

TP Info - Élections américaines

12 Décembre 2016

L'objectif de ce TP est de déterminer à quel point les sondages se sont trompés lors des dernières présidentielles américaines (2016, qui ont vu Donald Trump l'emporter à la grande surprise de la plupart des analystes "experts").

On suppose connues les modalités d'élections du président aux États Unis (élections état par état, le gagnant d'un état gagne un certain nombre de grands électeurs, le perdant rien; sauf dans le Maine et Nebraska où l'élection est par *county*. Celui qui a le plus de grands électeurs est élu Président).

On rappelle une règle d'or : plus l'ordi est pourri, plus on sauvegarde régulièrement.

1 Fichiers sources pour l'analyse

Une archive .rar contenant les documents du TP se trouvent ici : http://perso.ens-lyon.fr/maximilien.drevet/Lyce/TP_info/TP_elections_US_fichiers_eleves.rar

Dans le fichier vous avez 4 documents présent en 3 formats : en .ods ou .csv (préférez le .ods, à ouvrir avec un tableur, de préférence LibreOffice ou OpenOffice) et .xls (avec Excel) :

1. *president_general_polls_2016* : fichier contenant les sondages effectués durant la présidentielle américaine.
2. *president_results_2016* : les résultats définitifs États par États.
3. *polls_Florida_20161107* : les résultats des sondages en Floride effectués le 07 Novembre (au cas où vous n'arrivez pas à faire la partie 2.2).
4. *poll_analysis_2016* : la moyenne des résultats des sondages (effectués le 7 Novembre) dans chacun des 50 États américains (pour la partie 3).

1.1 Comprendre le tableur; élaguer les informations inutiles (document *president_general_polls_2016*)

Le tableur comporte 12 625 lignes : ce sont tous les sondages effectués. Toutes ces lignes ne nous intéressent pas. On ne gardera que les sondages les plus récents.

Il y a 27 colonnes. Seul nous intéressent l'état (*state*) où à été fait le sondage, la date de fin du sondage (*enddate*), la taille de l'échantillon (le nombre de personnes sondées, *samplesize*), et les résultats de Clinton et Trump : résultats brut (*rawpoll*) et ajustés (*adjpoll*).

Vous pouvez supprimer toutes les autres colonnes. Elles indiquent l'élection (ici uniquement présidentielle), la date de l'élection, les personnes sondées (*lv* signifie *likely voters*, les personnes que l'on suppose aller voter; aux États Unis les taux de participation sont autour de 50%), la qualité du sondage, les scores d'autres candidats, etc.

2 Analyse dans un des États clé ("swingstate") : la Floride

2.1 Récupérer les sondages récents

Trier les sondages restant par date de fin de sondage (*enddate*), afin de ne garder que les sondages qui se sont terminés le 7 Novembre 2016 (éventuellement aussi le 6 Novembre 2016). Le plus simple est de trier par ordre décroissant selon la colonne "*enddate*".

1. Sélectionner la colonne *enddate*
2. Aller dans Outils/tri
3. **Impérativement** sélectionner la colonne "étendre la sélection".
4. Vérifier les options et faire OK

2.2 Récupérer les sondages en Floride

Vous devriez avoir 324 sondages pour le 7 Novembre. On veut analyser ceux de la Floride. Une méthode pour les récupérer est de trier les sondages par ordre alphabétique de l'état dans lequel ils ont été effectué. Vous devriez avoir 6 sondages fait en Floride le 7 Novembre. Copier ces sondages dans une nouvelle feuille.

Question 2.1. Calculer le pourcentage moyen attribué par les sondages à Clinton et à Trump.

Question 2.2. Comparer avec le pourcentage réel. On admet que les sondages ont une marge d'erreur de 2 à 3%. On dit qu'ils se trompent s'ils sont en dehors de cette marge d'erreur. Se sont-ils trompés en Floride ?

Question 2.3. N'avoir récupéré que les sondages du 7 Novembre est-il suffisant ? Comment améliorerez vous l'analyse qui vient d'être faite ?

3 Analyse dans les autres États

Question 3.1. Comparer les prévisions des sondages (effectués le 7 Novembre) avec les résultats réels pour tous les États. Dans combien d'états les sondages se sont-ils trompés ?

4 Références

1. Le document contenant les sondages provient de FiveThirtyEight, un site spécialisé dans les statistiques. <https://www.r-bloggers.com/fivethirtyeight-polling-data-for-the-us-presidential-election/>
2. Vous pouvez trouver les résultats des élections sur Wikipedia, ou ici : <http://www.presidency.ucsb.edu/showelection.php?year=2016>