ 1853 : L'association de deux hommes

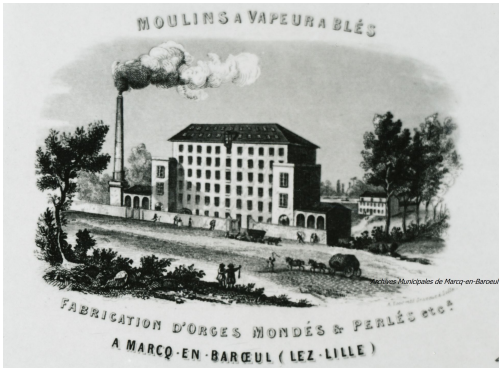
L'histoire de l'entreprise s'est construite autour de deux amis, Louis Lesaffre et Louis Bonduelle. Tout a commencé en 1853, lorsque les deux hommes s'associent pour construire une fabrique d'alcool de gains et de genièvre, à Marquette-lez-Lille dans le nord de la France.

Louis Bonduelle

Louis Lesaffre

⋆ 1863 : L'acquisition du premier moulin à Marcq-en-Barœul

Dix ans après leur association, les deux Louis se portent acquéreurs d'un moulin à Marcq-en-Barœul, qu'ils transforment en une distillerie d'alcool de grains. C'est à partir de ce site que se développera la Société Industrielle Lesaffre qui se révélera progressivement comme l'élément moteur et le support de l'essor industriel et commercial de la levure pour le groupe.



⋆ 1872 : Un nouveau procédé de fermentation

Le baron Max de Springer, rapportant de Vienne (Autriche) l'idée d'extraire de la levure des moûts de fermentation des grains et de la vendre aux boulangers, implante la première levurerie française à Maisons-Alfort. Ces derniers, à cette époque, utilisaient leurs propres levains, accompagnés parfois de levure résiduaire de brasserie.

⋆ 1873 : Début de la fabrication de la levure de boulangerie

Séduits par le procédé de fermentation, Louis Lesaffre et Louis Bonduelle se lancent dans la production de levure fraîche de boulangerie sur la site de Marcq-en-Barœul, à la place de l'ancien moulin.

⋆ 1895 : Naissance de la marque de levure L'hirondelle

Une hirondelle dont le dessin va évoluer au fil du temps, jusqu'à devenir l'emblème de Groupe en 2003

1895

⋆ **1901 : Scission de l'entreprise entre les trois branches familiales**
Les familles Lesaffre et Bonduelle décident de poursuivre séparément leurs activités.
L'entreprise est partagée en 3 entreprises distinctes : Bonduelle, Lesaffre & Cie (alcool et levure) et Lesaffre Frères (sucrerie et distillerie).
Bonduelle est aujourd'hui un acteur sur le marché du légume.

Emile Lesaffre

Louis Bonduelle fils

Louis Lesaffre fils

⋆ **1910 : Grand incendie de Marcq-en-Barœul**
L'usine de Marcq-en-Barœul subit un grand incendie qui la détruit totalment. Elle sera reconstruite par la suite.
⋆ **1923 : Crise de l'alcool de grains**
L'Etat a baissé brutalement le prix de l'alcool de grains qui met en péril l'industrie de la levure.
Pour rester dans la course, Lesaffre réagit en exploitant la mélasse, nouvelle matière première pour la levure. Après la cessation de l'activité « alcool » à Marquette-lez-Lille, les installations de maltage se trouvent vacantes. Il est décidé de les utiliser pour fabriquer des malts de brasserie. Cette usine prend le nom de « Grandes Malteries Modernes ».
⋆ **1930 : Traitement des eaux résiduaires par évaporation**
L'environnement est déjà une préoccupation majeure pour Lesaffre. C'est donc l'une des premières levureries au monde à opter pour la :solution d'évaporation en remplacement de l'épuration des rejets (couteuse et imparfaite). Cette technique induit une fermentation en milieu très concentré qui permet d'obtenir une : : : : qualité irréprochable.

⋆ **1936 Acquisition de la distillerie de pommes de Cérences (future Lesaffre Ingrédients Services)**
Préoccupés par l'approche d'une nouvelle guerre qui semble inéluctable, les dirigeants de Lesaffre cherchent à acheter une nouvelle usine qui leur permettra de continuer à produire en cas d'occupation d'une partie du territoire.
Leurs vues se portent principalement vers le sud (pour les distilleries de vin) et vers l'ouest (distilleries de pommes), car personne n'imagine alors que la totalité du territoire puisse être occupée ... C'est ainsi que l'entreprise acquiert la distllerie de pommes de Gérences, en Normandie.

⋆ **1939 - 1945 : Production de la 1ère levure sèche active**

Lors de la seconde guerre mondiale, la fabrication d'alcool est arrêtée dans les trois distilleries.
L'usine de Marcq met au point des produits à base de levure destinée à atténuer la pénurie alimentaire. Ainsi Lesaffre produit la première levure sèche active en adaptant un séchoir à pâtes alimentaires.
En 1944, les alliés bombardent par erreur les usines de Marcq et Cérances, ainsi que le chantier de la nouvelle distillerie à Buchères

⋆ **1947 : Premier tambour rotatif**

La sortie de la Seconde Guerre mondiale est marquée par d'importants progrès techniques, notamment avec le remplacement des filtres-presses par des filtres rotatifs sous vide. Lesaffre met au point le premier tambour rotatif pour la production de levure sèche en sphérule.

⋆ **1950 : Commercialisation de la production française en Afrique, Moyen-Orient et Asie**

Les exportations de Lesaffre se développent dans les années 50. Les premiers extraits de levure sont proposés sur le marché des ingrédients.

⋆ **1958 : Démarrage du séchage du lait par atomisation chez Lesaffre Nermandie**

Lesaffre Normandie, par son expérience acquise dans le séchage par atomisation des extraits de levure, est devenue un des rares spécialistes européens capables de sécher des produits réputés très difficiles.
La société s'est aussi spécialisée dans le conditionnement d'améliorants de panification.

⋆ **1963 : Partenariat avec le fabricant de levure italien STL**

Dès les années 60, l'internationalisation s'accélère. La société Trentina Lieviti (STL) rejoint le Groupe. C'est aujourd'hui la plus ancienne filiale de Lesaffre à l'étranger.

⋆ **1965 : Création de la division Recherche et Développement (R&D)**

L'entreprise conclut en 1956 un accord avec la Distillers C° Ltd (DCL), premier fabricant mondial de whisky.
Le concours technique de son centre de recherches est précieux ; mais, au bout de quelque années, l'encadrement technique de Lesaffre ayant été considérablement renforcé, l'entreprise peut bénéficier de sa propre organisation.
En 1965, une division Recherche & Développement est mise en place par Jean Goux, ingénieur agroalimentaire nouvellement recruté.

⋆ **1967 : Création de SIL-Fala**

Léon Lesaffre signe un accord qui prévoit la création d'une nouvelle Société : la Société Industrille de Levure Fala (SIL Fala). Celle-ci se construira au Port du Rhin en vue de produire dans un premier temps 18000 tonnes anuelles.

⋆ **1972 : Prise de participation majoritaire dans Fouls Springer**

Après cette prise de participation, Lesaffre acquière ce spécialiste des extraits de levure qui apporte une 4ème actité au Groupe.

⋆ **1973 : Première production de levure sèche instantanée**

Deux ans auparavant, la R&D avait inscrit à son programme de recherches, la conception d'un nouveau type de levure sèche. La levure sèche instantanée Saf-Instant est produite et commercialisée pour la 1ère fois en 1973.

⋆ **1974 : Démarrage du 1er Baking Center à Marcq**
Lesaffre crée son premier Baking Center, véritable fleuron du groupe.
⋆ **1976 : Début de l'activité Améliorants de panification**
Les produits sont d'abord des mixes incorporant de la levure vendus exclusivement à l'export.

⋆ **1981 : Création de Safmex (Mexique)**
Lesaffre construit la levurerue Safmex à Toluca.

⋆ **1985 : Lesaffre au 2e rang mondial**
Lesaffre contrôle 9 usines réparties sur 3 continents, dont six produisent des levures pour un total de 200000 tonnes (équivalent levure pressée) et plaçant le Groupe au 2e rang mondial.
⋆ **1985 : Les début de la nutrtion-santánimal chez Lesaffre**
L'idée de se tourner vers le marché prometteur de la nutririon-santé animale est née au Mexique avec la vente du tout premier produit : la Procréatine 7.

⋆ **1988 - 1989 : Poursuite de l'internationalisation**
Le groupe poursuit son développement commercial en Europe du Sud avec la création de Zymes en Grèce et d'Algist Bruggeman en Belgique (1988), de Lesaffre Far East en Asie (Hong-Kong en 1988) et de Saf Do Brasil en Amérique du Sud (1989).

⋆ **1990 : Fould Springer devient Bio Springer**
A Maisons-Alfort, le site accueille un site de production, une R&D et la commercialisation des extraits de levure.

⋆ **1990 : Nouvelle cuverie à la SIL (Marcq-en-Barœul - France)**
La mise en servise de la nouvelle cuverie de la Société Industrielle Lesaffre lui permet de devenir la plus importante unité de production mondiale de levure.

⋆ **1992 : Safisis**
Un an après avoir initié des travaux d'envergure dans le domaine des molécules aromatiques par la R&D, Lesaffre crée Safisis en partenariat avec l'IFP (Institut Français du Pétrole) pour exlpoiter la plateforme industrielle de biotechnologies de Soustons dans les Landes.
⋆ **1992 : Acquisition d'Ozmaya (Turquie) et DCL (Royaume-Uni)**

⋆**1993 - 1999 : Nouvelles associations et acquisitions**
Lesaffre crée des filiales sur tous les continents : Australie, Argentine, Chili, Etats-Unis, Maroc, Autriche, Suisse, Italie...et développe ses positions en Europe de l'est : Hongrie, Bulgarie, Russie, Ukraine, Pologne, Roumanie, République, Tchèque, Solvaquie, Croatie...

⋆ **1998 : Création de Lesaffre Feed Additives**
13 ans après ses débuts dans la nutrition-santé animale, Lesaffre crée la business unit LEsaffre Feed Additives pour réduire l'utilisation des antibiotiques tout en améliorant les performances et la santé des animaux d'élevage.

LFA LESAFFRE FEED ADDITIVES

⋆ **2000 : Acquisition de la SIL-Fala et implantation de Lesaffre Normandie en Pologne**
Lesaffre achète la fabrication alsacienne de levures et alcools SIL-Fala, mise en marche en 1921. Quand à Lesaffre Normandie, elle implante sa filiale spécialisée dans le séchage par atomisation en pologne. Lesaffre se recentre sur la levure avec la vente de l'activité alcool.

⋆ **2001 : Un développement croissant**

⇝ Acquisition de Red Star Yeast & Product aux Etats-Unis, où Lesaffre devient un acteur de référence sur les marchés de la levure de boulangerie et de sa filiale Astrico en Jordanie.

⇝ Achat de la levurie de Ming Guang en Chine après une période de leasing.

⇝ Prise de majorité dans le capital de Safviet au Vietnam et de Prodese au Brésil.

⇝ Création de Bonopan en Slovénie.

⇝ Céation de Saf Indonusa en Indonésie.

⇝ Création de Lesaffre International, la société de services de Lesaffre.

⋆ **2003 : Un acteur référent sur le plan mondial**
Pour ces 150 ans, Lesaffre est devenu un acteur mondial de référence dans le domaine de la levure de boulangerie, de même pour sa filiale Bio Springer pour les extraits de levure.
Le Groupe construit une nouvelle levurerie à Cedar Rapids (Etats-Unis). Lesaffre possède désormais une trentaine de levurerie dans le monde.

⋆ **2003 : Création de Fermentis**
Lesaffre crée Fermentis pour emmener le groupe vers plus d'innovation dans le domaine des boissons fermentées comme la bière, le vin et les spiritueux.

⋆ **2004 :Première Coupe Louis Lesaffre**
La coupe Louis Lesaffre est née en 2003 pour prendre en charge l'organisation des présélections de la Coupe du Monde de Boulangerie. Elle permet à Lesaffre de se positionner comme le véritable partenaire des boulangers du monde entier.

⋆ **2005 : Lancement de Kastalia**
Lesaffre commercialise la levure liquide réfrigérée « simple, sûre et efficace » qui facilite le travail des boulangers.

⋆ **2006 : Cession de l'activité malt**
Lesaffre se recentre sur la levure avec la vente de l'activité malt.
⋆ **2006 : Création de Lesaffre Human Care**
Fort de plus de 150 ans d'expérience et d'u savoir-faire unique sur le marché de la levure, Lesaffre crée Lesaffre Human Care, une business unit spécialisée dans les ingrédients issus de la levure, de ses dérivés et de bactéries pour investir le domaine de la nutrition-santé humaine.

⋆ **2007 : 2e édition de la Coupe Louis Lesaffre**
Lesaffre organise les sélections nationales et internationales pour la Coupe du monde de la Boulangerie.
⋆ **2007 : Des usines en construction**
Aux Etats-Unis, Lesaffre construit une usine d'extrait de levure dans l'Iowa ainsi qu'une unité de production en Chine.

⋆ **2008 : Partenaire du projet Futurol**
Lesaffre est partenaire de Futurol, un R&D sur le bioéthanol de deuxième génération.

⋆ **2009 : 3^{e} édition de la Coupe Louis Lesaffre**
La 3^{e} édition de la Coupe Louis Lesaffre rassemble les Meilleurs Boulangers du Monde. La Coupe Louis Lesaffre fait le tout du monde dans les plus grandes salons du moment pour sélectionner les équipes les plus talentuauses.

⋆ **2010 : Inauguration des usines aux Etats-Unis et en Chine**
Lesaffre met en service son usine de levure et d'extraits de levure à Laibin dans le Guangxi en Chine ainsi que son usine d'extraits de levure à cedar Rapids dans l'Iowa aux Etats-Unis.

Cedar Rapids - Iowa (USA)

Laibin - Région autonome du Guangxi (Chine)

⋆ **2011 : Lancement de Lynside Pro Gi**
Lesaffre Human Care lance la commercialisation d'une nouvelle pro-biotique destinée aux personnes souffrant d'inconfort intestinal.

⋆ **2012 : Une levurerie à Voronej**
Lesaffre est l'acquéreur d'une unité de production de levure à Voronej en Russie.

⋆ **2013 : 160 ans du Groupe**
Lesaffre fête ses 160 ans à Versailles.

⋆ **2014 : Une croissance dans le domaine de la bio-protection des plantes**
Deus acquisitions illustrent cette ambition de croissance :
– Omniabios en Italie, spécialiste de la purification de molécules
– Agrauxine en France, spécialiste du bio-contôle et de la bio-nutrition des plantes

⋆ **2014 :Création de LEAF Technologies**
Lesaffre met en place une nouvelle business unit LEAF Technologies dédiée au développement et à la vente de solutions de fermentation à forte valeur ajoutée pour les industries du bioéthanol et de la chimie bio-sourcée. Cette dernière acquière la start-up Butalco quelque mois plus tard.

⋆ **2014 : Une nouvelle plateforme technologique de R&D : Lesaffre Puri Center**
Pour répondre aux besoins d'innovation, Lesaffre se dote un second site de R&D dans la Métropole Lilloise : Lesaffre Puri Centre.

⋆ **2015 : 4^{e} Coupe Louis Lesaffre**
La Coupe Louis Lesaffre lance sa 4ème édition avec de nouveaux défis pour les candidats : spécialités revisitées, culture plus internationale du pain et plus adaptée aux tendances de la boulangerie d'aujourd'hui.

Acquisition d'une unité de production de levure en Turquie
Dotée d'une capacité de plus de 80000 tonnes, cette usine située à Lüleburgaz produit de le levure fraîche pour le marché domestique et de la levure sèche pour l'export. Cette acqusition permet à Lesaffre de renforcer sa présence en Turquie et sa stratégie de développement dans cette région du monde (Afrique, Moyen Orient et Asie Centrale).

⋆ **2015 : LFA renaît sous le nom de PHILEO**
Phileo, le business unit Nutrition-santé animal de Lesaffre adopte une nouvelle identité corporate. Ce changement de marque affirme son ambition de devenir un acteur majeur dans le domaine de la nutrition-santé animale.

⋆ **2015 :Acquisition d'une unité de production de levure en Egypte**
Lesaffre a finalisé l'acquisition d'Egybelg, une entité de production de levure en Egypte implantée à Nubaria, ville située à mi-chemin entre Alexandrie et le Caire. Dotée d'une capacité de plus de 45000 tonnes, cette usine produit de la levure fraîche pour le marché domestique et de levure sèche pour le marché domestique et l'export. Sa situation géographique au cœur d'une région betteravière lui donneun avantage compétitif indéniablepoir l'approvisionnement en matière premières.

⋆ **2015 : Lesaffre Zimbabwe, une nouvelle acquisition pour le Groupe**
Lesaffre Anchor Holdinds Private Limited, une société familiale d'investissement, ont signé le 10 septemmbre 2015 un accord pour la création d'une joint-venture au Zimbabwe, qui se matérialise par la création d'une nouvelle société Lesaffre Zimbabwe. Dotée d'une capacité annuelle de 6000 tonnes de levure, l'usine produit de la levure fraîche et de la levure sèche pour le marché local et de proximité. Elle emploie une centaine de personnes. Cette joint-venture permet à Lesaffre de renforcer sa présence sur la région australe de l'Afrique.

⋆ **2015 : Lesaffre prend une participation majoritaire dans l'entreprise italienne Gnosis**
Poursuivant sa stratégie de développement en nutrition et santé, Lesaffre annonce une entrée majoritaire au capital de Gnosis, entreprise italienne spécialisée dans le développement, la production et la vente d'ingrédients de fermentation pour les industries pharmaceutiques, neutraceutiques, cosmétiques et vétérinaires.

2015 : Safisis devient Ennolys
Afin d'affirmer son expertise en production d'ingrédients naturels par biotechnologie, l'entreprise Safisis devient Ennolys. Son développement s'accélérant, cette filiale de Lesaffre a souhaité symboliser ce dynamisme par une image corporate plus en phase avec ses activités d'aujourd'hui et ses ambitions pour demain.

⋆ **2015 : Lesaffre acquière une part majoritaire des titres de deux usines en Chine**
Lesaffre prend une participation majoritaire dans les entreprises chinoises Guangxi Sungain Yeast et Guangxi Sungain Biological Technology.

⋆ **2015 : Une nouvelle identité pour Bio Springer Biotec : Procelys**
Business unit dédiée à la fermentation de nutriments, Procelys fournit l'industrie de la biotechnologie au niveau mondial.

II Lesaffre Maroc

Crée en 1975, la SODERS (Societé des Derivés du Sucre) est depuis 1993 majoritairement détendue par le groupe français Lesaffre. Elle est ainsi devenue la première entreprise privatisée au Maroc. Elle bénéficie de l'expérience et de la maîtrise du leadeur mondial de la fabrication de levure de panification. Basée à Fès, elle emploie 170 personnes avec une superficie de deux hectares qui bénéficie d'une politique salariale attractive et de possibilités de formation continue d'un grand groupe, qui a su conserver les valeurs humaines d'une entreprise familiale.
Lesaffre fabrique et commercialise au Maroc de la levure et des améliorants de panification des marques suivantes :

◇ Jaouda pour la levure fraîche

◇ Rafiaa et Nevada pour le levure sèche, ainsi qu'un type spécial destiné pour saturer les besoins des forces armées royales (FAR) en levure.

◇ Ibis bleus et Magimix pour les améliorant, qui sont des produits qui apportent au consomateur le pain qu'il apprécie que ce soit en terme de volume, de texture et de couleur, d'aspect et couleur de croûte, de conservation et bien sûr le goût. Sa large gamme de produit en fait aujourd'hui le leadeur sur le marché des professionnels.

Bénéficiant de l'expertise et du savoir-faire du groupe Lesaffre, la soders possède un laboratoire d'analyse qui effectue chaque jour de nombreux tests physico-chimiques et bactériologiques. La qualité des levures est ainsi sans cesse évaluée afin d'optimiser leurs performances : force fermentative, pureté, stabilité et résistance par rapport au contexte climatique et il a reçu 2 trophées :

- le trophée du prestige arabe en 1984 à Barcelone.
- le trophée international de la qualité en 1985 à Madrid.

Par ailleurs, le servive qualité de la société Lesaffre Maroc assure un suivi des produits en faisant réaliser quotidiennement des contrôles depuis la réception des matières premières jusqu'à la livraison aux clients, il valide à chaque étape de fabrication la conformité des produits à un cahier de charge très strict.

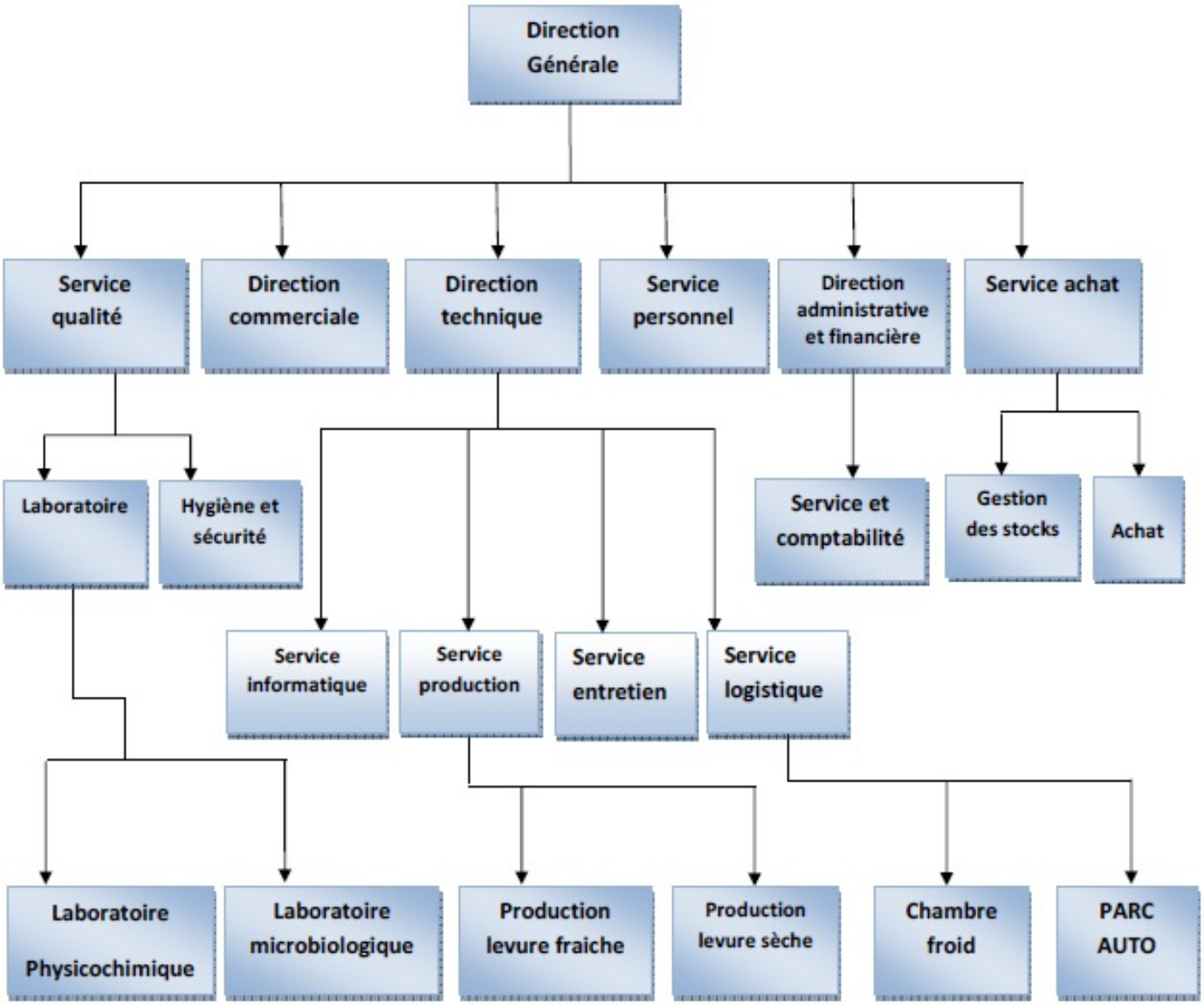
Enfin, une sensibilisation permanente des salariés de l'entreprise aux principes et règlements relatifs à l'hygiène permet de respecter des normes bactériologiques rigoureuses.

Entre 1993 et 2004, l'entreprise a investi 200 millions de dirhams dans la modernisation de ses outils de production.

En 2004, Lesaffre fait l'achat de SNA : société nouvelle de l'alimentation, elle est la spécialiste des produits de pâtisserie au Maroc. Elle commercialise la levure et les améliorants ainsi que toute une gamme de produits de pâtisserie et petit matériel de haute qualité.

En 2006 : création de la nouvelle station de traitement de la mélasse, et aussi d'un nouveau laboratoire moderne très sophistiqué.

II.1 Organigramme de la société :



II.2 Carte d'identité de la société :

Raison sociale: LESAFFRE MAROC
Directeur général: Mr Damien LESAFFRE
Forme juridique : société anonyme
Effectifs : 200 personnes (dont 20 cadres)
Secteur d'activité : L'agroalimentaire
Gamme de produits: Levures de panification et Améliorants...

FIGURE 1 – Carte d'identité de la Sté Lesaffre Maroc

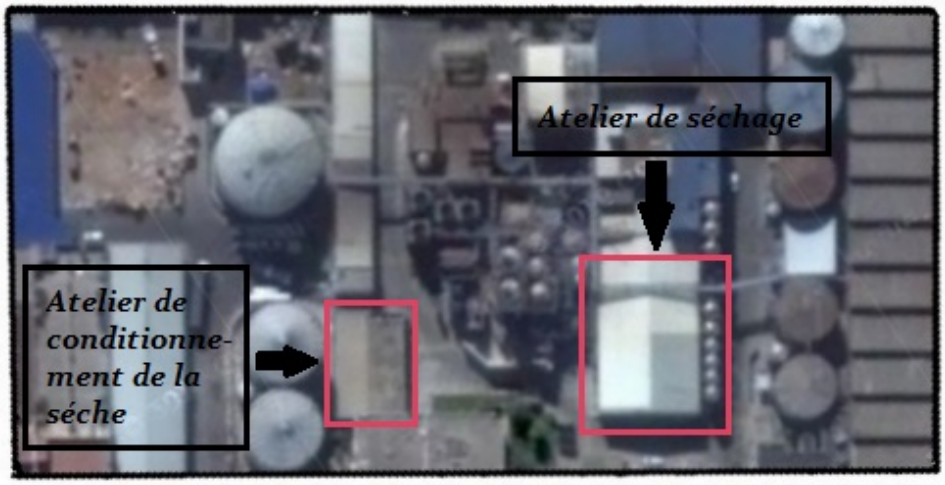


FIGURE 2 – Image satellite montrant l'usine Lesaffre Maroc.

II.3 La chaîne de production de la levure de boulangerie :

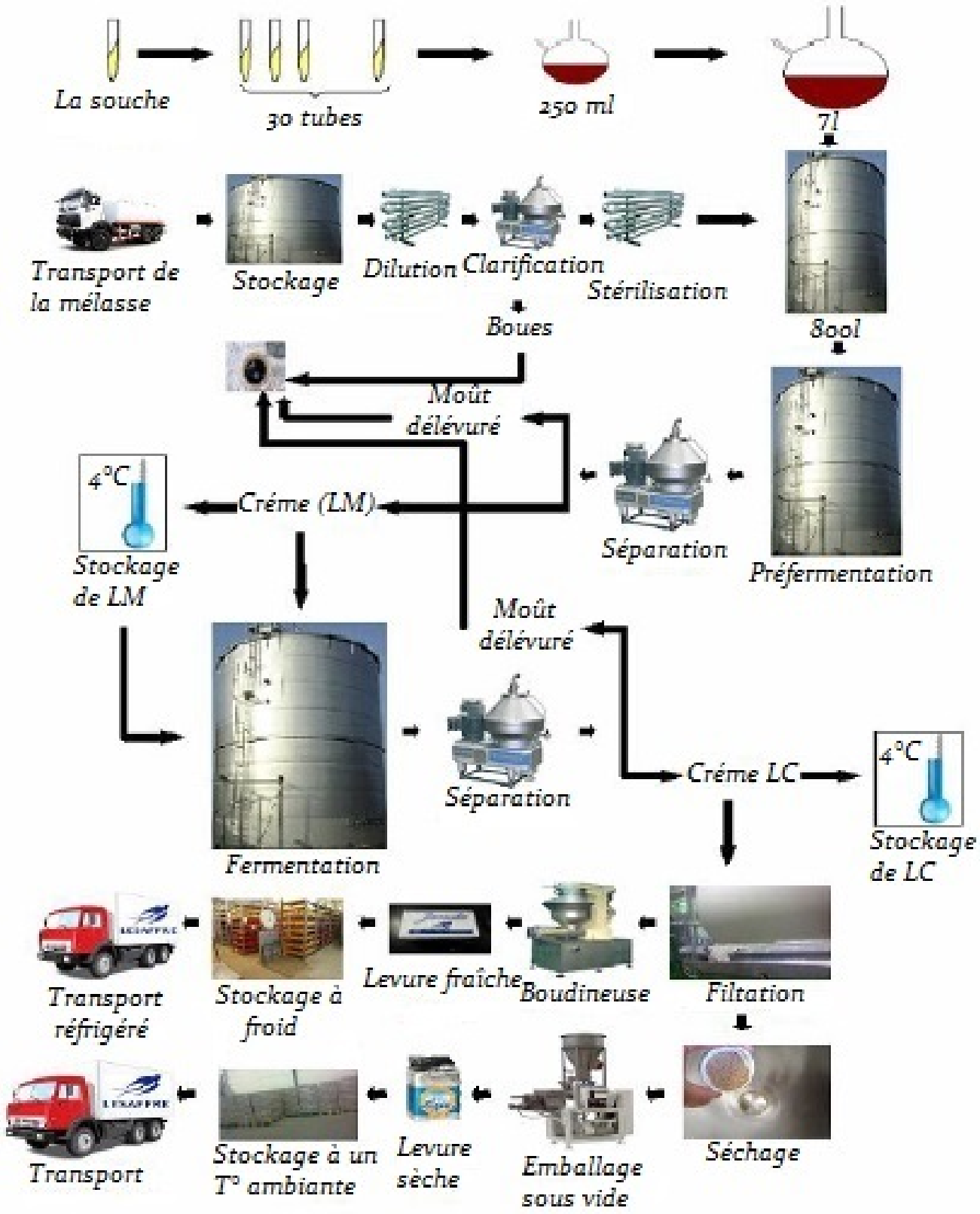


FIGURE 3 – Schéma récaputilatif de la chaîne de fabrication de la levure.

Partie théorique :

Régression linéaire

Introduction :

En statistique un modèle de **régression linéaire** est un modèle de régression d'une variable expliquée sur une ou plusieurs variables explicatives dans lequel on fait l'hypothèse que la fonction qui relie les variables explicatives à la variable expliquée est linéaire dans ses paramètres.
On parle aussi de modèle linéaire ou modèle de régression linéaire.

En général, le modèle linéaire de régression linéaire désigne un modèle dans lequel l'espérance conditionnelle de Y sachant X est une transformation affine de X. Ce pendant, on peut aussi considérer des modèles dans lequels c'est la médiane conditionnelle de Y sachant X ou n'importe quel quantile de la distribution de Y sachant X qui est une transformation affine de X.

Le modèle de régression linéaire est souvent estimé par la méthodes des moindres carrés mais il existe aussi de nombreuses autres méthodes pour estimer ce modèle. On peut par exemple estimer le modèle par maximum de vraisemblance ou encore par inférence bayésienne.

Bien qu'ils soient souvent présentés ensemble, le modèle linéaire et la méthode des moindres carrés ne désignent pas la même chose. Le modèle linéaire désigne une classe de modèles qui peuvent être estimés par un grand nombre de méthodes, et la méthode des moindres carrés désigne une méthode d'estimation. Elle peut être utilisée pour estimer différents types de modèles.

III Régression linéaire simple :

III.1 Modèle linéaire

III.1.1 Définition :

Un modèle de régression linéaire simple est de la forme :

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

où :
- Y est la variable dépendante (une variable aléatoire).
- β_0 et β_1 sont les coefficients (ordonnée à l'origine et pente).
- X est la variable indépendante (variable explicative).
- ε est une erreur aléatoire.

L'éspérance de Y pour chaque X est le point sur la droite d'équation

$$E(Y/X) = \beta_0 + \beta_1 X$$

Conditions d'application du modèle linéaire :

◇ La courbe de régression n'est autre que la courbe joignant les moyennes des distributions des Y_i.
◇ Pour chaque valeur de X, $E(\varepsilon) = 0$ et $V(\varepsilon) = \sigma^2$.
◇ $\varepsilon \sim N(0, \sigma^2)$.
◇ Les erreurs ε sont indépendantes (non corrélées).
◇ Les Y_i ne sont pas corrélées.

On cherche à
- Estimer les paramétres β_0, β_1 et σ^2.
- Vérifier si le modèle est adéquat.

III.1.2 Estimation des paramétres β_0 et β_1

Supposons que n paires d'observations $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ ont été faites. Subsitituant dans le modèle linéaire, on obtient ;

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \Rightarrow \qquad \varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$$

Les coefficients sont déterminés par la méthode des moindres carrés qui minimise la somme des carrés des erreurs :

$$L(b_0, b_1) = Min\left(\sum_{i=1}^{n} e_i^2\right) = \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2$$

On résout le système de deux équations a deux inconnues $\nabla L(b_0, b_1) = 0$.

$$\nabla L(b_0, b_1) = 0 \qquad \Rightarrow \quad \begin{cases} \dfrac{\partial L}{\partial b_0}(b_0, b_1) = \sum\limits_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0, \\ \\ \dfrac{\partial L}{\partial b_1}(b_0, b_1) = \sum\limits_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) = 0. \end{cases}$$

$$\Rightarrow \qquad \begin{cases} b_0 = \bar{Y} - b_1 \bar{X}, \\ \\ b_1 = \dfrac{\sum\limits_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum\limits_{i=1}^{n} X_i^2 - n\bar{X}^2}. \end{cases}$$

avec : $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ et $\bar{Y} = \frac{1}{n}\sum\limits_{i=1}^{n} Y_i$.

Propriété de β_0 et β_1 :
La droite de régression estimée est $\tilde{Y} = b_0 + b_1 X$.
Les variables aléatoires b_0 et b_1 sont des estimateurs de l'ordonnée à l'origine β_0 et de la pente β_1.

Théorème :

▷ $E(b_0) = \beta_0$ et $E(b_1) = \beta_1$ (estimateurs non biaisés).

III.1.3 Estimation du paramétre σ^2

La différence entre la valeur estimée $\tilde{Y}_i = b_0 + b_1 X_i$ et la valeur observée Y_i est appelée *résidu* et est dénotée $E_i = \tilde{Y}_i - Y_i$.
Théorème

◇ $S^2 = \dfrac{\sum\limits_{i=1}^{n}(Y_i - \tilde{Y}_i)^2}{n-2}$ est donc un estimateur sans biais de σ^2.

◇ Si σ est connue on a : $V(b_0) = \sigma^2\left[\dfrac{1}{n} + \dfrac{\bar{X}^2}{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}\right]$ et $V(b_1) = \dfrac{\sigma^2}{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}$.

Si σ est inconnue on a : $S(b_0) = S^2\left[\dfrac{1}{n} + \dfrac{\bar{X}^2}{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}\right]$ et $S(b_1) = \dfrac{S^2}{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}$.

Distributions pour b_0 et b_1

Théorème

◇ Si σ est connue on a : $\dfrac{b_0-\beta_0}{\sigma(b_0)} \backsim N(0,1)$ et $\dfrac{b_1-\beta_1}{\sigma(b_1)} \backsim N(0,1)$

◇ Si σ est inconnue on a : $\dfrac{b_0-\beta_0}{S(b_0)} \backsim T_{n-2}$ et $\dfrac{b_1-\beta_1}{S(b_1)} \backsim T_{n-2}$

Tests d'hypothèse pour β_0
La distribution

$$t_{exp} = \frac{b_0-\beta_0}{S(b_0)} \backsim T_{n-2}$$

permet de tester des hypothèses du type

$$\begin{array}{lll} H_0 & : & \beta_0 = 0 \\ H_1 & : & \beta_0 \neq 0 \end{array}$$

Soient α le seuil de signification du test et b_0 l'estimateur du test.
Sous l'hypothèse H_0 et les conditions d'applications du test :
Régle de décision :

$\rightsquigarrow$ On rejette H_0 si $\dfrac{b_0}{S(b_0)} \notin \left[-t_{\frac{\alpha}{2};n-2}; t_{\frac{\alpha}{2};n-2}\right]$.

$\rightsquigarrow$ Sinon on accepte H_0 c-à-d que β_0 n'est pas significatif dans ce cas le modèle sans constante est $Y = \beta_1 X + \varepsilon$ ($Y_i = \beta_1 X_i + \varepsilon_i$).On a alors :

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}, \qquad S^2 = \frac{\sum_{i=1}^{n} (Y_i - \tilde{Y}_i)^2}{n-1}, \qquad S^2(b_1) = \frac{S^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

III.1.4 Inférence concernant la moyenne de la distribution conditionnelle Y à $X = X_h$

Distribution d'échantillonnage $\tilde{Y}_h$
$\tilde{Y}_h = b_0 + b_1 X_h$ est une distribution normale de moyenne $E(\tilde{Y}_h) = \beta_0 + \beta_1 X_h$ et de variance $V(\tilde{Y}_h) = \sigma^2 \left[\dfrac{1}{n} + \dfrac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right]$.
Une estimation de $\sigma^2(\tilde{Y}_h)$ est : $S^2(\tilde{Y}_h) = S^2 \left[\dfrac{1}{n} + \dfrac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right]$ avec $S^2 = \dfrac{\sum e_i^2}{n-2}$.
Intervalle de confiance de $E(\tilde{Y}_h)$ à $X = \tilde{X}_h$: au seuil $1-\alpha$

$\rightsquigarrow$ Si σ est connue :

$$\tilde{Y}_h - Z_{\frac{\alpha}{2}} \sigma(\tilde{Y}_h) \leq E(\tilde{Y}_h) \leq \tilde{Y}_h + Z_{\frac{\alpha}{2}} \sigma(\tilde{Y}_h)$$

$\rightsquigarrow$ Si σ est inconnue :

$$\tilde{Y}_h - t_{\frac{\alpha}{2};n-2} S(\tilde{Y}_h) \leq E(\tilde{Y}_h) \leq \tilde{Y}_h + t_{\frac{\alpha}{2};n-2}(\tilde{Y}_h)$$

III.2 Analyse de la variance

On détermine dans quelle mesure la droite de régression est utile à expliquer la variabilité existante dans les observations.

Décomposition de la variation :

$$\underbrace{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}_{\text{Var. Totale}} = \underbrace{\sum_{i=1}^{n}(\tilde{Y}_i-\bar{Y})^2}_{\text{Var.expliquée par la droite}} + \underbrace{\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2}_{\text{Var. inexpliquée par la droite}}$$

On définit :

♢ La somme des carrés dûe à l'erreur par :

$$\begin{aligned} SC_{RES} &= \sum_{i=1}^{n} e_i^2 \\ &= \sum_{i=1}^{n}(\tilde{Y}_i - Y_i)^2 \end{aligned}$$

♢ La somme des carrés dûe à la régression par :

$$\begin{aligned} SCR &= \sum_{i=1}^{n}(\tilde{Y}_i-\bar{Y})^2 \\ &= b_1^2\sum_{i=1}^{n}(X_i-\bar{X})^2 \end{aligned}$$

♢ La somme des carrés totale par :

$$\begin{aligned} SCT &= SCR + SC_{RES} \\ &= \sum_{i=1}^{n}(Y_i-\bar{Y})^2 \end{aligned}$$

Le coefficient de détermination :

$$r^2 = \frac{SCR}{SCT} = \frac{\sum_{i=1}^{n}(\tilde{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

Tableau d'analyse de la variance :

Source de variation	Somme des carrés	D.d.l.	Carré Moyen
Régression	SCR	1	$CMR = \frac{SCR}{1}$
Résiduelle	SC_{RES}	$n-2$	$CM_{RES} = \frac{SC_{RES}}{n-2}$
Totale	SCT	$n-1$	

Signification de la régression

Il s'agit de tester les hypothèses :

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

Accepeter H_0 implique que l'on conclut qu'il n'y a pas cause à effet entre X et Y. Ceci peut signifier que

▷ La relation entre X et Y n'est pas linéaire.

▷ La variation de X influe peu ou pas sur la variation de Y.

Au contraire, rejeter H_0 implique que l'on conclut que la variation de X influe sur la variation de Y.

Le critère est : rejeter H_0 au seuil de signification α si $F_{exp} = \dfrac{CMR}{CM_{RES}} \succ F_{\alpha;1;n-2}$.

IV Régression linéaire multiple

IV.1 Modèle linéaire

IV.1.1 Définition :

Un modèle de régression linéaire multiple est de la forme :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$$

où :

◇ Y est la variable dépendante (une variable aléatoire).

◇ β_0, β_1, ... , β_k sont les $k+1$ paramètres du modèle.

◇ X_i représente la $i^{\text{ème}}$ valeur des k variables explicatives. On les considére comme des grandeurs certaines.

◇ ε est le fluctuation aléatoire non observable.

Hypothése fondamentales du modèle de regréssion multiple

1. ε est une v.a de moyenne 0 et de variance constante $= \sigma^2$.
2. Il n'existe aucune corrélation entre les ε_i ; $\quad cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$

◇ $X_1, ..., X_n$ sont des grandeurs certaines.

◇ Les ε_i sont distribuées normalment $\quad \varepsilon_i \backsim N(0, \sigma^2) \quad \forall i$

Remarque : Le nombre de données excéde le nombre de paramètres.

IV.1.2 Estimation des paramétres $\quad \beta_i \quad \forall i = 1, ..., n$

On se donne deux n-échantillons (X_n) et (Y_n) qui ne sont pas mutuellement indépendants où :

◇ (X_n) forme une suite de vecteurs de dimension $p \geqslant 1$,

◇ $X_i = (X_{i1}, ..., X_{ik})$ est la $i^{\text{ème}}$ composante de (X_n).

Le modèle s'écrit alors :

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} + \varepsilon_i \qquad i = 1, ..., n$$

Forme matricielle de régression multiple

$$Y = \beta X + \varepsilon$$

où :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & X_{12} & ... & X_{1k} \\ 1 & X_{21} & X_{22} & ... & X_{2k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & ... & X_{nk} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}.$$

Nous obtenons une estimation de B à l'aide de l'égalité suivante :

$$B = (X^T X)^{-1} X^T Y$$

Où X^T est la matrice transposée de X.
Une estimation de $E(Y_i) = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} V(Y_i) = \sigma^2$ est

$$\tilde{Y_i} = b_0 + b_1 X_{i1} + ... + b_k X_{ik}$$

$V(Y_i) = \sigma^2$
Soit $e_i = Y_i - \tilde{Y_i}$ le résidu
En appliquant la méthode du moindre carrées qui consiste à minimiser la somme des carrès résiduelle.

IV.2 Analyse de la variance en régression multiple

Calcule des sommes de carrés en régression multiple

Source de variation	Somme des carrées	D.d.l.	Carré moyen
Expliqué par la régression	$SCR = \sum_{i=1}^{n} (\tilde{Y_i} - \bar{Y})^2$	k	$CMR = \frac{SCR}{k}$
Résiduelle	$SC_{RES} = \sum_{i=1}^{n} (Y_i - \tilde{Y_i})^2$	$n-k-1$	$CM_{RES} = \frac{SC_{RES}}{n-k-1}$
Totale	$SCT = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$	$n-1$	

IV.2.1 Estimation de σ^2 :

CM_{RES} est une estimation ponctuelle de σ^2

$$CM_{RES} \quad = \quad S^2 \quad = \quad \frac{\sum_{i=1}^{n} (Y_i - \tilde{Y_i})^2}{n-k-1}$$

Le coefficient de détermination multiple R^2 :

$$R^2 \quad = \quad \frac{SCR}{SCT} \quad = \quad \frac{\sum_{i=1}^{n} (\tilde{Y_i} - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

Expressions matricielles pour les sommes des carrés et carrés moyens :

◇ La variation totale : $SCT = Y^TY - n\bar{Y}^2$ avec $n-1$ d.d.l.

◇ Somme des carrées due à la régression : $SCR = B^TX^TY - n\bar{Y}^2$ avec k d.d.l.

◇ Sommes des carrées résiduelle : $SC_{RES} = SCT - SCR = Y^TY - B^TX^TY$ avec $n-k-1$

◇ Le carré due à la régression : $CMR = \dfrac{SCR}{k} = \dfrac{B^TX^TY - n\bar{Y}^2}{k}$

◇ Le carré résiduelle : $CM_{RES} = \dfrac{SC_{RES}}{n-k-1} = \dfrac{Y^TY - B^TX^TY}{n-k-1} = S^2$

IV.2.2 Test de signification de la régression dans son ensemble

Nous voulons tester si la régression est significative dans son ensemble

$$\begin{array}{lll} H_0 & : & \beta_0 = \beta_1 = ... = \beta_k = 0 \\ H_1 & : & \text{au moins un des } \beta_j \text{ est différent de 0} \end{array}$$

$$F_{exp} = \frac{CMR}{CM_{RES}} \backsim \text{Fisher à } k \text{ et } n-k-1 \text{ d.d.l.}$$

Au seuil de signification α , sous l'hypothèse H_0 et les conditions d'applcation du test ;
Régle de décsion :
On rejette H_0 si $F_{exp} \geqslant F_{\alpha;k;n-k-1}$
Si on décidede favoriser H_1, nous concluons que la contribution de l'ensemble des variables pour expliquer les fluctuation de Y est significative.

IV.2.3 Contribution marginale et estimation par intervalle

Test de signification de chaque variable :
Sous les hypothèses de normalité des erreurs, $b_j \backsim N(\beta_j, \sigma^2)$

$$\begin{array}{lll} H_0 & : & \beta_j = 0 \\ H_1 & : & \beta_j \neq 0 \end{array}$$

Soit α le seuil de signification du test, sous H_0 et les conditions d'application du test, on a ;

$$\frac{b_j}{S(b_j)} \backsim T_{n-k-1}$$

Régle de décision :

$$t_{exp} = t_j = \frac{b_j}{s(b_j)}$$

$\rightsquigarrow$ Si $t_{exp} \in \left[-t_{\frac{\alpha}{2};n-k-1}, t_{\frac{\alpha}{2};n-k-1}\right]$, on accepte H_0.

$\rightsquigarrow$ Sinon on rejette H_0 et on accepte H_1, dans ce cas on dit que l'effet de la variable X_j est significatif.

La variance $S^2(b_j)$ s'obtient de la matrice des variances covariances des coefficients de régression ;

$$CM_{RES}.(X^TX)^{-1} = \begin{pmatrix} S^2(b_0) & Cov(b_0,b_1) & ... & Cov(b_0,b_k) \\ Cov(b_0,b_1) & S^2(b_1) & ... & Cov(b_1,b_k) \\ . & & & . \\ . & & & . \\ . & & & . \\ Cov(b_0,b_1) & Cov(b_1,b_k) & ... & S^2(b_k) \end{pmatrix}$$

Intervalle de confiance des β_j : au seuil $1-\alpha$

$$b_j - t_{\frac{\alpha}{2},n-k-1}S(b_j) \leq \beta_j \leq b_j + t_{\frac{\alpha}{2},n-k-1}S(b_j)$$

IV.2.4 Estimation de $E(Y_h)$ par intervalle de confiance

$$\tilde{Y_h} - t_{\frac{\alpha}{2},n-k-1}S(Y_h) \leq E(Y_h) \leq \tilde{Y_h} + t_{\frac{\alpha}{2},n-k-1}S(Y_h)$$

où $S(\tilde{Y_h})$ est l'écart type de $\tilde{Y_h}$

Intervalle de prévision pour Y_h :

$$\tilde{Y_h} - t_{\frac{\alpha}{2},n-k-1}S(d_h) \leq Y_h \leq \tilde{Y_h} + t_{\frac{\alpha}{2},n-k-1}S(d_h)$$

avec $d_h = Y_h - \tilde{Y_h}$ et $S^2(d_h) = S^2 + S^2(\tilde{Y_h}) = S^2\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right]$

Variance $S^2(\tilde{Y_h})$ sous forme matricielle :

$$S^2(\tilde{Y_h}) = (X_h^T(X^TX)^{-1}X_h)CM_{RES} \qquad \text{où} \qquad X_h = \begin{pmatrix} 1 \\ X_{h1} \\ . \\ . \\ . \\ X_{hk} \end{pmatrix}$$

Analyse des données

L'analyse des données est une famille de méthodes statistiques dont les principales caractéristiques sont d'être multidimensionnelles et descriptives. Certaines méthodes, pour la plupart géométriques, aident à faire ressortir les relations pouvant exister entre les différentes données et à en tirer une information statistique qui permet de décrire de façon plus succincte les principales informations contenues dans ces données. D'autres techniques permettent de regrouper les données de façon à faire apparaître clairement ce qui les rend homogénes, et ainsi mieux les connaître.

L'analyse des données permet de traiter un nombre très important de données et de dégager les aspects les plus intéressants de la structure de celles-ci. Le succès de cette discipline dans les dernières années est dû, dans une large mesure, aux représentations graphiques fournies. Ces graphiques peuvent mettre en évidence des relations difficilement saisies par l'analyse directe des données mais surtout, ces représentations ne sont pas liées à une opinion « a priori » sur les lois des phénomènes analysés contrairement aux méthodes de la statistique classique.

Les fondements mathématiques de l'analyse des données ont commencé à se développer au début du XX^{e} siècle, mais ce sont les ordinateurs qui ont rendu cette discipline opérationnelle, et qui en ont permis une utilisation très étendue. Mathématiques et informatique sont ici intimement liées.

Dans l'acception française, la terminologie « analyse des données » désigne un sous-ensemble de ce qui est appelé plus généralement la statistique multivariée. L'analyse des données est un ensemble de techniques descriptives, dont l'outil mathématique majeur est l'algèbre matricielle, et qui s'exprime sans supposer a priori un modèle probabiliste.

Elle comprend l'analyse en composantes principales (ACP), employée pour des données quantitatives, et ses méthodes dérivées : l'analyse factorielle des correspondances (AFC) utilisée sur des données qualitatives et l'analyse factorielle des correspondances multiples (AFCM ou ACM) généralisant la précédente. L'analyse canonique et l'analyse canonique généralisée, qui sont plus des cadres théoriques que des méthodes aisément applicables étendent plusieurs de ces méthodes et vont au-delà des techniques de description. L'Analyse Factorille Multiple est adaptée aux tableaux dans lesquels les variables sont structurées en groupes et peuvent être quantitatives et/ou qualitatives. La classification automatique, l'analyse factorielle discriminante (AFD) ou analyse discriminante permettent d'identifier des groupes homogènes au sein de la population du point de vue de variables étudiées.

La représentation des données multidimensionnelles dans un espace à dimension réduite est le domaine des analyses factorielles, analyse en composantes principales, analyse des correspondances multiples. Ces méthodes permettent de représenter le nuage de points à analyser dans un plan ou dans un espace à trois dimensions, sans trop de perte d'information, et sans hypothèse statistique préalable. En mathématiques, elles exploitent le calcul matriciel et l'analyse des vecteures et des valeurs propres.

V Analyse en composantes principales

V.1 Introduction

Lorsqu'on étudie simultanèment un nombre important de variables quantitatives, comment en faire un graphique global ? La difficulté vient de ce que les individus étudiés ne sont plus représentés dans un plan, mais dans un espace de dimension plus importante (par exemple 4). L'objectif de

l'ACP est de revenir à un espace de dimension réduite (par exemple 2) en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales.

C'est la matrice des variances-covariances (ou celle des corrélations) qui va permettre de réaliser ce résumé pertinent, parce qu'on analyse essentiellement la dispersion des données considérées. De cette matrice, on va extraire, par un procédé mathématique adéquat, les facteurs que l'on recherche, en petit nombre. Ils vont permettre de réaliser les graphiques désirés dans cet espace de petite dimension (le nombre de facteurs retenues), en déformant le moins possible la configuration globale des individus selon l'ensemble des variables initiales (ainsi remplacées par les facteurs).

C'est l'interprétation de ces graphiques qui permettra de comprendre la structure des données analysées. Cette interprétation sera guidée par un certain nombre d'indicateurs numériques et graphiques, applés aides à l'interprétation, qui sont là pour aider l'utilisateur à faire l'interprétation la plus juste et la plus objective possible.

V.2 Histoire

L'ACP prend sa source dans un article de Karl Pearson publié en 1901. Le père du test du χ^2 y prolonge ses travaux dans le domaine de la régression et des corrélations entre plusieurs variable. Pearson utilise ces corrélations non plus pour expliquer une variable à partir des autres (comme en régression), mais pour décrire et résumer l'information contenue dans ces variables.

Encore connue sous le nom de transformée de Karhunen-Loève ou de transformée de Hotelling, l'ACP a été de nouveau développée et formalisée dans les année 1930 par Harold Hotelling. La puissance mathématique de l'économiste et statisticien américain le conduira aussi à développer l'analyse canonique, généralisation des analyses factorielles dont fait partie l'ACP.

Les champs d'application sont aujourd'hui multiples, allant de la biologie à la recherche économique et sociale, et plus récemment le traitement d'images. L'ACP est majoritairement utilisée pour :
- Décrire et visualiser des données ;
- Les décorréler ; la nouvelle base est constituée d'axes qui ne sont pas corrélés entre eux ;
- Les débruiter, en considérant que les axes que l'on décide d'oublier sont des axes bruités.

V.3 La méthode :

On ne décrira ici, que l'une des variantes de cette méthode et qui est de loin la plus employée : *l'analyse en composantes principales normées* . On suppose que les données de départ sont non seulement hétérogènes quant à leur moyenne, mais le sont également quant à leur dispersion et à leur nature (disparité des unités de mesure). Pour ramener chaque variable à un cadre commun de comparabilité, on opérera sur chacune d'elle une transformation linéaire ramenant sa moyenne à zéro et sa variance à l'unité (variable centrée réduite).

V.4 Les données :

Elles se présentent dans un tableau ou matrice à n lignes et p colonnes que l'on notera X. Chacune des n lignes représente un individu et chacune des p colonnes une variable. A l'intersection de la $i^{\text{iéme}}$ ligne et de la $j^{\text{iéme}}$ colonne, on trouve x_{ij} valeur de l'individu i pour la variable j.

Pour une ACP, les variables sont quantitatives : la matrice X est donc constituée de valeurs numériques. Nous avons fait figurer au bas du tableau figure 4, la moyenne et l'écart-type des variables. Avec ces notations ;

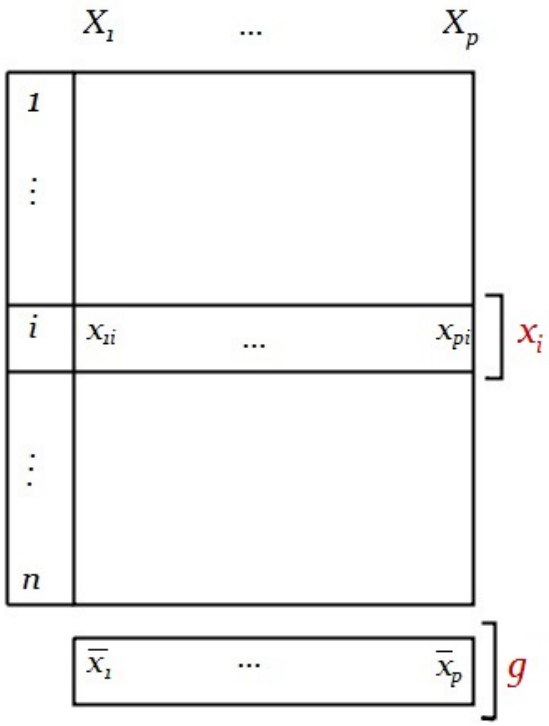


FIGURE 4 – Notations de l'ACP.

$N = \{x_1, ..., x_n\}$: Nuage de points associé aux données.
Centre de gravité du nuage N :

$$g = \frac{1}{n} \sum_{i=0}^{n} x_i$$

La moyenne de la $j^{\text{iéme}}$ variable est :

$$\bar{x_j} = \sum_{i=0}^{n} x_{ij}$$

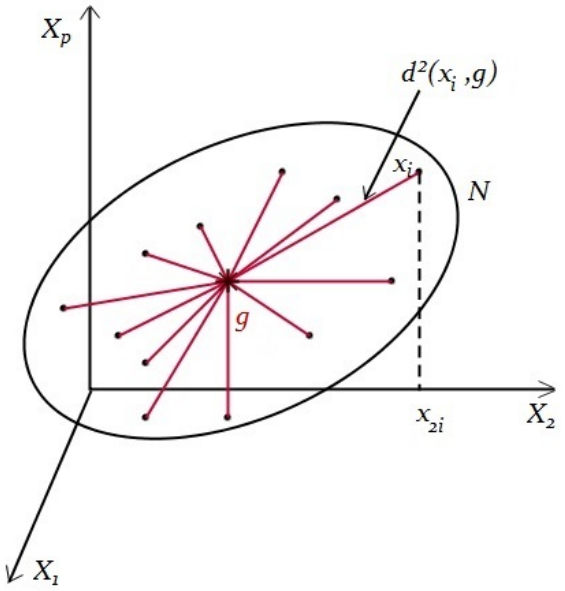
est son écart-type :

$$\sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x_j})^2}$$

V.5 Nuage de points :

Inertie totale :

$$\begin{aligned} I(N,g) &= \frac{1}{n}\sum_{i=1}^{n} d^2(x_i,g) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij}-\bar{x_j})^2 \\ &= \sum_{j=1}^{p}\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x_j})^2 \\ &= \sum_{j=1}^{p}\sigma_j^2 \end{aligned}$$



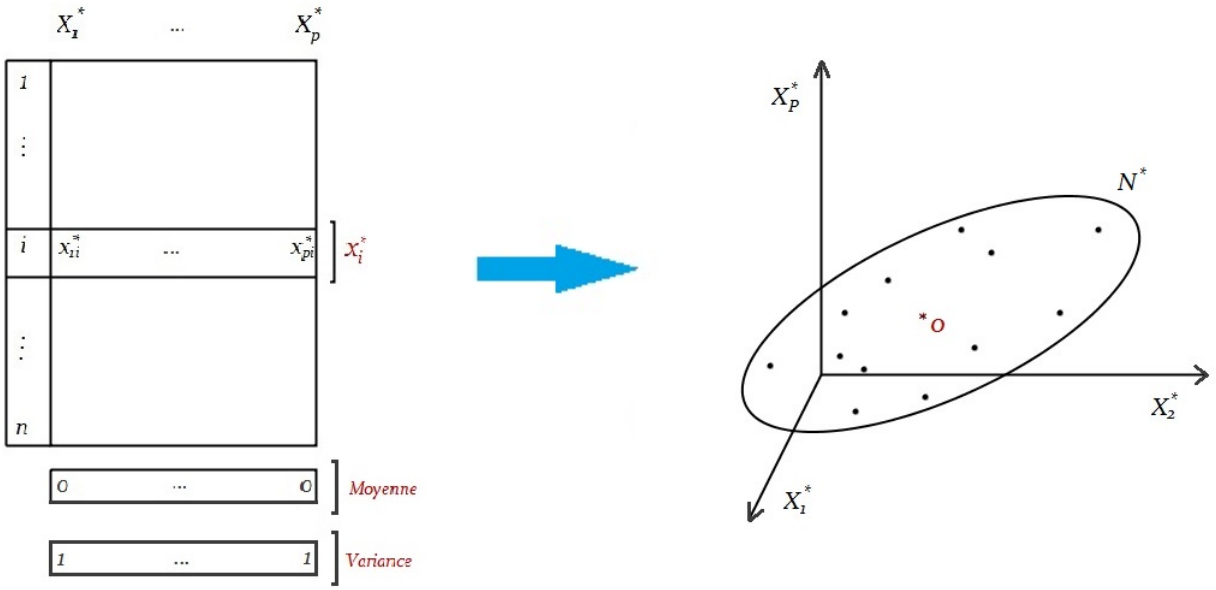
V.6 Centrage et réduction des données :

Pour neutraliser le problème des unités on remplace les données d'origine par les données centrées-réduites :

$X_1^* = \dfrac{X_1 - \bar{x_1}}{\sigma_1}$ de moyenne 0 et d'écart-type 1.

.
.
.

$X_p^* = \dfrac{X_p - \bar{x_1}}{\sigma_p}$



$N^* = \{x_1^*, ..., x_n^*\}$
Centre de gravité : $g^* = 0$, Inertie totale : $I(N^*, 0) = p$

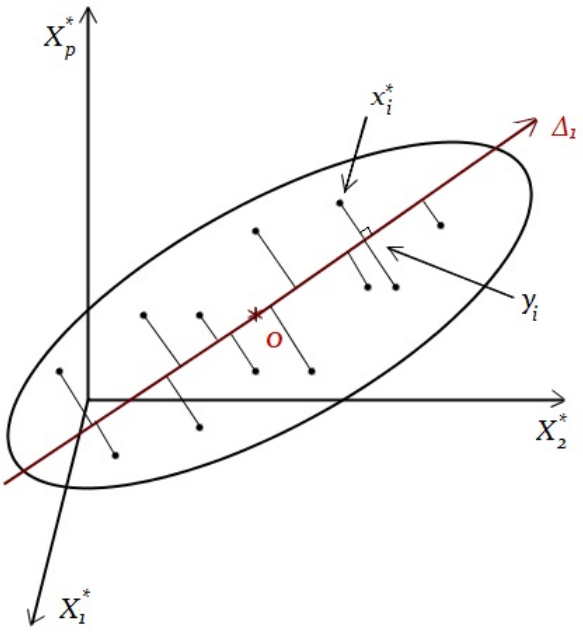
V.7 Détermination du premier axe principal Δ_1 :

On cherche l'axe Δ_1 passant le mieux possible au milieu du nuage N^*.
On cherche à minimiser l'inertie du nuage N^* par rapport à l'axe Δ_1 :

$$I(N^*, \Delta_1) = \frac{1}{n}\sum_{i=1}^{n} d^2(x_i^*, y_i)$$

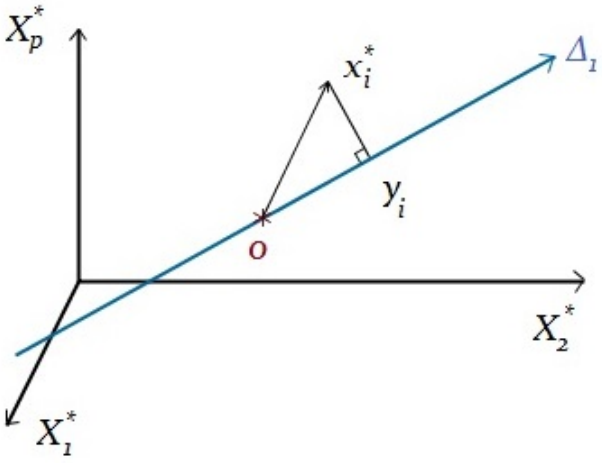
On cherche l'axe d'allongement Δ_1 du nuage N^*.
On cherche à maximiser l'inertie du nuage N^* projeté sur l'axe Δ_1 :

$$I(\{y_1, ..., y_n\}, 0) = \frac{1}{n}\sum_{i=1}^{n} d^2(y_i, 0)$$



De :

$$d^2(x_i^*, 0) = d^2(y_i, 0) + d^2(x_i^*, y_i)$$



On déduit :

$$\frac{1}{n}\sum_{i=1}^{n} d^2(x_i^*, 0) = \frac{1}{n}\sum_{i=1}^{n} d^2(y_i, 0) + \frac{1}{n}\sum_{i=1}^{n} d^2(x_i^*, y_i)$$

Inertie totale $= p$ = Inertie éxpliquee par Δ_1 + Inertie résiduelle

Inertie éxpliquee par Δ_1 : Maximiser

Inertie résiduelle : Minimiser

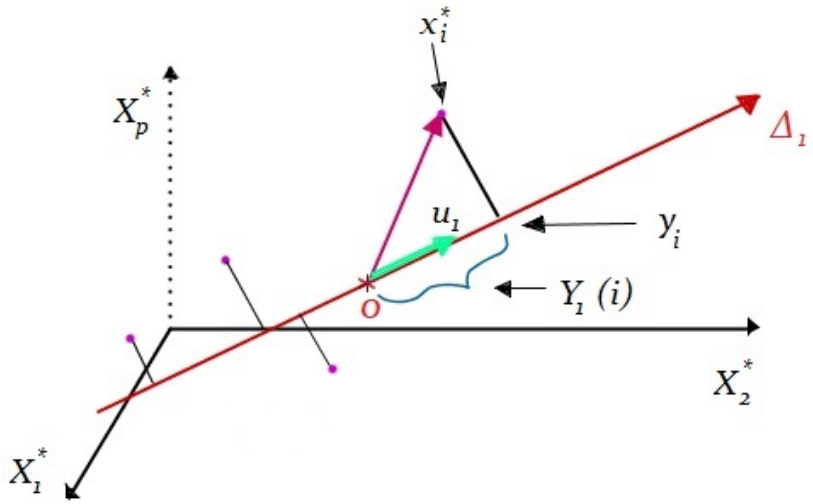
Propriétés de l'axe Δ_1 :

- L'axe Δ_1 passe par le centre de gravité 0 du nuage de points N^*.

- L'axe Δ_1 est engendré par le vecteur normé u_1, **vecteur propre** de la matrice des corrélations R associé à la plus grande valeur propre λ_1.
- L'inertie expliquée par l'axe Δ_1 est égal à λ_1.

– La part d'inertie expliquée par le premier axe principal Δ_1 est égal à $\frac{\lambda_1}{p}$.

V.7.1 Première composante principale Y_1



Y_1 est une nouvelle variable définie pour chaque individu i par :

$$\begin{aligned} Y_1(i) &= \text{longueur algébrique du segment } 0y_i \\ &= \text{coordonnée de } y_i \text{ sur l'axe } \Delta_1 \\ &= \text{produit scalaire entre les vecteurs } x_i^* \text{ et } u_1 \\ &= \boxed{\sum_{j=1}^{p} u_{1j} x_{ji}^*} \qquad \Longrightarrow \qquad \boxed{Y_1 = \sum_{j=1}^{p} u_{1j} X_j^*} \end{aligned}$$

V.7.2 Propriétés de la première composante principale Y_1 :

– $Y_1 = u_{11}X_1^* + u_{12}X_2^* + ... + u_{1p}X_p^*$

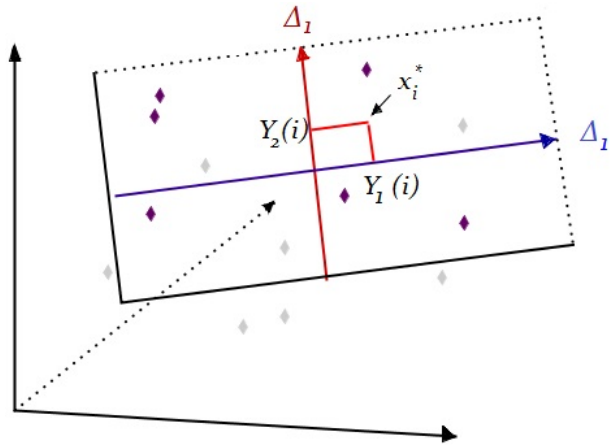
– Moyenne de $Y_1 = 0$

– Variance de Y_1 = Inertie expliquée par $\Delta_1 = \lambda_1$

– $Cor(X_j, Y_1) = \sqrt{\lambda_1} u_{1j}$

– $\frac{1}{p}\sum_{j=1}^{p} cor^2(X_j, Y_1) = \frac{\lambda_1}{p}$ est maximum

V.8 Deuxième axe principal Δ_2 :



- On recherche le deuxième axe principal Δ_2 **orthogonal** à Δ_1 et passant le mieux possible au milieu du nuage.
- Il **passe par le centre de gravié** 0 du nuage de points et est engendré par le vecteur normé u_2, **vecteur propre de la matrice des corrélations R associé à la deuxième plus grande valeur propre** λ_2.
- La deuxième composante principale Y_2 est définie par projection des points sur le deuxième axe principal.
- La deuxième composante principale Y_2 est centrée, de variance λ_2, et non corrélée à la première composante principale Y_1.

V.9 Représentation des variables :

Les « proximités » entre les composantes principales et les variables initiales sont mesurées par les covariances, et surtout les corrélations.

Dans le cas de variables x_i^* réduites ; $\|x_i^*\| = 1$, donc leur projections est à l'intérieur du cercle unité.

$r(\Delta_j, x_i^*)$ est le coefficient de corrélation linéaire entre Δ_j et x_i^*.

Le cercle des corrélations est la projection du nuage des variables sur le plan des composantes principales.

Les variables bien représentées sont celles qui sont proches du cercle, celles qui sont proches de l'origine sont mal représentées.

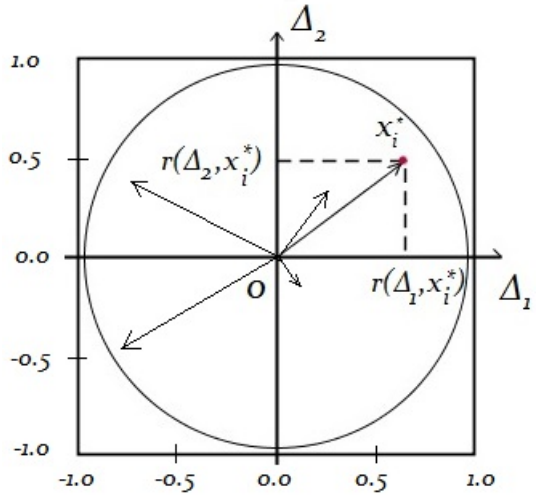


FIGURE 5 – Cercle des corrélations.

VI L'analyse factorielle des correspondances :

L'analyse factorielle des correspondances, notée AFC, est une analyse destinée au traitement des tableaux de données où les valeurs sont positives et homogènes comme les tableaux de contigence.

L'AFC a été introduite de façon complète dans les années 60 par JP BENZECRI. L'AFC est une ACP. les composantes principales sont toujours obtenues à partir de la distance entre les différents points des nuages multidimensionnels, mais les points ont des coordonnées qui ont subis une transformation préalable permettant de conserver une métrique identique à celle de l'ACP pour calculer ces distances.

Le but principal de l'AFC reste donc le même ; lire l'information contenue dans un espace multidimensionnel par une réduction de la dimension de cet espace tout en conservant un maximum de l'information contenu dans l'espace de départ.

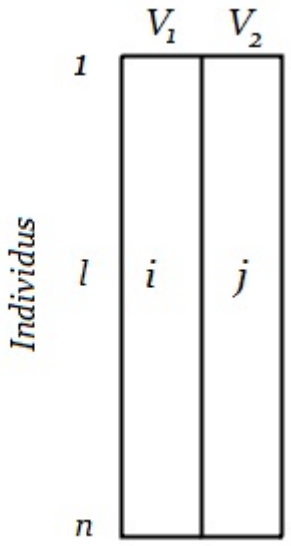
VI.1 Le tableau de données :

L'AFC s'applique essentiellement à des tableaux de contigence. C'est un tableau d'effectifs qui contient à l'intersection de la ligne i et de la colonne j des z_{ij} individus. Il s'agit de la ventilation d'une population totale M selon deux caractères quelconques X en ligne et Y en colonne. Ce sont donc des caractères qualitatifs nominaux et/ou ordinaux.

L'étude traditionnelle d'un tel tableau se concentre le plus souvent sur la dépendance ou l'indépendance entre les deux caractères. Elles s'effectue généralement en utilisant le test du χ^2 et plus particulièrement par l'analyse de la variance (le rapport de corrélation) et la régression lorsque les deux caratères sont qualitatifs ordinaux en classes.

Dans un tableau de contigence, les modalités des caractères sont exclusives les unes par rapport aux autres et exhaustives. Il en résulte que les sommes en ligne et en colonne du tableau ont un sens.

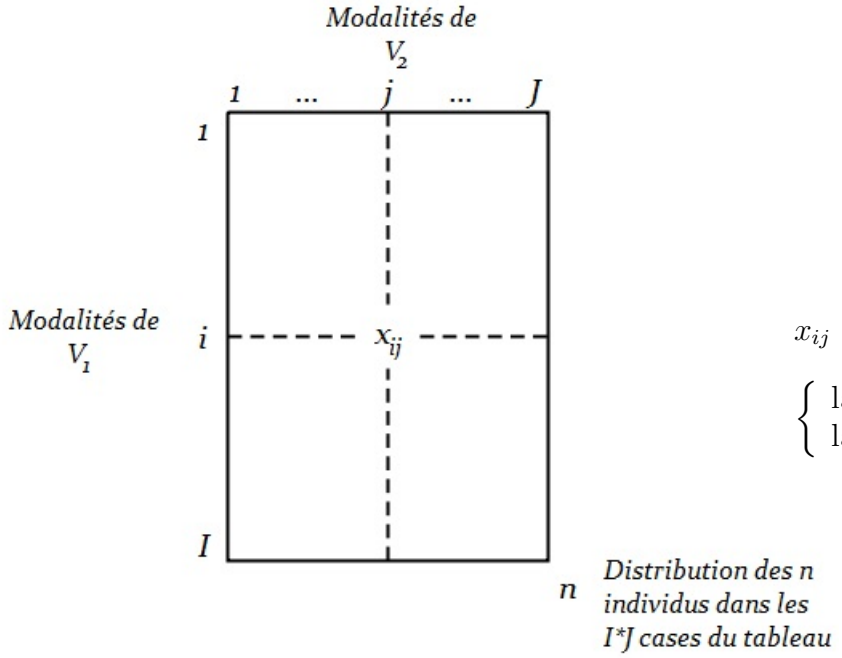
VI.1.1 Données :



n individus
Deux variables qualitatives V_1 et V_2
Variable qualitative et modalité

L'individu l possède :

$\begin{cases} \text{La modalité } i \text{ de } V_1, \\ \text{La modalité } j \text{ de } V_2. \end{cases}$



x_{ij} : nombre d'individus possédant ;

$\begin{cases} \text{la modalité } i \text{ de } V_1, \\ \text{la modalité } j \text{ de } V_2. \end{cases}$

FIGURE 6 – Tableau de contingence

Du tableau de contingence au tableau des probabilités

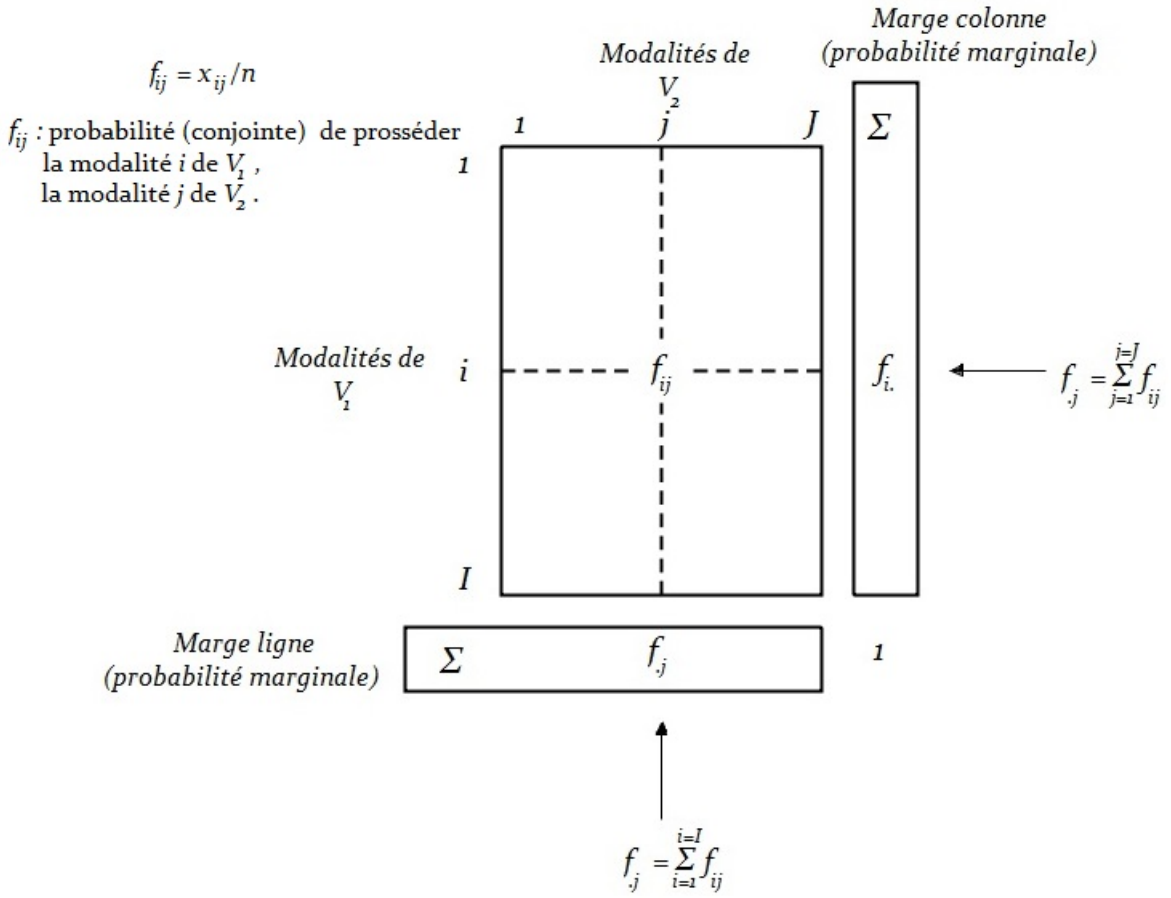


FIGURE 7 – Tableau des probabilité

Liaison entre V_1 et V_2 : écart entre les données observées et le modèle d'indépendance.

VI.2 Liaison et indépendance entre deux variables qualitatives :

Modèle d'indépendance
On dit que deus événements A et B sont indépendants si ; $P[A \text{ et } B] = P[A][B]$
Variables qualitatives indépendantes : $f_{ij} = f_{i.}f_{.j} \qquad \forall i, j$
Liaison entre deux variables qualitatives
Ecart entre les données observées (f_{ij}) et le modèle d'indépendance ($f_{i.}f_{.j}$)
$\rightsquigarrow$ Significativité de la liaison = significativité de cet écart : test du χ^2

$$\chi^2_{obs} = \sum_{i,j} \frac{(\text{effectif observé - effectif théorique})^2}{\text{effectif théorique}}$$

Avec : effectif théorique $= nf_{i.}f_{.j}$

$$\chi^2_{obs} = \sum_{i,j} \frac{(\text{probabilité observée - effectif théorique})^2}{\text{probabilité théorique}} = n\phi^2$$

Intensité de la liaison $= \phi^2$
Nature de la liaison = association entre modalités

L'AFC travaille sur le tableau des probabilités. Elle ne dit rien sur la significativité. Elle visualise la nature de la liaison entre les deux variables.

VI.3 Comment l'AFC appréhende-t-elle l'écart à l'indépendance ?

$\rightsquigarrow$ **Analyse par lignes** $\dfrac{f_{ij}}{f_{i.}} = f_{.j}$

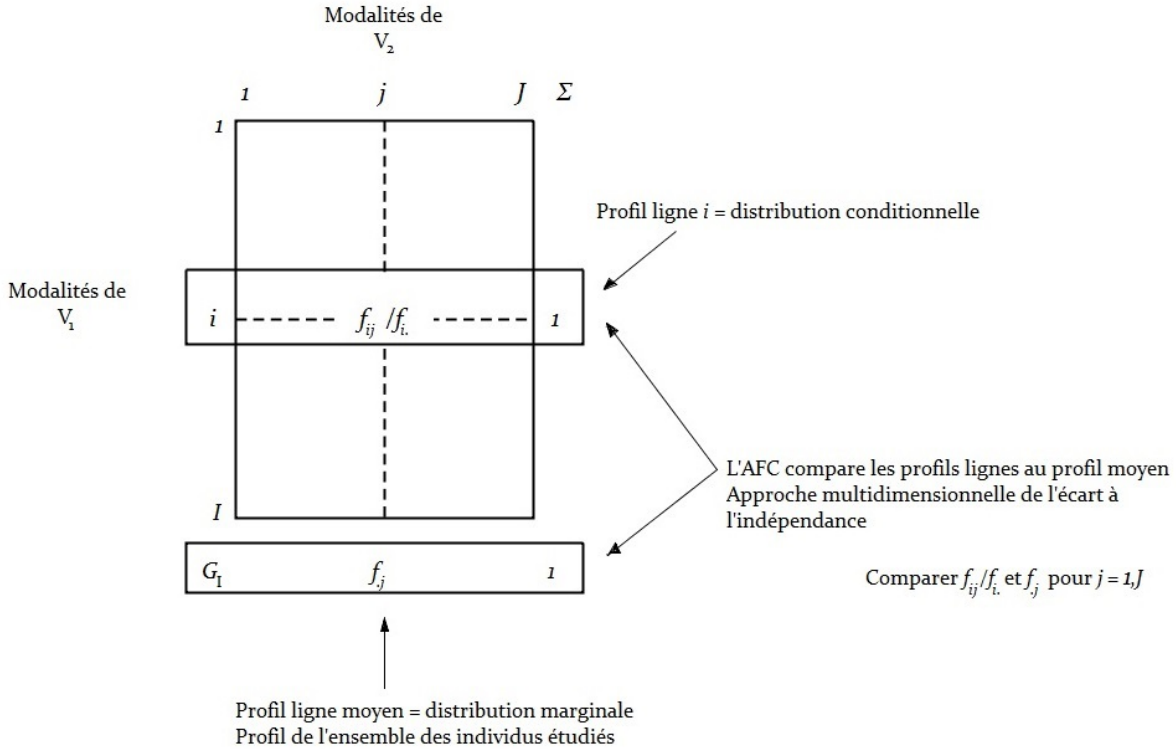


FIGURE 8 – Analyse par ligne

$\rightsquigarrow$ **Analyse par colonnes** $\dfrac{f_{ij}}{f_{.j}} = f_{i.}$

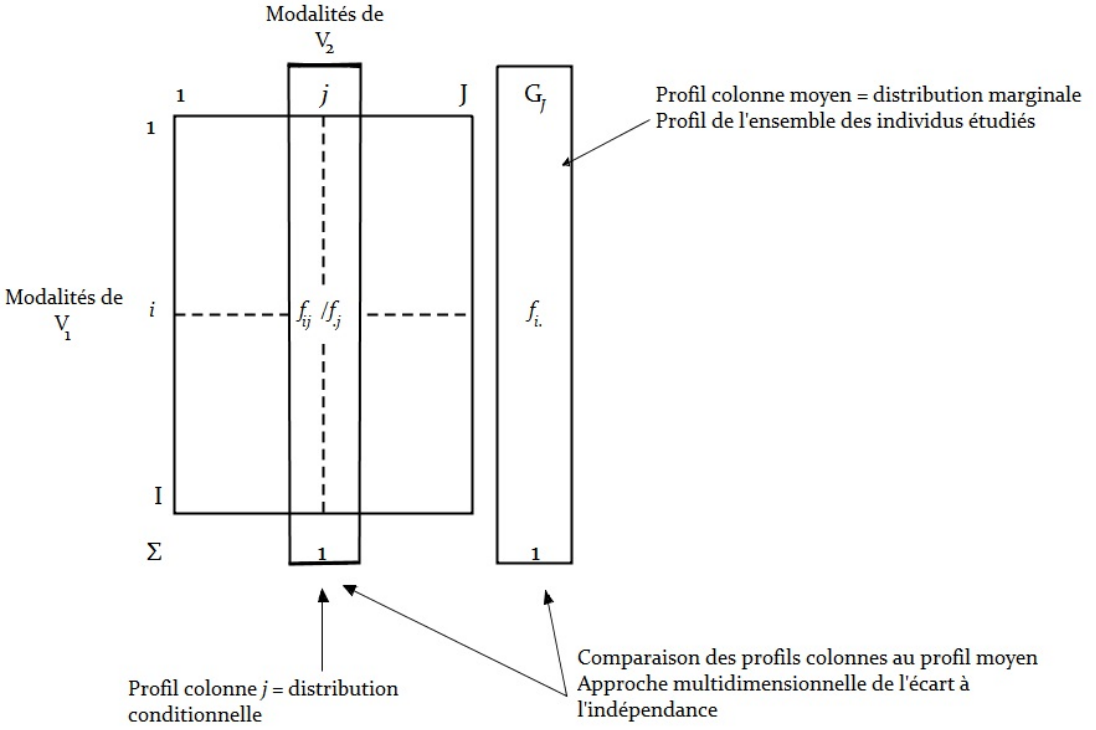
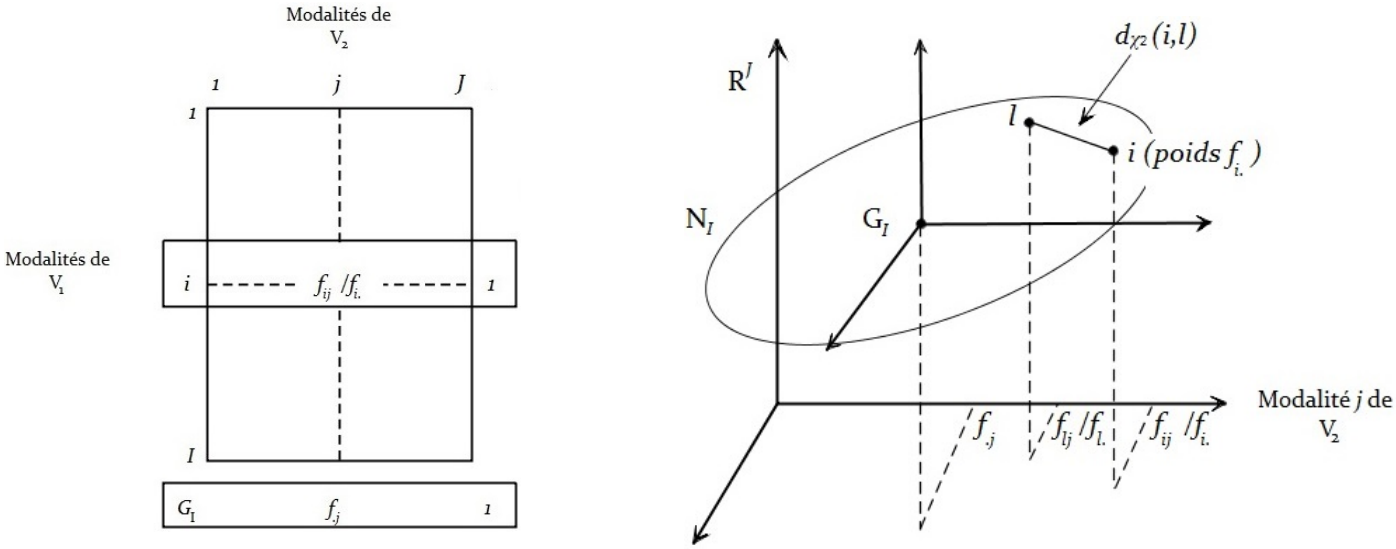


FIGURE 9 – Analyse par colonne

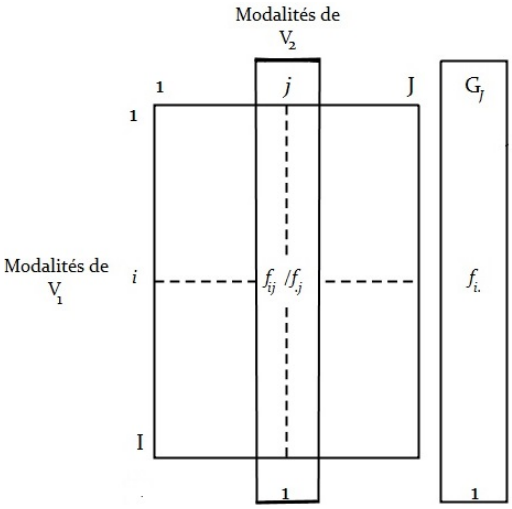
VI.4 Nuage des lignes et nuage des colonnes

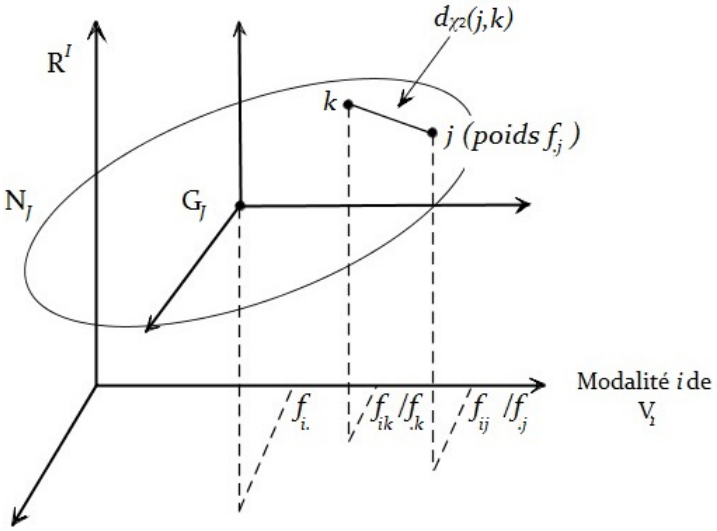
VI.4.1 Le nuage des profils lignes N_I



$$d^2_{\chi^2}(i,l) = \sum_{j=1}^{J} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{lj}}{f_{l.}} \right)^2$$

VI.4.2 Le nuage des profils colonnes N_J





$$d^2_{\chi^2}(j,k) = \sum_{i=1}^{I} \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ik}}{f_{.k}} \right)^2$$

Que se passe-t-il s'il y a indépendance ? ($f_{ij} = f_{.j} f_{i.}$)

▷ Les profils sont confondus avec le profil moyen.

▷ Chacun des deux nuages est réduit à un point : l'origine des axes.

▷ Plus les données s'écartent de l'indépendance et plus les profils s'écartent de l'origine.

$$\begin{array}{rcccl} \text{Inertie } (N_I/G_I) & = & \sum\limits_{i\in I} \text{Inertie } (i/G_I) & = & \sum\limits_{i\in I} f_{i.} d^2_{\chi^2}(i, G_I) \\ & = & \sum\limits_{i\in I}\sum\limits_{j\in J} \dfrac{1}{f_{i.} f_{.j}} \left[f_{ij} - f_{i.} f_{.j}\right]^2 & = & \phi^2 \\ & = & \dfrac{\chi^2}{n} & & \end{array}$$

Etudier l'inertie de N_I revient à étudier l'écart à l'indépendance.

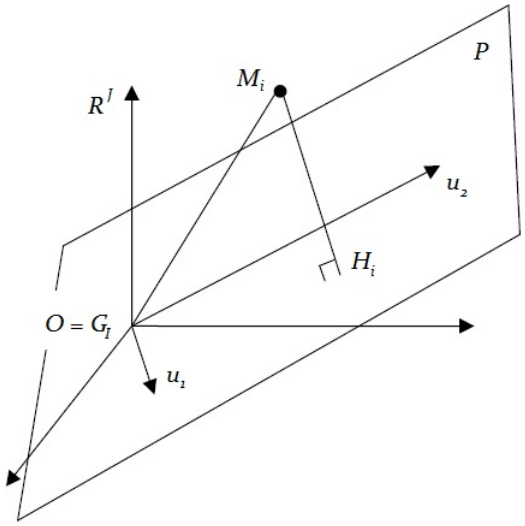
De même pour N_J :

$$\text{Inertie } (N_I/G_I) = \text{Inertie}(N_J/G_J) \quad \text{(dualité)}$$

En AFC, les lignes et les colennes jouent des rôles symétriques.

VI.5 Représentation des nuages des lignes et des colonnes

On décompose de l'inertie de N_I par analyse factorielle, et on projecte le nuage N_I sur une suit d'axes orthogonaux d'inertie maximum.
Meilleure représentation plane de N_I



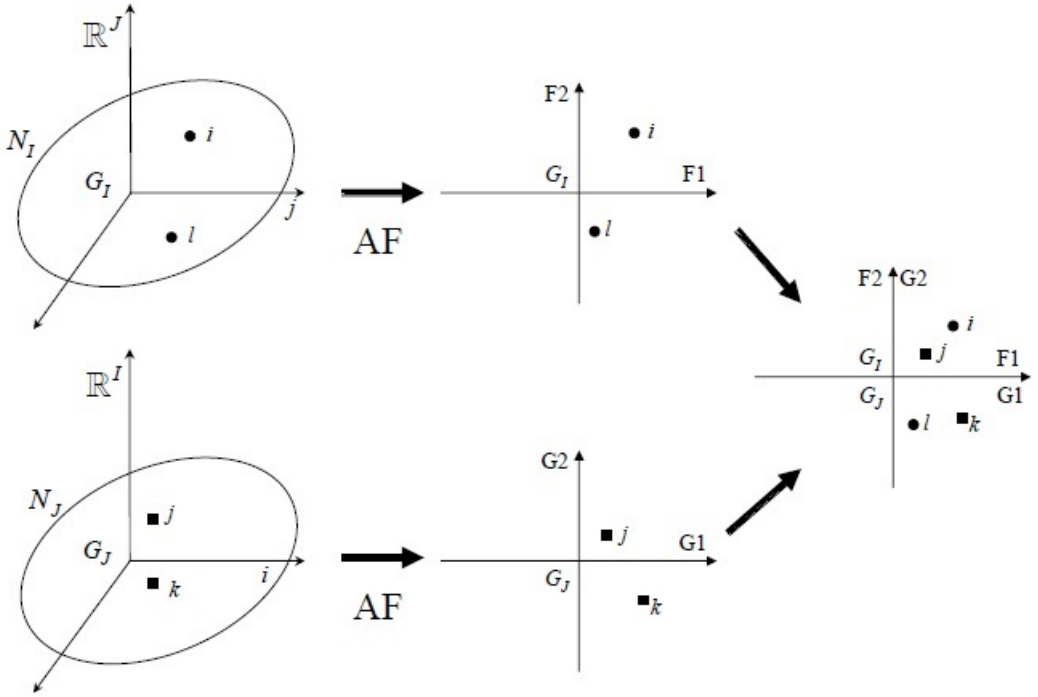
Trouver P tel que : $\underbrace{\sum_i \overbrace{f_{i.}OH_i^2}^{\text{inertie projectée de } i}}_{\text{inertie projetée de} N_I}$ maximum

Plan d'inertie maximum

$$\begin{cases} u_1 \text{ axe d'inertie maximum,} \\ u_2 \text{ axe d'inertie maximum avec } u_2 \perp u_1. \end{cases}$$

De même chose pour le nuage N_J.

Représentation simultanée de N_I et de N_J



VI.6 Interprétation simultanée des lignes et des colonnes

Relation de transition (= propriétés barycentriques)

$$F_s(i) = C_s \sum_j \frac{f_{ij}}{f_{i.}} G_s(j)$$

Oú :

⋆ $F_s(i)$: Coordonnée de la ligne i le long de l'axe de rang s ;

⋆ C_s : Constante associée à l'axe de rang s (en AFC : $1 \leq C_s$) ;

⋆ $\frac{f_{ij}}{f_{i.}}$: $j^{\text{éme}}$ élément du profil i ;

⋆ $G_s(j)$: Coordonnée de la colonne j le long de l'axe de rang s.

Le long de l'axe de rang s, la ligne i est au barycentre de toutes les colonnes, chaque colonne j étant affectée du poids $\frac{f_{ij}}{f_{i.}}$, $j^{\text{éme}}$ terme du profil lignr i.
Concrètement : une ligne est du côté des colonnes auxquelles elle s'associe le plus.
En permutant les rôles joués par les lignes et les colonnes, on obtient la seconde propriété barycentrique ; (une colonne est du côté des lignes auxquelles ellle s'associe le plus.)
clc : Même sur les données de petite taille, l'AFC appotre une visualisation synthétique de l'écart à l'indépendance qui aide la compréhension du tableau. A fortiori avec des grands tableaux.

Conclusion :

Dans cette partie on a donné un petit rappel sur la régression linéaire simple et multiple et puis on a passé à l'analyse des données qui comprend plusieur types d'analyse et parmis eux on a présenté l'analyse en composantes principales et l'analyse factorielle des correspondances.

Partie pratique :

VII Position du problème :

Lesaffre Maroc fait un grand effort pour qu'il puisse offrir ses produits dans un meilleur état sur le marché, dont le but de gagner la confiance de ses clients, pour cela l'équipe de l'atelier de séchage et l'ateleir de conditionnement essayent de résoudre le problème de l'existence d'une quatité d'air dans quelques sachets de la levure sèche, pour amiliorer la qualité du produit .

Dans cette partie nous voulons découvrir le (ou) les facteurs qui agissent sur l'absence du vide dans les sachets de levure sèche. Donc, on doit détailler le procédé de séchage et la ligne d'emballage de cette dernière ;

VII.1 Séchage

Dans cette étape, la levure filtrée et mélangée par l'émulsifiant, et puis elle est déchargée dans le bol de la machine de séchage. L'opération de séchage commence sous certaines conditions qu'il faut respecter soigneusement afin d'avoir les résultats souhaités (le pourcentage de la matière sèche de la levure et la densité). Notre travail consiste à modéliser la liaison entre la densité et le temps de remplissage dans le bol, et pour cela on a essayé plus possible de fixer les autres paramètres (la quantité de l'émulsifiant dans la charge, le débit d'air, ...) et on a varié le temps de remplissage de la levure dans le bol.

On veut expliquer la densité à partir du temps de remplissage de la levure dans le bol.

Temps de remplissage de la levure dans le bol (s) X	Densité Y
300	60.74
310	62.15
360	65.10
380	65.80
390	66.15
420	66.30
450	66.50

$\rightsquigarrow$ Le modèle de régression linéaire simple est

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{pour} \quad i = 1, ..., 7$$

$\sum_{i=1}^{7} X_i = 2610$; $\sum_{i=1}^{7} Y_i = 452,74$

$\bar{X} = 372,86$; $\bar{Y} = 64,68$

$\sum_{i=1}^{7} X_i^2 = 991100$; $\sum_{i=1}^{7} Y_i^2 = 29313,3826$

$\sum_{i=1}^{7} X_i Y_i = 169498$

$b_1 = 0,0385$; $b_0 = 50,32489$

$\rightsquigarrow$ Donc le modèle linéaire est :

$$\boxed{\tilde{Y}_i = 50,32489 + 0,0385 X_i}$$

On obtient donc :

Temps de remplissage de la levure dans le bol (s) X	Densité Y	$\tilde{Y}$
300	60.74	61.87489
310	62.15	62.25989
360	65.10	64.18489
380	65.80	64.95489
390	66.15	65.33989
420	66.30	66.49489
450	66.50	67.64989

$$\sum_{i=1}^{7}(Y_i - \tilde{Y}_i)^2 = 4.8682 \qquad \Longrightarrow \qquad S^2 = 0.9736$$

Test de signification des coefficients du modèle

$\rightsquigarrow$ Paramètre β_0 :

Hypothèse du test :
$$\begin{array}{lcl} H_0 & : & \beta_0 = 0 \\ H_1 & : & \beta_0 \neq 0 \end{array}$$

$$S^2(b_0) = S^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{7}(X_i - \bar{X})^2}\right] \quad \Rightarrow \quad S^2(b_0) = 7.6827 \quad \Rightarrow \quad \boxed{S(b_0) = 2.7718}$$

$$\Rightarrow \qquad \boxed{\frac{b_0}{s(b_0)} = 18.156}$$

Et on a : $t_{0.025;5} = 2.5706,$

Donc : $\frac{b_0}{s(b_0)} \notin [-t_{0.025;5}, t_{0.025;5}]$

d'où on rejette H_0 c-à-d que $\beta_0 \neq 0$.

Tableau d'analyse de la variance :

Source de variation	Somme des carrés	D.d.l.	carré moyen
Régression	$SCR = 26.5958$	1	$CMR = 26.5958$
Résidus	$SC_{RES} = 4.8571$	5	$CM_{RES} = 5.32$
Total	$SCT = 31.4529$	6	

Coefficient de détermination :

On a :
$$r^2 = \frac{SCR}{SCT}$$
$$\Rightarrow \qquad r^2 = 0.8456$$

$\Rightarrow$ 84.56% de la variabilité de la densité est expliquée par le modèle obtenu.

Signification de la régression :

L'hypothèse du test : $H_0 \quad : \quad \beta_1 = 0$
$H_1 \quad : \quad \beta_1 \neq 0$

$$F_{Exp} = \frac{CMR}{CM_{RES}} \qquad \Rightarrow \qquad \boxed{F_{Exp} = 27.3783}$$

Et on a : $F_{0.05;1;5} = 6.61,$

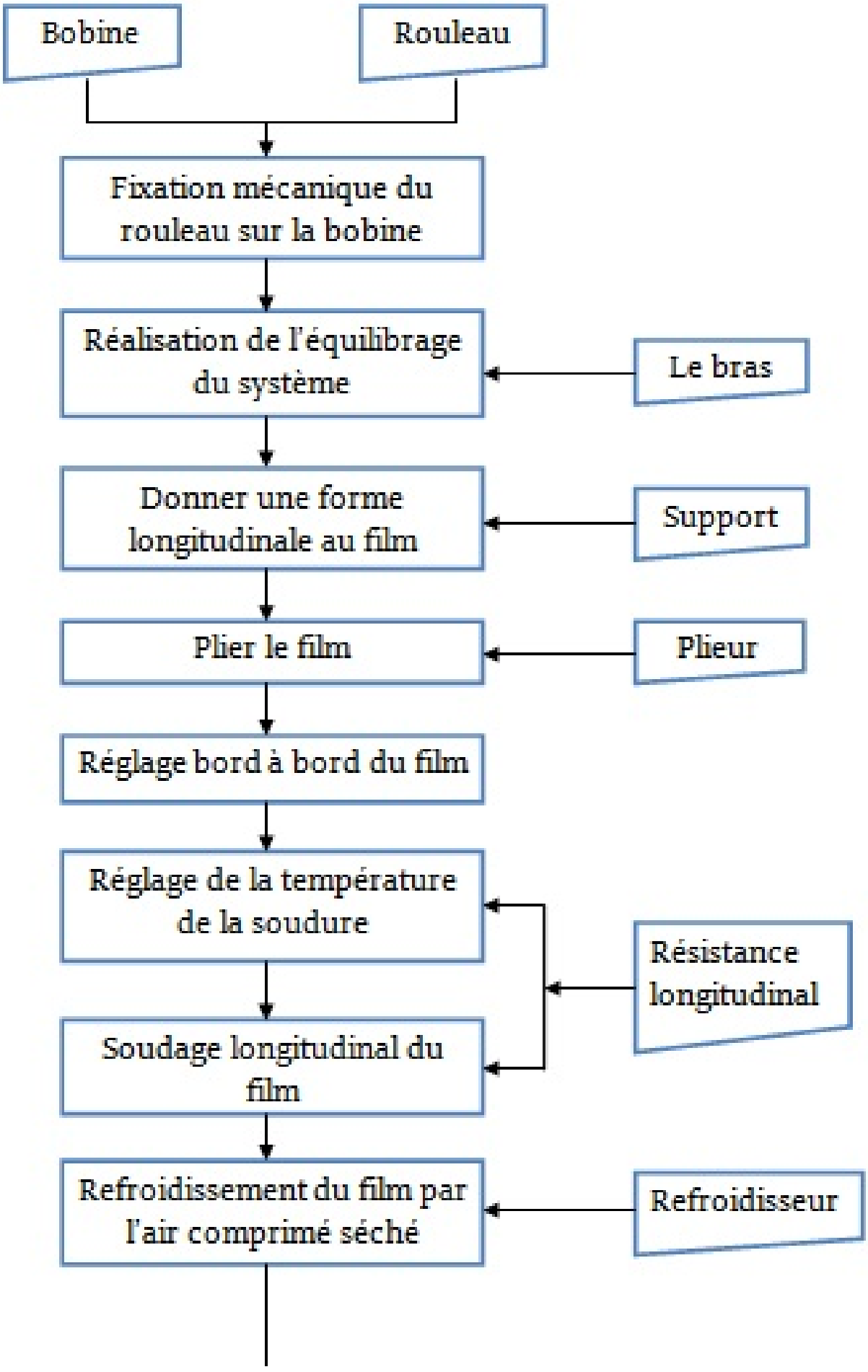
Donc : $F_{exp} \leq F_{0.05;1;5}$

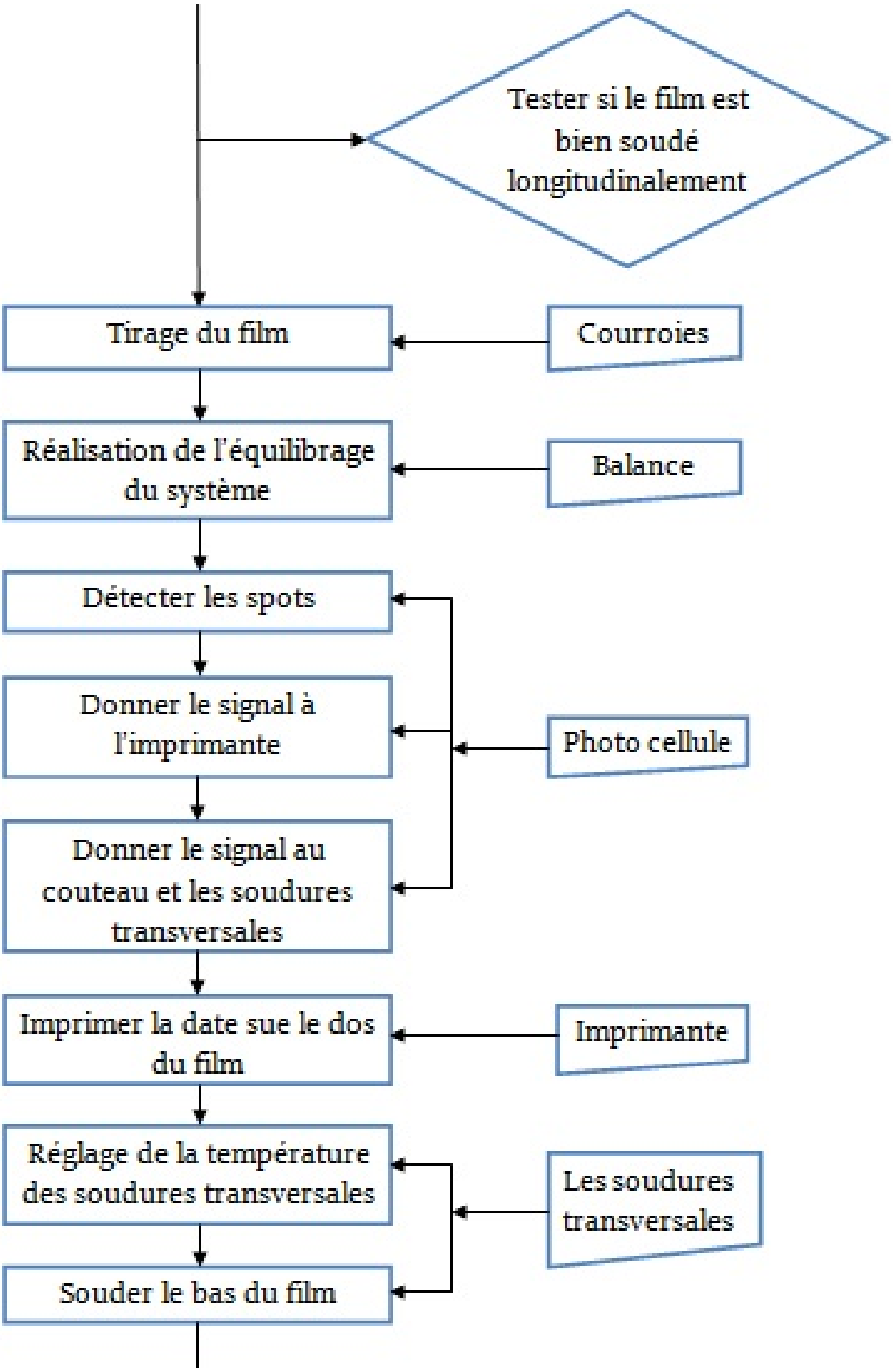
Alors on rejette H_0, c-à-d que $\beta_1 \neq 0$.

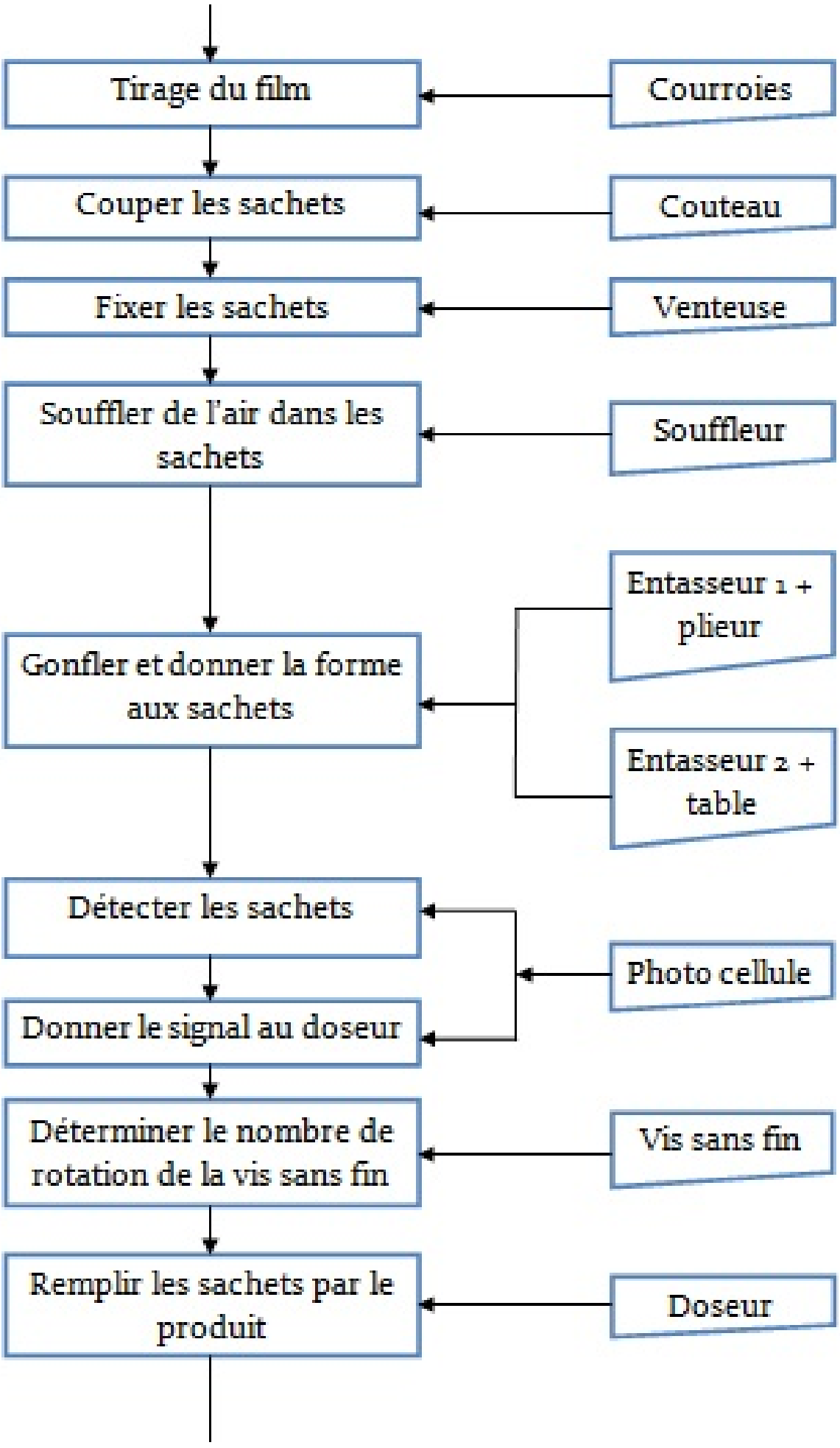
VII.2 Emballage sous vide :

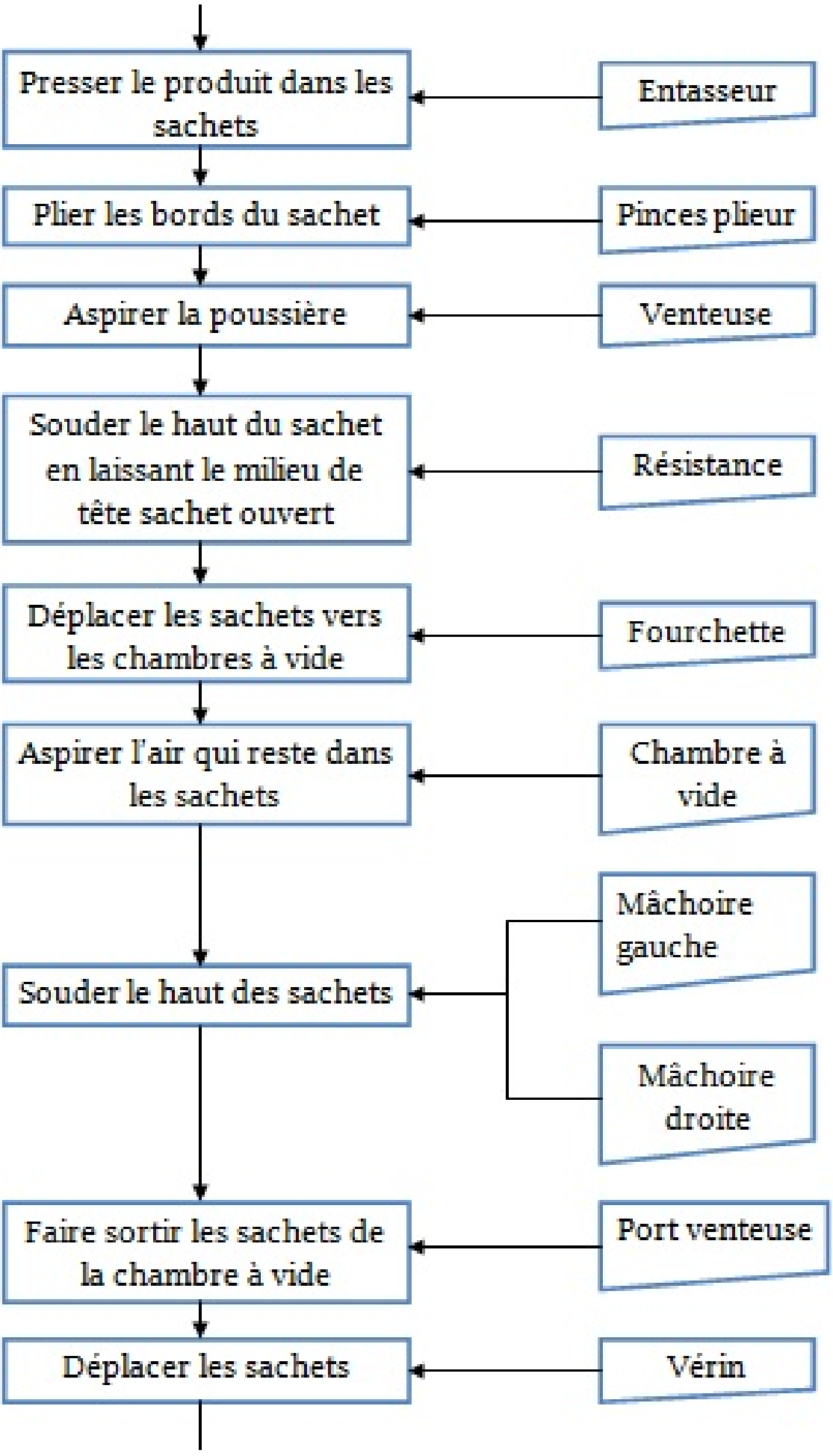
L'étape de l'emballage se décompose en plusieurs opérations dont le but d'emballer la levure séchée sous le vide pour éviter toutes réactions chimiques. Notre travail consiste à déterminer les facteurs principaux qui empêchent la réussite de cette étape. Pour cela on a détaillé la ligne d'emballage sous vide sous forme d'un organigramme :

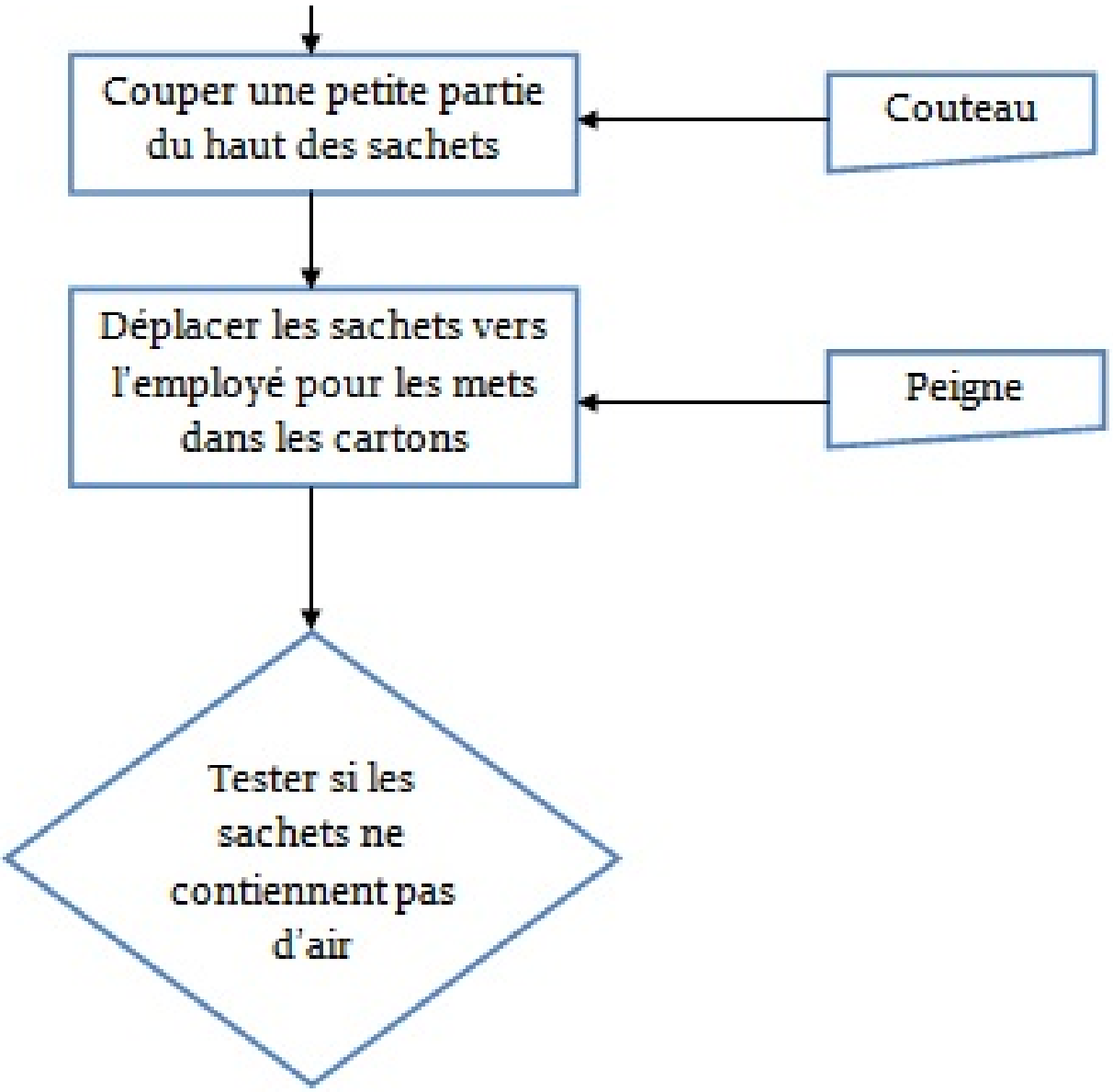
VII.2.1 Le procédé d'emballage sous vide




Tester si le film est bien soudé longitudinalement
Tirage du film
Courroies
Réalisation de l'équilibrage du système
Balance
Détecter les spots
Donner le signal à l'imprimante
Photo cellule
Donner le signal au couteau et les soudures transversales
Imprimer la date sue le dos du film
Imprimante
Réglage de la température des soudures transversales
Les soudures transversales
Souder le bas du film


Tirage du film
Courroies
Couper les sachets
Couteau
Fixer les sachets
Venteuse
Souffler de l'air dans les sachets
Souffleur
Gonfler et donner la forme aux sachets
Entasseur 1 + plieur
Entasseur 2 + table
Détecter les sachets
Photo cellule
Donner le signal au doseur
Déterminer le nombre de rotation de la vis sans fin
Vis sans fin
Remplir les sachets par le produit
Doseur


Presser le produit dans les sachets
Entasseur
Plier les bords du sachet
Pinces plieur
Aspirer la poussière
Venteuse
Souder le haut du sachet en laissant le milieu de tête sachet ouvert
Résistance
Déplacer les sachets vers les chambres à vide
Fourchette
Aspirer l'air qui reste dans les sachets
Chambre à vide
Souder le haut des sachets
Mâchoire gauche
Mâchoire droite
Faire sortir les sachets de la chambre à vide
Port venteuse
Déplacer les sachets
Vérin


Couper une petite partie du haut des sachets
Couteau
Déplacer les sachets vers l'employé pour les mets dans les cartons
Peigne
Tester si les sachets ne contiennent pas d'air

Après un suivi durant 18 jours de la ligne d'emballage, on a pris différentes valeurs de température de la soudure longitudinale (X_2), les trois soudures transversales (X_3, X_4 et X_5), la soudure de la tête sachet (X_6), et les deux mâchoires droite et gauche (X_7 et X_8) que le tableau ci-dessous les montrent, et on a considéré comme réponse le pourcentage des sachets mal emballé (Y) ;

Températures des soudures ($°C$)							Production (sachets)	Eliminer (sachets)	Y
X_2	X_3	X_4	X_5	X_6	X_7	X_8			
195.7	190.5	190	184.7	169	165	164.5	41320	10	0.0242
196.4	190.9	189.5	185.2	169.3	164.5	162.9	40272	19	0.0471
195.9	190.6	190	185.2	169	164.6	164.8	33456	5	0.0149
195	190.6	190	185	169	165	165	37680	45	0.1194
195	190.5	190	185	169	165	165	22464	24	0.1068
195.4	190.6	189.9	184.9	168.9	164.8	164.7	43824	39	0.0889
189.5	190.6	190.1	185.3	169	164.3	164.5	43200	57	0.1319
189.3	190.5	190	185.1	169	165.2	164.4	44064	7	0.0158
190.2	190.5	190	185	169	165.3	164.9	42384	32	0.0755
190.1	193.2	186.2	186.2	169	165.4	164.8	40464	7	0.0172
190.7	193.1	186.3	186.3	169	165.3	165.1	26976	21	0.0779
189.7	196.1	188.6	188.6	169	165.5	163.7	29280	13	0.0443
190.5	196	188.4	188.4	169	165.4	165.8	29328	20	0.0681
188.6	196	188.2	188.2	165.8	164.7	165.3	29712	8	0.0269
189.5	196.1	188.6	188.6	169	165.6	164.8	29088	7	0.024
189.6	196.1	188.5	188.5	169	165	165	39456	44	0.1115
189.6	196.1	188.5	188.5	169	165	165	28800	6	0.0208
190	196.1	188	188	169	164.8	165.5	42192	24	0.0568

Après l'application de l'ACP on obtient :

∗ La matrice des Corrélations ci-dessous avec une corrélation significative ; $r = 0.4683$ à 95%.

	X2_	X3_	X4_	X5_	X6_	X7_	X8_
X2_	1,0000						
X3_	-0,6586	1,0000					
X4_	0,4978	-0,6171	1,0000				
X5_	-0,6692	0,9918	-0,5460	1,0000			
X6_	0,2966	-0,2920	0,1509	-0,2718	1,0000		
X7_	-0,3780	0,3779	-0,4266	0,3695	0,1922	1,0000	
X8_	-0,3357	0,3438	-0,2844	0,3099	-0,2716	0,2002	1,0000

D'après la matrice des corrélation on conclut que ;

- ► X_2 est corrélé positivement avec X_4 et X_5 et il est corrélé négativement avec X_3.
- ► X_3 est corrélé positivement avec X_5 et négativement avec X_4.
- ► X_4 et X_5 sont corrélé négtivement.

∗ Les valeurs propres et les axes principaux :

Axe	Valeur Propre	% Inertie	% cumulé	Histogramme
1	3,5494	50,7033	50,7033	==========
2	1,2353	17,6469	68,3502	======
3	0,8141	11,6290	79,9791	====
4	0,5262	7,5174	87,4966	==
5	0,4878	6,9686	94,4652	==
6	0,3843	5,4901	99,9553	=
7	0,0031	0,0447	100,0000	

A partir du tableau ci-dessus on voit que l'axe 1 explique 50,7033 de l'enertie, l'axe 2 explique 17,6469 et l'axe 3 explique 11,6290.

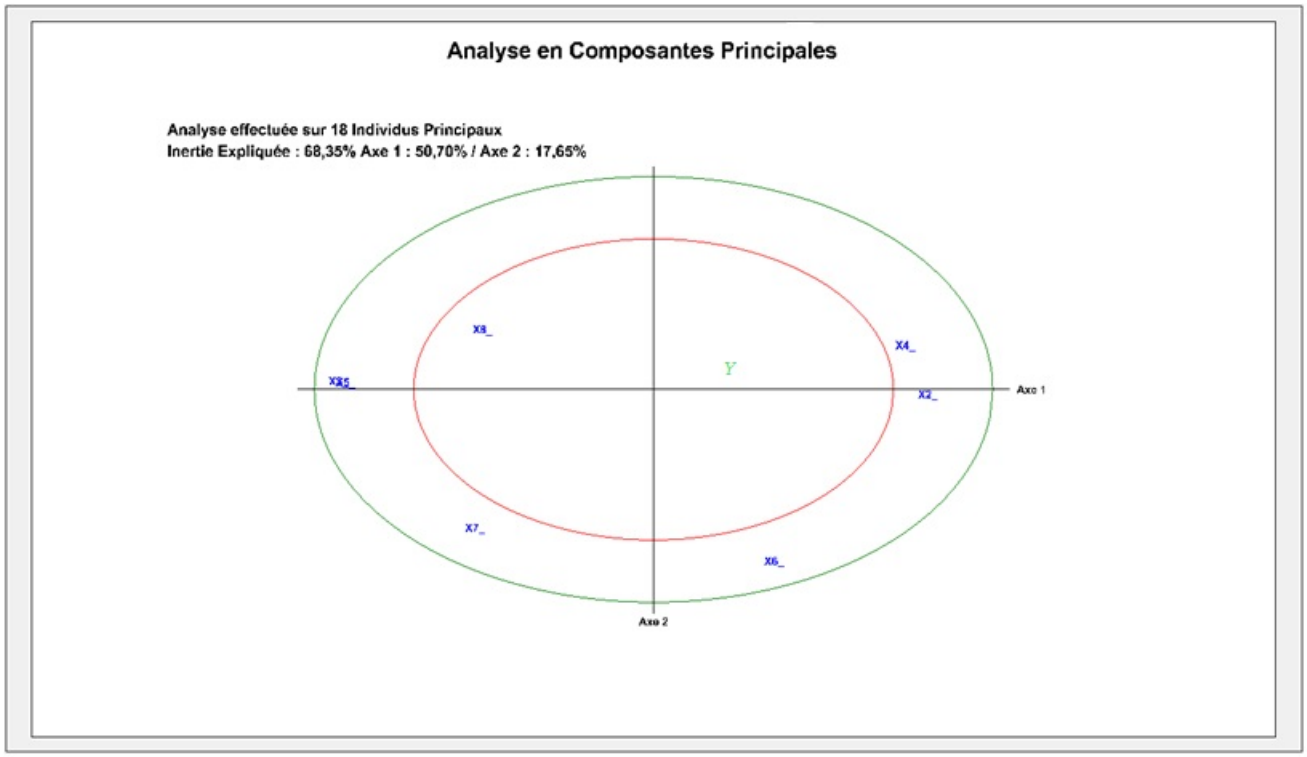
∗ Analyse des variables principales

Variables	Axe 1-coor	Axe 2-coor	Axe 3-coor	Axe 4-coor	Axe 5-coor	Axe 6-coor	Axe 7-coor
X2_	0,81	-0,04	-0,06	-0,07	-0,36	0,45	0,00
X3_	-0,93	0,02	0,23	0,19	-0,17	0,13	-0,04
X4_	0,74	0,19	0,00	0,55	0,31	0,12	0,00
X5_	-0,91	0,02	0,27	0,27	-0,12	0,13	0,04
X6_	0,36	-0,82	-0,04	0,27	-0,26	-0,23	0,00
X7_	-0,53	-0,67	-0,24	-0,09	0,36	0,28	0,00
X8_	-0,50	0,26	-0,79	0,17	-0,16	-0,03	0,00

D'après le tableau d'analyse des variables principales ci-dessus on a pris :

- ► Le facteur X_3 qui a la plus grande valeur sur l'axe 1 ;
- ► Le facteur X_6 qui ala plus grande valeur sur l'axe 2 ;
- ► Le facteur X_8 qui ala plus grande valeur sur l'axe 3.

∗ Le cercle des corrélations :



Le cercle des corrélations nous représente la projection de nos facteurs et de la réponse Y sur le plan défini par les deux premier axes principaux

On choisit donc X_3, X_6 et X_8 comme les facteurs les plus influents dans notre problèmatique. Et maintenant on cherche à modéliser la liaison entre nos facteurs et notre réponse Y, ce qui nécessite l'utilisation de la régression linéaire multiple ;

X_3	X_6	X_8	Y
190,5	169.0	164,5	0.0242
190,9	169.3	162,9	0.0471
190,6	169.0	164,8	0.0149
190,6	169.0	165.0	0.1194
190,5	169.0	165.0	0.1068
190,6	168.9	164,7	0.0889
190,6	169.0	164,5	0.1319
190,5	169.0	164,4	0.0158
190,5	169.0	164,9	0.0755
193,2	169.0	164,8	0.0172
193,1	169.0	165,1	0.0779
196,1	169.0	163,7	0.0443
196,0	169.0	165,8	0.0681
196,0	165.8	165,3	0.0269
196,1	169.0	164,8	0.0240
196,1	169.0	165.0	0.1115
196,1	169.0	165.0	0.0208
196,1	169.0	165,5	0.0568

$\rightsquigarrow$ Le modèle de la régression linéaire multiple est donnée par :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \text{pour} \quad i = 1, ..., 18$$

$\rightsquigarrow$ Le modèle de la regréssion multiple sous le forme matricielle

$$Y = \beta X + \varepsilon$$

Avec :

$$X = \begin{pmatrix} 1 & 190,5 & 169.0 & 164,5 \\ 1 & 190,9 & 169.3 & 162,9 \\ 1 & 190,6 & 169.0 & 164,8 \\ 1 & 190,6 & 169.0 & 165.0 \\ 1 & 190,5 & 169.0 & 165.0 \\ 1 & 190,6 & 168.9 & 164,7 \\ 1 & 190,6 & 169.0 & 164,5 \\ 1 & 190,5 & 169.0 & 164,4 \\ 1 & 190,5 & 169.0 & 164,9 \\ 1 & 193,2 & 169.0 & 164,8 \\ 1 & 193,1 & 169.0 & 165,1 \\ 1 & 196,1 & 169.0 & 163,7 \\ 1 & 196,0 & 169.0 & 165,8 \\ 1 & 196,0 & 165.8 & 165,3 \\ 1 & 196,1 & 169.0 & 164,8 \\ 1 & 196,1 & 169.0 & 165.0 \\ 1 & 196,1 & 169.0 & 165.0 \\ 1 & 196,1 & 169.0 & 165,5 \end{pmatrix} \quad ; \quad Y = \begin{pmatrix} 0.0242 \\ 0.0471 \\ 0.0149 \\ 0.1194 \\ 0.1068 \\ 0.0889 \\ 0.1319 \\ 0.0158 \\ 0.0755 \\ 0.0172 \\ 0.0779 \\ 0.0443 \\ 0.0681 \\ 0.0269 \\ 0.0240 \\ 0.1115 \\ 0.0208 \\ 0.0568 \end{pmatrix} \quad ; \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad ; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \\ \varepsilon_{17} \\ \varepsilon_{18} \end{pmatrix}$$

$\rightsquigarrow$ Pour obtenir une estimation de β, nous devons résoudre le système d'équation suivant :

$$X^T X B = X^T Y$$

La résolution de ce système d'équation s'obtient à l'aide de la matrice inverse de $X^T X$:

$$B = (X^T X)^{-1} X^T Y$$

On a :

$$X^T X = \begin{pmatrix} 18.0 & 3450.7 & 3039.0 & 2965.7 \\ 3450.7 & 661657.6 & 582604.2 & 568530.6 \\ 3039.0 & 582604.2 & 513094.3 & 500706.7 \\ 2965.7 & 568530.6 & 500706.7 & 488639.2 \end{pmatrix}$$

$$\Rightarrow \qquad (X^T X)^{-1} = \begin{pmatrix} 9714.962410 & -2.144258037 & -22.497189791 & -33.41544067 \\ -2.144258037 & 0.008533977 & -0.007118387 & 0.01037908 \\ -22.497189791 & -0.007118387 & 0.115654796 & 0.02631349 \\ -33.41544067 & 0.01037908 & 0.02631349 & 0.16377117 \end{pmatrix}$$

Et on a :

$$X^T Y = \begin{pmatrix} 1.0720 \\ 205.7309 \\ 181.0872 \\ 176.6893 \end{pmatrix}$$

$$\Rightarrow \qquad B = \begin{pmatrix} -4.80239399 \\ 0.00188093 \\ 0.01145000 \\ 0.01558756 \end{pmatrix}$$

Alors le modèle de la régression multiple devient :

$$\boxed{\tilde{Y}_i = -4.80239399 + 0.00188093 X_{i1} + 0.01145 X_{i2} + 0.01558756 X_{i3}}$$

$$Y^T Y = 0.09029446, \qquad \bar{Y} = 0.003308642, \qquad B^T X^T Y = 0.06640099$$

Tableau d'analyse de la variance :

Source de variation	Somme des carrés	D.d.l.	Carré moyen
Régression	$SCR = B^T X^T Y - n\bar{Y}^2 = 0.06620394$	3	$CMR = 0.02206798$
Résidus	$SC_{RES} = Y^T Y - B^T X^T Y = 0.02301175$	14	$CM_{RES} = 1.6437 \times 10^{-3}$
Total	$SCT = Y^T Y - n\bar{Y}^2 = 0.09009741$	17	

Coefficient de détermination multiple R^2

$$R^2 = \frac{SCR}{SCT} \qquad \boxed{R^2 = 0.734804}$$

C-à-d que 73.48% de la variation de Y est expliquée par le modèle obtenu.

Test de signification de la régression dans son ensemble :

Hypothèse du test :

$$\begin{array}{lll} H_0 & : & \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1 & : & \text{au moins un des } \beta_j \neq 0 \qquad j = 0, ..., 3 \end{array}$$

$$F_{exp} = \frac{CMR}{CM_{RES}} \qquad \Rightarrow \qquad \boxed{F_{Exp} = 13.4258}$$

Soit $\alpha = 0.05$ le seuil de signification du test, on a $F_{0.05;3;14} = 3.34$

Donc : $F_{Exp} \geq F_{0.05;3;14}$

D'où on rejette H_0, c-à-d que la contribution de l'ensemble des variables pour exliquer les fluctuation de Y est significative.

Contribution marginale et estimation par intervalle

La matrice des variances covariances de coefficints de la régression :

$$CM_{RES}(X^T X)^{-1} = \begin{pmatrix} 15.968483713 & -0.003524517 & -0.036978631 & -0.054924960 \\ -0.003524517 & 1.402730 \times 10^{-5} & -1.170049 \times 10^{-5} & 1.706009 \times 10^{-5} \\ -0.036978631 & -1.170049 \times 10^{-5} & 1.901018 \times 10^{-4} & 4.325148 \times 10^{-5} \\ -0.054924960 & 1.706009 \times 10^{-5} & 4.325148 \times 10^{-5} & 2.691907 \times 10^{-4} \end{pmatrix}$$

$\rightsquigarrow$ Test de signification de β_0

$$\begin{array}{lll} H_0 & : & \beta_0 = 0 \\ H_1 & : & \beta_0 \neq 0 \end{array}$$

Soit $\alpha = 0.05$ le seuil de signification du test, sous H_0 et les conditions d'application du test $\dfrac{b_0}{S(b_0)} \backsim T_{14}$.

Régle de décision : $t_0 = \dfrac{b_0}{S(b_0)} = -1.201783$ et on a $t_{0.025;14} = 2.1448$

Alors $t_0 \in [-t_{\frac{\alpha}{2};n-k-1}; t_{\frac{\alpha}{2};n-k-1}]$

Donc : on accepte H_0.

Intervalle de confiance au seuil $1 - \alpha$:

$$-13.37314 \leq \beta_0 \leq 3.768352$$

$\rightsquigarrow$ Test de signification de β_1

$$\begin{array}{lcl} H_0 & : & \beta_1 = 0 \\ H_1 & : & \beta_1 \neq 0 \end{array}$$

Soit $\alpha = 0.05$ le seuil de signification du test, sous H_0 et les conditions d'application du test $\dfrac{b_1}{S(b_1)} \backsim T_{14}$.

Régle de décision : $t_1 = \dfrac{b_1}{S(b_1)} = 0.5022103$ et on a $t_{0.025;14} = 2.1448$

Alors $t_1 \in [-t_{\frac{\alpha}{2};n-k-1}; t_{\frac{\alpha}{2};n-k-1}]$

D' où : on accepte H_0, c-à-d que l'effet de X_1 n'est pas significatif.

Intervalle de confiance au seuil $1-\alpha$:

$$-0.006151997 \leq \beta_1 \leq 0.009913857$$

$\rightsquigarrow$ Test de signification de β_2

$$\begin{array}{lcl} H_0 & : & \beta_2 = 0 \\ H_1 & : & \beta_2 \neq 0 \end{array}$$

Soit $\alpha = 0.05$ le seuil de signification du test, sous H_0 et les conditions d'application du test $\dfrac{b_2}{S(b_2)} \backsim T_{14}$.

Régle de décision : $t_2 = \dfrac{b_2}{S(b_2)} = 0.8304479$ et on a $t_{0.025;14} = 2.1448$

Alors $t_2 \in [-t_{\frac{\alpha}{2};n-k-1}; t_{\frac{\alpha}{2};n-k-1}]$

D' où : on accepte H_0, c-à-d que l'effet de X_2 n'est pas significatif.

Intervalle de confiance au seuil $1-\alpha$:

$$-0.01812195 \leq \beta_2 \leq 0.04102195$$

$\rightsquigarrow$ Test de signification de β_3

$$\begin{array}{lcl} H_0 & : & \beta_3 = 0 \\ H_1 & : & \beta_3 \neq 0 \end{array}$$

Soit $\alpha = 0.05$ le seuil de signification du test, sous H_0 et les conditions d'application du test $\dfrac{b_3}{S(b_3)} \backsim T_{14}$.

Régle de décision : $t_3 = \dfrac{b_3}{S(b_3)} = 0.9500536$ et on a $t_{0.025;14} = 2.1448$

Alors $t_3 \in [-t_{\frac{\alpha}{2};n-k-1}; t_{\frac{\alpha}{2};n-k-1}]$

D' où : on accepte H_0, c-à-d que l'effet de X_3 n'est pas significatif.

Intervalle de confiance au seuil $1-\alpha$:

$$-0.01960224 \leq \beta_3 \leq 0.05077736$$

Conclusion :

D'après l'étude statistique on conclut qu'on a un effet global, car lorsqu'on a testé l'effet de chaque facteur séparément on n'a pas trouvé d'effet marginal, ce qui nous dirige à penser à d'autres facteurs non pris en considération, du fait qu'on ne peut pas mesurer et à titre d'exemple : la poussière.

Annex

Application de l'AFC

Dans une Ancienne enquête du CREDOC(Centre de Recherche pour l'Etude et l'Observation des Conditions de Vie) (Nicole Tabard, 1974) ; on a 1724 femmes interrogées ;
Questions relatives à leur attitude à l'égard du travail féminin.

⇝ La famille idéale est celle où :

∗Les deux conjoints travaillent également	261	15.1
∗Le mari a un métier plus absorbant que celui de sa femme	555	32.2
∗Seul le marie travaille	908	52.7
	1724	100

⇝ Activité convenant le mieux à une mère de famille quand les enfants vont à l'école :

∗ Rester au foyer	284	16.5
∗ Travailler à mi-temps	1123	65.1
∗ Travailler à plein temps	317	18.4
	1724	100

Réponses simultanées à des questions d'opinion

La famille idéale est celle où :	Activité convenant le mieux à une mère de famille quand les enfants vont à l'école :			
	Rester eu foyer	Travailler à mi-temps	Travailler à plein-temps	Marge colonne
Les deux conjoints travaillent égalment	13	142	106	261
Le mari a un métier plus absorbant que celui de sa femme	30	408	117	555
Seul le mari travaille	241	573	94	908
Marge ligne	284	1123	317	1724

FIGURE 10 – Tableau croisant les réponses à deux questions d'opinion

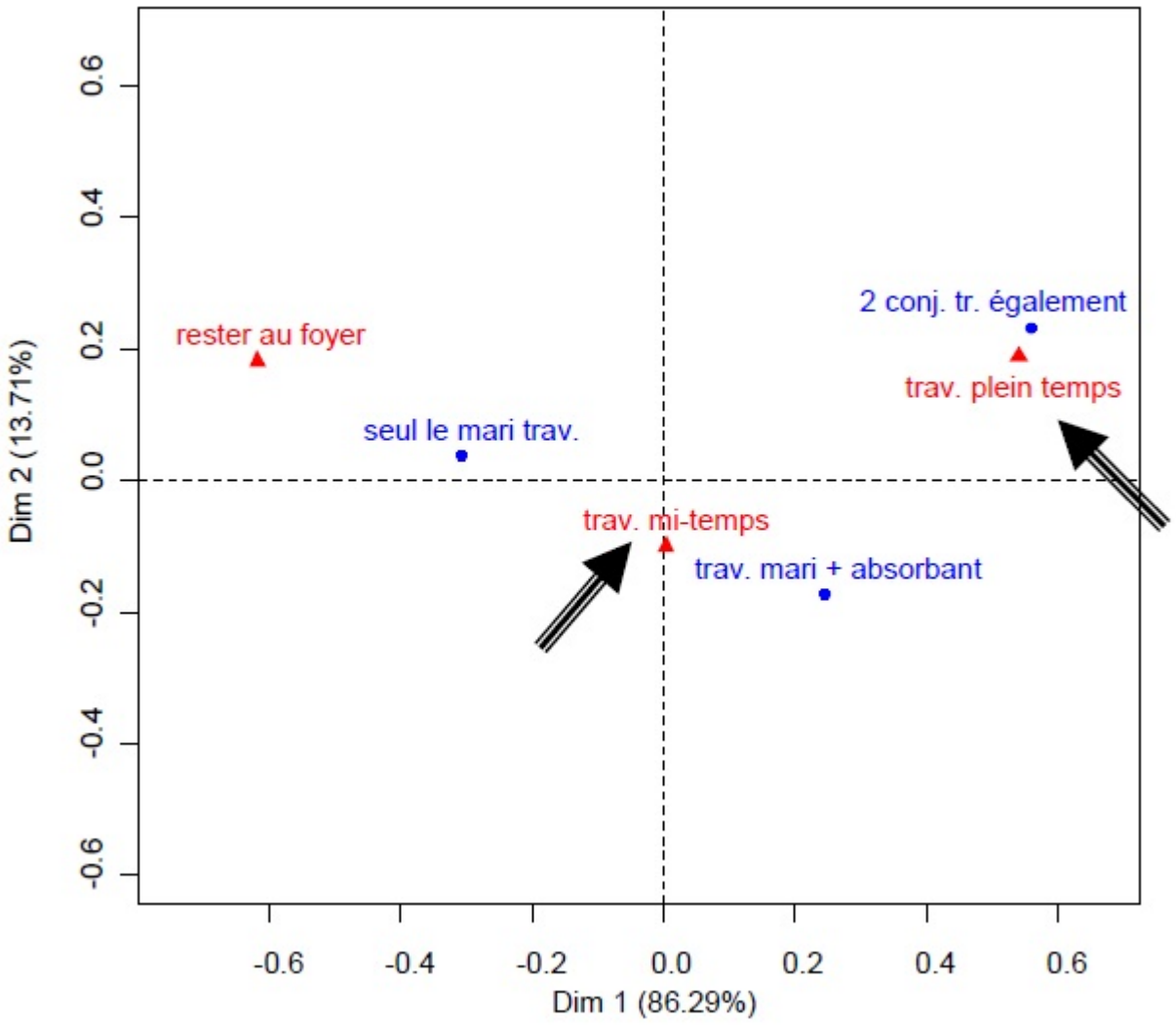
“Comparer un profil ligne au profil moyen”

		Activité mère de famille			
		Rester au foyer	Travailler à mi-temps	Travailler à plein-temps	$\sum$
La famille idéale	Les deux conjoints travaillent égalment	4,98	54,41	40,61	100
	Le mari a un métier plus absorbant que celui de sa femme	5,41	73,51	21,08	100
	Seul le mari travaille	26,54	63,11	10,35	100
G_I		16,47	65,14	18,39	100

“Comparer un profil colonne au profil moyen”

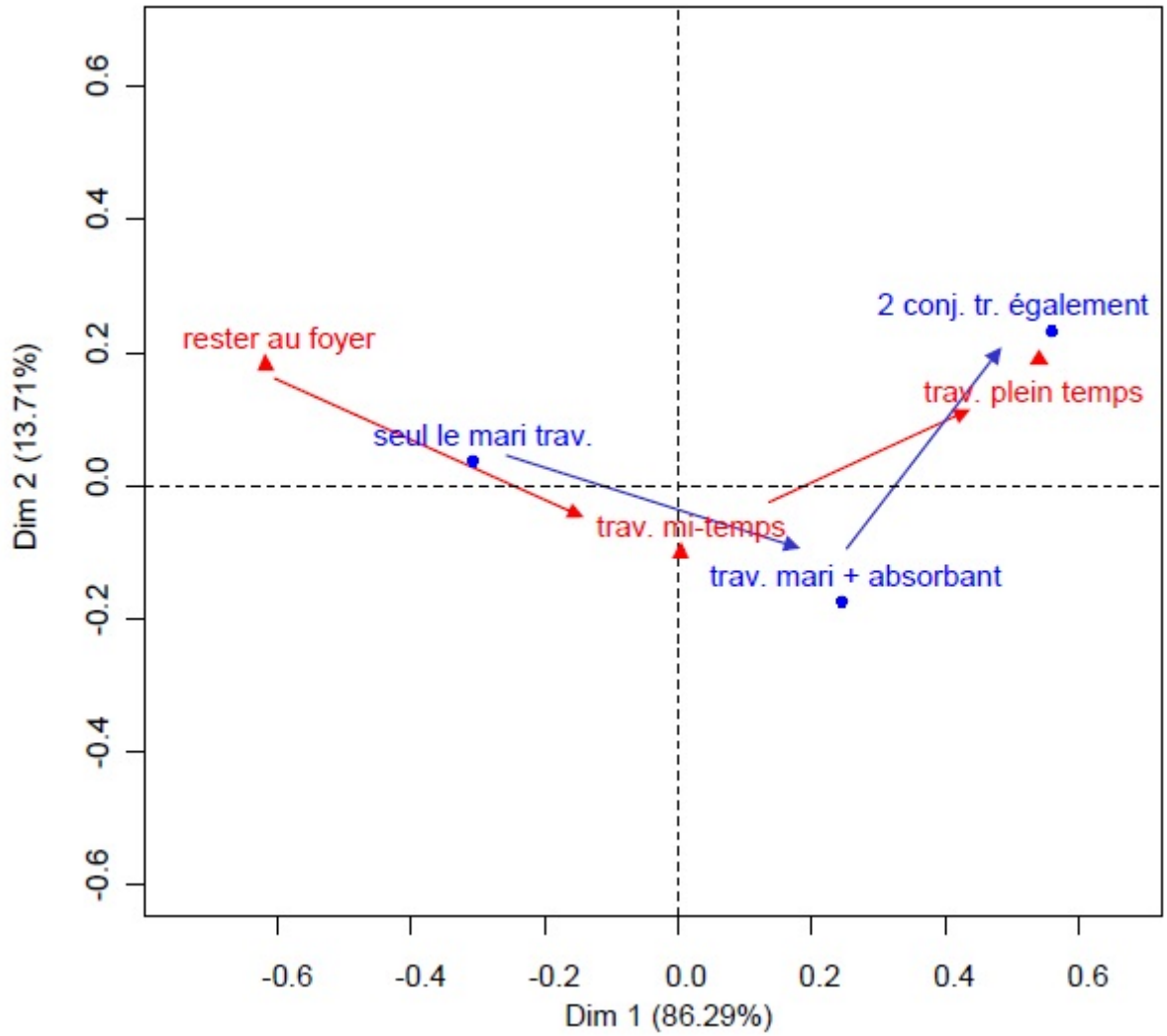
		Activité mère de famille			
		Rester au foyer	Travailler à mi-temps	Travailler à plein-temps	G_J
La famille idéale	Les deux conjoints travaillent égalment	4,58	12,64	33,44	15,14
	Le mari a un métier plus absorbant que celui de sa femme	10,56	36,33	36,91	32,19
	Seul le mari travaille	84,86	51,02	29,65	52,67
$\sum$		100	100	100	100

Régle d'interprétation :



	Rester au foyer	Travailler à mi-temps		Travailler à plein-temps		$O = G_J$
Les deux conjoints travaillent égalment	4,58	12,64		33,44		15,14
Le mari a un métier plus absorbant que celui de sa femme	10,56	36,33		36,91		32,19
Seul le mari travaille	84,86	51,02		29,65		52,67
$\sum$	100	100		100		100

Nommer le 1^{er} axe



Premier axe : Attitude à l'égard du travail féminin (ligne et colonnes)Dualité.

Propriété barycentrique

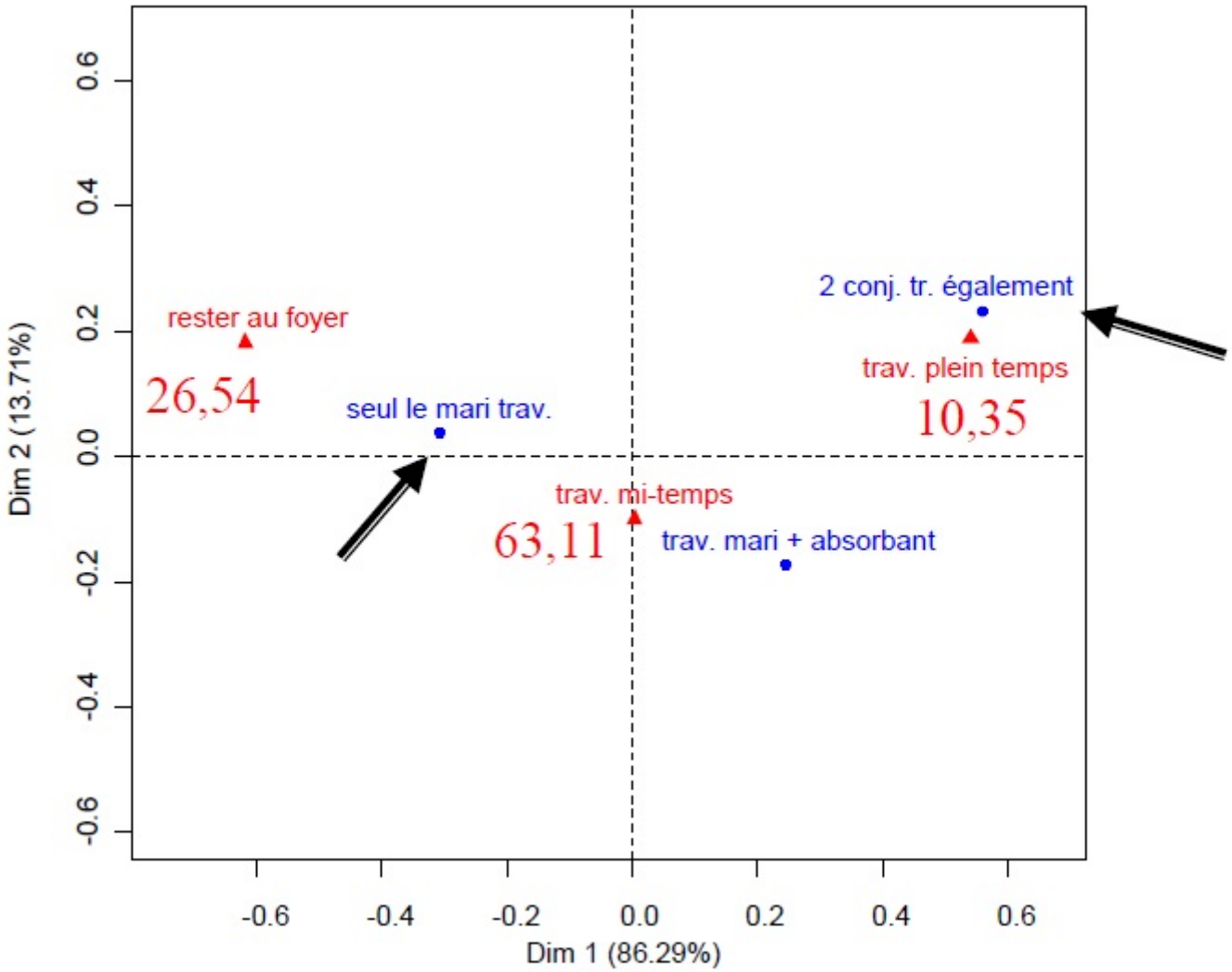
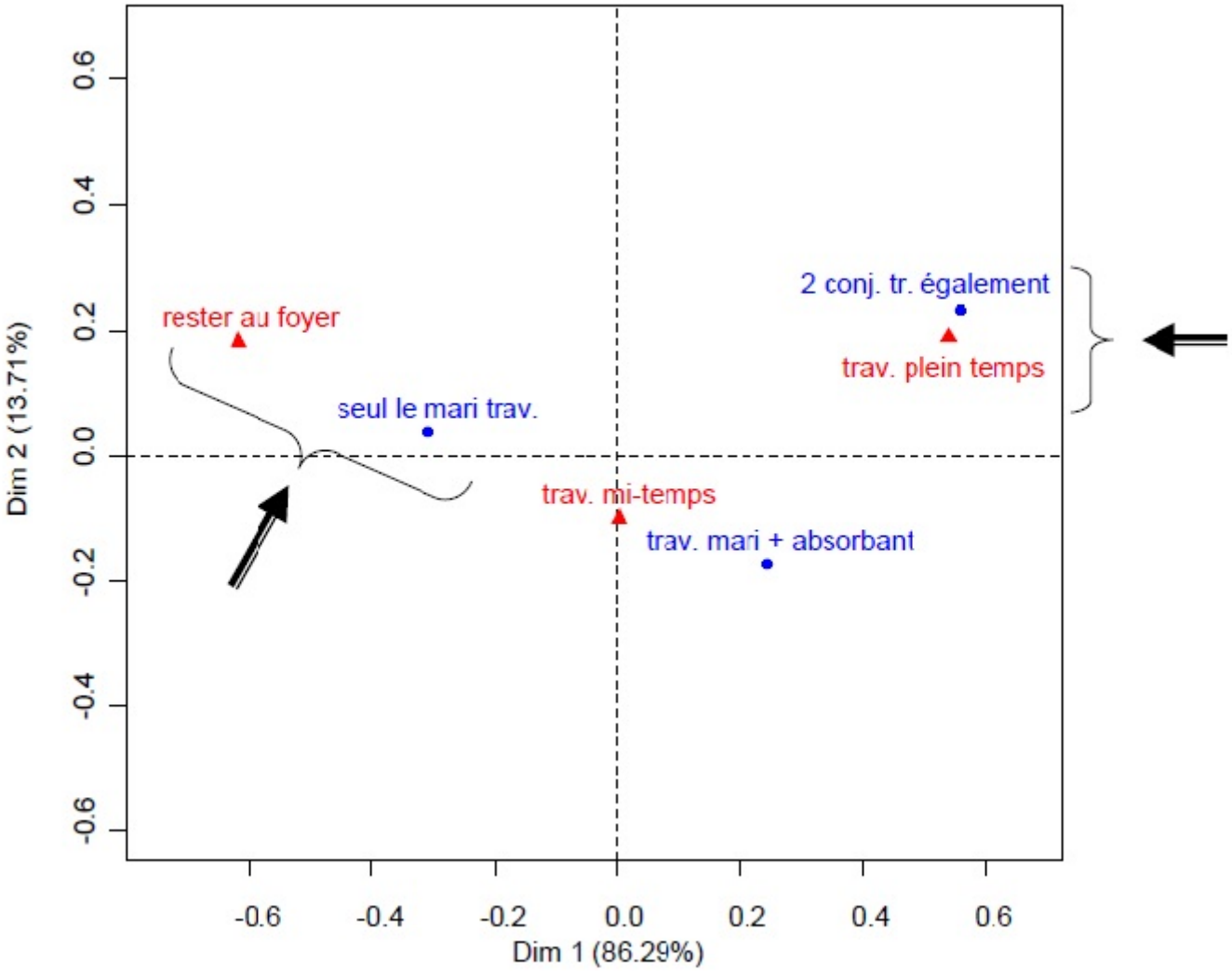


Tableau des Profils lignes

	Rester au foyer	Travailler à mi-temps	Travailler à plein-temps	$\sum$
Les deux conjoints travaillent égalment	4,98	54,41	40,61	100
Le mari a un métier plus absorbant que celui de sa femme	5,41	73,51	21,08	100
Seul le mari travaille	26,54	63,11	10,35	100
$O = G_I$	16,47	65,14	18,39	100

Interprétation simultanée des lignes et des colonnes



Le premier axe classe les modalités (des deux questions) depuis la plus défavorable au travail féminin jusqu'à la plus favorable au travail féminin.

La famille idéale est celle où :	Activité convenant le mieux à une mère de famille quand les enfants vont à l'école :			
	Rester eu foyer	Travailler à mi-temps	Travailler à plein-temps	Marge colonne
Les deux conjoints travaillent égalment	13	142	[106]	261
Le mari a un métier plus absorbant que celui de sa femme	30	[408]	117	555
Seul le mari travaille	[241]	573	94	908
Marge ligne	284	1123	317	1724

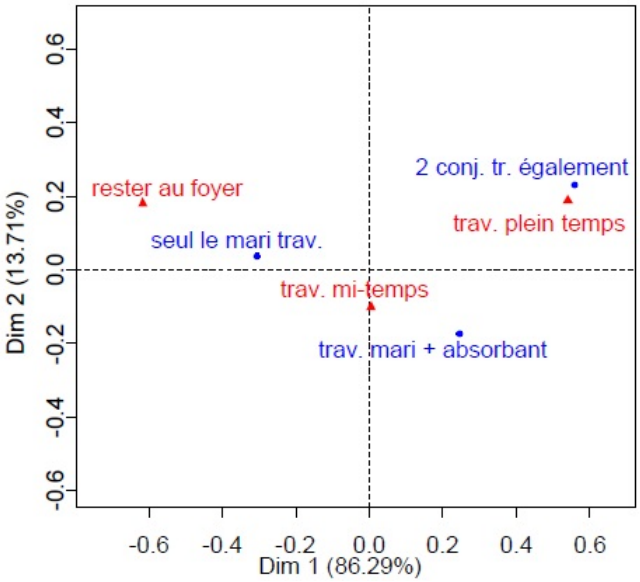
Tbleau croisant les réponses à deux questions d'opinion

L'AFC suggére que rester au foyer est plus défavorable à l'égard du travail féminin que seul le mari travaille.

$\rightsquigarrow$ Dans $\mathbb{R}^I$ d^2(rester au foyer, G_J) = .416

$\rightsquigarrow$ Dans $\mathbb{R}^J$ d^2(seul le mari travaille, G_I) = .097

La famille idéale est celle où :	Activité mère de famille :			
	Rester eu foyer	Trav. à mi-temps	Trav. à plein-temps	
2 conj. tr. égalment	13	142	[106]	261
trav. mari + absorbant	30	[408]	117	555
Seul le mari trav.	[241]	573	94	908
Marge ligne	284	1123	317	1724



Bilan

L'essentiel de l'écart à l'indépendance est structuré par l'attitude à l'égard du travail féminin. La position des modalités le long de l'échelle d'attitude éclaire leur signification.

La proximité de travailler à mi-temps avec le profil moyen suggère que cette modalité est « neutre », à la différence de l'autre modalité moyenne (plutôt favorable au travail féminin).

Conclusion

Au terme de notre analyse et en guise de conclusion, nous pouvons dire que ce stage bien limité dans le temps, nous permet de tester certaines de nos connaissances théoriquement acquises aux sein de la FST, et les traduire en pratique. Ce serait vraisemblablement une étape transitoire et précieuse vers la traduction de nos connaissances théoriques cumulées par le passé vers les aspects pratiques de la vie professionnelle.

Dans le but d'une meilleure optimisation des ressources disponibles et de réduction des coûts des pertes qui sont la clé de survie de l'entreprise, nous avons essayé dans un premier temps de faire la description des machines mises en service. Cette démarche a été suivie d'une description de toute la ligne d'emballage (ICA), afin de savoir d'où vient le problème d'air dans quelque sachets d'emballage de la levure sèche, le second problème traiter dans le rapport est celui de la variabilité de la densité de la levure sèche.

Pour le problème d'emballage nous avons commencé par une analyse des différents facteurs soupçonnés influant sur le résultat d'emballage. Parmi ces facteurs nous avons choisi ceux expliquant plus la variabilité du procédé. Après une modélisation multiple des données, notre modèle est globalement significatif. Cependant, nous avons poussé notre étude plus loin pour savoir l'effet de chaque variable sur notre réponse. Nous avons constaté que les facteurs étudiés n'ont pas une influence marginale sur le problème posé c.à.d. qu'il 'y a présence d'autre facteurs non mesurable qui ont un effet sur la réponse Y, par exemple la poussière qui empêche la clôture de la tête sachet ce qu'il résulte l'absence du vide dans les sachets finals. Pour la résolution du second problème, nous avons donné une modélisation de la densité de la levure sèche en fonction du temps de remplissage.

Par-ailleurs cette période de stage effectuée à Lesaffre-Maroc m'a permis d'acquérir des connaissances sur le plan technique et humain. Le contact que j'ai eu le plaisir d'avoir avec le personnel de la société en général à tous niveaux hiérarchiques confondus, m'a permis de comprendre les gens, de développer mon sens de la communication, de mieux gérer le système hiérarchique de la société, et d'apprendre à me connaitre moi-même.

Cette expérience a été très enrichissante tant sur le niveau professionnel que personnel.

Référence Bibliographique :

[**1**] Mme EZZAKI Fatima, Cours de processus Stochastique et optimisation, Master MACS, FST fès ;

[**2**]Mme AMMOR Ouafae, Cours de statistique décisionnelle, Master MACS, FST fès ;

[**3**] Mme Houda Benbrahim, Cours d'Analyse des Données, ENSIAS Rabat ;

[**4**] Jérôme Pagès, Cours d'AFC, , Laboratoire de mathématique appliquées Agrocampus, Rennes ;

[**5**] www.wikipédia.org ;

[**6**] www.lesaffre.com