République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option :**Statistique**

Par

Bourhefir Massilya

Titre :

Statistique descriptive des données climatiques

Membres du Comité d'Examen :

Dr.	**CHERFAOUI Mouloud**	U.M.K.B.	Président
Dr.	**Abdelli Jihane**	U.M.K.B.	Encadreur
Dr.	**OUANOUGHI Yasmina**	U.M.K.B.	Examinateur

Dédicace

Avant de dire pour qui je dédie ce modeste travail, je glorifié ALLAH le tout puissant de m'avoir donnée courage et patience qui n'ont permis d'accomplir ce travial. Je dédie ce modeste travail.

À mes parents pour leur soutient, leur confiance dans ma vie et pour tous ses sacrifier.

À mes chères sœurs : Ranida, Nour, Fadoi,

À mon cher frère walid,

À mon mari houssam,

À tous mes amis, Samiha, Mariem, Bouthiana, Abir et à tous ce que j'ai connus.

REMERCIEMEN

J'espère que ce modeste travail, sera une source important pour les étudiants en future à exigé. Inversement, je ne dois en aucune manière oublier de remercier tous les personnes qui m'aident dant ce travail.

J'exprime ma plus grand reconnaissance et mes plus remerciements à mon encadreur madame ABDDELI Jihane pour sa patience, disponibilité et conseil, elle a toujeurs sa me guider et me motiver. Sa confiance en me capacité et sa présence inconditionnelle était pour moi une grande source de motivation.

Nous tenons à remercier tous les membres de jury pour l'acceptation de mon travail en acceptant leur jugement.

Mes remerciements vos aussi à tous les enseignants de département de mathématique.

Enfin, je n'oublié pas de remercier mes amis collègues ; grâce à qui ma vie universitaire à été très plaisant et joyeuse.

Table des matières

Table des figures

Introduction

L'objectif de la Statistique Descriptive est de décrire de façon synthétique et parlante des données observées pour mieux les analyser. Le terme« statistique » est issu du latin «statisticum», c'est-à-dire qui a trait à l'État. Ce terme a été utilisé, semble-t-il pour la première fois,à l'époque de Colbert, par Claude Bouchu, intendant de Bourgogne, dans une «Déclaration des biens, charges, dettes et statistiques des communautés de la généralité de Bourgogne de 1666 à 1669 ».

Par contre, l'apparition du besoin « statistique » de posséder des données chiffrées et précises,précède sa dénomination de plusieurs millénaires. À son origine, il est le fait de chefs d'États(ou de ce qui en tient lieu à l'époque) désireux de connaître des éléments de leur puissance : population, potentiel militaire, richesse, . . .

L'objectif de ce travail est d'étudier, comparer et contraster deux régions différentes de l'Algérie en termes de température et vent.

Dans ce memoire, nous avons examiné les statistiques descriptives des deux types et certaines des lois de probabilité, et les avons appliquées dans les données climatiques.
Ce travail se partage en en troix chapitres :

Le premier chapitre est consacré à une introduction générale sur les branche de la statistique qui est la statistique descriptive univariée et la statistique descriptive bivariée.

Dans le deuxième chapitre, nous mentionnerons quelques lois de probabilités discrèts et continues.

Enfin dans le troixièmme chapitre, on va appliquer tout ce qu'on a vu dans les deux chapitres sur les données climatiques(vent, températeur)pour comparer deux régions différenttes à l'Algérie.

Chapitre 1

Sur la théorie de statistique descriptive

La statistique est l'étude de la collecte de données, leur analyse, leur traitement et l'interprétation des résultats graphique avec des méthodes de classement tels que les tableaux statistique et les graphiques.

Dans ce chapitre, on s'intéresse d'introduire les notions de base de la statistique descriptive (l'effectif,la fréquence, la fréquence cumulée...etc).

1.1 Définitions fondamentales

1.1.1 La statistique descriptive

Le principe de la statistique descriptive est la description des données étudiées à l'aide de moyens appropriés, qu'ils correspondent à des valeurs calculées (moyenne, médiane, quartiles...), ou à des représentations graphiques (Histogramme, le diagramme en bâtons,...) et comme un exemple, l'étude de la distrubution des salaires dans une entreprise. L'objectif de la statistique descriptive est de résumer l'échantillon par deux moyens : L'approche numérique et graphique.

Elle se compose de deux domaines distincts :

La statistique descriptive univarées : Correspond à l'analyse d'un seul caractère, c'est l'étude de la population selon une seule variable (la taille, le poids...).

La statistique descriptive multivariée : Est l'étude de la relation qui peut exister entre deux ou plusieurs variables, que l'on traite avec des méthode comme l'analyse factorielle, par exemple la relation entre la taille et le poids.

1.1.2 Vocabulaires statistiques

- Population : L'ensemble des unités concernés par l'etude statistique.
- Individu (unité statistique) : Élément de la population étudiée.
- Echantillon : L'échantillon est un sous ensemble de la population statistique.
- Taille de l'échantillon : Cardinal du sous-ensemble correspondant.
- Enquête (statistique) : Opération consistant à obeserver l'ensemble des individus d'un échantillon.
- Données : L'ensemble des valeurs de variables mesurérs sur les individus de la population.
- Les modalités : Les modalités sont les différentes situations x_i possible du caractère (ou les valeurs possibles de X), où chaque caractère possède deux ou plusieurs modalités.

Dans la section suivante, on va expliquer la notion d'un caractère statistique.

1.1.3 Un caractère(ou une variable)

Définition 1.1 *Une variable X est un moyen de décrire chacun des indididus de la populaton étudiée, par exemple : àge, sexe, taille,...ect. Elle a deux types :*

Une variable quantitative : Une variable est dite quantitative, si l'ensemble des observation est un ensemble des nombres numériques qu'ils peuvent être ordonnés. On distingue deux types de variable quantitative :

– Une variable quantitative discrète : Elle ne prend que des valeurs dans $\mathbb{N}$. Elle représente par un nombre fini de valeurs, par exemple le nombre d'enfants par famille.

– Une variable quantitative continue : Elle est dite continue, lorsque ces modalité ne sont pas des valeurs précises, mais des intervalles $[a, b]$ de nombre réels. par exemple le poids,la taille,...etc.

Une variable qualitative : Une varible est dite qualitative, lorsque les modalités d'une variable sont des catégories que l'on désigne par noms, qu'elles ne peuvent pas s'exprimer par des nombres. Par exemple, la couleur de peau est une variable a pour modalité : blanc, noir,...etc. Et elle a deux catégories :

– Une variable qualitative ordinale : Elle est dite ordinale quand les modalités peuvent être naturellement ordonnées, par exemple : niveau d'études, grade,...etc.
– Une variable qualitative nominale : Elle est dite nominale lorsque ses modalités ne peuvent être classées de façon naturelle par exemple : la variable couleurs des yeux et la variable sexe,...etc.

1.1.4 Effectifs, fréquence, fréquence cumulées

Effectif

Définition 1.2 *L'effectif d'une modalité x_i est le nombre de fois, où cette modalité apparait dans la série statistique, on le note n_i. On a :*

$$n_i = card\left\{\omega \in \Omega,\ X(\omega) = x_i\right\}.$$

Effectif total

Définition 1.3 *L'effectif total est la somme des effectifs de chaque valeur, on le note n :*

$$n = \sum_{i=1}^{p} n_i = card(\omega).$$

Fréquence(fréquence relative)

Définition 1.4 *La fréquence d'une modalité est l'effectif divisé par le nombre d'unités* d'observation.

$$f_i = \frac{n_i}{n}$$

Remarque 1.1 *La valeur de la fréquence est toujours comprise entre* 0 *et* 1*.*

Effectif et fréquence cumulés :

Définition 1.5 *L'effectif cumulé croissant (la fréquence cumulée croissante) d'une modalité de rang i est la somme de tous les n_i (de toutes les f_i) jusqu'au rang i compris. Ils sont notés*

respectivement ECC et FCC :

$$ECC = \sum_{j=1}^{i} n_j \quad , \quad FCC = \sum_{j=1}^{i} f_j$$

Définition 1.6 *L'effectif cumulé décroissant(la fréquence cumulée décroissante) d'une modalité de ranq i et la somme de tous les n_i (de toutes les f_j), à partir de la dernière valeur jusqu'au rang i compris. Ils sont notés respectivement ECD et FCD :*

$$ECD = \sum_{j=i}^{p} n_j \qquad , \quad FCD = \sum_{j=i}^{p} f_j$$

où p représente le nombre de classes.

Nous pouvons représenter les donnéess statistiques avec des tableaux et des séries statistiques et en plus du graphique.

1.2 Représentation des données

Il existe plusieurs méthodes de la description statistique : des représentations par des tableau statistiques numériques, des représentations graphiques, celui nous permet de visualiser l'information rapidement, et avoir une vue plus globale du phénomène étudié.

1.2.1 Série statistique

Définition 1.7 *On appele série statistique la suite des valeurs prises par une variable X sur les unités d'observetion.*

Le nombre d'unités d'observation est noté n.

Les valeurs de la variable X sont notées :

$$x_1,\ \dots,\ x_i, \dots,\ x_n.$$

1.2.2 Tableau statistique

Définition 1.8 *un tableau statistique est un moyen d'organiser et ranger par ordre croissant (ou décroisent) les données brutes de la série statistique, pour les bien représenté.*

1.3 Représentation graphique

En générale, la représentation graphique des données relatives à un caractère unique est une synthèse de l'information qui fai apparaitre la forme globale de la distribution des données. La nature de graphique dépent du type de variable.

1.3.1 Cas d'une variable quantitative

Pour les valeurs quantitatives, il existe deux types de représentation graphique qui sont :

Les diagrammes différentiels

Cas d'une variable quantitative discrète :

Le diagramme en bâtons : Ce diagramme comporte deux axes, un axe horizontal qui représente les valeurs de la variable, et un axe vertical qui représente les effectifs ou les fréquences. A chaque valeur on associe un segment (bâton) dont sa hauteur est proportionnelle à l'effectif ou à la fréquence de cette modalité.

Cas d'une variable quantitative continue :

L'histogramme : Un histogramme est un graphique permettant de représenter les séries continues dont les valeurs du caractère étudié ont été regroupées en « classes ». Chaque classe est alors représentée par un rectangle dont la base est proportionnelle à son amplitude et de hauteur telle que l'aire du rectangle soit proportionnelle à son effectif(ou fréquence).

Les diagrammes cumulatifs :

Les diagramme cumulatifs permattent de visualiser l'évolution des fréquences cumulées ou les effectifs cumulés croissants ou décroissants. Ils sont obtenus à partir de la fonction de répartition empirique.

1.3.2 Cas d'une variable qualitative

Lorsque le caractère est qualitatif, on utilise le tableau de fréquence pour construire les graphiques, qui permettent de représenter la série statistique.

Diagramme en barres :

Le diagramme en barres est un ensemble de rectangles de même largeur, séparés par un espace, qu'ils sont placés sur un axe horizontal (cette droite n'orient pas car les modalités ne sont pas numériquement et elles n'ont pas de relation d'ordre), chaque rectangle représentant une modalité, sa hauteur est proportionnelle à l'effectif ou à la fréquence de la modalité.

1.3.3 Statistique descriptive univarée

On entend par statistique univariée l'étude d'une seule variable, que celle-ci soit quantitative ou qualitative. La statistique univariée fait partie de la statistique descriptive.

Paramètres caractéristiques

L'objectif d'une étude statistique est aussi de résumer et visualiser les données par des paramètre ou des indicateurs caractéristiques, qu'ils sont séparés par trois types : les paramètre de position, les paramètres de dispèrtion et les paramètre de forme.

Paramètres de position les paramètre de position donnent une idée sur la position des données, ils permettent à indiquer autour de quelle valeur central se situent ces données.

Mode

Définition 1.9 *Le mode noté M_o, (ou la classe modale) est la valeur la plus fréquente de la serie. Il peut y en avoir plusieurs.*

Moyenne

Définition 1.10 *La moyenne, noté $\bar{X}$, d'une serie statistique est le quotient de la somme de toutes les valeurs de cette série par son effectif total.*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i$$

La médiane

Définition 1.11 *La médiane une serie statistique ordonnée est la valeur qui partage la serie statistique en deux partiés égales.*

Quantiles

Définition 1.12 *Les quantiles sont des caractéristiques de position partageant la série statistique ordonnée en quatre parties égaux. Les quartiles Q_1 et Q_3 qui sont les quantiles d'ordre 0,25 et 0,75 (Q_2 est la médiane).*

Paramètres de dispersion Ces paramètres permettent de mesurer la variable (la dispertion) des données, autour d'une valeur centrale, et de trouver un indicateur de cette dispersion.

Étendue

Définition 1.13 *L'étendue est la différence entre la plus petite et la plus grande des* valeurs du caractère, elle est notée E :

$$E = x_{\max} - x_{\min}$$

Variance

Définition 1.14 *la variance notée $V(X)$ ou encore $Var(X)$ d'une série statistique est*

la moyenne des carrés des écarts des valeurs du caractère x à la moyenne de la série statistique :

$$Var(X) = \frac{1}{N} \sum_{j=1}^{p} n_j \left(x_j - \overline{x} \right)$$

x_j étant remplacé par c_j dans le cas d'une variable continue, N est l'effectif total de la série statistique et p le nombre de classes de la série statistique.

La distance interquartile La distance interquartile est la différence entre le troisième et le premier quartile :

$$IQ = x_{3/4} - x_{1/4}.$$

Écart-type

Définition 1.15 *l'écart-type, noté souvent d'une série statistique est la racine carrée de la variance de cette série statistique.*

$$\sigma = \sqrt{var\left(x\right)}$$

Paramètres de forme Les paramètre de forme permettent de décrire la forme de la distribution statistique par la symétrique et l'aplatissement. On les définit que pour les variables quantitatives.

• **Coefficient d'asymétrie de Fisher :** Le moment centré d'ordre trois est défini par :

$$m_3 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3$$

Il peut prendre des valeurs positives, négatives ou nulles. L'asymétrie se mesure au moyen du coefficient d'asymétrie de Fisher :

$$g_1 = \frac{m_3}{s_x^3}$$

où S_x^3 est le cube de l'écart-type.

• **Coefficient d'asymétrie de Pearson :** Le coefficient d'asymétrie de Pearson est basé sur une comparaison de la
moyenne et du mode, et est standardisé par l'écart-type :

$$A_p = \frac{\bar{x} - x_M}{S_x}$$

• **Coefficient d'asymétrie de Yule :** Le coefficient d'asymétrie de Yule est basé sur les positions des 3 quartiles
(1^{er} quartile, médiane et troisième quartile), et est normalisé par la distance interquartile

$$A_y = \frac{x_{3/4} + x_{1/4} - 2x_{1/2}}{x_{3/4} - x_{1/4}}$$

1.3.4 Statistique descriptive bivariée

La statistique bivariée est l'analyse descriptive des variables deux à deux.

L'objectif de cette étude statistique est d'étudier sur même population de N individu, deux caractères différents (ou modalités différentes), et de rechercher s'il existe un lien ou corrélation entre ces deux variables.

Série statistique bivariée

Définition 1.16 *On appelle série statistique à deux variable x et y, toute liste de couples $(x_1; y_1), (x_2; y_2), ..., (x_n; y_n)$.*

n est le nombre de ces couples. On distingue deux cas, où les deux variables sont quantitatives ou les deux variables sont qualitatives.

Deux variables quantitatives Dans ce cas, chaque couple est composé de deux valeurs numériques. Un couple de nombres peut toujours être représenté comme un point dans un plan $(x_1, y_1), ..., (x_i, y_i), ..., (x_n, y_n)$.

Analyse des variables Les variables x et y peuvent être analysées séparément. On peut calculer tous les paramètres dont les moyennes et les variance :

$$\begin{aligned} \overline{x} &= \tfrac{1}{n} \sum_{i=1}^{n} x_i, \quad s_x^2 = \tfrac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2, \\ \overline{y} &= \tfrac{1}{n} \sum_{i=1}^{n} y_i, \quad s_y^2 = \tfrac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2. \end{aligned}$$

Ces paramètres sont appelés paramètres marginaux.

Covariance la covariance est un indicateur qui mesure la liaison entre deux varibles X et Y.

Définition 1.17 *on appelle la covariance, la moyenne des produits des écarts de X et Y à leur moyenne.*

$$cov(x, y) = S_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \overline{xy}.$$

Remarque 1.2 *La covariance peut prendre des valeurs positives, négatives ou nulles. Quand $x_i = y_i$, pour tout $i = 1, ..., n$, la covariance est égale à la variance.*

Corrélation Le coefficient de corrélation est la covariance divisée par les deux écart-types :

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Le coefficient de détermination est le carré du coefficient de corrélation :

$$r_{xy}^2 = \frac{s_{xy}^2}{s_x^2 s_y^2}$$

Remarque 1.3 – *Le coefficient de corrélation mesure la dépendance linéaire entre deux variables.*

- $-1 \leq r_{xy} \leq 1$.
- $-1 \leq r_{xy}^2 \leq 1$.
- $r = 0$ *si la corrélation linéaire observée est nulle.*
- $r = \pm 1$ *si la corrélation linéaire observée est parfaite. Les points sont alignés.*
- *Le coefficient est positif si la liaison est positive.*
- *Le coefficient est négative si la liaison est négative.*

Droite de régression La droite de régression est la droite qui ajuste au mieux un nuage de points au sence des moindres carrés. L'équation de la droite est : $y = ax + b$, en utilisant la droite de régression pour prédire y_i à partir de x_i.

Exemple 1.1 *La droite de régression :*

Résidus et valeurs ajustées

Résidus

Définition 1.18 *Le résidu e_i est l'erreur que l'on commet. Si les coefficients a et b étaient connuson, pourrait calculer les résidus de la régression définis par :*

$$e_i = y_i - a - bx_i$$

Les résidus peuvent être positifs ou négatifs.

Remarque 1.4 *Les résidus sont les différences entre les valeurs observées et les valeurs ajustées de la variable dépendante*

$$e_i = y_i - y_i^*$$

Valeurs ajustées

Définition 1.19 *Les valeurs ajustées sont obtenues au moyen de la droite de régression :*

$$y_i^* = a + bx_i$$

Sommes de carrés et variances

Définition 1.20 *On appelle somme des carrés totale la quantité :*

$$SC_{TOT} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

La variance marginale peut alors être définie par :

$$S_y^2 = \frac{SC_{TOT}}{n} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Définition 1.21 *On appelle somme des carrés de la régression la quantité :*

$$SC_{REGR} = \sum_{i=1}^{n} (y_i^* - \bar{y})^2$$

Et la variance de régression est la variance des valeurs ajustées :

$$S_{y^*}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i^* - \bar{y})^2$$

Définition 1.22 *On appelle somme des carrés des résidus (ou résiduelle) la quantité :*

$$SC_{RES} = \sum_{i=1}^{n} e_i^2$$

Et la variance résiduelle est la variance des résidus :

$$S_e^2 = \frac{SC_{RES}}{n} = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Chapitre 2

Lois de probabilitées

Le but de ce chapitre est de voir si la variable aléatoire suit une certaine loi de probabilité. Les lois de probabilités sont des objets mathématiques qui permettent aux statisticiens de fabriquer des modéles pour décrire des phénomènes où le hasard intervient.
On note par (Ω, A, P) l'espace de probabilité des variables aléatoires. On distingue deux types de lois de probabilité, lois discrètes et lois continues.

2.1 Lois discrètes

2.1.1 Variable aléatoire discrète

On dit que X est une variable aléatoire discréte s'il prend des valeurs discrètes $\{x_1, x_2, x_3, ..., x_n, ...\}$.

Définition 2.1 *Soit X une variable aléatoire discrète, on définit la probabilité de x_i par :*

$$f_X(x_i) = p(X = x_i).$$

Cette dernière vérifie les propriétés suisantes :

1. $\forall x_i \in \mathbb{R} : 0 \leq f_X(x_i) \leq 1.$
2. $\sum_i f_X(x_i) = 1.$

2.1.2 Loi de Bernoulli

C'est la loi de la réussite ou de l'échec, codée 1 et 0.

Définition 2.2 *Soient X une v.a et p un paramètre y compris sur l'intervalle $[0,1]$. On dit que X suit la loi de Bernoulli c'est -à-dire (i.e) $X \rightarrow B(p)$.*

On définit :

$$p[x=0]=q$$

et

$$p[x=1]=p$$

avec

$$p+q=1$$

1. $E(X)=p$.
2. $V(X)=p(1-q)$.

2.1.3 Loi Binomiale

Définition 2.3 *Une variable aléatoire est dite suivre une loi binomiale de paramètres n et p, notée $X \rightarrow B(n,p)$, si elle peut être considérée comme la somme de variables aléatoires de Bernoulli, indépendantes et de même paramètre p.*

$$p(X=k)=C_n^k p^k (1-p)^{n-k} \quad , X \in \{0,1,.......,n\}.$$

Avec $k \in \{0,1,2,...,n\}$, n est un entier donné et p est un réel tel que $0<p<1$.

$C_n^k = \frac{n!}{k!(n-k)!}$ est le coefficient du binôme.

1. $E(X)=np$.
2. $V(x)=np(1-q)$.

avec

$$p + q = 1.$$

Remarque 2.1 *Une variable de Bernoulli est un cas particulier d'une variable binomiale.*

$$X \sim B(1, p)$$

Définition 2.4 Exemple 2.1 *On lance une pièce de monnaie, bien équilibre* 12 *fois. X est la v.a definie par le nombre de face obtenues. Donc $X \in \{0,\ 1,\ 2,\ 3, ..., 12\}$ et $P = \frac{1}{2}$ (bien équilibré).*

$$P(X = k) = C_n^k p^k q^{n-k} = C_{12}^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{12-k} = C_{12}^k \left(\frac{1}{2}\right)^{12}.$$

$$E(X) = np = 12/2 = 6.$$

$$V(X) = npq = 12(1/2)(1/2) = 12/4 = 3.$$

2.1.4 Loi de Poisson

Définition 2.5 *Soit X une variable aléatoir à valeurs dans $\mathbb{N}$. On dit que X suit une loi de Poisson de paramètre λ ($\lambda \in \mathbb{R}^+$), notée $P(\lambda)$, si :*

$$p(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \forall k \in \mathbb{N}.$$

Avec

1. $E(X) = \lambda$.
2. $V(x) = \lambda$.

La loi de poisson est une approximation de la loi binomiale $B(n, p)$ $\overrightarrow{n \to \infty}$ $p(\lambda)$ avec p $\overrightarrow{n \to \infty}$ 0 et np $\overrightarrow{n \to \infty}$ λ.

Remarque 2.2 *Cette loi est souvent utilisée dans la modélisation des files d'attente : nombre de clients en attente à une caisse de supermarché ; nombre de conducteurs passés à un péage pendant une période de temps...*

2.1.5 Loi Géométrique

Définition 2.6 *On dit qu'une variable X suit la loi géométrique $G(p)$ de paramètre p si x est à valeur dans $\mathbb{N}$ et si :*

$$p\left(X=x\right)=p\left(1-p\right)^{x-1},\ x=1,2,...,\infty.$$

1. $V\left(X\right)=\frac{q}{p^2}$.
2. $V\left(X\right)=\frac{q}{p^2}$

Remarque 2.3 *On remarque que X suit la loi géométrique lorsque X compte le nombre d'essais avant le premier succès dans une répétition de tirages de Bernoulli.*

2.1.6 Loi Hypergéométrique

Définition 2.7 *Soient n, N et M trois entiers positifs tels que $n \leq N$, $m < N$. On dit que X suit une loi hypergéométrique de paramètres N, n et m, ce que l'on note $X \rightarrow H\left(N,n,m\right)$ si :*

$$p\left(X=k\right)=\frac{C_m^k C_{N-m}^{n-k}}{C_N^n},\ X\in\left\{0,1,...,n\right\}.$$

X admet alors une espérance et une variance :

$$E(X)=\frac{nm}{N}\ \text{ et } V(x)=\frac{N-n}{N-1}\frac{nm}{N}\left(1-\frac{m}{N}\right).$$

2.2 Lois continues

Cette partie est consacrée aux lois continues ou on va citer quelques une d'entres elles (uniforme, exponentielle,gamma, ...ect).

2.2.1 Variable aléatoire continue

On dit que X est une v.a.r continue, si elle admet une fonction f définie de $\mathbb{R}$ dans $\mathbb{R}$ satisfaisant les conditions suivantes :

1. $f_X(\varkappa) \geq 0$.
2. $\int_{\mathbb{R}} f_X(\varkappa)\, dx = 1$.

2.2.2 Loi Uniforme $U[a, b]$

Définition 2.8 *On dit que X suit une loi uniforme sur l'intervalle fini $[a, b]$, si sa densité est constante sur $[a, b]$ et nulle à l'extérieur de cet intervalle :*

$$f(x) = \begin{cases} \frac{1}{b-a} & si \quad a \leq x \leq, \\ 0 & sinon. \end{cases}$$

Et sa fonction de répartition est :

$$f(x) = \begin{cases} 0 & si \quad x < a, \\ \frac{x-a}{b-a} & si \quad a \leq x \leq b, \\ 0 & si \quad x \succ b. \end{cases}$$

On a

1. $E(X) = \frac{a+b}{2}$.
2. $V(X) = \frac{(b-a)^2}{12}$.

2.2.3 Loi Exponentielle

Définition 2.9 *On dit qu'une variable continue X suit une loi exponentielle de paramètres $\lambda > 0$ et θ lorsque sa densité de probabilité est :*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x > 0, \\ 0 & \quad sinon. \end{cases}$$

On a

1. $E(X) = \frac{1}{\lambda}$.
2. $V(X) = \frac{1}{\lambda^2}$.

2.2.4 Loi de Gamma $\Gamma(r, \lambda)$

Définition 2.10 *Soit X_1, $X_2, ..., X_r$ une suite de r variables aléatoires i.i.d. de loi $\xi(\lambda)$ et soit $T = \sum_{i=1}^{r} X_i$ que T suit une loi de densité :*

$$f(x) = \begin{cases} \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

1. $E(X) = \frac{r}{\lambda}$.
2. $V(X) = \frac{r}{\lambda^2}$

2.2.5 Loi de Cauchy

Définition 2.11 *Une variable aléatoire X suit la loi de Cauchy si elle est absolument continue et admet pour densité de probabilité :*

$$f(x) = \frac{1}{\pi} \times \frac{1}{1 + x^2}.$$

Remarque 2.4 *Les intégrales qui définissent l'espérance et la variance de cette loi ne convergent pas, de sorte qu'une variable aléatoire de Cauchy ne possède ni espérance, ni variance.*

2.2.6 Loi Normale

Définition 2.12 *Soit X une variable aléatoire rélle. X suit une loi normale de paramères (μ, σ), notée $X \to N(\mu, \sigma^2)$, si X admet comme densité :*

$$\forall x,\ \mu \in \mathbb{R},\ \sigma \in R^{+*} :$$
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

1. $E(X) = \mu$.
2. $V(X) = \sigma^2$.

2.2.7 Loi Khi-deux χ^2_ν

Soient Z_1, Z_2,..., Z_ν une suite des v.as indépendantes de même loi $N(0;1)$.
Alors la v.a $Z = \sum\limits_{i=1}^{\nu} Z_i^2$ suit la loi du Khi-deux à ν degrés liberté, notée χ^2_ν. Et la densité sa loi est :

$$f_z(x) = \frac{1}{2^\nu \Gamma(\nu/2)} x^{(\nu/2)-1} \exp(-x/2),\ x \in \left]0, +\infty\right[.$$

On peut remarque que la loi χ^2_ν est la loi $\Gamma(1/2, \nu/2)$, alors

1. $E(Z) = \nu$.
2. $V(Z) = 2\nu$.

2.2.8 Loi Student $St(\nu)$

La loi Student et le quotient entre une v.a suit une loi $N(0;1)$ et la racine carrée d'une v.a suit la loi χ^2_ν.

Définition 2.13 *Soient Z une v.a suit la loi $N(0;1)$ et Q v.a suit la loi χ^2_ν à ν degrés de liberté, telle que Q v.a indépendante de v.a Z, alors*

$$T = \frac{Z}{\sqrt{Q/\nu}}$$

Suit la loi Student à ν degrés de liberté, notée $St(\nu)$.

La fonction de densité sa loi est donnée par.

$$f_T(x) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)} \left(1 + x^2/\nu\right)^{-(\nu+1)/2}.$$

Avec

1. $E(X) = 0$ si $\nu > 1$.
2. $V(X) = \nu/(\nu - 1)$ si $\nu > 2$.

Chapitre 3

Application sur les données climatiques

Le climat a montré des fluctuations entre les années entre les périodes chaudes ou les périodes sèches, froides ou humides, et ces changement sont susceptibles d'être humides associé à l'atmosphère, c-à-d les gaz, que le climat de l'avenir dépendra du développement de ces gaz à l'époque, il est maintenant possible de faire des projections en utilisant la modalisation climatique.

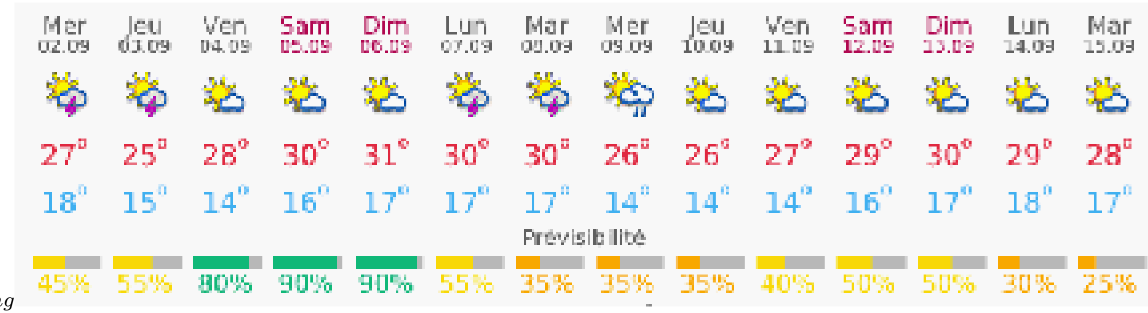
Dans ce chapire, nous examinons les changements climatiques de températeur et le vitesse du vent dans deux régions différentes de l'Algerie(Biskra et Batna).

3.1 Étude discreptive des données climatiques de wilaya de Batna et Biskra

3.1.1 Température de la wilaya de Batna

Ce graphique montre les tendances météo des 14 jours à venir pour Batna (Batna, Algérie) par des symboles météo journaliers, les températures minimales et maximales ainsi que les quantités et les probabilités de précipitations.

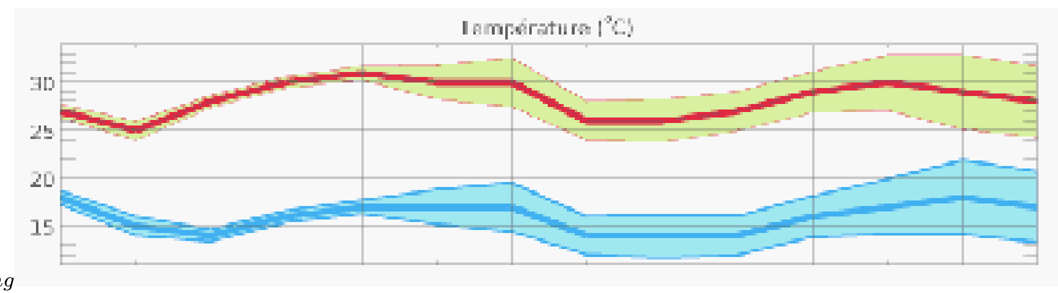
Sur la courbe d'évolution des températures, la variation est présentée par des couleurs. Une batna$_1$4day



$_2.png$

FIG. 3.1 – Température dans la wilaya de Batna du 19/08/2020 aux 01/09/2020

batna$_1$4day



$_3.png$

FIG. 3.2 – Graphe de température dans la wilaya de Batna du 19/08/2020 aux 01/09/2020

plus grande variation signifie une plus grande incertitude dans les prévisions. La ligne en gras montre l'évolution la plus probable.

Pour les précipitations, la variation est présentée sous forme de "T". Après 6-10 jours, des divergences significatives sont visibles.

Les prévisions sont constituées de modèles "ensemble". De ce fait, plusieurs modèles avec différentes variables de départ seront calculés afin d'estimer au mieux l'incertitude des conditions météorologiques.

- Chaque modèle a une couleur attribuée qui est utilisée dans tous les diagrammes. La légende à côté du diagramme liste les noms de modèles et les couleurs correspondantes.
- Le premier diagramme montre les températures prévues par chaque modèle. La durée du jour, du lever au coucher du soleil, est indiquée en jaune clair. La ligne pointillée représente la moyenne de tous les modèles.

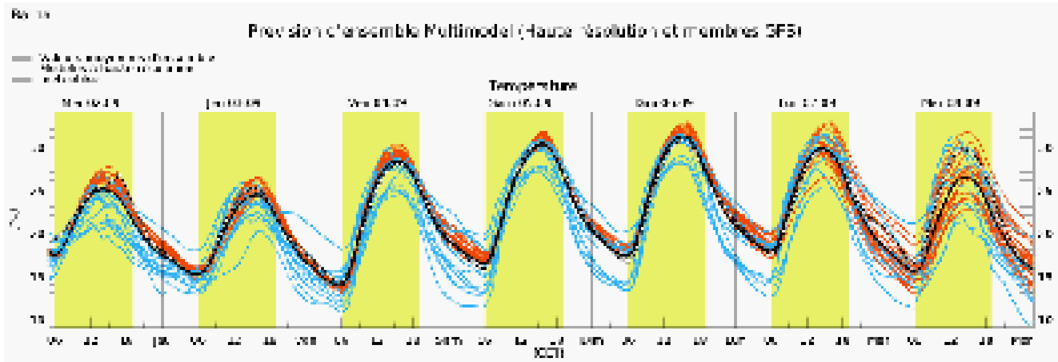


FIG. 3.3 – Comparaison des modèles météorologiques pour la wilaya de Batna

3.1.2 Vitesse du vent de la wilaya de Batna

- Les variables affichées sont des sorties directes du modèle et non post-processées selon l'altitude et la position exacte du lieu choisi.
- La plupart du temps, les prévisions sont de bonne qualité mais parfois elles manquent de précisons et quelques fois sont complètement fausses. Il serait intéressant de connaître à l'avance si la prévision est susceptible d'être bonne, mais comment y parvenir ? Toutes les prévisions météorologiques sont calculées par des modèles numériques, et parfois celles-ci diffèrent de façon significative selon le modèle, ce qui indique l'incertitude de modélisation et la difficulté de faire tout le temps une prévision précise des variables météorologiques.

3.1.3 Température de la wilaya de Biskra

Ce graphique montre les tendances météo des 14 jours à venir pour Biskra (Wilaya de Biskra, Algérie) par des symboles météo journaliers, les températures minimales et maximales ainsi que les quantités et les probabilités de précipitations.

- Sur la courbe d'évolution des températures, la variation est présentée par des couleurs. Une plus grande variation signifie une plus grande incertitude dans les prévisions. La ligne en gras montre l'évolution la plus probable.
- Les prévisions sont constituées de modèles "ensemble". De ce fait, plusieurs modèles avec différentes variables de départ seront calculés afin d'estimer au mieux l'incertitude des conditions météorologiques.

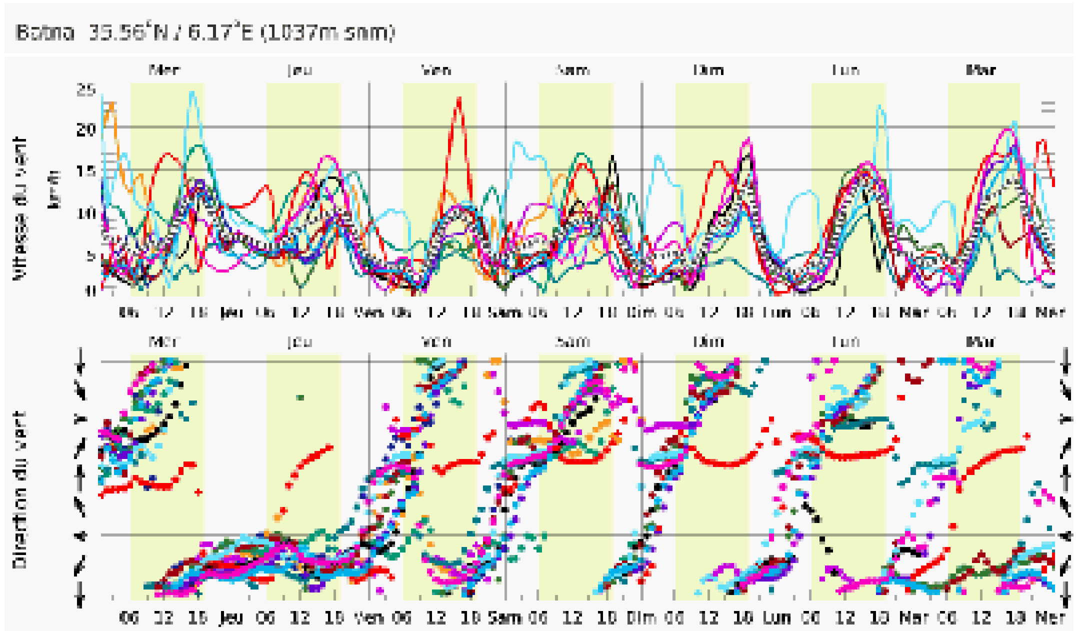


FIG. 3.4 – Comparaison des modèles météorologiques de vitesse du vent pour la wilaya de Batna

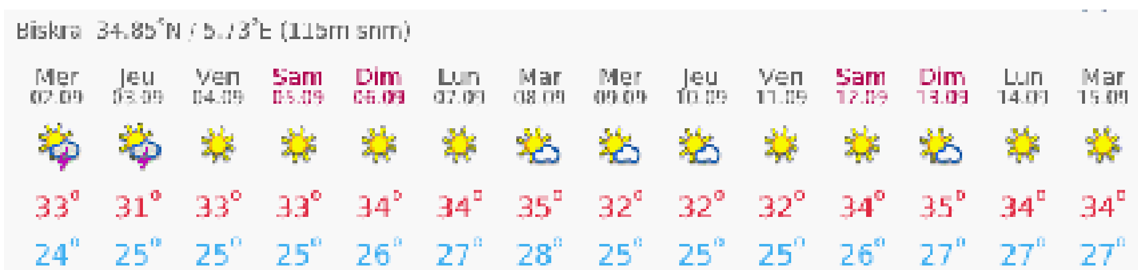


FIG. 3.5 – Température dans la wilaya de Biskra du 19/08/2020 aux 01/09/2020

– Dans le graphique du haut, la température prévue pour Biskra est indiquée en bleu clair pour différents modèles à haute résolution et en rouge pour les membres de l'ensemble GFS. La ligne noire représente la moyenne de toutes les prévisions et la ligne pointillée la prévision consensuelle de meteoblue comme indiqué dans nos prévisions météo.

– Chaque modèle a une couleur attribuée qui est utilisée dans tous les diagrammes. La légende à côté du diagramme liste les noms de modèles et les couleurs correspondantes.

– Le premier diagramme montre les températures prévues par chaque modèle. La durée du jour, du lever au coucher du soleil, est indiquée en jaune clair. La ligne pointillée représente la moyenne de tous les modèles.

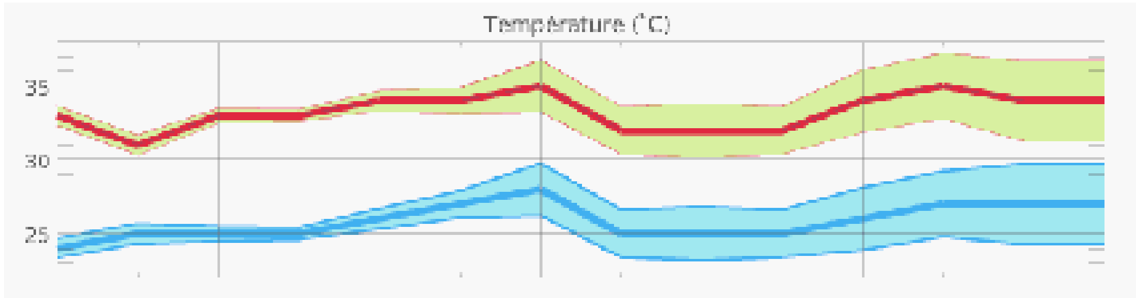


FIG. 3.6 – Température dans la wilaya de Biskra du 19/08/2020 aux 01/09/2020

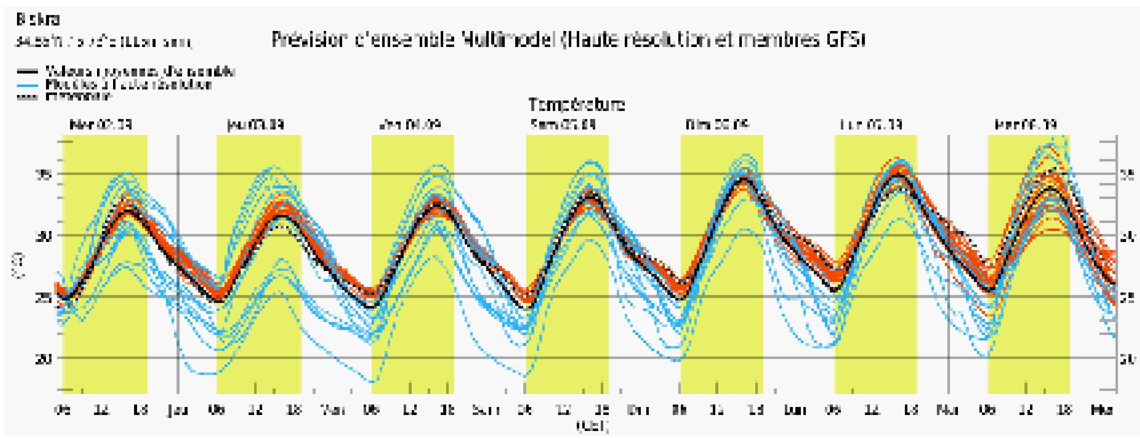


FIG. 3.7 – Comparaison des modèles météorologiques pour la wilaya de Biskra

3.1.4 Vitesse du vent de la wilaya de Biskra

– Les lignes bleues correspondent aux prévisions calculées par différents modèles météorologiques à haute résolution. Sont également représentés les membres d'une prévision d'ensemble traditionnelle, où le même modèle météorologique (GFS) est exécuté plusieurs fois avec des conditions initiales légèrement différentes, pour refléter les incertitudes dans les observations requises pour exécuter un modèle de prévision. Les membres du GFS ont été réduits et les biais ont été corrigés pour correspondre aux conditions météo locales ; les données des modèles à haute résolution ne sont pas modifiées.

– Le graphique indique la prévision du vent calculée par les modèles à haute résolution (bleu clair) et par la prévision d'ensemble (vert). Le résumé quotidien de la direction du vent sous la forme d'une rose des vents est également montré. Les segments plus grands indiquent

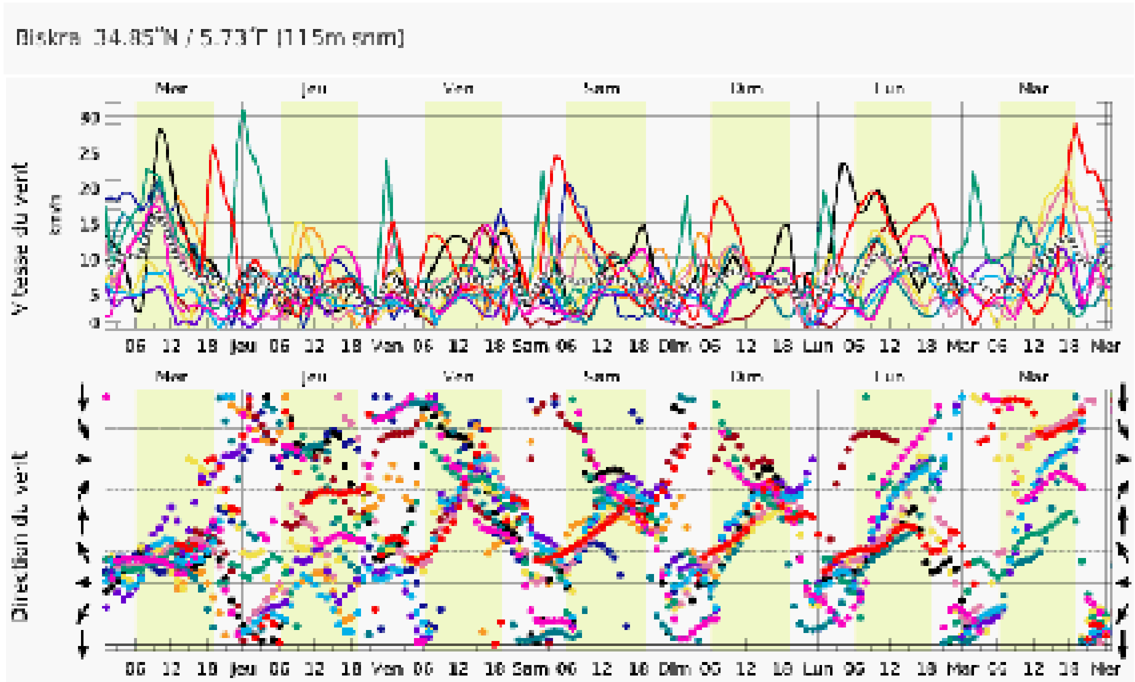


FIG. 3.8 – Comparaison des modèles météorologiques de vitesse du vent pour la wilaya de Biskra

que cette direction du vent est plus probable et plus fréquente au cours de la journée que les directions ayant des segments plus petits. Si vous avez plusieurs segments de taille à peu près égales, cela signifie que la prévision de la direction du vent est très incertain. S'il y a principalement deux directions opposées, cela indique souvent une circulation de vent thermique où le vent souffle d'une direction différente pendant le jour que la nuit.

3.2 représentation des données climatiques :

Un tableau contenant la température pour la wilaya de Biskra et Batna en 14 jours.

/	jours	T.Biskra	T.Batna
1	mercredi	33	27
2	jeudi	31	25
3	vendredi	33	28
4	samedi	33	30
5	dimanche	34	31
6	lundi	34	30
7	mardi	35	30
8	mercredi	32	26
9	jeudi	32	26
10	vendredi	32	27
11	samedi	34	29
12	dimanche	35	30
13	lundi	34	29
14	mardi	34	28

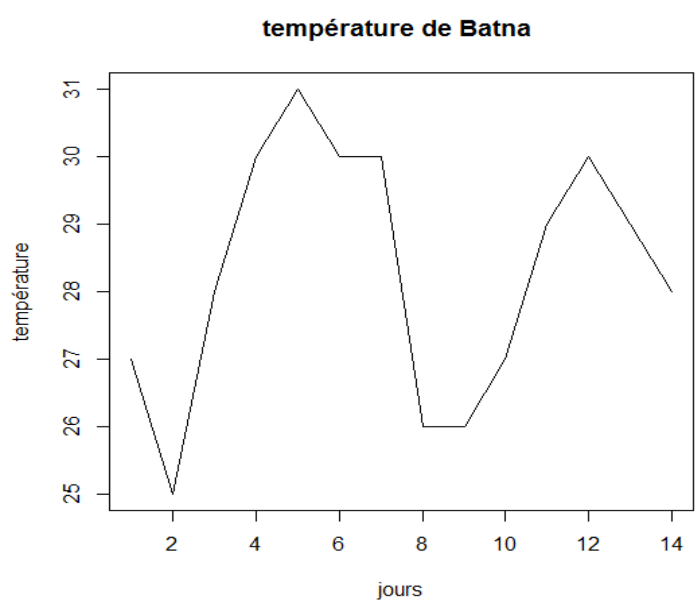


FIG. 3.9 – Courbe des températures de Batna

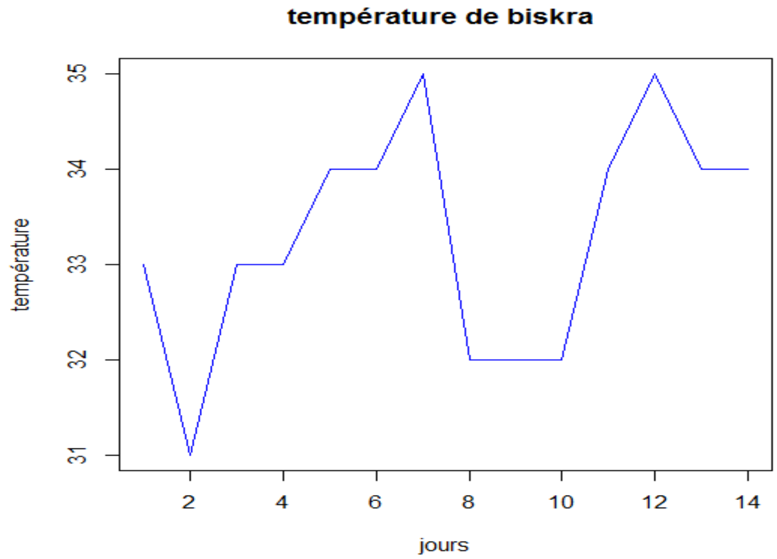


FIG. 3.10 – Graphique représente les changements de température pour Biskra dans les 14 jours

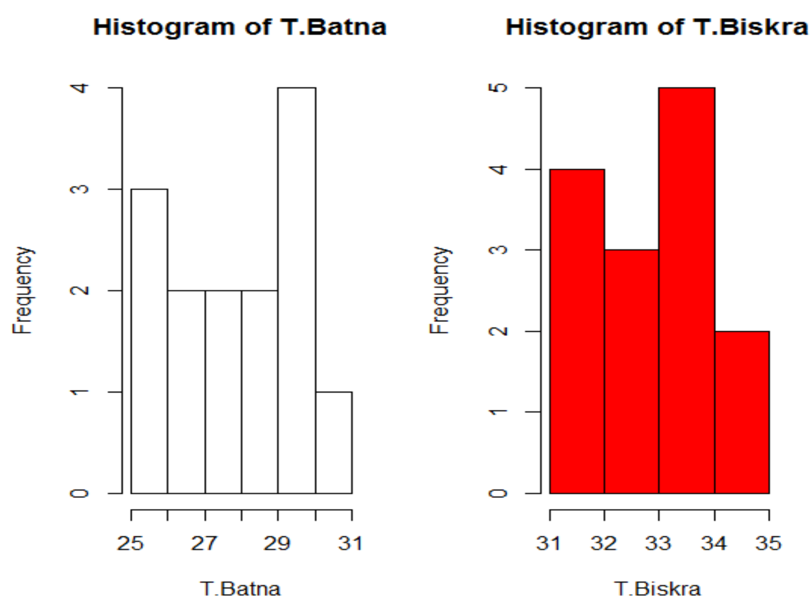


FIG. 3.11 – Histogramme de la fréquence des changements de température pour Biskra et Batna

Conclusion

Dans ce memoire on a étude la statistique descriptive dans les deux cas univariée et bivarée d'ou ona présentés les lois des probabilités les plus connues. Puis, on a fait une application sur logiciel R des données climatique (vent, température), pour la wilaya de Biskra et Batna.

Bibliographie

[1] Antoine Ayache & Julien Hamonier.Cours de Statistique Descriptive.

[2] Berkane Hassiba.(Juin 2016).Analyse statistique descriptive multivariée.Memoire mastère.Université Mohamed kheider.

[3] Camelia Goga et Catherine Labruere.Introduction au logiciel R.Université de Bourgogne.Ecole Doctorale Dijon-2009.

[4] Myriam Maumy.Statistique et Probabilités pour Licence STPI Première Annéee.

[5] Ouanoughi Yasmina.(juin2013).Les lois de probabilités.Memoire mastère.Université Mohamed kheider.

[6] Sayah Abdallah.(2011-2012).Les loi de probabilités.Memoire mastère. Université Mohamed kheider.

[7] Sayah Abdallah.(juin 2018).La statistique descriptive univariée.Memoire mastère. Université Mohamed kheider.

[8] S. Betteka (2018/2019).Statistique Descriptive Univariée.

[9] Shuyan LIU.Année 2013-2014.Notes de cours Statistique avec le logiciel R.

[10] Yves Tillé.15 décembre 2010.Résumé du Cours de Statistique Descriptive.