



Projet Statistiques – CIPA3

Compte rendu d'analyse

Comment les aléas météorologiques affectent-ils l'abstention aux élections présidentielles françaises sous la V^{ème} République ?

MAXIME CHAUVEAU

ALLAN CUEFF

2024

Dépôt du code : https://github.com/allan-cff/projet_stats

INTRODUCTION	3
METHODOLOGIE	3
Provenance des données	3
Données des élections présidentielles de 2002 à 2022	3
Données météorologiques quotidiennes	4
Choix arbitraires	5
Tests statistiques réalisés	5
PROCEDE	6
Extraction des données	6
Données des élections présidentielles de 2002 à 2022	6
Données météorologiques par station météo de 1950 à 2022	6
Réarrangement des données	7
Données des élections	7
Données météorologiques	7
Hypothèses	8
Pluviométrie	8
Température maximale moyenne	8
Régression univariée	8
Tests de corrélation linéaire	9
RESULTATS	10
Influence des précipitations journalières sur l'abstention	10
Influence de la température maximale journalière sur l'abstention	11
CONCLUSION	12
Difficultés rencontrées	12
Critique des résultats	12

Introduction

L'objectif de ce projet est de récupérer, comprendre, et analyser les jeux de données provenant de data.gouv.fr sur les élections présidentielles de 2002 à 2022. Le but est ainsi de réaliser une première étude statistique de ces résultats, en utilisant les notions vues en cours et en utilisant le langage de programmation R.

Lors de nos différentes recherches sur le sujet, nous sommes tombés sur des articles de [BFM](#), du [Figaro](#) et du [Parisien](#) qui certifient que la météo a un impact sur les élections. De plus, une [étude](#) sur le sujet, basée sur les élections présidentielles de 1988 à 2002 a été publiée par deux chercheurs français. Les conclusions de ces trois journaux sont-elles correctes ? Nous avons décidé de mener une étude sur l'impact du climat sur l'abstention aux élections présidentielle de 2002 à 2022.

Pouvons-nous confirmer ou non ce qu'avancent les journaux sur l'impact de la météo sur les abstentions aux présidentielles de 2002 à 2022 ? Est-il possible de prévoir le taux d'abstention à l'avance grâce aux prévisions météorologiques ? Existe-t-il des départements plus touchés que d'autres par l'abstention météorologique ?

Pour répondre à cette question, nous allons comparer les résultats des abstentions par départements en fonction de la météo le jour des élections. Ainsi, essayer de déterminer une corrélation ou non entre les données climatologiques des sondes météo et l'absentéisme des citoyens dans les bureaux de vote.

Méthodologie

Provenance des données

Données des élections présidentielles de 2002 à 2022

Les données que nous avons utilisées pour mener notre étude sont celles les résultats définitifs des élections présidentielles émises par le ministère de l'Intérieur sur le site data.gouv.fr

Les URL utilisées pour télécharger les ressources utilisées sont les suivantes :

- Présidentielles 2002 Tours 1 et 2 :
 - Métadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-2002-resultats-par-bureaux-de-vote/>
 - Données : <https://www.data.gouv.fr/fr/datasets/r/17fbaf03-33f2-48d6-95e8-aaa13809e432>
- Présidentielles 2007 Tours 1 et 2 :
 - Métadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-2007-resultats-par-bureaux-de-vote/>

- Données : <https://www.data.gouv.fr/fr/datasets/r/83e6c566-e313-4dbb-a504-8467c0952697>
- Präsidentielles 2012 Tours 1 et 2 :
 - Mätadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-2012-resultats-par-bureaux-de-vote-1/>
 - Données : <https://www.data.gouv.fr/fr/datasets/r/206b1668-2b31-46da-b957-9857f94fe85c>
- Präsidentielles 2017 Tour 1 :
 - Mätadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-des-23-avril-et-7-mai-2017-resultats-definitifs-du-1er-tour-1/>
 - Données : <https://www.data.gouv.fr/fr/datasets/r/2776519f-a940-46f0-99f4-1a3a1374193b>
- Präsidentielles 2017 Tour 2 :
 - Mätadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-des-23-avril-et-7-mai-2017-resultats-definitifs-du-2nd-tour/>
 - Données : <https://www.data.gouv.fr/fr/datasets/r/0e50ba6a-8175-4455-9e4a-09a7783dc547>
- Präsidentielles 2022 Tour 1 :
 - Mätadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-des-10-et-24-avril-2022-resultats-definitifs-du-1er-tour/>
 - Données : <https://www.data.gouv.fr/fr/datasets/r/18847484-f622-4ccc-baa9-e6b12f749514>
- Präsidentielles 2022 Tour 2 :
 - Mätadonnées : <https://www.data.gouv.fr/fr/datasets/election-presidentielle-des-10-et-24-avril-2022-resultats-definitifs-du-2nd-tour/>
 - Données : <https://www.data.gouv.fr/fr/datasets/r/e7b263e5-bae2-43cc-8944-c8daae6f7ff6>

Données météorologiques quotidiennes

Les données météorologiques proviennent de Météo France. Elles sont disponibles sur data.gouv.fr à l'adresse <https://www.data.gouv.fr/fr/datasets/donnees-climatologiques-de-base-quotidiennes/>, nous les avons téléchargées à l'adresse :

https://object.files.data.gouv.fr/meteofrance/data/synchro_ftp/BASE/QUOT/Q_{NUMERO_DEPARTEMENT}_previous-1950-2022_RR-T-Vent.csv.gz

En remplaçant le {NUMERO_DEPARTEMENT} par le département souhaité.

Choix arbitraires

Nous avons choisi pour cette étude d'analyser uniquement les données de pluviométrie et de température maximale journalière. En effet nous avons accès aux températures minimales, moyennes et maximales journalières pour chaque station de France, seulement, les températures minimales et moyennes reflètent des valeurs prélevées la nuit lorsque les bureaux de vote sont fermés.

Quant à la manière d'agréger ces données, nous allons faire une moyenne des données de toutes les stations de chaque département, afin de comparer ces valeurs moyenne au taux d'abstention aux élections dans ce même département.

Nous avons choisi dix départements de manière aléatoire répartis dans la France entière. Il est possible d'adapter le code pour analyser tous les départements français en s'armant de patience.

Tests statistiques réalisés

Pour nos exploitations de données, on va faire des tests de corrélation entre, dans un premier temps, la pluviosité et le taux d'abstention, puis la température maximale et le taux d'abstention. Le but est de pouvoir affirmer ou réfuter une théorie initiale qui suppose que les citoyens aillent plus ou moins voter en fonction de la météo.

Dans un premier temps, on va déterminer notre hypothèse H_0 , comme étant :

- Il n'y a aucun lien entre la pluviosité le jour du vote et le taux d'abstention.
- Il n'y a aucun lien entre la température maximale le jour du vote et le taux d'abstention.

Ainsi, on pourra poser à l'inverse H_1 , comme étant :

- Il y a un lien entre la pluviosité le jour du vote et le taux d'abstention.
- Il y a un lien entre la température maximale le jour du vote et le taux d'abstention.

Avec les données obtenues, on pourra calculer le coefficient de corrélation linéaire r , qui nous permettra de déterminer le paramètre statistique t , pour chaque département. Ainsi, on pourra déterminer pour quels départements il y a corrélation ou non. Ces calculs nous permettront donc de déterminer si l'abstention a un lien avec la météo.

On pourra aussi essayer d'estimer, avec une incertitude, un pourcentage d'abstention en fonction de la météo sur certains départements, en particulierité sur ceux pour lesquels on aura pu trouver les meilleures corrélations en calculant les coefficients de régression linéaire (**régression univariée uniquement possible si nos abstentions suivent une loi normale, sinon on ne pourra pas prédire de taux d'abstention**).

Procédé

Extraction des données

Données des élections présidentielles de 2002 à 2022

Nous avons remarqué que le téléchargement des données directement dans R avec la fonction `url()` posait des problèmes à la fois de timeout (chargement des données trop lente à l'ISEN) ou de problèmes liés à l'encodage (caractères non reconnus). C'est pourquoi nous avons fait le choix pour ces données de les importer depuis des fichiers sur le disque.

Pour les CSV des élections de 2002 à 2012, il est nécessaire de réécrire les headers dans R

Enfin, pour les élections de 2002 à 2012, les données 'Inscrits' et 'Exprimés' sont dupliquées pour chaque candidat, nous avons donc filtré pour ne garder que les lignes concernant un candidat étant allé au second tour.

```
data2002 <- read.csv(textConnection(c("
N_tour;code_dep;code_comm;nom_comm;N_bureau;Inscrits;Votants;Exprimes;N_cand;
nom_cand;pren_cand;code_cand;voix_cand
", readLines(dataURL2002)[-(0:17)])), sep=";")
%>% filter(N_cand == 5)
data2007 <- read.csv(textConnection(c("
N_tour;code_dep;code_comm;nom_comm;N_bureau;Inscrits;Votants;Exprimes;N_cand;
nom_cand;pren_cand;code_cand;voix_cand
", readLines(dataURL2007)[-(0:17)])), sep=";")
%>% filter(N_cand == 12)
data2012 <- read.csv(textConnection(c("
N_tour;code_dep;code_comm;nom_comm;NA;NA2;N_bureau;Inscrits;Votants;Exprimes;
N_cand;nom_cand;pren_cand;code_cand;voix_cand
", readLines(dataURL2012))), sep=";")
%>% filter(N_cand == 11)
```

Figure 1: Code utilisé pour importer depuis le disque les données des élections présidentielles de 2002 à 2022

Données météorologiques par station météo de 1950 à 2022

Pour les données météorologiques, mieux organisées en CSV que celles des élections, nous avons fait le choix de les télécharger directement avec R afin de ne télécharger que les données concernant les départements sur lesquels nous

travaillerons.

```
for (i in seq_along(departement)) {  
  
  downloadedCSV <- gzcon(url(paste("https://object.files.data.gouv.fr/meteofrance/data/synchro_ftp/BASE/QUOT/",  
                                "Q_", departement[i], "_previous-1950-2022_RR-T-Vent.csv.gz", sep="")))  
  # téléchargement fichier pluie pour un département  
  dat <- read.csv(textConnection(readLines(downloadedCSV)), sep=";") # Lecture du csv  
}
```

Figure 2: Code utilisé pour télécharger les données météo par département dans une boucle for

Réarrangement des données

Données des élections

Nous avons en entrée divers tableaux en fonctions des années, nous appliquons donc des fonctions sur ces tableaux pour les normaliser sous la forme de tableaux à 3 colonnes, « code_dep, Inscrits, Abstentions ». Ces tableaux, un par tour d'élection, sont rangés par ordre chronologique dans une liste. Enfin nous itérons sur cette liste avec nos départements sélectionnés pour ranger les taux d'abstention par département dans une matrice en fonction du département et de la date de vote du tour de l'élection concernée.

```
> print(matrice)  
      [2002-1] [2002-2] [2007-1] [2007-2] [2012-1] [2012-2] [2017-1] [2017-2] [2022-1] [2022-2]  
[44]      0.26      0.18      0.13      0.13      0.16      0.16      0.16      0.22      0.22      0.24  
[29]      0.25      0.17      0.13      0.13      0.16      0.15      0.17      0.21      0.22      0.23  
[61]      0.26      0.18      0.14      0.15      0.17      0.17      0.18      0.22      0.25      0.24  
[33]      0.28      0.18      0.13      0.13      0.17      0.17      0.20      0.23      0.21      0.25  
[13]      0.27      0.19      0.16      0.15      0.20      0.19      0.22      0.27      0.27      0.29  
[69]      0.26      0.19      0.14      0.15      0.18      0.19      0.22      0.25      0.22      0.27  
[75]      0.30      0.17      0.13      0.14      0.20      0.17      0.16      0.22      0.22      0.26  
[59]      0.29      0.22      0.19      0.18      0.22      0.22      0.23      0.27      0.28      0.29  
[37]      0.26      0.19      0.14      0.14      0.17      0.17      0.19      0.23      0.24      0.25  
[67]      0.26      0.20      0.17      0.16      0.20      0.20      0.20      0.24      0.24      0.25
```

Figure 3: Matrice des taux d'abstention aux élections présidentielles en fonction du jour et du département

Données météorologiques

Nous avons en entrée les données de pluviométrie par jour et par station dans chaque département. Nous faisons donc une moyenne de la pluviométrie de toutes les stations du département en omettant celles sans mesure sur la journée de vote. Nous rangeons ensuite ces valeurs moyennes de pluviométrie par jour et par département dans une matrice. Le procédé est le même pour la température max. Il suffit de remplacer l'entête de colonne sélectionnée à « RR » pour la pluviométrie ou « TX » pour la température max.

```
> print(matriceRR)
      [2002-1] [2002-2] [2007-1] [2007-2] [2012-1] [2012-2] [2017-1] [2017-2] [2022-1] [2022-2]
[44]    0.02    0.41    0.04    0.04    6.42    0.11    0.01    0.23    0.09    7.83
[29]    0.01    0.19    0.11    0.02    9.66    3.09    0.04    0.19    0.06    2.18
[61]    0.01    0.46    0.00    0.06    4.01    1.94    0.06    0.03    0.09    0.07
[33]    0.00    0.01    0.01    0.14    3.62    0.23    0.00    0.08    0.05    0.11
[13]    0.00    0.01    0.05    0.10    0.01    1.01    0.02    0.00    0.00    0.68
[69]    0.00    1.53    0.00    0.03    1.01    0.88    0.00    2.38    0.10    8.19
[75]    0.00    1.52    0.09    0.00    2.14    0.11    0.00    0.30    0.03    0.00
[59]    0.00    8.90    0.00    0.96    4.90    0.37    0.02    0.30    0.08    0.02
[37]    0.00    0.06    0.00    0.01    2.19    0.42    0.00    1.03    0.10    6.11
[67]    0.02    0.39    0.00    0.06    2.61    5.14    0.00    4.79    0.09    6.01
```

Figure 4: Vue de la matrice de pluviométrie par jour d'élection par département en mm

```
> print(mat_tmp)
      [2002-1] [2002-2] [2007-1] [2007-2] [2012-1] [2012-2] [2017-1] [2017-2] [2022-1] [2022-2]
[44]    20.28    14.96    24.38    17.14    13.99    15.73    19.09    18.03    16.07    18.81
[29]    18.34    15.02    20.48    16.89    12.58    14.61    17.43    17.14    13.90    15.67
[61]    18.90    11.22    24.64    16.04    12.05    12.00    17.63    16.11    13.71    18.39
[33]    22.36    17.33    26.94    20.13    14.19    18.35    21.52    20.58    19.91    16.91
[13]    21.17    16.26    25.08    23.80    17.48    21.79    18.26    20.93    18.10    16.66
[69]    16.79    9.79    26.10    19.11    14.41    19.05    16.25    15.01    13.51    16.04
[75]    20.19    11.37    25.74    19.61    13.12    16.01    18.70    14.10    14.33    18.98
[59]    19.00    9.55    22.93    17.92    12.31    11.40    15.19    11.72    13.44    18.68
[37]    19.62    13.13    26.70    19.03    14.04    15.62    18.62    17.90    15.47    19.07
[67]    16.31    8.98    23.67    23.41    11.16    15.59    15.81    13.00    11.37    10.04
```

Figure 5: Vue de la matrice de température max moyenne par jour d'élection par département en °C

Hypothèses

Pluviométrie

Pour l'étude du lien entre la pluviométrie journalière moyenne dans le département, nous avons formulé deux hypothèses :

- H0 : La pluviométrie moyenne dans le département le jour du vote n'est pas corrélée au taux d'abstention à ce même vote.
- H1 : La pluviométrie moyenne dans le département le jour du vote est corrélée au taux d'abstention à ce même vote.

Température maximale moyenne

Pour l'étude du lien entre la température maximale moyenne journalière dans le département, nous avons formulé deux hypothèses :

- H0 : La température maximale moyenne dans le département le jour du vote n'est pas corrélée au taux d'abstention à ce même vote.
- H1 : La température maximale moyenne dans le département le jour du vote est corrélée au taux d'abstention à ce même vote.

Régression univariée

Afin d'espérer pouvoir faire la régression univariée, il faut regarder si nos abstentions de nos départements suivent la loi normale. Pour ce faire, on va essayer d'afficher un histogramme :

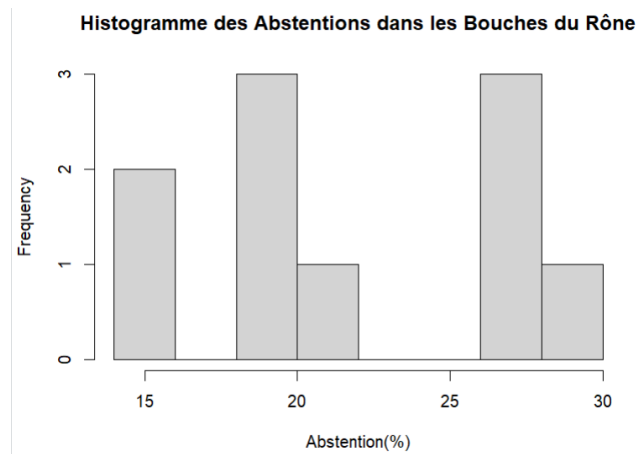


Figure 6 : abstention dans les Bouches du Rhône entre 2002 et 2020

Malheureusement, après avoir testé tous nos départements, aucun ne suit une loi normale. On en conclut donc qu'on ne pourra pas faire la régression univariée. On va donc s'arrêter à l'étude de la corrélation linéaire.

Tests de corrélation linéaire

Une fois qu'on a nos matrices de moyennes de pluie et de températures par département et par élections, on va pouvoir s'attaquer aux calculs de corrélation. Pour ce faire, on va créer des fonctions qui permettront de calculer automatiquement le taux de corrélation et le paramètre statistique. On aura ainsi 10 taux r et 10 paramètres t car 10 départements analysés. On décide de les incorporer dans des vecteurs, pour simplifier l'affichage. Ainsi on obtient :

```
> print(r_RR)
[1] 0.20964558 -0.17793494 -0.30764931 -0.19441340 0.03029630 0.5427412
8 -0.09717123 -0.28592907 0.36375637 0.34833427
> print(t_RR)
[1] 0.60644399 0.51143740 0.91451794 0.56058018 0.08573024 1.82772267 0.27
614858 0.84396440 1.10452483 1.05106595
> print(r_TX)
[1] -0.006448621 -0.078153395 -0.037108943 -0.114311941 -0.492628428 -0.45
3410800 -0.147539730 -0.072689888 -0.199746947 -0.621784780
> print(t_TX)
[1] 0.01823983 0.22172938 0.10503229 0.32545639 1.60112613 1.43883880 0.42
192284 0.20614338 0.57658939 2.24553056
```

On constate ci-dessus les valeurs obtenues sur RStudio pour tous les départements. On va pouvoir maintenant passer à l'analyse de ces données, pour essayer d'en tirer des conclusions sur nos hypothèses.

Résultats

Influence des précipitations journalières sur l'abstention

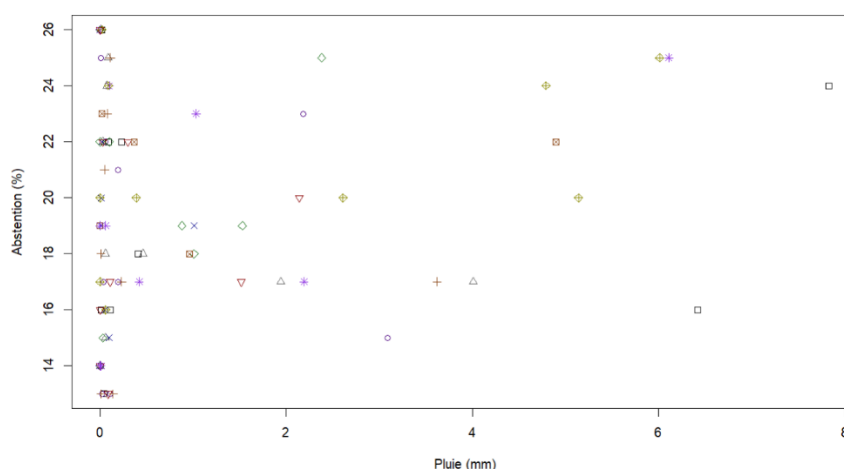


Figure 7 : nuage de points, abstention par département en fonction de la pluie

On commence par calculer la moyenne de nos paramètres paramètre statistique, ça va nous permettre de le comparer avec la valeur théorique dans la table de Student. On a $t_{moy} = 0.78$. On peut ainsi comparer cette valeur avec la table : on a ainsi 0.706, qui correspond à 50%. Cela nous permet de dire qu'à 50%, la moitié des départements ne réfute pas H_0 . On en conclut à 50% que dans la moitié des départements analysés, les citoyens sont influencés par la pluie pour aller voter.

Tableau de l'hypothèse validée par département au risque de 90%

Département	44	29	61	33	13	69	75	59	37	67
Hypothèse validée	H1	H1	H1	H1	H1	H1	H1	H1	H1	H1

Tableau de l'hypothèse validée par département au risque de 50%

Département	44	29	61	33	13	69	75	59	37	67
Hypothèse validée	H0	H0	H1	H0	H1	H1	H0	H1	H1	H1

Tableau de l'hypothèse validée par département au risque de 30%

Département	44	29	61	33	13	69	75	59	37	67
Hypothèse validée	H0	H0	H0	H0	H1	H1	H0	H0	H1	H0

Pour le département des Bouches du Rhône, le paramètre statistique est de 1.899, supérieur à t_α à risque de 10%. En utilisant la table de Student, on peut donc conclure que l'hypothèse H1 est valide pour ce département. On peut donc dire qu'il y a une corrélation entre l'abstention aux présidentielles entre 2002 et 2022 dans les Bouches du Rhône et la pluviométrie du jour du scrutin au risque de **10%**.

Influence de la température maximale journalière sur l'abstention

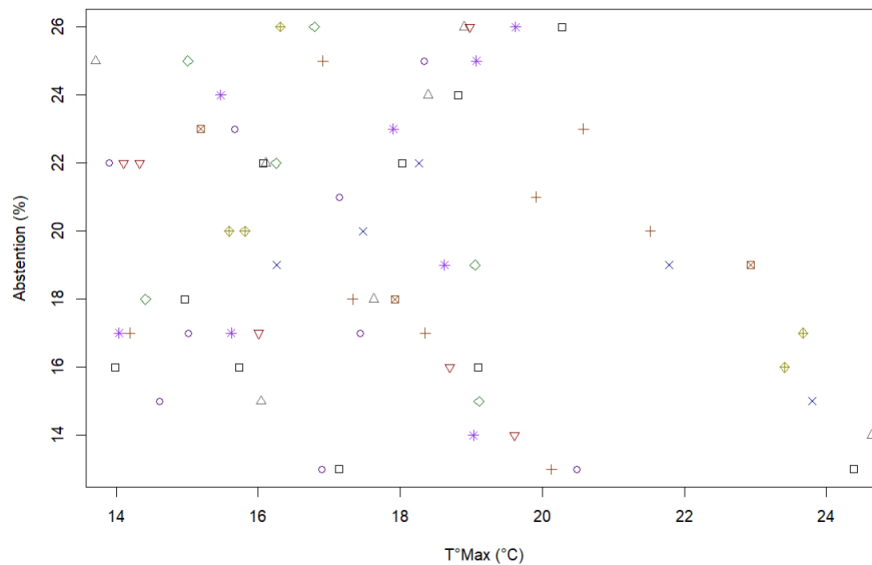


Figure 8 : nuage de points, abstention par département en fonction de la température max

On commence par calculer la moyenne de nos paramètres paramètre statistique. On a $t_{moy} = 0.72$. On peut ainsi comparer cette valeur avec la table : on a ainsi 0,706, qui correspond à 50%. Cela nous permet de dire qu'à 50%, la moitié des départements ne réfute pas H_0 . On en conclut à 50% que dans la moitié des départements analysés, les citoyens sont influencés par les températures maximales journalières.

Tableau de l'hypothèse validée par département au risque de 90%

Département	44	29	61	33	13	69	75	59	37	67
Hypothèse validée	H0	H1	H0	H1	H1	H1	H1	H1	H1	H1

Tableau de l'hypothèse validée par département au risque de 50%

Département	44	29	61	33	13	69	75	59	37	67
Hypothèse validée	H0	H0	H0	H0	H1	H1	H0	H0	H0	H1

Tableau de l'hypothèse validée par département au risque de 20%

Département	44	29	61	33	13	69	75	59	37	67
Hypothèse validée	H0	H0	H0	H0	H1	H1	H0	H0	H0	H1

Tableau de l'hypothèse validée par département au risque de 10%

[illegible]

Pour le département du Bas-Rhin, le paramètre statistique est de 2.246, supérieur à t_{α} à risque de 10%. En utilisant la table de Student, on peut donc conclure que l'hypothèse H1 est valide pour ce département. On peut donc dire qu'il y a une corrélation entre l'abstention aux présidentielles entre 2002 et 2022 dans le Bas-Rhin et la température maximale du jour du scrutin au risque de **10%**.

Conclusion

Avec ce projet, on a pu comprendre qu'on pouvait assez facilement faire dire ce qu'on voulait aux chiffres. En effet, on a vu qu'on pouvait soit affirmer soit réfuter notre corrélation en fonction de la certitude qui variait. Ainsi, pour dire qu'il y a corrélation dans tous les départements analysés, on se place à 90% quand on regarde par rapport à la pluie. Ce n'est cependant pas possible pour la température. A l'inverse, pour affirmer qu'il n'y a aucune corrélation dans l'ensemble des départements analysés, on se place à 5% de risque quand on regarde par rapport à la pluie et 5% de risque par rapport à la température. Cette étude a pu nous permettre d'avoir un esprit critique sur ce que les journaux grand public, et même les études scientifiques, peuvent avancer et d'avoir nos propres conclusions sur l'impact de la météo sur le taux d'abstention aux présidentielles.

Difficultés rencontrées

Le principal frein que nous avons rencontré dans ce projet fut les données à importer venant des csv. En effet, ces derniers n'étant pas formés de la même manière en fonction des élections, il a fallu trouver une solution pour récupérer les données qui nous intéressaient pour chaque élection, en revoyant la méthode d'extraction à chaque fois.

Le croisement des données entre la météo et les abstentions nous ont demandé aussi une petite réflexion, mais le plus compliqué restera la récupération des données qui nous intéressaient.

Critique des résultats

Il reste important de préciser que les valeurs d'absentéisme sont bien évidemment influencées par des facteurs multiples. Ainsi nous avons inclus dans nos études les records d'absentéisme du premier tour de 2002 à 28,4% et les records de participation de 2007 à 83,97%. De plus, nous avons remarqué que la majorité des valeurs de pluviométrie étaient à 0mm les jours de scrutin. Enfin, nous n'avons sélectionné que 10 départements, ce qui ne nous permet pas de déterminer quel est le département le plus influencé par la météo. Une marge de progression possible serait de refaire ces expériences mais cette fois pour tous les départements de France, ce qui nécessite de biens plus grandes capacités de calculs et une très bonne connexion pour récupérer les données.