



UNIVERSITÉ ABDREHMANE MIRA - BEJAIA
Faculté des Sciences Exactes
Département de Recherche Opérationnelle

Mémoire de fin de cycle

Pour l'obtention du diplôme de Master en Mathématiques Appliquées

Spécialité : Mathématiques Financières

Analyse de la Relation "Offre \sim Prix" en Matières Céréales : Cas de l'Entreprise Portuaire de Béjaïa

Réalisé par :

Mr. SOUKANE Mohamed

Soutenu le 16 juillet 2022, devant le jury composé de :

Président :	Mr. ASLI Larbi	MCA	Université de Béjaïa
Promoteur :	Mme. BOUCHEBBAH Kahina	MAA	Université de Béjaïa
Co-promoteur :	Mme. BOURIHANE Djedjiga	Cadre	EPB
Examineur :	Mr. SOUFIT Massinissa	MCB	Université de Béjaïa
Examineur :	Mme. BOUZERIA Zahra	Doctorante	Université de Béjaïa

Promotion : 2021/2022

Remerciements

Tout d'abord , je remercie Allah, le Très Haut, de m'avoir donné la santé, la volonté, le courage et la patience nécessaires pour la concrétisation de ce travail.

Je tiens à exprimer ma profonde gratitude et mes remerciements à **Mme. Bouchebah Kahina**, pour son encadrement de qualité, sa motivation professionnelle, pour sa patience et son encouragement. Son œil critique m'a été très précieux pour structurer mon travail et pour améliorer la qualité des différentes sections.

Je remercie également ma promotrice de stage **Mme. Bourihane Djedjiga**, pour son soutien, son encadrement et ses conseils. elle a veillé à me fournir un bon environnement de travail, qui m'a permis d'atteindre ces résultats.

Je tiens à exprimer ma reconnaissance envers les membres du jury, et les remercier pour le temps et les efforts qu'ils consacrent à l'évaluation de mon travail.

En guise de reconnaissance, je tiens à remercier, très sincèrement, ma famille pour leur soutien inconditionnel, à la fois moral et financier et pour leur patience.

Enfin, je voudrais remercier toute personne qui a contribué explicitement ou implicitement, de loin ou de près, à la réalisation de ce mémoire.

Dédicaces

C'est avec beaucoup d'amour que je dédie ce modeste travail :

♡ À mes très chers parents,

♡ À mon cher frère,

♡ À ma grande mère,

♡ À Amine, Hammou, Hamida, Dyhia, Karima,

♡ À tout mes amis et amies,

♡ À tout le service statistique de l'EPB,

♡ À tous ceux qui, par un mot, m'ont donné la force de continuer.

SOUKANE Mohamed

TABLE DES MATIÈRES

Introduction générale	1
1 Présentation de l'organisme d'accueil	4
1.1 Historique de l'Entreprise Portuaire de Béjaia	5
1.1.1 Généralités sur le port de Bejaia	5
1.1.2 Historique et cadre juridique	6
1.2 Situation géographique du Port	7
1.3 Statut de l'entreprise	7
1.4 Activités de l'entreprise	8
1.5 Description des services	9
1.6 Surfaces d'entrepôts	10
1.7 Présentation de l'organigramme de l'EPB	11
1.7.1 Direction Générale "DG"	11
1.7.2 Direction Générale Adjointe Fonctionnelle "DGAF"	11
1.7.3 Direction Générale Adjointe Opérationnelle "DGAO"	12
1.7.4 Direction Remorquage "DR"	13
1.7.5 Direction Domaine et Développement "DDD"	13
1.7.6 Direction du Système de Contrôle Interne "DSCI"	14
1.7.7 Direction Logistique "DL"	14
1.7.8 Direction Marketing "DM"	15
1.7.9 Direction Digitalisation et Numérique "DDN"	15
1.7.10 Direction des Ressources Humaines "DRH"	16
1.7.11 Direction Finance et Comptabilités "DFC"	16
1.7.12 Direction Achats "DA"	16
1.7.13 Direction Exploitation "DE"	17

1.7.14	Direction Maintenance “DMA ”	17
1.7.15	Direction de la Capitainerie “DC ”	17
1.8	Types de navires	18
1.8.1	Les navires spécialisés	18
1.8.2	Les navires non spécialisés	18
1.9	Conclusion et Position du problème	21
2	Régression Linéaire Simple	22
2.1	La régression linéaire simple	23
2.2	Modèle	23
2.3	Estimation des paramètres par la méthode des Moindres Carrés Ordinaires (MCO)	24
2.4	Evaluation de la qualité de la régression	27
2.5	Prévisions	29
2.6	Tests et intervalles de confiance	29
2.6.1	Test sur la pente β_1	29
2.6.2	Intervalle de confiance pour β_1	31
2.6.3	Test sur l’ordonnée à l’origine β_0	31
2.6.4	Intervalle de confiance pour β_0	33
2.6.5	Intervalles et régions de confiance pour les coefficients de régression lorsque la variance σ^2 est connue	33
2.6.6	Intervalle de prédiction	34
2.7	Test de signification global de Fisher	35
2.8	Cas particulier	37
3	Modèles Mathématiques, Applications et Interprétations des Résultats	44
3.1	Applications sous le logiciel R	45
3.1.1	Présentation du logiciel R	45
3.1.2	Justification du choix de l’environnement de programmation	45
3.2	Collecte de données Financières	45
3.3	Le Blé Dur	46
3.3.1	Description des données statistiques sur le blé dur	46
3.3.2	Inférence sur les deux paramètres du modèle β_0 et β_1	46
3.3.3	Interprétation des résultats obtenus et discussion	48
3.3.4	Prédiction et comparaison avec des données réelles	49
3.4	Le Blé Tendre	49
3.4.1	Description des données statistiques sur le blé tendre	50
3.4.2	Inférence sur les deux paramètres du modèle β_0 et β_1	50

3.4.3	Interprétation des résultats obtenus et discussion	52
3.4.4	Prédiction et comparaison avec des données réelles	53
3.5	Le Maïs	53
3.5.1	Description des données statistiques sur le maïs	54
3.5.2	Inférence sur les deux paramètres du modèle β_0 et β_1	54
3.5.3	Interprétation des résultats obtenus et discussion	55
3.5.4	Prédiction et comparaison avec des données réelles	56
3.6	L'Orge	56
3.6.1	Description des données statistiques sur l'orge	57
3.6.2	Inférence sur les deux paramètres du modèle β_0 et β_1	57
3.6.3	Interprétation des résultats obtenus et discussion	59
3.6.4	Prédiction et comparaison avec des données réelles	59
	Conclusion générale	61
	Bibliographie	63
	Annexes	65
A	Notions de base	67
A.1	Espace Probabilisé	67
A.2	Variable Aléatoire	67
A.3	Couple de Variable Aléatoire	68
A.4	Espérance Mathématique	69
A.5	Variance et Ecart-type	70
A.6	Covariance	71
A.7	Echantillon	72
A.8	Moyenne Empirique	72
A.9	Variance Empirique	73
A.10	Covariance	74
A.11	Loi de Probabilité	74
A.11.1	Loi Normale	75
A.11.2	Loi du χ^2 , Student et de Fisher	75

LISTE DES FIGURES

1.1	Plan de développement du port de Béjaia	8
1.2	Organigramme de l'EPB	12
1.3	Les types de navires	19
1.4	Vraquier	19
2.1	Droite de régression, résidus, linéarité.	27
2.2	Illustration de la formule $DT = DA + DR$	28
2.3	Table de ANOVA	36
3.1	Graphe de Y_{BD} en fonction de X_{BD} et la droite de régression	47
3.2	Graphe de Y_{BT} en fonction de X_{BT} et la droite de régression	51
3.3	Graphe de Y_M en fonction de X_M et la droite de régression	54
3.4	Graphe de Y_{OR} en fonction de X_{OR} et la droite de régression	57

LISTE DES TABLES

1.1	Bassins du port	8
3.1	La variable Prix (ou Cours) mesurée en €	46
3.2	La variable Quantité importée (ou Offre) mesurée en Tonne	46
3.3	Inférence sur les deux paramètres du modèle β_0 et β_1 cas du blé dur (MCO)	46
3.4	La variable Prix (ou Cours) mesurée en €	50
3.5	La variable Quantité importée (ou Offre) mesurée en Tonne	50
3.6	Inférence sur les deux paramètres du modèle β_0 et β_1 cas du blé tendre .	50
3.7	Description statistique sur les résidus ε_i	52
3.8	La variable Prix (ou Cours) mesurée en €	54
3.9	La variable Quantité importée (ou Offre) mesurée en Tonne	54
3.10	Description statistique sur les résidus ε_i	55
3.11	La variable Prix (ou Cours) mesurée en €	57
3.12	La variable Quantité importée (ou Offre) mesurée en Tonne	57
3.13	Description statistique sur les résidus ε_i	58

INTRODUCTION GÉNÉRALE

L'Algérie subit depuis près d'un siècle une forte dynamique démographique qui a rompu divers équilibres économiques, sociaux et naturels et a induit une haute pression sur les ressources agricoles disponibles. Résultat, le pays connaît actuellement un déséquilibre important de la balance commerciale agricole, conduisant à une dépendance alimentaire extrêmement forte vis-à-vis de l'étranger, notamment pour les céréales. [4]

Les céréales constituent un produit aussi vital que stratégique en Algérie où le mode alimentaire est basé essentiellement sur leur consommation. La superficie consacrée traditionnellement aux céréales varie de 3 à 3,5 millions d'hectares. Le blé dur occupe une place privilégiée suite à son utilisation dans l'alimentation quotidienne de la population sous diverses formes. [4]

Les rendements restent faibles et très variables d'une année à l'autre, à l'image de la production qui varie de 4.9 à 20 millions de quintaux/an pour la même période. La culture des céréales d'hiver demeure encore difficile à maîtriser tant que celle-ci reste confrontée et soumise à plusieurs contraintes (les spécificités climatiques, les déficits pluviométriques, les disponibilités en terres agricoles, les caractéristiques et les particularités environnementales de l'Algérie, etc.). [3]

L'offre et la demande des produits agricoles sur le marché, notamment les matières céréales, sont caractérisées par leurs fluctuations qui dépendent du temps, certains produits par exemple ne peuvent être disponibles que dans certaines saisons. Par ailleurs, l'offre et la demande des produits céréales sur le marché présentent une fluidité

telle que la demande s'adapte rapidement aux fluctuations de l'offre et de même l'offre s'adapte facilement à la demande. Les prix des céréales sont déterminés principalement par le volume de l'offre (qualité et quantité) et le niveau du coût de revient apporte très peu d'influence sur eux.

La présence d'un nombre élevé d'opérateurs financiers sur les marchés des matières premières agricoles, et leur impact sur l'instabilité des prix, ont contribué à alimenter le débat sur la formation des cours. Une première école de pensée, incluant entre autres la Commodities Futures Trading Commission, soutient que l'accession des agents financiers à ces marchés n'a pas accentué la volatilité des prix agricoles. Il s'agit d'une vision traditionnelle selon laquelle les activités spéculatives vont dans le sens des volumes d'opérations de couverture, et que le comportement naturellement instable de ces marchés est entretenu par un nombre d'opérateurs financiers de plus en plus élevé. [1]

Pour répondre à cette préoccupation, des données statistiques mensuelles sur les quantités des céréales importées¹ et les cours correspondants² ont été utilisés afin de bâtir des modèles linéaires reliant ces deux critères. En s'appuyant sur le caractère des données disponibles, leurs critères statistiques et la nature du problème posé, nous justifions le choix de l'approche d'analyse de données statistiques³.

La régression est l'une des méthodes les plus connues et les plus appliquées en statistique pour l'analyse de données quantitatives. Elle est utilisée pour établir une liaison entre une variable quantitative et une ou plusieurs autres variables quantitatives, sous forme d'un modèle. Si nous nous intéressons à la relation entre deux variables, nous parlerons de régression simple en exprimant une variable en fonction de l'autre. Si la relation porte entre une variable et plusieurs autres variables, nous parlerons de régression multiple. [6]

L'origine de la régression remonte au dix-neuvième siècle et sa première application a été réalisée par le scientifique anglais Sir Francis Galton [10], le cousin de Darwin. Dans le cadre d'une étude biologique sur l'hérédité, Galton souhaitait démontrer l'existence d'une relation entre le diamètre de graines de pois de senteur et le diamètre moyen des

¹Les quantités des céréales importées sont collectées au sein du Département Statistiques de la Direction Exploitation de l'Entreprise Portuaire de Béjaia.

²Les cours mensuels des céréales sur le marché sont disponibles sur le site Web <https://www.agri-mutuel.com/cotations>.

³Dans notre travail, nous utilisons la méthode de régression linéaire simple.

graines de la descendance. [6]

En ce qui concerne le contenu de ce manuel, nous y trouvons une présentation de l'organisme d'accueil; l'histoire de l'Entreprise Portuaire de Béjaia, sa situation géographique et sa structure directionnelle font l'objet du premier chapitre. Des concepts et techniques des bases de l'analyse de régression simple et la méthode d'estimation des paramètres, consistant à estimer la droite de régression par une droite qui minimise une fonction des résidus, la plus connue est la méthode des moindres carrés, elle y est présentée dans le deuxième chapitre. Enfin, le troisième chapitre est voué à la présentation et développement des modèles mathématiques de prévision et cela se réalisera à l'aide des données statistiques collectées de chacun des quatre types de céréales, à savoir : le blé dur, le blé tendre, le maïs et l'orge. A travers chaque type de modèles développés, des prévisions futures sur les quantités importées en fonction de la flambée des cours actuels, les résultats obtenus et leurs interprétations seront présentés à la fin de ce dernier chapitre.

CHAPITRE 1

PRÉSENTATION DE L'ORGANISME D'ACCUEIL

La réalisation du port de Bejaia dans la composante actuelle débuta en 1834, elle fut achevée en 1987. L'Entreprise Portuaire de Bejaia occupant aujourd'hui la première position pour ce qui est du trafic céréalier, l'activité est dotée d'équipements les plus performants sur le marché ; qui ont fait que les rendements de la manipulation des céréales sont passés de 50T/heure à 750T/heure. L'EPB peut ainsi traiter de gros céréaliers de 60.000 tonnes.

Le port de Béjaia est réputé mixte ; hydrocarbures et marchandises générales y sont traités. L'aménagement moderne des superstructures, le développement des infrastructures, l'utilisation de moyens de manutention et de techniques adaptés à l'évolution de la technologie des navires et enfin ses outils de gestion moderne, a fait évoluer le port de Béjaia depuis le milieu des années 1990.

Le port de Béjaia est classé premier port d'Algérie concernant le transit et la manutention des marchandises générales. Il est également le premier port du bassin sud méditerranéen certifié ISO 9001:2000 pour l'ensemble de ses prestations, et à avoir ainsi installé un système de management de qualité. Cela constitue une étape dans le processus d'amélioration continue de ses prestations au grand bénéfice de ses clients. L'Entreprise Portuaire a connu d'autres succès depuis, elle est notamment certifiée à la Norme ISO 14001:2004 et au référentiel OHSAS 18001:2007, respectivement, pour l'environnement,

l'hygiène et la sécurité au travail.

1.1 Historique de l'Entreprise Portuaire de Béjaia

1.1.1 Généralités sur le port de Bejaia

Au cœur de l'espace méditerranéen, la ville de Bejaia possède de nombreux sites naturels et vestiges historiques, datant de plus de 10 000 ans, ainsi qu'une multitude de sites archéologiques, recelant des trésors anciens remontant à l'époque du néolithique.

Dans l'antiquité, Amsyouen, habitants des flans surplombant la côte, ne fréquentaient la côte que pour pêcher. Les premières nefes qui visitèrent nos abris naturels furent les phéniciennes, ils y installèrent des comptoirs.

La Saldae romaine leur succéda, et devint port d'embarquement de blé. Ce n'est qu'au onzième siècle que la berbère Begaieth, devenue Ennaciria, prit une place très importante dans le monde de l'époque. Le port de Bejaia devint l'un des plus importants de la Méditerranée, ses échanges étaient très denses. L'histoire retiendra également à cette époque, que par Fibonacci de Pise, fils d'un négociant pisan, les chiffres aujourd'hui universellement utilisés s'étendirent dans le monde à partir de Bejaia.

La réalisation des ouvrages du port débuta en 1834. Les infrastructures actuelles ont commencé à être érigées à la fin du siècle dernier. Les ouvrages de protection furent entamés en 1870 et ceux d'accostage en 1879. La construction de l'ancien port fut achevée en 1911. Le nouveau quai long de 300 m fût achevée en 1987. C'est en 1960 que fût chargé le premier pétrolier au port de Bejaia.

L'aménagement moderne des espaces et des installations spécialisées, l'implantation de nouvelles industries et l'introduction d'outils modernes de gestion ont fait évoluer le Port de Bejaia vers la troisième génération et le classent aujourd'hui deuxième port d'Algérie en marchandises générales.

Ayant acquis la certification ISO 9002 en 2000 puis la transition à la norme 9001 version 2000 en 2002 et la 14001 en 2004, l'entreprise aspire pour fin 2007 à inclure l'OHSAS 18000 pour faire aboutir son projet d'être le premier port africain à se doter d'un système de management intégré.

1.1.2 Historique et cadre juridique

Le décret n°82-285 du 14 Août 1982 publié dans le journal officiel n° 33 porta la création de l'Entreprise Portuaire de Bejaia ; entreprise socialiste à caractère économique ; conformément aux principes de la charte de l'organisation des entreprises, aux dispositions de l'ordonnance no 71-74 du 16 Novembre 1971 relative à la gestion socialiste des entreprises.

L'entreprise, réputée commerçante dans ses relations avec les tiers, fut régie par la législation en vigueur et soumise aux règles édictées par le sus mentionné décret.

Pour accomplir ses missions, l'entreprise est substituée à l'Office National des Ports (ONP), à la Société Nationale de Manutention (SO.NA.MA) et pour partie à la Compagnie Nationale Algérienne de Navigation (CNAN).

Elle fut dotée par l'Etat, du patrimoine, des activités, des structures et des moyens détenus par l'ONP, la SO.NA.MA et de l'activité Remorquage, précédemment dévolue à la CNAN, ainsi que des personnels liés à la gestion et au fonctionnement de celles-ci.

En exécution des lois n° 88.01, 88.03 et 88.04 du 02 Janvier 1988 s'inscrivant dans le cadre des réformes économiques et portant sur l'autonomie des entreprises, et suivant les prescriptions des décrets n°88.101 du 16 Mai 1988, n°88.199 du 21 Juin 1988 et n°88.177 du 28 Septembre 1988.

L'Entreprise Portuaire de Bejaia, entreprise socialiste, est transformée en Entreprise Publique Economique, Société par Actions (EPE-SPA) depuis le 15 Février 1989, son capital social fut fixé à dix millions (10.000.000) de dinars algériens par décision du conseil de la planification n°191/SP/DP du 09 Novembre 1988.

L'assemblée générale extraordinaire du 22 Juin 2002 décida de l'augmentation du capital social qui fut fixé à un montant de Soixante millions (60.000.000) de dinars algériens.

En date du 30 Juin 2003, la même assemblée augmenta le capital social de l'Entreprise Portuaire de Bejaia à hauteur de six cent millions (600.000.000) de dinars algériens. Il est divisé en six mille (6.000) Actions de cent mille (100.000) dinars de valeur nominale entièrement libérées et souscrites pour le compte de l'Etat, détenues à 100% par la Société de Gestion des Participations de l'Etat «Ports », par abréviation « SOGEPORTS ».

En 2006 l'assemblée générale extraordinaire a augmentée le capital social de l'Entreprise Portuaire par l'incorporation des réserves facultatives, et qui fut fixé à un montant de Un milliard deux cent millions de dinars (1.200.000.000 DA).

En 2007 l'assemblée générale extraordinaire a augmenté le capital social de l'Entreprise Portuaire par l'incorporation des réserves facultatives, et qui fut fixé à un montant d'un milliard sept cent millions de dinars (1.700.000.000 DA).

En 2014 l'assemblée générale extraordinaire a augmenté le capital social de l'Entreprise Portuaire par l'incorporation des réserves facultatives, et qui fut fixé à un montant de trois milliard cinq cent millions de dinars (3.500.000.000 DA).

L'aménagement moderne des superstructures, le développement des infrastructures, l'utilisation de moyens de manutention et de techniques adaptés à l'évolution de la technologie des navires et enfin ses outils de gestion moderne, ont fait évoluer le Port de Béjaia depuis le milieu des années 1990 pour être classé aujourd'hui second port d'Algérie.

1.2 Situation géographique du Port

Position : Latitude N 36° 45' 24", longitude E 005° 05' 50"; le port de Béjaia est délimité par :

- Au nord par la route nationale N°9.
- Au sud par les jetées de fermeture et du large sur une largeur de 2750 m.
- A l'est par la jetée Est.
- A l'ouest par la zone industrielle de Béjaia.

Connue pour être l'une des meilleures de la côte algérienne, la rade de Béjaia offre d'excellentes potentialités en matière de protection et des fonds propices à un bon mouillage, avec des profondeurs allant de 10 m à plus de 20 m . Abrisée de tous les vents sauf du nord-est à l'est, la rade est limitée par une ligne imaginaire s'étendant du Cap Carbon au Cap Aokas.

1.3 Statut de l'entreprise

L'entreprise portuaire de Béjaia est régie par les disposition du décrets n°82.285 d'août 1982 publié au journal officiel n°33 du 17 août 1982. Depuis février 1989 elle a été

érigée en entreprise publique économique - société par action - dans le cadre de la loi sur l'autonomie des entreprises. Son capital social a été fixé à 60.000.000.00 DA. Elle a pour missions essentielles la gestion, l'exploitation et le développement du domaine et ce en vue de promouvoir les échanges extérieurs du pays dans les meilleures conditions économiques.

Les bassins	Superficie	Profondeur
Bassins de l'avant-port	75 hectares	[10-13,5] m
Bassins du vieux port	26 hectares	[6-8] m
Bassins de l'arrière-port	55 hectares	[10-12,5] m

Table 1.1: Bassins du port

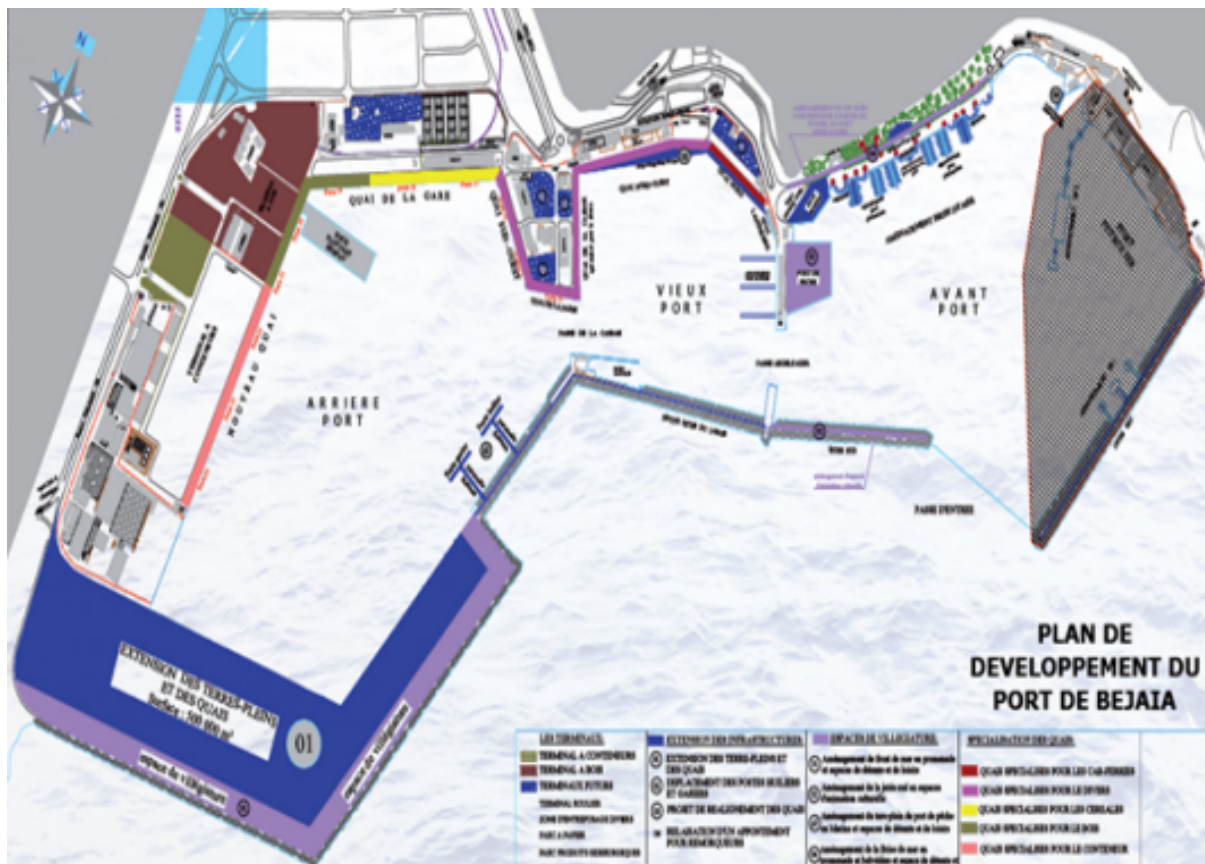


Figure 1.1: Plan de développement du port de Béjaia

1.4 Activités de l'entreprise

L'entreprise portuaire couvre un vaste domaine d'activités qui concerne le traitement dans les meilleures conditions de délai, de coût et de sécurité, l'ensemble des passages des navires et des marchandises.

Elle exerce à cet effet, en plus de sa vocation naturelle d'autorité portuaire, les activités suivantes :

1. La gestion, l'exploitation et le développement des infrastructures et superstructures.
2. La manutention et l'acconage des marchandises en transit par le port de Béjaia.
3. Le pilotage, le remorquage et le lamanage des navires dans les limites du port de Béjaia.
4. La sécurité portuaire dans les limites du domaine portuaire.
5. Transit des passagers et leurs véhicules par la gare maritime.
6. La mise à disposition des infrastructures nécessaires aux activités relatives aux hydrocarbures.

1.5 Description des services

1. **L'acheminement des navires de la rade vers le quai** : Dans certains cas exceptionnels, d'arrivée massive en rade, les navires restent en attente dans la zone de mouillage (rade) jusqu'à obtention de l'autorisation de rejoindre un poste à quai. Cette dernière est délivrée après une conférence de placement qui se tient quotidiennement au niveau de la Direction Capitainerie.
L'acheminement des navires se fait par des opérations d'aide à la navigation identifiée par le remorquage, le pilotage et le lamanage.
2. **Le remorquage** : Il consiste à tirer ou à pousser le navire, pour effectuer les manœuvres d'accostage, de décalage ou d'appareillage du navire. Il consiste également à effectuer les opérations de convoyage et d'aide dans l'exécution d'autres manœuvres.
3. **Le pilotage** : Il est assuré de jour comme de nuit par la Direction Capitainerie et est obligatoire à l'entrée et à la sortie du navire. Il consiste à assister le commandant dans la conduite de son navire à l'intérieur du port.
4. **Le lamanage** : Il consiste à amarrer ou désamarrer le navire de son poste d'accostage.
5. **Les opérations de manutention et d'acconage pour les marchandises** :
Ces opérations sont :

- Les opérations d'embarquement et de débarquement des marchandises.
- La réception des marchandises.
- Le transfert vers les aires d'entreposage, hangars et terre-pleins, ports secs.
- La préservation ou la garde des marchandises sur terre-pleins ou hangar et hors port.
- Pointage des marchandises.
- La livraison aux clients.
- La manutention et l'acconage sont assurés, par un personnel formé dans le domaine. Il est exercé de jour comme de nuit, réparti sur deux vacations de 6h à 19h avec un troisième shift optionnel qui s'étale entre 19h et 01h du matin. Pour des cas exceptionnels, ce dernier peut s'étaler jusqu'à 7 h du matin.
- D'autres prestations sont également fournies aux navires et aux clients telles que :
 - Enlèvement des déchets des navires et assainissement des postes à quai.
 - Pesage des marchandises (ponts bascules).
 - Location de remorqueurs ou vedettes (pour ravitaillement des navires, assistance médicale, assistance et sauvetage en haute mer).

1.6 Surfaces d'entreposages

Le port de Béjaïa dispose de 400.000 m² de surface non couverte et de 17500m² de surface couverte, ces surfaces sont réparties comme suit :

- Un magasin cale .Neuf hangars banals pour marchandises diverses.
- Six hangars et terres plaines formant le Centre de Transit des Marchandises Dangereuses, CTMD.
- Un dock silo de stockage d'huile de table.
- Une gare maritime.
- Un terminal à conteneurs.
- Deux enclos pour véhicules de tourisme.
- Une surface de 400.000 m² de terre pleines nue.
- Une station de déballastage pour les pétroliers.

- Un centre de conditionnement de ciments d'une capacité de 30.000 tonnes.
- Un centre anti-pollution marine.

L'entreposage de la marchandise d'ébarquée est aussi l'une des principales activités du port. En effet, la marchandise déchargée sera affectée à des aires de stockage selon des critères :

- Les marchandises non périssables sont admises à l'entreposage en terre-pleins.
- Les produits dangereux sont acheminés au CTMD conformément aux règles et modalités d'admission des produits dangereux.
- Les marchandises conteneurisées sont acheminées et entreposées dans le parc à conteneurs.

1.7 Présentation de l'organigramme de l'EPB

L'entreprise portuaire est structurée au nombre de neuf directions, elles-mêmes subdivisées en départements et services.

Les activités de l'entreprise sont organisées suivant un schéma de décentration des pouvoirs de décision au profit des directeurs centraux et de leurs chef de départements.

Un projet de réorganisation de l'entreprise a été présenté au conseil d'Administration en date du 17 février 2020, qui s'est soldé par une validation (résolution n°04). Des ajustements ont été apportés à l'organisation actuelle par la fusion, la suppression ainsi que la création de certaines directions. vu la mise en place de cette nouvelle organisation, des décisions de mobilité ont été établies au profit du personnel concerné ; d'où la nécessité d'actualiser les fiches emplois avant de procéder à leur diffusion.

1.7.1 Direction Générale "DG"

Le transfert du Département Affaires Juridiques à la Direction Générale. Le rehaussement du Département Audit interne et Contrôle de Gestion en « Direction du Système de Contrôle Interne » ;

1.7.2 Direction Générale Adjointe Fonctionnelle "DGAF"

La dissolution de la Direction du Management Intégré et le transfert du Département Mangement Intégré à la « Direction des Ressources Humaines » ;

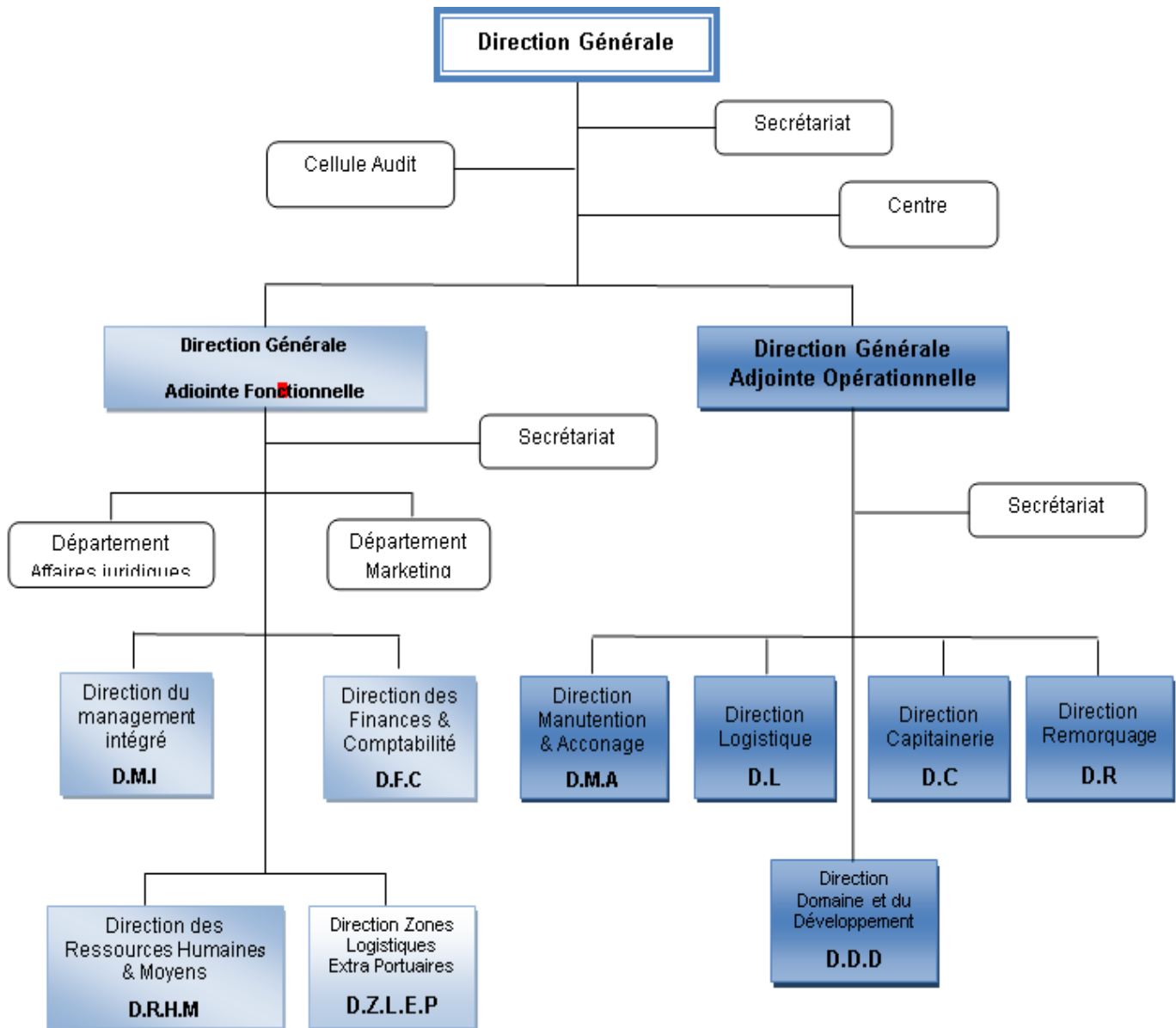


Figure 1.2: Organigramme de l'EPB

Le changement de dénomination de la Direction des Systèmes d'information en « Direction Digitalisation et Numérique » ;

1.7.3 Direction Générale Adjointe Opérationnelle “DGAO”

1. La création du « Service Gestion du Patrimoine », rattaché à la Direction Générale Adjointe Opérationnelle,
2. Le rattachement du Bureau de Sûreté Portuaire B.S.P au DGAO,
3. La création de la « Direction Exploitation » par la fusion des directions de la

Manutention et Acconage, les Zones Logistiques Extra portuaires, le Département Exploitation des Engins de la Direction Logistique,

4. La fusion des Directions : Remorquage et Capitainerie en «Direction Capitainerie»,
5. La fusion de la Direction du Domaine et Développement et la Direction de la Gare Maritime en « Direction Domaine et Développement »,
6. La centralisation des achats/approvisionnements et maintenances des différentes structures en une seule «Direction Achats et Maintenances» : un groupe de travail a été mis en place ayant pour missions la proposition d'une organisation opérationnelle.

1.7.4 Direction Remorquage “DR”

Cette direction assure les opérations de remorquage des navires à l'entrée comme à la sortie, elle couvre diverses missions qui sont :

- Le remorquage portuaire et hauturier.
- L'assistance sauvetage des navires et engins en péril.
- L'avitaillement et le transport de matériels.

Autre intervention qui fait partie du cadre des secours tels que, la protection de l'environnement et la lutte contre les incendies.

1.7.5 Direction Domaine et Développement “DDD”

Elle a pour mission essentielle la mise en œuvre des programmes d'entretien et d'investissement de l'entreprise “travaux et acquisition du matériel”. Elle est subdivisée en trois départements :

Département travaux

Il est chargé de l'entretien des infrastructures, de balayage des quais, enlèvement de déchet des navires et l'alimentation de ce dernier en eau potable. Il est composé de quatre services :

- Service domaine.
- Service infrastructure.

- Service superstructure.
- Service marché.

Département Maintenance

Il s'occupe de la maintenance curative et préventive des engins, la sécurité, la fiabilité pendant leur utilisation, il est chargé aussi du développement et l'extension des moyens et équipements.

Département Approvisionnements

Il s'occupe de la gestion, du renouvellement des pièces de recharge, ainsi que l'achat des matériaux et outils pour les travaux de manutentions.

1.7.6 Direction du Système de Contrôle Interne "DSCI"

La Direction du Système de Contrôle Interne comprend trois (03) services :

1. Service Audit Interne.
2. Service Management des Risques.
3. Service Contrôle de Gestion.

1.7.7 Direction Logistique "DL"

Consciente des enjeux économiques liés à la région et a son hinterland, l'entreprise portuaire de Béjaia a donné la priorité au développement et à la modernisation de la manutention mécanisée, permettant ainsi d'assurer des prestations à même de répondre à ses objectifs et aux soucis de la clientèle en offrant une meilleure qualité de services, dans les meilleurs délais et à moindre coût. C'est dans cet esprit que l'entreprise à modéliser une organisation. Elle s'est basée sur une approche stratégique axée principalement sur :

- Le développement d'une stratégie globale de survie axée sur : la diversification, la spécialisation et la différenciation .
- L'option de filialisation et de développement de nouvelles prestations et de nouveaux métiers.
- La spécialisation des métiers et la gestion par terminaux.

- La modernisation du matériel et l'introduction de nouvelles technologies (manutention, pointage portuaire, etc) .
- La modernisation des procédés de gestion portuaire .
- Un système de management incluant les trois volets : qualité, environnement et santé et sécurité au travail.

La direction logistique englobe dorénavant le service engins et grues, issu de la direction manutention et acconage, le département maintenance et le département approvisionnements issus tous deux de la direction domaine et développement. Les principaux métiers exercés par la direction logistique sont :

- L'approvisionnement en pièces de rechange.
- La maintenance des équipements.
- La planification des affectations.

1.7.8 Direction Marketing “DM”

Elle est composée de deux services :

- **Service étude** : Il est chargé de tous les travaux d'étude et de planification concernant l'entreprise. Cette structure élabore les opérations de marketing visant à augmenter les parts de marché de l'EPB.
- **Service informatique** : Il est chargé de la gestion de tous les systèmes informatiques de l'entreprise à savoir les réseaux, et les différentes applications utilisées dans la gestion du port.

1.7.9 Direction Digitalisation et Numérique “DDN”

La Direction des Systèmes d'information est, désormais, dénommée Direction Digitalisation et Numérique, rattachée à la Direction Générale Adjointe Fonctionnelle. La Direction comportera trois (03) départements :

1. Département Numérique.
2. Département Génie Logiciel.
3. Département Infrastructures et Systèmes.

1.7.10 Direction des Ressources Humaines “DRH”

Cette direction est chargée d'exécuter toutes les tâches liées à la gestion et au développement des structures, et d'appliquer rigoureusement les lois de gestions et les règlements intérieurs de l'entreprise.

Elle a comme principal département celui du personnel et de la formation qui s'occupe de la gestion administrative du personnel, ainsi que l'établissement de la paie, le suivi des affaires sociales, le suivie de carrière de l'effectif et l'élaboration des actions de formations afin d'atteindre une grande maîtrise des métiers pour adapter le personnel à l'exigence de la modernisation.

1.7.11 Direction Finance et Comptabilités “DFC”

Cette direction gère les finances et les dépenses de l'entreprise, elle est constituée de deux départements à savoir :

Département Comptabilité

Il s'occupe de la comptabilité générale et analytique de l'entreprise, de l'établissement des synthèses des journaux auxiliaires, effectue la balance générale des comptes et le bilan fiscal à chaque fin d'exercice.

Département des Finance

Il a comme principale préoccupation d'assurer le recouvrement des créances, de veiller au bon équilibre financier de l'entreprise, ainsi la prise en charge et la versificatrice des régularités des dépenses, enfin il assure les niveaux d'endettement et des créances.

1.7.12 Direction Achats “DA”

La Direction Achats comprend trois départements :

1. Département Passations de Commandes.
2. Département Passation de Marchés.
3. Département Approvisionnements.
4. et la cellule Planification et Performance.

1.7.13 Direction Exploitation “DE”

L'acconage est une opération qui consiste à assurer la réception, la reconnaissance à terre des marchandises ainsi que de leurs gardes jusqu'à leur embarquement ou livraison au destinataire.

La “DE” est subdivisée en cinq petits départements :

1. Département Manutention et Acconage.
2. Département Affectation de Moyens.
3. Département Statistiques et Commercial.
4. Département Logistique d'IGHIL OUBEROUAK.
5. Département Logistique de TIXTER.

et deux grands départements distincts :

Département Commercial

Il est chargé de l'établissement des factures, des prestations fournies aux clients, ainsi que la détermination du chiffre d'affaires.

Département Manutention

La manutention portuaire comprend les opérations d'embarquement et de débarquement des marchandises, de la mise et de reprise des marchandises sous Hangars, sur terre pleine et magasins et cela dépend bien sûr de la volonté du client.

1.7.14 Direction Maintenance “DMa”

Cette Direction comporte deux départements, à savoir :

- Département Maintenance Engins.
- Département Maintenance Navale.

1.7.15 Direction de la Capitainerie “DC”

Les missions de la Capitainerie instrument de l'autorité portuaire comme est stipulé dans le code maritime algérien dans ses articles 890 et 891 que sont exercées au niveau des ports des missions de puissance publique et de service public et des activités commerciales.

Les missions de puissance publique sont exercées par les organes d'état habilités à cet effet, tandis que le développement, l'entretien, la gestion, la préservation et la conservation du domaine public portuaire constituent des missions de service public dévolu à des institutions créées à cet effet dénommées (autorités portuaires).

La Capitainerie, via ses agents, veille à la préservation et la conservation du domaine public portuaire et d'une manière générale de tout ce qui se rapporte à la prévention, la sécurité et la lutte contre les sinistres de toute nature ainsi que de la régulation du mouvement de la navigation dans les limites maritimes du domaine public portuaire.

On y trouve actuellement les structures suivantes :

- Département police et sécurité.
- Service sécurité terrestre.
- Service facturation.
- Service CTMD et hydrocarbures.
- Station pilotage.
- Département police de la navigation.
- Bureau de sûreté portuaire.

1.8 Types de navires

Le chargement de la marchandise peut s'effectuer sur deux types de navires comme l'indique la FIGURE 1.3 :

1.8.1 Les navires spécialisés

Sont réservés au transport d'un type particulier de marchandise tels que : les vraquiers, pétroliers ou céréaliers, les navires polythermes destinés au transport de denrées périssables.

1.8.2 Les navires non spécialisés

Ils sont destinés pour transporter toute sorte de marchandises diverses ; emballées et conditionnées (sacs, cartons, caisses ; mais aussi cadres conteneurs, etc...). Tels que les cargos conventionnels qui disposent de leurs moyens de manutention, les portes conteneurs adaptés au volume des conteneurs, et les navires rouliers équipés d'une rampe

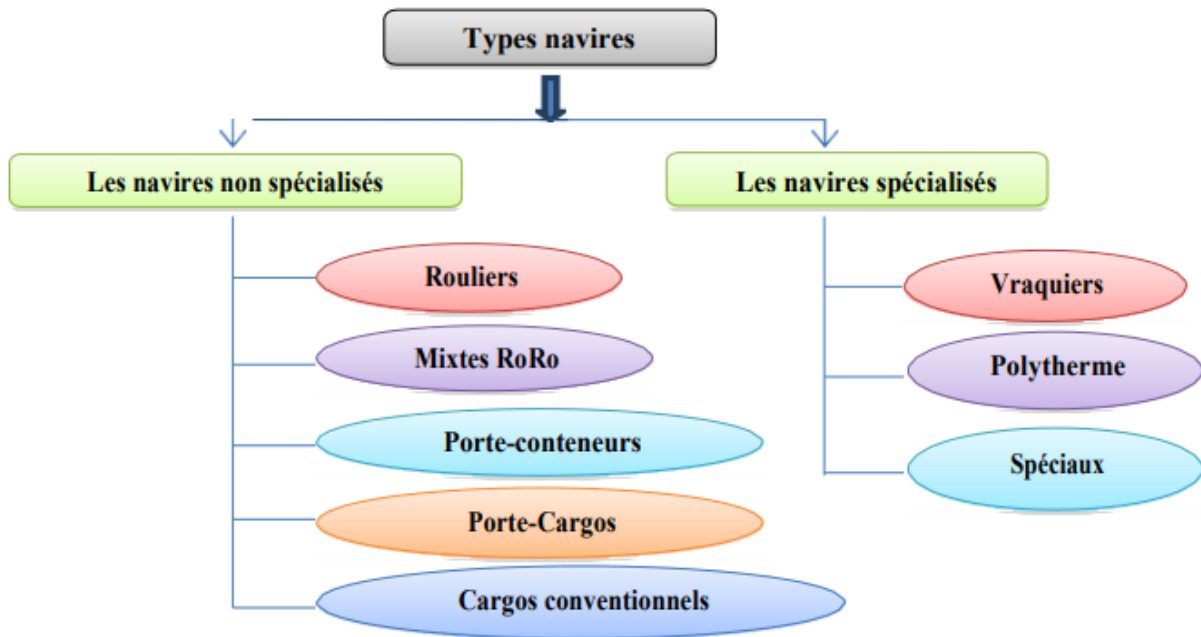


Figure 1.3: Les types de navires

arrière permettent la manutention de tous les types de marchandises. Cette technique de roulage est aussi appelée la technique Ro-Ro (Roll on – Roll off). Les navires mixtes (RoRo + conteneur) combinent l'intérêt du conteneur avec celle du roulage. Les porte-barges destinés au transport combiné fluvial/maritime.

Voici les différents navires et leurs principal fonction :

Les vraquiers :

Les vraquiers comprennent les transporteurs de vrac solides (minéraliers, céréaliers, etc) et les navires citernes, essentiellement les pétroliers, les transporteurs de gazliquéfiés.



Figure 1.4: Vraquier

Les navires polythermes :

Les navires polythermes s'adaptent au transport sous température dirigée. Ils sont destinés au transport denrées périssables. Ce type de navire est aujourd'hui remplacé par les conteneurs reefers. L'usage des conteneurs isothermes et frigorifiques étant jugé préférable, car ils sont plus souple.

Les navires spéciaux :

Il en existe de toutes sortes : car-ferries, des navires spécialisés dans le transport de décharges lourdes ou encombrantes, telles que les grumes de bois, les pipe-lines; etc.

Les porte-conteneurs :

Les porte-conteneurs sont des navires puissants et rapides, les portes boites en jargon de marin. Ce sont les navires qui transportent exclusivement des conteneurs ils peuvent transporter jusqu'à 3000 conteneurs (le plus grand porte conteneur au monde appartient au groupe CMA CGM qui fait environ 400 m de long et peut transporter 217 000 conteneurs).

Les navires rouliers :

Les navires rouliers sont équipés d'une rampe arrière relevable qui prend appui sur la descente du port. Cette technique dérivée de celle du ferry-boat et appelée « Roll-on/Roll-off » ou « Ro-Ro », présente l'avantage de permettre la manutention d'à peu près tous les types de marchandises. On distingue alors :

Le roulage direct (manutention de charge sur roues ou chenilles) : semi-remorques routières, voitures, wagons parfois.

Le roulage indirect : dans lequel on utilise un matériel spécialisé pour charger les marchandises à bord : remorque-esclaves pour les colis lourds ou encombrants, chariot sur pneus équipés d'un palonnier pour déplacer les conteneurs.

Les navires mixtes Ro-Ro + conteneur :

Les navires mixtes Ro-Ro + conteneur combinent l'intérêt de conteneur et de la manutention par portique, avec celle de roulage pour les colis non conteneurisables et pour les ports non équipés. Leurs avantages et donc la souplesse. Le souci de polyvalence amène à équiper aussi ses navires de prises pour conteneurs frigorifiques, de cales pour marchandises sèches manutentionnées en conventionnel, et parfois deciternes.

Les navires porte-barges :

Les navires porte-barges sont des navires munis d'appareils permettant la manutention

des barges fluviales destinées au transport océanique de celles-ci.

Les cargos conventionnels :

Les cargos conventionnels disposent à bord de leurs propres moyens de manutention (bigues, grues, palans). Les marchandises sont chargées en cale par des panneaux coulissants. Ces navires classiques perdent peu à peu de leur importance.

Dans le cas des céréales leurs importations ou exportations se font par le biais des vraquiers ou céréaliers qui sont des navires spécialisés.

1.9 Conclusion et Position du problème

La crise de blé et des céréales en général à donner lieu à des mesures législatives très nombreuses et très complexes; il est presque impossible de les connaître, à qui ne s'est pas astreint à se tenir au courant minutieusement, au jour le jour. Pour ne donner qu'un exemple, celui de la France, le nombre des lois, décrets, arrêtés, règlements relatifs aux céréales, depuis le début de la crise ukrainienne, dépasse actuellement la centaine, les gouvernants des autres grands pays à céréales n'ont guère été moins féconds. Tous ces facteurs ont influé considérablement sur le prix de ces matières premières et qui ont fait presque doubler son prix, donnant place à un bon nombre de spéculations et autres supputations de la part des agents économiques. Le problème majeur est que ces instabilités donnent lieu à une balance économique plus lourde sur les pays consommateurs qui sont les principaux acheteurs sur le marché mondial. Pour chercher un équilibre économique, ces états doivent prévoir une montée des cours de ces matières premières et c'est sur cette dernière que notre travail se portera. Nous ferons une analyse de la relation offre \sim Prix, en matière céréale : cas de l'Entreprise Portuaire de Béjaïa.

CHAPITRE 2

RÉGRESSION LINÉAIRE SIMPLE

L'analyse de régression peut être définie comme la recherche de la relation stochastique que lie deux ou plusieurs variables. Son champ d'application peut toucher à-peu-près tous les domaines, parmi lesquels on peut citer la physique, l'astronomie, la biologie, la chimie, la médecine, la géographie, la sociologie, l'histoire, l'économie, la linguistique et le droit.

La régression est une des méthodes les plus connues et les plus appliquées en statistique pour l'analyse de données quantitatives. Elle est utilisée pour établir une liaison entre une variable quantitative et une ou plusieurs autres variables quantitatives, sous la forme d'un modèle. Si nous nous intéressons à la relation entre deux variables, nous parlerons de régression simple en exprimant une variable en fonction de l'autre. Si la relation porte entre une variable et plusieurs autres variables, nous parlerons de régression multiple. La régression linéaire simple et multiple sont des classes particulières de modèle de régression.

Le but est d'expliquer une variable Y , appelée variable endogène (ou expliquée) par une (X) ou plusieurs variables exogènes (ou explicatives) (X_1, X_2, \dots, X_n) à travers une fonction affine. Ces modèles de régression font partie d'une classe plus large de modèle appelé modèle linéaire général.

Ce présent chapitre définit clairement en quoi consiste la régression simple. [2], [6], [13]

2.1 La régression linéaire simple

En introduisant un modèle de régression linéaire avec une seule variable explicative, nous parlons de régression simple. Nous définissons ainsi la droite de régression qui se traduit par $Y = \beta_1 X + \beta_0$ que nous devons estimer à partir des données d'un échantillon en utilisant un coefficient de corrélation entre -1 et 1 et un coefficient nulle qui indique qu'il n'y a aucune relation entre ces variables. Une situation courante dans n'importe quel domaine en science est d'avoir à sa disposition deux ensembles de données de taille n , y_1, y_2, \dots, y_n et x_1, x_2, \dots, x_n , obtenus expérimentalement ou mesurés sur une population. Le problème de la régression consiste à rechercher une relation pouvant éventuellement exister entre les X et les Y , par exemple de la forme $Y = f(X)$. Lorsque la relation recherchée est affine, c'est-à-dire de la forme $Y = \beta_1 X + \beta_0$, nous parlons de régression linéaire. Mais même si une telle relation est effectivement présente, les données mesurées ne vérifient pas, en général, cette relation exactement. Pour tenir compte dans le modèle mathématique des erreurs observées, on considère les données y_1, y_2, \dots, y_n comme étant des réalisations d'une variable aléatoire Y et parfois aussi les données x_1, x_2, \dots, x_n comme étant des réalisations d'une variable aléatoire X . Nous disons que la variable Y est la variable dépendante ou variable expliquée et que la variable X est la variable explicative. [13]

2.2 Modèle

Considérons un échantillon d'observations pairées $(x_i, y_i), i = 1, \dots, n$ indépendants et identiquement distribués. Par ce modèle de régression linéaire, nous cherchons à prédire les valeurs de Y_i en fonction des valeurs prises par, la relation :

$$Y_i = \beta_1 X_i + \beta_0 + \varepsilon_i$$

Où

- Y est la variable dépendante (variable endogène quantitative),
- β_0 et β_1 sont les coefficients (représenté par ordonnée à l'origine et pente du modèle),
- X est la variable indépendante (variable exogène),
- ε l'erreur aléatoire (permet de résumer toute l'information non prise en compte dans la relation linéaire entre Y et X).

Les ε_i sont des quantités aléatoires inobservables qui représentent les erreurs. Ainsi les y_i sont des variables aléatoires (car elles dépendent des ε_i), alors que les x_i sont considérés

comme des nombres fixés. En supposant l'espérance des ε_i nulle, on a ainsi :

$$\begin{aligned} E(y_i) &= \beta_0 + \beta_1 x_i + E(\varepsilon_i) \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

et

$$\begin{aligned} \text{Var}(y_i) &= \text{Var}(\beta_0 + \beta_1 x_i + \varepsilon_i) \\ &= \text{Var}(\varepsilon_i) \end{aligned}$$

Hypothèse 2.1.

Ces hypothèses permettent de déterminer les propriétés des estimateurs et de mettre en places des outils de statistique (tests d'hypothèses, intervalles de confiance).

1. Hypothèses sur X et Y

Ce sont des grandeurs numériques mesurées sans erreur X qui est une donnée (exogène) dans le modèle. Y est aléatoire (c'est-à-dire la seule erreur qu'on peut avoir sur Y provient des insuffisances de X à expliquer ses valeurs dans le modèle).

2. Hypothèses sur le terme aléatoire

$H_1 : E(\varepsilon_i) = 0 \forall_i$ (En moyenne les erreurs s'annulent).

$H_2 : \text{Var}(\varepsilon_i) = \sigma^2 \forall_i$ (La variance de l'erreur est constante et ne dépend pas de l'observation).

$H_3 : \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \forall_{i \neq j}$ (les erreurs sont indépendantes).

$H_4 : \text{Cov}(X_i, \varepsilon_j) = 0 \forall_{i \neq j}$ (l'erreur est indépendante de la variable exogène).

$H_5 : \varepsilon_i \equiv N(0, \sigma^2) \forall_i$

(\equiv c'est-à-dire *i.i.d* : sont des variables aléatoires qui suivent toutes la même loi de probabilité et sont indépendantes).

2.3 Estimation des paramètres par la méthode des Moindres Carrés Ordinaires (MCO)

Estimer β_0 et β_1 revient à minimiser la sommes des carrés des erreurs :

$$\text{Min} \sum_{i=1}^n \varepsilon_i^2 = \text{Min} \sum_{i=1}^n (Y_i - \beta_1 X_i - \beta_0)^2 = \text{Min} \sum_{i=1}^n S_i^2$$

2.3. ESTIMATION DES PARAMÈTRES PAR LA MÉTHODE DES MOINDRES
CARRÉS ORDINAIRES (MCO)

Pour que cette fonction ait un minimum, il faut que les deux dérivées par rapport à β_0 et β_1 soient nulles :

$$\begin{aligned}
 \frac{\delta S_i}{\delta \beta_0} = 0 &\Rightarrow 2 \sum_{i=1}^n (-1)(Y_i - \beta_1 X_i - \beta_0) = 0 \\
 &\Rightarrow -2 \sum_{i=1}^n (Y_i - \beta_1 X_i - \beta_0) = 0 \\
 &\Rightarrow \sum_{i=1}^n Y_i - \beta_1 \sum_{i=1}^n X_i - n\beta_0 = 0 \\
 &\Rightarrow \hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n Y_i - \frac{\beta_1}{n} \sum_{i=1}^n X_i \\
 &\Rightarrow \hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}
 \end{aligned}$$

Tel que $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$

$$\begin{aligned}
 \frac{\delta S_i}{\delta \beta_1} = 0 &\Rightarrow 2 \sum_{i=1}^n (-X_i)(Y_i - \beta_1 X_i - \beta_0) = 0 \\
 &\Rightarrow \sum_{i=1}^n X_i Y_i = \beta_1 \sum_{i=1}^n X_i^2 + \beta_1 \sum_{i=1}^n X_i
 \end{aligned}$$

$$\begin{aligned}
 \sum_{i=1}^n X_i Y_i &= \beta_1 \sum_{i=1}^n X_i^2 + (\bar{Y} - \beta_1 \bar{X}) \sum_{i=1}^n X_i \\
 \Rightarrow \sum_{i=1}^n X_i Y_i &= \beta_1 \sum_{i=1}^n X_i^2 + \bar{Y} \sum_{i=1}^n X_i - \beta_1 \bar{X} \sum_{i=1}^n X_i \\
 \Rightarrow \beta_1 \left(\sum_{i=1}^n X_i - \bar{X} \sum_{i=1}^n X_i \right) &= \sum_{i=1}^n X_i Y_i - \bar{Y} \sum_{i=1}^n X_i \\
 \Rightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^n X_i Y_i - \bar{Y} \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2 - \bar{X} \sum_{i=1}^n X_i} \\
 \Rightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2} \\
 \Rightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}
 \end{aligned}$$

Donc, les estimateurs des MCO du modèle de régression linéaire simple $Y_i = \beta_1 X_i + \beta_0 + \varepsilon_i$ sont :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{et} \quad \hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}$$

Propriétés 2.1. 1. $E(\hat{\beta}_0) = \beta_0$ donc $\hat{\beta}_0$ est sans biais.

2. $E(\hat{\beta}_1) = \beta_1$ donc $\hat{\beta}_1$ est sans biais.

3. $Var(\hat{\beta}_0) = \hat{\sigma}^2(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}$

$$4. \text{Var}(\widehat{\beta}_1) = \widehat{\sigma}^2(\widehat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$5. \text{Cov}(\widehat{\beta}_0, \widehat{\beta}_1) = 0$$

$$6. s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$7. s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$8. \text{Cov}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$9. \widehat{\beta}_1 = \frac{\text{Cov}_{xy}}{s_x^2}$$

$$10. \text{L'estimateur de la variance } \sigma^2 \text{ est } S^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \widehat{y}_i)^2 = \frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2$$

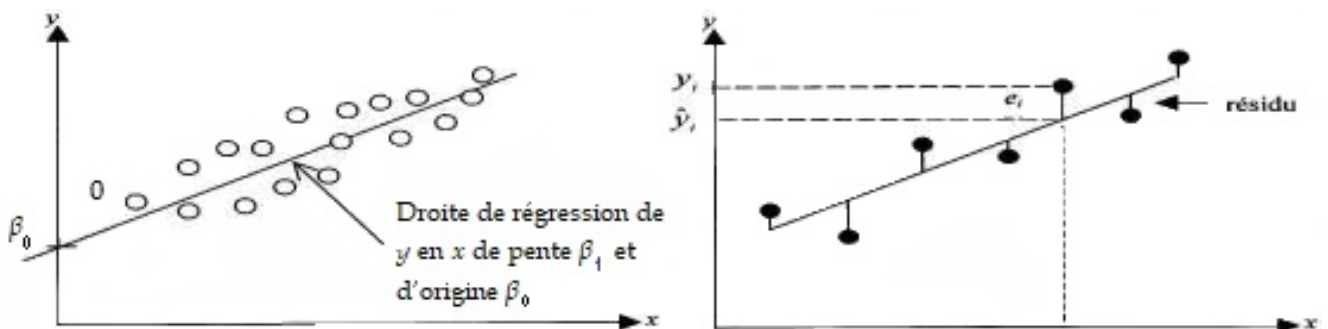


Figure 2.1: Droite de régression, résidus, linéarité.

2.4 Evaluation de la qualité de la régression

Pour mesurer la qualité de l'approximation d'un nuage $(x_i, y_i)_{i=1, \dots, n}$ par sa droite des moindres carrés (après tout nous pouvons toujours faire passer une droite par n'importe quel nuage!), nous calculons son coefficient de corrélation linéaire défini par

$$r_{xy} = \frac{\text{Cov}_{xy}}{s_x s_y}$$

C'est un nombre compris entre -1 et +1, qui vaut +1 (resp. -1) si les points du nuage sont exactement alignés sur une droite de pente β_1 positive (resp. négative). Ce coefficient est

une mesure de la dispersion du nuage. Nous considérons que l'approximation d'un nuage par sa droite des moindres carrés est de bonne qualité lorsque $|r_{xy}|$ est proche de 1 (donc r_{xy} proche de +1 ou de -1) et de médiocre qualité lorsque $|r_{xy}|$ est proche de 0. En pratique, nous estimons souvent la régression acceptable lorsque $|r_{xy}| \geq \frac{\sqrt{3}}{2} = \sqrt{0.75} = 0.866\dots$. Parfois nous préférons calculer non plus r_{xy} mais son carré noté $R^2 = r_{xy}r_{xy}$ car nous avons la relation suivante

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{(DT)} = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{(DA)} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{(DR)}$$

qui exprime que la dispersion totale de Y (DT) est égale à la dispersion autour de la régression (DA) plus la dispersion due à la régression (DR). Or on peut vérifier que l'on a $R^2 = \frac{DR}{DT}$, c'est-à-dire que le R^2 représente la part de la dispersion totale de Y que l'on peut expliquer par la régression.

Ainsi, si nous obtenons une valeur de $R^2 = 0,85$ (et donc $|r| = 0.92\dots$), cela signifie que la modélisation par la droite des moindres carrés explique 85% de la variation totale, ce qui est un très bon résultat.

Cependant, même avec un R^2 excellent (proche de 1), le modèle linéaire peut encore être rejeté. En effet, pour être assuré que les formules données $\hat{\beta}_1$ et $\hat{\beta}_0$ fournissent de bonnes estimations de la pente et de l'ordonnée à l'origine de la droite de régression, il est nécessaire que les résidus ϵ_i soient indépendants et distribués aléatoirement autour de zéro. Ces hypothèses ne sont pas forcément faciles à vérifier. Un tracé des résidus et un examen de leur histogramme permet de détecter une anomalie grossière mais il faut faire appel à des techniques statistiques plus élaborées pour tester réellement ces hypothèses.

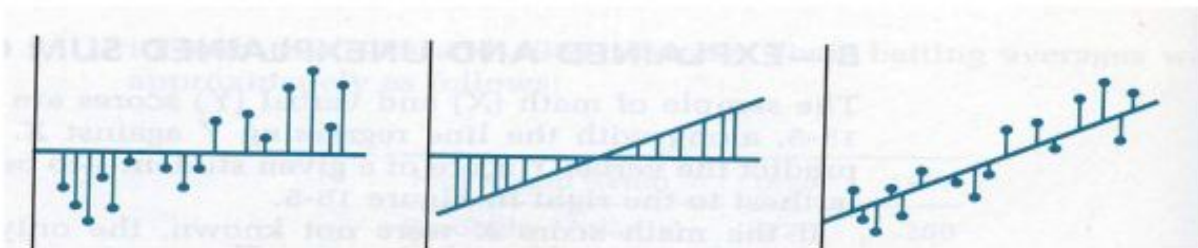


Figure 2.2: Illustration de la formule $DT = DA + DR$.

2.5 Prévisions

Si $y = \hat{\beta}_1 x + \hat{\beta}_0$ est la droite des moindres carrés d'un nuage de points (x_i, y_i) $i = 1, \dots, n$, nous appellons les valeurs prédites de y par le modèle $\hat{y}_i = \hat{\beta}_1 x_i + \hat{\beta}_0$.

Notons, cependant, que s'il peut sembler naturel d'utiliser une valeur prédite pour compléter les données initiales dans l'intervalle des valeurs de X , nous nous garderons de prédire, sans de multiples, précautions supplémentaires des valeurs de X en dehors de cet intervalle. En effet, il se peut que la relation entre X et Y ne soit pas du tout linéaire mais qu'elle nous soit apparue comme telle à tort parce que les x_i sont proches les uns des autres.

2.6 Tests et intervalles de confiance

Nous allons voir dans cette section comment effectuer des tests d'hypothèses sur les paramètres β_1 et β_0 du modèle de régression simple, ainsi que sur la manière de construire des intervalles de confiance. [6]

2.6.1 Test sur la pente β_1

L'estimateur $\hat{\beta}_1$ est normalement distribué, d'espérance mathématique β_1 et de variance que le nous notons ici par :

$$\sigma^2(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Il s'ensuit que la quantité :

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma(\hat{\beta}_1)}$$

Suit une loi normale centrée réduite. Or cette quantité ne peut pas être utilisé pour un problème de test d'hypothèses puisque nous ne connaissons pas la valeur de σ . En pratique, nous estimons $\sigma^2(\hat{\beta}_1)$ par :

$$S^2(\hat{\beta}_1) = \frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Où S^2 est l'estimateur sans biais de σ^2 défini en **Propriétés 2.1**. Nous pouvons alors montrer que la quantité :

$$\frac{\hat{\beta}_1 - \beta_1}{S(\hat{\beta}_1)}$$

Suit une loi de student à $(n - 2)$ degrés de liberté. Il s'ensuit que dans un problème de test d'hypothèse bilatéral où nous désirons tester l'hypothèse nulle :

$$H_0 : \text{''}\beta_1 = b_1\text{''}$$

Contre l'hypothèse alternative :

$$H_1 : \text{''}\beta_1 \neq b_1\text{''}$$

Nous utilisons la statistique :

$$t_c = \frac{\hat{\beta}_1 - b_1}{S(\hat{\beta}_1)}$$

Au seuil de signification α , nous rejettons l'hypothèse H_0 si :

$$|t_c| > t_{(\alpha/2, n-2)},$$

où la valeur $t_{(\alpha/2, n-2)}$ est le $(1 - \alpha/2)$ quantile d'une loi de Student à $(n - 2)$ degrés de liberté qui se trouve dans une table de Student.

Un test d'hypothèse particulièrement intéressant est le test de l'hypothèse nulle :

$$H_0 : \text{''}\beta_1 = 0\text{''}$$

contre l'hypothèse alternative :

$$H_1 : \text{''}\beta_1 \neq 0\text{''}$$

En effet, le non-rejet de l'hypothèse nulle implique un modèle avec un seul paramètre :

$$y_i = \beta_0 + \varepsilon_i,$$

Par contre, si cette hypothèse H_0 est rejetée, c'est-à-dire si :

$$|t_c| = \frac{|\hat{\beta}_1|}{S(\hat{\beta}_1)} > t_{(\alpha/2, n-2)},$$

Nous disons que la relation entre les x_i et y_i est significative au seuil de signification α .

2.6.2 Intervalle de confiance pour β_1

L'intervalle de confiance au niveau $(1 - \alpha)$ pour le paramètre β_1 est définie par :

$$\left[\widehat{\beta}_1 - t_{(\alpha/2, n-2)} \times S(\widehat{\beta}_1), \widehat{\beta}_1 + t_{(\alpha/2, n-2)} \times S(\widehat{\beta}_1) \right]$$

Cet intervalle est construit de telle sorte qu'il contienne le paramètre inconnu β_1 avec une probabilité de $(1 - \alpha)$.

Rappelons que les hypothèses H_0 et H_1 peuvent également être testées à l'aide de l'intervalle de confiance; nous rejettons alors l'hypothèse :

$$H_0 : " \beta_1 = b_1 "$$

contre l'hypothèse :

$$H_1 : " \beta_1 \neq b_1 "$$

au seuil de signification α si et seulement si la valeur testée b_1 ne se trouve pas dans l'intervalle de confiance de β_1 au niveau $(1 - \alpha)$.

2.6.3 Test sur l'ordonnée à l'origine β_0

Nous avons vu que l'estimateur $\widehat{\beta}_0$ est normalement distribué, d'espérance β_0 et de variance :

$$\sigma^2(\widehat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

Il s'ensuit que la quantité :

$$\frac{\widehat{\beta}_0 - \beta_0}{\sigma(\widehat{\beta}_0)}$$

suit une loi normale centrée réduite. Or cette quantité ne peut pas être utilisée pour un problème de test d'hypothèses puisque nous ne connaissons pas la valeur de σ . En pratique, nous estimons $\sigma^2(\widehat{\beta}_0)$ par :

$$S^2(\hat{\beta}_0) = \frac{S^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

Nous pouvons alors montrer que la quantité :

$$\frac{\hat{\beta}_0 - \beta_0}{S(\hat{\beta}_0)}$$

suit une loi de student à $(n - 2)$ degrés de liberté.

Il s'ensuit que dans un problème de test d'hypothèse bilatéral où nous désirons tester l'hypothèse nulle :

$$H_0 : \text{''}\beta_0 = b_0\text{''}$$

Contre l'hypothèse alternative :

$$H_1 : \text{''}\beta_0 \neq b_0\text{''}$$

Nous pouvons utiliser la statistique :

$$t_c = \frac{\hat{\beta}_0 - b_0}{s(\hat{\beta}_0)}$$

Nous rejettons l'hypothèse H_0 au seuil de signification α si :

$$|t_c| > t_{(\alpha/2, n-2)},$$

Un test d'hypothèses souvent effectué est le test de l'hypothèse nulle :

$$H_0 : \text{''}\beta_0 = 0\text{''}$$

contre l'hypothèse alternative :

$$H_1 : \text{''}\beta_0 \neq 0\text{''}$$

Le non-rejet de cette hypothèse nulle implique un modèle avec un seul paramètre :

$$y_i = \beta_1 x_i + \varepsilon_i,$$

Par contre, si cette hypothèse H_0 est rejetée, c'est-à-dire si :

$$|t_c| = \frac{|\widehat{\beta}_0|}{S(\widehat{\beta}_0)} > t_{(\alpha/2, n-2)},$$

La constante β_0 du modèle de régression $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ est dite significative au seuil de signification α .

2.6.4 Intervalle de confiance pour β_0

Un intervalle de confiance au niveau $(1 - \alpha)$ pour le paramètre β_0 est définie par :

$$\left[\widehat{\beta}_0 - t_{(\alpha/2, n-2)} \times S(\widehat{\beta}_0), \widehat{\beta}_0 + t_{(\alpha/2, n-2)} \times S(\widehat{\beta}_0) \right]$$

Cet intervalle est construit de telle sorte qu'il contienne le paramètre inconnu β_0 avec une probabilité de $(1 - \alpha)$.

Rappelons que les hypothèses H_0 et H_1 peuvent également être testées à l'aide de l'intervalle de confiance; on rejette alors l'hypothèse :

$$H_0 : " \beta_0 = b_0 "$$

contre l'hypothèse :

$$H_1 : " \beta_0 \neq b_0 "$$

au seuil de signification α si et seulement si la valeur testée b_0 ne se trouve pas dans l'intervalle de confiance de β_0 au niveau $(1 - \alpha)$.

2.6.5 Intervalles et régions de confiance pour les coefficients de régression lorsque la variance σ^2 est connue

Lorsque la variance σ^2 est connue, un intervalle de confiance de niveau de confiance $(1 - \alpha)$ pour β_0 est

$$\left[\widehat{\beta}_0 - t_{(\frac{1-\alpha}{2}, n-2)} \times \sigma(\widehat{\beta}_0), \widehat{\beta}_0 + t_{(\frac{1-\alpha}{2}, n-2)} \times \sigma(\widehat{\beta}_0) \right]$$

Un Intervalle de confiance de niveau $(1-\alpha)$ pour β_1 donné par :

$$\left[\widehat{\beta}_1 - t_{(\frac{1-\alpha}{2}, n-2)} \times \sigma(\widehat{\beta}_1), \widehat{\beta}_1 + t_{(\frac{1-\alpha}{2}, n-2)} \times \sigma(\widehat{\beta}_1) \right]$$

2.6.6 Intervalle de prédiction

Lorsque les hypothèses sont vérifiées, Nous pouvons également calculer des intervalles de prédiction. Un intervalle de prédiction au niveau $(1 - \alpha)$ est construit de telle sorte qu'il contienne avec une probabilité de $(1 - \alpha)$ la valeur (future) de la variable Y d'un individu pour lequel nous aurons $X = x$. Nous notons ici une telle quantité par $y_f(x)$. Si les hypothèses du modèle sont correctes, nous avons donc :

$$y_f(x) = \beta_0 + \beta_1 x + \varepsilon_f = \mu_y(x) + \varepsilon_f,$$

où ε_f est une erreur (future) distribuée selon une loi normale d'espérance nulle de variance σ^2 . Un intervalle de prédiction pour $y_f(x)$ au niveau $(1 - \alpha)$ est alors donné par :

$$[\mu_y(x) - \Phi^{-1}(\alpha/2) \times \sigma, \mu_y(x) + \Phi^{-1}(\alpha/2) \times \sigma]$$

Cette formule est pourtant inutilisable en pratique, puisque β_0 , β_1 et σ^2 sont inconnus. Notons alors que la quantité :

$$\widehat{y}_f(x) = \beta_0 + \beta_1 x - \varepsilon_f = \widehat{y}_f(x) - \varepsilon_f$$

est distribuée selon une loi normale d'espérance $\mu_y(x)$ et de variance

$$\sigma^2(\widehat{y}_f(x)) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

qui peut être approximée par

$$S^2(\widehat{y}_f(x)) = S^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Nous pouvons montrer alors que la quantité :

$$\frac{\widehat{y}_f(x) - \mu_y(x)}{S(\widehat{y}_f(x))} = \frac{\widehat{y}(x) - y_f(x)}{S(\widehat{y}_f(x))}$$

suit une loi de Student à $(n-2)$ degrés de liberté. Il s'en suit qu'un intervalle de prédiction au niveau $(1 - \alpha)$ pour $y_f(x)$ est donné par :

$$[\widehat{y}(x) - t_{(\alpha/2, n-2)} \times S(\widehat{y}_f(x)), \widehat{y}(x) + t_{(\alpha/2, n-2)} \times S(\widehat{y}_f(x))]$$

Notons que contrairement à un intervalle de confiance, la longueur d'un intervalle de prédiction ne tend pas vers zéro lorsque la taille de l'échantillon augmente. Dans ce cas, l'intervalle de prédiction s'approche alors de l'intervalle

$$\mu_y(x) \pm \Phi^{-1}(\alpha/2).$$

2.7 Test de signification global de Fisher

Nous avons définis trois sommes des carrées SC_{tot} , SC_{reg} , SC_{res} . Dans cette section, nous allons voir comment on peut utiliser ces sommes de carrés pour tester l'hypothèse :

$$H_0 : \text{''}\beta_1 = 0\text{''}$$

Si cette hypothèse est vérifiée, nous pouvant montrer que les espérances de ces sommes de carrés valent, respectivement :

$$E(SC_{tot}) = (n - 1)\sigma^2,$$

$$E(SC_{reg}) = \sigma^2,$$

$$E(SC_{res}) = (n - 2)\sigma^2.$$

Où

SC_{tot} est la somme des carrés totale.

SC_{reg} est la somme des carrés régression.

SC_{res} est la somme des carrés résiduelle.

Les quantités suivantes sont alors des estimateurs sans biais de σ^2 :

$$MC_{tot} = \frac{SC_{tot}}{n - 1},$$

$$MC_{reg} = \frac{SC_{reg}}{1},$$

$$MC_{res} = \frac{SC_{res}}{n - 2}.$$

Où MC symbolise la moyenne des carrés. Les nombres $(n - 1)$, 1 et $(n - 2)$ formant le dénominateur de ces estimateurs représentent les degrés de liberté associés à ces sommes de carrés, c'est-à-dire le nombre de termes linéairement indépendants impliqués dans chacune de ces sommes. Les degrés de libeté représentent aussi le nombre d'éléments dans la somme des carrés moins le nombre de paramètres estimés dans cette somme. Notons que lorsque H_0 n'est pas vérifiée, seul MC_{res} est encore un estimateur sans biais de σ^2 . Il s'agit en fait de l'estimateur S^2 défini auparavant. Lorsque l'hypothèse H_0 est vérifiée, les quantités $\frac{SC_{tot}}{\sigma^2}, \frac{SC_{reg}}{\sigma^2}, \frac{SC_{res}}{\sigma^2}$ suivent des lois de χ^2 avec, respectivement, $(n - 1)$, 1 et $(n - 2)$ degrés de liberté. De plus, les quantités $\frac{SC_{reg}}{\sigma^2}, \frac{SC_{res}}{\sigma^2}$ sont indépendantes. Il s'ensuit que lorsque l'hypothèse H_0 est vérifiée, la statistique :

$$F_c = \frac{\left(\frac{SC_{reg}}{1 \cdot \sigma^2}\right)}{\left(\frac{SC_{res}}{(n-2) \cdot \sigma^2}\right)} = \frac{SC_{reg}}{SC_{res}} \cdot \frac{n-2}{1} = \frac{MC_{reg}}{MC_{res}}$$

suit une loi de Fisher avec 1 et $(n - 2)$ degrés de liberté. Nous rejettons ainsi l'hypothèse H_0 au seuil de signification α si :

$$F_c > F_{(\alpha, 1, n-2)},$$

où la valeur critique $F_{(\alpha, 1, n-2)}$ est le $(1 - \alpha)$ quantile d'une loi de Fisher avec 1 et $(n - 2)$ degrés de liberté que nous trouvons dans une table de Fisher. Cette procédure est appelée analyse de la variance. Nous présentons, généralement, les quantités nécessaire au calcul de la statistique de test F_c dans un tableau d'analyse de variance, abrégé par **tableau d'ANOVA** et qui prend la forme de ce tableau ci-dessous:

Nous remarquons que ce test qui utilise la statistique F_c est équivalent au test

Source de la variance	Degré de liberté	Somme des carrés	Moyenne des carrés
Régression	1	$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$MC_{reg} = \frac{SC_{reg}}{1}$
Résiduelle	$n - 2$	$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$MC_{res} = \frac{SC_{re}}{n-2}$
Total	$n - 1$	$SC_{res} = \sum_{i=1}^n (y_i - \bar{y})^2$	

Figure 2.3: Table de ANOVA

présenté dans la section précédente pour la même hypothèse qui utilise la statistique t_c . En effet, comme :

$$\begin{aligned}
SC_{reg} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
&= \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\
&= \sum_{i=1}^n (\bar{y} + \beta_1(x_i - \bar{x}) - \bar{y})^2
\end{aligned}$$

alors

$$\begin{aligned}
F_c &= \frac{SC_{reg}}{SC_{res}/(n-2)} \\
&= \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{s^2} \\
&= \frac{\hat{\beta}_1^2}{s^2 \beta_1} \\
&= t_c^2
\end{aligned}$$

L'équivalence des deux tests est alors la conséquence du fait que :

$$F_{(\alpha,1,n-2)} = t_{(\alpha/2,n-2)}^2.$$

2.8 Cas particulier

Nous avons vu dans ce chapitre que lorsque nous ne rejetton pas les hypothèses $\beta_0 = 0$ ou $\beta_1 = 0$, nous pouvons être amené à considérer des modèles avec un seul paramètre. Nous allons montrer maintenant comment ces modèles se résolvent avec la méthode des moindres carrés.

Modèle sans variable explicative

Partons du modèle :

$$y_i = \beta_0 + \varepsilon_i,$$

En considérant également les hypothèses émises, nous avons donc les y_i qui sont indépendants et normalement distribués avec :

$$E(y_i) = \beta_0,$$

$$Var(y_i) = \sigma^2.$$

A partir d'un estimateur $\hat{\beta}_0$ de β_0 , nous définissons les valeurs estimées :

$$\hat{y}_i = \hat{\beta}_0,$$

et les résidus :

$$\varepsilon_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0.$$

L'estimateur $\hat{\beta}_0$ obtenu par la méthode des moindres carrés est le point qui minimise la fonction définie par la somme des carrés des résidus :

$$z = f(\beta_0) = \sum_{i=1}^n (y_i - \beta_0)^2$$

c'est-à-dire le point qui annule sa dérivée :

$$\frac{\delta z}{\delta \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0)$$

$\hat{\beta}_0$ est donc la solution de :

$$\sum_{i=1}^n y_i = n\beta_0.$$

On trouve ainsi :

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}.$$

On a donc ici :

$$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = SC_{tot},$$

et

$$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - y_i).$$

L'estimateur $\hat{\beta}_0 = \hat{y}_i$ est normalement distribué, d'espérance :

$$E(\hat{\beta}_0) = E(\bar{y}) = E(y_i) = \beta_0$$

et de variance :

$$\sigma^2(\hat{\beta}_0) = Var(\bar{y}) = \frac{Var(y_i)}{n} = \frac{\sigma^2}{n}.$$

Par ailleurs, l'espérance de la somme des carrés des résidus vaut :

$$E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) = E\left(\sum_{i=1}^n (y_i - \bar{y})^2\right) = \sigma^2(n-1).$$

Nous obtenons ainsi un estimateur sans biais de σ^2 en posant :

$$S^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n-1}$$

est un estimateur sans biais de $\sigma^2(\hat{\beta}_0)$ qui vaut :

$$S^2(\hat{\beta}_0) = \frac{s^2}{n}.$$

Nous pouvons montrer que la quantité :

$$\frac{\hat{\beta}_0 - \beta_0}{S(\hat{\beta}_0)}.$$

où $S(\hat{\beta}_0)$ est la racine carrée de $S^2(\hat{\beta}_0)$, suit une loi de student avec $(n-1)$ degrés de liberté.

Nous rejettons ainsi l'hypothèse nulle :

$$H_0 : \text{''}\beta_0 = b_0\text{''}$$

pour l'hypothèse alternative :

$$H_0 : \text{''}\beta_0 \neq b_0\text{''}$$

au seuil de signification α si :

$$|t_c| = \frac{|\hat{\beta}_0 - b_0|}{S(\hat{\beta}_0)},$$

si

$$|t_c| > t_{\alpha/2}(2, n-1)$$

où la valeur critique $t_{\alpha/2}(2, n-1)$ est le $(1\alpha/2)$ quantile d'une loi de Student avec $(n-1)$ degrés de liberté que nous trouvons dans une table de Student. Nous voyons ainsi que cette procédure est identique à la procédure du t-test classique utilisée pour inférer sur la moyenne β_0 des y_i .

Modèle sans constante

Partons maintenant du modèle :

$$y_i = \beta_1 x_i + \varepsilon_i.$$

En considérant également les hypothèses émises, nous avons donc les y_i qui sont indépen-

dants et normalement distribués avec :

$$E(y_i) = \beta_1 x_i,$$

$$\text{Var}(y_i) = \sigma^2.$$

A partir d'un estimateur $\hat{\beta}_1$ de β_1 , nous définissons les valeurs estimées :

$$\hat{y}_i = \hat{\beta}_1 x_i.$$

et les résidus :

$$\varepsilon_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 x_i.$$

L'estimateur $\hat{\beta}_1$ obtenu par la méthode des moindres carrés est le point qui minimise la fonction définie par la somme des carrés des résidus :

$$z = f(\beta_0) = \sum_{i=1}^n (y_i - \beta_1 x_i)^2,$$

c'est-à-dire, le point qui annule sa dérivée :

$$\frac{\delta z}{\delta \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_1 x_i).$$

$\hat{\beta}_1$ est donc la solution de :

$$\sum_{i=1}^n x_i y_i = \beta_1 \sum_{i=1}^n x_i^2.$$

Nous trouvons ainsi :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

La particularité d'un modèle sans constante est que la droite des moindres carrés :

$$\hat{y}(x) = \hat{\beta}_1 x$$

ne passe pas par le point (\bar{x}, \bar{y}) et que nous n'avons pas :

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i.$$

Par conséquent, la somme des résidus n'est pas nulle. Il en résulte également que nous n'avons pas non plus :

$$SC_{tot} = SC_{reg} + SC_{res}.$$

L'estimateur $\widehat{\beta}_1$ est normalement distribué d'espérance :

$$\begin{aligned} E(\widehat{\beta}_1) &= \frac{\sum_{i=1}^n x_i E(y_i)}{\sum_{i=1}^n x_i^2} \\ &= \frac{\beta_1 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} \\ &= \beta_1. \end{aligned}$$

et de variance :

$$\begin{aligned} \sigma^2(\widehat{\beta}_1) &= \frac{\sum_{i=1}^n x_i^2 \text{Var}(y_i)}{\sum_{i=1}^n x_i^2} \\ &= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2}. \end{aligned}$$

Les valeurs estimées y_i sont également normalement distribué d'espérance :

$$\begin{aligned} E(y_i) &= E(\widehat{\beta}_1 x_i) \\ &= x_i E(\widehat{\beta}_1) \\ &= x_i \beta_1. \\ &= E(y_i) \end{aligned}$$

et de variance :

$$\text{Var}(\widehat{y}_i) = \text{Var}(\widehat{\beta}_1 x_i) = x_i^2 \text{Var}(\widehat{\beta}_1) = \frac{x_i^2 \sigma^2}{\sum_{i=1}^n x_i^2}$$

Par ailleurs nous pouvons calculer l'espérance de la somme des carrés des résidus :

$$\begin{aligned}
 E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) &= E\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2\right) \\
 &= \sum_{i=1}^n \text{Var}(y_i) - \sum_{i=1}^n \text{Var}(\hat{y}_i) \\
 &= n\sigma^2 - \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} \\
 &= \sigma^2(n-1)
 \end{aligned}$$

Nous obtenons ainsi un estimateur sans biais de σ^2 en posant :

$$S^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}$$

Nous avons ainsi un estimateur :

$$S^2(\hat{\beta}_1) = \frac{S^2}{\sum_{i=1}^n x_i^2}$$

de $\sigma^2(\hat{\beta}_1)$ nous pouvons montrer que la quantité :

$$\frac{\hat{\beta}_1 - \beta_1}{S(\hat{\beta}_1)}$$

où $S(\hat{\beta}_1)$ suit une loi de Student avec $(n-1)$ degrés de liberté. Nous rejettons ainsi l'hypothèse :

$$H_0 : \text{"}\beta_1 = b_1\text{"}$$

contre l'hypothèse alternative :

$$H_1 : \text{"}\beta_1 \neq b_1\text{"}$$

au seuil de signification α si :

$$|t_c| = \frac{|\hat{\beta}_1 - b_1|}{S(\hat{\beta}_1)} > t_{(\alpha/2, n-1)}.$$

Remarque.

Le modèle de régression multiple est une généralisation du modèle de régression simple lorsque les variables explicatives sont en nombre fini. En effet, elle est utilisée pour décrire les variations d'une variable endogène associée à plusieurs variables exogènes.

Conclusion

La régression linéaire est une méthode de modélisation permettant d'établir une relation linéaire entre une variable continue dite "variable expliquée" ou dépendante et un ensemble d'autres variables continues dites "variables explicatives" ou indépendantes. Plus spécifiquement, elle propose un modèle explicatif qui permet de prédire la variable dépendante en fonction des variables indépendantes. La régression peut servir à remplacer une variable difficile à observer par une autre variable qui elle est relativement simple à mesurer. Elle peut aussi servir à comprendre les liens existants entre les variables pour établir les principales causes d'un phénomène.

CHAPITRE 3

MODÈLES MATHÉMATIQUES, APPLICATIONS ET INTERPRÉTATIONS DES RÉSULTATS

Introduction

Dans ce présent chapitre, nous faisons une analyse de la relation entre les quantités importées et le prix d'une céréale, en considérant quatre type de céréales, à savoir : le blé dur, le blé tendre, le maïs et l'orge. L'objectif de cette étude est d'examiner l'impact de la flambée des prix des matières en céréales sur la capacité de l'offre, en considérant le cas de l'Entreprise Portuaire de Béjaia.

Afin de réaliser cette analyse sur les quatre espèces de céréales déjà mentionnées, des modèles mathématiques linéaires reliant la quantité importée et le cours correspondant, ont été élaborés.

3.1 Applications sous le logiciel R

Le chapitre précédent est dédié aux différentes techniques classiques de la régression linéaire simple, ainsi, des résultats théoriques qui en découlent. En passant à la pratique, le traitement de données et d'analyse statistique est appliqué sous le logiciel R. Ce qui fait l'objet de ce dernier chapitre.

3.1.1 Présentation du logiciel R

Le logiciel R est un langage de programmation libre et un logiciel de statistique. Il est largement utilisé par les statisticiens, les datas mineurs, les data scientists afin de développer des logiciels statistiques et l'analyse des données.

R est une implémentation du langage de programmation S. Il fait partie de la liste des paquets GNU, il était écrit en "C", "Fortran" et "R", naît en Aout 1993 comme un projet de recherche de "Ross Ihaka" et "Robert Gentleman" à l'université d'Auckland. Sa première version "R 1.0.0" est publiée le 29 février 2000. La "version 2.0" est publiée le 4 octobre 2004 et la "version 3.0" le 3 avril 2013. En 2015, plusieurs acteurs économiques importants comme IBM, Microsoft ou encore la société RStudio créent le R Consortium pour soutenir la communauté R et financer des projets autour de ce langage à ce moment. Le langage compte plus de 200 bibliothèques développées par la communauté scientifique qui utilise R.

3.1.2 Justification du choix de l'environnement de programmation

Le choix s'est porté à l'emploi du logiciel R en raison que ce dernier est un environnement de traitement de données. En outre, Le logiciel R est complet, c'est-à-dire, il est dédié à la manipulation de données, à la construction des graphes et à la réalisation des tests statistiques.

3.2 Collecte de données Financières

Pour étudier les relations entre les quantités¹ importées et les cours² correspondants de matières en céréale, nous utilisons des données mensuelles de prix sur la période de

¹Les quantités des céréales importées sont collectées au sein du Département Statistiques de la Direction Exploitation de l'Entreprise Portuaire de Béjaia.

²Les cours mensuels des céréales sur le marché sont disponibles sur le site Web <https://www.agri-mutuel.com/cotations>.

janvier 2016 à décembre 2020, également pour les quantités des céréales importées, soit 60 observations. Le choix de ce pas de temps permet de réduire l'impact éventuel du bruit de haute fréquence, comme c'est souvent le cas pour les données financières.

3.3 Le Blé Dur

Le blé dur³ (*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.) est une espèce de blé caractérisée par son amande (intérieur du grain) dure et vitreuse. Cultivée depuis la Préhistoire, cette céréale est riche en protéines, y compris en gluten. Il ne peut donner que des pains peu levés mais convient parfaitement à la fabrication des pâtes alimentaires, couscous et boulghour.

3.3.1 Description des données statistiques sur le blé dur

X_{BD}	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
€	195.00	216.60	234.50	235.10	252.10	282.50

Table 3.1: La variable Prix (ou Cours) mesurée en €

Y_{BD}	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
tonne	0.00	0.00	0.00	6420.00	3676.00	57000.00

Table 3.2: La variable Quantité importée (ou Offre) mesurée en Tonne

3.3.2 Inférence sur les deux paramètres du modèle β_0 et β_1

	Valeur estimée	Intervalle de confiance	Coefficient de corrélation
Inférence sur β_0	$\hat{\beta}_0 = -98286.9$	$\beta_0 \in [-122685.8911, -73887.8314]$	$R = 0.5620$
Inférence sur β_1	$\hat{\beta}_1 = 445.5$	$\beta_1 \in [342.1036, 548.8298]$	

Table 3.3: Inférence sur les deux paramètres du modèle β_0 et β_1 cas du blé dur (MCO)

L'équation de régression linéaire s'écrit sous la forme

$$Y_{BD} = \hat{\beta}_0 + \hat{\beta}_1 X_{BD} = -98286.9 + 445.5 X_{BD}$$

³Source : <https://fr.wikipedia.org>

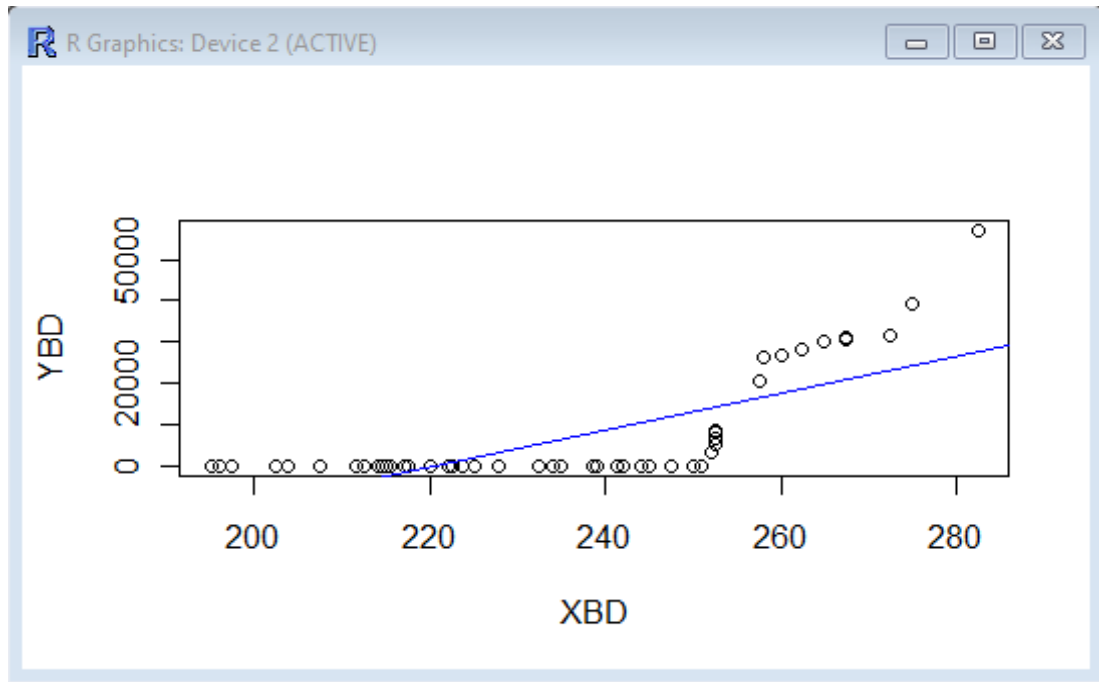


Figure 3.1: Graphe de Y_{BD} en fonction de X_{BD} et la droite de régression

Tests statistiques

- **Test sur l'ordonnée à l'origine β_0**

La réalisation de T_{β_0} est $|t_{\beta_0}| = \frac{|-98286.86|}{12189.05} = 8.064$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student⁴ la valeur critique $t_{(\alpha/2, n-2=58)} = 2.002$

Nous remarquons que $|t_{\beta_0}| = 8.064 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejettons l'hypothèse $H_0 : \beta_0 = 0$.

- **Test sur la pente β_1**

La réalisation de T_{β_1} est $|t_{\beta_1}| = \frac{|445.47|}{51.64} = 8.627$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique $t_{(\alpha/2, n-2=58)} = 2.002$

Nous remarquons que $|t_{\beta_1}| = 8.627 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejettons l'hypothèse $H_0 : \beta_1 = 0$.

- **Test global de significativité de la régression**

Il apparaît raisonnable de tester la significativité du modèle bâti, c'est-à-dire tester l'hypothèse $H_0 : \beta_1 = 0$ contre l'hypothèse $H_1 : \beta_1 \neq 0$, excepter la constante

⁴Pour la lecture des quantiles de la loi de Student et de la loi de Fisher-Snedecor, utiliser le lien suivant : <https://www.supagro.fr/cnamlr/statnet/tables.htm>

à l'origine β_0 . Ce test est basé sur la statistique de Fisher (à 1 et 58 degrés de liberté) $F_c = \frac{SC_{reg}/1}{SC_{res}/(n-2)} = 74.42$. La sortie R nous donne la valeur critique $p-value = 5.502e-12$ qui est inférieure au seuil de signification $\alpha = 0.05$, donc nous rejettons ainsi l'hypothèse $H_0 : \beta_1 = 0$ au seuil de signification $\alpha = 0.05$.

Etude sur les résidus du modèle bâti

- **Description statistique sur les résidus ε_i**

Min	1er Quantile	Median	Moyenne	3ème Quantile	Max	Variance
-13525.30	-6696.70	-718.10	0.00	6270.10	29442.50	8689

- **Test sur la normalité des résidus ε_i ⁵**

Ici nous utilisons le test de normalité de Shapiro-Wilk. La sortie R nous donne la valeur de la statistique de Shapiro-Wilk $W = 0.9495$ et la valeur critique $p-value = 0.014$. Puisque $p-value = 0.0147 < \alpha = 0.05$ alors nous rejetons l'hypothèse de normalité des résidus.

3.3.3 Interprétation des résultats obtenus et discussion

Dans le cas de la céréale "blé dur", le coefficient de corrélation entre les deux variables offre et prix est $R=0,5620$.

Cette valeur indique ainsi une relation linéaire positive existante, mais faible et probablement anecdotique, c'est-à-dire, les deux variables "offre" et "prix" sont liées à 56,20%.

Le test global de Fisher indique que le modèle linéaire bâti est significatif. Cependant, le test de Shapiro-Wilk nie la normalité des résidus du modèle.

⁵Le test le plus recommandé est celui de Shapiro-Wilk basé sur la comparaison de l'écart-type à une combinaison linéaire $\sum a_i w_i$ des étendues successives $w_1 = x_n - x_1, w_2 = x_{n-1} - x_2$, etc. Les coefficients de la combinaison dépendent de n et sont données dans des tables, ainsi que la région critique. Les calculs sont programmés dans la plus part des logiciels statistiques. [13]

3.3.4 Prédiction et comparaison avec des données réelles

Mois	Prix (X_{BDRel} €)	Quant Préd (tonnes)	Inter de conf	Quant import (tonnes)
Janvier	272.50	23102.81	[5143.75 , 41061.87]	0.00
Février	260.00	18425.41	[668.63 , 36182.19]	0.00
Mars	267.50	20875.48	[3020.68 , 38730.27]	30597.00
Avril	261.00	17979.94.61	[239.08 , 35720.80]	6545.00
Mai	262.50	18648.14	[883.18, 36413.10]	63200.00
Juin	267.50	20875.48	[3020.68 , 38730.27]	0.00
Juillet	277.50	25330.14	[7252.64, 43407.64]	26250.00
Août	347.50	56512.81	[35473.75, 77551.87]	0.00
Septembre	415.00	86581.81	[61017.96 , 112145.66]	0.00
Octobre	442.50	98832.15	[71131.39, 126532.90]	0.00
Novembre	455.50	104400.48	[75687.93, 133113.03]	0.00
Décembre	247.50	101059.48	[72956.76, 129162.19]	30000.00

Nous remarquons que les données se rapprochent des valeurs des intervalles de confiance prédits pendant la période allant de janvier à juillet. Cependant, nous remarquons que pour le mois de décembre, les données réelles sont éloignées des intervalles de confiance prédits. Ce qui peut être expliqué ainsi; le changement (la flambée) des prix de la céréale “blé dur” ne provoque pas une augmentation ou une baisse des quantités importées.

3.4 Le Blé Tendre

Le blé tendre⁶ ou froment (*Triticum aestivum*) est une espèce de plantes monocotylédones de la famille des Poaceae (graminées), sous-famille des Pooideae. Domesticquée au Proche-Orient, il y a environ 6 000 ans, cette plante cultivée (ou cultigène), est actuellement l'espèce de blé la plus cultivée dans le monde (95 %), tant en termes de surface que de tonnage. L'expression « blé tendre » s'oppose à celle de « blé dur » qui désigne la céréale méditerranéenne par excellence, prisée pour la confection de semoules et de pâtes.

⁶Source : <https://fr.wikipedia.org>

3.4.1 Description des données statistiques sur le blé tendre

X_{BT}	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
€	149.80	162.90	171.80	175.30	186.20	207.50

Table 3.4: La variable Prix (ou Cours) mesurée en €

Y_{BT}	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
tonne	0.00	31500.00	58290.00	55920.00	78480.00	108900.00

Table 3.5: La variable Quantité importée (ou Offre) mesurée en Tonne

3.4.2 Inférence sur les deux paramètres du modèle β_0 et β_1

	Valeur estimée	Intervalle de confiance	Coefficient de corrélation
Inférence sur β_0	$\widehat{\beta}_0 = -217693.00$	$\beta_0 \in [-236570.940, -198814.981]$	$R = 0.9360$
Inférence sur β_1	$\widehat{\beta}_1 = 1561.00$	$\beta_1 \in [1453.434, 1667.977]$	

Table 3.6: Inférence sur les deux paramètres du modèle β_0 et β_1 cas du blé tendre

L'équation de régression linéaire s'écrit sous la forme

$$Y_{BT} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{BT} = -217693.00 + 1561.00 X_{BT}$$

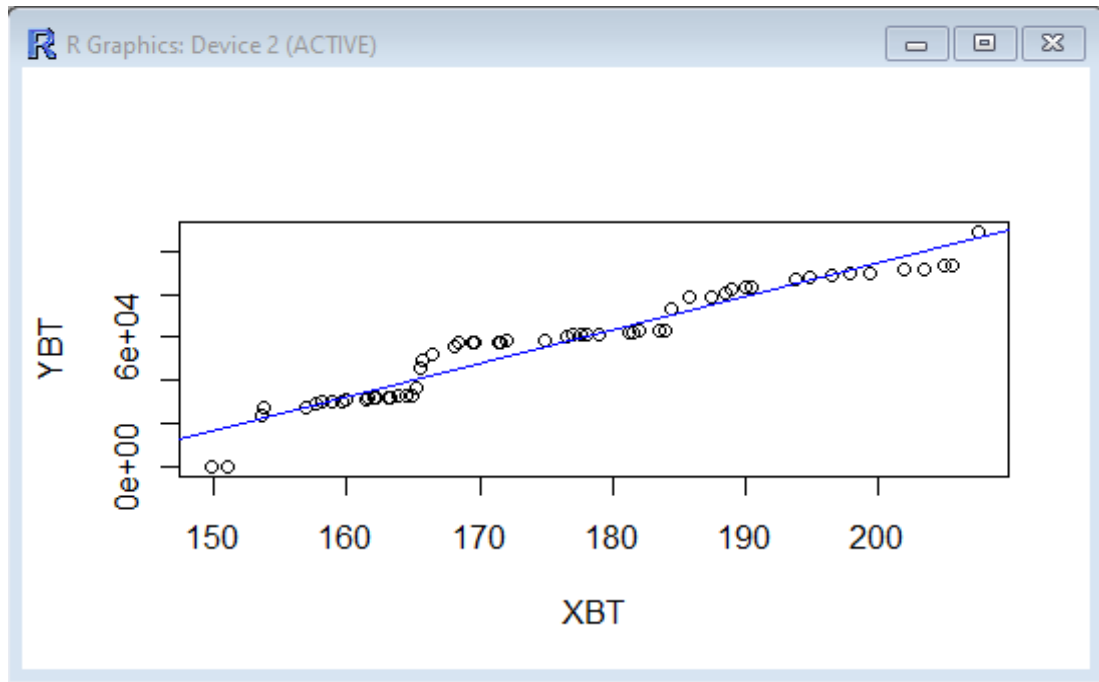


Figure 3.2: Graphe de Y_{BT} en fonction de X_{BT} et la droite de régression

Tests statistiques

- **Test sur l'ordonnée à l'origine β_0**

La réalisation de T_{β_0} est $|t_{\beta_0}| = \frac{|-217692.96|}{9430.89} = 23.08$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique $t_{(\alpha/2, n-2=58)} = 2.002$

Nous remarquons que $|t_{\beta_0}| = 23.08 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejettons l'hypothèse $H_0 : \beta_0 = 0$.

- **Test sur la pente β_1**

La réalisation de T_{β_1} est $|t_{\beta_1}| = \frac{|1560.71|}{53.59} = 29.12$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique $t_{(\alpha/2, n-2=58)} = 2.002$

Nous remarquons que $|t_{\beta_1}| = 29.12 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejettons l'hypothèse $H_0 : \beta_1 = 0$.

- **Test global de significativité de la régression**

Il apparaît raisonnable de tester la significativité du modèle bâti, c'est-à-dire tester l'hypothèse $H_0 : \beta_1 = 0$ contre l'hypothèse $H_1 : \beta_1 \neq 0$, excepter la constante à l'origine β_0 . Ce test est basé sur la statistique de Fisher (à 1 et 58 degrés de

liberté) $F_c = \frac{SC_{reg}/1}{SC_{res}/(n-2)} = 848.2$. La sortie R nous donne la valeur critique $p-value < 2.2e-16$ qui est inférieure au seuil de signification $\alpha = 0.05$, donc nous rejettons ainsi l'hypothèse $H_0 : \beta_1 = 0$ au seuil de signification $\alpha = 0.05$.

Etude sur les résidus du modèle bâti

Min	1er Quantile	Median	Moyenne	3ème Quantile	Max	Variance
-17973.60	-3721.90	113.60	0.00	4196.20	12306.40	6362.00

Table 3.7: Description statistique sur les résidus ε_i

- **Test sur la normalité des résidus ε_i**

Ici nous utilisons le test de normalité de Shapiro-Wilk. La sortie R nous donne la valeur de la statistique de Shapiro-Wilk $W = 0.9771$ et la valeur critique $p-value = 0.318$. Puisque $p-value = 0.318 > \alpha = 0.05$ alors nous ne rejetons pas l'hypothèse de normalité des résidus.

3.4.3 Interprétation des résultats obtenus et discussion

Dans le cas de la céréale "blé tendre", le coefficient de corrélation entre les deux variables offre et prix est $R=0,9360$.

Cette valeur indique ainsi une relation linéaire positive existante très forte, soit à 93,60%. Le test global de Fisher indique que le modèle linéaire bâti est significatif. De plus, le test de Shapiro-Wilk affirme la normalité des résidus du modèle.

3.4.4 Prédiction et comparaison avec des données réelles

Mois	Prix (X_{BTRel} €)	Quant Préd (tonnes)	Inter de conf	Quant import (tonnes)
Janvier	223.50	131124.70	[117283.23 , 144966.20]	59940.00
Février	225.50	134246.10	[120323.10 , 148169.10]	30820.00
Mars	219.00	124101.50	[110432.99 , 137770.00]	94352.00
Avril	218.50	123321.20	[109670.93 , 136971.40]	88500.00
Mai	219.00	124101.50	[110432.99 , 137770.00]	63871.00
Juin	209.50	109274.80	[95921.31 , 122628.30]	80406.00
Juillet	208.00	106933.80	[93623.54 , 120244.00]	0.00
Août	243.50	162338.80	[147561.52 , 177116.10]	63000.00
Septembre	248.50	170142.30	[155092.35 , 185192.30]	55500.00
Octobre	271.00	205258.20	[188819.70 , 221696.70]	93800.00
Novembre	292.50	238033.00	[220101.37 , 255964.70]	16359.00
Décembre	283.00	223986.70	[206715.10 , 241258.20]	57800.00

Nous remarquons que les données réelles ne se rapprochent pas des valeurs des intervalles de confiance prédits. Ces résultats de prédiction peuvent être expliqués ainsi ; le changement (la flambée) des prix de la céréale “blé tendre” ne provoque pas une augmentation ou une baisse des quantités importées.

3.5 Le Maïs

Le maïs⁷ (*Zea mays* L., ou *Zea mays* subsp. *mays* (autonyme)), appelé aussi blé d’Inde, est une plante herbacée tropicale annuelle de la famille des Poacées (graminées), largement cultivée comme céréale pour ses grains riches en amidon, mais aussi comme plante fourragère. Le terme désigne aussi le grain de maïs lui-même.

⁷Source : <https://fr.wikipedia.org>

3.5.1 Description des données statistiques sur le maïs

X_M	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
€	145.60	162.40	165.70	166.50	171.30	190.50

Table 3.8: La variable Prix (ou Cours) mesurée en €

Y_M	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
tonne	41750.00	153600.00	196800.00	200500.00	247500.00	388900.00

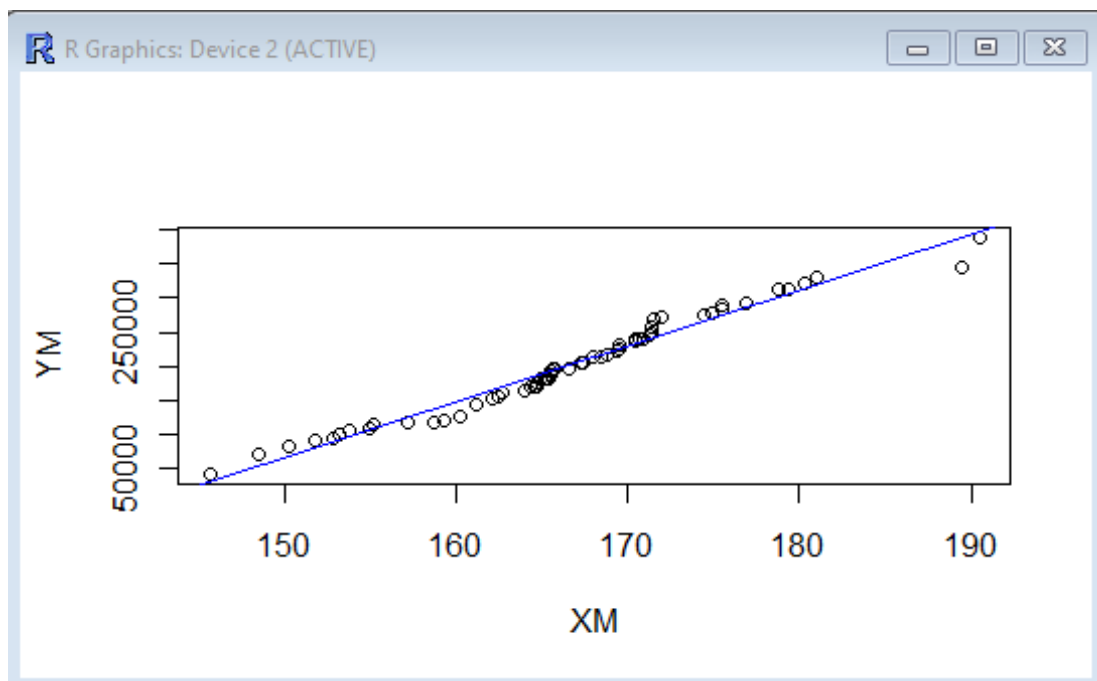
Table 3.9: La variable Quantité importée (ou Offre) mesurée en Tonne

3.5.2 Inférence sur les deux paramètres du modèle β_0 et β_1

	Valeur estimée	Intervalle de confiance	Coefficient de corrélation
Inférence sur β_0	$\hat{\beta}_0 = -1159639.00$	$\beta_0 \in [-1220595.855, -1098683.002]$	$R = 0.9718$
Inférence sur β_1	$\hat{\beta}_1 = 8168.00$	$\beta_1 \in [7802.265, 8533.278]$	

L'équation de régression linéaire s'écrit sous la forme

$$Y_M = \hat{\beta}_0 + \hat{\beta}_1 X_M = -1159639.00 + 8168.00 X_M$$

Figure 3.3: Graphe de Y_M en fonction de X_M et la droite de régression

Tests statistiques

- **Test sur l'ordonnée à l'origine β_0**

La réalisation de T_{β_0} est $|t_{\beta_0}| = \frac{|-1159639.40|}{30452.10} = 38.08$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique

$$t_{(\alpha/2, n-2=58)} = 2.002$$

Nous remarquons que $|t_{\beta_0}| = 38.08 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejetons l'hypothèse $H_0 : \beta_0 = 0$.

- **Test sur la pente β_1**

La réalisation de T_{β_1} est $|t_{\beta_1}| = \frac{|8167.80|}{182.60} = 44.73$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique

$$t_{(\alpha/2, n-2=58)} = 2.002$$

Nous remarquons que $|t_{\beta_1}| = 44.73 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejetons l'hypothèse $H_0 : \beta_1 = 0$.

- **Test global de significativité de la régression**

Il apparaît raisonnable de tester la significativité du modèle bâti, c'est-à-dire tester l'hypothèse $H_0 : \beta_1 = 0$ contre l'hypothèse $H_1 : \beta_1 \neq 0$, excepter la constante à l'origine β_0 . Ce test est basé sur la statistique de Fisher (à 1 et 58 degrés de liberté) $F_c = \frac{SC_{reg}/1}{SC_{res}/(n-2)} = 2001.00$. La sortie R nous donne la valeur critique $p\text{-value} < 2.2e-16$ qui est inférieure au seuil de signification $\alpha = 0.05$, donc nous rejetons ainsi l'hypothèse $H_0 : \beta_1 = 0$ au seuil de signification $\alpha = 0.05$.

Etude sur les résidus du modèle bâti

Min	1er Quantile	Median	Moyenne	3ème Quantile	Max	Variance
-42343.00	-8060.00	1414.00	0.00	8523.00	27850.00	12690

Table 3.10: Description statistique sur les résidus ε_i

- **Test sur la normalité des résidus ε_i**

Ici nous utilisons le test de normalité de Shapiro-Wilk. La sortie R nous donne la valeur de la statistique de Shapiro-Wilk $W = 0.9733$ et la valeur critique $p\text{-value} = 0.2109$. Puisque $p\text{-value} = 0.2109 > \alpha = 0.05$ alors nous ne rejetons pas l'hypothèse de normalité des résidus.

3.5.3 Interprétation des résultats obtenus et discussion

Dans le cas de la céréale "Maïs", le coefficient de corrélation entre les deux variables offre et prix est $R=0,9718$.

Cette valeur indique ainsi une relation linéaire positive existante très forte, soit à 97,18%.

Le test global de Fisher indique que le modèle linéaire bâti est significatif. De plus, le test de Shapiro-Wilk affirme la normalité des résidus du modèle.

3.5.4 Prédiction et comparaison avec des données réelles

Mois	Prix (X_{MRel} €)	Quant Préd (tonnes)	Inter de conf	Quant import (tonnes)
Janvier	205.50	518837.60	[489530.50 , 548144.60]	302295.00
Février	212.00	571928.10	[541395.70 , 602460.50]	218102.00
Mars	211.00	563760.30	[533425.30 , 594095.30]	105366.00
Avril	217.00	612766.90	[581202.50 , 644331.30]	223998.00
Mai	242.50	821045.10	[783268.80, 858821.30]	258606.00
Juin	248.50	870051.70	[830635.40 , 909468.00]	73616.00
Juillet	242.00	816961.20	[779319.10, 854603.30]	163229.00
Août	243.50	829212.90	[791167.10, 867258.60]	258290.00
Septembre	222.00	653605.80	[620939.70 , 686271.90]	138517.00
Octobre	242.50	821045.10	[783268.80, 858821.30]	222488.00
Novembre	248.00	865967.80	[826690.30, 905245.40]	134460.00
Décembre	240.50	804709.60	[767467.30, 841951.80]	179871.00

Nous remarquons que les données réelles ne se rapprochent pas des valeurs des intervalles de confiance prédits. Ces résultats de prédiction peuvent être expliqués ainsi ; le changement (la flambée) des prix de la céréale “Maïs” ne provoque pas une augmentation ou une baisse des quantités importées.

3.6 L'Orge

L'orge⁸ commune (*Hordeum vulgare*) est une céréale à paille, plante herbacée annuelle de la famille des Poaceae, sous-famille des Pooideae. Elle fait partie des plus anciennes céréales cultivées. Bien adaptée au climat méditerranéen du fait de sa rusticité, elle constituait ainsi la principale céréale cultivée dans l'Antiquité grecque et était consommée sous forme de galette ou de bouillie (maza). L'orge pousse aussi bien sous les tropiques qu'à 4 500 m d'altitude au Tibet.

⁸Source :<https://fr.wikipedia.org>

3.6.1 Description des données statistiques sur l'orge

X_{OR}	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
€	102.40	138.30	149.40	149.40	164.10	194.50

Table 3.11: La variable Prix (ou Cours) mesurée en €

Y_{OR}	Min	1er Quantile	Median	Moyenne	3ème Quantile	Max
tonne	0.00	0.00	0.00	9645.00	13860.00	86790.00

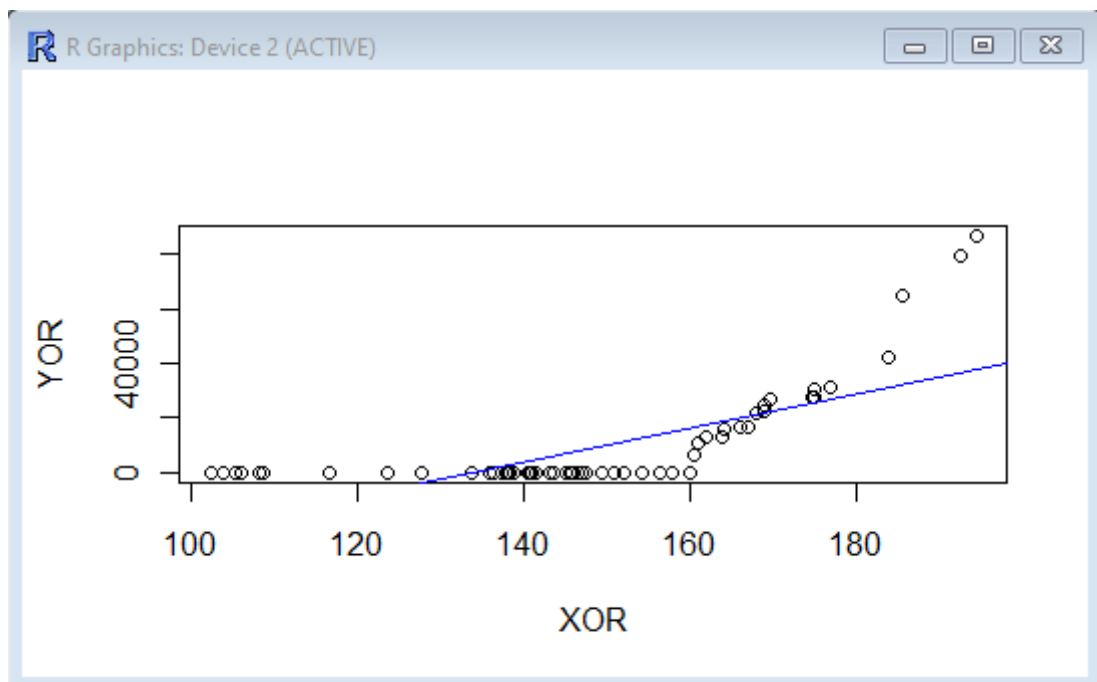
Table 3.12: La variable Quantité importée (ou Offre) mesurée en Tonne

3.6.2 Inférence sur les deux paramètres du modèle β_0 et β_1

	Valeur estimée	Intervalle de confiance	Coefficient de corrélation
Inférence sur le paramètre β_0	$\hat{\beta}_0 = -83784.20$	$\beta_0 \in [-108016.3492, -59552.1174]$	$R = 0.5118$
Inférence sur le paramètre β_1	$\hat{\beta}_1 = 625.30$	$\beta_1 \in [464.8127, 785.8688]$	

L'équation de régression linéaire s'écrit sous la forme

$$Y_{OR} = \hat{\beta}_0 + \hat{\beta}_1 X_{OR} = -83784.20 + 625.30 X_{OR}$$

Figure 3.4: Graphe de Y_{OR} en fonction de X_{OR} et la droite de régression

Tests statistiques

- **Test sur l'ordonnée à l'origine β_0**

La réalisation de T_{β_0} est $|t_{\beta_0}| = \frac{|-83784.20|}{12105.70} = 6.921$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique $t_{(\alpha/2, n-2=58)} = 2.002$

Nous remarquons que $|t_{\beta_0}| = 6.921 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejettons l'hypothèse $H_0 : \beta_0 = 0$.

- **Test sur la pente β_1**

La réalisation de T_{β_1} est $|t_{\beta_1}| = \frac{|625.30|}{80.20} = 7.798$

Au seuil de signification $\alpha = 0.05$, nous lisons sur la table statistique de Student la valeur critique $t_{(\alpha/2, n-2=58)} = 2.002$

Nous remarquons que $|t_{\beta_1}| = 7.798 > t_{(\alpha/2, n-2=58)} = 2.002$. Par conséquent, nous rejettons l'hypothèse $H_0 : \beta_1 = 0$.

- **Test global de significativité de la régression**

Il apparaît raisonnable de tester la significativité du modèle bâti, c'est-à-dire tester l'hypothèse $H_0 : \beta_1 = 0$ contre l'hypothèse $H_1 : \beta_1 \neq 0$, excepter la constante à l'origine β_0 . Ce test est basé sur la statistique de Fisher (à 1 et 58 degrés de liberté) $F_c = \frac{SC_{reg}/1}{SC_{res}/(n-2)} = 60.80$. La sortie R nous donne la valeur critique $p-value = 1.338e - 10$ qui est inférieure au seuil de signification $\alpha = 0.05$, donc nous rejettons ainsi l'hypothèse $H_0 : \beta_1 = 0$ au seuil de signification $\alpha = 0.05$.

Etude sur les résidus du modèle bâti

Min	1er Quantile	Median	Moyenne	3ème Quantile	Max	Variance
-16352.00	-7972.00	-3340.00	0.00	4001.00	48941.00	13390

Table 3.13: Description statistique sur les résidus ε_i

- **Test sur la normalité des résidus ε_i**

Ici nous utilisons le test de normalité de Shapiro-Wilk. La sortie R nous donne la valeur de la statistique de Shapiro-Wilk $W = 0.8439$ et la valeur critique $p-value = 2.023e - 06$. Puisque $p-value = 2.023e - 06 < \alpha = 0.05$ alors nous rejetons l'hypothèse de normalité des résidus.

3.6.3 Interprétation des résultats obtenus et discussion

Dans le cas de la céréale “orge”, le coefficient de corrélation entre les deux variables offre et prix $R=0,562$.

Cette valeur indique ainsi une relation linéaire positive existante, mais faible et probablement anecdotique, c'est-à-dire, les deux variables “offre” et “prix” sont liées à 51,18%.

Le test global de Fisher indique que le modèle linéaire bâti est significatif. Cependant, le test de Shapiro-Wilk nie la normalité des résidus du modèle.

3.6.4 Prédiction et comparaison avec des données réelles

Mois	Prix (X_{ORRel} €)	Quant Préd (tonnes)	Inter de conf	Quant import (tonnes)
Janvier	206.50	45348.63	[16812.23 , 73885.03]	0.00
Février	208.50	46599.31	[17958.18 , 75240.44]	0.00
Mars	200.50	41596.58	[13354.80 , 69838.37]	12546.00
Avril	203.50	43472.61	[15087.22 , 71858.00]	27495.00
Mai	211.00	48162.66	[19386.12, 76939.21]	57979.00
Juin	207.00	45661.30	[17099.02 , 74223.58]	0.00
Juillet	202.50	42847.27	[14510.57, 71183.96]	0.00
Août	228.50	59106.13	[29247.55, 88964.70]	32500.00
Septembre	227.50	58480.78	[28690.11 , 88271.46]	27450.00
Octobre	246.00	70049.59	[38892.51, 101206.67]	0.00
Novembre	264.50	81618.39	[48881.83, 114354.95]	31499.00
Décembre	254.50	75364.98	[43506.84, 107223.13]	0.00

Nous remarquons que les données se rapprochent des valeurs des intervalles de confiance prédits pendant la période allant de janvier à août. Cependant, nous remarquons que pour les mois de septembre et novembre, les données réelles sont éloignées des intervalles de confiance prédits. Ce qui peut être expliqué ainsi; le changement (la flambée) des prix de la céréale “orge” ne provoque pas une augmentation ou une baisse des quantités importées.

Conclusion

Dans ce présent chapitre, nous considérons quatre types de céréales, à savoir : le blé dur, le blé tendre, le Maïs et l'orge. Pour chacun d'entre eux, nous présentons une étude

analytique, une modélisation mathématique qui reflète le cas réel et nous développons des modèles linéaires adéquats qui englobent toute l'information entre les cours et les quantités des céréales importées. Pour ce faire, nous avons utilisé le logiciel R qui est robuste et destiné pour l'analyse des données statistiques. Nous concluons qu'une flambée des prix d'une céréale ne provoque pas une augmentation ou une baisse des quantités importées.

CONCLUSION GÉNÉRALE

Les céréales sont des ressources vitales et stratégiques en Algérie. Leur consommation ne cesse de croître à cause d'une forte dynamique démographique qu'a connue le pays depuis son indépendance. Aujourd'hui, l'essentiel de ces céréales est importé et l'écart entre l'offre et la demande s'accroît de plus en plus. La dépendance alimentaire est extrêmement forte vis-à-vis de l'étranger, notamment pour les céréales. Les statistiques sur la filière céréalière font ressortir des craintes sur l'avenir et les perspectives sont pour le moins, inquiétantes. De plus, les spécificités climatiques, les déficits pluviométriques, les disponibilités en terres agricoles, les caractéristiques et les particularités environnementales de l'Algérie, compliquent la situation. [4]

La production des céréales, jachère comprise, occupe environ 80% de la superficie agricole utile (SAU) du pays. La superficie emblavée annuellement en céréales se situe entre 3 et 3,5 millions d'ha. Les superficies annuellement récoltées représentent 38% des emblavures. Elle apparaît donc comme une spéculation dominante. [5]

Dans notre travail, nous avons fait une étude examinant l'impact de la flambée des prix de certains type de céréales sur la capacité de l'offre, en considérant le cas de l'Entreprise portuaire de Béjaia. Pour ce faire, nous avons construit des modèles mathématiques linéaires par la méthode de régression simple à partir des données financières récoltées sur une période de cinq ans (de janvier 2016 à décembre 2020).

Les Prévisions ont été comparées avec des données réelles collectées sur une année (de janvier à décembre 2021). Nous pouvons conclure que l'Algérie fait toujours recours à l'importation pour combler le déficit en matières céréales malgré la flambée de leurs cours, due aux plusieurs facteurs (politiques, économiques, environnementaux, crises sanitaire, etc), sur les marchés économiques.

Dans les faits, l'Algérie a importé en moyenne plus de 12 millions de tonnes de céréales par an au cours des cinq dernières années, alors que la production annuelle était d'environ 4,92 millions de tonnes, dont 3,3 millions de tonnes de blé. En 2021, les épisodes de sécheresse qui ont frappé l'Afrique du Nord ont affecté la production de céréales.⁹

En Algérie, « la production céréalière totale en 2021 est estimée à 3,5 millions de tonnes, ce qui est inférieur à la moyenne quinquennale et environ 38 % de moins que l'année précédente », rapporte la FAO¹⁰. Toujours selon l'organisation internationale, « le pays importe du blé de France, du Canada, d'Allemagne, des États-Unis d'Amérique, d'Espagne et du Mexique. Pour la première fois depuis 2016, la fédération de Russie a expédié du blé en Algérie en juin 2021 ».

Le pays dispose encore d'importantes potentialités pour développer les cultures céréalières. C'est dans ce contexte que le 16 janvier 2022 la présidence algérienne a annoncé sa décision d'augmenter les prix d'achat de céréales aux producteurs. L'objectif est d'inciter les agriculteurs à fortement augmenter la production et améliorer la sécurité alimentaire du pays.¹¹

⁹Source : <https://www.jeuneafrique.com>

¹⁰Food and Agriculture Organization : Organisation des Nations Unies pour l'alimentation et l'agriculture, spécialisée dans l'aide au développement et chargée d'améliorer le niveau de vie, l'état nutritionnel et la productivité agricole. Intervient notamment en cas de famine, de crise alimentaire et agricole. Source : <https://www.fao.org/>

¹¹Source : <https://www.terre-net.fr>

BIBLIOGRAPHIE

- [1] V. Meuriot A. S. Diallo and M. Terraza. *Analyse d'une nouvelle émergence de l'instabilité des prix des matières premières agricoles*. Tech. rep. Document de Recherche, DR n°2012-01 2012.
- [2] J. M. Azaïs and J. M. Bardet. *Le modèle linéaire par l'exemple : Régression, analyse de la variance et plans d'expérience illustrés avec R et SAS*. Deuxième édition. DUNOD, France, 2021.
- [3] A. Benbelkacem and K. Kellou. “Évaluation du progrès génétique chez quelques variétés de blé dur (*Triticum turgidum* L.var. durum) cultivées en Algérie”. In: *Options Méditerranéennes Série A.40* (2000), pp. 105–110.
- [4] M. Chabane and J. M. Boussard. “La production céréalière en Algérie : Des réalités d'aujourd'hui aux perspectives stratégiques de demain”. PhD thesis. Centre universitaire de recherches politiques, université de Picardie., 2012.
- [5] A. Djermoun. “La production céréalière en Algérie : les principales caractéristiques”. In: *Revue Nature et Technologie Série A.n°01* (Juin 2009), pp. 45–53.
- [6] Y. Dodge. and V. Rousson. *Analyse de régression appliquée*. Deuxième édition. DUNOD, France, 2004.
- [7] *Données unidimensionnelles*. URL: <https://membres-ljk.imag.fr/Bernard.Ycart/smel/cours/sd/node4.html>.
- [8] *Données unidimensionnelles*. URL: <https://mistis.inrialpes.fr/software/SMEL/cours/sd/node3.html>.

- [9] *Données unidimensionnelles*. URL: <https://mistis.inrialpes.fr/software/SMEL/cours/sd/node12.html>.
- [10] F. A. Galton. “Theory of heredity”. In: *Contemporary Review* 27 (1875), 80–95. Traduction française : F. Galton Théorie de l’hérédité. La Revue Scientifique 1876, 17 : 198–205.
- [11] H. Guerin. *Couple de variables aléatoires - Notion d’indépendance*. Tech. rep.
- [12] *Probabilités et variables aléatoires*. 2012. URL: <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-1-inf-probas.pdf#>.
- [13] G. Saporta. *Probabilités, analyse des données et Statistique*. Deuxième édition. Editions TECHNIP, France, 2006.

ANNEXES

- [1] V. Meuriot A. S. Diallo and M. Terraza. *Analyse d'une nouvelle émergence de l'instabilité des prix des matières premières agricoles*. Tech. rep. Document de Recherche, DR n°2012-01 2012.
- [2] J. M. Azaïs and J. M. Bardet. *Le modèle linéaire par l'exemple : Régression, analyse de la variance et plans d'expérience illustrés avec R et SAS*. Deuxième édition. DUNOD, France, 2021.
- [3] A. Benbelkacem and K. Kellou. “Évaluation du progrès génétique chez quelques variétés de blé dur (*Triticum turgidum* L.var. durum) cultivées en Algérie”. In: *Options Méditerranéennes Série A.40* (2000), pp. 105–110.
- [4] M. Chabane and J. M. Boussard. “La production céréalière en Algérie : Des réalités d'aujourd'hui aux perspectives stratégiques de demain”. PhD thesis. Centre universitaire de recherches politiques, université de Picardie., 2012.
- [5] A. Djermoun. “La production céréalière en Algérie : les principales caractéristiques”. In: *Revue Nature et Technologie Série A.n°01* (Juin 2009), pp. 45–53.
- [6] Y. Dodge. and V. Rousson. *Analyse de régression appliquée*. Deuxième édition. DUNOD, France, 2004.
- [7] *Données unidimensionnelles*. URL: <https://membres-ljk.imag.fr/Bernard.Ycart/smel/cours/sd/node4.html>.
- [8] *Données unidimensionnelles*. URL: <https://mistis.inrialpes.fr/software/SMEL/cours/sd/node3.html>.

- [9] *Données unidimensionnelles*. URL: <https://mistis.inrialpes.fr/software/SMEL/cours/sd/node12.html>.
- [10] F. A. Galton. “Theory of heredity”. In: *Contemporary Review* 27 (1875), 80–95. Traduction française : F. Galton Théorie de l’hérédité. La Revue Scientifique 1876, 17 : 198–205.
- [11] H. Guerin. *Couple de variables aléatoires - Notion d’indépendance*. Tech. rep.
- [12] *Probabilités et variables aléatoires*. 2012. URL: <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-1-inf-probas.pdf#>.
- [13] G. Saporta. *Probabilités, analyse des données et Statistique*. Deuxième édition. Editions TECHNIP, France, 2006.

ANNEXE A

NOTIONS DE BASE

Dans cette partie nous aborderons certaines définitions vus dans ce mémoire, nous les aborderons de manières brèves et objectives.[7, 8, 9, 11, 12]

A.1 Espace Probabilisé

Définition A.1.1. Soit Ω un ensemble.

Une famille Λ de parties de Ω (soit un sous-ensemble de $P(\Omega)$) est appelée tribu sur Ω si et seulement si elle vérifie les propriétés suivantes :

$$\Omega \in \Lambda, \tag{A.1}$$

$$\forall A \in \Lambda, \bar{A} \in \Lambda,$$

$$\forall (A_n)_{n \in \mathbb{N}} \in \Lambda^{\mathbb{N}}, \bigcup_{n \in \mathbb{N}} A_n \in \Lambda.$$

A.2 Variable Aléatoire

Définition A.2.1. Une variable X à valeurs dans un ensemble X fini ou dénombrable est appelée variable discrète. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités :

$$P_X(x) = P(X = x), x \in X$$

Ainsi, pour toute partie A de X , on a alors :

$$P_X(A) = P(X \in A) = \sum_{x \in A} P(X = x) \text{ et } P_X(X) = \sum_{x \in X} P(X = x) = 1$$

Soit X une variable qui prend un nombre infini non dénombrable de valeurs. Si F_X est une fonction continue, on dit que X est une variable continue. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités $P(a < X < b)$, pour tout $a < b$.

Si l'on peut écrire la fonction de répartition d'une variable continue sous la forme :

$$F_X(t) = \int_{-\infty}^t f_X(x).dx$$

où f_X est une fonction de \mathbb{R} dans \mathbb{R} , alors on dit que f_X est la densité de probabilité de la variable X .

Ceci implique que l'on a pour tout $a < b$:

$$P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x).dx$$

Cette intégrale étant positive pour tout $a < b$, il en résulte que $f_X \geq 0$. De plus, puisque $\lim_{t \rightarrow +\infty} F_X(t) = 1$, on a :

$$\int_{-\infty}^{+\infty} f_X(x).dx = 1.$$

A.3 Couple de Variable Aléatoire

on va considérer le cas de vecteurs aléatoires c'est à dire des variables aléatoires à valeurs dans \mathbb{P}^d où $d \in \mathbb{N}^*$. Pour simplifier, nous considérerons seulement le cas de la dimension $d = 2$. Un vecteur aléatoire Z de \mathbb{R}^2 sera décrit dans la suite par son abscisse X et son ordonnée Y i.e. $Z = (X, Y)$. On utilise aussi le terme « couple aléatoire » pour un vecteur aléatoire de dimension 2. On appelle couple aléatoire – ou vecteur aléatoire de dimension deux – toute paire $Z = (X, Y)$ de variables aléatoires réelles. On a donc :

$$Z : (\Omega, \mathbb{F}, \mathbb{P}) \rightarrow \mathbb{R}^2, \omega \mapsto Z(\omega) = (X(\omega), Y(\omega)).$$

Sur \mathbb{R} , nous avons considéré la plus petite tribu engendrée par les intervalles i.e. la tribu borélienne. Pour travailler dans \mathbb{R}^2 , on introduit la tribu borélienne de \mathbb{R}^2 , $B(\mathbb{R}^2)$, qui

est « la plus petite tribu » sur \mathbb{R}^2 contenant tous les pavés $I \times J$ où I et J sont deux intervalles de \mathbb{R} . Cette tribu contient tous les pavés du type AB où A et B sont deux boréliens de \mathbb{R} mais elle contient également des ensembles beaucoup plus complexes qui ne sont pas des produits cartésiens. Si Z est un couple aléatoire, on montre que pour tout borélien C de \mathbb{R}^2 , l'image réciproque de C par Z , $Z^{-1}(C) = \{\omega \in \Omega, Z(\omega) \in C\}$, est un élément de \mathcal{F} .

A.4 Espérance Mathématique

Définition A.4.1. Soit X une variable et h une application de \mathbb{R} dans \mathbb{R} . Donc $h(X)$ est elle aussi une variable aléatoire.

Si X est discrète à valeurs dans un ensemble X , l'espérance de $h(X)$ est la quantité.

$$E(h(X)) = \sum_{x \in X} h(x)P(X = x),$$

pourvu que cette série converge (dans le cas où X est infini. Si X est continue et admettant une densité f_X , l'espérance de $h(X)$ est la quantité.

$$E(h(X)) = \int_{-\infty}^{+\infty} h(x)f_X(x).dx,$$

à condition que cette intégrale soit convergente.

Notons que si $h(x) = x$, on obtient $E(X)$ appelée espérance mathématique (ou moyenne) de la variable X . Par ailleurs, si l'on définit la variable suivante :

$$1_{\{X \in A\}} = \begin{cases} 1 & \text{si } X \in A (A \subset \mathbb{R}) \\ 0 & \text{sinon} \end{cases}$$

qui est appelée fonction caractéristique de l'événement $X \in A$, alors l'espérance de cette variable est :

$$E(1_{X \in A}) = P(X \in A) = P_X(A)$$

d'où le lien étroit entre probabilité et espérance.

Propriétés

1. L'espérance est linéaire : pour tout $\alpha, \beta \in \mathbb{R}$, et pour toutes variables X et Y

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

2. Si X est une variable constante égale à $a \in \mathbb{R}$, c'est-à-dire pour tout $\omega \in \Omega$, $X(\omega) = a$, alors $P(X = a) = 1$ et $E(X) = a$.
3. . L'espérance d'une variable positive est positive. En particulier, si $X \geq Y$ (ce qui signifie que pour tout $\omega \in \Omega$, $X(\omega) \geq Y(\omega)$), alors $E(X) \geq E(Y)$.

L'espérance d'une variable X est un indicateur de "localisation" de sa loi :

$$E(X) \simeq \text{"valeur moyenne de } X\text{"}$$

Néanmoins, la connaissance de l'espérance mathématique donne peu de renseignements sur cette variable ainsi, il faut étudier "l'étalement" de sa loi, c'est-à-dire la dispersion de la variable X autour de sa moyenne $E(X)$.

A.5 Variance et Ecart-type

Pour rendre positifs les écarts entre X et son espérance $E(X)$, un autre outil plus facile à manipuler que la valeur absolue, est à notre disposition : la mise au carré. On ne va donc pas calculer la moyenne des écarts mais la moyenne des écarts au carré. C'est ce qu'on appelle la variance.

La variance de la variable X est la quantité :

$$\text{Var}(X) = E[(X - E(X))^2].$$

Propriétés

- $\text{Var}(X) = E(X^2) - (E(X))^2$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$ pour tout $a, b \in \mathbb{R}$. En particulier, $\text{Var}(X + b) = \text{Var}(X)$.

Afin d'être en mesure de comparer, en termes d'ordre de grandeur, variance et espérance, il faut prendre la racine carrée de la variance.

C'est ce qu'on appelle l'écart-type.

La racine carrée de $Var(X)$, notée σ_X , est appelée écart-type de X .

Remarque.

- Si X est une variable telle que $E(X) = \mu$ et $Var(X) = \sigma^2$, alors la variable $Y = (X - \mu)/\sigma$ est d'espérance nulle et de variance 1. On dit que Y est centrée (d'espérance nulle) et réduite (de variance 1).
- Le moment d'ordre k est défini par :

$$m_k = \int_{-\infty}^{+\infty} (x - \mu)^k f(x) . dx.$$

$\gamma_1 = \frac{m_3}{\sigma^3}$ est le coefficient d'asymétrie (Skewness).

$\gamma_2 = \frac{m_4}{\sigma^4} - 3$ est le coefficient d'aplatissement (Kurtosis).

A.6 Covariance

La covariance permet d'estimer la dépendance entre deux variables aléatoires.

La covariance de deux variables X et Y est :

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

L'espérance $E(XY)$ est calculée à partir de la loi jointe de (X, Y) :

1. Dans le cas discret

$$E(XY) = \sum_{x \in D_X, y \in D_Y} xy P(X = x, Y = y)$$

2. Dans le cas continu

$$E(XY) = \int \int_{\mathbb{R}^2} xy f_{(X,Y)}(x, y) . dx dy.$$

Remarque Soient X et Y deux variables aléatoire Alors :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

A.7 Echantillon

On appelle échantillon une suite de variables aléatoires indépendante, de même loi de même moyenne et de même variance. On écrit : X_1, X_2, \dots, X_n l'échantillon de taille n issu de la variable aléatoire X de la fonction de répartition $F(x)$ de densité de probabilité $f(x)$ de moyenne m et de variance σ^2 . La réalisation d'un échantillon est la suite (x_1, x_2, \dots, x_n) .

A.8 Moyenne Empirique

La statistique la plus évidente à calculer sur un échantillon numérique, celle dont l'interprétation est la plus intuitive, est la moyenne empirique.

La moyenne empirique d'un échantillon est la somme de ses éléments divisée par leur nombre. Si l'échantillon est noté (x_1, \dots, x_n) , sa moyenne empirique est :

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n).$$

La moyenne est donc le centre de gravité des données, affectées de coefficients égaux pour chaque individu. Elle peut être considérée comme une valeur centrale, même si elle n'est pas égale à une des modalités. Tout comme le centre de gravité, la moyenne est associative. Si on réunit deux échantillons, de tailles respectives n_x et n_y , de moyennes respectives \bar{x} et \bar{y} , alors la moyenne du nouvel échantillon sera $(n_x\bar{x} + n_y\bar{y})/(n_x + n_y)$.

Si (x_1, \dots, x_n) est un échantillon et si on pose pour tout $i = 1, \dots, n$, $y_i = ax_i + b$, où a et b sont deux constantes, alors la moyenne empirique de l'échantillon (y_1, \dots, y_n) est $\bar{y} = a\bar{x} + b$. En particulier, si $a = 1$ et $b = -\bar{x}$, le nouvel échantillon a une moyenne nulle. Centrer les données c'est leur retrancher la moyenne empirique de manière à la ramener à 0.

Un cas particulier important est celui des données binaires. On est souvent amené à construire un échantillon binaire à partir d'un échantillon numérique, ne serait-ce que

pour le regroupement en classes. Si (x_1, \dots, x_n) est un échantillon à valeurs réelles et si A est un sous ensemble de \mathbb{R} (un intervalle dans le cas d'une classe), notons $1_A(x)$ sa fonction indicatrice qui vaut 1 si $x \in A$, 0 sinon. Alors $(1_A(x_1), \dots, 1_A(x_n))$ est un échantillon binaire, dont la moyenne empirique est appelée la fréquence empirique de A . C'est simplement la proportion des valeurs de (x_1, \dots, x_n) qui appartiennent à A .

Un des inconvénients de la moyenne empirique, vue comme valeur centrale d'un échantillon, est d'être sensible aux valeurs extrêmes. Une valeur manifestement très différente des autres est souvent qualifiée de valeur aberrante. Qu'elle soit ou non le résultat d'une erreur dans le recueil ou la transcription, on ne peut pas la considérer comme représentative. Supposons que sur un échantillon de 10 valeurs, toutes soient de l'ordre de 10, sauf une, qui est de l'ordre de 1000. La moyenne empirique sera de l'ordre de 100, c'est-à-dire très éloignée de la plupart des valeurs de l'échantillon. Pour palier cet inconvénient, on peut décider ne pas tenir compte des valeurs extrêmes dans le calcul de la moyenne. On obtient alors une moyenne élaguée.

A.9 Variance Empirique

Les notions de variance et d'écart-type servent à quantifier la dispersion d'un échantillon autour de sa moyenne. La définition est la suivante : Soit (x_1, \dots, x_n) un échantillon, et \bar{x} sa moyenne empirique. On appelle variance de l'échantillon, la quantité, notée s^2 , définie par :

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

On appelle écart-type de l'échantillon la racine carrée de la variance.

L'avantage de l'écart-type sur la variance est qu'il s'exprime, comme la moyenne, dans la même unité que les données. On utilise parfois le coefficient de variation, qui est le rapport de l'écart-type sur la moyenne.

Pour mesurer la dispersion d'un échantillon autour de sa moyenne, on pourrait trouver plus naturelle une autre mesure d'écart, par exemple l'écart absolu moyen dont nous parlerons plus loin. La raison pour laquelle la définition ci-dessus est la plus naturelle tient dans la proposition suivante.

A.10 Covariance

Si deux caractères quantitatifs x et y sont mesurés sur n individus, on peut considérer l'échantillon bidimensionnel comme un nuage de n points dans \mathbb{R}^2 . Différentes caractéristiques statistiques permettent de résumer l'information contenue dans sa forme. Si \bar{x} et \bar{y} désignent les moyennes empiriques des deux caractères, le point (\bar{x}, \bar{y}) est le centre de gravité du nuage. Les variances empiriques s_x^2 et s_y^2 traduisent la dispersion des abscisses et des ordonnées. Pour aller plus loin dans la description, il faut calculer la covariance.

On appelle covariance de x et y , et on note c_{xy} la quantité :

$$c_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Cette définition étend celle de la variance, dans la mesure où $s_x^2 = c_{xx}$. La covariance est symétrique ($c_{xy} = c_{yx}$) et bilinéaire : si x et x' sont deux échantillons de taille n , a et a' deux réels et $x^* = ax + a'x'$, alors :

$$c_{x^*y} = ac_{xy} + a'c_{x'y}.$$

Comme conséquence de la bilinéarité, on a :

$$s_{x+y}^2 = s_x^2 + s_y^2 + 2c_{xy}.$$

Pour le calcul pratique, on utilise la formule suivante :

La covariance de x et y s'écrit :

$$c_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}.$$

La covariance est la moyenne des produits moins le produit des moyennes.

A.11 Loi de Probabilité

La loi de probabilité d'une variable aléatoire permet de connaître les chances d'apparition des différentes valeurs de cette variable.

On se place sur l'espace de probabilité (Ω, \mathbb{P}) .

Soit X une variable aléatoire. La loi de probabilité de X est définie par la fonction F_X , appelée fonction de répartition de la variable X , définie par :

$$F_X : \mathbb{R} \rightarrow [0, 1]$$

$$x \mapsto \mathbb{P}(X \leq x).$$

On dit que deux variables aléatoires X et Y ont la même loi si elles ont la même fonction de répartition $F_X = F_Y$.

A.11.1 Loi Normale

Soient $\mu \in \mathbb{R}$ et $\sigma > 0$. On dit que X suit une loi normale de paramètres (μ, σ^2) , notée $N(\mu, \sigma^2)$, si la loi de X a pour densité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

la densité de la loi normale présente un axe de symétrie vertical pour $x = \mu$; il n'existe pas d'expression analytique de la fonction de répartition de X qui est approchée par le calcul numérique de l'intégrale de la densité. La loi normale s'applique à de nombreux phénomènes, en physique, en économie (erreurs de mesure), biologie ; c'est une conséquence du théorème central limite, elle est la forme limite de nombreuses lois discrètes.

Ainsi, toute grandeur résultat d'un ensemble ou d'une "somme" de plusieurs variables indépendantes et de même loi se distribue approximativement suivant une loi normale. Il faut cependant remarquer que les variables utilisées dans les domaines technologique, économique, biologique sont bien souvent positives.

Pour que la loi normale puisse être représentative d'un tel phénomène, il faut que la probabilité théorique d'obtenir des valeurs négatives de la variable soit très faible. Il faut en particulier éviter d'utiliser cette modélisation pour les queues des distributions.

A.11.2 Loi du χ^2 , Student et de Fisher

Par définition, la variable aléatoire, somme des carrés de variables indépendantes $N(0, 1)$ suit une loi du χ^2 à v degrés de liberté. Deux autres lois jouent des rôles importants en statistique.

La loi de Student à n degrés de liberté, $T(n)$ est la loi du rapport $X/(\sqrt{Y/n})$, où les

variables aléatoires X et Y sont indépendantes, X de loi $N(0, 1)$, Y de loi $\chi^2(n)$. Elle a pour densité :

$$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Elle décrit la distribution de la moyenne empirique d'un échantillon gaussien.

La loi de Fisher de paramètres n et m (entiers positifs), est la loi du rapport $(X/n)/(Y/m)$, où X et Y sont deux variables aléatoires indépendantes, de lois respectives $\chi^2(n)$ et $\chi^2(m)$. Elle caractérise la distribution de rapports de variances et est très présente en théorie des tests (analyse de variance et modèle linéaire). L'expression de sa densité est définie par un rapports de fonctions $\Gamma(x)$.

Résumé

L'Algérie subit depuis près d'un siècle une forte dynamique démographique qui a rompu divers équilibres économiques, sociaux et naturels et a induit une haute pression sur les ressources agricoles disponibles ; Les céréales constituent un produit aussi vital que stratégique en Algérie où le mode alimentaire est basé essentiellement sur leur consommation. Les crises liées directement ou indirectement à l'offre du blé et des céréales, en général et les crises économique en particulier, donnent lieu à une instabilité des prix. Le problème majeur est que ces instabilités donnent lieu à une balance économique plus lourde sur les pays consommateurs qui sont les principaux acheteurs sur le marché mondial et pour chercher un équilibre économique, ces états doivent prévoir une montée des cours de ces matières premières et c'est sur cette dernière que ce présent travail se portera. Une analyse de la relation Offre \sim Prix en matières céréales : cas de l'Entreprise Portuaire de Béjaïa à l'aide de la Régression Linéaire Simple, a été faite. Cette étude examinera l'impact des cours des céréales sur les quantités importées.

Mots clés : Offre, Prix (Cours), Céréales, Crise Economique, Régression Linéaire Simple.

Abstract

Algeria has been subjected for almost a century to a strong demographic dynamic which has broken various economic, social and natural balances and has induced high pressure on available agricultural resources ; Cereals are a product that is as vital as it is strategic in Algeria, where the diet is essentially based on their consumption. Crises linked directly or indirectly to the supply of wheat and cereals, in general and economic crises in particular, give rise to price instability. The major problem is that these instabilities give rise to a heavier economic balance on the consumer countries which are the main buyers on the global market and to seek an economic balance, these states must foresee a rise in the prices of these raw materials and it is on the latter that this work will focus. An analysis of the relation Supply Price in cereal materials : case of the Port Company of Béjaïa using Simple Linear Regression, was made. This study will examine the impact of cereal prices on the quantities imported.

Keywords : Supply, Price, Cereals, Economic Crisis, Simple Linear Regression.