

Mathématique Projet semestriel

Groupe :

Charles Logeais, Nadejda Dorosenco, Julian Alizay, Duncan Lopes

Table des matières

Exercice 1 : Diagonalisation	2
Exercice 2 : Approximation des données.....	2
Exercice 3 : Etude de cas (1)	8
Modèle additif :	9
Modèle multiplicatif :	10
Exercice 4 : Etude de cas (2)	11
Exercice 5 : ACP	13
Analyse mensuelle.....	13
Analyse annuelle	14

Exercice 1 : Diagonalisation

On peut écrire le problème sous la forme suivante :

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \\ w_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_n \\ v_n \\ w_n \end{pmatrix}$$

A l'aide de la diagonalisation on peut exprimer la suite u_n, v_n, w_n en fonction de l'état initial u_0, v_0, w_0 et de n :

$$X_n = PD^nP^{-1}X_0$$

Avec $X_0 = (u_0, v_0, w_0) = (1, 2, 3)$, $X_n = (u_n, v_n, w_n)$

Et

$$D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad P = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} \quad P^{-1} = \begin{pmatrix} -0.5 & 0 & -0.5 \\ 0.5 & 0 & -0.5 \\ 0 & -1 & 0 \end{pmatrix}$$

Si $n = 3$:

$$\begin{aligned} u_3 &= 55 \\ v_3 &= -2 \\ w_3 &= 53 \end{aligned}$$

Exercice 2 : Approximation des données

Les exemples de génération des points suivants seront faits en considérant les paramètres suivants :

n (nombre de points) : 1 000

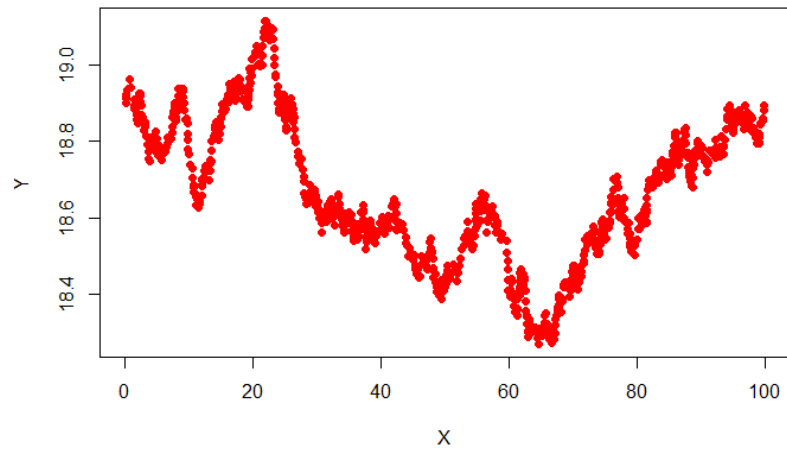
x_min : 0

x_max : 100

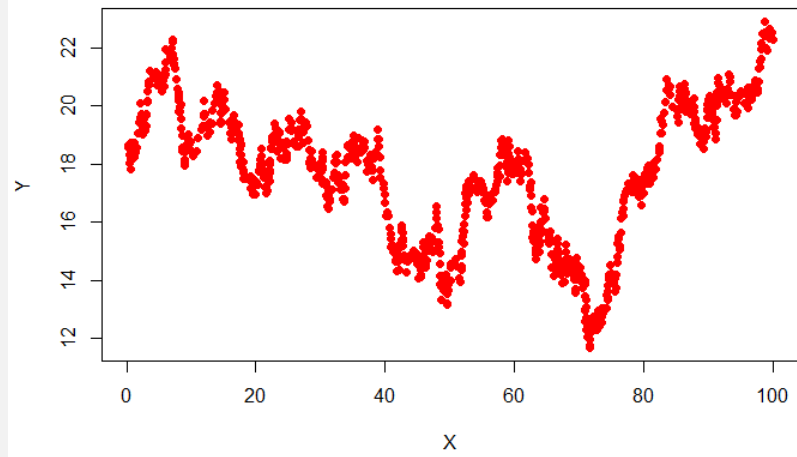
y_min : 0

y_max : 50

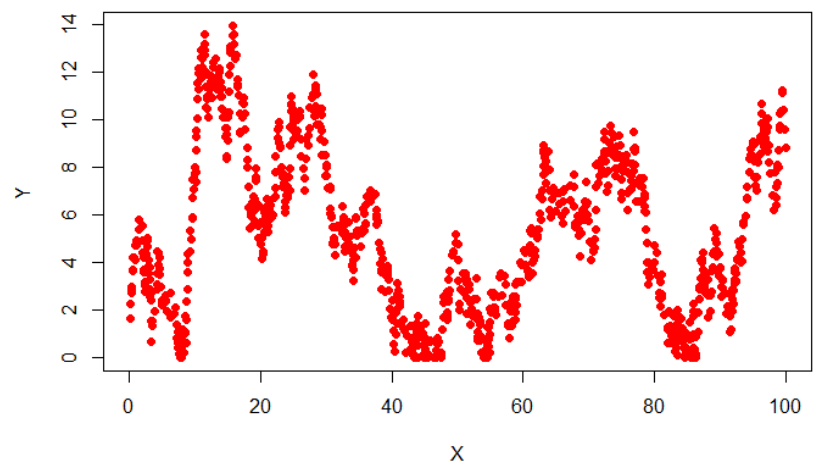
Nuage de point pour
 $\epsilon = 0.03$



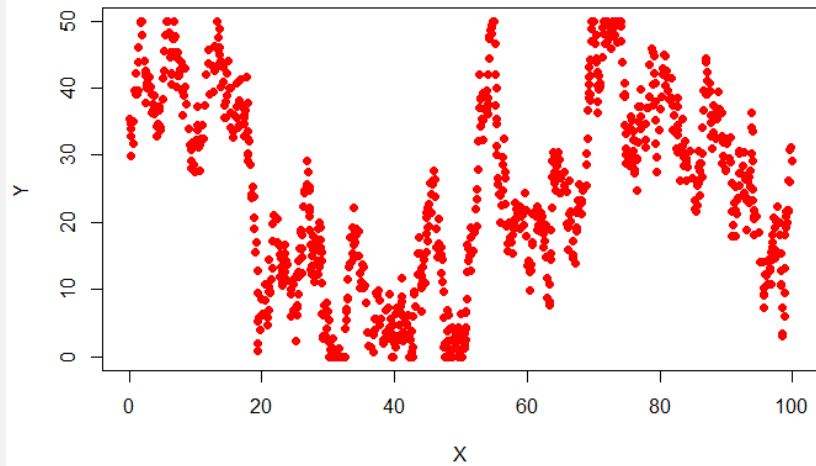
Nuage de point pour
 $\epsilon = 0.5$



Nuage de point pour
 $\epsilon = 1$



Nuage de point pour
 $\epsilon = 5$



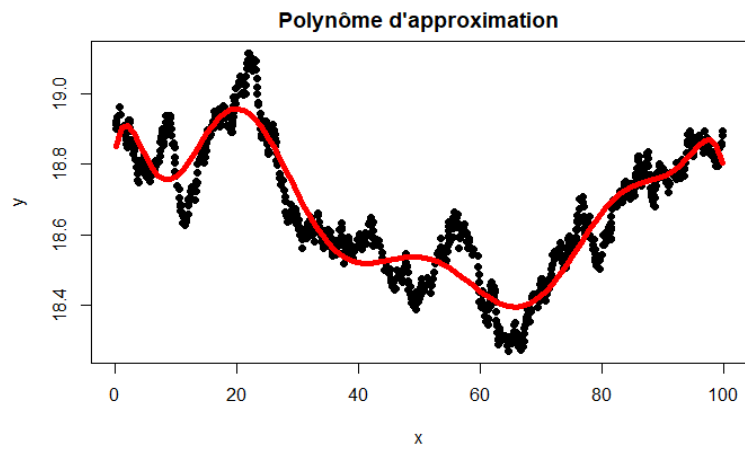
2) On peut choisir le degré du polynôme en regardant le nombre de changement de direction que nous voyons sur le nuage de point. Plutôt que d'une approche visuelle nous décidons d'appliquer un algorithme qui détecte un changement significatif sur l'axe Y.

Ce changement significatif est caractérisé par la formule : sensibilité * ϵ

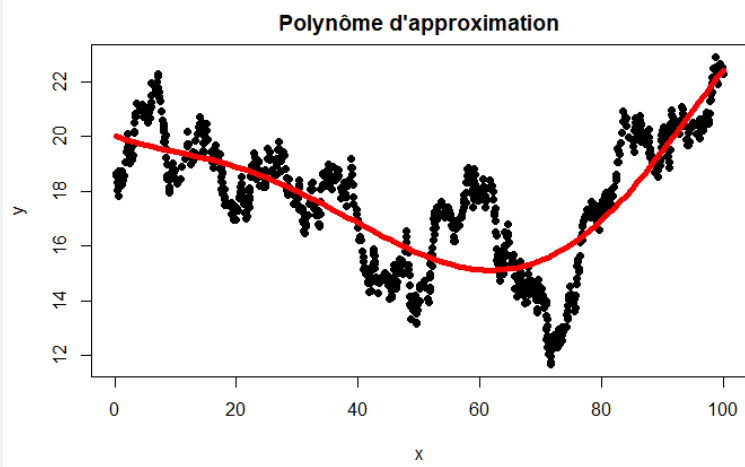
La sensibilité est un paramètre que l'on peut modifier au besoin, par défaut il est mis à 5

3)

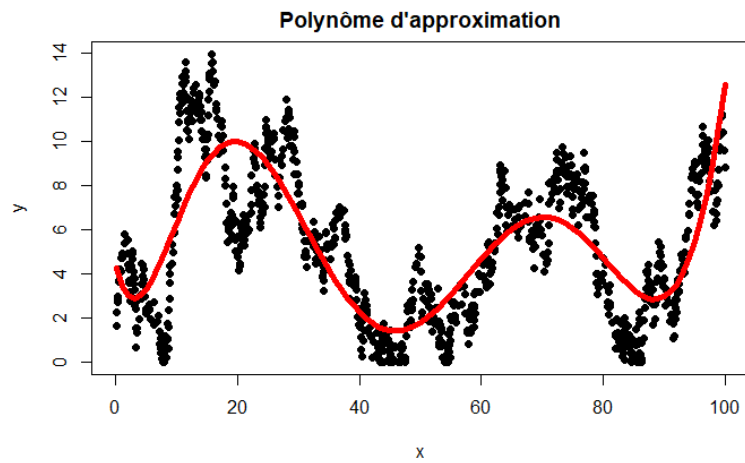
Polynôme d'approximation
pour $\epsilon = 0.03$ | degré = 10



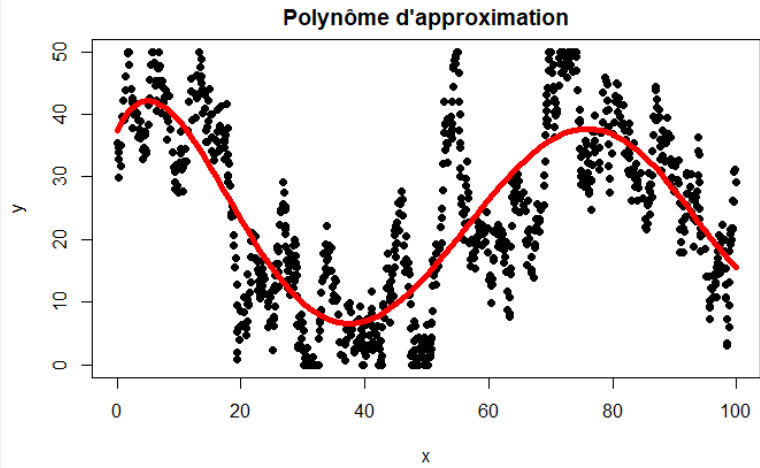
Polynôme d'approximation
pour $\epsilon = 0.5$ | degré = 5



Polynôme d'approximation
pour $\epsilon = 1$ | degré = 7

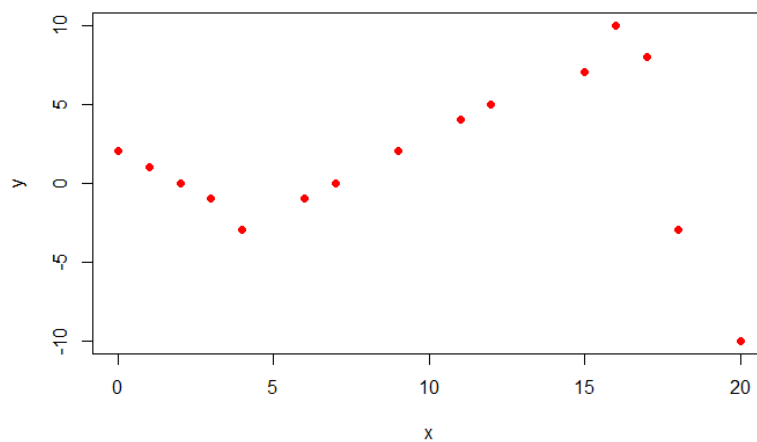


Polynôme d'approximation
pour $\epsilon = 5$ | degré = 5



4)

Nuage de point :



On peut se rendre compte que $\epsilon = 11$

Comme l'Epsilon est haut, nous allons augmenter la sensibilité de la détection pour le polynôme.
Nous détectons un polynôme de rang 3

Ses paramètres sont:

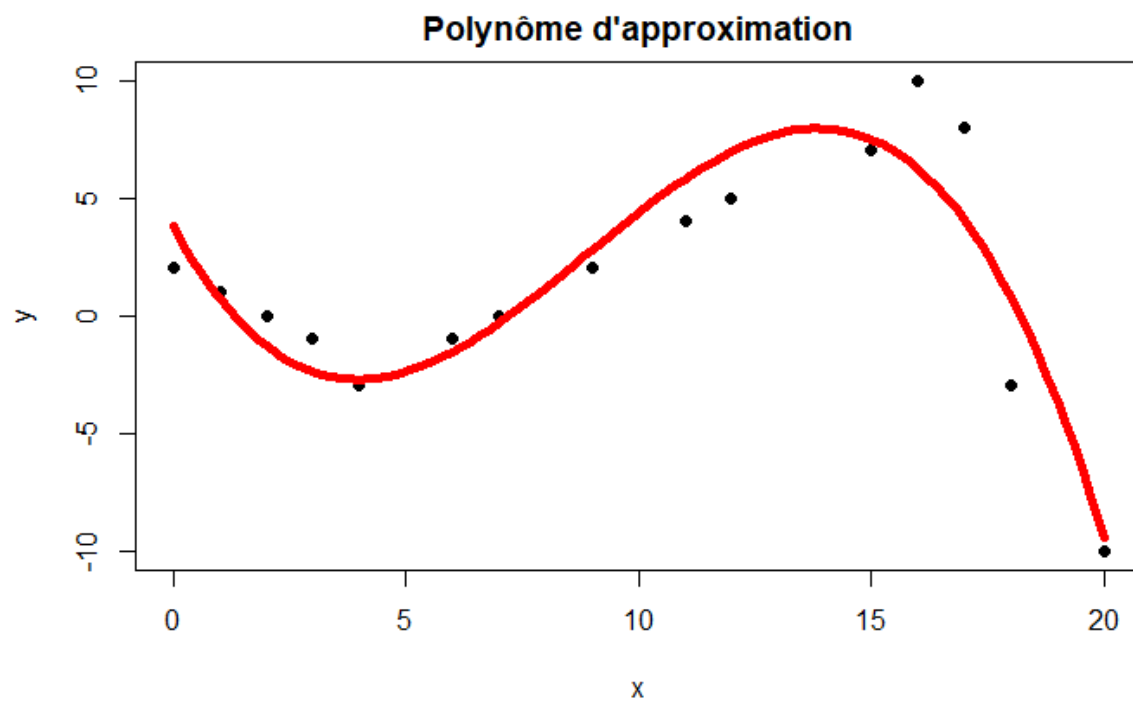
(Intercept) poly(x, degree, raw = TRUE)1 poly(x, degree, raw = TRUE)2 poly(x, degree, raw = TRUE)3

3.78 (c)

-3.62 (x)

0.59 (x2)

-0.02 (x3)



Les prévisions sont :

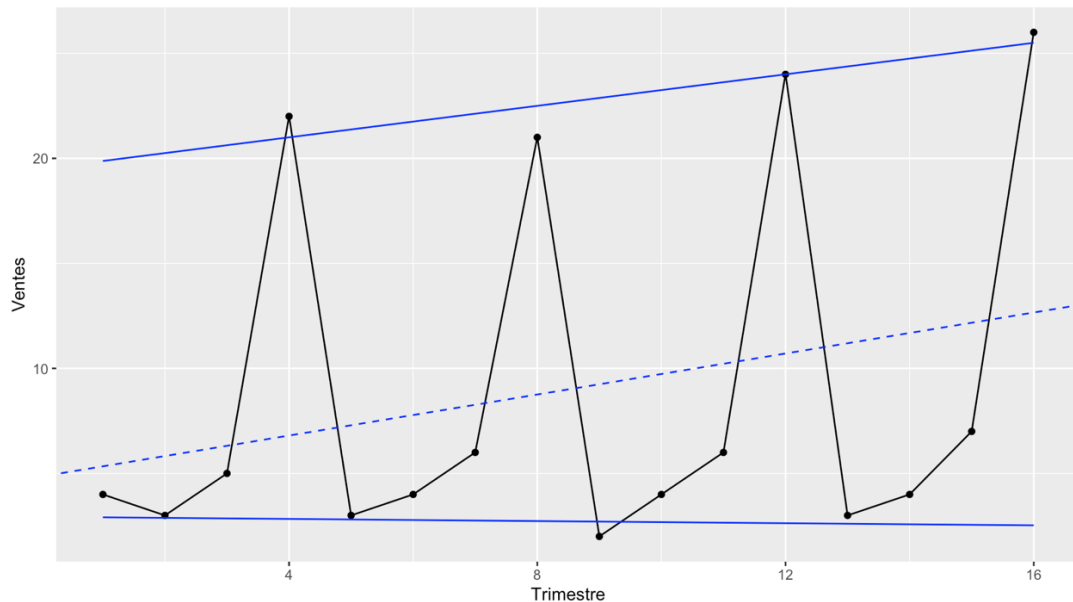
" pour $x = 22$ est de **-25.51**"

" pour $x = 25$ est de **-62.92**"

" pour $x = 50$ est de **-1456.13**"

Exercice 3 : Etude de cas (1)

On peut représenter le graphe chronologique des ventes trimestrielles entre 2019 et 2022 de dindes comme ceci :



Droite d'ajustement affine (*droite en pointillé bleu sur le graphique précédent*) est :

$$y = 0.11 x + 7.81$$

En bleu sont représentées les droites passant par la moyenne des 4 points les plus hauts et les 4 points les bas des ventes de dinde par période (ici une période correspond à une année).

Si les droites sont parallèles, le modèle est additif sinon multiplicatif. Ici, il est difficile de dire quel modèle correspond le mieux. Il faudrait plus de données pour pouvoir trancher. Nous avons donc testé les deux modèles pour voir si l'un correspond mieux aux données.

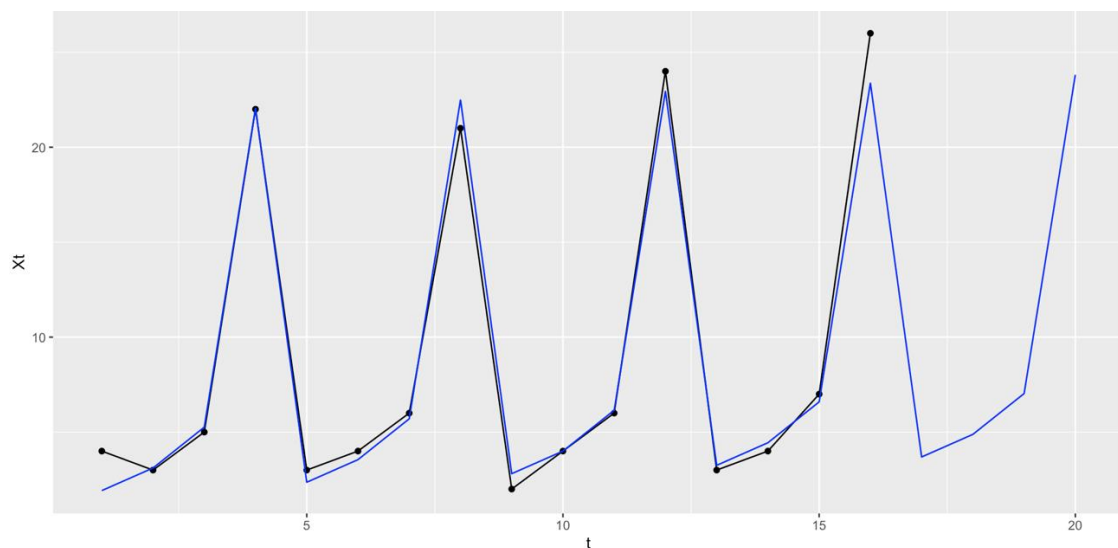
Modèle additif :

Tableau des données :

	t	Xt	Mt	Tt	St_chapeau	St	St_prime	Xt_prime	Xt_prev	erreur
1	1	4	NA	7.92	NA	-6.08	-6.00	10.00	1.92	2.08
2	2	3	NA	8.03	NA	-5.00	-4.92	7.92	3.11	0.11
3	3	5	8.38	8.14	-3.38	-2.96	-2.88	7.88	5.26	0.26
4	4	22	8.38	8.25	13.62	13.71	13.79	8.21	22.04	0.04
5	5	3	8.62	8.36	-5.62	-6.08	-6.00	9.00	2.36	0.64
6	6	4	8.62	8.47	-4.62	-5.00	-4.92	8.92	3.55	0.45
7	7	6	8.38	8.58	-2.38	-2.96	-2.88	8.88	5.70	0.30
8	8	21	8.25	8.69	12.75	13.71	13.79	7.21	22.48	1.48
9	9	2	8.25	8.81	-6.25	-6.08	-6.00	8.00	2.81	0.81
10	10	4	8.62	8.92	-4.62	-5.00	-4.92	8.92	4.00	0.00
11	11	6	9.12	9.03	-3.12	-2.96	-2.88	8.88	6.15	0.15
12	12	24	9.25	9.14	14.75	13.71	13.79	10.21	22.93	1.07
13	13	3	9.38	9.25	-6.38	-6.08	-6.00	9.00	3.25	0.25
14	14	4	9.75	9.36	-5.75	-5.00	-4.92	8.92	4.44	0.44
15	15	7	NA	9.47	NA	-2.96	-2.88	9.88	6.59	0.41
16	16	26	NA	9.58	NA	13.71	13.79	12.21	23.37	2.63
17	17	NA	NA	9.69	NA	-6.08	-6.00	NA	3.69	NA
18	18	NA	NA	9.80	NA	-5.00	-4.92	NA	4.88	NA
19	19	NA	NA	9.91	NA	-2.96	-2.88	NA	7.03	NA
20	20	NA	NA	10.02	NA	13.71	13.79	NA	23.81	NA

Prévision pour le trimestre 20 : **23.81**

Voici la superposition des courbes X_t en noir et T_t+St en (avec la prédiction pour les trimestres suivants) :



On obtient une erreur de **0.7** entre la courbe estimée et les données réels en faisant la moyenne de la somme des différences en valeur absolue.

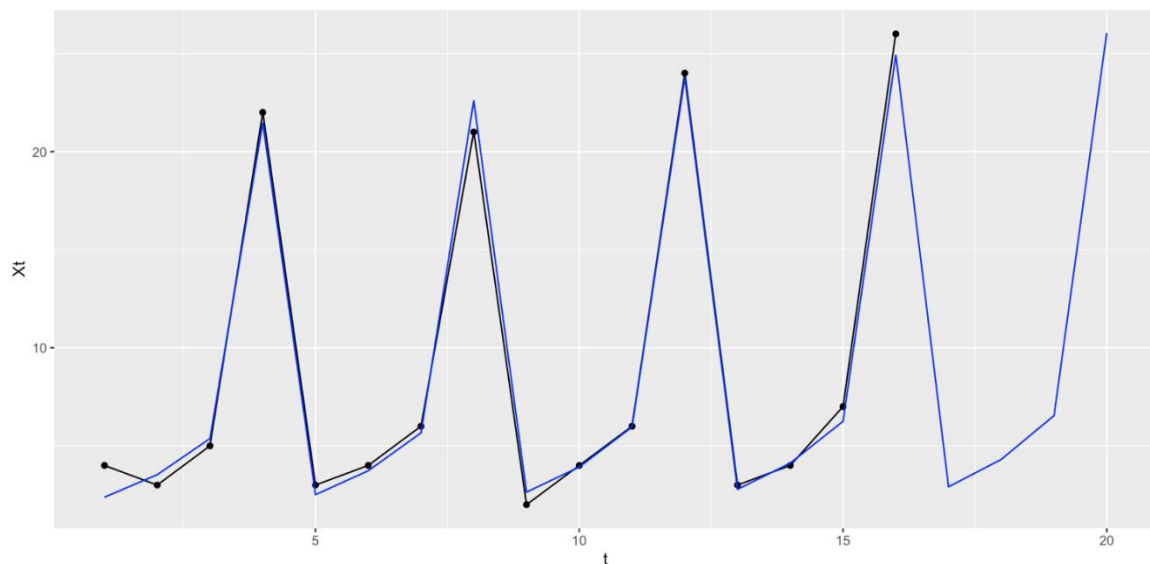
Modèle multiplicatif :

Tableau des données :

	t	Xt	Mt	Tt	St_chapeau	St	St_prime	Xt_prime	Xt_prev	erreur
1	1	4	NA	7.92	NA	0.30	0.30	13.33	2.38	1.62
2	2	3	NA	8.03	NA	0.44	0.44	6.82	3.53	0.53
3	3	5	8.38	8.14	0.60	0.66	0.66	7.58	5.37	0.37
4	4	22	8.38	8.25	2.63	2.59	2.60	8.46	21.45	0.55
5	5	3	8.62	8.36	0.35	0.30	0.30	10.00	2.51	0.49
6	6	4	8.62	8.47	0.46	0.44	0.44	9.09	3.73	0.27
7	7	6	8.38	8.58	0.72	0.66	0.66	9.09	5.66	0.34
8	8	21	8.25	8.69	2.55	2.59	2.60	8.08	22.59	1.59
9	9	2	8.25	8.81	0.24	0.30	0.30	6.67	2.64	0.64
10	10	4	8.62	8.92	0.46	0.44	0.44	9.09	3.92	0.08
11	11	6	9.12	9.03	0.66	0.66	0.66	9.09	5.96	0.04
12	12	24	9.25	9.14	2.59	2.59	2.60	9.23	23.76	0.24
13	13	3	9.38	9.25	0.32	0.30	0.30	10.00	2.78	0.22
14	14	4	9.75	9.36	0.41	0.44	0.44	9.09	4.12	0.12
15	15	7	NA	9.47	NA	0.66	0.66	10.61	6.25	0.75
16	16	26	NA	9.58	NA	2.59	2.60	10.00	24.91	1.09
17	17	NA	NA	9.69	NA	0.30	0.30	NA	2.91	NA
18	18	NA	NA	9.80	NA	0.44	0.44	NA	4.31	NA
19	19	NA	NA	9.91	NA	0.66	0.66	NA	6.54	NA
20	20	NA	NA	10.02	NA	2.59	2.60	NA	26.05	NA

Prévision pour le trimestre 20 : **26.05**

Voici la superposition des courbes X_t en noir et $Tt \cdot St$ en bleu :

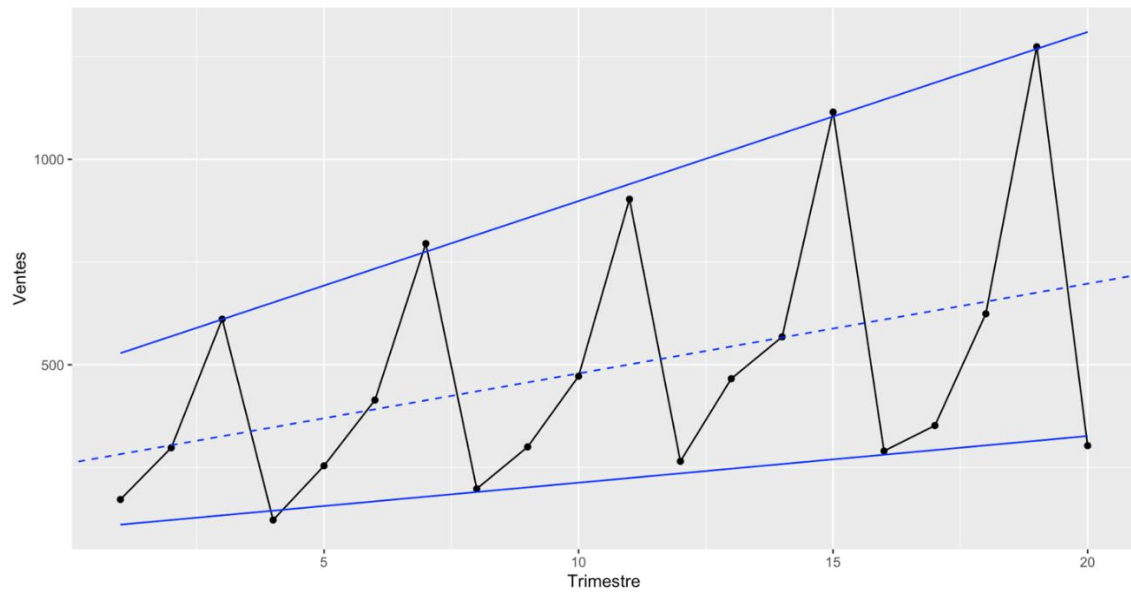


On obtient une erreur de **0.56**

Il y a une erreur de **0.7** pour le modèle additif et **0.56** pour le modèle multiplicatif.

Le modèle multiplicatif semble être plus adapté car il colle mieux aux données mais cela n'est pas suffisant pour nous permettre de conclure sur le choix du modèle. Des données supplémentaires sont nécessaires.

Exercice 4 : Etude de cas (2)



Ici on peut voir que le modèle est bien **multiplicatif** car les deux droites divergent.

Droite d'ajustement affine (*droite en pointillé bleu sur le graphique précédent*) est :

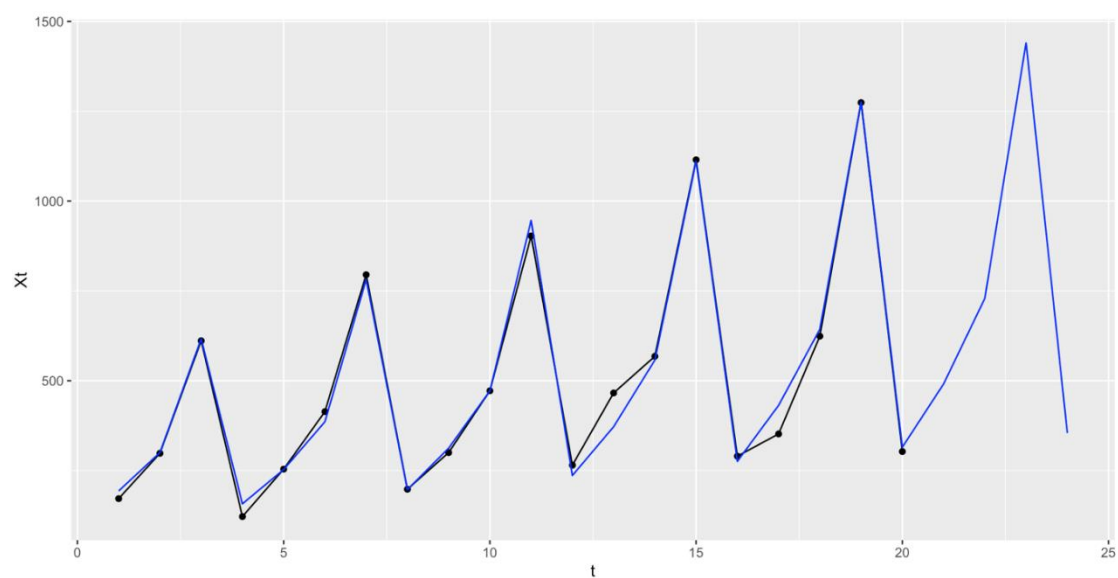
$$y = 21.84 x + 260.47$$

Tableau des données :

t	Xt	Mt	Tt	St_chapeau	St	St_prime	Xt_prime	Xt_prev	erreur
1	172	NA	284.29	NA	0.68	0.68	252.94	193.32	21.32
2	298	NA	306.17	NA	0.98	0.98	304.08	300.05	2.05
3	611	311.00	328.06	1.96	1.88	1.88	325.00	616.75	5.75
4	122	335.75	349.95	0.36	0.45	0.45	271.11	157.48	35.48
5	254	373.25	371.84	0.68	0.68	0.68	373.53	252.85	1.15
6	414	405.75	393.72	1.02	0.98	0.98	422.45	385.85	28.15
7	795	421.00	415.61	1.89	1.88	1.88	422.87	781.35	13.65
8	198	434.00	437.50	0.46	0.45	0.45	440.00	196.88	1.12
9	300	454.75	459.39	0.66	0.68	0.68	441.18	312.39	12.39
10	472	476.62	481.27	0.99	0.98	0.98	481.63	471.64	0.36
11	903	505.75	503.16	1.79	1.88	1.88	480.32	945.94	42.94
12	265	538.50	525.05	0.49	0.45	0.45	588.89	236.27	28.73
13	466	577.00	546.94	0.81	0.68	0.68	685.29	371.92	94.08
14	568	606.62	568.82	0.94	0.98	0.98	579.59	557.44	10.56
15	1115	595.50	590.71	1.87	1.88	1.88	593.09	1110.53	4.47
16	290	588.25	612.60	0.49	0.45	0.45	644.44	275.67	14.33
17	352	615.12	634.49	0.57	0.68	0.68	517.65	431.45	79.45
18	624	636.62	656.37	0.98	0.98	0.98	636.73	643.24	19.24
19	1274	NA	678.26	NA	1.88	1.88	677.66	1275.13	1.13
20	303	NA	700.15	NA	0.45	0.45	673.33	315.07	12.07
21	NA	NA	722.04	NA	0.68	0.68	NA	490.99	NA
22	NA	NA	743.93	NA	0.98	0.98	NA	729.05	NA
23	NA	NA	765.81	NA	1.88	1.88	NA	1439.72	NA
24	NA	NA	787.70	NA	0.45	0.45	NA	354.47	NA

Prévision pour le trimestre 23 : **1439.72**

Voici la superposition des courbes X_t en noir et $Tt*St$ en bleu :



Le modèle fit très bien les données réelles.

Exercice 5 : ACP

Pour l'analyse en composante principale sur un jeu de donnée de pluviométrie pour les villes en France, on a fait 3 sélections d'ensemble de variables :

- Les volumes de précipitations par mois
- Les jours de pluies par mois
- Toutes les variables annuelles : Insolation, quantités de précipitations, jours de pluies, température moyenne, amplitude des températures, latitude et longitude

Pour choisir les variables on a séparé celles annuelles et mensuelles. Les quantités de précipitations et jours de pluies pour les variables mensuelles.

Analyse mensuelle

La figure 1 représente les cercles de corrélations des variations mensuelles des jours de pluies et quantités de précipitations. Les nombres de jours de pluie sont représentés par un seul axe principal d'environ 90% et les quantités de précipitations par 2 axes principaux 60% et 28% avec une inertie totale de 88%.

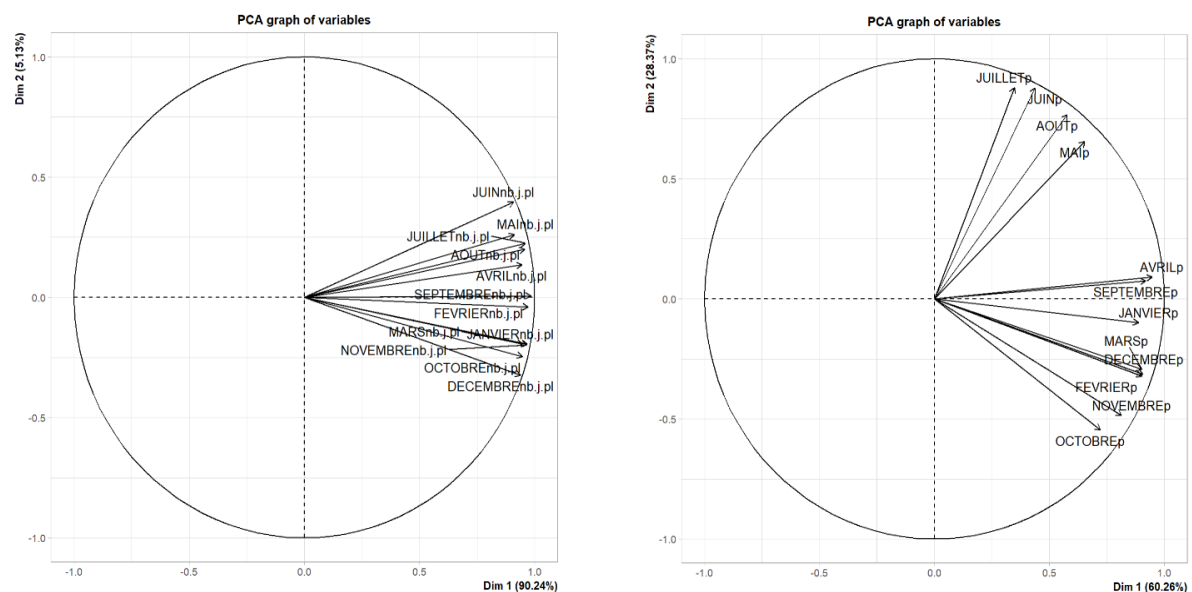


Figure 1 Représentation du cercle de corrélation pour les variables des jours de pluies à gauche et des quantités de précipitations à droite

A l'aide de la figure 2, on peut dire que l'axe principale pour les nombres de jours de pluie par mois représente le nombre de jour de pluie moyen. Les 2 villes les plus éloignées sur l'axe qui sont Marseille et Brest sont les villes avec le plus grand et le plus petit nombre de jour de pluie sur l'année.

Pour les quantités de précipitations par mois, on a le premier axe en abscisse qui représente la quantité de précipitation moyenne. Les 2 villes les plus éloignées sur l'axe qui sont Marseille et Biarritz sont les villes avec la quantité la plus grande et plus petite de précipitation sur l'année. Le 2ème axe en ordonné représente la variation de précipitation entre l'été et les autres mois de l'année. Les 2 villes les plus éloignées sur l'axe qui sont Strasbourg et Ajaccio ont des variations opposées. Pour Strasbourg il y a plus de précipitation en été tandis que pour Ajaccio c'est l'inverse.

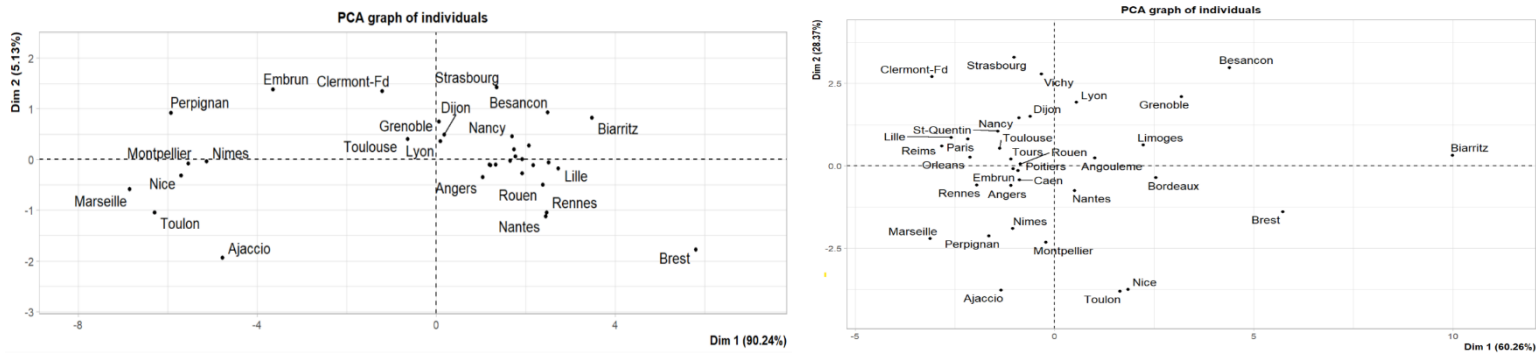


Figure 2 Représentation des villes dans l'espace réduit à 2 dimensions par la méthode ACP pour les variables des jours de pluies en haut et des quantités de précipitations en bas

Les cercles de corrélations permettent de dire que le nombre de jour de pluie de chaque mois est fortement corrélé à tous les autres mois. Ceci permet de dire que toutes les villes suivent la même évolution du nombre de jour de pluie par mois. Par exemple s'il pleut 2 fois plus en janvier qu'en juillet pour la ville X, on observera une tendance similaire pour la ville Y.

Par contre pour les quantités de précipitations, on remarque un groupe de variable qui correspond au mois d'été qui se distinguent des autres mois. Ceci permet de dire que les mois d'été évoluent différemment des autres mois. En reprenant l'exemple précédent, s'il y a 2 fois plus de précipitations au mois de janvier qu'en juillet dans la ville X, cela ne permet pas d'en conclure un comportement similaire pour la ville Y.

On peut conclure qu'il y a moins de jours de pluie en été que durant le reste de l'année pour toutes les villes mais ce n'est pas forcément le cas pour les quantités de précipitations. Par exemple pour Strasbourg il y a plus de quantité de pluie qui tombe en été mais pour Ajaccio c'est l'inverse.

Analyse annuelle

La figure 3 représente les corrélations entre les variables annuelles. 2 axes principaux d'inertie 54% pour l'abscisse et 27% pour l'ordonnée pour un totale de 81%.

On observe 2 groupes distincts :

- Latitude, Nombre annuel de jours de pluie, Insolation annuelle et Température moyenne annuelle (corrélation entre 0.73 et 0.94)
- Longitude, Amplitude annuelle des températures (0.72)

Et la variable Précipitations annuelles qui n'est pas corrélée.

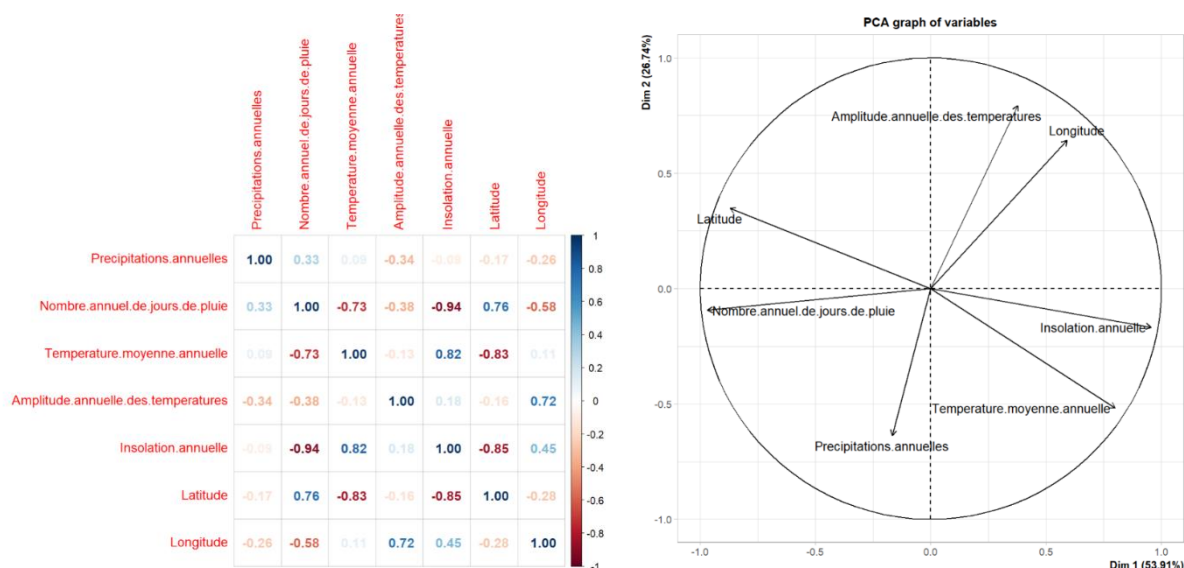


Figure 3 Représentation de la matrice de corrélation à gauche et du cercle de corrélation à droite pour les variables annuelles.

Le premier groupe représente le climat, plus la latitude augmente (on se dirige vers le nord de la France), plus le nombre de jours de pluie augmente et l'insolation diminue avec la température moyenne. La figure 5 montre que ce sont ces variables qui représentent l'axe des abscisses.

Le second groupe représente l'amplitude des températures, plus on se dirige vers l'est plus l'amplitude augmente. La figure 5 montre que ce sont ces variables qui représentent l'axe des ordonnées.

En observant la figure 4 représentant les villes dans ce nouvel espace, on distingue les villes ou « il fait beau » (Marseille, Nîmes, Montpellier ...) qui sont situées sur la côte méditerranéenne. 2 villes se distinguent sur l'axe des ordonnées, Brest et Biarritz étant donné leur position sur la cote ouest et leurs amplitudes des températures les plus faibles.

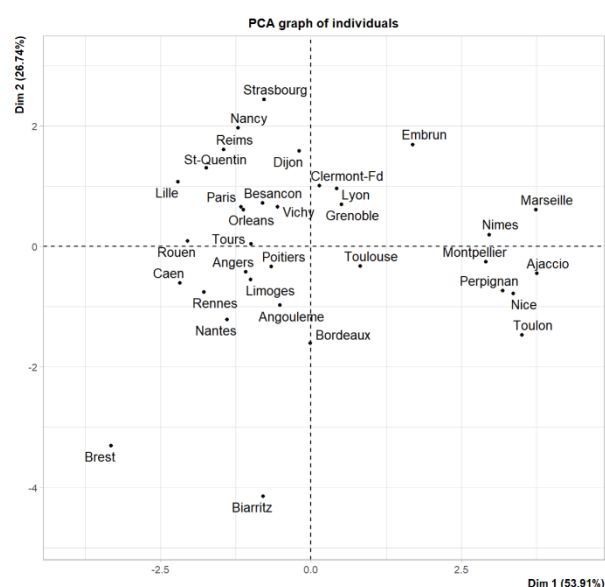


Figure 4 Représentation des villes dans l'espace réduit à 2 dimensions par la méthode ACP pour les variables annuelles

D'après la figure 6, les individus qui contribuent le plus aux 2 axes principaux sont les villes aux extrémités de chaque axe. Pour l'axe 1, les villes sur la côte méditerranéenne mais aussi Brest car c'est la ville où il pleut le plus de jours dans l'année. Pour l'axe 2, Brest et Biarritz car ce sont elles ont les plus faibles amplitudes de températures mais aussi Strasbourg qui possède l'amplitude et la longitude parmi les plus grandes.

Pour conclure, dans cette analyse annuelle on a pu mettre en évidence les villes où il fait « le plus beau » mais aussi que l'amplitude des températures dépend de la longitude ce qui n'est pas triviale.

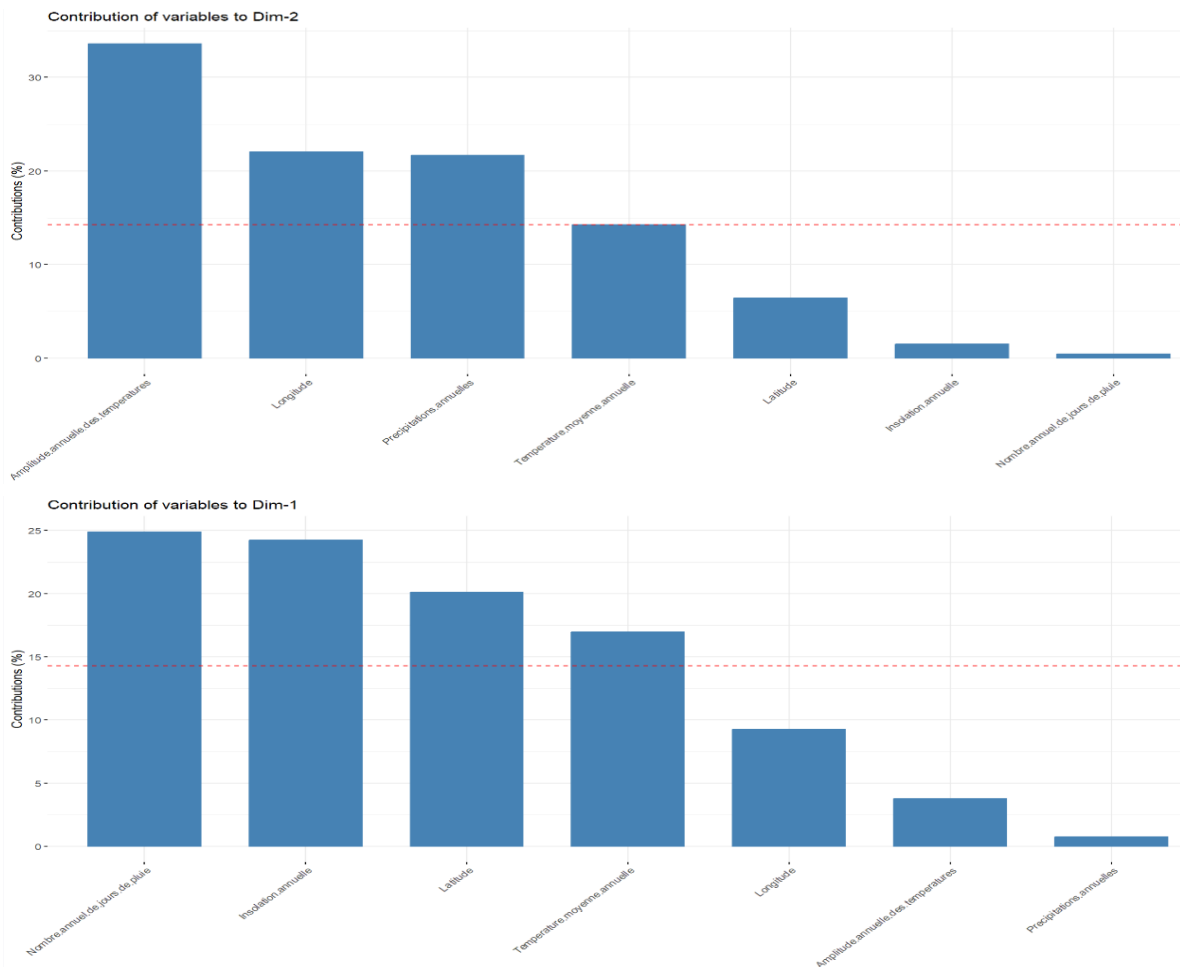


Figure 5 Représentation de la contribution des variables aux 2 axes principaux

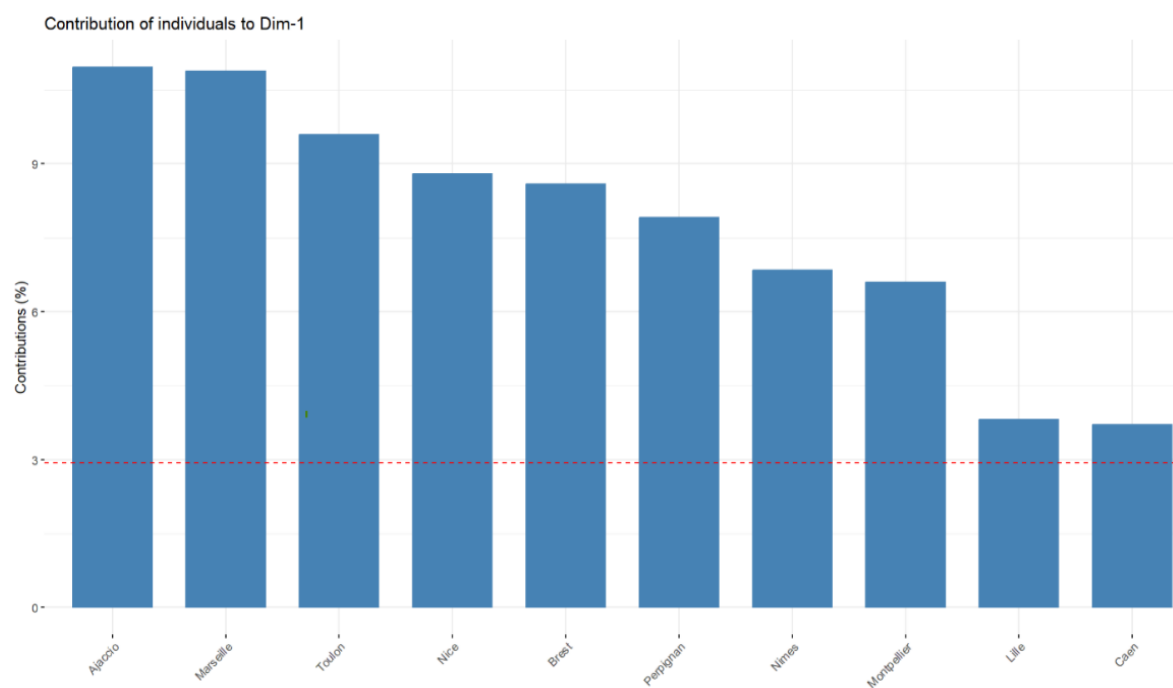
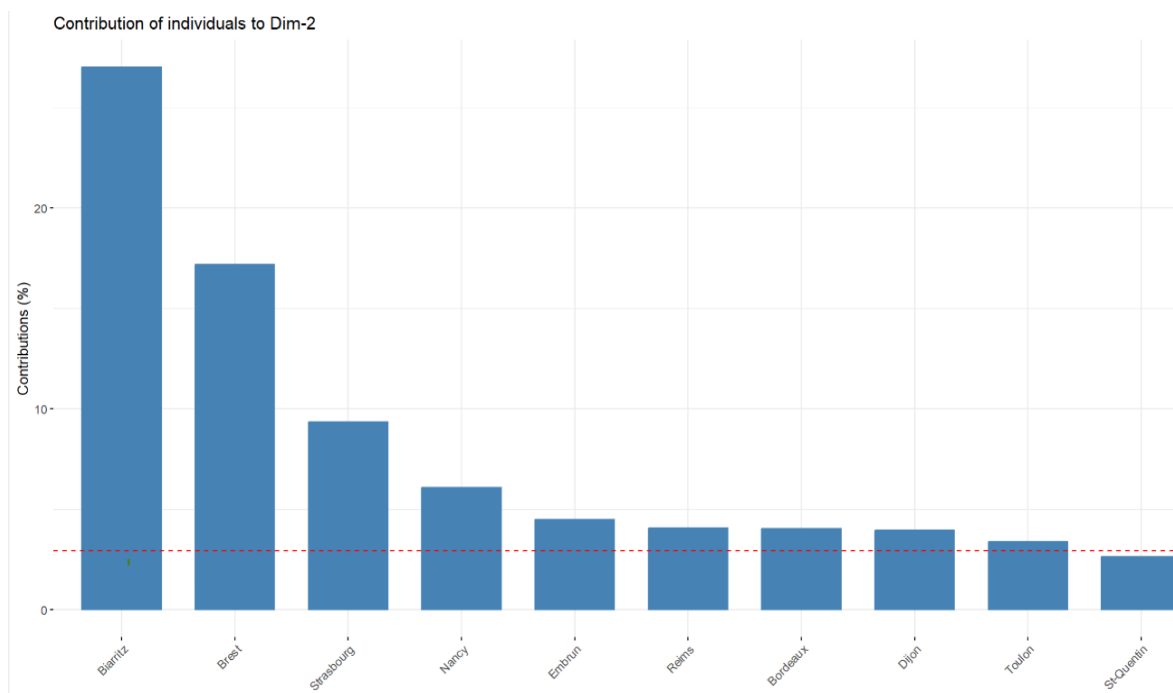


Figure 6 Représentation de la contribution des individus aux 2 axes principaux

