N^o d'ordre: 84/2024-D/MT

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie HOUARI BOUMEDIENE

Faculté Mathématiques

THÈSE DE DOCTORAT EN SCIENCES

Présentée pour l'obtention du grade de **DOCTEUR**

En : Mathématiques

Spécialité : Probabilité & Statistique

Par : BOUCHAFAA ASMA

Thème

Analyse de la qualité de vie des données spatiales

Soutenue publiquement, le **26/05/2024**, devant le jury composé de :

1.	Mme.	H. SAGGOU	Professeur à l'USTHB	Présidente
2.	Mme.	K.DJABALLAH	Professeur à l' USTHB	Directrice
3.	Mme.	M.OURBIH	Professeur au Centre universitaire de Tipaza	Examinatrice
4.	M.	Y.BOULAROUK	Maitre de conférence A au Centre universitaire de Mila	Examinateur
5.	M.	S. BENJRADA	Maitre de conférence B à l'Université de Constantine 3	Invite

I cannot thank my dear parents, sisters, and brothers enough, who have always believed in me, supported me, and helped me many times with their encouragement or availability; I would not be here without them.
I dedicate this modest work to: My father, BOUCHAFAA BelKACEM(may allah have mercy on him and grant him al jannah), you are always in my heart. My mother, KHERRAT Fatima, My husband, Belaroussi Samir, My daughters, Sirine and Sabrina Hanane My son ,rayane My father-in-law, Belaroussi Mouloud, My mother-in-law, Bouchafa Nouara My sisters: Louiza and Karima Safia Rafika, My brothers: Kamel Sofiane Badr Abderahim Djeber Oussama, To all the family of Bouchafaa and and the family of Belaroussi, To all my friends and colleagues for their encouragement, and their support.

Acknowledgements

At the end of this work, I am grateful to my supervisor, Professor DJABALLAH Khedidja, whom I would like to thank particularly for her confidence, which allowed me to access the world of scientific research.

I warmly thank her for her human qualities, wisdom, availability, and great patience she has devoted to guiding, helping, improving, and completing my work. My words are not enough to express my sincere gratitude to her.

My sincere thanks to Ms. SAGGOU Hafida, Professor at USTHB, for her particular interest and valuable contribution to this work. Her comments and advice have helped me immensely. I am honored that she chairs the jury for this thesis.

I would like to warmly thank Ms. M. OURBIH M, Professor at Centre Universitaire de Tipaza, and Mr. BOULAROUK Y, MCA at the University of Mila, for reviewing my work and for their comments, which helped improve the manuscript. I am honored to have them as part of my jury.

Special thanks go to Mr. BENJARADA Salih, MCB at the University of Constantine, for his help and for accepting the invitation to be part of the jury.

I would also like to thank all the members of the M.S.T.D. laboratory and colleagues who helped me directly or indirectly in my work.

Abstract

Most regression methods rely on assumptions about the conditional distribution of the dependent variable given the explanatory variables. Assuming normality of the error variables can simplify the estimator considerably. In this thesis, we propose a linear regression model with an intercept, assuming non-normal errors. We consider the case where the errors follow an exponential distribution. The maximum likelihood estimate of the parameter in the model is developed under this hypothesis. We describe the theoretical properties of the proposed estimator, including its limit distribution. Furthermore, we estimate the regression parameter using the standard least squares method for comparison.

Furthermore, analysis of cereal production allows one to make decisions about the importance of certain products and the water resources. The study con- sidered here is based on statistical methods to model the production of cereal. A principal component analysis was used and the results obtained revealed a classification of regions according to their cereal production. It appears that durum wheat production was explained jointly by precipitation and irrigation. However, the variations in the production of the bread wheat, oat and barley can only be explained by precipitation. The results showed that the crop yield depended heavily on rainfall and very little on irrigation.

Keywords: Regression, Estimation,Exponential Distribution, Intercept-only,Convergence. cereal, irrigation, rainfall, regression, principal component, forcasting.

Résumé

La plupart des données des méthodes de régression ne sont basées que sur des hypothèses concernant la distribution conditionnelle de la variable dépendante, compte tenu des explicatifs variables. Si nous supposons la normalité des variables d'erreur, l'estimateur peut être considérablement simplifié.Dans cette Thèse , nous proposons un modèle de régression linéaire uniquement à l'origine sous l'hypothèse de non-normalité. Nous considérons ici que les erreurs suivent la loi exponentielle.L'Estimation possible maximum du paramètre dans le modèle est développée sous cette hypothèse. Nous décrivons les propriétés théoriques de l'estimateur proposé, y compris sa distribution limite.De plus, le paramètre de régression est estimé par la méthode standard des moindres carrés pour faire une comparaison.

De plus, l'analyse de la production céréalière permet de trancher sur l'importance de certains produits et sur les ressources en eau. L'étude envisagée ici s'appuie sur des méthodes statistiques pour modéliser la production de blé dur, de blé panifiable, d'orge et d'avoine. La première méthode utilisée est l'analyse en composantes principales. Elle a été appliquée pour classer les données afin de déterminer l'importance relative des différentes régions pour l'évaluation de la production céréalière. . Il apparaît que la production de blé dur s'explique conjointement par les précipitations et l'irrigation Cependant, les variations de production du blé tendre, de l'avoine et de l'orge ne peuvent s'expliquer que par les précipitations.

Mots Clés: Régression, Estimation, Distribution exponentielle, Interception uniquement, Convergence.

cereal, irrigation, rainfall, regression, principal compo- nent, forcasting.

Contents

List of Figures

List of Tables

Introduction

In this thesis, we have addressed two types of modeling, namely:

- The linear model, which encompasses several statistical analysis methods for approximating a variable based on other correlated variables. By extension, the term is also used to define certain curve-fitting methods.

- Exploratory statistical analysis methods are used to shape vast datasets, extract structures from them, and validate these structures. They fall under multidimensional exploratory statistics. The method used here is primarily Principal Component Analysis (PCA).

These methods generalize classical descriptive statistics and use fairly intuitive mathematical tools, but more complex than the means, variances, and empirical correlation coefficients of descriptive statistics. The basic analysis methods investigated here are : Principal Component Analysis and the linear model(linear modeling).

The first case describes the relationship between a response variable and one or more predictor variables. It is used to analyze a well-formulated hypothesis. The linear model evaluates whether there is a significant correlation between the response variable and the explanatory variable(s). This is done by evaluating whether the mean value of the response variable significantly differs between different values of the explanatory variables. In almost all cases, this later do not explain all the variation in the response variable. The remaining unexplained variation is the residuals or the error. For the results of a linear model to be interpretable, the residuals ε are assumed to follow a normal distribution with a mean of 0 and a variance of σ^2, so that the majority of residuals have values close to 0 (i.e., the error is very small), and their distribution is symmetrical. The errors are usually assumed to follow a multivariate normal distribution. If the errors do not follow a multivariate normal distribution, generalized linear models can be used to relax the assumptions. The Gauss-Markov assumptions and the normality assumptions ensure particularly interesting properties of the estimators of the model coefficients. It is also essential that the residuals are independent, meaning that there is no missing structure

in the model (such as spatial or temporal autocorrelation). In other words, each residual is independent of any other residual. We also assume the hypothesis of non-collinearity of the explanatory variables. This assumption supposes that none of the explanatory variables in the model can be written as a linear combination of the other variables. This condition is often expressed by the fact that the design matrix has the maximum rank. Note that if the non-collinearity assumption is not satisfied, the estimation of the model is impossible (it would require inverting a singular matrix), whereas for all other assumptions, the estimation is possible but yields a biased and/or inefficient estimator (with non-minimal variance), but there are possible corrections. The normality of errors is not obligatory, but it allows drawing good properties.

One of the most used statistical methods in applied sciences is the method of linear regression. It allows evaluating how an increase in one variable is associated with a more or less significant effect on the increase or decrease of another variable. Its origin dates back to the method of least squares errors Legendre (1805). It is then associated with the correlation between two normal distributions by Galton (1886) in the case of simple regression. The extension to partial correlations of a dependent variable with at least two other variables was invented by Yule (1897) . It is called multiple regression. For this purpose, we must first discuss the question of statistical inference:How can we derive,from a specific number of observations ,a certain probability regarding the relationships that may exist among multiple variables . an inference method adapted to the regression method was proposed by Fraser (1991). For example, the elimination of explanatory variables with zero simple correlation with the dependent variable is the starting point for selecting explanatory variables in a recent version of Spirtes et al. (2000) algorithm proposed by Bühlmann et al. (2010).

The theoretical model is a family of functions $f(x;\theta)$ of one or more variables x, indexed by one or more unknown parameters θ. The least squares method allows selecting among these functions the one that best reproduces the experimental data. In this case, it is called adjustment by the least squares method. The model is written as:

$$y = f(x;\theta) + \varepsilon$$

If the parameters θ have a physical meaning, the adjustment procedure also provides an indirect estimation of the value of these parameters.

Once the model is posed, the next question is the estimation of the unknown parameters of the model. The parameters come in two types: those related to the expectation and contained in the vector θ, and the parameter σ^2 which measures the variability that remains when the total variability of the observations is removed from everything explained

by the model.

The linear model can be estimated by different methods : maximum likelihood, the least squares method, the method of moments, or by Bayesian methods. The least squares method is very popular and often presented with the linear model. To fit the model, the most common method to minimize the variation of a model is to sum the squared residuals. This method is called Ordinary Least Squares (OLS). We seek the estimators that minimize the sum of the squares of the distances from the observations to the fitted curve. The function f is convex, so the parameters are obtained by setting the partial derivatives of the function f to zero, which gives a system of equations with multiple unknowns. The "best" parameters are those that minimize the variation in the response variable. The most common method to minimize the variation of a model is to sum the squared residuals. This method is called Ordinary Least Squares (OLS).

In addition, to Ordinary Least Squares (OLS), the Maximum Likelihood (ML) method also allows estimating the parameters of a model, assuming that the true distribution is known. If the principle for OLS is to find the parameter that minimizes the sum of the squared errors, the Maximum Likelihood method seeks to find the parameter with the highest probability of reproducing the true values of the sample (those actually observed), i.e., finding the most likely value of the parameter of a population based on a given sample. In other words, the ML method is based on the idea that if we are faced with possible different values for a parameter, we will choose the value with which the model would most probably generate the observed sample.

Several works have been carried out on the models described above particularly on the regression model . Some of them include Erkel-Rousse and Le Gallo (2002), focused more specifically on the detection of multicollinearity using indicators proposed by D.A. Belsley (1991) , the condition number and the variance decomposition table. Courcoul et al. (2010) introduced an estimation algorithm for any pattern of missing data. For more details see Legendre (1805), Galton (1886) ,Yule (1897) , Spirtes et al. (2000),Bühlmann et al. (2010) Erkel-Rousse and Le Gallo (2002), Belsley (1991).

The second case, Principal Component Analysis (PCA), is a classic technique in data analysis for exploratory study or compression of large quantitative data tables. The books by Jolliffe (1990) and Diday (1982) . Let p real statistical variables X^j $(j = 1, ..., p)$ be observed on n individuals $(i = 1, ..., n)$, these measurements are grouped in a matrix X of order $(n \times p)$. The matrix X denotes the data table resulting from the observation of p quantitative variables X^j on n individuals.

PCA plays a crucial role in serving as theoretical; this method serves as a theoretical foundation for other factorial multidimensional statistical methods. From a mathematical point of view, PCA corresponds to approximating a matrix of order (n, p) by a matrix of

the same dimensions but of rank $q < p$. PCA is also the search for q normalized linear combinations of X^j, uncorrelated, and whose sum of variances is maximal.

The first axis is the one that should best summarize the multidimensional shape of the cloud. It is a one-dimensional vector subspace on which all points are projected, generated by a director vector u passing through the origin, and whose coordinates are sought.

The objectives pursued by PCA are:

- Optimal graphical representation of individuals (rows of table X), minimizing deformations of the cloud of points, in a subspace of dimension $n \times q$ (with $q < p$),
- Graphical representation of variables in a subspace, explicitly highlighting the initial relationships between these variables,
- Dimension reduction or approximation of X by a table of rank q ($q < p$).

Principal Component Analysis (PCA) is a tool for compressing and synthesizing the information contained in a table, very useful when dealing with a large amount of quantitative data to process and interpret.

The correlation coefficient ranges between 0 (not correlated at all) and 1 (strongly correlated). If this value is close to 1, then the point is well represented on the axis. Points located near the center are generally poorly represented by the factorial plane. Their interpretation cannot be done with confidence.

Limitations of PCA: As PCA uses the correlation coefficient, it can only measure linear relationships between variables.

This thesis is divided into three chapters.

1. First, we have presented a general introduction where some previous works were cited, and our contributions in the studied themes were presented.

2. A development of the statistical methods used in the document is proposed in the First chapter.

3. Then, in chapter 2, we study regression models without predictors to see what they can tell us about the nature of the constant term. Understanding the constant term in these simpler models will help us understand both the constant term and other regression coefficients in more complex models. Continuous variables can deviate from normality in terms of asymmetry, kurtosis, and even higher-order moments. Many applications have variable positive response effects. These variables usually

have right-skewed distributions. We propose a linear regression model only at the origin under the assumption of non-normality. Here, we assume that the errors follow the exponential distribution.

4. Regression methods and Principal Component Analysis for the statistical analysis of data on cereal production in Algeria are introduced in chapter 3. The aim of this chapter is to analyze parameters such as precipitation and irrigation that influence crop yield and establish a relationship between these parameters. The data used are cereal production, irrigation (irrigated areas), and rainfall. These data have undergone statistical analysis. Firstly, Principal Component Analysis was applied to classify the data to determine the relative importance of different regions in evaluating cereal production. Secondly, a regression analysis was used to analyze the factors and their effects on crop yield. This work highlights the estimation of agricultural production. Particularly concerning the development of regression models using climatological parameters as independent variables and crop yield as the dependent variable. We then discuss the importance of forecasting for estimating crop yield. In this context, the document also covers Principal Component Analysis, a statistical procedure used to reduce a number of correlated variables to a smaller number of uncorrelated variables called principal components. These hidden latent variables are called factors or components, hence the name analysis.

5. We finish with a conclusion and perspectives.

Chapter 1

Linear Model

1.1 Definitions and Estimations

1.1.1 Definition

This chapter introduces the concept of a linear model: explaining Y as an affine function of X. If we assume that there is a cause-and-effect relationship between X and Y, the random phenomenon represented by X can be used to predict that represented by Y. The purpose of such a model is multiple and depends on the context. It could be to seek an answer to a question like: Does a quantitative variable X have an influence on the quantitative variable Y? Or to find a predictive model of Y based on X. The model assumes that, on average, $E(Y)$ is an affine function of X.

Remark. For simplicity, we assume that X is deterministic. In the case of X is random, the model is written conditionally on the observations of X.

The following definition holds:

Definition 1.1. A p-multidimensional linear model (X, θ, V) is defined as a random variable Y taking values in $\mathbb{R}^{np}$ such that:

- $\underset{n\times p}{EY} = \underset{n\times k}{X} \quad \underset{k\times p}{\theta}$ where X is a linear mapping $\mathbb{R}^k \to \mathbb{R}^n$, and θ is a parameter in $\mathbb{R}^{kp}$.
- The rows of Y are pairwise uncorrelated and have variance V.

If we write

$$Y = X\theta + \varepsilon,$$

ε is a random variable with zero mean, and its rows ε_i are uncorrelated and have variance V.

If the variable X can be used to predict Y, we are led to search for a prediction formula of Y given X, of the form $\widehat{Y} = f(X)$, with no bias $E(Y - \widehat{Y}) = 0$, and to evaluate the magnitude of the prediction error measured by the variance of $\varepsilon = Y - \widehat{Y}$. We naturally seek to minimize this variance. In our theoretical study, we will look for the ideal prediction formula, in the least squares sense, particularly if this formula is linear. In this type of model, X will be called an explanatory variable, and Y will be called a response variable.

So, we are interested in a random variable Y that depends on a random vector X in $\mathbb{R}^{np}$, and we seek to explain the variation of Y based on the variations of X. For this purpose, we consider a linear model where $E(Y|X)$ is linear. We denote ε as the random variable that represents the noise associated with the model. As in simple regression, we have $E(\varepsilon|X) = 0$ and $E(\varepsilon) = 0$. Additionally, we assume that the variance σ^2 of ε does not depend on X.

Remark. The model is said to be Gaussian if $Y \sim \mathcal{N}_{np}$; then the rows of Y are pairwise independent.

Throughout, we will assume that the model is regular, i.e. $\mathbf{X}$ has maximal rank k, which requires $n \geq k$.

Remark. The singular case can be treated similarly to the unidimensional model, by introducing constraints $C\theta = 0$ or using $(X'X)^{-1}$.

Moreover, we will see later that the support of V is almost surely determined if $n \geq p + k$; thus, if $n \geq p + k$, we can assume that $r = p$, and therefore, $|r| \neq 0$.

Let $L_{p\times r}$ be linear: $\mathbb{R}^P \to \mathbb{R}^n$. Then

$$YL = X\theta L + \varepsilon L,$$

is a r-dimensional linear model with parameters $\{\theta L, L'VL\}$. Indeed,

$$Var(\varepsilon_i L) = L'VL,$$

and

$$\operatorname{Cov}(\varepsilon_i L, \varepsilon_j L) = L' \operatorname{Cov}(\varepsilon_i, \varepsilon_j) L = 0,$$

for $i \neq j$.

In particular, if $Y = (U_1, \ldots, U_p), \theta = (\theta_1 \ldots, \theta_p), \varepsilon = (e_1 \ldots, e_p)$, then

$$y_j = x\theta_j + e_j,$$

for $j = 1, \ldots, p$, are unidimensional linear models, which is crucial for practical calculations.

The parameters of the model are θ and σ^2. Adjusting the model consists of estimating these parameters. The parametrization $E(Y) = X\theta$ is identifiable if the decomposition $E(Y) = X\theta$ with respect to the exogenous vectors $X_1, X_2, \cdots, X_p$ is unique. In this case, θ_k can be interpreted as the coordinate of $E(Y)$ with respect to the k-th exogenous variable X_k. This condition is equivalent to $X_1, X_2, \cdots, X_p$ being linearly independent in $\mathbb{R}^n$, or equivalently, X being full rank equal to p. The vector subspace $E(X)$ of $\mathbb{R}^n$ spanned by $X_1, X_2, \cdots, X_p$ is then of dimension p, and it is the space to which the mean $E(Y)$ belongs.

Nonidentifiability: Without identifiability, the representation $E(Y) = X\theta$ in θ is not unique, and θ is neither interpretable nor estimable. To make a parameterization identifiable, it suffices to select a subfamily of linearly independent regressors that span the space of the mean $E(X)$.

After specifying the necessary assumptions and the model terms, the concepts of parameter estimation in the model, prediction by confidence interval, and the significance of hypothesis tests **will be** discussed.

1.1.2 Estimation

1.1.2.1 Ordinary Least Squares Estimation

The estimation of the parameters in this model is based on n simultaneous observations of the variables X^j and Y performed on n individuals assumed to be independent. It will be assumed throughout that the model is identifiable. The estimation of β using Ordinary Least Squares (OLS) is a value $\widehat{\theta}$ that minimizes the sum of squared residuals:

$$SC(\beta) = \|Y - X\theta\|^2 = \sum_{i=1}^{n} (y_i - x_i\theta)^2.$$

The function $\beta \to SC(\theta)$ is strictly convex. In fact, the matrix

$$\frac{\partial^2 SC(\theta)}{\partial \theta^2} = 2 \times \ X^t X,$$

is positive definite since X is full rank. Hence, $\widehat{\theta}$ is unique, and the gradient of $SC(\theta)$ vanishes:

$$\frac{\partial SC(\widehat{\theta})}{\partial \theta} = 0 \Rightarrow X^t X\widehat{\theta} = X^t Y,$$

X^tX being invertible, the OLS estimation is:

$$\widehat{\theta} = (X^tX)^{-1}X^tY,$$

$E(Y) = X\theta$ belongs to $\mathcal{E}_X$, the space of the mean of Y. Since $\widehat{\theta}$ is the unique value minimizing $\|Y - X\theta\|^2$, $\widehat{Y} = X\widehat{\theta}$ is the orthogonal projection of Y (a vector in $\mathbb{R}^n$) onto $\mathcal{E}_X$ (a subspace of dimension p).

The only assumptions used so far concern the expectations, variances, and covariances of the variables (y_i). To dive further in the statistical study, for example, to test a sub-hypothesis, validate a model, or construct a confidence interval on a parameter, an additional assumption specifying the distribution of y_i must be added. Gaussian models allow answering these questions.

Furthermore, we obtain a very simple expression for its covariance matrix $Var(\widehat{\theta})$. Recall that the covariance matrix of the random vector $\widehat{\theta}$, also known as the variance-covariance matrix or dispersion matrix, is defined as:

$$Var(\widehat{\theta}) = E((\widehat{\theta} - E\widehat{\theta})(\widehat{\theta} - E\widehat{\theta})') = E(\widehat{\theta}\widehat{\theta\prime}) - E(\widehat{\theta})E(\widehat{\theta})'.$$

Since β is of dimension p, its covariance matrix is of dimension $p \times\ p$. Additionally, for any matrix A of size $m \times\ p$ and any deterministic vector B of dimension m, we have:

$$Var(A\widehat{\beta} + B) = AVar(\widehat{\theta})A'.$$

1.1.2.2 Maximum Likelihood Estimation

Let's assume that rank V = p.

Then, in the Gaussian case, the estimator of the maximum likelihood $(\hat{\theta}, \hat{V})$ maximizes:

$$\log L = -\frac{n}{2}\log 2\pi + \frac{n}{2}\log|V|^{-1} - \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_i\theta)\,V^{-1}\,(Y_i - X_i\theta)'$$

where Y_i and X_i respectively represent the i^{th} rows of Y and X.

Let:

$$\log L = -\frac{n}{2}\log 2\pi + \frac{n}{2}\log|V|^{-1} - \frac{1}{2}\operatorname{tr} V^{-1}(Y - X\theta)'(Y - X\theta).$$

Thus:

$$\frac{\partial}{\partial\theta}\log L = -\hat{V}^{-1}(Y - X\theta)'\mathrm{X} = 0$$

$$\Rightarrow$$

$$X'X\hat{\theta} = X'Y.$$

Note that $V^{-1} = (v^{ij})$ and observe that:

$$\begin{cases} \frac{\partial}{\partial v^{ij}} \log|v^{-1}| = \frac{1}{|v^{-1}|}\frac{\partial |v^{-1}|}{\partial v^{ij}} = \begin{cases} 2v^{ij} & i \neq j \\ v^{ii} & i = j \end{cases} \\ \frac{\partial}{\partial v^{ij}} \operatorname{tr}(V^{-1}S) = \begin{cases} 2s_{ij} & i \neq j \\ s_{ij} & i = j \end{cases} \quad \text{where} \quad S = (Y - x\hat{\theta})'(Y - x\hat{\theta}); \end{cases}$$

We then obtain:

$$\frac{n}{2}v_{ij} = \frac{1}{2}s_{ij'},$$

or

$$\hat{v} = \frac{1}{n}s.$$

Thus, in the regular case:

$$\hat{\theta} = (X'X)^{-1}X'Y,$$

and

$$\hat{V} = \frac{1}{n}(Y - X\widehat{\theta})'(Y - X\hat{\theta}).$$

Remark. It is known that, if f is sufficiently regular, the maximum likelihood estimator $\widehat{f(\theta)}$ of $f(\theta)$ is equal to $f(\hat{\theta})$.

In particular, we will obtain the maximum likelihood estimators of the eigenvalues of V, [of the canonical correlations associated with a partition $\left(X_{(1)}, X_{(2)}\right)\ldots$] by substituting $\frac{1}{n}S$ for V in the algorithms.

Proposition 1.1. *Let V be of rank p; $(\hat{\theta}, S)$ is sufficient for (θ, V), and minimal sufficient in the absence of constraints on θ and V, meaning θ varies in $\mathbb{R}^{kp}$ and V in the open cone of positive definite symmetric $p \times p$ matrices.*

The exponent of the density is written as follows:

$$-\frac{1}{2}\operatorname{tr}\left(V^{-1}(Y - X\theta)'(Y - X\theta)\right) = -\frac{1}{2}\operatorname{tr}\left(V^{-1}\left((Y - X\hat{\theta})'(Y - X\hat{\theta}) - \hat{\theta}'X'X\theta - \theta'X X\hat{\theta} + \theta'X'X\theta\right)\right)$$

and we conclude by the factorization theorem. The minimal complete property follows from the theory of regular exponential families.

Definition 1.2. Let $Y_1 \dots Y_r$ be independent variables with respective laws $\mathcal{N}_p(n_i, V)$, and let $M = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_r \end{pmatrix}$. Then

$$\underset{p\times p}{W} = \sum_{i=1}^{r} Y_i' Y_i,$$

follows a non-central Wishart distribution with r degrees of freedom, denoted as $W_p'(r, V, M)$.

Example 1.1. *The centered Wishart distribution is obtained when $M = 0$, denoted as $W_p(r, V)$, and it generalizes the X' and χ^2 distributions.*

We no longer assume that $\underline{\text{rank}(V) = p}$.

Theorem 1.1. *If $Y = X\theta + \varepsilon$ is a Gaussian linear model, then:*

1. *$\hat{\theta} \sim \mathcal{N}_{kp}(\theta, V \otimes (X'X)^{-1})$, i.e., $Cov(\hat{\theta}_j, \hat{\theta}_h) = v^{jh}(X'X)^{-1}$,*
2. *$\hat{\theta}$ and S are independent,*
3. *$S \sim W_p(n-k, V)$.*

Then:

1. $\hat{\theta} = (X'X)^{-1} X'Y$, thus $\hat{\theta} \sim \mathrm{N_{kp}}$; with $E\hat{\theta} = (X'X)^{-1} X'EY = \theta$;

$$\begin{aligned} \operatorname{Cov}\left(\hat{\theta}_j, \hat{\theta}_h\right) &= \operatorname{Cov}\left((X'X)^{-1} X'U_j, (X'X)^{-1} X'U_h\right) \\ &= (X'X)^{-1} X' \operatorname{Cov}(U_j, U_h) X (X'X)^{-1} \\ &= (X'X)^{-1} v^{jh}. \end{aligned}$$

2. Let U be the subspace of $\mathbb{R}^n$ image of $\mathbb{R}^k$ under X.

   As $\mathbb{R}^k$ and $\mathbb{R}^n$ have their respective canonical bases, let y_j be the vector with coordinates U_j and y_j^* be the one with coordinates $x\hat{\theta}_j, j = 1, \dots, p$.

   Then, y_j^* is the orthogonal projection of y_j onto V, similar to the one-dimensional case, and

$$s = \left\langle y_j - y_j^*, y_h - y_h^* > \right).$$

Let $(e_i)_{i=1,\dots,n}$ be an orthonormal basis of $\mathbb{R}^n$, where $(e_1, \dots, e_k)$ forms a basis of U; in this basis, y_j and y_j^* respectively have coordinates $z_{ij}, i = 1, \dots, n$, and $z_{ij}, j = 1, \dots, k$, with 0 for $j > k$. Thus

$$S = \sum_{k+1}^{n} z_i' z_i \quad ,$$

when we denote $z_i = (z_{il}, \dots z_{ip})$.

Since $Z = P'Y$ where P is the orthogonal change of basis matrix from the canonical basis to (e_i), the rows of Z are independent with covariance matrix V, which implies the independence of

$$X\hat{\theta} = P \begin{pmatrix} z_1 \\ z_k \\ 0 \end{pmatrix} \quad \text{and} \quad S = \sum_{i=k+1}^{n} z_i' z_i.$$

3. $EZ = EP'Y = P'X\theta = EP'X\hat{\theta} = \begin{pmatrix} EZ_1 \\ EZ_k \\ 0 \end{pmatrix}$, so $EZ_i = 0$ for $i > k$; hence $S \sim W_p(n-k, V)$.

The maximum likelihood estimators are equivalent to the Least Squares estimators of β. This can be shown in the case of linear regression. However, some properties are only valid under the assumption of normality of the residuals.

In the case of a sample of size n, we obtain in particular:

Corollary 1.1. *$(\bar{X}, S)$ is exhaustive for (ξ, V), and minimal exhaustive in the absence of prior assumptions on (ξ, V).*

Corollary 1.2. *(Multidimensional Fisher's theorem).*

1. *$\bar{X} \sim \mathcal{N}_p\left(\xi, \frac{1}{n}V\right)$;*
2. *$S \sim W_p(n-1, V)$;*
3. *$\bar{X}$ and S are independent.*

We have:

Proposition 1.2. *In the general (non-Gaussian) case, $\hat{\theta}$ and $\frac{1}{n-k}S$ are unbiased estimators of θ and V. In the Gaussian case, they are furthermore optimal among unbiased estimators (in the absence of constraints on θ and V, with rank $V = p$ or $n - k \geq p$).*

The demonstration of the previous theorem shows that, in the general case, the Z_i are uncorrelated, with zero mean and variance V.

Thus,

$$E\hat{\theta} = (X'X)^{-1}X'X\theta = \theta,$$

and

$$ES = (n-k)EZ_i'Z_i = (n-k)V.$$

The optimality in the Gaussian case follows from the completeness property of the statistics $(\hat{\theta}, \mathrm{S})$.

Remark. An analogue of the Gauss-Markov theorem can be shown in the general non-Gaussian case.

1.1.2.3 Confidence Intervals and Regions

Software and some references provide confidence intervals (CIs) for the parameters taken separately. However, these confidence intervals do not take into account the dependence of the parameters, which would lead to constructing confidence regions (CR) instead.

Assuming the validity of known assumptions (random fluctuations of the explained variable y following a normal distribution with constant variance regardless of the value of the explanatory variable X), the least squares method allows us to estimate the average value of Y for a given x and have, in the form of $y = f(x)$, an average estimation law of y as a function of x.

The obtained law is based on a sample; hence the result is random and only an estimation of the real law (of suitably chosen form) that best describes this relation in the infinite population that can be conceptualized. Therefore, it is necessary to specify, regarding this average estimation:

1. the uncertainty related to the knowledge of this estimation line itself,
2. ε the possible fluctuations of the variable y for a given x, around this line.

One of the purposes of studying the correlation that may exist between two variables is to forecast the average value of one of them for a given value of the other. A study of numerical data allows us to specify this through the determination of a curve.

1.2 Hypothesis Testing on the Columns of θ

Let ω be a vector subspace of $\mathbb{R}^k$ of dimension h; we denote H_ω as the hypothesis "$\theta_j \in \omega, \forall j = 1, \ldots, p,$" which is linear in the columns of θ. By translating the v_j, we can

transform an affine hypothesis into a linear one.

1.2.1 Estimation of θ and V under the hypothesis H_ω

Let $\hat{\theta}_\omega$ and $\widehat{V}_\omega$ (or equivalently $S_\omega = n\hat{V}_\omega$) be the maximum likelihood estimators of θ and V under H_ω. If ω is the image of $R : \mathbb{R}^h \to \mathbb{R}^k - R$ with rank h, then we have $\theta = R\xi$ under $H_{\omega'}$, which leads to the model:

$$Y = XR\xi + \varepsilon.$$

In this new linear model, we obtain:

$$\begin{cases} \hat{\xi}_\omega = (R'X'XR)^{-1} R'X'Y, \\ n\hat{V}_\omega = S_\omega = \left(Y - XR\hat{\xi}_\omega\right)' \left(Y - XR\hat{\xi}_\omega\right). \end{cases}$$

Thus,

$$\begin{aligned} \hat{\theta}_\omega &= R\left(R'X'XR\right)^{-1} R'X'Y, \\ S_\omega &= \left(Y - X\hat{\theta}_\omega\right)' \left(Y - X\hat{\theta}_\omega\right). \end{aligned}$$

If ω is defined as the kernel of $C : \mathbb{R}^k \to \mathbb{R}^{k-h}$ with rank $k - h$, then we need to estimate θ under the constraint $C\theta = 0$, so we maximize

$$\log\left(L + \frac{1}{2}\operatorname{Tr} C\theta\phi'\right),$$

where $\phi_{(k-h)\times p}$ is a Lagrange multiplier.

By differentiating with respect to θ_{ij}, we obtain:

$$nX'X\hat{\theta}_\omega S_\omega^{-1} - nX'Y S_\omega^{-1} + C'\phi = 0,$$

which leads to:

$$\hat{\theta}_\omega = \left(X'X\right)^{-1} X'Y - \left(X'X\right)^{-1} C'\phi\frac{S_\omega}{n}.$$

From $C\hat{\theta}_\omega = 0$, we get

$$\frac{1}{n}\phi S_\omega = \left(C\left(X'X\right)^{-1} C'\right)^{-1} C\hat{\theta},$$

where $C\left(X'X\right)^{-1}C'$) is invertible since C has rank $k - h$. Thus,

$$\hat{\theta}_\omega = \left[I_k - \left(X'X\right)^{-1} C'\left(C\left(X'X\right)C'\right)^{-1} C\right]\hat{\theta}.$$

1.2.2 Testing the Hypothesis H_ω in the Gaussian case.

ω Let V and V_ω be the respective images of $\mathbb{R}^k$ and ω under X in $\mathbb{R}^n$ ($\mathbb{R}^k$ and $\mathbb{R}^n$ being equipped with their canonical bases). We can choose an orthonormal basis $\{e_i\}_{i=1,\dots,n}$ of $\mathbb{R}^n$ such that $\{e_i\}_{i=1,\dots,h}$ and $\{e_i\}_{i=1,\dots,k}$ are respectively bases of V_ω and V.

If P denotes the orthogonal matrix for the change of basis from the canonical basis of $\mathbb{R}^n$ to the basis $\{e_i\}$, and $Z = P'Y$, then we have:

$$Z = \underbrace{\begin{pmatrix} \xi_1 \\ \xi_k \\ 0 \end{pmatrix}}_{\mathrm{p}} {}_{n-k} + \varepsilon^*,$$

where $\varepsilon^* = P'\varepsilon$. Thus, the rows of ε^* are independent, with variance V and expectation $E\varepsilon^* = \mathrm{P'E}\varepsilon = 0$.

The hypothesis H_ω is then equivalent to " $\xi_{\mathrm{h}+1} = \ldots = \xi_{\mathrm{k}} = 0$ ".

We can restrict ourselves to look for tests that are functions of the exhaustive statistic $\left(Z_1, \dots, \mathrm{Z}_k, \sum_{i=k+1}^{\hbar} Z_i' Z_i = \mathrm{S}\right)$, where $Z_1, \dots, Z_k$ and S are independent. Moreover, comparing with the search for the test F, we can see that:

1. Tests invariant under translations of $\mathbb{R}^k$ that leave ω invariant are functions of (Z_{h+1}, Z_k, S),
2. Tests invariant under orthogonal transformations of v_ω depend only on $B_\omega = \sum_{i=h+1}^{k} Z_i' Z_i = S_\omega - S$ and S,
3. Tests invariant under regular transformations of $\mathbb{R}^p$ (homotheties in the case $p = 1 \dots$) are only functions of the roots λ of $|B_\omega - \lambda S| = 0$ (i.e., the eigenvalues of $B_\omega S^{-1}$).

Indeed, through the regular transformation $\Gamma_{p\times p}$, $\mathrm{Z}_i^* = \mathrm{Z}_i\Gamma$, $B_\omega^* = \Gamma' B_\omega \Gamma$, and $S^* = \Gamma' S\Gamma$, thus $|\Gamma' B_\omega \Gamma - \lambda \Gamma' \mathrm{S}\Gamma| = |\Gamma|^2 |\, B_\omega - \lambda S \,|$.

Conversely, suppose that $|B_\omega - \lambda S| = 0$ and $|B_\omega^* - \lambda S^*| = 0$ have the same roots, i.e., $B_\omega S^{-1}$ and $B_\omega^* \mathrm{S}^{*-1}$ have the same eigenvalues. There exist regular matrices T and T^* such that: $\mathrm{TST}' = I_p = \mathrm{T}^*\mathrm{S}^*\ \mathrm{T}^*, \mathrm{TBT}' = \Lambda$ diagonal and $\mathrm{T}^*\ \mathrm{B}^*\ \mathrm{T}^{*'} = \Lambda^*$ diagonal (since $S, B_\omega, S^*, B_\omega^*$ are symmetric...).

We have $\Lambda = T B_\omega T' \underbrace{\left(T^{-1} S^{-1} T^{-1}\right)}_{I_P} = T B_\omega S^{-1} T^{-1}$ thus the diagonal elements of Λ are the eigenvalues of $B_\omega S^{-1}$; the same applies to Λ^*.

The conclusion is $\Lambda = \Lambda^*$, up to a permutation of the eigenvalues, hence a specific choice of T and T^*. Consequently, $S^* = \Gamma' S \Gamma$ and $B^*_\omega = \Gamma' B_\omega \Gamma$ by taking $\Gamma = (T^{-1}T^*)'$. Thus, we obtain:

Theorem 1.2. *The tests of the hypothesis* Hω *invariant under transformations* $1, 2, 3$*, are functions solely of the roots of* $|B\omega - \lambda S| = 0$ *where* $B_\omega = S_\omega - S = \left(\hat{\theta}\omega - \hat{\theta}\right)' X'X \left(\hat{\theta}\omega - \hat{\theta}\right)$ *is independent of* S *and follows the law* $W_p^{\prime(k-h,V,M)}$ (*with* $M = 0$ *under* H_ω).

Remark. Under H_ω, the set of roots $(\lambda_1, \ldots, \lambda_p)$ of $|B_\omega - \lambda S| = 0$ is a freen statistic for (θ, V). indeed, under $H_\omega{}'$, $B_\omega \sim W_p(k-h, V)$, $S \sim W_p(n-k, V)$ and they are independent, hence

$$V^{-\frac{1}{2}} B_\omega V^{-\frac{1}{2}} \sim W_p\left(k-h, I_p\right) \text{ and } V^{-\frac{1}{2}} S V^{-\frac{1}{2}} \sim W_p\left(n-k, I_p\right)$$

and the λ_i are solutions of $\left|\mathrm{V}^{-\frac{1}{2}}\ \mathrm{BV}^{-\frac{1}{2}} - \lambda \mathrm{V}^{-\frac{1}{2}} \mathrm{SV}^{-\frac{1}{2}}\right| = 0$.

As $n \longrightarrow +\infty$, $\frac{1}{n} S \xrightarrow{P} V$, and the $n\lambda_i$, eigenvalues of $nS^{-\frac{1}{2}} B S^{-\frac{1}{2}}$, converge in probability and in distribution to the eigenvalues of $V^{-\frac{1}{2}} B_\omega V^{-\frac{1}{2}} \sim W_p\left(k-h, I_p\right)$.

Remark. There will be a uniformly most powerful test among the invariant tests when only one root λ is non-zero, i.e.,

$$1 = \operatorname{rank} B_\omega S^{-1} = \operatorname{rank} B_\omega = \operatorname{rank} \begin{pmatrix} Z_{h+1} \\ \vdots \\ Z_k \end{pmatrix} = \inf(p, k-h).$$

- The case $p = 1$ is the unidimensional model, and $\lambda = \frac{B_\omega}{S} = \frac{S_\omega - S}{S}$; thus, we obtain the test F.
- In the case $p > 1$, we distinguish the situations $k - h = 1$ and $k - h > 1$; in the second situation, the tests will be based on real functions ψ that are increasing in the roots $\lambda_1, \ldots, \lambda_p$ of $|B_\omega - \lambda S| = 0$.

1. <u>The case $k - h = 1$.</u> Hotelling's Test.

The unique non-zero eigenvalue is $\operatorname{tr}\left(B_\omega S^{-1}\right) = \operatorname{tr}\left(Z_k' Z_k S^{-1}\right) = Z_k S^{-1} Z_k'$ and it is known that

$$Z_k S^{-1} Z_k' \times \frac{n-k-p+1}{p} \sim F'\left(p; n-k-p+1, \xi_k V^{-1} \xi_k'\right) \quad .$$

Hence:

Corollary 1.3. *In the case $k-h=1$, the test for hypothesis H_ω defined by the rejection region*

$$\left\{ Z_k S^{-1} Z_k' \geq \frac{p}{n-k-p+1} F_{\alpha,p,n-k-p+1} \right\},$$

is UPP invariant of level α, and called T^2-test or Hotelling's Test.

One can derive a test for affine hypotheses, a confidence region. More generally, we have:

2. The case $k-h>1$. Hotelling-Lawley test.

Let

$$\psi_{HL}(\lambda_1,\ldots,\lambda_p)=\sum_{i=1}^{p}\lambda_i=\mathrm{tr}(B_\omega S^{-1}).$$

Under $\mathrm{H}_{\omega'}$, the distribution of ψ_{HL} depends only on $n-k, p, k-h$, and we can define $tr_{\alpha,p,n-k,k-h}$ such that $P_{H_\omega}(\psi \geq tr_{\alpha,p,n-k,k-h})=\alpha$; Moreover

$$n\times\psi_{HL} \underset{\substack{\mathrm{H}_\omega \\ \mathrm{n}\to+\infty}}{\xrightarrow{\mathcal{L}}} \mathrm{tr}\left(\mathrm{V}^{-1/2}\mathrm{B}_\omega\mathrm{V}^{-1/2}\right)$$

has the $\chi^2(p(k-h))$ distribution $\left(\text{since}\mathrm{V}^{-1/2}\mathrm{B}_\omega\mathrm{V}^{-1/2}\sim \mathrm{W_p(k-h},I_p)\right)$, hence the asymptotic approximation of $tr_{\alpha,p,n-k,k-h}\cdot$ can be concluded:

Proposition 1.3. *A level α test of the hypothesis H_ω is defined by the rejection region*

$$\left\{ tr(B_\omega S^{-1}) \geq tr_{\alpha,p,n-k,k-h} \right\}.$$

Let $\underset{q\times p}{\varphi}=\underset{q\times k}{C}\underset{k\times p}{\theta}$, $V_{\hat{\varphi}}=C\left(X'X\right)^{-1}C'$, and $\hat{\varphi}=C\hat{\theta}$, then:

Corollary 1.4. *A level α test of the hypothesis $"\varphi=\varphi_0"$ is defined by the rejection region*

$$\mathrm{tr}\left((\hat{\psi}-\varphi_0)v_{\hat{\varphi}}^{-1}\left(\hat{\varphi}-\varphi_0\right)S^{-1}\right)\geq \mathrm{tr}_{\alpha,p,n+k,k-h}.$$

It suffices to verify that $B_\omega=\hat{\psi}\times V_{\hat{\varphi}}^{-1}\hat{\varphi}$. In the linear case $\varphi_0=0$, then we make the translation $\varphi\to\varphi-\varphi_0$.

If $f=(f_1,\ldots,f_p)$ is a $q\times p$ matrix, let $[f]=\begin{pmatrix} f_1 \\ \vdots \\ f_p \end{pmatrix}$ 1e $q\times p$ vector.

Corollary 1.5. *$[\psi]$ admits the confidence ellipsoid in $\mathbb{R}^{qp}$ at level α:*

$$R_\alpha(\hat{\varphi},S)=\left\{[f]\in\mathbb{R}^{\mathrm{qp}}/([\hat{\varphi}]-[f])'\left(\mathrm{S}\otimes V_{\hat{\varphi}}\right)^{-1}([\hat{\varphi}]-[f])<tr_\alpha\right\}.$$

Indeed, we have

$$P\left(\left\{f/\operatorname{tr}[(\hat{\varphi}-f))'V_{\hat{\varphi}}^{-1}(\hat{\varphi}-f)s^{-1}\right)\right] < \operatorname{tr}\alpha\right) \geq 1-\alpha,$$

and also

$$\begin{aligned}\operatorname{tr}\left[(\hat{\varphi}-f)'V_{\hat{\varphi}}^{-1}(\hat{\varphi}-f)S^{-1}\right] &= \sum_{ij}(\hat{\varphi}_i-f_i)'\,V_{\hat{\varphi}}^{-1}\,(\hat{\varphi}_j-f_j)\,S^{ij} \\ &= [\hat{\varphi}-f]'\cdot(S\otimes V_{\hat{\varphi}})^{-1}\,[\hat{\varphi}-f].\end{aligned}$$

Applying Scheffé's method to the above ellipsoid, we obtain:

Corollary 1.6. *When l ranges over the set of linear forms on $\mathbb{R}^{qp}$, we simultaneously have at level α the set of inequalities:*

$$|\ell'([\hat{\varphi}]-[\varphi])| < \left[\ell'\,(S\otimes V_{\hat{\varphi}})\,\ell\cdot\operatorname{tr}_{\alpha,p,n-k,q}\right]^{1/2}.$$

This corollary allows obtaining simultaneous confidence intervals for φ_{ij}; below we will obtain (cf. Roy's test) an even better result by restricting ourselves to decomposable linear forms (of the type $a\otimes b$).

3. The case $k-h>1$. Maximum likelihood test.

Proposition 1.4. *The maximum likelihood test of the hypothesis H_ω has the rejection region*

$$\left\{\frac{|S|}{|S+B|}\leq U_{\alpha,p,n-k,k-h}\right\},$$

and is an invariant test (Wilks' test).

Indeed, we have

$$\sup(L_{\theta,V}) = C(V)\frac{n^{n/2}}{|S|^{n/2}}\exp\left(-\frac{np}{2}\right),$$

and

$$\sup_{H_\omega}(L_{\theta,V}) = C(V'\times\frac{n^{n/2}}{|S_\omega|^{n/2}}\exp\left(-\frac{np}{2}\right)).$$

Hence,

$$\begin{aligned}\Lambda_n = \left(\frac{|S|}{|S_\omega|}\right)^{n/2} &= \left(\frac{|S|}{|S+B_\omega|}\right)^{n/2} = \left|I+B_\omega s^{-1}\right|^{-n/2} \\ &= \left[\prod_{i=1}^{p}(1+\lambda_i)\right]^{-n/2}.\end{aligned}$$

Asymptotically, we can use, under $H_{\omega'}$,

$$-2\log\Lambda_n \xrightarrow[n\to+\infty]{\mathcal{L}} X^2(p(k-h)).$$

We can also observe that, under H_ω,

$$\lambda_i \xrightarrow[n\to+\infty]{\mathcal{P}} 0,$$

then

$$n\log(1+\lambda_i) \underset{as}{\overset{\mathcal{L}}{\simeq}} n\lambda_i,$$

thus

$$-2\log(\Lambda_n) \underset{as}{\overset{\mathcal{L}}{\simeq}} n \times tr(B_\omega S^{-1}).$$

the Wilks and Hotelling-Lawley tests are asymptotically equivalent.

4. The case $k-h \geq 1$. Roy's test.

Let $\lambda_{\max} = \sup_{1\leq i\leq p} \lambda_i$ and $\lambda_{\max,\alpha,p,n-k,k-h}$ defined by $\mathrm{P}_{\mathrm{H}_\omega}\{\lambda_{\max} \geq \lambda_{\max,\alpha,\mathrm{p,n-k,k-h}}\} = \alpha$, then we have:

Proposition 1.5. *A level α test of H_ω is defined by the rejection region $\{\lambda_{\max} \geq \lambda_{\max,\alpha,p,n-k,k-h}\}$, called Roy's test.*

Let's review some notations: C is surjective, $\underset{q\times\mathrm{p}}{\varphi} = \underset{\mathrm{q\times k}}{\mathrm{C}} \quad \underset{\mathrm{k\times p}}{\theta}$, $V_{\hat{\psi}} = C\left(X'X\right)^{-1}C'$, and $\hat{\varphi} = C\hat{\theta}$.

Corollary 1.7. *The ellipsoids in $\mathbb{R}^q$:*

$$\{f \times b/fq \times p \text{ matrix and } b'(\hat{\psi}-f)'V_{\hat{\varphi}}^{-1}(\hat{\varphi}-f)b \leq b'Sb \times \lambda_{\max,\alpha}\},$$

are simultaneous confidence regions for the φ b at level α when b' ranges over the set of linear forms on $\mathbb{R}^p$.

Let $A = (\hat{\varphi}-\varphi_0)'{V_{\hat{\varphi}}}^{-1}(\hat{\varphi}-\varphi_0)$ and $B = S$;

$$\sup_{\mathrm{b}\in\mathbb{R}^P} \frac{b'Ab}{b'Bb} = \sup\left(\text{valeurs propres } AB^{-1}\right).$$

the refined test "$\varphi = \varphi_0$" associated with Roy's test admits the rejection region :

$$\left\{\sup_b \frac{b'Ab}{b'Bb} \geqslant \lambda_{\max,\alpha}\right\}$$

Then, the confidence region for φ at level is given by α

$$\left\{ \sup_b \frac{b'Ab}{b'Bb} < \lambda_{\max,\alpha} \right\}$$

hence

$$\frac{b'Ab}{b'Bb} < \lambda_{\max,\alpha} \quad \forall b.$$

Corollary 1.8. *the ellipsoids*

$$\xi_j = \left\{ f_j \in \mathbb{R}^q / \frac{(\hat{\varphi}_j - f_j)' \cdot V_{\hat{\varphi}}^{-1} (\hat{\varphi}_j - f_j)}{S_{jj}} \leqslant \lambda_{\max,\alpha} \right\}$$

are confidence region for $\varphi_1, \ldots, \varphi_p$ at level α simultaneously .

We choose $b_j = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \leftarrow j, \quad j = 1, \ldots, p$, in **corollary 1**.

Corollary 1.9. *The intervals*

$$\left\{ a' f b \in \mathbb{R} / \left| a'(\hat{\varphi} - f)b \right| \leq \left[(b'Sb) \times (a' V_{\hat{\varphi}} a) \, \lambda_{\max,\alpha,p,n-k,q} \right]^{1/2} \right\},$$

are simultaneous confidence regions for the $a'\varphi b$ at level α when a and b respectively range over $\mathbb{R}^q$ and $\mathbb{R}^p$.

Ab being fixed in **corollary 1**, we apply the inverse of Scheffé's method: for a fixed, the intersection of the bands in $\mathbb{R}^q$ when b varies is an ellipsoid in $\mathbb{R}^{\mathrm{q}}$.

Corollary 1.10. *Let* $\varphi = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_{\mathrm{q}} \end{pmatrix}$, *the ellipsoids in* $\mathbb{R}^{\mathrm{p}}$

$$\varphi_i = \{ g \in \mathbb{R}^p / \left(\hat{\psi}_i - g' \right) S^{-1} \left(\hat{\psi}_i - g' \right)' \leq v^{ii} \times \lambda_{\max,\alpha,p,n-k,q} \}$$

are simultaneous confidence regions for $\psi'_1, \ldots, \psi'_{\mathrm{q}}$ at level α.

It suffices to choose $a_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \dot{0} \end{pmatrix} \leftarrow i, i = 1, \ldots, q$, in **corollary 5**.

The confidence region obtained using the Hotelling-Lawley test allowed obtaining simultaneous confidence regions for <u>all</u> linear forms of $\psi_{ij'}$, and in particular the previous ones by substituting tr_α with $\lambda_{\max,\alpha}$; Roy's test allows obtaining simultaneous confidence intervals associated only with the decomposable linear forms $a \otimes b$. In return, these intervals are shorter and thus "better" since $\lambda_{\max,\alpha,p,n-k,q} \leq tr_{\alpha,p,n-k,q}$. Indeed, we have

$$\left\{ \sum_{i=1}^{p} \lambda_i \leq \mathrm{tr}_\alpha \right\} \subset \{\lambda_{\max} \leq \mathrm{tr}_\alpha\},$$

so

$$P(\lambda_{\max} \leq \mathrm{tr}_\alpha) \geq 1 - \alpha = P(\lambda_{\max} \leq \lambda_{\max,\alpha}).$$

Proposition 1.6. *The T^2-test of the hypothesis "$\theta_1 = \theta_2$" has, at level α, the rejection region:*

$$\left\{ \frac{n_1 n_2}{n_1 + n_2} (\bar{Y} - \bar{Z}) (S_Y + S_Z)^{-1} (\bar{Y} - \bar{Z})' \geqslant \frac{p}{n_1 + n_2 - p - 1} F_{\alpha,p,n_1+n_2-p-1} \right\}.$$

Given an additional variable $T \sim \mathcal{N}_p(\theta, V)$ where $\theta \in \{\theta_1, \theta_2\}$, we want to decide

$$\theta = \theta_1 \text{ or } \theta = \theta_2.$$

This is a <u>discrimination</u> problem. According to the Neyman-Pearson lemma, the most powerful test at a given level is to decide:

- $\theta = \theta_1$ if $\frac{L_{\theta_1,V^{(T)}}}{L_{\theta_2,V^{(T)}}} > k$ and
- $\theta = \theta_2$ otherwise.

$$\begin{aligned} \log\left(\frac{L_{\theta_1,\,\mathrm{V}^{(T)}}}{\mathrm{L}_{\theta_2,V^{(T)}}}\right) &= -\frac{1}{2}\left[(T-\theta_1)\,\mathrm{V}^{-1}\,(T-\theta_1)' - (T-\theta_2)\,\mathrm{V}^{-1}\,(T-\theta_2)'\right] \\ &= (\theta_1-\theta_2)\,\mathrm{V}^{-1}T' - \frac{1}{2}\left(\theta_1\,\mathrm{V}^{-1}\theta_1' - \theta_2\,\mathrm{V}^{-1}\theta_2'\right). \end{aligned}$$

Since the parameters θ_1, θ_2, and V are not known, we consider the "discriminant" function: $(\bar{Y} - \bar{Z})S^{-1}T'$, which is linear in T, and we decide

- $\theta = \theta_1$ if $(\bar{Y} - \bar{Z})S^{-1}T' > K$, and
- $\theta = \theta_2$ otherwise.

If we demand the equality of the Type I and Type II errors (symmetry of the problem), we can see that the hyperplane $(\bar{Y} - \bar{Z})S^{-1}T' = C$ is an oblique affinity hyperplane for the confidence ellipsoids of θ_1 and θ_2 respectively centered at $\bar{Y}$ and $\bar{Z}$.

4. Linear hypothesis tests in line. Potoff and Roy's Model.

Let $\mathrm{Y} = \mathrm{X}\theta + \varepsilon$ be a multidimensional linear model where $\underset{k\times p}{\theta} = \begin{pmatrix} t_1 \\ \vdots \\ t_k \end{pmatrix}$ and let $\bar{\omega}$ be a subspace of $\mathbb{R}^p$ with dimension r.

Definition 1.3. The hypothesis

$$H_{\bar{\omega}} : t_i \in \bar{\omega}, i = 1, \ldots, k,$$

is called a linear hypothesis in line.

Example 1.2. *Let $\{\mathrm{Y}_1, \ldots, \mathrm{Y}_n\}$ be an n-sample from the $\mathcal{N}_{\mathrm{p}}(\theta, \mathrm{V})$ distribution with $\theta = (\theta_1, \ldots, \theta_p)$; the hypothesis $\theta_1 = \ldots = \theta_p$ is a linear hypothesis in line with $\bar{\omega} = \{\underbrace{(X, \ldots, X)}_{p} / X \in \mathbb{R}\}$ of dimension 1.*

The subspace $\bar{\omega}$ can be defined as the image of an injective application $R : \mathbb{R}^r \longrightarrow \mathbb{R}^p$ or the kernel of a surjective application $T : \mathbb{R}^{\mathrm{P}} \longrightarrow \mathbb{R}^{\mathrm{p-r}}$.

The associated matrices (in canonical bases) will be respectively $r \times p$ and $p \times (p - r)$ as they are operators on the right (on rows).

Thus, under $H_{\bar{\omega}}$, the model can be written as

$$Y = \underset{k\times r}{X} \; \varphi \; R + \varepsilon,$$

from which:

Definition 1.4. The Potoff and Roy model (resp. Gaussian model) is given by n random p-vectors $Y_1, \ldots, Y_n$ that are uncorrelated (resp. independent Gaussian) with common variance V such that, if

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad EY = \underset{\substack{n\times k \\ n>k}}{X} \quad \underset{k\times r}{\varphi} \quad \underset{\substack{r\times p \\ \mathrm{p>r}}}{\mathrm{R}} \qquad \begin{array}{c} X : \mathbb{R}^k \to \mathbb{R}^n \text{ (on the left)} \\ R : \mathbb{R}^r \to \mathbb{R}^p \text{ (on the right).} \end{array}$$

We assume that X and R are injective (otherwise we add identifiability constraints). If $\bar{\omega}$ is defined as $\operatorname{Im} R$ and $\operatorname{Ker} T$, then there exist $D_{p-r\times p}$ and $E_{p\times r}$ such that

$$\begin{pmatrix} R \\ D \end{pmatrix} (ET) = (ET) \begin{pmatrix} R \\ D \end{pmatrix} = I_p.$$

In fact, $RT = 0$, we just need to choose $D = (TT')^{-1} T$ and $E = R' (RR')^{-1}$. Then

$$\begin{aligned} Y^* &= Y(E,T) \\ &= X\theta(E,T) + \varepsilon(E,T) \\ &= \left(Y^*_{(1)}, Y^{\pi}_{(2)}\right) \\ &= X(\varphi_1, \varphi_2) + \left(\varepsilon^*_{(1)}, \varepsilon^*_{(2)}\right), \end{aligned}$$

is an equivalent multidimensional linear model to the initial model.

$V^* = \begin{pmatrix} \mathrm{E}' \\ \mathrm{T}' \end{pmatrix} \mathrm{V}(E,T)$ in "the canonical" position for $\mathrm{H}_{\bar{\omega}}$ becomes here "$\varphi_2 = 0$".

a) Estimation in the Gaussian case.

In what follows, we assume a Gaussian model and we denote by:
$\hat{\theta} = (X'X)^{-1} X'Y$, $S = (Y - X\hat{\theta})'(Y - X\hat{\theta})$, $S^* = \begin{pmatrix} E' \\ T' \end{pmatrix} S(E,T)$, and $\begin{pmatrix} V^*_{11} & V^*_{12} \\ V^*_{21} & V^*_{22} \end{pmatrix}$, etc., the partitions associated with $\left(Y^*_{(1)}, Y^*_{(2)}\right)$, V^*, etc.

Proposition 1.7. *The maximum likelihood estimator $(\hat{\theta}_{\bar{\omega}}, S_{\bar{\omega}})$ of (θ, nV) under the hypothesis $H_{\bar{\omega}}$ is given by:*

$$\begin{aligned} \hat{\theta}_{\bar{\omega}} &= \hat{\theta} S^{-1} R' \left(RS^{-1}R'\right)^{-1} R \\ &= \hat{\theta}\left(I_p - T\left(T'ST\right)^{-1} T'S\right), \end{aligned} \tag{1.1}$$

and

$$S_{\bar{\omega}} = (Y - X\theta_{\omega})' (Y - X\theta_{\omega}).$$

We could directly calculate this (cf. Anderson). However, we prefer to start from the model in the canonical position and return to the initial model.

Lemma 1.1. *We have*

$$\mathrm{Y}^*_1 = \mathrm{X}\tau + \mathrm{Y}^*_2\beta + \mathrm{e}^*,$$

where Y^*_2 and e^* are independent, $e^* \sim \mathcal{N}_p(0, V^*_e \otimes I_n)$, $V^*_e = V^*_{11} - V^*_{12}V^{*-1}_{22}V^*_{21}$, $\tau = \varphi_1 - \varphi_2\beta$, and $\beta = V^{-1^*}_{22} V^*_{21}$.

Indeed, $\varepsilon_{i1}^* = \varepsilon_{i2}^*\beta + e_i^* \quad i = 1, \ldots, n$, and we conclude with $\varepsilon_2^* = \dot{Y_2^*} - X\varphi_2'$ using the results from Chapter 0 (**proposition 4**). We note that the parameter $(\varphi_1, \varphi_2, V_{11}^*, V_{12}^*, V_{22}^*)$ is in one-to-one correspondence with $(\tau, \varphi_2, \beta, V_{22}^*, V_e^*)$ in the model:

$$\begin{cases} Y_1^\star = X\tau + Y_2^*\beta + e^*, & \text{equivalent to the previous,} \\ Y_2^* = X\varphi_2 + \varepsilon_2^*, \end{cases}$$

Let $L_{\varphi_1,\varphi_2,V}(y_1^*, y_2^*)$, $L_{\tau,\beta,V_e^*}(y_1^*/y_2^*)$, and $L_{\varphi_2,V_{22}^*}(y_2^*)$ be the respective densities of (Y_1^*, Y_2^*), Y_1^* conditional on Y_2^*, and Y_2^*.

Then

$$\sup_{\varphi_1,\varphi_2,V} L(Y_1^*, Y_2^*) = \sup_{\tau,\beta,V_e^*\varphi_2,V_{22}^*} \sup L_{\varphi_2,V_{22}^*},$$

which implies

$$\hat{\tau} = \hat{\varphi}_1 - \hat{\varphi}_2\hat{\beta} \text{ and } \hat{\beta} = S_{22}^{*-1}S_{21}^* \ .$$

Under $H_{\bar{\omega}}$

$$\sup_{\varphi_1,V} L_{\varphi_1,0,V}(Y_1^*, Y_2^*) = \sup_{\tau,\beta,V_e^*} L_{\tau,\beta,V_e^*} \sup_{\varphi_2=0,V_{22}^*} L_{0,V_{22}^*}.$$

Hence,

$$\hat{\tau} = \hat{\varphi}_{1\bar{\omega}} \left(\text{car}\, \hat{\varphi}_{2\bar{\omega}} = 0\right).$$

Thus,

$$\begin{aligned} \hat{\varphi}_{1\bar{\omega}} = \hat{\tau} &= \hat{\varphi}_1 - \hat{\varphi}_2 S_{22}^{*-1} S_{21}^* \\ &= \widehat{(\varphi_1, \varphi_2)} \begin{pmatrix} I_r \\ -S_{22}^{*-1} \times S_{21}^* \end{pmatrix}. \end{aligned} \tag{1.2}$$

Thus,

$$\hat{\theta}_{\bar{\omega}} = (\hat{\varphi}_{1\bar{\omega}}, 0) \begin{pmatrix} R \\ D \end{pmatrix} = \left[\hat{\theta}(E,T) \begin{pmatrix} I_r & \\ -S_{22}^{*-1} & S_{21}^- \end{pmatrix}, 0\right] \begin{pmatrix} R \\ D \end{pmatrix}$$

$$= \hat{\theta}(I_p - T\left(T'ST\right)^{-1}T'S)(E,T) \begin{pmatrix} R \\ D \end{pmatrix}.$$

Since $S_{22}^* = T'ST, S_{21}^* = T'SE$, the last equation can be written as

$$\hat{\theta}_{\bar{\omega}} = \hat{\theta}(I_p - T\left(T'ST\right)^{-1}T'S).$$

The second formula results from the fact that $\left(I_p - T\left(T'ST\right)^{-1}I'S\right)$, idempotent, is a projector with rank $p-(p-r)=r$, thus with image $\operatorname{Im} R$ since $R\left(I_P - T\left(T'ST\right)^{-1}T'S\right) = R$ and kernel $\operatorname{Im} T'S$ since $T'S\left(I_p - T\left(T'SE\right)^{-1}T'S = 0\right.$; hence $I_p - T\left(T'ST\right)^{-1}T'S$ is equal to the projector $S^{-1}R'\left(RS^{-1}R'\right)^{-1}R$ which enjoys the same properties. The last equation results from the classical calculation of $\widehat{V}_H$ under the hypothesis H **(cf §I)**.

Corollary 1.11. *Under the hypothesis $H_{\bar{\omega}}$, $\hat{\theta}_{\bar{\omega}}$ is an unbiased estimator of θ.*

Indeed, we have

$$\begin{aligned} E\hat{\theta}_{\bar{\omega}} &= E\hat{\theta}\cdot\left(I_p - ET\left(T'ST\right)^{-1}T'S\right) \\ &= \theta\left(I_p - TE\left(T'ST\right)^{-1}T'S\right) = \theta, \end{aligned}$$

due to the independence of $\hat{\theta}$ and S and to $\theta \mathrm{T} = 0$ under $H_{\bar{\omega}}$.

The preceding lemma can be used to solve the following problem: Are the p coordinates of the variables $y_i, i = 1, \ldots, n$, of a multidimensional linear model necessary to test a linear column hypothesis H_ω?

Let $Y = (\underbrace{Y_{(1)}}_{S}, \underbrace{Y_{(2)}}_{p-S})$ and let $\theta = \left(\theta_{(1)}, \theta_{(2)}\right), V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ be the associated partitions.

Since $EY_{(1)} = X\tau + X\theta_{(2)}\beta$ and X injective, we deduce that $H_\omega : \theta_j \in \omega, j = 1, \ldots, p$ is equivalent to $\theta_j \in \omega, j = s+1, \ldots, p$ and $\tau_j \in \omega, j = 1, \ldots, s$.

Definition 1.5. If $Y = (U_1, \ldots, U_p)$, the hypothesis $\tau_j \in \omega, j = 1, \ldots, s$ is called the absence of an additional information on H_ω provided by $U_j, j = 1, \ldots, s$.

$$Y_{(1)} = X\tau + Y_{(2)}\beta + e^* = (\underbrace{X}_{k}, \underbrace{Y_{(2)}}_{p-S})\begin{pmatrix} \overbrace{\tau}^{S} \\ \beta \end{pmatrix} + e^*,$$

where $Y_{(2)}$ and e^* are independent.

This is a (random) fixed-effects model where we need to test $\tau_j \in \omega, j = 1, \ldots, s$, which is equivalent to the column linear hypothesis $\begin{pmatrix} \tau_j \\ \beta_j \end{pmatrix} \in \begin{Bmatrix} \omega \\ \oplus \\ \mathrm{R}^{\mathrm{p}-S} \end{Bmatrix}, j = 1, \ldots, \mathrm{s}.$

Thus, it suffices to calculate $S_{(2)}$ and $S_{(2)_\omega}$. By **proposition 4 from** §0 : $S_{(2)} = S_{22} - S_{21}S_{11}^{-1}S_{12}$

$$S_{(2)}\left\{\tau_j \in \omega\right\} = S_{22(\omega)} - S_{21(\omega)}S_{11(\omega)}^{-1}S_{12(\omega)}.$$

We will then perform a T^2-test if $k - \dim(\omega) = 1$, or a Hotelling-Lawley, Wilks, or Roy test otherwise. The degrees of freedom for $S_{(2)}$ and $S_{(2)_\omega}$ are respectively $n - (k + p - s)$ and $n - (\dim \omega + p - s)$.

b) Test of the hypothesis $H_{\bar{\omega}}$ (Gaussian case).

i) Let's search for the maximum likelihood test in the model in the canonical situation; then $H_{\bar{\omega}}$ is $\varphi_2 = 0$. We have seen that

$$L_{\varphi_1,\varphi_2,V}\big(Y^*_{(1)}, \mathrm{Y}^*_{(2)}\big) = L_{\tau,\beta,V^*_e} L_{\varphi_2,V^*_{22}}.$$

Hence

$$\frac{\sup\limits_{\varphi_1,\varphi_2=0,V} L_{\varphi_1,\varphi_2,V}(Y^*_{(1)}, Y^*_{(2)})}{\sup\limits_{\varphi_1,\varphi_2,V} L_{\varphi_1,\varphi_2,V}(Y^*_{(1)}, Y^*_{(2)})} = \frac{\sup\limits_{\varphi_2=0,V^*_{22}} L_{0,V^*_{22}}(Y^*_{(2)})}{\sup\limits_{\varphi_2,V^*_{22}} L_{\varphi_2,V^*_{22}}(Y^*_{(2)})}.$$

This is the maximum likelihood test of the linear hypothesis in column, $\varphi_2 = 0$ in the restricted model

$$Y^*_{(2)} = X\varphi^*_{(2)} + \varepsilon^*_{(2)},$$

i.e., $YT = X\theta T + \varepsilon T$.

ii) More generally, let's search for tests invariant under regular linear transformations in $\mathbb{R}^p$ parallel to $\bar{\omega}$; we can restrict the tests to be functions of the exhaustive statistic $(\hat{\varphi}_1, \hat{\varphi}_2, S)$. Invariance implies that these tests will depend only on $(\hat{\varphi}_2, S^*_{22})$, which reduces to a test of the linear hypothesis in column $\varphi_2 = 0$ in the restricted model $Y^*_{(2)} = YT = X\theta T + \varepsilon T$.

Therefore:

1) If $k = 1$, there exists an invariant T^2-test for the restricted model.

The hypothesis $H_{\bar{\omega}}$, gives

$$\hat{\theta}T\left(T'ST\right)^{-1}T'\hat{\theta}' \times \frac{n-r}{r} \sim \mathcal{F}(r, n-r),$$

hence the test.

Example 1.3. *Test of $\theta_1 = \ldots = \theta_p$ for a sample of size n from the $\mathcal{N}_p(\theta, V)$ distribution. Choose*

$$T = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ -1 & \dots & -1 \end{pmatrix};$$

$$\textit{then} \quad \begin{array}{c} Y_i^* = (Y_{i1} - Y_{ip}, \dots, Y_{ip-1} - Y_{ip}) \\ \varphi_2 = (\theta_1 - \theta_p, \dots, \theta_{p-1} - \theta_p) \end{array}$$

The hypothesis to test becomes $\varphi_2 = 0$, and we conclude with a T^2-test.

2) If $p - r = 1$, the restricted model is one-dimensional. Under $\mathrm{H}_{\bar{\omega}'}$

$$\frac{T'\hat{\theta}'X'X\hat{\theta}T/k}{T'ST/(n-k)} \sim \mathcal{F}(k, n-k),$$

and we conclude with an F-test.

3) In the general case, we will use the Hotelling-Lawley test, the Wilks' test, or the Roy's test, etc...

c) Study of the hypothesis H : : "$\mathrm{C}\theta\mathrm{T} = 0$ " (Gaussian case).

1) Let $C_{k-h\times k} : \mathbb{R}^k - \mathbb{R}^{k-h}$ (on the left) with rank $k-h$ and $T_{p\times p-r} : \mathbb{R}^p - \mathbb{R}^{p-r}$ (on the right) with rank $p-r$. It is sometimes useful to consider the hypothesis $C\theta T = 0$, which generalizes the linear hypotheses in columns ($T = I_p$) or in rows ($C = I_k$).

For example, if $T = \begin{pmatrix} 0 \\ I_{p-r} \end{pmatrix}$, it is a linear hypothesis on the last $(p-r)$ columns of θ:

$$\mathrm{C}\theta_j = 0, j = r+1, \dots, p.$$

Alternatively,

$$Y^* = Y(E,T) = X(\varphi_1, \varphi_2) + \left(\varepsilon^*_{(1)}, \varepsilon^*_{(2)}\right),$$

is the model in canonical position (for the hypothesis $\theta T = 0$).

From $L_{\varphi_1,\varphi_2,V}(Y^*_{(1)}, Y^*_{(2)}) = L_{\tau,\beta_1 V_e^*}\left(Y^*_{(1)}/Y^*_{(2)}\right) \times L_{\varphi_2,V_{22}^*}\left(Y^*_{(2)}\right)$, we can deduce, as in the case of the linear hypothesis in rows:

- $\hat{\tau} = \hat{\varphi}_1 - \hat{\varphi}_2\hat{\beta}$ with $\hat{\beta} = S_{22}^{*-1}S_{21}^* = (T'ST)^{-1}T'SE$.
- $\hat{\tau}_H = \hat{\varphi}_{1_H} - \hat{\varphi}_{2_H}\hat{\beta}_H$ and $\widehat{\tau}_H = \widehat{\tau}$, $\hat{\beta}_H = \hat{\beta}$, hence,
- $\hat{\varphi}_{1_H} = \hat{\varphi}_1 + (\hat{\varphi}_{2_H} - \hat{\varphi}_2)\hat{\beta}$.

But $\hat{\varphi}_{2_H}$ is the maximum likelihood estimator of φ_2 under $\mathrm{C}\varphi_2 = 0$ in the restricted model

$$\mathrm{Y}^*_{(2)} = \mathrm{X}\varphi_2 + \varepsilon^*_{(2)}$$

since $\hat{\varphi}_{2_H}$ maximizes $L_{\varphi_2,\ \mathrm{V}_{22}}\left(\mathrm{Y}^*_{(2)}\right)$ under $\mathrm{C}\varphi_2 = 0$, so:

$$\hat{\varphi}_{2_H} = \left[I_k - (X'X)^{-1} C' \left(C\,(X'X)^{-1} C'\right)^{-1} C\right] \hat{\varphi}_2 = (I_k - A)\,\hat{\varphi}_2.$$

Thus,

$$\begin{aligned}\left(\widehat{\varphi_{1_H}, \varphi_{2_H}}\right) &= \left(\hat{\varphi}_1 - A\hat{\varphi}_2\hat{\beta}, (I_k - A)\,\hat{\varphi}_2\right) = (\hat{\varphi}_1, \hat{\varphi}_2) - A\left(\hat{\varphi}_2\hat{\beta}, \hat{\varphi}_2\right) \\ &= (\hat{\varphi}_1, \hat{\varphi}_2) - A\hat{\varphi}_2\left(\hat{\beta}, I_{p-r}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{\theta}_H &= \hat{\theta} - A\hat{\theta}T\left((T'ST)^{-1} T'SE, I_{p-r}\right)\begin{pmatrix} C \\ D \end{pmatrix} \\ &= \hat{\theta} - A\hat{\theta}\left(T\,(T'ST)^{-1} T'S\right)(E, T)\begin{pmatrix} C \\ D \end{pmatrix} \\ &= \hat{\theta} - A\hat{\theta}\left(T\,(T'ST)^{-1} T'S\right).\end{aligned}$$

Therefore,

Proposition 1.8. *Under the hypothesis* $\mathrm{H} : \mathrm{C}\theta\mathrm{T} = 0$*, the maximum likelihood estimator of* (θ, nV) *is given by:*

$$\begin{aligned}\hat{\theta}_H &= \hat{\theta} - (X'X)^{-1} C' \left(C\,(X'X)^{-1} C'\right)^{-1} C\hat{\theta}T\,(T'ST)^{-1} T'S \\ S_H &= \left(Y - X\hat{\theta}_H\right)' \left(Y - X\hat{\theta}_H\right).\end{aligned}$$

Corollary 1.12. *Under* H, $\hat{\theta}_{\mathrm{H}}$ *is an unbiased estimator of* θ*.*

We still have $E\hat{\theta}_H = E\hat{\theta} - (X'X)^{-1} C'(C\,(X'X)^{-1} \underbrace{CE\hat{\theta}ET}_{C\theta T=0}(T'ST)^{-1} T'S = \theta$ due to the independence of $\hat{\theta}$ and S.

2) The likelihood ratio test for $C\theta T = 0$ is obtained using:

$$\frac{\sup\limits_{\varphi_1, C\varphi_2=0, V^*} L_{\varphi_1,\varphi_2,V^*}(Y^*_{(1)}, Y^*_{(2)})}{\sup\limits_{\varphi_1,\varphi_2,V^*} L_{\varphi_1,\varphi_2,V^*}(Y^*_{(1)}, Y^*_{(2)})} = \frac{\sup\limits_{C\varphi_2=0, V^*_{22}} L_{\varphi_2,V^*_{22}}(Y^*_{(2)})}{\sup\limits_{\varphi_2,V^*_{22}} L_{\varphi_2,V^*_{22}}(Y^*_{(2)})}$$

This is equivalent to the likelihood ratio test for the linear hypothesis in column

$C\varphi_2 = 0$ in the restricted model $Y^*_{(2)} \equiv \mathrm{YT} = X\varphi_2 + \varepsilon T$, which is

$$\left(\frac{|S^*_{22}|}{|S_{22_H}|}\right)^{n/2} = \frac{T'ST^{n/2}}{\left|T'ST + T'\hat{\theta}'C'\left(C\left(X'X\right)^{-1}C'\right)^{-1}C\hat{\theta}T\right|^{n/2}}$$

because

$$S^*_{22_H} = T'S_HT = T'\left(Y - X\hat{\theta}_H\right)'\left(Y - X\widehat{\theta}_H\right)T = T'ST + T'\hat{\theta}C'\left(C\left(X'X\right)^{-1}C'\right)^{-1}C\hat{\theta}T$$

.

3) More generally, if we search for invariant tests under (affine) regular transformations of $\mathbb{R}^p$ parallel to $\bar{\omega} = \mathrm{Ker}\, T$, we can limit the tests to be functions of the exhaustive statistic $(\hat{\theta}, S)$, and thus by invariance, functions of $(\hat{\varphi}_2, S^*_{22})$: these are tests for the hypothesis $C\varphi_2 = 0$ in the restricted linear model $YT = X\varphi + \varepsilon T$. The invariance considerations of **paragraph 3** imply that, depending on the cases, we can perform a F test, a T^2 test, or a test such as Hotelling-Lawley, Wilks or Roy.

d) Structure hypotheses on θ (Gaussian case).

Given the multidimensional linear model

$$Y = X\theta + \varepsilon$$

and $C : \mathbb{R}^k \longrightarrow \mathbb{R}^{k-h}$ with rank $k-h$, consider the "structure" hypothesis $\varphi_r : C\theta$ has rank $\leq r''$ with a given $r < \inf(p, k-h)$. In particular, for $C = I_k, \varphi_r$ is the hypothesis θ has rank $\leq r''$.

The hypothesis φ_r is equivalent to: there exists $T_{p\times(p-r)}$ with rank $p-r$ such that $C\theta T = 0$.

The symmetric hypothesis θT has rank $\leqslant m$ for $Tp \times (p-r)$ with rank $p-r$ given and $m < \inf(p, -r, k)$ is reduced to the previous one by considering the restricted model $\mathrm{YT} = \mathrm{X}\varphi_2 + \varepsilon T$; the hypothesis then becomes φ_2 has rank m. Let's find the maximum likelihood estimators $\hat{\theta}_\varphi, \hat{V}_\varphi\, (= nS_\varphi), \widehat{T}_\varphi$ under the hypothesis φ; at fixed T, by maximizing over θ and V, we obtained in the previous paragraph:

$$\sup_{C\theta T=0} L_{\theta,V} / \sup_{\theta,V} L_{\theta,V} = \frac{|T'ST|^{n/2}}{\left|T'ST + T'\hat{\theta}'C'\left(C\left(X'X\right)^{-1}C'\right)^{-1}C\hat{\theta}T\right|^{n/2}}.$$

Thus, we need to maximize this ratio over T, where $\sup_{\theta,V} L_{\theta,V}$ is a constant.

Lemma 1.2. *Let E and F be positive definite symmetric matrices, and $\phi_1 \geqslant \ldots \geqslant \phi_p > 0$ be the roots of $|E - \phi F| = 0$, i.e., the eigenvalues of the positive definite symmetric matrix DED' where $F = (D'D)^{-1}$. Let $\xi_1, \ldots, \xi_p$ be linearly independent eigenvectors associated with $\phi_1, \ldots, \phi_p$. When T varies among the $(p \times (p-r))$ matrices with rank $(p-r)$, then*

$$\sup_T \frac{|T'ET|}{T'FT} \mid$$

is attained for $T_M = (\xi_1, \ldots, \xi_{p-r})$ and equals

$$\prod_{i=1}^{p-r} \phi_i.$$

Here $E = S$ and

$$F = S + \hat{\theta}' C' \left(C \left(X'X \right)^{-1} C' \right)^{-1} C\hat{\theta} = S_\omega = S + B_\omega,$$

with the notations from **Paragraph 3** concerning the hypothesis $\mathrm{H}_\omega : \mathrm{C}\theta = 0$. The $(p-r)$ largest roots ϕ_i of $|S - \phi (S + B_\omega)| = 0$ correspond, with $\phi_i = \frac{1}{1+\lambda_i}$, to the $(p-r)$ smallest non-negative roots of $|B_\omega - \lambda S| = 0$. We can conclude:

Theorem 1.3. *Let S_ω be the maximum likelihood estimator of nV under the hypothesis $C\theta = 0$, and $B_\omega = S_\omega - S$; let $0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_p$ be the roots of $|B_\omega - \lambda S| = 0$ and $\xi_1, \ldots \xi_p$ be orthogonal eigenvectors for E and F associated with λ_i.*

1. $\hat{T}_\varphi$ has a solution $(\xi_1, \ldots, \xi_{p-r})$ [not unique as only the space defined by $\{\xi_1, \ldots, \xi_{p-r}\}$ is well-defined];

2. $\hat{\theta}_\varphi$ and $\hat{V}_\varphi$ can be obtained from the formulas in the previous paragraph by replacing T with $\hat{T}_\varphi$;

3. $\Lambda_\varphi = \underset{\varphi}{Sup} / SupL = |\prod_{i=1}^{p-r} (1 + \lambda_i)|^{-n/2}$.

Corollary 1.13. *The rejection region of the likelihood ratio test for the hypothesis φ_r : $C\theta$ has rank $\leqslant r$ against $C\theta$ has rank $> r$ is of the form*

$$\{\Lambda_\varphi \leqslant C_\alpha\}.$$

Note that under the hypothesis φ_r, we have:

$$P\left(\Lambda^{2/n}(T) \leqslant U_{\alpha,p-r,n-k,h}\right) = \alpha,$$

for at least one T; since $\Lambda_y^{2/n} = \underset{T}{Sup} \Lambda^{2/n}(T)$, it follows:

$$P\left(\Lambda_\varphi^{2/n} \leqslant U_{\alpha,p-r,n-k,h}\right) \leqslant \alpha \text{ under } \varphi_r.$$

The choice $C_\alpha = U_{\alpha,p-r,n-k,h}$ leads to a "conservative" test with a level no greater than α. We can also calculate C_α using the asymptotic approximation (in a model to be specified where $\frac{1}{n} X_n X_n' \underset{n\to+\infty}{\longrightarrow} A$):

$$-2\log\Lambda_\varphi \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} X^2((p-r)(k-h-r)) \text{ under } \varphi_r.$$

Chapter 2

Intercept-only Model under Non-normality.

2.1 Introduction

The main use of regression is to illuminate on a supposed linear relationship between predictor variables and an outcome variable Anderson (1958). Regression is an old and established statistical method, with a background that is more relevant for its role of traditional explanatory modeling than for prediction. With the advent of big data, regression is widely used to train a model for predicting outcomes, rather than explaining the data. In this case, the main items of interest are the fitted outcome values. Usually the first step is to determine if there is a relationship between the outcome and the predictors. The null hypothesis refers that there is no relationship between any of the predictors and the response. If the null hypothesis is accepted, we retain the intercept-only model. Generally, we suppose that the errors are independent and normally distributed with a zero mean and a variance σ^2. In these case, it should be noted that there is no slope so the intercept is estimated by the mean response y. At first glance, it does not seem that studying regression without predictors would be very useful. Certainly, we are not suggesting that using regression without predictors is a major data analysis tool. Often a model with intercept and predictors is compared to an intercept-only model (sometimes known as the mean model) to test whether the predictors add over and above the interceptonly.

The mean model may seem overly simplistic, but the sample mean is a simple but very powerful descriptor; it is counted among the most basic ways to describe, analyze and summarize information about a phenomenon, in gaussian case. In the absence of explanatory variables, the mean can be a model by itself. The mean model is often the starting point for constructing forecasting models for time series data, including random

walk models. For example, a brief look at the intercept-only model, consider a time series presenting the daily closing price of the Dow Jones Industrial Average over a some period. Suppose we wish to create a regression model for this time series. But we don't know what factors influence the Closing Price. Neither do we want to assume any inflation, trend or seasonality in the data set. In the absence of any assumptions about inflation, trend, seasonality or the presence of explanatory variables, the best we can do is the intercept-only model. In the intercept-only model, all forecasts take the value of the intercept. If we can find some mathematical transformation (e.g., differencing, logging, deflating, etc.) that converts the original time series into a sequence of values that are independently and identically distributed, we can use the mean model to obtain forecasts and confidence limits for the transformed series, and then reverse the transformation to obtain corresponding forecasts and confidence limits for the original series. What we do get is the mean of the outcome, i.e. the expected value of y when you do not control for anything. Another reason for doing this is that some packages require the user to define a base model.

We do think that it is worthwhile to look at regression models without predictors to see what they can tell us about the nature of the constant. Understanding the regression constant in these simpler models will help us to understand both the constant and the other regression coefficients in later more complex models. In the case where there are no predictors, the equation reduces to,

$$y_i = a + u_i \quad i = 1, \ldots, n. \tag{2.1}$$

The disadvantage of parametric modeling is the requirement that the structural model and error distribution be correctly specified. In some cases, for example, if the observations come from a discrete distribution or the deviations from the mean have a strong dissymmetry, the hypothesis of normality is no longer tenable. Violation of the normality assumption sometimes may be attributed to the skewed nature of the dependent variable. The data distribution may deviate from a Gaussian distribution in multiple ways. Rather than being continuous, data may be discrete, such as integer counts or even binomial character states. Continuous variables may deviate from normality in terms of skewness (showing a long tail on one side), kurtosis (curvature leading to light or heavy tails), and even higher-order moments. For example, measures of actuating asymmetry are distributed half-normally. Many applications have positive response variables. Such variables usually have distributions right-skewed. The boundary at zero limits the left tail of the distribution.

We suppose that residues u_i are i.i.d.having exponential distribution $\mathcal{E}(\theta^{-1})$, $\theta > 0$

with $E(u_1) = \theta, \mathrm{Var}(u_1) = \theta^2$. The exponential probability distribution describes the probabilities with which a random variable u takes on values, where u can only be positive. More precisely, the probability density function of the law exponential for value u is given by

$$f(u) = \theta^{-1} \exp\left(-\theta^{-1} u\right). \tag{2.2}$$

For example, package of R contains a data set on the number of lynx caught per year in Canada between 1821 and 1934. The shape of the histogram has a decreasing tendency, the values of the observations are all positive; it makes us think of an exponential law. Suppose then that the observations $(x_1, \ldots, x_n)$ with $n = 114$ are the realizations of random variables $X_1, \ldots, X_n$ independent of exponential law $\mathcal{E}(\theta)$ with $\theta > 0$.

Departures from normality may have several causes. For example, they may be due by the presence of outlying values in the responses, among others reasons. For this, several researchers proposed to perform regression analysis using a model that assumes a non gaussian distribution for the error terms. In Huber (1996) and Tiku et al (1986), it has been mentioned that the underlying distribution is, in most situations, basically non normal. Tiku et al. (2008) who constructed a model with a variety of bivariate non-normal distributions by using the conditional method. Tiku et al. (2001) considered the simple linear regression model and considered several non-normal distributions for the random error, both symmetric and skew. They obtained the modified maximum likelihood estimators of parameters.Islam and Tiku (2005) derive the modified maximum likelihood estimators of the parameters in multiple linear regression models and compare them with the least squares estimators and the Huber (1981) M-estimators. Hung T. Nguyen (2015) emphasize problems where fuzzy data appear naturally and need to be used and analyzed properly within the context of applied statistics. In Djaballah-Djeddour and Tazerouti (2022) the problem of checking the linearity of a regression relationship is addressed. A test based on a Hermite transform characterization of conditional expectations was developed without assuming the normality of the errors. Research focused on the broad class of elliptic distributions, and in particular on the multivariate t-distribution. For example, Zellner (1976) ;Sutradhar and Ali (1986) ;Galea et al. (1997) ;Liu (2000) ;Díaz-García et al. (2003) investigate cases performed within the elliptic distribution family, in order to analyse more complex situations, such as data with missing values in the response variables, see also Liu (2004) , and with monotone missing response variables as in Batsidis and Zografos (2008) . problem was also approached, within a Bayesian framework, by assuming a multivariate skewed and heavy-tailed distribution for the error terms see Ferreira and Steel (2007).Carroll and Hall (2004) suggest two new methods, which are applicable to both deconvolution and regression with errors in explanatory

variables, for non parametric inference. The two approaches involve kernel or orthogonal series methods. They are based on defining a low order approximation to the problem at hand, and proceed by constructing relatively accurate estimators of that quantity rather than attempting to estimate the true target functions consistently.

The paper is structured as follows: the estimates are presented in Section 2 as well as the finite and asymptotic properties of the obtained estimator. Section 3 provides simulations. Section 4 is devoted to the proofs.

2.2 Main results

We will consider two estimators for a, the maximum likelihood estimator and the ordinary least squares estimator. The least squares technique has traditionally been justified by two assumptive arguments, it provides the maximal likelihood regression coefficients, if the errors are Gaussian and of all unbiased linear estimators, least squares have a minimal variance. Both of these properties have at times been adduced to call least squares the best of regression techniques. Because least squares possess in addition, the attribute of the computational facility, this method long has reigned as the foremost tool in reducing data to mathematically descriptive relationships. The first argument above assumes a normal distribution of the error terms. We argue that this supposition is often unwarranted and we can show that significant gains in likelihood maybe achieved when the regression technique allows for the more general class of error distribution.

When the probability distribution of errors is assumed, it is possible, to obtain consistent and efficient estimates (minimum variance) of the parameters using the maximum likelihood method. This technique could be widely applied to non-Gaussian noise problems.

2.2.1 Maximum Likelihood Estimator

We consider likelihood based inference for the parameter a. For that, let us calculate $\log L(y_1, \ldots, y_n; a)$ and look for the solution that maximizes this quantity.

Proposition 2.1. *The log-likelihood of y is given by*

$$L(y_1, \ldots, y_n, a, \theta) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}(y_i - a)\right) 1_{\{\inf y_i - a \geq 0\}}. \tag{2.3}$$

Suppose θ known, then the estimate of a is given by

$$\widehat{a}_{mle} = \inf_{1\leq i\leq n} y_i. \tag{2.4}$$

Proof. Given the assumed structural model (2.19) and the known error distribution, the conditional distribution of y can be derived. We have $u_i \sim \mathcal{E}xp\left(\theta^{-1}\right)$ which gives by definition $f_u(z) = \frac{1}{\theta}\exp\left(-\frac{z}{\theta}\right)1_{\{z\geq 0\}}$ and $F_u(z) = 1-\exp\left(-\frac{z}{\theta}\right)$. The distribution function of y is thus deduced as follows:

$$\begin{aligned} F_y(t) &= P(y\leq t) = P(a+u\leq t) \\ &= 1-\exp\left(-\frac{t-a}{\theta}\right). \end{aligned}$$

with $t-a\geq 0$. Therefore

$$f_y(t) = \frac{1}{\theta}\exp\left(-\frac{t-a}{\theta}\right)1_{\{t\geq a\}}. \tag{2.5}$$

The log-likelihood is then written

$$\begin{aligned} L\left(y_1,\ldots,y_n,a,\theta\right) &= \prod f_y\left(y_i\right) \\ &= \frac{1}{\theta^n}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}\left(y_i-a\right)\right)\mathbf{1}_{\{\inf y_i\geq a\}} \end{aligned}$$

Let us note

$$g(a) = \frac{1}{\theta^n}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}\left(y_i-a\right)\right).$$

For every fixed θ and with $a\in\,]-\infty,\inf_{1\leq i\leq n} y_i]$. The derivative of g is

$$g'(a) = \frac{1}{\theta^{n+1}}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}\left(y_i-a\right)\right).$$

The derivative $g'(a)$ is always positive $\forall\theta>0$. This implies that g is increasing from $-\infty$ to inf y_i. Consequently, the maximum likelihood estimator of a is reached in:

$$\widehat{a}_{mle} = \inf_{1\leq i\leq n} y_i. \tag{2.6}$$

Given the assumed structural model (2.19) and the known error distribution, the conditional distribution of y can be derived. We have $u_i \sim \mathcal{E}xp\left(\theta^{-1}\right)$ which gives by definition $f_u(z) = \frac{1}{\theta}\exp\left(-\frac{z}{\theta}\right)1_{\{z\geq 0\}}$ and $F_u(z) = 1-\exp\left(-\frac{z}{\theta}\right)$. The distribution function of y is

thus deduced as follows:

$$\begin{aligned} F_y(t) &= P(y \leq t) = P(a + u \leq t) \\ &= 1 - \exp\left(-\frac{t-a}{\theta}\right). \end{aligned}$$

with $t - a \geq 0$. Therefore

$$f_y(t) = \frac{1}{\theta} \exp\left(-\frac{t-a}{\theta}\right) 1_{\{t \geq a\}}. \tag{2.7}$$

The log-likelihood is then written

$$\begin{aligned} L(y_1, \ldots, y_n, a, \theta) &= \prod f_y(y_i) \\ &= \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum_{i=1}^{n} (y_i - a)\right) \mathbf{1}_{\{\inf y_i \geq a\}} \end{aligned}$$

Let us note

$$g(a) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum_{i=1}^{n} (y_i - a)\right).$$

For every fixed θ and with $a \in]-\infty, \inf_{1 \leq i \leq n} y_i]$. The derivative of g is

$$g'(a) = \frac{1}{\theta^{n+1}} \exp\left(-\frac{1}{\theta} \sum_{i=1}^{n} (y_i - a)\right).$$

The derivative $g'(a)$ is always positive $\forall \theta > 0$. This implies that g is increasing from $-\infty$ to inf y_i. Consequently, the maximum likelihood estimator of a is reached in:

$$\widehat{a}_{mle} = \inf_{1 \leq i \leq n} y_i. \tag{2.8}$$

□

Corollary 2.1. *2 If θ is unknown, it can be estimate by*

$$\widehat{\theta} = \bar{y} - \inf_{1 \leq i \leq n} y_i. \tag{2.9}$$

Proof. To search for the global max, it suffices to maximize the function $L(y_1, \ldots, y_n, a, \theta)$ or $\log L(y_1, \ldots, y_n, a, \theta)$ with respect to θ. We place ourselves outside the case (which is of zero probability whatever the value of the parameter) where $\sum_{i=1}^{n} (x_i - \inf_{1 \leq i \leq n} (x_i)) =$

0 (which means that all x_i are equal). Suppose that a is fixed

$$\log L\left(y_1,\ldots,y_n,a,\theta\right) = -n\log\theta - \frac{1}{\theta}\sum_{i=1}^{n}\left(y_i-a\right). \tag{2.10}$$

Condition necessary: $\frac{d\log L}{d\theta}=0$. We have

$$\frac{d\log L}{d\theta} = -\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}\left(y_i-a\right) \Longrightarrow n\widehat{\theta} = \sum_{i=1}^{n} y_i - na \Longrightarrow \widehat{\theta} = \bar{y}-a$$

. a is unknown, we can replace it by its estimator:

$$\widehat{\theta} = \bar{y} - \inf_{1\le i\le n} y_i. \tag{2.11}$$

Then we check that the second derivative at this point is negative, which ensures that the critical point is indeed a maximum. By calculating the second derivative, we get

$$\frac{d^2\log L}{d\theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}\left(y_i-a\right) \tag{2.12}$$

$\Rightarrow$

$$\begin{aligned}\frac{n}{\widehat{\theta^2}} - \frac{2}{\widehat{\theta^3}}\sum_{i=1}^{n}\left(y_i-a\right) &= -\frac{1}{\widehat{\theta}}\left[-\frac{n}{\widehat{\theta}}+\frac{1}{\widehat{\theta^2}}\sum_{i=1}^{n}\left(y_i-a\right)\right] - \frac{1}{\widehat{\theta^3}}\sum_{i=1}^{n}\left(y_i-a\right)\\ &= -\frac{1}{\widehat{\theta^3}}\sum_{i=1}^{n}\left(y_i-a\right).\end{aligned}$$

Since $\theta>0$ and $\sum_{i=1}^{n} y_i >$ na because all $y_i > a \Rightarrow \frac{d^2\log L}{d\theta^2} < 0$. □

The distribution of $\widehat{a}_{mle}$: it is helpful to know the law of an estimator. It allows us to calculate the characteristics of the estimator and construct a confidence interval for the parameter. In some cases, that distribution can be determined directly from observed random variables law. The distribution function of $\widehat{a}_{\text{mle}}$ is:

$$\begin{aligned}F_{\widehat{a}_{mle}}(t) &= P\left(\inf_{1\le i\le n} y_i \le t\right) = 1 - P\left(\inf_{1\le i\le n} y_i \ge t\right)\\ &= 1-\prod_{i=1}^{n} P\left(y_i\ge t\right)\\ &= 1-\left[1-F_Y(t)\right]^n = 1-\left(e^{-\frac{(t-a)}{\theta}}1_{\{(t-a)\ge 0\}}\right)^n\\ &= 1-e^{-n\frac{(t-a)}{\theta}}1_{\{t\ge a\}}.\end{aligned}$$

From where

$$f_{\widehat{a}_{mle}}(t) = \frac{n}{\theta} \exp\left(-\frac{n(t-a)}{\theta}\right) 1_{\{t>a\}}. \tag{2.13}$$

It's the exponential distribution $\mathcal{E}\S\sqrt{}\left(\frac{n}{\theta}\right)$. An unknown parameter can have more than one estimator. When we use point estimates, we want them to have certain properties. These properties are important in choosing the best estimator for the parameter, that is, the one that comes closest to the true parameter. The bias of $\widehat{a}_{mle}$ is defined as $E\left(\widehat{a}_{mle}\right) - a$. It is the distance between the average of collection estimates and the single parameter being estimated. The bias also is the expected value of the error, since $E\left(\widehat{a}_{mle}\right) - a = E\left(\widehat{a}_{mle} - a\right)$. The relationship between bias and variance is analogous to the relationship between accuracy and precision. Then the error for one estimate is high, do not mean the estimator is biased. The ideal situation is to have an estimate, unbiased, with low variance, and with few outliers. We calculate of expectation and variance of $\widehat{a}_{mle}$.

- Knowing that $\widehat{a}_{mle}$ can be written as $\widehat{a}_{mle} = Z + a$ where Z follows the law $\mathcal{E}\S\sqrt{}\left(\frac{n}{\theta}\right)$, we deduce that

$$E\left(\widehat{a}_{\text{mle}}\right) = EZ + a = \frac{\theta}{n} + a. \tag{2.14}$$

- It is therefore, obvious that $\widehat{a}_{mle}$ is a biased estimator of a. The variance of $\widehat{a}_{mle}$ is simply the expected value of the squared sampling deviations; that is,

$$\text{Var}\left(\widehat{a}_{mle}\right) = E\left(\widehat{a}_{mle} - E\left(\widehat{a}_{mle}\right)\right)^2 = \frac{\theta^2}{n^2}. \tag{2.15}$$

Every time a sample is taken, we lose some part of the information about the population. That inevitably results in an error in the estimate. Therefore, if we want a very high level of precision, we must take a sample of a size sufficient to extract sufficient information from the population to make the estimate with the desired precision.

The bias of $\widehat{a}_{mle}$ is $\frac{\theta}{n}$ tends to 0 when $n \to \infty$, we deduce that $\widehat{a}_{\text{mle}}$ is asymptotically unbiased. The variance of $\widehat{a}_{mle}$ is $\frac{\theta^2}{n^2}$. The variance $\text{Var}\left(\widehat{a}_{mle}\right)$ tends to 0 when n tends to infinity.

The Mean Squared Error (MSE) : It is used to indicate how far, on average, the collection of estimates are from the single parameter being estimated. If the MSE is relatively low then the estimators are likely more highly clustered (than highly dispersed)

around the a.

$$\begin{aligned}\mathrm{MSE}\left(\widehat{a}_{\mathrm{mle}}\right) &= E\left(\widehat{a}_{mle}-a\right)^2 \\ &= \left(\text{ biais }\left(\widehat{a}_{\mathrm{mle}}\right)\right)^2+\mathrm{Var}\left(\widehat{a}_{\mathrm{mle}}\right) \\ &= \left(\frac{\theta}{n}\right)^2+\frac{\theta^2}{n^2} \\ &= \frac{2\theta^2}{n^2}\to 0 \text{ when } n\to\infty.\end{aligned}$$

Note the difference between MSE and variance. A consistent sequence of estimators is a sequence of estimators that converge in probability to the quantity being estimated as the sample size grows without bound. In other words, increasing the sample size increases the probability of the estimator being close to the population parameter. Mathematically, a sequence of estimators $\{t_n; n\geq 0\}$ is a consistent estimator for parameter a if and only if, for all $\varepsilon>0$, no matter how small we have

$$\lim_{n\to\infty} P\left(|t_n-a|>\varepsilon\right)=0. \tag{2.16}$$

The consistency defined above may be called weak consistency. The sequence is strongly consistent, if it converges almost surely to the true value.

Proposition 2.2. *We have that $\sqrt{n}\left(y_{(1)}-a\right)\to 0$ in probability when $n\to\infty$, where $y_{(1)}=\inf_{1\leq i\leq n} y_i$. This shows the consistency of the estimator with*

$$y_{(1)}=\widehat{a}_{mle}=\inf_{1\leq i\leq n} y_i=t_n,$$

and $|t_n-a|=y_{(1)}-a$ because $y_{(1)}>a$. A convergent estimator deviates from the parameter with a low probability, if the sample size is large enough.

Proof. We write

$$P\left(\sqrt{n}\left|y_{(1)}-a\right|>\delta\right)=P\left(\left|\inf_{1\leq i\leq n} y_i-a\right|>\frac{\delta}{\sqrt{n}}\right) \text{ and } \inf_{1\leq i\leq n} y_i\geq a.$$

Then we get

$$\begin{aligned}P\left(\sqrt{n}\left|y_{(1)}-a\right|>\delta\right) &= P\left(\inf_{1\leq i\leq n} y_i>\frac{\delta}{\sqrt{n}}+a\right) \\ &= 1-P\left(\inf_{1\leq i\leq n} y_i\leq\frac{\delta}{\sqrt{n}}+a\right) \\ &= 1-F_{\widehat{a}_{mv}}\left(\frac{\delta}{\sqrt{n}}+a\right) \\ &= e^{-\frac{\sqrt{n}}{\theta}\delta}\end{aligned}$$

. Thus,

$$\lim_{n\to\infty} P\left(\sqrt{n}\left|y_{(1)}-a\right|>\varepsilon\right)=0. \tag{2.17}$$

□

Lemma 2.1. *We prove that* $\frac{1}{\log n}y_{(1)}\to\theta$ *in probability,* $n\to\infty$*. Let* $\delta>0$*, this amounts to showing that*

$$P\left(\left|\frac{1}{\log n}y_{(1)}-\theta\right|>\delta\right)\to 0 \text{ when } n\to\infty. \tag{2.18}$$

Proof. We prove (10) by splitting the event into two events

- $P\left(\frac{1}{\log n}y_{(1)}\leq\theta-\delta\right)\to 0$ when $n\to\infty$,
- $P\left(\frac{1}{\log n}y_{(1)}\geq\theta+\delta\right)\to 0$ when $n\to\infty$.

The first is rewritten as:

$$\begin{aligned}
P\left(\frac{1}{\log n}y_{(1)}\leq\theta-\delta\right) &= 1-P\left(\frac{1}{\log n}y_{(1)}\geq\theta-\delta\right)\\
&= 1-\left(P\left(y_1\geq(\theta-\delta)\log n\right)\right)^n\\
&= 1-\left(1-P\left(y_1\leq(\theta-\delta)\log n\right)\right)^n\\
&= 1-\left(e^{-\frac{((\theta-\delta)\log n-a)}{\theta}}\right)^n
\end{aligned}$$

. We assume that $\delta<\theta$. The second is rewritten as:

$$\begin{aligned}
P\left(\frac{1}{\log n}y_{(1)}\geq\delta+\theta\right) &= P\left(y_{(1)}>(\delta+\theta)\log n\right)\\
&\leq nP\left(y_1>(\delta+\theta)\log n\right),
\end{aligned}$$

this increase (majoration) by sub-additivity of P is sufficient, and

$$\begin{aligned}
nP\left(y_1>(\delta+\theta)\log n\right) &= n-\left(1-e^{-\frac{(\delta+\theta)\log n-a}{\theta}}\right)n\\
&= ne^{-\frac{(\delta+\theta)\log n-a}{\theta}}=ne^{\frac{a}{\theta}}e^{-\frac{(\delta+\theta)\log n}{\theta}}\\
&= e^{\frac{a}{\theta}}n^{-\frac{(\delta+\theta)}{\theta}+1}=e^{\frac{a}{\theta}}n^{-\frac{\alpha}{\theta}}
\end{aligned}$$

□

One of the main uses of the idea of an asymptotic distribution is in providing approximations to the cumulative distribution functions of statistical estimators. An asymptotical distribution estimator is a consistent estimator whose distribution around the true parameter $\widehat{a}_{mle}$ approaches some distribution with standard deviation shrinking in proportion

to n as the sample size n grows. Specifically, the asymptotic distribution of the maximum likelihood estimator is derived. In the asymptotic analysis of the ML estimators, the main challenge is to find the asymptotic distribution of the estimator. We get the asymptotic distribution of the estimator $\widehat{a}_{mle}$ using the theorem below.

Theorem 2.1. *The asymptotal distribution of $\widehat{a}_{mle}$ is given in terms of Gumbel's distribution. We have*

$$\left(y_{(1)} - \theta \log n\right) \to Z \text{ in distribution} \tag{2.19}$$

when $n \to \infty$, where the distrution of Z is defined by

$$F_Z(t) = 1 - \exp\left(-e^{-\frac{t-a}{\theta}}\right).$$

Proof. We can already predict that for n sufficiently large the expectation of $\widehat{a}_{ml}$ will be close to a. This needs to be clarified. We know that

$$\lim_{n\to\infty} \left(1 - \frac{e^{-a}}{n}\right)^n = e^{-e^{-a}}$$

The sequence of general term $y_{(1)} - \theta \log n$ converges in distribution towards a limit that one seeks to determine.

$$\begin{aligned} P\left(y_{(1)} - \theta \log n \leq t\right) &= P\left(y_{(1)} \leq t + \theta \log n\right) \\ &= 1 - \left(P\left(y_{(1)} \geq t + \theta \log n\right)\right)^n \\ &= 1 - \left(1 - e^{-\frac{(t+\theta \log n - a)}{\theta}}\right)^n \\ &= 1 - \left(1 - \frac{e^{-\frac{t-a}{\theta}}}{n}\right)^n \end{aligned}$$

We know that

$$\lim_{n\to\infty} \left(1 - \frac{e^{-\frac{t-a}{v}}}{n}\right)^n = \exp\left(-e^{-\frac{t-a}{\theta}}\right) = G\left(\frac{t-a}{\theta}\right)$$

where G is the distribution function of Gumbel's law. The cumulative distribution function of the Gumbel distribution is :

$$F_{\text{Gumbel}}\left(x; \mu, \beta\right) = \exp\left(-e\left(\frac{\mu - x}{\beta}\right)\right) \tag{2.20}$$

The standard Gumbel distribution is the case where $\mu = 0$ et $\beta = 1$. The Gumbel distribution is used to model the distribution of the maximum (or the minimum) of a

number of samples of various distributions. We conclude that, when $n \to \infty$

$$y_{(1)} - \theta \log n \to Z \text{ in distribution} \tag{2.21}$$

Where the distribution of Z is defined by

$$P(Z \geq t) = 1 - F_Z(t) = \exp\left(-e^{-\frac{t-a}{\theta}}\right) = F_{Gumbel}(a; t, \theta)$$

Gumbel has shown that the maximum value in a sample of a random variable following an exponential distribution minus the logarithm of the sample size approaches the Gumbel distribution closer with increasing sample size. □

Thus the limiting distribution of the maximum likelihood estimator is linked with Gumbel's distribution. This distribution can be analyzed and used to understand the high sample behavior of the $\widehat{a}_{mle}$. In hydrology, Gumbel's law is used to analyze variables such as monthly and annual maximum values of daily precipitation and river flow volumes, and also to describe droughts.
A second estimator is developed below, for the purpose of comparison. 2.2 Ordinary Least Squares In statistics, ordinary least squares (OLS) is a linear least-squares method for estimating the unknown parameters in a linear regression model. OLS method chooses the parameters by minimizing the sum of the squares of the differences between the observed dependent variable in the given data set and those predicted by the linear function of the independent variable. The resulting estimator can be expressed by a simple formula, especially in the case of simple linear regression. There are several methods for constructing an estimator; among them, the least-squares method (OLS) and the maximum likelihood method are the most used.

Proposition 2.3. *Assume that θ is known, the OLS estimator $\widehat{a}_{ols}$ of a is $\bar{y} - \theta$, where $\bar{y}$ is the empirical mean of y_i.*

Proof. The error term accounts for the variation in the dependent variable that the independent variables do not explain. For the model to be unbiased, the average value of the error term must equal zero.

Let $y_i = a + u_i = a + \theta + \varepsilon_i$ which implies $E\varepsilon_i = 0$ and $E(Y_i) = a + \theta$. Note $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. The OLS method consists of minimizing:

$$S(a) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - \theta)^2 \tag{2.22}$$

The solution to the problem of minimization (28), denoted $\widehat{\alpha}_{\text{ols}}$, is given by

$$\widehat{a}_{\text{ols}} = \frac{1}{n}\sum_{i=1}^{n} y_i - \theta \tag{2.23}$$

Let's calculate the expectation and variance of $\widehat{a}_{\text{ols}}$

- $E\left(\widehat{a}_{ols}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} Ey_i - \theta = a$
- $\text{Var}\left(\widehat{a}_{\text{ols}}\right) = \text{Var}(\bar{y}) = \frac{\theta^2}{n}$.

□

Remark. Note that the OLS estimator of a is different from the previously determined maximum likelihood estimator.

There is a reason why we should not use OLS, because there is a violation of the usual assumptions. Indeed, OLS assumes that the mismatch between what is expected and observed is $E\left(u_i\right) = 0$. Alas in our case $E\left(u_i\right) = \theta$.

The error term accounts for the variation in the dependent variable that the independent variables do not explain. For the model to be unbiased, the average value of the error term must equal zero. The OLS estimator is identical to the maximum likelihood estimator (MLE) under the normality assumption for the error terms.

That assumption is not necessaire for the validity of the OLS method. However, if we assume that the normality assumption does not hold, then some properties must have to be added. In that case, we can get an OLS estimator.

The least-squares estimators are point estimates of the linear regression model parameters. However, generally, we also want to know how close those estimates might be to the reals values of parameters. The expectation and variance of $\widehat{a}_{ols}$ are

- $E\left(\widehat{a}_{ols}\right) = a$ and
- $\text{Var}\left(\widehat{a}_{ols}\right) = \frac{\theta^2}{n}$.

It is therefore obvious that $\widehat{a}_{ols}$ is an unbiased estimator of a. An unbiased estimator gives us estimates of the unknown parameter that, on average, are around that parameter. The bias being equal to zero, we deduce the MSE :

$$\text{MSE}\left(\widehat{a}_{ols}\right) = \text{Var}\left(\widehat{a}_{ols}\right) = \frac{\theta^2}{n}. \tag{2.24}$$

The variance $\text{Var}\left(\widehat{a}_{ols}\right)$ tends to 0 when n tends to infinity. We conclude that $\widehat{a}_{ols}$ is an efficient estimator of a. Finally, the MSE obtained for $\widehat{a}_{mle}$ tends to 0 fast than that obtained with the estimator $\widehat{a}_{ols}$, we conclude that the estimator $\widehat{a}_{mle}$ is better than $\widehat{a}_{ols}$.

We have established that the constant in an OLS regression model has something to do with the mean of the response variable. In particular, in interceptonly models, the intercept is almost equal to the average of the response variable. If the data errors are Gaussian and independent, the OLS estimators will be maximum likelihood estimators and will be unbiased and of minimal variance. However, if the noise is not Gaussian, the OLS adjustment will give parameter estimates that may be biased. Even when normality does not hold, the Gauss-Markov theorem states that the best linear unbiased estimator of regression coefficients is still yielded by OLS estimation, so long as the errors have expectation zero, are uncorrelated, and have equal variance.

2.3 Simulation study

This section is established with the intention of examining the performance quality of the estimators under study over a finite sample size n.

The simulation is conducted for a certain value of the parameter to be estimated, namely a.

Tableau 2.1: Simulation results: the values of the MSE (for both OLS and MLE estimators) with the corresponding Bias.

Estimators	OLS		MLE	
sample size n	MSE	Bias	MSE	Bias
50	4.1301×10^{-2}	0.0132	3.3325×10^{-4}	0.1454
200	2.4581×10^{-3}	0.0059	6.7712×10^{-5}	0.0131
500	2.2041×10^{-4}	0.0011	2.3865×10^{-6}	0.0027
1000	2.4891×10^{-4}	0.0042	3.6711×10^{-7}	-0.0085

2.3.1 Design of simulation

We generate data that fits our model as follows:

- Generate n iid random variables $\{\varepsilon_i\}_{i=1}^{n}$ from $\mathcal{E}xp\,(\theta^{-1})$,
- $y_i = a + \theta + \varepsilon_i$ for all $i = 1, \ldots, n$. Where the intercept a is chosen arbitrarily.

Remark. The intercept a and the scale parameter θ have been appropriately chosen in order to have a good model. Actually, the intercept would be better to be moderately and the exponential parameter small. This helps to have a better comparison.

The Mean Square Error (MSE) is chosen to be a criterion for quantifying the performance of our estimators. The MSE of an estimator $\hat{\theta}$ with respect to an unknown

parameter θ is defined as

$$\text{MSE}(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2.$$

2.3.2 Consistency Results

To give an overview of the influence of the sample size n on the quality of fit, the least square (OLS) and maximum likelihood estimators (MLE) $\hat{a}_{OLS}$ and $\hat{a}_{MLE}$ respectively was implemented from Model 1 which contaminated by exponential errors with $\theta = 0.25$. See Figure 1. To exhibit more comparison of the influence of the sample size n on the estimation fit, the values of MSE and Bais are computed from Model 1 and summarized in Table 1. The first column displayed for the different values of n, the second column displayed for the results(MSE and Bias) of $\hat{a}_{\text{ols}}$. While the last column is for $\hat{a}_{\text{mle}}$.

Remark. From the simulation results in Table 1 and Figure 1, we can see that the quality of fit depends directly on the estimation method and the sample size n. Actually, the larger the sample size, the better the quality of performance will be. Furthermore, the quality of fit declines substantially for the Ordinary Least Squares method compared to the Maximum likelihood method but it increases with a sufficiently large sample size.

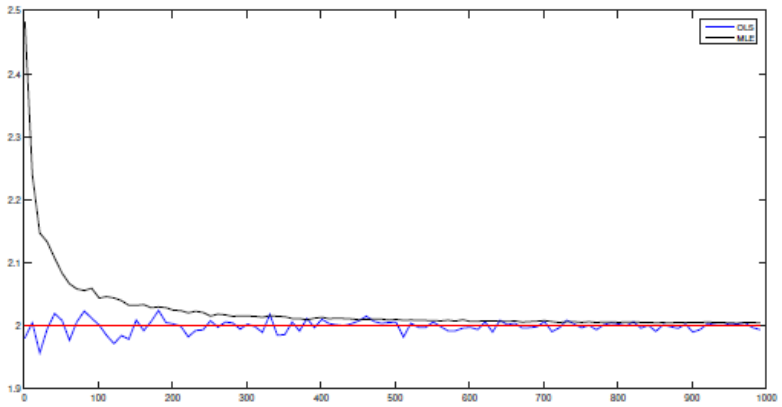


Figure 2.1: Simulation results of estimation from Model 1 for $\theta = 0.25$ and $a = 2$. The red line corresponds to the true intercept, the blue line corresponds to estimation by the Last Square Error method, the black line corresponds to the estimation by the Maximum Likelihood Method.

2.3.3 Asymptotic normality

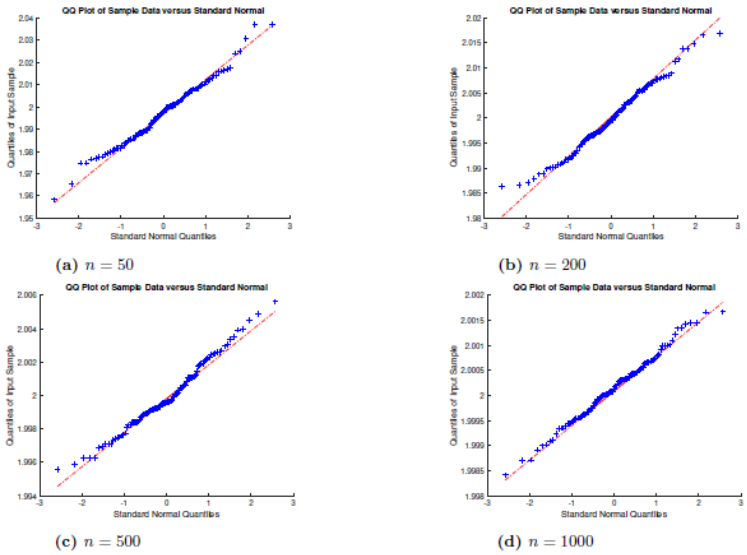
In this subsection, we examine the asymptotic normality of the understudy estimators throughout normal-probability plots. For this aim, we only consider the estimation given by the Ordinary Least Squares method. And for better comparison this estimator was

implemented here for $\theta = 0.25, m = 100$ iterations, and $n = 50, 200, 500$ and 1000 . The results of this numerical implementation are summarized in Figure 2.

Remark. From graphs summarised in Figure 2, we can see that for the asymptotic normality, the estimator provides good performance for a large sample size. This indicates that the convergence in distribution becomes better more and more along with n .

2.4 Implementation to The Number of Lynx in Canada

The data set R package contains data on the number of lynx caught per year in Canada during the period of 1821-1934. Figure 1 presents the corresponding histogram; It is may easy to think of an exponential distribution. In fact, it has a decreasing tendency, furthermore, the observations are all of positive values. From the graph in Figure 4, we can see that the data might contain an invariable intercept. Hence, our goal here is to compare the adjustment of our



(a) $n = 50$ (b) $n = 200$

(c) $n = 500$ (d) $n = 1000$

Figure 2.2: The normal-probability plots of the ordinary least squares estimator for $n = 50n = 200, 500$ and $1000, \theta = 0.25\%$.

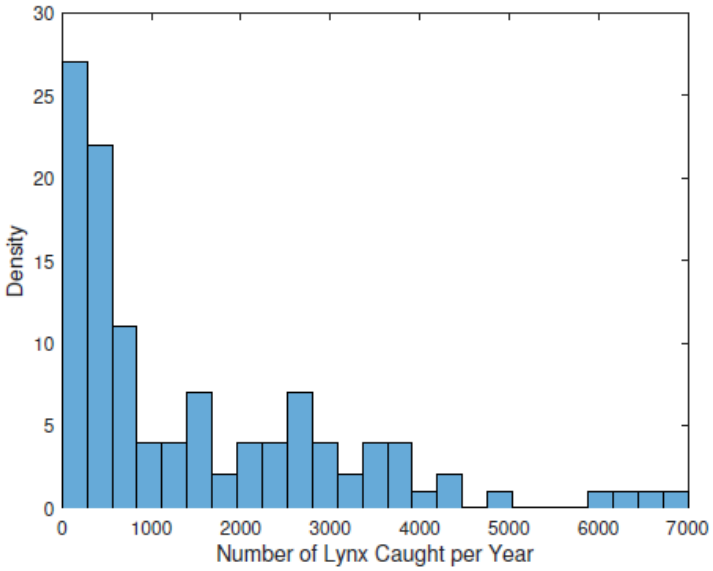


Figure 2.3: Histogram for data of the number of lynx caught per year in Canada from 1821 to 1934 .

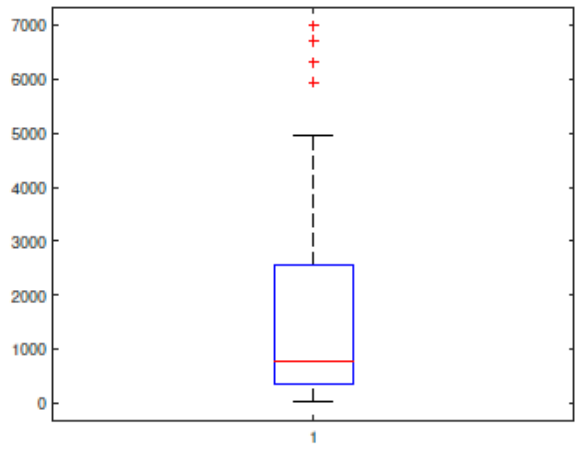


Figure 2.4: Boxplot of the number of lynx caught per year in Canada from 1821 to 1934.

data under the two possible models, namely:

1. The presence of the intercept: In this situation, we suppose that the data are coming from the following model

$$\text{lynx}_i = a + \varepsilon_i \quad \text{for all } i = 1, \ldots, 114. \tag{2.25}$$

2. Ignoring the intercept: We assume that the data are from

$$\text{lyn}_i = \varepsilon_i \quad \text{for all } i = 1, \ldots, 114, \tag{2.26}$$

where lynx x_i stands for a number of lynx caught in the i-th year between 1821 and 1934. The random variables $\{\varepsilon_i\}_{i=1}^{114}$ are supposed to be iid and follow an exponential distribution of shape parameter μ to be estimated.

Estimation At this level, we are interested to estimate the exponential parameter θ and use this latter to conclude the intercept estimator.

In the situation where the intercept is considered, and from the result in Corollary 2, we have

$$\hat{\theta} := \left(\overline{\text{lyn}\, x} - \inf\left(\text{lyn}\, x_i\right)\right) = 0.00066710. \tag{2.27}$$

Where for the second model, we have

$$\hat{\theta} := \overline{\text{lyn}\, x} = 0.00065018.. \tag{2.28}$$

Using the result in previous sections 2.1 and 2.2, we have

$$\hat{a}_{\text{ols}} = 37.9919, \tag{2.29}$$

$$\hat{a}_{mle} = 39. \tag{2.30}$$

To provide a more clear comparison, we consider the following data

$$\text{lyn}\, x_i' = \text{lyn}\, x_i - \hat{a}_{mle} \text{ for all } i = 1, \ldots, 114. \tag{2.31}$$

So, the aim now is to compare the adjustment of our new data and see which model explains more clearly the number of lynx caught per year in Canada.

Remark. Figures 5-6 reveal that the data lynxí coincides better with an exponential model with an intercept compared to a free-exponential model(ie without intercept). Hence, we can conclude that the data set on the number of lynx caught per year in Canada on the period between 1821 and 1934 refers to an exponential model with intercept.

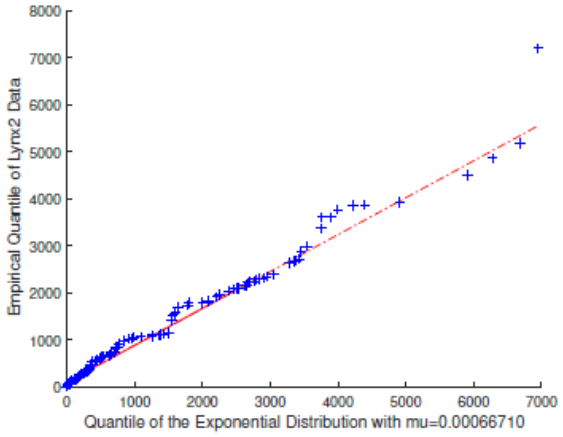


Figure 2.5: The quantile-quantile plot of the quantiles of the sample data lynx2 versus the theoretical quantile values from an exponential distribution with $\mu = 0.00066710$

.

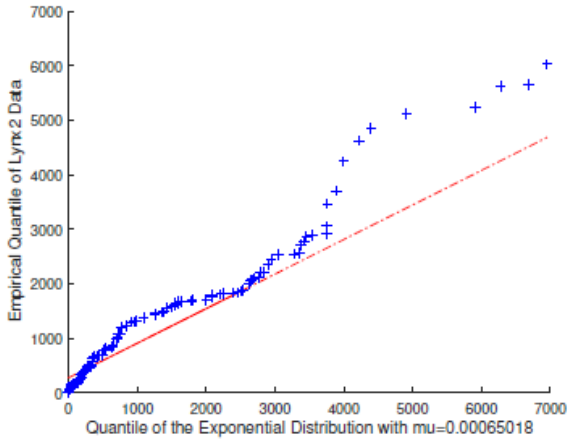


Figure 2.6: The quantile-quantile plot of the quantiles of the sample data lynx2 versus the theoretical quantile values from an exponential distribution with $\mu = 0.00065018$

.

2.5 Conclusion

The estimated parameter differs for the two methods. The advantage of the likelihood method proposed is significant. Both techniques could be employed to construct consistent estimators.

But in many contexts (if the errors are Gaussian e.g), consistency is from, a practical viewpoint, an attainable goal. We have illustrated the performance via numerical studies.

Chapter 3

Statistical Analysis of Data on Cereal Production in Algeria.

3.1 Introduction

Agro-climatic constraints, combined with the recent effects of climate change, are weighing in on the development of Algerian agriculture. Research on the impact of climate changes howed declining rainfall levels. The weather projections suggest that Algeria will have a sharp increase in aridity which makes it more vulnerable to water stress and desertification. Agro-climatic in Algeria models predict that climate change will modify the water cycle, contributing to a degradation of agricultural land, a decline in agricultural production and yields, and a loss of biodiversityBessaoud et al. (2019).

Algeria is a dry country that belongs to the arid-semi-arid triangle. It is one of the regions of the world located in a climate transition zone, and which, therefore, is subject to both the influence of humid and temperate zones (in winter) and the influence of the desert. On a map, Algeria occupies a considerable area, but arable land is in limited supply. The fertile lands susceptible to appropriation by farmers are scattered over vast areas almost desert. Olive trees and vines trace the lines of Mediterranean agriculture in Algeria. Cereals, legumes, fodder and spring cereals are also grown in the best watered areas Bessaoud (1999).

Irrigation has been a central feature of agriculture for over 5,000 years. When it comes to irrigation, agriculture is the first area that comes to mind. Irrigation is the artificial process of applying controlled amounts of water to land to assist in production of crops. It use to increase yields, but also to compensate for the lack of rainwater. Thereby, if the rainwater is not sufficient, this alternative is the only solution. Agriculture that relies only on direct rainfall is referred to as rain-fed.

Agriculture can overcome, through the extension of supplementary irrigation, the problem of water stress which prevails almost throughout the agricultural year. Indeed, this beneficial technique would allow Algeria to increase its yields, especially for wheat crops, despite the lack of water resources. In some cereal regions, the yield of durum wheat per hectare can reach substantial quantities thanks to the extension of irrigated areas. The agricultural strategy is fundamentally focused on the development of strategic sectors, in particular soft wheat, corn, sugar crops and oil seeds, which still constitute the bulk of national food imports. To achieve this objective, the country intends to extend these crops, particularly in remote areas (Sahara), through incentive measures. Improving cereal production is crucial in developing the standard of living in Algeria. Actually, it should be part of any future strategy for the country. Most of the arable land in Algeria is in a Mediterranean climate, where droughts are common and rainfall is distributed unevenly throughout the year. Research on the impact of climate variability and irrigation on cereal production is necessary due to the effects on the uneven performance of crops. The objective of this work is to analyze the parameters like rainfall and irrigation that influences the yield of crop and to establish a relationship among these parameters.

The data used are cereal production, irrigation (areas irrigated) and rainfall. Those data were subjected to a statistical analysis. First, Principal component analysis was applied to classify the data in order to determine the relative importance of the various regions to the evaluation of cereal production. Second, Regression analysis was used to analyze the factors and their effects on crop yield.

This work highlights the estimation of crop production. Especially with regard to the development of regression models using climatologically parameters as independent variables and crop yield as a dependent variable. One subsequently discusses the significance of prevision for crop yield estimation.

In this context, the paper also covers Principal Component Analysis, a statistical procedure used to reduce a number of correlated variables into a smaller number of uncorrelated variables called principal components. These hidden latent variables are called factors or components hence the name of analysis.

The theme developed in this work has been treated by previous authors. A brief overview is given in the following. Knowing the water level of the Akosombodam is heavily related to the hydroelectric power. In Asare et al. (2018), the principal component regression was applied to the input variables in the goal of reducing the large number to a few main components in order to explain the variations in the original data set (see Sellam and Poovammal (2016)) analyzes the environmental parameters such as area under cultivation, annual rainfall, and food price index that influences the yield of the crop and establishes a relation between these parameters. An analysis of published data

on genetic relations between dry grain yields and protein content of cereals is presented in Simmonds (1995). In all, 106 usable regressions across geneotypes were assembled. Result of principal component analysis in Camelo-Méndez et al. (2012) is an equation with seven variables, area, perimeter, length, width, thickness, sphericity and color, which was useful for distinguishing between nine different cultivars. InBodroža-Solarov et al. (2014), gas chromatographic-mass spectrometric analytical data for both fractions were analyzed by multivariate statistical techniques to model classes of different common wheat and spelt cultivars.

The forthcoming section presents used methods. Section 3 discusses the obtain results. Conclusion is proposed in Section 4 .

3.2 Methodology

3.2.1 Regression analysis

Regression methods continue to be an area of active research. It is the method of using observations to quantify the relationship between a target variable, also referred to as a dependent variable and a set of independent variables. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. The conditional mean of the response given the values of the explanatory variables is assumed to be an affine function of those values. Linear regression models are often fitted using the least squares approach. Ordinary least squares (OLS) find the value of parameters that minimizes the sum of squared errors. The Gauss-Markov theorem states that OLS produces estimates that are better than estimates from all other linear model estimation methods when the assumptions hold true. The coefficients computed are the ones that best fits the data. The analysis done using linear regression is based on the identification of factors whose depends the dependent variable. The contributions related to the regression model are extensive; the main reference used here is book Coursol (1981) , **?**, and Saporta (2006). Regression analysis is widely used for prediction, error reduction and forecasting. It can be used to infer causal relationships between the independent and dependent variables.

3.2.2 Principal Component Analysis

In the book Anderson (1958), principal components are linear combinations of random or statistical variables which have special properties in terms of variances. For example, the first principal component is the normalized linear combination (the sum of squares of

the coefficients being one) with maximum variance. In effect, transforming the original vector variable to the vector of principal components amounts to a rotation of coordinate axes to a new coordinate system that has inherent statistical properties. The principal components turn out to be the characteristic vectors of the covariance matrix. Thus the study of principal components can be considered as putting into statistical terms the usual developments of characteristic roots and vectors (for positive semidefinite matrices). In statistical practice, the method of principal components is used to find the linear combinations with large variance. In many exploratory studies the number of variables under consideration is too large to handle. Since it is the deviations in these studies that are of interest, a way of reducing the number of variables to be treated is to discard the linear combinations which have small variances and study only those with large variances (see Anderson (1958)).

According to Jollite's see Jolliffe (1990), the central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. Suppose that x is a vector of p random variables, and that the variances of the p random variables and the structure of the covariance or correlations between the p variables are of interest. The first step is to look for a linear function $\alpha_1 x$ of the elements of x having maximum variance, where α_1 is a vector of p constants. Next, look for a linear function $\alpha_2 x$, uncorrelated with $\alpha_1 x$ having maximum variance, and so on. Up to pPCs could be found, but it is hoped, in general, that most of the variation in x will be accounted for by m PCs, where $m \ll p$. The reduction in complexity achieved by transforming the original variables to PCs. More generally, if a set of $p(> 2)$ variables has substantial correlations among them, then the first few PCs will account for most of the variation in the original variables. Conversely, the last few PCs identify directions in which there is very little variation; that is, they identify near-constant linear relationships among the original variables. Having defined PCs, we need to know how to find them. Consider, for the moment, the case where the vector of random variables x has a known covariance matrix Σ. This is the matrix whose $(i,j)^{th}$ element is the (known) covariance between the i^{th} and j^{th} elements of x when $i \neq j$, and the variance of the j^{th} element of x when $i = j$. The more realistic case, where Σ is unknown, follows by replacing Σ by a sample covariance matrix S. It turns out that for $k = 1, 2, \cdots, p$, the kth PC is given by $z_k = \alpha_k x$ where α_k is an eigenvector of Σ corresponding to its k^{th} largest eigenvalue λ_k. Furthermore, if α_k is chosen to have unit length ($\alpha_k' \alpha_k = 1$), then $\text{var}(z_k) = \lambda_k$, where $\text{var}(z_k)$ denotes the variance of z_k. The

k^{th}PC will be taken to mean the PC with the kth largest variance, with corresponding interpretations for he ' k th eigenvalue' and ' k^{th} eigenvector' (see Bülthoff (1996)).

Assume n observations have been made on p variables, and let $\widetilde{X}$ denote the $n \times p$ matrix of observations. Let x_{ij} denote the observation in the i^{th} row and j^{th} column of $\tilde{X}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, p$. Before PCA, column sample means are subtracted from all observations. The PCA could be performed directly on these mean-centred data or after division with column sample means or after division with column sample standard deviations. Let X denote the $n \times p$ matrix used for PCA Diday (1982).

Sometimes variables are highly correlated in such a way that they contain redundant information. This allows the reduction of variables. Finding the directions of the data that contain the greatest variance is achieved by decomposing the sample correlation matrix into eigenvalues. The eigenvalues are ordered in decreasing order; this allows finding the main components in order of significance Diday (1982). These principal component analysis methods exist in many books, including Anderson (1958), Diday (1982), Jolliffe (1990) and Dillon (1984).

3.3 Application

3.3.1 Principal component analysis

The study begins with the use of a descriptive statistical method, namely; the principal component analysis.

3.3.1.1 Principal component analysis wilayas vs productions

Principal component analysis (PCA) allows to summarize and to visualize the information in a data set containing individuals (observations) described by multiple correlated quantitative variables. This analysis assumes that the directions with the largest variances are the most important. Principal components represent the directions of the data that explain a maximal amount of variance, that is to say, the lines that capture most information of the data. Computing the eigenvectors and ordering them by their eigenvalues in descending order, allow finding the principal components in order of significance. The dataset contains the harvest of 4 types of cereals for 48 wilayas.

The correlation circle shows the correlations between the components and the initial variables. To interpret each component, we must compute the correlations between the original data and each principal component. The distance between variables and the origin measures the quality of the variables. Variables that are away from the origin are

well represented.

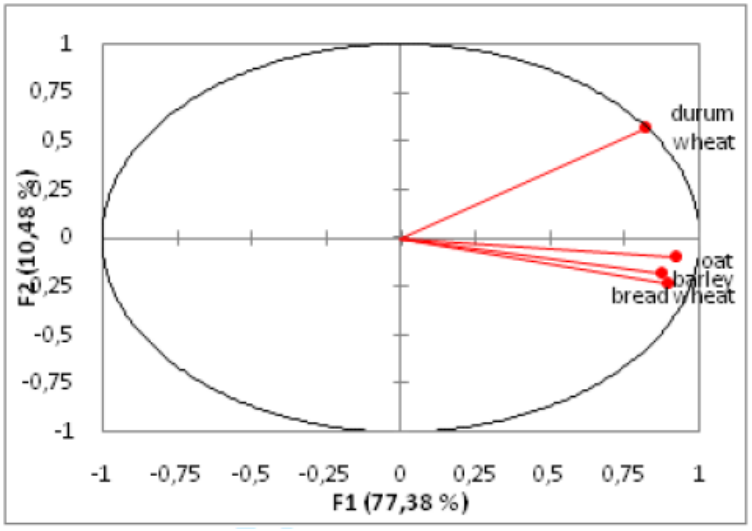


Figure 3.1: Correlation circle

The group of three very tightly knit variable markers for 3 variables "productions" barley, bread wheat and oat, suggests a group of highly correlated variables. In dataset, the variable durum wheat is pointing up and to the right, and then the rest of the variables are bunched up together pointing down and to the right. Since all the variables are pointing to the right, they are correlating on this first principal component F1, it represents all cereal productions. The more interesting might be the second principal component F2, since that is where we see a clear division between the durum wheat on the one hand and all the other cereal productions on the other hand.

The dimension with the most explained variance is called F1 and plotted on the horizontal axes; the second-most explanatory dimension is called F2 and placed on the vertical axis. Inside this 2-dimensional circle, the original 4 variables are projected in red onto this 2-dimensional factor space. If 2 red lines are pointing in the same direction then they are highly correlated, if they are orthogonal they are unrelated and if they are pointing in opposite directions they are negatively correlated.

The eigenvalues $\lambda_k, k = 1, 2, \ldots p$ of the sample covariance or correlation matrix reflect the relative importance of the principal components. The first two PCs account for 77.378% and 10.478%, respectively, of the total variation in the dataset, so the two-dimensional scatter plot of the 48 wilayas given in Figure 1 is a very good approximation to the original scatter-plot in four-dimensional space. It is, by definition, the best two-dimensional plot preserving variance of the data, representing more than 88% of the total variation. The percentage of total variation is an obvious measure of how good a two-dimensional representation is. There is, therefore, some distortion in the two-dimensional

representation.

The interpretation of the principal components is based on finding which variables are most strongly correlated with each component. The PCs can then be interpreted on the basis of which variables they are most correlated in either a positive or negative direction. The very high proportion of variability explained by the two-dimensional principal subspace provides solid ground for conclusions. The correlation between a variable and a principal component (PC) is used as the coordinates of the variable on the principal component. The representation of variables differs from that of observations: observations are represented by their projections, but variables are represented by their correlations.

All of the loadings in the first PC have the same sign, so it is a weighted average of all variables, representing "overall size". In Figure 2, large productions are on the right and small productions on the left. The first principal component increases with increasing production. The second PC has negative loadings for three variables and positive loading for the durum wheat variable, representing an aspect of the "shape" of production. This second component is a contrast of durum wheat (0.873) against bread wheat (-0.417), oats (-0.201) and barley (-0.175). Wilayas near the top of Figure 2 have higher durum wheat productions, relative to their other productions, than those towards the bottom. The relatively compact cluster of points in the bottom half of Figure 2 is thought to correspond to small productions. Such PCA plots are often used to find potential clusters. Based on Figure 2 , it is clear that the wilayas of Sétif, Tiaret and Sidi-Bel-Abbès form a distinct cluster on the right. These wilayas are characterized by large productions. Projecting the marker for "Tiaret" onto the positive direction of all variable markers suggests that wilaya Tiaret (on the right of the plot) has a large cereal production. Inspection of the data matrix confirms that it is the largest production on two of the four variables, and close to largest for the durum wheat. The South Wilayas (on the right) have small productions. Individuals whose markers are close to the origin have values close to the mean for all variables, such as Saida and Constantine. Compact cluster structure in Figure 2 in the left of the plot, is formed by the wilayas which produces little or no cereal.

The latest component explains only 0.005 of the variability in the data. This component may not be important enough to include. The third component is correlated with barley. This component could be viewed as a measure of the barley production. The third and fourth principal component explain very small percentages of the total variation, so it would be surprising if it found that they were very informative and separated the groups or revealed apparent patterns. Thus, $PC3 - PC4$ can be ignored, they contribute little (12%) to explaining the variance, and express the data in two dimensions instead of four.

Conclusion : it is essentially the wilayas of the east and center produce more durum wheat than the other cereals; while in the west it is bread wheat as well barley and oats grown there. It should also be noted that southern wilayas have low yields in different types of cereals (practically nonexistent).

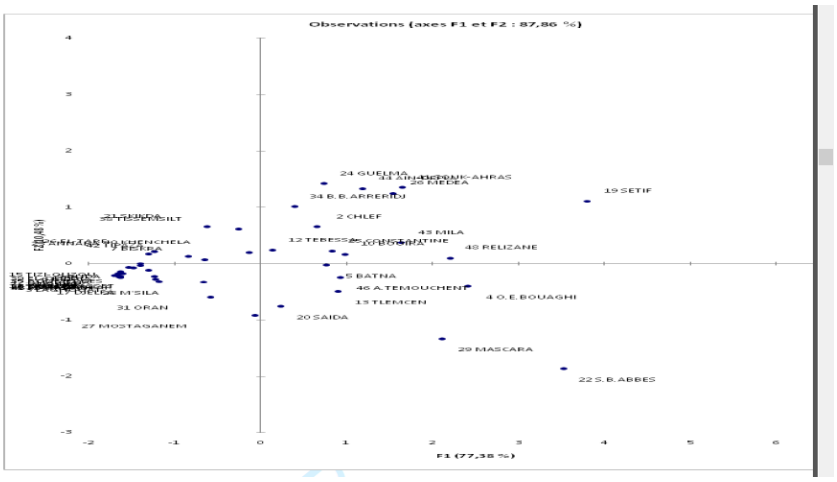


Figure 3.2: Representation of data points on the two first components

3.3.1.2 Principal component analysis years vs productions

It may be interested in describing and analyzing how years differ in the cereal productions. The following analysis concerns the principal component analysis carried out on the data table contains years and types of cereals. The table is of dimension (13×4), the productions are in quintals. The principal component analysis was done on the correlation matrix, even though it could be argued that, since all measurements are made in the same units, the covariance matrix might be more appropriate. The correlation matrix was preferred because the covariance matrix gives greater weight to larger, and hence more variable, measurements, such as durum wheat production.

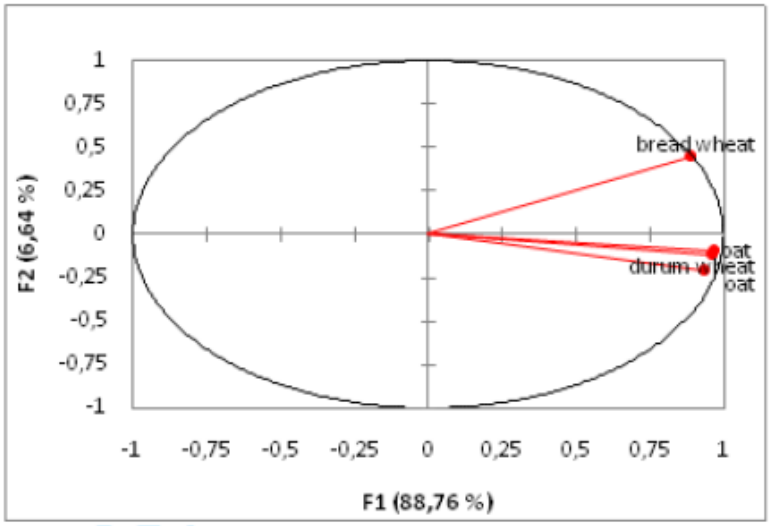


Figure 3.3: correlation circle

The group of three very tightly knit variable markers for 3 variables "productions" barley, bread wheat and oat, suggests a group of highly correlated variables. In dataset, the variable durum wheat is pointing up and to the right, and then the rest of the variables are bunched up together pointing down and to the right. Since all the variables are pointing to the right, they are correlating on this first principal component $F1$, representing all cereal productions. The more interesting might be the second principal component F2, since that is where we see a clear division between the durum wheat on the one hand and all the other cereal productions on the other hand.

The first eigenvalue is equal to 3.550 and the PC1 provides 88.579% of the initial information (Table 3). This means that if we represent the data on this axis, then we will have 88.579% of the total variability, which will be preserved. Using the first two PCs, we obtain 95.403% of the total inertia of the initial data table. Consequently, the projection on the first two axes offers the quality of a faithful representation of the initial data. The principal component analysis constructs low-dimensional plots of a set of data from information about similarities or dissimilarities between observations. In any case, a description of the sample, rather than inference about the underlying population, is often required, and PCs describe the major directions of variation within a sample, regardless of the sample size. The first PC has positive coefficients for all variables and simply reflects overall "size" of the individuals.

A graph of these data with respect to the first two PCs has been given in Figure 4 , and it was noted that the first PC succeeds in separating years with high productions from years with low productions. The second PC accounts for slightly less than 20% of the total variation. This second PC contrasts some of the productions with others, and

can often be interpreted as defining certain aspects of "shape". The second PC can be interpreted as a contrast between 'bread wheat' and the others' productions. In general, the first two PCs account for a substantial proportion of total variation 86.5%. Because there are relatively strong correlations among the 4 variables, the effective dimensionality of the 4 variables is around 1 or 2 , a substantial reduction then occurs. If a good representation of the data exists in a small number of dimensions, then the principal component analysis will find it, since the first two PCs give the 'bestfitting' 2-dimensional subspace. Interpreting observations consists of examining their coordinates and especially their resulting graphical representation called the first principal plane (Figure 4). The aim is to see how the observations are scattered, which observations are similar and which observations differ from the others. The use of the results of the analysis of the variables allows for the interpretation of the observations. For example, the first component is strongly correlated with the original variables; this means that years with large positive coordinates on the axis 1$(2009, 2012)$ are characterized by the fact that productions have values much larger than average (the origin of the axes represents the center of gravity of the data cloud). And vice versa years with negative coordinates on the axis 1$(2000, 2008)$ are characterized by the fact that productions have values much lower than average.

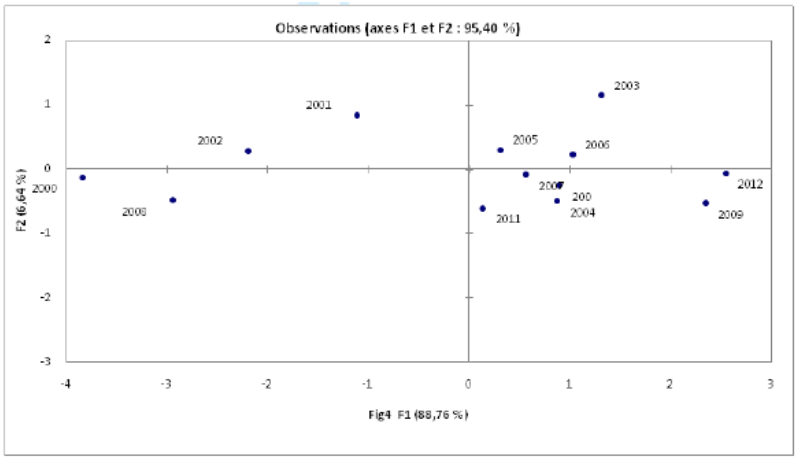


Figure 3.4: Representation of data points on the two components.

An examination of the evolution of cereal production reveals three clusters. The first group contains the years when there was high cereal productions (2009, 2012), the second consists of years with low productions $(2000, 2008, 2002$ and $2001)$ and the third includes the years when productions was average (the remainder). The coverage rate of national production remained constant over the whole period from 2000 to 2012 ; with the exception of a few exceptional campaigns such as 2008 , a particularly dry year and 2009 , the year in which production was the best of all time. Thus, we see that the country

has managed to "maintain" the same level since 2000 , with limited variations. This is an important achievement, in the sense that the performance in terms of production volume does not record such large and abrupt variations from one year to the next; apart from the two years mentioned.

The principal component analysis differs from linear regression in that the principal component analysis minimizes the perpendicular distance between a data point and the principal component, whereas linear regression minimizes the distance between the response variable and its predicted value.

3.3.2 Regressions Analysis

3.3.2.1 The regression production of durum wheat vs. irrigation

Regression model is widely used for prediction, error reduction and forecasting. It can be used to infer causal relationships between the independent and dependent variable. Regression analysis is a way of mathematically sorting out which of those independent variables does indeed have an impact on dependent variable.

First, a regression analysis of durum wheat production vs. irrigation is performed. High quality durum wheat is grown in areas with a relatively dry climate, with warm days and cool nights during the growing season. The Highlands (Hauts Plateaux) are the main cereal zones of Algeria. Durum wheat produced in moist conditions tends to have lower vitreous grain content, making it less suitable for making pasta. The crop considered for analysis is durum wheat because it is the most common crop cultivated in many areas of Algeria. The production of durum wheat y_i and irrigation x_i were measured over a period of 13 years. For linear regression, the model is as follows

$$y_i = b + ax_i + e_i$$

The random variable e_i is the error term in the model. In this context, error does not mean mistake but is a statistical term representing random fluctuations, measurement errors for example. Using the observed values x and y, parameters can be estimated and inferences such as hypothesis tests will be made. Also, the estimated model can be used to predict the value of y, for a particular value of x, in which case a measure of predictive precision may also be of interest. The correlation coefficient between production and irrigation is

$$r = 0.66408$$

The existence of a linear relationship between these two variables is, therefore, proven.

The estimates of the two parameters provide us the equation regression (see Table 6)

$$y = 162.71x + 5.6215 \times 10^5$$

This means that increasing irrigation by one unit increases durum production by 162.71. It can be said that there is a significant linear relationship between durum wheat production and irrigation. It is important to check if the model fits well with the data. Indeed, one of the objectives is to be able to predict the value of y, knowing a value of the variable x. However, if the fit is wrong, there is no hope of getting a good prediction. The Figure 5 presents the residuals which are the differences between the observed value of the dependent variable and the predicted value.

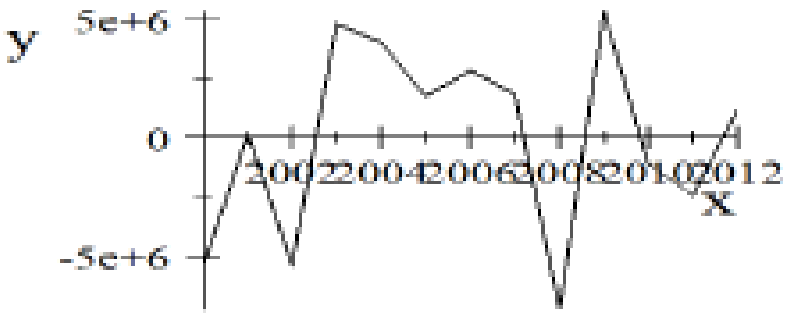


Figure 3.5: Residues

The biggest difference is during 2008 ; in other words, the production of durum wheat produced during 2008 is much lower than the expected production in view of the irrigation carried out. The opposite occurred during the year 2009 . Production exceeded "forecasts" that could be made in relation to production of durum wheat. 3.2.2 The regression production of durum wheat vs. rainfall The model is as follows:

$$y_i = b + ax_i + e_i$$

The random variable e_i is the error term in the model. The correlation coefficient between production and precipitation is $r = 0.85809$. This is favorable to the existence of a linear relationship between these two variables. Since there is a relationship between these two variables, then it is possible to build a model that predicts cereal production based on rainfall. The results are given in Table 7. We deduce the model

$$y = 35401x - 1593170 \tag{3.1}$$

Again, from table 7 , the p-value (0.0001748) being very low, the hypothesis $H_0 : a = 0$ is rejected. Therefore, the existence of a significant linear relationship between durum wheat production and rainfall is accepted.The results of R^2 clearly indicate that the crop's yield is highly dependent on the rainfall. Similarly, it is found that durum wheat yield plays a good role as a response variable for the explanatory variable rainfall.

Glancing at (3.1), it is noticeable that probably the production is higher when it rains a lot. If $x = 0$ then $y = -1593170$. The other regressions are summarized below.

- Irrigation does not affect the production of bread wheat.
- Rainfall affects the production of bread wheat.
- Irrigation does not affect barleyproduction.
- Rainfall affects the barley production.
- Irrigation does not affect oat production.
- Rainfall affects oat production.

Unequivocally, the null hypothesis can be rejected; the slope of the line is not null in the case of rainfall. On the other hand, the null hypothesis that the slope is zero is accepted almost systematically in the case of irrigation. This can be interpreted as follows: Irrigation does not affect the production of cereals other than the production of durum wheat; there is only rainfall that influences all productions.

Conclusion: Algerian agriculture relies mainly on rainfall rather than irrigation.

3.3.2.2 The regression production of durum wheat vs. irrigation and rainfall

Simple linear regression is useful, but it is oftentimes desirable to see how several variables can be used to predict a dependent variable. The response y is often influenced by more than one predictor variable. For example, the yield of a crop may depend on the amount of rainfall and irrigation. The model in this case, is written as

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + e_i$$

Where x_1 denote the variable "rainfall" and x_2 the variable "irrigation". The parameters estimators were calculated; theresults are given in Table 8 .Thus, the model is:

$$y_i = -7.2160 \times 10^6 + 29136 x_{1i} + 93.9 x_{2i}$$

The conclusion is that the coefficient of the first variable is significantly different from 0 (p-value 0.00028). In other words, the rainfall variable influences durum wheat production. Idem, the variable "irrigation" influences the production of durum wheat (Table 8). Yield prediction is one of the most critical issues faced in the agricultural sector. Uncertainties in the weather conditions lead irregularities in the production of the crops. Regression analysis is used to establish the relationship crop yield among these two factors and to identify their influence.

Conclusion and Discussion

This thesis consists of two parts namely: a part which studies the estimation in the intercept only regression model. The second concerns the study of cereal production in Algeria according to irrigation.

3.4 Intercept-only Model under Non-normality.

The main use of regression is to illuminate on a supposed linear relationship between predictor variables and an outcome variable. Generally, we suppose that the errors are independent and normally distributed with a zero mean and a variance σ^2. In the absence of explanatory variables, the mean can be a model by itself. The mean model is often the starting point for constructing forecasting models for time series data, including random walk models. In these case, it should be noted that there is no slope so the intercept is estimated by the mean response y. Often a model with intercept and predictors is compared to an intercept-only model to test whether the predictors add over and above the intercept only. In the intercept-only model, all forecasts take the value of the intercept. If we can find some mathematical transformation (e.g., differencing, etc.) that converts the original series into a sequence of values that are independently and identically distributed, we can use the mean model to obtain forecasts and confidence limits for the transformed series, and then reverse the transformation to obtain corresponding forecasts and confidence limits for the original series.

Violation of the normality assumption sometimes may be attributed to the skewed nature of the dependent variable. The data distribution may deviate from a Gaussian distribution in multiple ways. Continuous variables may deviate from normality in terms of skewness (showing a long tail on one side), kurtosis (curvature leading to light or heavy tails), and even higher-order moments. Many applications have positive response variables. Such variables usually have distributions right-skewed. The boundary at zero limits the left tail of the distribution. Departures from normality may have several causes. For example, they may be due by the presence of outlying values in the responses, among

other reasons. For this, several researchers proposed to perform regression analysis using a model that assumes a non gaussian distribution for the error terms.

The least squares technique has traditionally been justified by two assumptive arguments, it provides the maximal likelihood regression coefficients, if the errors are Gaussian and of all unbiased linear estimators, least squares have a minimal variance. Both of these properties have at times been adduced to call least squares the best of regression techniques. Because least squares possess in addition, the attribute of the computational facility, this method long has reigned as the foremost tool in reducing data to mathematically descriptive relationships. The first argument above assumes a normal distribution of the error terms. We argue that this supposition is often unwarranted and we can show that significant gains in likelihood maybe achieved when the regression technique allows for the more general class of error distribution. When the probability distribution of errors is assumed, it is possible, to obtain consistent and efficient estimates (minimum variance) of the parameters using the maximum likelihood method. This technique could be widely applied to non-Gaussian noise problems.

We have established that the constant in an OLS regression model has something to do with the mean of the response variable. In particular, in intercept only models, the intercept is almost equal to the average of the response variable. If the data errors are Gaussian and independent, the OLS estimators will be maximum likelihood estimators and will be unbiased and of minimal variance. However, if the noise is not Gaussian, the OLS adjustment will give parameter estimates that may be biased. Even when normality does not hold, the Gauss-Markov theorem states that the best linear unbiased estimator of regression coefficients is still yielded by OLS estimation, so long as the errors have expectation zero, are uncorrelated, and have equal variance. OLS method chooses the parameters by minimizing the sum of the squares of the differences between the observed dependent variable in the given data set and those predicted by the linear function of the independent variable. The resulting estimator can be expressed by a simple formula, especially in the case of simple linear regression. There are several methods for constructing an estimator; among them, the least-squares method (OLS) and the maximum likelihood method are the most used.

One of the main uses of the idea of an asymptotic distribution is in providing approximations to the cumulative distribution functions of statistical estimators. An asymptotical distribution estimator is a consistent estimator whose distribution around the true parameter approaches some distribution with standard deviation shrinking in proportion to n as the sample size n grows. Specifically, the asymptotic distribution of the maximum likelihood estimator was derived. In the asymptotic analysis of the ML estimators, the main challenge is to find the asymptotic distribution of the estimator. We has get the

asymptotic distribution of the estimator.

1. The simulation is conducted for a certain value of the parameter to be estimated. But in many contexts (if the errors are Gaussian e.g), consistency is from, a practical viewpoint, an attainable goal. We have illustrated the performance via numerical studies.

2. The data set R package contains data on the number of lynx caught per year in Canada during the period of 1821-1934. The aim is to compare the adjustment of the data and see which model explains more clearly the number of lynx caught per year in Canada. The data lynx coincides better with an exponential model with an intercept compared to a free-exponential model (ie without intercept). Hence, we can conclude that the data set on the number of lynx caught per year in Canada on the period between 1821 and 1934 refers to an exponential model with intercept. The estimated parameter differs for the two methods. The advantage of the likelihood method proposed is significant. Both techniques could be employed to construct consistent estimators.

3.5 Statistical Analysis of Data on Cereal Production in Algeria.

Agro-climatic constraints, combined with the recent effects of climate change, are weighing in on the development of Algerian agriculture. Research on the impact of climate changes howed declining rainfall levels. The weather projections suggest that Algeria will have a sharp increase in aridity which makes it more vulnerable to water stress and desertification. Agro-climatic in Algeria models predict that climate change will modify the water cycle, contributing to a degradation of agricultural land, a decline in agricultural production and yields, and a loss of biodiversity.

The data used are cereal production, irrigation (areas irrigated) and rainfall. Those data were subjected to a statistical analysis. First, Principal component analysis was applied to classify the data in order to determine the relative importance of the various regions to the evaluation of cereal production. Second, Regression analysis was used to analyze the factors and their effects on crop yield. This work highlights the estimation of crop production. Especially with regard to the development of regression models using climatologically parameters as independent variables and crop yield as a dependent variable. One subsequently discusses the significance of prevision for crop yield estimation. In this context, the paper also covers Principal Component Analysis, a statistical procedure

used to reduce a number of correlated variables into a smaller number of uncorrelated variables called principal components. These hidden latent variables are called factors or components hence the name of analysis.

The regression showed that rainfall has a strong influence on cereal production, unlike irrigation. This latest gives significant results for durum wheat but not for other types of cereals, despite the intensive irrigation introduced in some wilaya as Biskra. This wilaya is the most irrigated but cereal production is very low. Except the wilaya of Biskra is more known for its production of dates and therefore it is probably "the most irrigated wilaya" for this type of production. Conversely the wilaya of Sidi Bel Abbès is ranked among the first 3 wilayas that produce the most grain, but in terms of irrigation, it is among the last.

Using a principal components analysis, it was possible to determine a classification; the wilayas forme 3 groups according to their productions. Barley production is found to be much higher in the highlands than in other regions better known for the production of durum wheat in the center of the country and common wheat and oats in the west of the country. It is also noted that production differs from one type of grain to another. Farmers tend to prefer to invest in durum wheat and bread wheat because they have a better return on investment by exploiting them rather than growing oats. The second Principal component analysisenabled us to detect the years when production was weakest, such as the year 2008 , which experienced a severe drought. This gave a classification of years in relation to grain yields. There are two groups of years, namely the first years (2000-2002 and 2008) where the cereal yield is low and the more recent years (2003-2012) with a higher yield.

3.6 Perspectives

For a futur work, it would be intersting to Consider the context of a spatial study, where participants are interviewed about their Quality of Life, at a fixed time (cross sectional). The interviews consist, usually, to fulfill a questionnaire in which they are asked multiple choice questions, built in order to measure, the latent trait. The Graded Response model (GRM) and SAS Studio will be used to analyze the data.

Bibliography

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*, volume 2. Wiley New York.

Asare, I. O., Frempong, D. A., and Larbi, P. (2018). Use of principal components regression and time-series analysis to predict the water level of the akosombo dam level.

Batsidis, A. and Zografos, K. (2008). Multivariate linear regression model with elliptically contoured distributed errors and monotone missing dependent variables. *Communications in Statistics—Theory and Methods*, 37(3):349–372.

Belsley, D. A. (1991). A guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, 4(1):33–50.

Bessaoud, O. (1999). L 'algérie agricole: de la construction du territoire à l'impossible émergence de la paysannerie. *Insaniyat/. Revue algérienne d'anthropologie et de sciences sociales*, (7):5–32.

Bessaoud, O., Pellissier, J.-P., Rolland, J.-P., and Khechimi, W. (2019). Rapport de synthèse sur l'agriculture en algérie. projet d'appui a l'initiative enpard mediterranée.

Bodroža-Solarov, M., Vujić, Đ., Ačanski, M., Pezo, L., Filipčev, B., and Mladenov, N. (2014). Characterization of the liposoluble fraction of common wheat (triticum aestivum) and spelt (t. aestivum ssp. spelta) flours using multivariate analysis. *Journal of the Science of Food and Agriculture*, 94(13):2613–2617.

Bühlmann, P., Kalisch, M., and Maathuis, M. H. (2010). Variable selection in high-dimensional linear models: partially faithful distributions and the pc-simple algorithm. *Biometrika*, 97(2):261–278.

Bülthoff, A. (1996). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*.

Camelo-Méndez, G. A., Camacho-Díaz, B. H., del Villar-Martínez, A. A., Arenas-Ocampo, M. L., Bello-Pérez, L. A., and Jiménez-Aparicio, A. R. (2012). Digital image analysis of diverse mexican rice cultivars. *Journal of the Science of Food and Agriculture*, 92(13):2709–2714.

Carroll, R. J. and Hall, P. (2004). Low order approximations in deconvolution and regression with errors in variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(1):31–46.

Courcoul, A., Vergu, E., Denis, J.-B., and Beaudeau, F. (2010). Spread of q fever within dairy cattle herds: key parameters inferred using a bayesian approach. *Proceedings of the Royal Society B: Biological Sciences*, 277(1695):2857–2865.

Coursol, J. (1981). *Technique statistique des modèles linéaires: Aspects théoriques*. CIMPA.

Díaz-García, J. A., Galea Rojas, M., and Leiva-Sánchez, V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communications in statistics-Theory and Methods*, 32(3):625–641.

Diday, E. (1982). *Eléments d'analyse de données*. Bordas Editions.

Dillon, William R et Goldstein, M. (1984). *Analyse multivariée : méthodes et applications*.

Djaballah-Djeddour, K. and Tazerouti, M. (2022). Test for linearity in non-parametric regression models. *Austrian Journal of Statistics*, 51(1):16–34.

Erkel-Rousse, H. and Le Gallo, F. (2002). Product quality, national trade performances, and the estimation of trade price-elasticities.

Ferreira, J. T. and Steel, M. F. (2007). A new class of skewed multivariate distributions with applications to regression analysis. *Statistica Sinica*, pages 505–529.

Fraser, D. (1991). Statistical inference: Likelihood to significance. *Journal of the American Statistical Association*, 86(414):258–265.

Galea, M., Paula, G. A., and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(1):71–79.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263.

Huber, P. J. (1996). *Robust statistical procedures*. SIAM.

Islam, M. Q. and Tiku, M. L. (2005). Multiple linear regression model under nonnormality. *Communications in Statistics-Theory and Methods*, 33(10):2443–2467.

Jolliffe, I. T. (1990). Principal component analysis: a beginner's guide—i. introduction and application. *Weather*, 45(10):375–382.

Legendre, A. (1805). New methods for the determination of comet orbits. *Paris: F. Didot*.

Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers*, 41:211–224.

Liu, S. (2004). On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *Journal of Applied Probability*, 41(A):393–405.

Nguyen, H. T. (2015). Statistics of fuzzy data: a research direction for applied statistics. *Thailand Statistician*, 13(1):1–31.

Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions technip.

Sellam, V. and Poovammal, E. (2016). Prediction of crop yield using regression analysis. *Indian Journal of Science and Technology*.

Simmonds, N. W. (1995). The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture*, 67(3):309–315.

Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimale, V., and Wimberly, F. (2000). Constructing bayesian network models of gene expression networks from microarray data.

Sutradhar, B. C. and Ali, M. M. (1986). Estimation of the parameters of a regression model with a multivariate t error variable. *Communications in Statistics-Theory and Methods*, 15(2):429–450.

Tiku, M. L., Islam, M. Q., and Sazak, H. S. (2008). Estimation in bivariate nonnormal distributions with stochastic variance functions. *Computational statistics & data analysis*, 52(3):1728–1745.

Tiku, M. L., Islam, M. Q., and Selçuk, A. S. (2001). Nonnormal regression. ii. symmetric distributions. *Communications in Statistics-Theory and Methods*, 30(6):1021–1045.

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, 60(4):812–854.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, 71(354):400–405.

Tables

Table 1: Eigen analysis of the Correlation Matrix

	F1	F2	F3	F4
Eigenvalue	3.095	0.419	0.292	0.194
Proportion	0.77378	0.10478	0.07292	0.04852
Cumulative	0.77378	0.87855	0.95148	100

Table2: Correlation between variables and factors

	F1	F2	F3	F4
Durum wheat	0.465	0.869	-0.131	-0.108
Bread wheat	0.505	-0.417	-0.546	-0.523
Barley	0.504	-0.175	0.815	-0.225
Oats	0.524	-0.201	-0.142	0.809

Table 3: Eigen analysis of the Correlation Matrix

	F1	F2	F3	F4
Eigenvalue	3.550	0.266	0.125	0.059
Proportion	88.759	6.644	3.121	1.475
Cumulative	88.759	95.403	98.525	100

Table 4: Correlation between variables and factors

	F1	F2	F3	F4
Durum wheat	0.961	-0.114	-0.207	-0.143
Bread wheat	0.893	0.447	0.052	-0.015
Oats	0.940	-0.212	0.264	-0.038
Barley	0.972	-0.093	-0.098	0.192

Table 5: Data			
	Durum wheat production	Rainfall	Irrigation
2000	4863340	211.324	59910
2001	12388650	509.6775	71890
2002	9509670	289.322	88430
2003	18022930	666.9	77940
2004	20017000	524.184	95126
2005	15687090	418.028	82303
2006	17728000	539.514	88250
2007	15289985	574.676	79430
2008	8138115	342.53	90846
2009	20010378	574.116	86846
2010	18089739	511.876	114615
2011	19274740	569.352	130110
2012	24071180	590.416	137567

Table 6: Parameter estimation

Coefficients	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	5.619e + 05	5.248e + 06	0.107	0.9167
x	1.627e + 02	5.524e + 01	2.946	0.0133*

Residual standard error: 4273000 on 11 degrees of freedom Multiple R-squared: 0.441, Adjusted R-squared: 0.3902 F-statistic: 8.678 on 1 and 11DF, p-value: 0.01331

Table 7: Parameter estimation

Coefficients	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	-1593170	3211206	-0.496	0.629568
x	35401	6388	5.542	0.000175***

Residual standard error: 2935000 on 11 degrees of freedom

Multiple R-squared: 0.7363, Adjusted R-squared: 0.7123

F-statistic: 30.71 on 1 and 11 DF, p-value: 0.0001748

Table 8: Parameter estimation

Coefficients	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	−7.204e + 06	3.107e + 06	-2.318	0.04290*
x_1	2.910e + 04	5.338e + 03	5.452	0.00028***
x_2	9.371e + 01	3.170e + 01	2.956	0.01438*

Residual standard error: 2249000 on 10 degrees of freedom Multiple R-squared: 0.8593, Adjusted R-squared: 0.8311 F-statistic: 30.53 on 2 and 10DF, p-value: $5.519e-05$

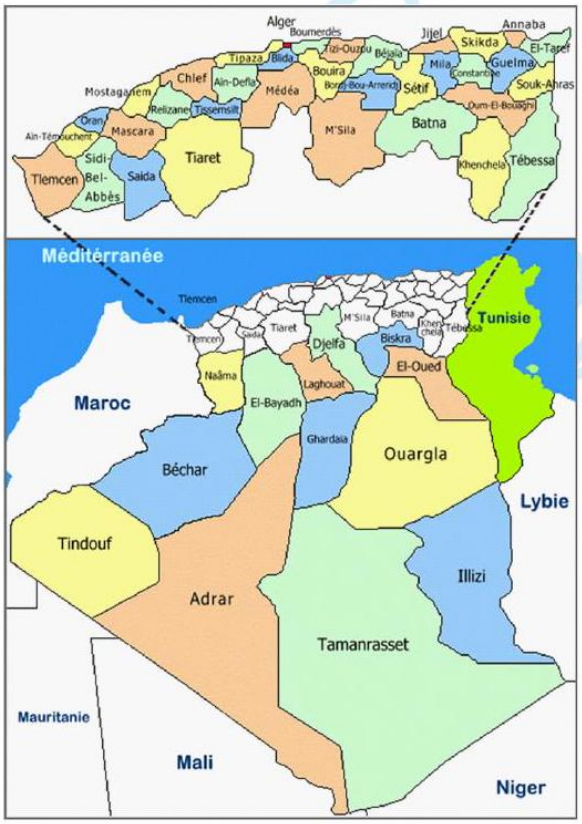
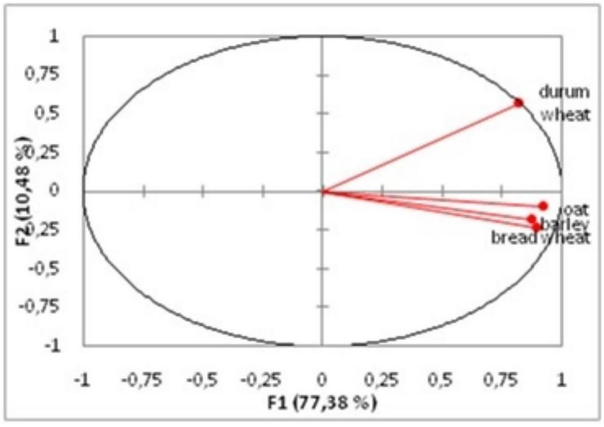
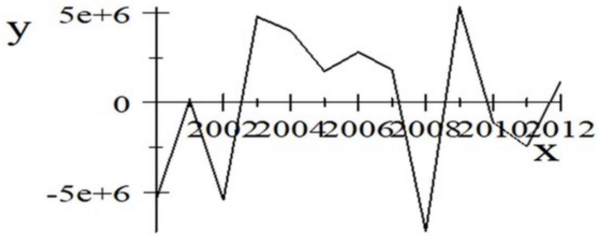
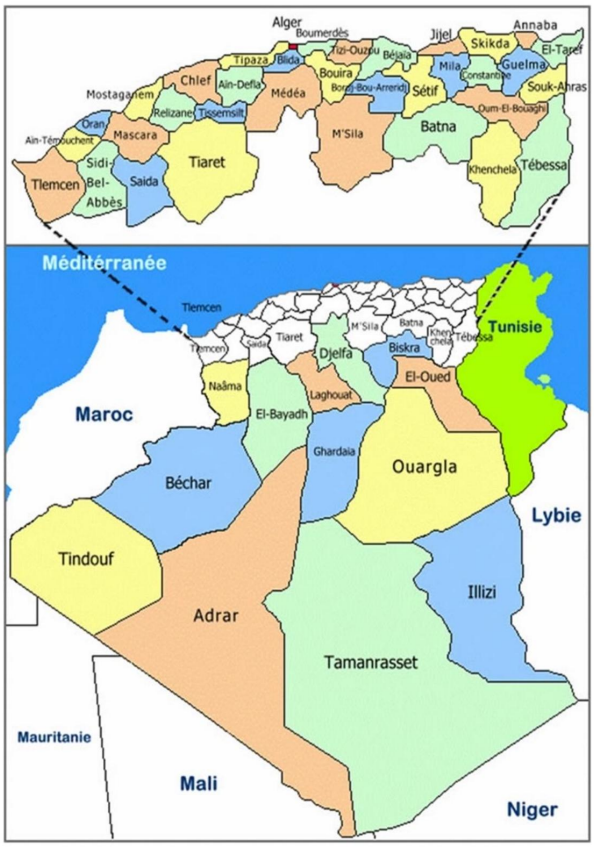


Figure 6: Map

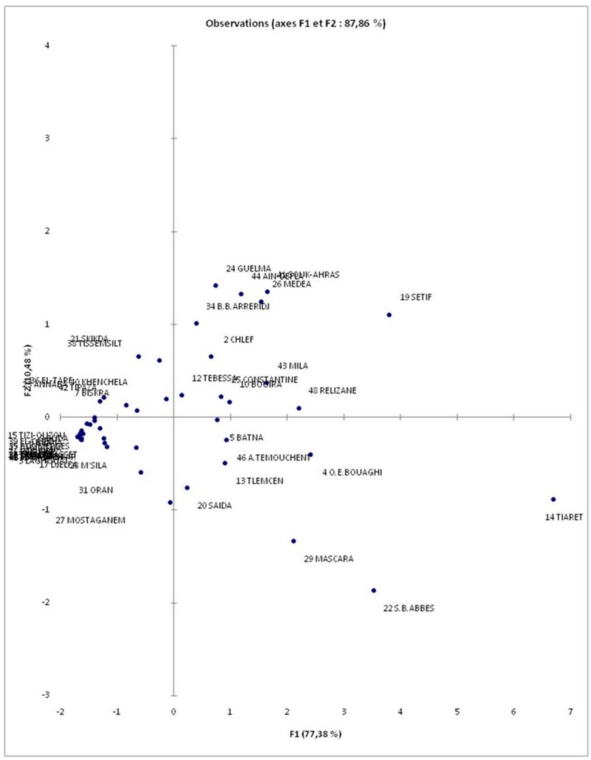






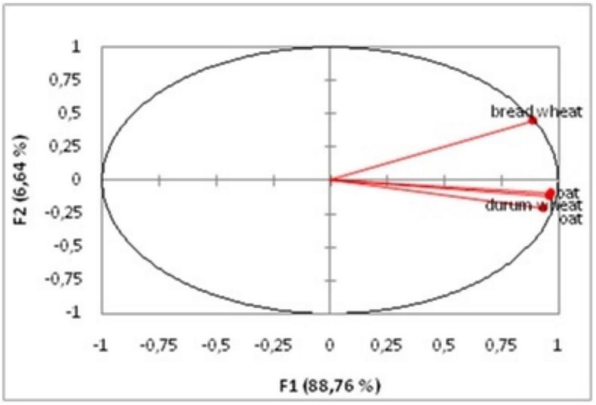
Map

195x278mm (96 x 96 DPI)

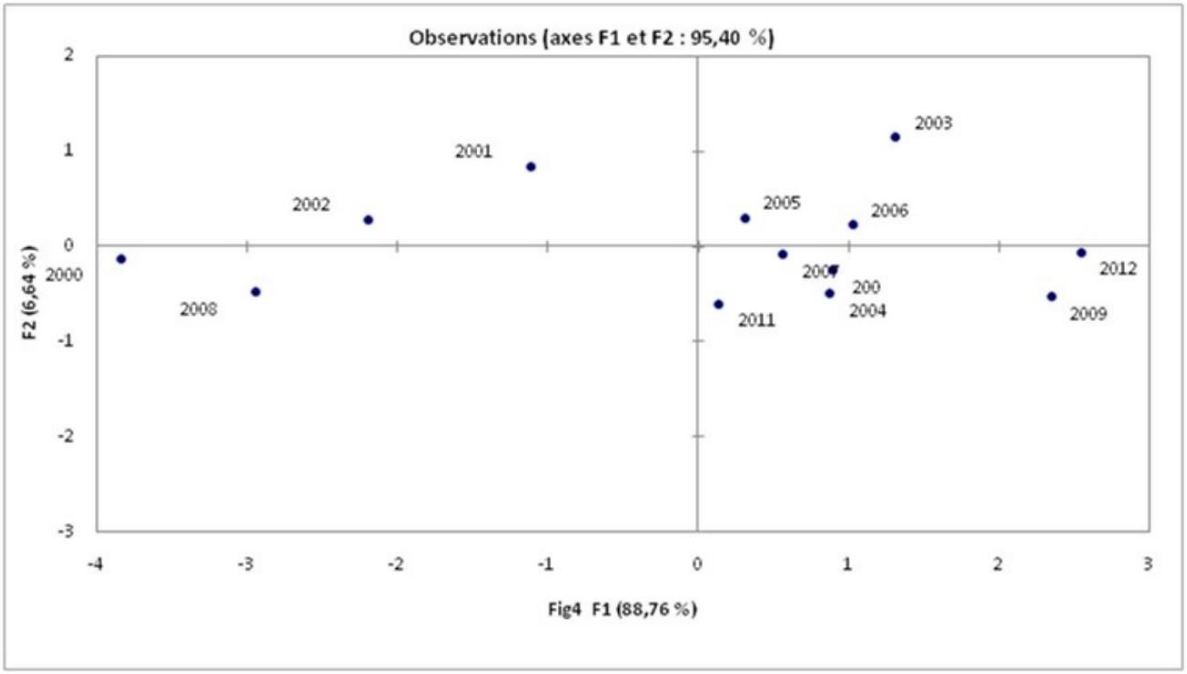


D

Representation of data points on the two first components 201 × 258 mm (96 x 96 DPI)



Correlation circle



Representation of data points on the two components.