



**FACULTÉ
DES SCIENCES
D'ORSAY**

Mathématiques et applications

Analyse de taux maladies par ACP et Relative Importance Analysis

Par LIAO Puwei

Tuteur ZENG Xiaolin

Juillet 2021

Table des matières

1	Introduction	2
2	Stockage de jeux de données	3
2.1	Division de temps et régions	3
2.2	Créations de variables	4
2.2.1	Variable taux d'infection globale en France	4
2.2.2	Variables taux de vaccination régionaux	5
2.2.3	Variables taux de vaccination par classement âge	6
2.2.4	Variables taux d'infection régionaux	7
2.2.5	Variables taux d'infection par classement d'âge	8
2.2.6	Variables température	8
2.2.7	Variables politique	9
2.3	Mini conclusion de création de variables	9
3	ACP	9
3.1	Normalisation des données et test de l'applicabilité de l'analyse factorielle	9
3.2	Test de validité	11
3.3	Analyse ACP	11
3.3.1	Critère de coude	12
3.3.2	Critère de Kaiser	12
3.4	Composantes principaux	13
4	Relative Importance Analysis	14
4.1	Présentation de Relative Importance Analysis	14
4.2	Processus de la méthode	14
4.3	Application de fonction standard rwa() en R	16
4.4	Groupe 1	16
4.5	Groupe 2	16
4.6	Groupe 3	17
4.7	Mini conclusion	17
5	ANOVA à 1 facteur	17
5.1	Groupe 1 : ANOVA à 1 facteur pour les taux infection régionaux dans différents groupes politique	18
5.1.1	Conditions d'application	18
5.1.2	Processus d'ANOVA à 1 facteur pour Taux d'infection régionaux dans différents groupes de variable politique en SPSS	19
5.2	Groupe 2 : ANOVA à 1 facteur pour les V_i^r dans différents groupes P_k	24
5.2.1	Conditions d'application	24
5.2.2	Processus d'ANOVA à 1 facteur pour les V_i^r dans différents groupes P_k en SPSS	25
6	Conclusion finale	30
7	Références	31

1 Introduction

La maladie coronavirus 2019 (COVID-19), abréviation de "nouvelle pneumonie à coronavirus" et nommée "maladie coronavirus 2019" par l'Organisation mondiale de la santé, désigne la pneumonie causée par la nouvelle infection à coronavirus 2019.

Selon les données disponibles sur les cas, la pneumonie à nouveau coronavirus se caractérise par de la fièvre, une toux sèche et un malaise, quelques patients présentant des symptômes des voies respiratoires supérieures et des symptômes gastro-intestinaux tels que congestion nasale, écoulement nasal et diarrhée. Il convient de noter que les patients gravement malades ou en état critique peuvent avoir une fièvre modérée ou faible, voire aucune fièvre significative, au cours de la maladie. Les patients atteints de formes légères ne présentent qu'une faible fièvre et un léger malaise, sans manifestation de pneumonie. Parmi les cas admis jusqu'à présent, la plupart des patients ont un bon pronostic, tandis que quelques patients sont dans un état critique. Le pronostic est moins bon chez les personnes âgées et chez celles qui présentent une maladie chronique sous-jacente. Chez les enfants, les symptômes sont relativement légers. Rosalind Eggo, Nicholas Davies et leurs collègues de l'École d'hygiène et de médecine tropicale de Londres, au Royaume-Uni, ont utilisé des données démographiques provenant de 32 régions de six pays (Chine, Italie, Japon, Singapour, Canada et Corée du Sud) et des données provenant de six études publiées sur les taux d'infection estimés et la gravité des symptômes par groupe d'âge, à partir desquelles ils ont mis au point un modèle d'évaluation des risques basé sur l'âge. modèle de transmission basé sur la technologie. Grâce à ce modèle, les auteurs ont pu estimer simultanément la susceptibilité à la maladie et les symptômes cliniques à différents âges. Dans toutes les régions, la sensibilité des moins de 20 ans était à peu près la moitié de celle des 20 ans et plus ; seuls 21 pour cent des personnes infectées dans le groupe d'âge le plus jeune, celui des 10-19 ans, présentaient des symptômes cliniques, et 69 pour cent des personnes infectées dans le groupe des 70 ans et plus en présentaient. Par conséquent, il y a de bonnes raisons de faire des spéculations : parce que la capacité du système immunitaire de personnes d'âges différents n'est pas tout à fait la même la prévalence de l'infection par le COVID-19 est fortement dépendante de l'âge.

Par ailleurs , En l'absence d'un ralentissement de l'épidémie, le nombre total estimé de cas cliniques dans chaque ville varie en fonction de l'âge médian de la population locale. Les auteurs ont constaté que les villes dont l'âge de la population était plus élevé présentaient plus de cas cliniques estimés par habitant ; les villes dont l'âge de la population était plus bas présentaient également plus d'infections asymptomatiques (ou infections mineures). Cependant, les estimations des auteurs concernant le nombre sous-jacent d'infections (le nombre d'infections qu'une personne infectée peut transmettre au cours d'une infection) ne différaient pas significativement selon l'âge médian. Les pays dont la population est jeune - comme de nombreux pays à faible revenu - peuvent également présenter des taux d'infection par habitant plus faibles. Toutefois, les auteurs ont également noté que les comorbidités associées aux pays à faible revenu peuvent également influencer la gravité de la maladie et de plus, nous avons constaté que la prévalence de l'infection par le COVID-19 était associée au climat de la zone concernée et à ses habitants. Donc il y a de bonnes raisons de faire des spéculations : la prévalence de l'infection par le COVID-19 est fortement dépendante de région.

En réponse à la variation significative des données sur l'efficacité des vaccins, les experts estiment qu'elle est influencée par de multiples facteurs potentiels, notamment la taille de l'essai clinique, les critères de recrutement des patients, la population cible, la durée d'observation après la vaccination, le délai entre les doses et la gravité de l'épidémie sur chaque site de vaccination ; en outre, la sensibilité des nouveaux réactifs de détection de l'acide nucléique de la couronne et les différences dans les méthodes statistiques peuvent affecter les relevés finaux de l'efficacité des vaccins.

En fait, l'efficacité de plusieurs des vaccins de New Crown qui ont été approuvés fait encore l'objet de controverses. Le plus controversé d'entre eux est le vaccin mis au point par l'Université d'Oxford et AstraZeneca, qui a présenté deux données d'efficacité différentes en fonction de la dose injectée. Lorsque deux doses complètes du vaccin ont été administrées à des participants adultes de tous âges, l'efficacité du vaccin était de 62 pour cent ; toutefois, dans un sous-groupe de participants adultes jusqu'à 55 ans qui ont reçu une dose et demie et une dose complète, l'efficacité du vaccin est passée à 90 pourcenmt, ce qui donne au vaccin un taux d'efficacité moyen d'environ 70 pour cent.Et nous avons consulté des chercheurs qui ont noté que l'âge et les différences régionales semblent être les influences dominantes sur l'efficacité des vaccins à l'heure actuelle.

Sur cette base, nous avons décidé d'utiliser la méthode ACP pour étudier l'incidence hebdomadaire en France, avec l'âge et la région comme variables, afin de voir quels aspects dominant l'incidence hebdomadaire.

2 Stockage de jeux de données

2.1 Division de temps et régions

Pour avoir notre propres variables , nous commençons par stockage de notre propre jeux de données.Toutes les données dont nous avons besoin tournent autour du temps, des taux de vaccination, de la région, de l'âge et des taux d'infection.

Par rapport au temps, nous décidons de choisir de premier janvier 2021 à le 15 juin 2021 parce que cette période inclut le taux d'infection pour la période où il n'y a pas de vaccination de la population et le taux d'infection pour la période où il y a à la fois un taux de vaccination et un taux d'infection régional correspondant.Et de plus , nous décidons de créer des variables hebdomadaires dont il y en a 24 moments données

Par rapport au région, nous avons décidé de diviser la France en 9 par rapport à densité de population, la raison est que nous intéressont à si densité de population influence beaucoup le taux de Maladie covid 19.Et numérotions-leur les numéros 1 à 9 pour faciliter créer des notre propres variables plus tard. Précisément,pour les régions 1 à 6 ce sont les découpage de régions plus dense, pour les restes, c'est à dire 7-9 sont les régions moins dense.

Par rapport à l'âge, nous avons décidé de classer les gens en jeune et vieux , ce qui ont inférieurement à 54 ans,nous leur classons en groupe jeunes ce qui est numérotés en 1.Et pour ce qui ont supérieurement à 54 ans, nous leur classons en groupe vieux ce qui est numérotés en 2.

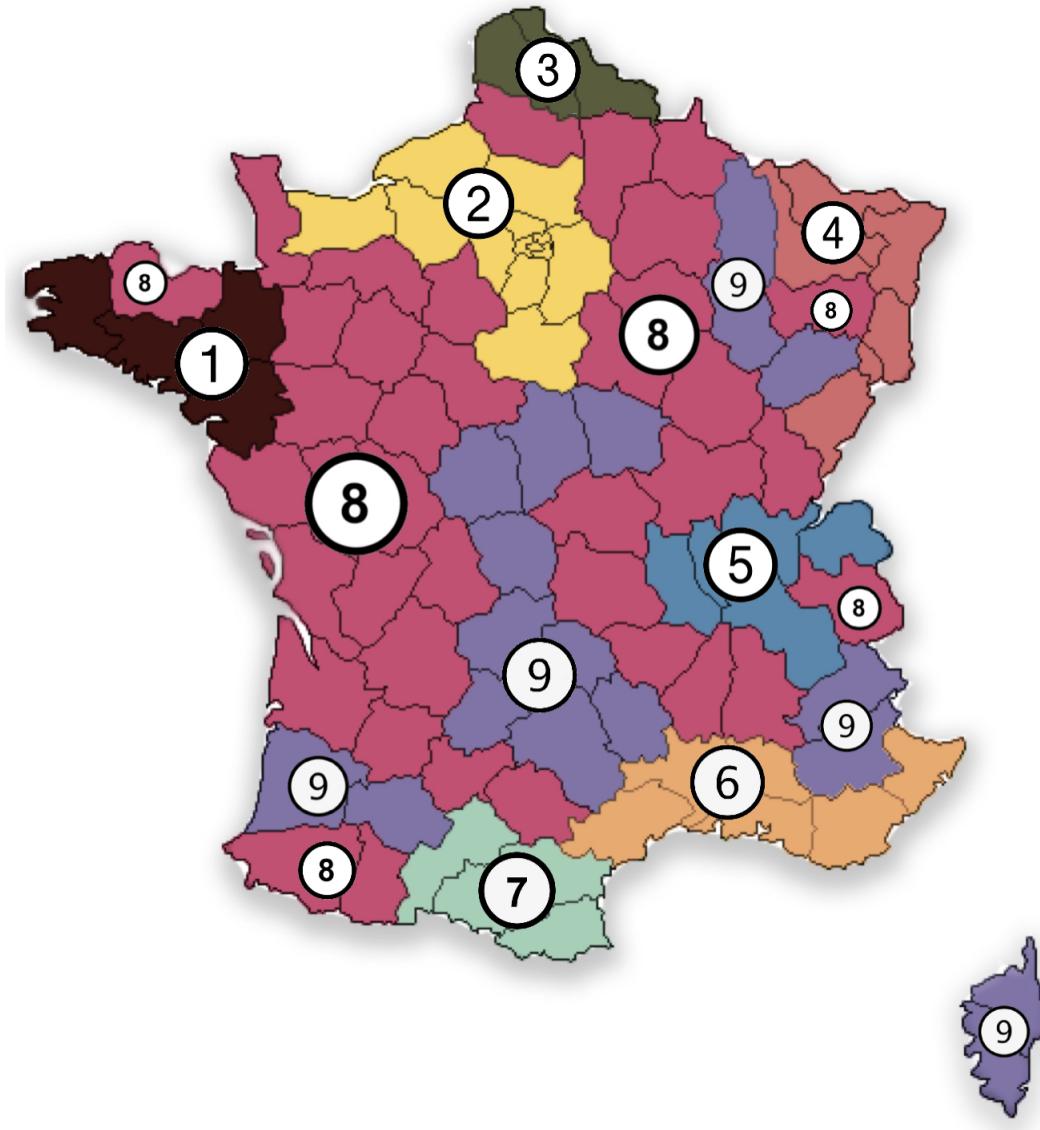


FIGURE 1 – Division de France en 9 parties

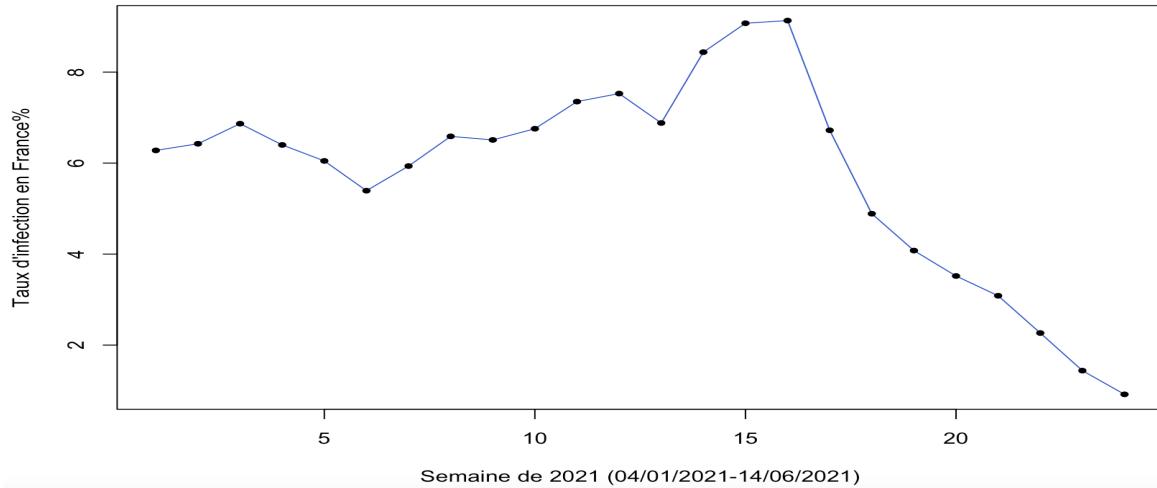
2.2 Créations de variables

2.2.1 Variable taux d'infection globale en France

Notre premier variable est le taux d'infection globale en France hebdomadaire ,notons :

$$Y_k, \forall k \in \{1..24\}$$

ce qui Reflète le taux d'infection français quotidien.Avec cette variable, nous pouvons voir l'ampleur du changement dans le taux d'infection français hebdomadaire, et cette variable sera utilisée comme notre variable d'analyse cible.



D'après cette graphe, Nous pouvons tout d'abord observer qu'en 2021, de début janvier à mi-avril, l'incidence tend à être plus ou moins stable en France, ce qui peut être lié aux mesures de fermeture des villes et de couvre-feu en France, qui ont quelque peu supprimé l'augmentation de l'incidence en France. Et de plus nous pouvons constater qu'à partir de la 16^e semaine environ, c'est-à-dire le 19 avril, le taux d'infection en France a commencé à afficher une tendance significative à la baisse, et nous supposons que le changement de tendance du taux d'infection est étroitement lié à la vaccination généralisée.

2.2.2 Variables taux de vaccination régionaux

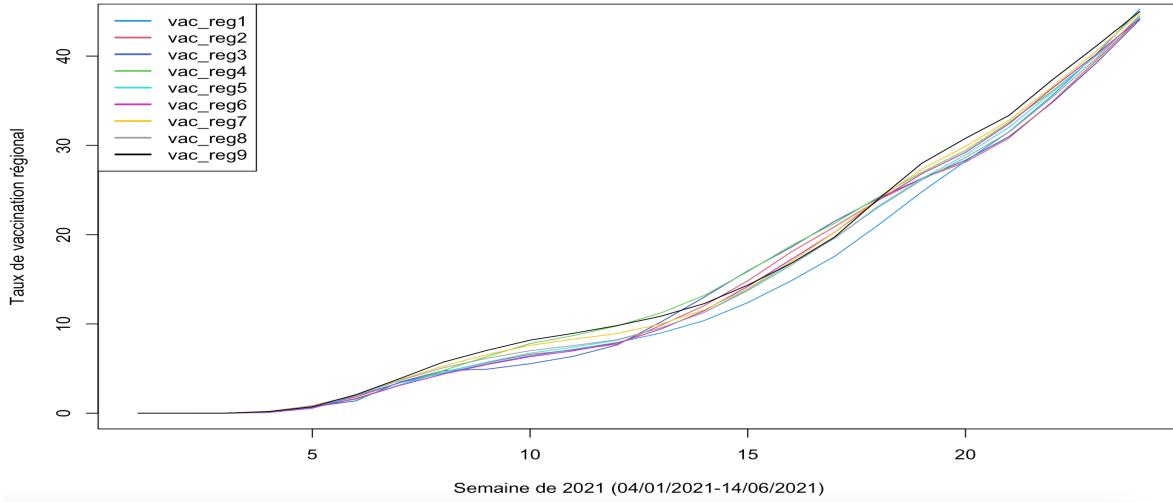
Après avoir collecté les jeux de données par rapport au taux de vaccination de chaque département, nous numérotions d'abord chaque département par numéro 1 à 9 en utilisant la fonction which en R.

Avec ce que l'on a, nous pouvons donc commencer à créer ces variables :

$$V_{i \ k}^r$$

$\forall i \in \{1, \dots, 9\}$: i ^{ème} partie de région

$\forall k \in \{1..24\}$: k i ^{ème} semaine



Voici les gourbes de taux de vaccination V_i^r , Nous pouvons clairement voir que Les taux de vaccination dans chaque région sont presque identiques avec la même transformation numérique et la même tendance, et nous pouvons presque supposer que la politique de vaccination est mise en œuvre à peu près de la même manière dans chaque région.

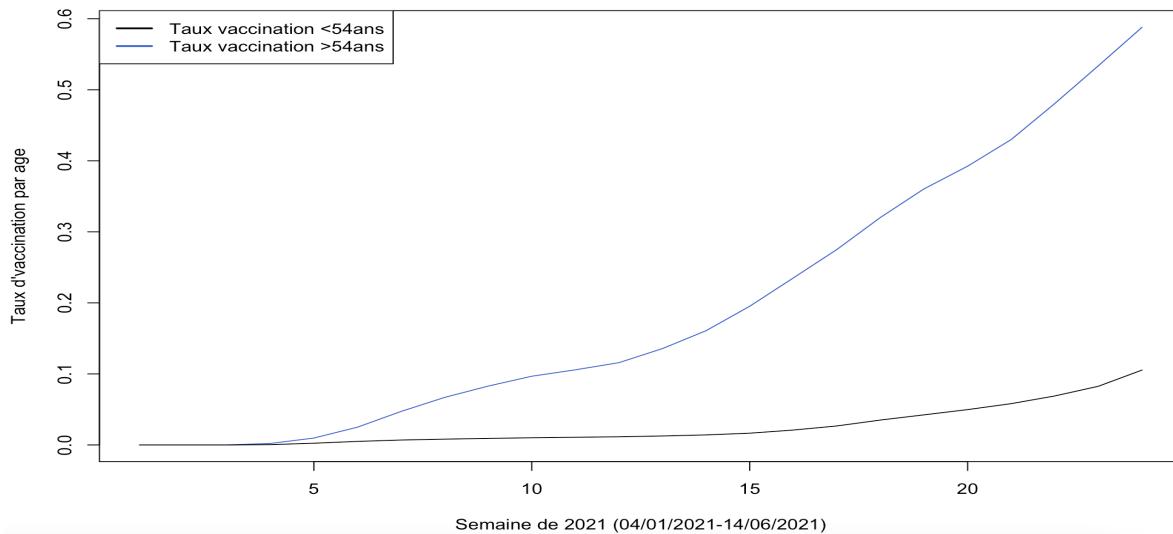
2.2.3 Variables taux de vaccination par classement âge

Créons les variables

$$V_{j \cdot k}^a$$

$\forall j \in \{1, 2\}$: j ème classement d'âge

$\forall k \in \{1..24\}$: k ième semaine



Par ces 2 courbes , nous constatons que le taux de vaccination est beaucoup plus élevé chez les plus de 54 ans que chez les moins de 54 ans, et que la tendance est à l'augmentation beaucoup plus rapide que chez les moins de 54 ans. Les taux de vaccination des personnes âgées de plus de 54 ans ont atteint jusqu'à présent environs de 60 pour cent. Mais regardons plutôt le groupe des moins de 54 ans, dont le taux de vaccination global est très faible, beaucoup plus faible que celui du groupe des plus de 54 ans, ce qui est peut-être indissociable de la politique de vaccination initiale en France, puisque la vaccination a été proposée en priorité à la population âgée, et ce n'est qu'après un certain temps, c'est-à-dire vers le mois d'avril, que la vaccination a commencé pour les personnes âgées de plus de 55 ans. Pour les plus de 18 ans, le droit à la vaccination commence à la fin du mois de mai, ce qui peut être l'une des raisons du taux de vaccination nettement inférieur chez les moins de 55 ans.

2.2.4 Variables taux d'infection régionaux

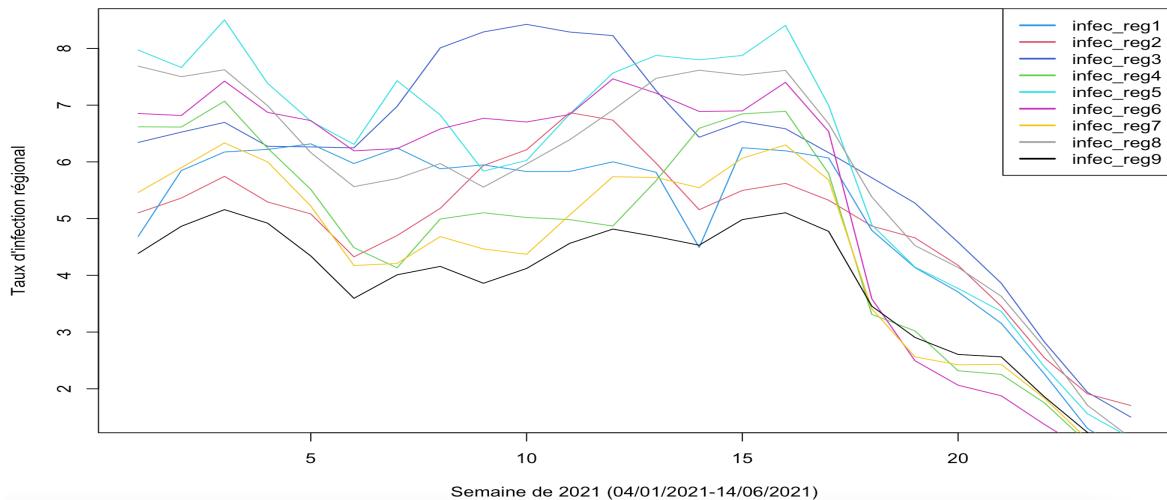
De même façons,pour les variables taux d'infection régionaux,nous numérotions d'abord chaque département par numéro 1 à 9 en utilisant la fonction which en R.

Avec ce que l'on a, nous pouvons donc commencer à créer ces variables :

$$X_{i k}^r$$

$\forall i \in \{1, \dots, 9\}$: i ème partie de région

$\forall k \in \{1..24\}$: k ième semaine



D'après ces 9 courbes, Nous pouvons constater que Les taux d'infection sont également plus faibles dans les zones moins densément peuplées que dans les zones plus densément peuplées, peut-être parce que le virus se propage plus rapidement et plus largement dans les zones plus densément peuplées.Nous trouvons que la densité de population est corrélé avec le taux d'infection.

2.2.5 Variables taux d'infection par classement d'âge

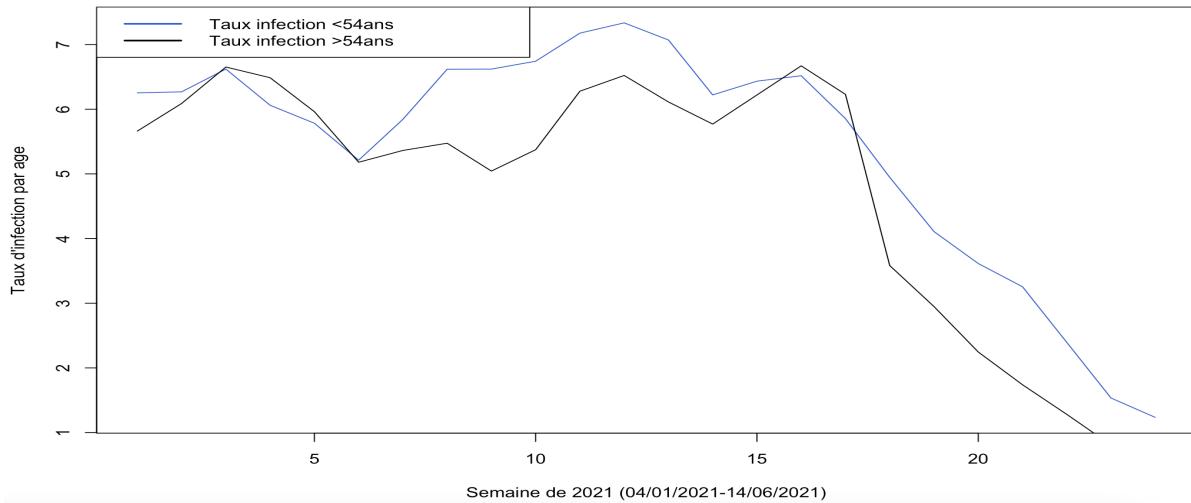
De même façons, pour les variables taux d'infection par classement d'âge.

Avec ce que l'on a, nous pouvons donc commencer à créer ces variables :

$$X_{j,k}^a$$

$\forall i \in \{1, 2\}$: j ème classement d'âge

$\forall k \in \{1..24\}$: k ième semaine



D'après ces 2 courbes, Nous pouvons constater que dans chaque classement d'âge sont presque identiques , cela peut s'expliquer par le fait que la différence des taux d'infection entre les groupes d'âge n'est pas forcément significative.

2.2.6 Variables température

La variable température est en unité degré Celsius, ce qui représente la température hebdomadaire moyenne de la France, la raison est que nous soupçonnons qu'il existe une corrélation non négligeable entre la propagation du virus et la température, c'est pourquoi nous insistons pour que la température soit l'une des variables de notre étude. Nous notons :

$$T_k$$

$\forall k \in \{1, .., 24\}$: k ième semaine de 2021

2.2.7 Variables politique

La variable politique est en unité booléenne, plus précisément, 0 représente confinement, 1 représente couvre-feu et 2 représenter déconfinement, c'est aussi une variable non négligeable car les facteurs politiques d'origine humaine ont un impact non négligeable sur la propagation du virus. Nous notons :

$$P_k$$

$\forall k \in \{1, \dots, 24\}$: k ième semaine de 2021

2.3 Mini conclusion de création de variables

Jusqu'à cette étape, nous avons réussi à créer nos variables, $\forall k \in \{1..24\}$:

$$\begin{aligned} & Y_k \\ & V_i^r, \forall i \in \{1, \dots, 9\} \\ & V_j^a, \forall j \in \{1, 2\} \\ & X_i^r, \forall i \in \{1, \dots, 9\} \\ & X_j^a, \forall i \in \{1, 2\} \\ & T_k \\ & P_k \end{aligned}$$

A partir de ces variables , nous pouvons commencer à faire nos études par rapport à méthode ACP et relative importance analysis, c'est bien ce que l'on va faire les sections suivantes.

3 ACP

L'analyse en composantes principales (ACP) qui est une tentative de remplacer les nombreux indicateurs originaux présentant une certaine corrélation en les regroupant dans un nouvel ensemble d'indicateurs composites non corrélés entre eux.Dans nos sections suivantes, nous décidons d'utiliser logiciel SPSS pour réaliser une analyse ACP.

3.1 Normalisation des données et test de l'applicabilité de l'analyse factorielle

A partir de cette étape, nous normalisons d'abord notre jeu de données en raison de toutes les variables sont pas tout à fait à même unité, donc c'est obligatoire de d'abord normaliser notre jeu de données pour que l'analyse suivante soit correcte .Et de plus, il nous faut tester que les données satisfont les pré-requis d'ACP, nous avons testé l'applicabilité de l'analyse factorielle avec le test de KMO et le test de Bartlett pour l'analyse factorielle et nous trouvons une matrice de corrélation qui n'est pas définie positive et les valeurs propres calculés sont suivantes :

		Matrice de corrélation																								
		X_i	T_i	I^a_1	I^a_2	I^r_1	I^r_2	I^r_3	I^r_4	I^r_5	I^r_6	I^r_7	I^r_8	I^r_9	V^a_1	V^a_2	V^r_1	V^r_2	V^r_3	V^r_4	V^r_5	V^r_6	V^r_7	V^r_8	V^r_9	S_i
X_i	1	-0.628	0.931	0.936	0.867	0.878	0.853	0.921	0.935	0.928	0.933	0.936	0.946	-0.83	-0.74	-0.774	-0.73	-0.71	-0.71	-0.75	-0.74	-0.75	-0.75	-0.75	-0.563	
T_i	-0.628	1	-0.768	-0.795	-0.79	-0.643	-0.75	-0.75	-0.78	-0.84	-0.74	-0.71	-0.73	0.905	0.978	0.962	0.975	0.977	0.982	0.974	0.976	0.975	0.972	0.976	0.978	
I^a_1	0.931	-0.768	1	0.954	0.926	0.963	0.964	0.88	0.933	0.96	0.916	0.926	0.948	-0.93	-0.86	-0.889	-0.86	-0.86	-0.85	-0.87	-0.87	-0.87	-0.87	-0.87	-0.713	
I^a_2	0.936	-0.795	0.954	1	0.934	0.886	0.867	0.943	0.974	0.985	0.977	0.951	0.981	-0.92	-0.88	-0.9	-0.87	-0.86	-0.88	-0.88	-0.885	-0.89	-0.89	-0.753		
I^r_1	0.867	-0.79	0.926	0.934	1	0.883	0.903	0.836	0.899	0.916	0.88	0.87	0.925	-0.93	-0.87	-0.897	-0.87	-0.86	-0.88	-0.87	-0.87	-0.878	-0.88	-0.73		
I^r_2	0.878	-0.643	0.963	0.886	0.883	1	0.953	0.789	0.842	0.873	0.84	0.862	0.895	-0.85	-0.76	-0.797	-0.77	-0.76	-0.75	-0.78	-0.77	-0.77	-0.776	-0.76	-0.598	
I^r_3	0.853	-0.747	0.964	0.867	0.903	0.953	1	0.751	0.826	0.889	0.788	0.811	0.847	-0.89	-0.83	-0.856	-0.84	-0.84	-0.82	-0.84	-0.83	-0.84	-0.83	-0.672		
I^r_4	0.921	-0.748	0.88	0.943	0.836	0.789	0.751	1	0.956	0.934	0.975	0.966	0.956	-0.87	-0.82	-0.846	-0.81	-0.8	-0.81	-0.83	-0.82	-0.83	-0.83	-0.735		
I^r_5	0.935	-0.778	0.933	0.974	0.899	0.842	0.826	0.956	1	0.961	0.974	0.979	0.977	-0.92	-0.86	-0.889	-0.86	-0.84	-0.85	-0.87	-0.87	-0.874	-0.88	-0.752		
I^r_6	0.928	-0.836	0.96	0.985	0.916	0.873	0.889	0.934	0.961	1	0.957	0.929	0.952	-0.93	-0.9	-0.918	-0.89	-0.89	-0.91	-0.91	-0.91	-0.907	-0.91	-0.781		
I^r_7	0.933	-0.735	0.916	0.977	0.88	0.84	0.788	0.975	0.974	0.957	1	0.972	0.986	-0.88	-0.82	-0.848	-0.81	-0.8	-0.83	-0.82	-0.83	-0.832	-0.84	-0.712		
I^r_8	0.936	-0.714	0.926	0.951	0.87	0.862	0.811	0.966	0.979	0.929	0.972	1	0.979	-0.9	-0.82	-0.854	-0.82	-0.8	-0.81	-0.83	-0.82	-0.83	-0.835	-0.83	-0.706	
I^r_9	0.946	-0.731	0.948	0.981	0.925	0.895	0.847	0.956	0.977	0.952	0.986	0.979	1	-0.91	-0.83	-0.864	-0.83	-0.81	-0.82	-0.84	-0.84	-0.845	-0.84	-0.705		
V^a_1	-0.834	0.905	-0.926	-0.921	-0.93	-0.853	-0.89	-0.87	-0.92	-0.93	-0.88	-0.9	-0.91	1	0.969	0.985	0.97	0.964	0.964	0.974	0.971	0.973	0.976	0.972	0.881	
V^a_2	-0.735	0.978	-0.862	-0.875	-0.87	-0.763	-0.83	-0.82	-0.86	-0.9	-0.82	-0.83	0.969	1	0.997	0.999	0.998	0.999	1	1	1	1	0.999	0.998	0.941	
V^r_1	-0.774	0.962	-0.889	-0.9	-0.9	-0.797	-0.86	-0.85	-0.89	-0.92	-0.85	-0.85	-0.86	0.985	0.997	1	0.996	0.993	0.995	0.998	0.997	0.998	0.999	0.941		
V^r_2	-0.728	0.975	-0.862	-0.868	-0.87	-0.767	-0.84	-0.81	-0.86	-0.89	-0.81	-0.82	-0.83	0.977	0.999	0.996	1	0.999	0.999	0.999	0.999	0.999	0.999	0.958		
V^r_3	-0.714	0.977	-0.855	-0.858	-0.86	-0.764	-0.84	-0.8	-0.84	-0.89	-0.8	-0.8	-0.81	0.964	0.998	0.993	0.999	1	0.998	0.997	0.998	0.997	0.997	0.995		
V^r_4	-0.714	0.982	-0.846	-0.86	-0.86	-0.746	-0.82	-0.81	-0.85	-0.89	-0.8	-0.81	-0.82	0.964	0.999	0.995	0.999	0.998	1	0.999	0.999	0.998	0.998	0.966		
V^r_5	-0.748	0.974	-0.872	-0.883	-0.88	-0.775	-0.84	-0.83	-0.87	-0.91	-0.83	-0.83	-0.84	0.974	1	0.998	0.999	0.997	0.999	1	1	1	1	0.954		
V^r_6	-0.744	0.976	-0.871	-0.88	-0.87	-0.774	-0.84	-0.82	-0.87	-0.91	-0.82	-0.82	-0.84	0.971	1	0.997	0.999	0.998	0.999	1	1	1	1	0.954		
V^r_7	-0.749	0.975	-0.868	-0.884	-0.87	-0.768	-0.83	-0.83	-0.87	-0.91	-0.83	-0.83	-0.84	0.973	1	0.998	0.999	0.997	0.998	1	0.999	1	1	0.956		
V^r_8	-0.75	0.972	-0.873	-0.885	-0.88	-0.776	-0.84	-0.83	-0.87	-0.91	-0.83	-0.84	-0.85	0.976	1	0.999	0.999	0.997	0.998	1	0.999	1	1	0.954		
V^r_9	-0.748	0.976	-0.865	-0.886	-0.88	-0.762	-0.83	-0.84	-0.88	-0.91	-0.84	-0.83	-0.84	0.972	0.999	0.998	0.997	0.995	0.998	0.999	0.999	1	0.999	0.959		
S_i	-0.563	0.978	-0.713	-0.753	-0.73	-0.598	-0.67	-0.74	-0.75	-0.78	-0.71	-0.71	-0.71	0.881	0.96	0.941	0.958	0.96	0.966	0.954	0.954	0.956	0.954	0.959	1	

Cette matrice n'est pas définie positive.

FIGURE 2 – Matrice de corrélation

valeur propre
22.28347227
1.79904786
0.52795278
0.11786223
0.10316585
0.06958804
0.03955797
0.02077371
0.0137489
0.00872725
0.004873
0.00405785
0.00360013
0.00239894
0.00185338
0.00138668
0.00106589
0.00049709
0.00038685
0.00023478
0.0001171
-0.00039181
-0.00094461
-0.00146313
-0.001569

FIGURE 3 – Valeurs propres de la matrice

Dans ce cas, nous trouvons que la plupart de valeurs propres sont très proche de 0, alors nous pouvons dire que qu'il existe certains colonnes de cette matrice qui sont très corrélés entre eux, donc notre ACP n'a

pas un sens de réaliser, nous allons alors éliminer certains variables pour que notre ACP a un sens correct.

Après plusieurs essais, On trouve que les variables X_i^r et V_i^r sont très corrélées, donc nous les enlevé dans l'analyse ACP. Nous avons aussi enlevé la variable P_k , en raison que la politique est une variable qualitative ce qui ne peut pas être traité par analyse ACP. Nous avons finalement décidé d'analyser ces variables en enlevant les variables indiquées précédemment par la méthode ANOVA à 1 facteur dans la section ANOVA.

A partir d'ici,Nous pouvons alors procéder la nouvelle ACP avec les variables restants.

Matrice de corrélation							
	taux_maladie_national	temperature	taux_infection_age1	taux_infection_age2	taux_vaccination_age1	taux_vaccination_age2	
Corrélation	taux_maladie_national	1.000	-.628	.931	.936	-.834	-.735
	temperature	-.628	1.000	-.768	-.795	.905	.978
	taux_infection_age1	.931	-.768	1.000	.954	-.926	-.862
	taux_infection_age2	.936	-.795	.954	1.000	-.921	-.875
	taux_vaccination_age1	-.834	.905	-.926	-.921	1.000	.969
	taux_vaccination_age2	-.735	.978	-.862	-.875	.969	1.000

FIGURE 4 – Matrice de corrélation

Dans cette nouvelle matrice, nous constatons que toutes les valeurs sont supérieures à 0.5 en valeur absolue, passons nous à test de validité pour voir que si notre analyse ACP est réalisable ou pas.

3.2 Test de validité

Indice KMO et test de Bartlett		
Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin.		.769
Test de sphéricité de Bartlett	Khi-deux approximé ddl Signification de Bartlett	296.639 15 .000

FIGURE 5 – Test de validité

A l'aide de KMO et Barlette,nous trouvons que l'indice KMO est plus grand que 0.5 ce qui implique que il y a une corrélation entre ces variables et de plus,la signification de barlett est très significative qui est 0, nous pouvons dire que les variables sont corrélées. Grâce à KMO et Barlette, nous pouvons conclure que ces variables choisis sont adaptées pour analyse factorielle et notre analyse ACP est réalisable.

3.3 Analyse ACP

Tout d'abord, nous allons décider les axes principaux à retenir en utilisant le critère de coude et le critère de Kaiser :

3.3.1 Critère de coude

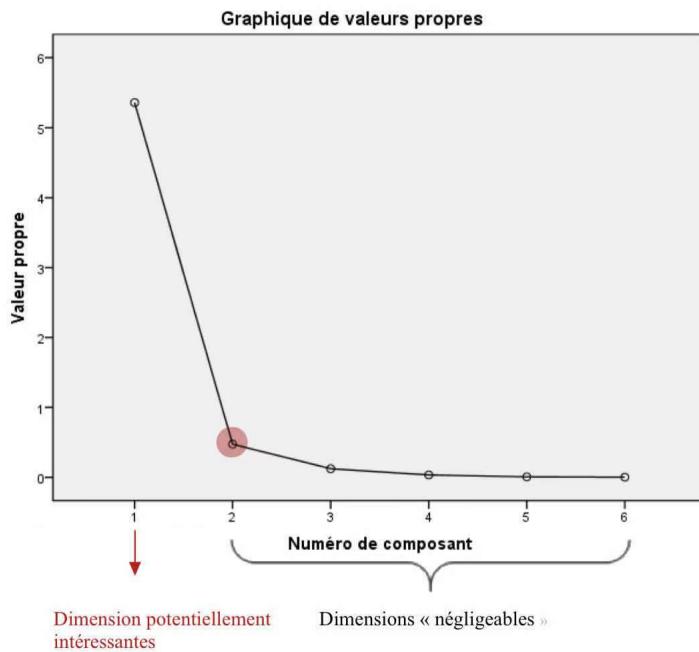


FIGURE 6 – Critère de coude

D'après ce graphe, nous constatons que sauf le premier axe, tout les autres axes compose une inertie très petit, donc ici nous décidons de garder que le permier axe

3.3.2 Critère de Kaiser

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs retenus		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	5.346	89.107	89.107	5.346	89.107	89.107
2	.529	8.817	97.924			
3	.057	.955	98.879			
4	.036	.592	99.471			
5	.029	.489	99.960			
6	.002	.040	100.000			

Méthode d'extraction : Analyse en composantes principales.

FIGURE 7 – Critère de Kaiser

D'après ce graphe, variance totale expliquée, nous constatons que c'est toujours la première composante qui compose le plus parmis les 5 autres composantes en variance,et comme notre acp est normé, et de plus notre premier composante est beaucoup plus grand que 1, donc on retient la première composante comme notre axe principal.

3.4 Composantes principaux

Comme notre données sont pas tous en même unité, donc nous décidons de les normaliser pour que notre analyse d'ACP soit correcte, donc ici donc cherchons plutôt avec la matrice de corrélation que la matrice de covariance :

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs retenus		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	5.346	89.107	89.107	5.346	89.107	89.107
2	.529	8.817	97.924			
3	.057	.955	98.879			
4	.036	.592	99.471			
5	.029	.489	99.960			
6	.002	.040	100.000			

Méthode d'extraction : Analyse en composantes principales.

FIGURE 8 – Variance totale expliquée

D'après ce tableau, visiblement que la première composante conserve 89.107% de la variance totale.

Matrice des composantes^a

	Composante
	1
taux_vaccination_age1	.982
taux_infection_age2	-.968
taux_infection_age1	-.962
taux_vaccination_age2	.958
temperature	.896
taux_maladie_national	-.894

Méthode d'extraction : Analyse en composantes principales.

a. 1 composantes extraites.

FIGURE 9 – Matrice des composantes

D'après cette matrice des composantes, nous trouvons que les variables taux vaccination pour tout âges, et la variable température dans la matrice sont très proches de 1, donc nous pouvons dire qu'ils sont très corrélés avec l'axe principal. Et ils sont aussi bien présentés dans le premier axe. Nous pouvons aussi dire que le taux de vaccination par âge et température contribuent positivement avec la suppression de la propagation du virus, et en même temps, le taux d'infection par âge et le taux d'infection national contribue négativement avec la suppression de la propagation du virus.

Dans la section suivante, nous visons à regarder quelle variables influencent le plus le taux d'infection nationale avec une nouvelle méthode Relative importance analysis.

4 Relative Importance Analysis

Notre objectif était de voir quelles variables restantes influençaient le plus le taux d'infection global en France, ce que la méthode ACP précédente n'a pas permis de réaliser. Donc , ici nous avons décidé d'utiliser une méthode qui s'appelle relative importance analysis

4.1 Présentation de Relative Importance Analysis

Les poids relatifs des variables prédicteurs dans une régression multiple sont difficiles à déterminer, et les poids relatifs (également appelés importance relative) sont définis ici comme la proportion de chaque variable prédicteur qui contribue à coefficient de détermination linéaire de Pearson , noté R², qui s'exprime suivant :

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

où n est le nombre de mesures, y_i la valeur de la ième mesure, \hat{y}_i la valeur prédictive correspondante et \bar{y}_i la moyenne des mesures.

Mais en effet, ces mesures sont difficiles à mettre en œuvre lorsque le nombre de prédicteurs dépasse cinq. Mais par contre ,Relative Importance Analysis est efficace sur le plan informatique pour tout nombre de prédicteursle calcul des poids relatifs devient plus complexe lorsque l'on considère l'effet des variables elles-mêmes et en combinaison avec d'autres variables du modèle.Considérez à la fois sa contribution unique et sa contribution lorsqu'elle est combinée à d'autres variables. Pour éviter toute confusion, nous utilisons le terme de "poids relatif" pour désigner cette contribution, plutôt que celui d'importance relative.

Dans le cas particulier où toutes les variables prédictives ne sont pas corrélées, le calcul des poids relatifs est relativement simple, mais il est généralement vrai que l'effet confondant de la corrélation entre les variables prédictives rend les coefficients de régression standardisés inexplicables en termes de poids relatifs.

4.2 Processus de la méthode

Tout d'abord, l'idée de la méthode est que les poids relatifs d'un ensemble de variables peuvent être approchés en créant un ensemble de variables qui sont fortement corrélées avec les variables originales, mais non corrélées entre elles. La variable dépendante peut être régressée sur les nouvelles variables non corrélées afin d'approximer les poids relatifs des variables originales.

Supposons qu'il existe un vecteur $n \times 1 : y$, ce qui est notre cible, et une matrice $n \times p$ de rang n : X. En supposant que X a au moins autant de lignes que de colonnes, il est possible de trouver une décomposition en valeurs singulières de X :

$$X = P \Delta Q'$$

où P contient les vecteurs propres de XX' , Q contient les vecteurs propres de $X'X$ et Δ est une matrice diagonale contenant les valeurs singulières de X.

Cette méthode nous propose que la meilleure approximation orthogonale de base originale est donnée par :

$$Z = PQ'$$

En effet, les variables orthogonales ainsi créées minimisent la somme des carrés résiduels entre les variables originales et orthogonales. Dans ce cas, lorsque nous laissons y régresser sur Z , le vecteur des poids de Z par rapport à y est donné par la formule suivante :

$$\begin{aligned}\boldsymbol{\beta}^* &= (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y} = (\mathbf{Q}\mathbf{P}'\mathbf{P}\mathbf{Q}')^{-1} \mathbf{Q}\mathbf{P}'\mathbf{y} = \mathbf{I}^{-1}\mathbf{Q}\mathbf{P}'\mathbf{y} \\ &= \mathbf{Q}\mathbf{P}'\mathbf{y}.\end{aligned}$$

Avec les expressions des poids relatifs aux variables orthogonales, nous devons dériver les expressions des poids relatifs aux variables originales par une transformation linéaire entre les variables orthogonales et les variables originales. Il nous faut régressions Z à X :

$$\begin{aligned}\boldsymbol{\Gamma} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} = (\mathbf{Q}\Delta\mathbf{P}'\mathbf{P}\Delta\mathbf{Q}')^{-1} \mathbf{Q}\Delta\mathbf{P}'\mathbf{P}\mathbf{Q}' = (\mathbf{Q}\Delta^2\mathbf{Q}')^{-1} \mathbf{Q}\Delta\mathbf{Q}' \\ &= \mathbf{Q}\Delta^{-1}\mathbf{Q}',\end{aligned}$$

où $\boldsymbol{\Gamma}$ est la matrice des poids de régression de Z sur X

De plus, nous avons aussi la matrice des poids de régression de X sur Z . C'est à dire, au lieu de régresser Z à X , nous régressons X à Z :

$$\begin{aligned}\boldsymbol{\Lambda}^* &= (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X} = (\mathbf{Q}\mathbf{P}'\mathbf{P}\mathbf{Q}')^{-1} \mathbf{Q}\mathbf{P}'\mathbf{P}\Delta\mathbf{Q}' = \mathbf{I}^{-1}\mathbf{Q}\Delta\mathbf{Q}' \\ &= \mathbf{Q}\Delta\mathbf{Q}'\end{aligned}$$

D'après tout ce que nous avons, nous pouvons alors déduire la nouvelle mesure de poids relatives de X à Y . La mesure que nous utilisons est que la proportion de la contribution d'un x_j particulier à y peut être estimée en multipliant la proportion de la variance représentée par x_j pour chaque z_k , c'est β_k^{*2} , k ième élément de matrice $\boldsymbol{\beta}^*$, et en additionnant les produits.

En utilisant la notation $\boldsymbol{\Lambda}^{*[2]} = \|\lambda_{jk}^{*2}\|$ et $\boldsymbol{\beta}^{*[2]} = \|\beta_k^{*2}\|$, le nouveau vecteur de poids relatifs $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)'$, est exprimé comme la formule suivante :

$$\boldsymbol{\varepsilon} = \boldsymbol{\Lambda}^{*[2]}\boldsymbol{\beta}^{*[2]}$$

La somme des poids relatifs est à son tour égale à la corrélation multiple au carré du modèle, de sorte que les poids relatifs finaux peuvent être exprimés en pourcentage de la variance prévisible associée à x_j en divisant ε_j par R^2 et en multipliant par 100

4.3 Application de fonction standard rwa() en R

Il existe une fonction standard en R qui s'appelle rwa(), qui nous permet de progresser le même processus Relative Importance Analysis, ici nous décidons de grouper les variables en 3 groupe en raison qu'il existe une corrélation très forte entre eux, les 3 groupe que nous décidons de grouper c'est taux d'infection par régions, et deuxième taux de vaccination par régions et le troisième groupe pour le reste.

4.4 Groupe 1

```
$rsquare  
[1] 0.9204963
```

```
$result  
variables Raw.Relweight Rescaled.RelWeight Sign  
1 taux_infection_region1 0.08665581 9.414032 +  
2 taux_infection_region2 0.09941184 10.799809 +  
3 taux_infection_region3 0.09772156 10.616182 +  
4 taux_infection_region4 0.11188556 12.154917 +  
5 taux_infection_region5 0.10521563 11.430315 +  
6 taux_infection_region6 0.10115045 10.988686 +  
7 taux_infection_region7 0.10676086 11.598185 +  
8 taux_infection_region8 0.10666283 11.587535 +  
9 taux_infection_region9 0.10503175 11.410340 +
```

FIGURE 10 – Résultat du groupe 1

Voici le résultat du groupe 1, la colonne « Rescaled.RelWeight » s'explique la contribution de variables à notre cible : taux d'infection nationale, ici, nous voyons que les variables taux d'infection par région ont tous une contribution positive avec taux d'infection nationale, de plus les valeur de « Rescaled.RelWeight » sont comparables. Nous pouvons dire que les taux d'infection par région contribuent toutes à peu près au même niveau au cible. la petite différence est peut-être que la densité de population de chaque région n'est pas même ce qui influence la vitesse de propagation de virus.

4.5 Groupe 2

```
$rsquare  
[1] 0.9684849
```

```
$result  
variables Raw.Relweight Rescaled.RelWeight Sign  
1 taux_vaccination_region1 0.11980786 12.37065 -  
2 taux_vaccination_region2 0.10311483 10.64702 -  
3 taux_vaccination_region3 0.09934609 10.25789 -  
4 taux_vaccination_region4 0.10168553 10.49944 -  
5 taux_vaccination_region5 0.10937568 11.29348 -  
6 taux_vaccination_region6 0.10725666 11.07469 -  
7 taux_vaccination_region7 0.10886814 11.24108 -  
8 taux_vaccination_region8 0.10979387 11.33666 -  
9 taux_vaccination_region9 0.10923629 11.27909 -
```

FIGURE 11 – Résultat du groupe 2

Voici le résultat du group 2, nous pouvons trouver que toutes les variables ont une contribution négative au cible, de plus les valeur de « Rescaled.RelWeight » sont comparables. Nous pouvons dire que taux de

vaccination par régions contribuent toutes à peu près au même niveau au cible. Et la petite différence est peut-être que la densité de population de chaque région n'est pas même ce qui influence la vitesse de propagation de virus.

4.6 Groupe 3

\$rsquare				
[1] 0.9462502				
\$result				
	variables	Raw.Relweight	Rescaled.Relweight	sign
1	temperature	0.06956312	7.351451	-
2	taux_infection_age1	0.24508806	25.900979	+
3	taux_infection_age2	0.24711002	26.114660	+
4	taux_vaccination_age1	0.15685437	16.576416	-
5	taux_vaccination_age2	0.10601371	11.203561	-
6	temps	0.05464279	5.774666	-
7	politique	0.06697811	7.078266	-

FIGURE 12 – Résultat du groupe 3

D'après le résultat du group 3, nous trouvons que sauf taux infection par âge toutes les autres variables contribue négativement pour cible, mais par contre on trouve que taux infection par âge influence le plus au cible du groupe 3, pour les autres variables, Nous constatons que le temps a un effet négatif sur le taux d'infection, ce qui peut être lié à la diffusion de la vaccination et à la production d'anticorps dans la population. Étonnamment, le degré d'assouplissement des politiques a également un effet négatif sur le taux d'infection, peut-être parce que la diminution des taux d'infection a entraîné un assouplissement des politiques, ou parce que l'assouplissement des politiques a permis aux gens de se sentir mieux, ce qui a amélioré le système cardiovasculaire et renforcé le système immunitaire, car il existe une forte corrélation entre le bien-être émotionnel et la fonction cardiovasculaire et le système immunitaire. Nous avons également constaté qu'il existe une corrélation négative entre la température et le taux de transmission. Peut-être que l'augmentation de la température réduit le taux de transmission du virus, ou peut-être que c'est simplement l'augmentation de la température dans le temps qui crée l'illusion que l'augmentation de la température réduit le taux de transmission, sans se rendre compte que c'est juste l'effet du temps.

4.7 Mini conclusion

Nous concluons que les les variables vaccination et infection par classement région contribuent toutes à peu près au même niveau au cible, la raison est peut-être que la différence de densité de population, et de plus, nous trouvons que le nombre de personnes vaccinées , la tranche d'âge de la population vaccinée , et aussi la température influence négativement le taux infection nationale, et de plus , les jeunes sont vaccinés donne une influence négative a l'infection nationale.

5 ANOVA à 1 facteur

Dans cette session, nous avons voulu savoir si les taux d'infection et les taux de vaccination étaient affectés par la politique gouvernementale.C'est à dire, nous allons faire les comparaison de moyennes des

variables taux d'infection et taux vaccination dans différents groupe de variable politique. Nous avons choisi le logiciel SPSS (version 21.0) pour réaliser notre ANOVA à 1 facteur.

En raison que la variable politique possède plus que 2 niveaus, donc nous avons plus que 2 groupes à comparer, donc c'est impossible d'utiliser le test de student pour comparer leur moyennes, donc dans notre cas, nous allons procéder la méthode ANOVA à 1 facteur.

L'analyse de la variance à 1 facteur est une méthode statistique qui nous permet de comparer une moyenne d'une variable aléatoire quantitative dans différents groupes, avec cette méthode , nous sommes capable de trouver s'il y a des différences significatives d'une variables quantitatives dans différents niveaux de groupe, et cette méthode possédant au moins 3 niveaux de groupes.

5.1 Groupe 1 : ANOVA à 1 facteur pour les taux infection régionaux dans différents groupes politique

5.1.1 Conditions d'application

Indépendance Chaque échantillon est composé d'observations indépendantes, les échantillons sont indépendants, et nos variables sont de variables continues.

Normalité

Test de Kolmogorov-Smirnov à un échantillon									
	taux_infection_region1	taux_infection_region2	taux_infection_region3	taux_infection_region4	taux_infection_region5	taux_infection_region6	taux_infection_region7	taux_infection_region8	taux_infection_region9
N	24	24	24	24	24	24	24	24	24
Paramètres normaux ^{a,b}	Moyenne 4.998399771	4.893241729	6.059006279	4.664248755	6.053274747	5.390242938	4.402293241	5.757589852	3.850780976
	Ecart-type 1.633358540	1.339831151	1.910515044	1.928683301	2.200662417	2.395197080	1.684273678	1.903851821	1.225795556
Differences les plus extrêmes	Absolue .317	.181	.230	.167	.202	.340	.154	.172	.176
	Positive .210	.085	.108	.106	.133	.193	.126	.155	.143
	Négative .317	-.181	-.230	-.167	-.202	-.340	-.154	-.172	-.176
Z de Kolmogorov-Smirnov	1.551	.888	1.126	.819	.990	1.665	.755	.842	.864
Signification asymptotique (bilatérale)	.016	.410	.158	.513	.281	.008	.619	.477	.445

a. La distribution à tester est gaussienne.

b. Calculée à partir des données.

FIGURE 13 – Test de Kolmogorov-Smirnov à un échantillon pour les taux d'infection régionaux

Après avoir la réalisation un test de Kolmogorov-Smirnov à un échantillon pour ces 9 variables, nous avons trouvé 2 résultat qui ne sont pas significatifs (inférieur à 0.05), nous pouvons en conclure que pour la région 1 et 6 ils sont pas normaux, ici pour les rendre normal, nous allons le faire à l'aide de logiciel SPSS :

Syntaxe SPSS de la fonction : *COMPUTE x_{new} = 1/(max(x) + 1 - x)*

Test de Kolmogorov-Smirnov à un échantillon									
	new_taux_infection_region1	new_taux_infection_region6	taux_infection_region2	taux_infection_region4	taux_infection_region3	taux_infection_region5	taux_infection_region7	taux_infection_region8	taux_infection_region9
N	24	24	24	24	24	24	24	24	24
Paramètres normaux ^{a,b}	.4682	.5095	4.893241729	4.664248755	6.059006279	6.053274747	4.402293241	5.757589852	3.850780976
Ecart-type	.18844	.27312	1.339831151	1.928683301	1.910515044	2.200662417	1.684273678	1.903851821	1.225795556
Déviations les plus extrêmes	.244	.159	.181	.167	.230	.202	.154	.172	.176
Absolue	.124	.159	.085	.106	.108	.133	.126	.155	.143
Positive	.124	.159	.085	.106	.108	.133	.126	.155	.143
Négative	-.244	-.139	-.181	-.167	-.230	-.202	-.154	-.172	-.176
Z de Kolmogorov-Smirnov	1.195	.779	.888	.819	1.126	.990	.755	.842	.864
Signification asymptotique (bilatérale)	.115	.578	.410	.513	.158	.281	.619	.477	.445

a. La distribution à tester est gaussienne.

b. Calculée à partir des données.

FIGURE 14 – Test de Kolmogorov-Smirnov après normalisation

Après avoir normalisé, nous avons tout les variables qui ont une valeur significatif, donc nous pouvons commencer analyse ANOVA à 1 facteur.

5.1.2 Processus d'ANOVA à 1 facteur pour Taux d'infection régionaux dans différents groupes de variable politique en SPSS

Test d'homogénéité des variances

Test d'homogénéité des variances				
	Statistique de Levene	ddl1	ddl2	Signification
new_taux_infection_region1	1.612	2	21	.223
taux_infection_region2	2.869	2	21	.079
taux_infection_region3	7.708	2	21	.003
taux_infection_region4	.648	2	21	.533
taux_infection_region5	1.315	2	21	.290
new_taux_infection_region6	2.298	2	21	.125
taux_infection_region7	.757	2	21	.481
taux_infection_region8	.267	2	21	.768
taux_infection_region9	.354	2	21	.706

FIGURE 15 – Test des homogénéité

D'après le tableau du test d'homogénéité des variances,nous avons trouvé que presque tout les variables ont une signification grande (supérieure à 0.05), sauf le 3ème région,c'est à dire, il n'y a pas de différence significative pour taux d'infection de 3ème région dans 3 différents niveaux de politique.Mais par contre, pour les 8 autres variables, ils ont tous une différence significative dans 3 différents niveaux de politique.

Résultat d'ANOVA

ANOVA à 1 facteur

		Somme des carrés	ddl	Moyenne des carrés	F	Signification
new_taux_infection_region1	Inter-groupes	.515	2	.257	17.897	.000
	Intra-groupes	.302	21	.014		
	Total	.817	23			
taux_infection_region2	Inter-groupes	30.292	2	15.146	28.924	.000
	Intra-groupes	10.997	21	.524		
	Total	41.288	23			
taux_infection_region3	Inter-groupes	63.180	2	31.590	31.938	.000
	Intra-groupes	20.771	21	.989		
	Total	83.952	23			
taux_infection_region4	Inter-groupes	59.334	2	29.667	23.760	.000
	Intra-groupes	26.222	21	1.249		
	Total	85.556	23			
taux_infection_region5	Inter-groupes	84.305	2	42.153	32.686	.000
	Intra-groupes	27.082	21	1.290		
	Total	111.387	23			
new_taux_infection_region6	Inter-groupes	.854	2	.427	10.401	.001
	Intra-groupes	.862	21	.041		
	Total	1.716	23			
taux_infection_region7	Inter-groupes	44.989	2	22.494	23.319	.000
	Intra-groupes	20.257	21	.965		
	Total	65.246	23			
taux_infection_region8	Inter-groupes	60.597	2	30.299	27.944	.000
	Intra-groupes	22.770	21	1.084		
	Total	83.367	23			
taux_infection_region9	Inter-groupes	25.662	2	12.831	30.286	.000
	Intra-groupes	8.897	21	.424		
	Total	34.559	23			

FIGURE 16 – ANOVA à 1 facteur

D'après le résultat de ANOVA, nous avons trouvé que toutes les valeurs ont une signification inférieure à 0.05, donc nous pouvons dire qu'il existe au moins une moyenne de ces trois niveaux de politique est différente des autres.

Mais par contre, nous ne savons qu'il existe des différences, mais nous ne savons pas qui diffère de qui, donc ici d'autres comparaisons multiples sont nécessaires.les taux d'infection régionaux se varie au mesure de changement de politique, mais ici nous ne savons pas quelle période a un taux d'infection plus fort que l'autre période, donc c'est nécessaire de procéder des comparaisons multiples des moyennes.

Comparaisons multiples Ici nous passons à l'étape comparaisons multiples d'où les variances de 3 niveau de politique sont significativement égales pour la variable taux d'infection région 3, ici nous allons comparer leurs moyennes.

Nous avons choisi la méthode LSD (Least—SignificantDifference) pour les comparaisons multiples. C'est une méthode le plus sensible qui nous permet de comparer s'il y a une différence significative pour les moyennes. Mais pour les 8 autres régions, nous allons utiliser la méthode Tamhane pour les comparaisons multiples.

Notre objectif est de savoir quelle groupe de politique a une moyenne plus petite en taux d'infection pour les 9 régions avec comparaisons multiples.

Ici pour les comparaisons multiples, on utilise soit la méthode LSD soit la méthode Tamhane, si la signification est inférieure à 0,05 alors il existe une différence significative entre ces deux groupes. Et après nous pouvons regarder le signe de la colonne « Différence de moyennes(I-J) », si le signe de cette colonne est positif, alors la moyenne du premier groupe est supérieure à celle du deuxième groupe, sinon la moyenne du deuxième groupe est supérieure à celle du premier groupe.

Les comparaisons multiples de moyennes de X_1^r dans différents groupes P_k

Test d'homogénéité des variances				
	Statistique de Levene	ddl1	ddl2	Signification
new_taux_infection_region1	1.612	2	21	.223

FIGURE 17 – Extrait du tableau Test d'homogénéité des variances pour X_1^r

Comparaisons multiples					
Variable dépendante	(I) politique	(J) politique	Différence de moyennes (I-J)	Erreurs standard	Signification
new_taux_infection_region1	LSD	0	.07932	.05703	.179
n1		2	.40083*	.07022	.000
	1	0	-.07932	.05703	.179
		2	.32151*	.06383	.000
	2	0	-.40083*	.07022	.000
		1	-.32151*	.06383	.000
	Tamhane	0	.07932	.06251	.538
		2	.40083*	.05327	.000
	1	0	-.07932	.06251	.538
		2	.32151*	.04273	.000
	2	0	-.40083*	.05327	.000
		1	-.32151*	.04273	.000

FIGURE 18 – Extrait du tableau de Comparaisons multiples pour X_1^r

Nous passons au tableau comparaisons multiples, comme le valeur de signification est 0.223 ce qui est plus grand que 0.05, donc ici nous utilisons la méthode Tamhane. On trouve que les valeurs de signification de P_0 et P_2 , P_1 et P_2 , P_2 et P_0 , P_2 et P_1 sont plus petit que 0.05, donc nous pouvons dire qu'il existe une différence significatives de moyennes entre groupe P_0 et groupe P_2 , groupe P_1 et groupe P_2 , groupe P_2 et groupe P_0 , groupe P_2 et groupe P_1 . On trouve que la valeur de signification de P_0 et P_1 est 0,538 ce qui est plus grand que 0,05, donc nous pouvons dire qu'il n'y a pas de différences significatives de moyennes entre groupe P_0 et groupe P_1 . Passons à la colonne « Différence de moyennes », nous trouvons que les différences de moyennes de $P_0 - P_2$ et $P_1 - P_2$ sont positives, et les différences de moyennes de $P_2 - P_0$ et $P_2 - P_1$ sont négatives.

Nous pouvons en conclure que la moyenne de groupe P_0 est plus grande que celle de groupe P_2 , et la moyenne de groupe P_1 est plus grande que celle de groupe P_2 .

D'après ce résultat , nous pouvons dire qu'il n'y a pas de différence significative pour le taux d'infection pendant les périodes confinement, déconfinement et couvre-feu.

Diagrammes des moyennes

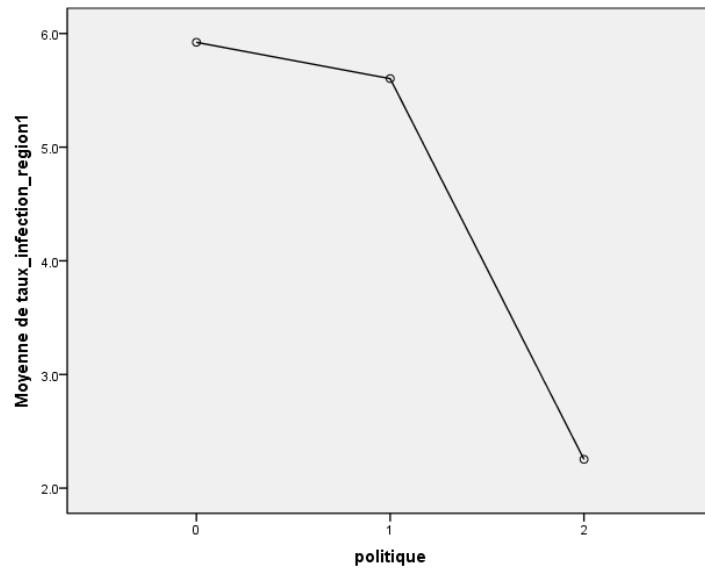


FIGURE 19 – Diagramme des moyennes de X_1^r

D'après ce diagramme des moyennes de X_1^r ,nous pouvons voir que le taux d'infection décroît très rapidement, la raison est peut-être grâce au confinement et couvre-feu.

Diagrammes des moyennes

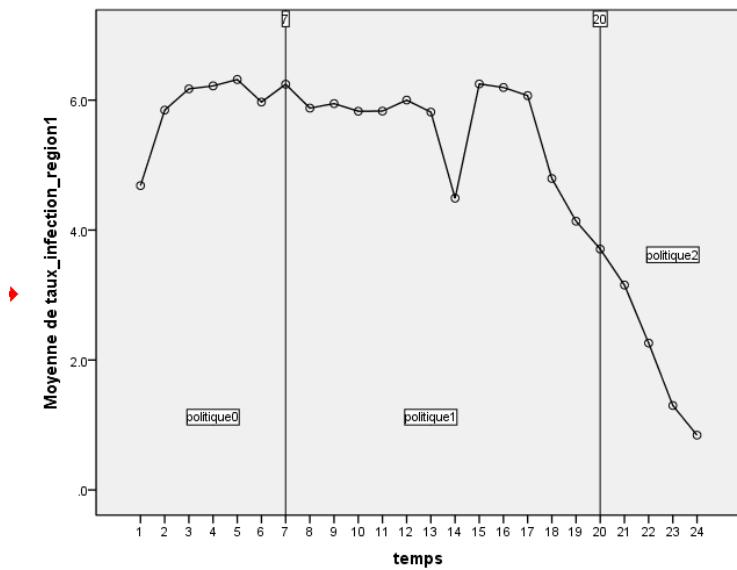


FIGURE 20 – La variation du taux d'infection hebdomadaire du région 1

D'après ce graphe, nous pouvons voir que pendant le moment de politique 1 et politique 2, cela reste presque le même niveau mais quand on passe au politique 3, cela décroît très rapide,c'est la même conclu-

sion que la graphe précédente.

Les comparaisons multiples de moyennes des autres région dans différents groupes P_k

Après avoir fait comparaisons pour la première région, nous passons à la comparaison des 8 autres régions, et voici le résultat des comparaisons multiples et les variations du taux d'infection de ces 9 régions par différentes niveaux politiques et aussi par le temps.

ANOVA à 1 facteur des taux d'infection pendant la période de mise en œuvre des différentes politiques							
		N	Moyenne	Ecart-type	F	Signification	Comparaisons multiples
taux_infection_region2	0	7	5.087144454	0.462020749	28.924	0	0>2,1>2
	1	12	5.66975227	0.697204593			
	2	5	2.758152618	1.045076335			
taux_infection_region3	0	7	6.475304691	0.276351025	31.938	0	0>2,1>2
	1	12	7.115008083	1.115142572			
	2	5	2.941784171	1.287846012			
taux_infection_region4	0	7	5.811166682	1.134524092	23.76	0	0>2,1>2
	1	12	5.257661304	1.228289239			
	2	5	1.634373542	0.689748013			
taux_infection_region5	0	7	7.425384765	0.737569768	32.686	0	0>2,1>2
	1	12	6.757260014	1.304214223			
	2	5	2.442756084	1.129944696			
taux_infection_region7	0	7	5.327775674	0.856619291	23.319	0	0>2,1>2
	1	12	4.967366323	1.123693043			
	2	5	1.750442439	0.700836119			
taux_infection_region8	0	7	6.747760491	0.922627298	27.944	0	0>2,1>2
	1	12	6.466900982	1.013494937			
	2	5	2.669004247	1.261280498			
taux_infection_region9	0	7	4.467562095	0.552719887	30.286	0	0>2,1>2
	1	12	4.329502444	0.657933353			
	2	5	1.838355889	0.758664448			
new_taux_infection_region	0	7	0.5913	0.13123	17.897	0	0>2,1>2
	1	12	0.512	0.13181			
	2	5	0.1905	0.04346			
new_taux_infection_region	0	7	0.6133	0.17384	10.401	0.001	0>2,1>2
	1	12	0.6022	0.24863			
	2	5	0.142	0.01212			

Remarque : 0 représente confinement, 1 représente couvre-feu et 2 représente déconfinement

FIGURE 21 – Tableau résumé des Comparaisons multiples de X_k^r

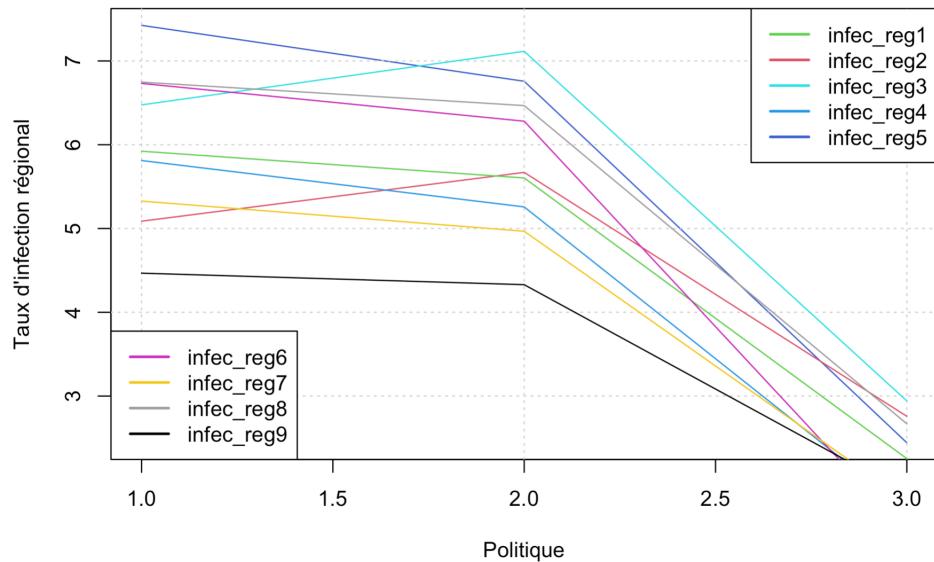


FIGURE 22 – Diagramme des moyennes des X_k^r

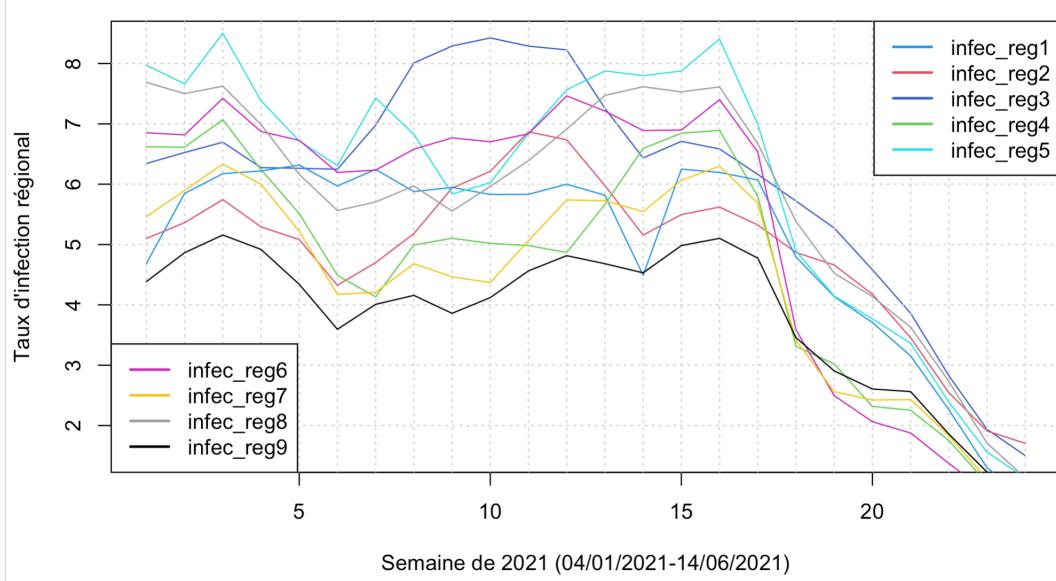


FIGURE 23 – La variation du taux d'infection par semaine

Selon la colonne « Comparaisons multiples », et ces 2 graphes de courbes, nous pouvons trouver une même conclusion que la région 1,c'est à dire, il n'y pas de différence significative entre la période confinement et couvre-feu, mais quand on est arrivé au déconfinement, cela décroît très rapidement avec une signification très significative.Nous pouvons conclure que les actions de gouvernement est mise en œuvre par le tendance de taux .

5.2 Groupe 2 : ANOVA à 1 facteur pour les V_i^r dans différents groupes P_k

5.2.1 Conditions d'application

Indépendance

Ici, pour chaque échantillon est d'observations indépendantes, les échantillons sont indépendants, et les variables sont continues.

Normalité

Test de Kolmogorov-Smirnov à un échantillon										
	taux_vaccination_region1	taux_vaccination_region2	taux_vaccination_region3	taux_vaccination_region4	taux_vaccination_region5	taux_vaccination_region6	taux_vaccination_region7	taux_vaccination_region8	taux_vaccination_region9	
N	24	24	24	24	24	24	24	24	24	24
Paramètres normaux ^{a,b}	Moyenne	13.68033	14.22	14.2517	14.87688	14.16700	14.25388889	14.67736111	14.19	15.00563
	Ecart-type	13.806789	13.731	13.79610	13.936776	13.910321	14.06533429	14.04860379	13.729	14.169777
Differences les plus extrêmes	Absolue	.178	.182	.184	.145	.171	.176	.172	.174	.160
	Positive	.178	.182	.184	.145	.171	.176	.172	.174	.160
	Négative	-.161	-.150	-.151	-.143	-.154	-.155	-.148	-.151	-.145
Z de Kolmogorov-Smirnov		.872	.890	.902	.711	.839	.865	.845	.852	.783
Signification asymptotique (bilatérale)		.433	.407	.390	.693	.482	.443	.473	.462	.573

a. La distribution à tester est gaussienne.

b. Calculée à partir des données.

FIGURE 24 – Test de Kolmogorov-Smirnov à un échantillon des variables V_i^r

D'après le test de Kolmogorov-Smirnov à un échantillon, on voit que toutes les valeurs de signification asymptotique des V_i^r sont plus grande que 0,05, donc on peut dire que les variables V_i^r sont normales. Ensuite, nous pouvons commencer ANOVA à 1 facteur.

5.2.2 Processus d'ANOVA à 1 facteur pour les V_i^r dans différents groupes P_k en SPSS

Test d'homogénéité des variances

Test d'homogénéité des variances

	Statistique de Levene	ddl1	ddl2	Signification
taux_vaccination_region1	5.159	2	21	.015
taux_vaccination_region2	7.652	2	21	.003
taux_vaccination_region3	8.520	2	21	.002
taux_vaccination_region4	7.202	2	21	.004
taux_vaccination_region5	6.603	2	21	.006
taux_vaccination_region6	7.174	2	21	.004
taux_vaccination_region7	5.660	2	21	.011
taux_vaccination_region8	6.026	2	21	.009
taux_vaccination_region9	4.766	2	21	.020

FIGURE 25 – Test des homogénéités

D'après le tableau du test d'homogénéité des variances, nous trouvons que toutes les valeurs de signification plus petit que 0,05. Donc on peut dire que les variances des trois groupes de V_i^r sont significativement

égales.

Résultat d'ANOVA

D'après le résultat de ANOVA, nous avons trouvé que toutes les valeurs ont une signification inférieure à 0.05, donc nous pouvons dire qu'il existe au moins une moyenne de ces trois niveaux de politique est différente des autres.

Mais par contre, nous ne savons qu'il existe des différences, mais nous ne savons pas qui diffère de qui, donc ici d'autres comparaisons multiples sont nécessaires. les taux de vaccination régionaux se varie au mesure de changement de politique, mais ici nous ne savons pas quelle période a un taux de vaccination plus fort que l'autre période, donc c'est nécessaire de procéder des comparaisons multiples des moyennes.

ANOVA à 1 facteur						
		Somme des carrés	ddl	Moyenne des carrés	F	Signification
taux_vaccination_region1	Inter-groupes	3722.608	2	1861.304	59.060	.000
	Intra-groupes	661.823	21	31.515		
	Total	4384.431	23			
taux_vaccination_region2	Inter-groupes	3532.223	2	1766.111	46.109	.000
	Intra-groupes	804.365	21	38.303		
	Total	4336.588	23			
taux_vaccination_region3	Inter-groupes	3527.823	2	1763.911	43.588	.000
	Intra-groupes	849.824	21	40.468		
	Total	4377.647	23			
taux_vaccination_region4	Inter-groupes	3729.969	2	1864.985	53.111	.000
	Intra-groupes	737.406	21	35.115		
	Total	4467.375	23			
taux_vaccination_region5	Inter-groupes	3719.399	2	1859.700	53.423	.000
	Intra-groupes	731.032	21	34.811		
	Total	4450.432	23			
taux_vaccination_region6	Inter-groupes	3764.150	2	1882.075	50.283	.000
	Intra-groupes	786.024	21	37.430		
	Total	4550.173	23			
taux_vaccination_region7	Inter-groupes	3809.611	2	1904.805	54.815	.000
	Intra-groupes	729.744	21	34.750		
	Total	4539.355	23			
taux_vaccination_region8	Inter-groupes	3617.605	2	1808.803	52.934	.000
	Intra-groupes	717.588	21	34.171		
	Total	4335.193	23			
taux_vaccination_region9	Inter-groupes	3929.890	2	1964.945	59.967	.000
	Intra-groupes	688.109	21	32.767		
	Total	4617.999	23			

FIGURE 26 – ANOVA à 1 facteur

Comparaisons multiples

Ici toutes les variances des trois groupes de V_i^r sont significativement égales, donc nous choisissons la méthode LSD (Least—SignificantDifference) pour faire les comparaisons multiples.

Les comparaisons multiples de moyennes de V_1^r dans différents groupes P_k

Test d'homogénéité des variances				
	Statistique de Levene	ddl1	ddl2	Signification
taux_vaccination_region1	5.159	2	21	.015

FIGURE 27 – Extrait du tableau Test d'homogénéité des variances pour V_1^r

Comparaisons multiples					
Variable dépendante	(I) politique	(J) politique	Déférence de moyennes (I-J)	Erreurs standard	Signification
taux_vaccination_region1	LSD	0	-10.938381	2.669922	.001
		2	-35.312914*	3.287138	.000
		1	10.938381*	2.669922	.001
		2	-24.374533*	2.988204	.000
		2	35.312914*	3.287138	.000
		1	24.374533*	2.988204	.000
	Tamhane	0	-10.938381	1.947462	.000
		2	-35.312914*	3.073974	.001
		1	10.938381*	1.947462	.000
		2	-24.374533*	3.568851	.001
		2	35.312914*	3.073974	.001
		1	24.374533*	3.568851	.001

FIGURE 28 – Extrait du tableau de Comparaisons multiples pour V_1^r

Ici nous trouvons que notre valeur de signification est 0,015 ce qui est inférieure à 0,05, donc nous utilisons donc la méthode LSD.

Les valeurs de signification de P_0 et P_1 , P_0 et P_2 , P_1 et P_0 , P_1 et P_2 , P_2 et P_0 , P_2 et P_1 sont inférieures à 0,05, donc on peut dire qu'il y a une différence significative entre groupe P_0 et groupe P_1 , groupe P_0 et groupe P_2 , groupe P_1 et groupe P_2 . Si on passe aux valeurs de la colonne « Différence de moyennes », nous pouvons constater que les différences de moyennes de $P_0 - P_1$, $P_0 - P_2$ et $P_1 - P_2$ sont négatives. Donc nous pouvons en dire que P_1 est plus grande que celle de groupe P_0 , la moyenne de groupe P_2 est plus grande que celle de groupe P_0 et la moyenne de groupe P_2 est plus grande que celle de groupe P_1 .

D'après ce résultat, nous pouvons conclure que le taux de vaccination de région 1 dans la période de déconfinement est plus grande que celui dans la période de couvre-feu, le taux de vaccination de la période de couvre-feu est plus grande que la période de confinement, la raison est peut-être que les mesures gouvernementales se généralisent, entraînant une augmentation des taux de vaccination et une meilleure acceptation des vaccins.

Diagrammes des moyennes

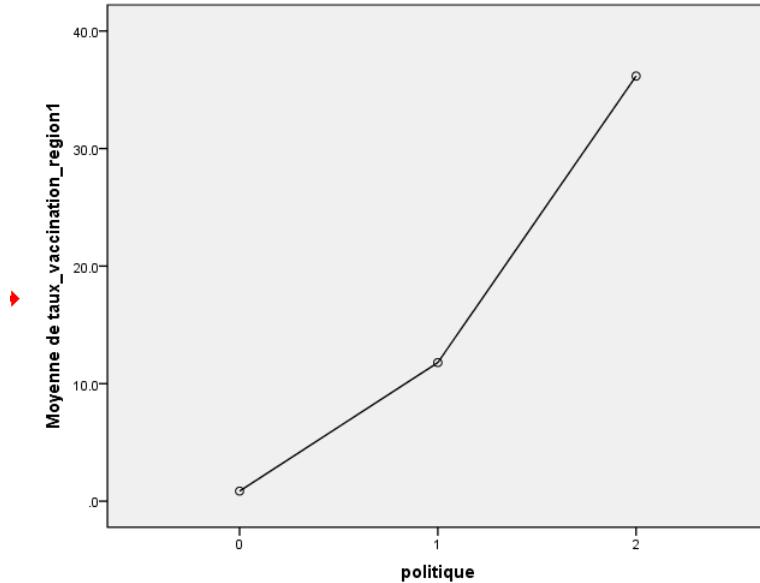


FIGURE 29 – Diagramme des moyennes de V_1^r

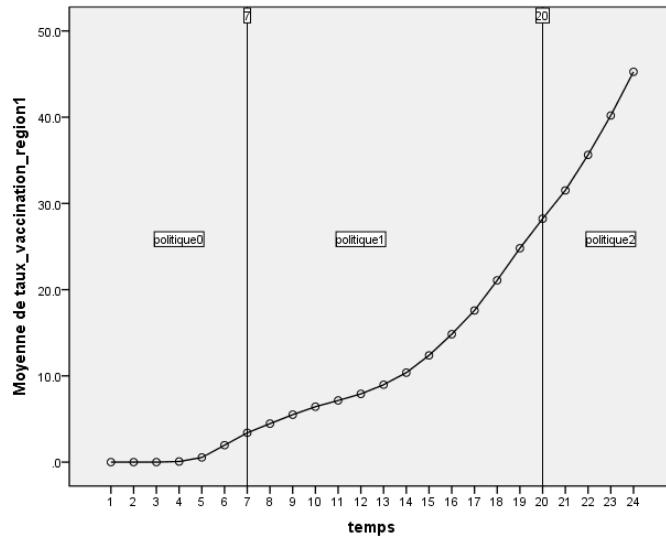


FIGURE 30 – La variation du taux de vaccination hebdomadaire du région 1

D'après ce diagramme ,nous trouvons même résultat que précédentes, c'est à dire le taux de vaccination devient plus en plus grand et croît plus rapidement.

Les comparaisons multiples de moyennes des autres V_i^r dans différents groupes P_k
 Ici, on procède les mêmes processus pour les 8 autres régions, et voici le résultat :

ANOVA à 1 facteur des taux de vaccination pendant la période de mise en œuvre des différentes politiques

		N	Moyenne	Ecart-type	F	Signification	Comparaisons multiples
taux_vaccination_region1	0	7	0.85429	1.329774	59.06	0	0<1, 0<2, 1<2
	1	12	11.79267	6.517664			
	2	5	36.1672	6.781113			
taux_vaccination_region2	0	7	0.86	1.3	46.109	0	0<1, 0<2, 1<2
	1	12	13.12	7.538			
	2	5	35.52	6.504			
taux_vaccination_region3	0	7	0.8229	1.27971	43.588	0	0<1, 0<2, 1<2
	1	12	13.24	7.86753			
	2	5	35.48	6.30714			
taux_vaccination_region4	0	7	0.80786	1.181189	53.111	0	0<1, 0<2, 1<2
	1	12	14.07625	7.319351			
	2	5	36.495	5.91043			
taux_vaccination_region5	0	7	0.78514	1.203368	53.423	0	0<1, 0<2, 1<2
	1	12	12.76433	7.204387			
	2	5	36.268	6.152409			
taux_vaccination_region6	0	7	0.777619048	1.188700994	50.283	0	0<1, 0<2, 1<2
	1	12	12.85527778	7.600164466			
	2	5	36.477333333	5.961505403			
taux_vaccination_region7	0	7	0.975238095	1.429512955	54.815	0	0<1, 0<2, 1<2
	1	12	13.39861111	7.230072332			
	2	5	36.929333333	5.968034428			
taux_vaccination_region8	0	7	0.96	1.403	52.934	0	0<1, 0<2, 1<2
	1	12	12.83	7.031			
	2	5	35.96	6.364			
taux_vaccination_region9	0	7	0.97643	1.475236	59.967	0	0<1, 0<2, 1<2
	1	12	13.8075	7.029027			
	2	5	37.522	5.735232			

Remarque : 0 représente confinement, 1 représente couvre-feu et 2 représente déconfinement

FIGURE 31 – Tableau résumé des Comparaisons multiples de V_k^r

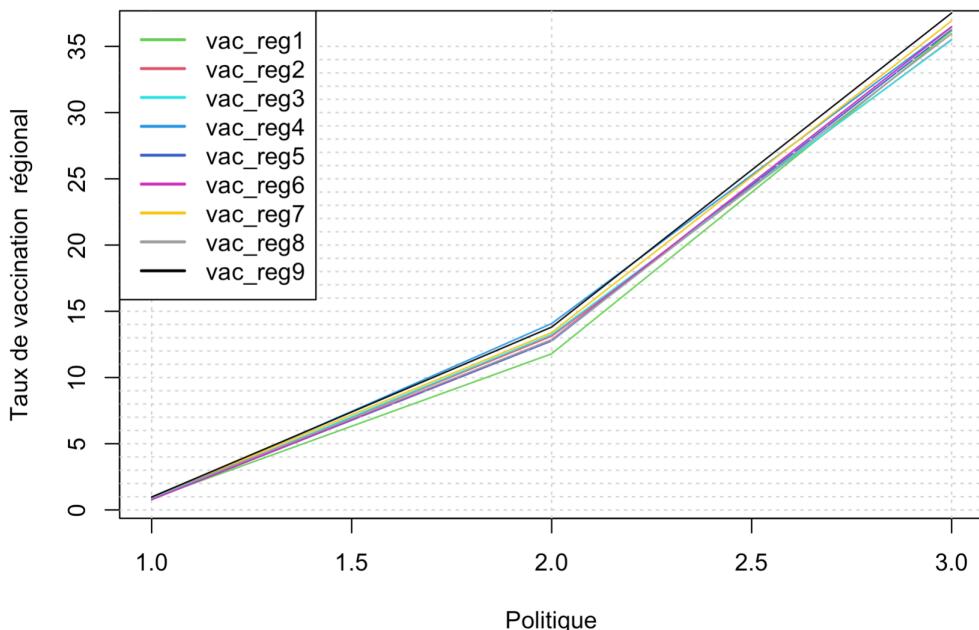


FIGURE 32 – Diagramme des moyennes des V_k^r

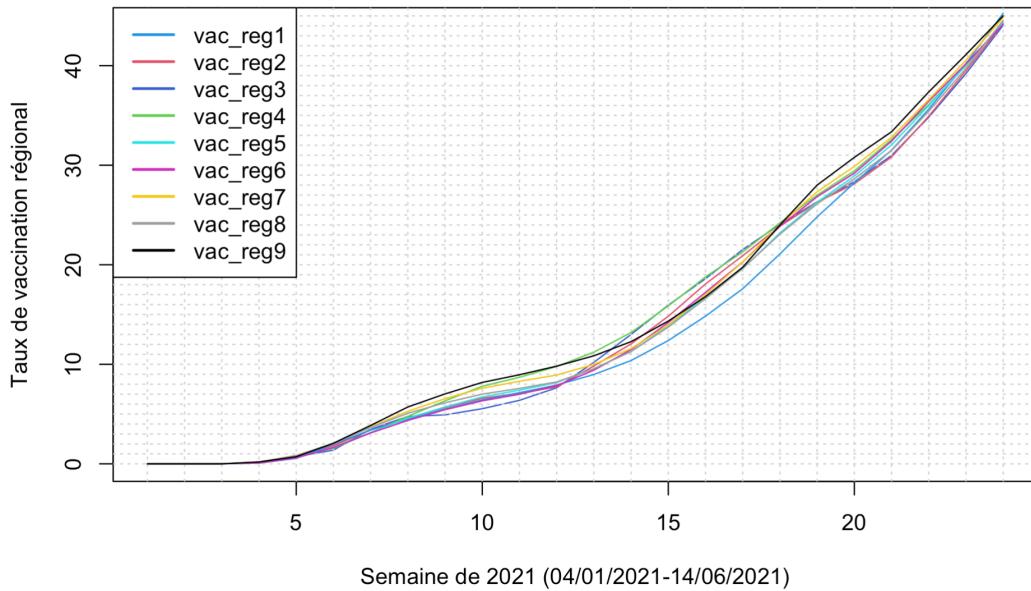


FIGURE 33 – La variation du taux de vaccination par semaine

D’après ce graphe, nous pouvons trouver presque la même résultat que la région 1, nous concluons que le taux de vaccination de toutes les régions dans la période de déconfinement est plus grande que celui dans la période de couvre-feu, le taux de vaccination de toutes les régions dans la période de couvre-feu est plus grande que celui dans la période de confinement, la raison est peut-être que les mesures gouvernementales se généralisent, entraînant une augmentation des taux de vaccination et une meilleure acceptation des vaccins.

6 Conclusion finale

Dans la partie ACP, le taux de vaccination par âge et la température contribuent positivement avec la suppression de la propagation du virus, et le taux d’infection par classement d’âge et le taux d’infection national contribue négativement avec la suppression de la propagation du virus.

Dans la partie RWA, plus le nombre de personnes vaccinées , plus la tranche d’âge de la population vaccinée et aussi plus la température est élevée, plus la suppression de la propagation du virus est fort et de plus . Nous avons trouvé aussi que les jeunes sont vaccinés contribue négativement à la propagation des virus.

Dans la partie ANOVA, nous trouvons que le taux d’infection diminue et le taux de vaccination augmente, et les actions de gouvernement devient plus en plus ouvert au mesure de temps.

7 Références

Présentation de COVID-19 :

https://fr.wikipedia.org/wiki/Maladie_%C3%A0_coronavirus_2019

Données originales :

<https://www.data.gouv.fr/fr/>

Actions du gouvernement :

<https://www.gouvernement.fr/info-coronavirus/les-actions-du-gouvernement>

Chiffres clés de COVID-19 :

<https://news.google.com/covid19/map?hl=frmid=%2Fm%2F0f8l9cgl=FRceid=FR%3Afr%>

Présentation ACP :

<https://www.youtube.com/watch?v=FgakZw6K1QQlist=PLGbayVYnCbodh-unkk0rDS98MdsZelTZLindex=3t=1112s>

Application ACP en SPSS :

<https://www.youtube.com/watch?v=R0oM6BinEE4>

Présentation RWA :

Jeff W. Johnson. 2000. A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression, *Multivariate Behavioral Research*.

Processus RWA :

<https://www.scotttonidandel.com/rwa-web>

Application ANOVA en SPSS :

<https://www.youtube.com/watch?v=768TC9zOvCg>