PLAN :

CDb01\_c + CDa05\_c + CBb01\_c + SB02\_c + CA02\_c + ABo12\_c

PROBLÉMATIQUE: L‘élection d’Obama a t-elle été déterminée par des facteurs différents de ceux usuels tel que la socio démographie .

INTRO : Facteurs émotionnels importants d’après les études et campagne axée dessus. A l’inverse en littérature classique c’est les facteurs socio démographiques. Études des biais principaux( erreur de mesure, échantillon, causalité inverse)

I) Caractéristiques Socio-Démographiques

Ajout des facteurs socio

II) Caractéristiques "Émotionnelles"/”Subjectives”

1. ajout du party ID\*
   1. (biais, précision, ccl) par rapport au blocs antérieur
   2. étude de nouveau bloc
2. ajout du bloc Ideology \*
3. ajout du bloc Emotional \*

CONCLUSION : Interprétation des impacts de chaque BLOC et conclusion est ce que Obama a gagné grâce aux émotions en ralliant de nouveaux votants

---------------------------------------------------------------------------------------------------------------------------

PARTIE STATISTIQUES

Premier grand bloc (**Objectif**) : Age, Sex, Income, Educ, Race

Second grand bloc (**Subjectif**) : Party id, Bloc : Emotional Response, Bloc : Ideology

Pense voter pour Obama : RCa10 (recodé en 1(yes) ) (Don’t know = 0)

Facteurs d’actualité déterminants :

* ABo12 : Obama share values ( variable ordinale de 0 à 10)
* CA02 : Country going in right direction (recodé en 1(right direction))trop de N(1345)
* SB01 : Trust in black people ( trop de valeurs N(1473)
* CBb01\_c: Taxes
* CDa05 : Favor negotiating w/ enemy trop de valeurs N (2632)
* CDb01 : Withdraw Iraq troops
* ABo05 : Obama est-il un bon leader ( Abo05 à 11 sont des variables très corrélées on a donc choisit une variable parmis ces 6 pour eviter une perte de precision)

Ideology :

* RD01 : Last Vote
* MA04 : Cons-LiberalC
* MA01 : Id Party (Pas recodé (juste enlever don’t know et no answer) )

Caractéristiques Socio-Démographiques:

* WA01\_c: sexe ( recodé en 1(femme)) Pas très utile (moy 50% pour les deux)
* WA02\_c: âge ( variable discrète)
* WA03\_c: éducation ( variable ordinale transformée en indicatrices, grade 8 or lower comme reference) Pas déterminante
* WA04\_c: salaire (trop de valeurs manquantes 1700)Possible biais de selection entre ceux qui ont répondu et ceux qui ne l’ont pas fait
* WC03\_c: ethnie ( 9 valeurs)

Remarques :

---------------------------------------------------------------------------------------------------------------------------

QUESTIONS

Comment peut-on mettre en place des vecteurs de variables pour coder nos modèles ?

→ Codage par bloc a voir +tard

Si 2 groupes de réponses on peut recoder en indicatrice ?

→ A voir mais pas recommandé

Lorsque l’on fait une regression le nombre de valeurs indiquées ne correspond pas aux valeurs manquantes.

Pour l’anova : hypothèse d'homoscédasticité. mais pour nous les groupes définis par X dont souvent d’effectif 1 => variance = 0 ??

Méthode ascendante : Recommandé mais est ce que l’on perdrait pas l’interet du sujet en se retrouvant limité aux variables “classiques” ?? Solution avec Fisher

Difference de significativité importante pour la modalité “blanc” si l’on considère seulement noir ou blanc , ou toutes les ethnies (AIC !)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AB04 | AB12 | CA02 | SB01 | SB02 | CD 05 | CD01b | RD01 | CD01d | MA04 | CB01 | MA01 | WA01 | WA02 | WA03 | WA04 | WC03 |
| P | P | P | P | P | C | C | C | C | C |  | C | C | C | C |  | C |
| :( | :( | :) | :// | :// | :) | :)))) | :) | :(( | :)) |  | :) | :) | :/// | :( |  | :) |

## Bloc 1 (Socio-démographiques) : WA01\_c + WA02\_C + WA03\_c + WA04\_c + WC03\_c

## Bloc 2 (Fixe) : RD01\_c + MA04\_c + MA01\_c

## Bloc 3 (Flex) : CDb01\_c + CDa05\_c + CBb01\_c + SB01\_c + CA02\_c + ABo12\_c

test = subset(Obama,!is.na(Obama$WA01\_c) & !is.na(Obama$WA02\_c)& !is.na(Obama$WA03\_c)& !is.na(Obama$WC03\_c)& !is.na(Obama$RD01\_c)

& !is.na(Obama$MA04\_c)& !is.na(Obama$MA01\_c)& !is.na(Obama$CDb01\_c)& !is.na(Obama$CBb01\_c)& !is.na(Obama$ABo12\_c))

---------------------------------------------------------------------------------------------------------------------------

**Initialisation des données :**

Obama %>% replace\_with\_na(replace = list(ABo04\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(ABo12\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(CA02\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(SB01\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(SB02\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(CDa05\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(CDb01\_c = c("no answer", "don't know", "none of these"))) -> Obama

Obama %>% replace\_with\_na(replace = list(RD01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(CDd01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(MA04\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(CBb01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(MA01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(WA02\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(WA03\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(WA04\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(WC03\_c = c("no answer", "don't know"))) -> Obama

Obama$RD01\_c = recodeVar(Obama$RD01\_c , c("nader") , c("other")) ## on transforme nader en other car il n'est ni républicain ni démocrate

Obama$RCa10 = recodeVar(Obama$RCa10 , c("Yes", "No", "Don't know") , c(1, 0, 0))

Obama$CA02\_c = recodeVar(

Obama$CA02\_c,

c("right direction", "wrong track", "don't know", "no answer"),

c("1", "0", "0", "0")

)

#Obama$WA01\_c = recodeVar(Obama$WA01\_c , c("Female", "Male") , c('1','0'))

Obama$RCa10 = as.integer(Obama$RCa10)

Obama$WC03\_c = recodeVar(

Obama$WC03\_c,

c(unique(Obama$WC03\_c)),

c("black", "white", "other", "other", "other", "other", "other", NA)

)

Obama$CDa05\_c = recodeVar(Obama$CDa05\_c,

c(unique(Obama$CDa05\_c)),

c("oppose", "neutral", "favor", "oppose", "favor", NA))

Obama$CDd01\_c = recodeVar(Obama$CDd01\_c,

c(unique(Obama$CDd01\_c)),

c("favor", "oppose", "oppose", "favor", NA))

Obama$MA04\_c = recodeVar(

Obama$MA04\_c,

c(unique(Obama$MA04\_c)),

c(

"moderate",

"liberal",

"liberal",

"conservative",

"conservative",

NA

)

)

Obama$WA03\_c = recodeVar(

Obama$WA03\_c,

c(unique(Obama$WA03\_c)),

c(

"Some education",

"Lot of education",

"Some education",

"Lot of education",

"No education",

"No education",

"Some education",

"Lot of education",

"Some education",

NA

)

)

data1$ABo12\_c = recodeVar(data1$ABo12\_c , c('0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10') , c('No', 'No', 'No', 'No', 'No', 'somewhat', 'somewhat', 'yes','yes', 'yes', 'yes'))

data1$SB02\_c = recodeVar(data1$SB02\_c , c('0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10') , c('No', 'No', 'No', 'No', 'No', 'somewhat', 'somewhat', 'yes','yes', 'yes', 'yes'))

data1$SB01\_c = recodeVar(data1$SB01\_c , c('0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10') , c('No', 'No', 'No', 'No', 'No', 'somewhat', 'somewhat', 'yes','yes', 'yes', 'yes'))

data1$ABo04\_c = recodeVar(data1$ABo04\_c , c('very conservative', 'very liberal'), c('somewhat conservative', 'somewhat liberal'))

data1$WA04\_c = recodeVar(data1$WA04\_c, c(unique(data1$WA04\_c)),c("li","hi","mi","mi","hi","li","mi",NA,"hi","li"))

## Factor-isation

data1$ABo04\_c <- as.factor(data1$ABo04\_c)

data1$ABo12\_c <- as.factor(data1$ABo12\_c)

data1$ABo12\_c =relevel(data1$ABo12\_c, "yes")

data1$SB01\_c <- as.factor(data1$SB01\_c)

data1$ABo12\_c <- as.factor(data1$ABo12\_c)

data1$CA02\_c <- as.factor(data1$CA02\_c)

data1$CA02\_c = relevel(data1$CA02\_c, "wrong track" )

data1$SB02\_c <- as.factor(data1$SB02\_c)

data1$SB02\_c =relevel(data1$SB02\_c, "yes")

Obama$RD01\_c = as.factor(Obama$RD01\_c)

Obama$RD01\_c = relevel(Obama$RD01\_c, "kerry")

Obama$CDb01\_c <- as.factor(Obama$CDb01\_c)

Obama$CDb01\_c <-

relevel(Obama$CDb01\_c, "withdraw as soon as possible")

Obama$CDd01\_c <- as.factor(Obama$CDd01\_c)

Obama$CDd01\_c <- relevel(Obama$CDd01\_c, "favor")

Obama$MA04\_c <- as.factor(Obama$MA04\_c)

Obama$MA04\_c <- relevel(Obama$MA04\_c, "liberal")

Obama$WA03\_c <- as.factor(Obama$WA03\_c)

Obama$WA03\_c <- relevel(Obama$WA03\_c, "No education")

################################################################################

CODE PABLO :

################################################################################

rm(list=ls())

setwd("~/LICENCE 3/SEMESTRE 2/PROJET OBAMA")

data1 <- read.csv("Obama\_data.csv", sep= ",", header= TRUE)

unique(data1$WC03\_c) ## regarde les valeurs de la variable

library(doBy)

library(naniar)

data1$RCa10 = recodeVar(data1$RCa10 , c("Yes", "No", "Don't know") , c('1','0','0'))

data1%>% replace\_with\_na(replace = list(ABo04\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(ABo12\_c = c( 998, 999))) -> data1

data1%>% replace\_with\_na(replace = list(CA02\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(SB01\_c = c( 998, 999))) -> data1

data1%>% replace\_with\_na(replace = list(SB02\_c = c( 998, 999))) -> data1

data1%>% replace\_with\_na(replace = list(CDa05\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(CDb01\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(RD01\_c = c("no answer", "don't know"))) -> data1

data1$RD01\_c = recodeVar(data1$RD01\_c , c("nader") , c("other")) ## on transforme nader en other car il n'est ni républicain ni démocrate

data1%>% replace\_with\_na(replace = list(CDd01\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(MA04\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(CBb01\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(MA01\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(WA02\_c = c(998, 999))) -> data1

data1%>% replace\_with\_na(replace = list(WA03\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(WA04\_c = c("no answer", "don't know"))) -> data1

data1%>% replace\_with\_na(replace = list(WC03\_c = c("no answer", "don't know"))) -> data1

##Regression Obama cons-liberal

table(data1$ABo04\_c)

data1$ABo04\_c <- as.factor(data1$ABo04\_c)

summary(reg <- lm(RCa10 ~ ABo04\_c, data1))

##Regression Obama share values

table(data1$ABo12\_c)

data1$ABo12\_c <- as.factor(data1$ABo12\_c)

data1$ABo12\_c =relevel(data1$ABo12\_c, "yes")

summary(reg <- lm(RCa10 ~ ABo12\_c, data1))

##Regression Trust in blacks

table(data1$SB01\_c)

data1$SB01\_c <- as.factor(data1$SB01\_c)

summary(reg <- lm(RCa10 ~ SB01\_c, data1))

##Regression Trust in whities

table(data1$SB02\_c)

data1$ABo12\_c <- as.factor(data1$ABo12\_c)

summary(reg <- lm(RCa10 ~ ABo12\_c, data1))

## Regression country going in right direction

unique(data1$CA02\_c)

data1$CA02\_c <- as.factor(data1$CA02\_c)

data1$CA02\_c = relevel(data1$CA02\_c, "wrong track" )

summary(reg <- lm(RCa10 ~ CA02\_c, data1))

##Régression de salaire

unique(data1$WA04\_c)

data1$WA04\_c = recodeVar(data1$WA04\_c, c(unique(data1$WA04\_c)),c("li","hi","mi","mi","hi","li","mi",NA,"hi","li"))

data1$WA04\_c = as.factor(data1$WA04\_c)

table(data1$WA04\_c)

data1$WA04\_c = relevel(data1$WA04\_c,"li")

summary(lm(RCa10 ~ WA04\_c,data1 ))

CODE CYRIAC :

## Mise en place des données

Obama\_data <-

read.csv("C:/Users/cyria/Downloads/Obama\_data.csv",

sep = ",",

header = TRUE)

Obama <- Obama\_data

Obama %>% replace\_with\_na(replace = list(ABo04\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(ABo12\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(CA02\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(SB01\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(SB02\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(CDa05\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(CDb01\_c = c("no answer", "don't know", "none of these"))) -> Obama

Obama %>% replace\_with\_na(replace = list(RD01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(CDd01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(MA04\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(CBb01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(MA01\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(WA02\_c = c(998, 999))) -> Obama

Obama %>% replace\_with\_na(replace = list(WA03\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(WA04\_c = c("no answer", "don't know"))) -> Obama

Obama %>% replace\_with\_na(replace = list(WC03\_c = c("no answer", "don't know"))) -> Obama

Obama$RD01\_c = recodeVar(Obama$RD01\_c , c("nader") , c("other")) ## on transforme nader en other car il n'est ni républicain ni démocrate

Obama$RCa10 = recodeVar(Obama$RCa10 , c("Yes", "No", "Don't know") , c(1, 0, 0))

Obama$CA02\_c = recodeVar(

Obama$CA02\_c,

c("right direction", "wrong track", "don't know", "no answer"),

c("1", "0", "0", "0")

)

#Obama$WA01\_c = recodeVar(Obama$WA01\_c , c("Female", "Male") , c('1','0'))

Obama$RCa10 = as.integer(Obama$RCa10)

Obama$WC03\_c = recodeVar(

Obama$WC03\_c,

c(unique(Obama$WC03\_c)),

c("black", "white", "other", "other", "other", "other", "other", NA)

)

Obama$CDa05\_c = recodeVar(Obama$CDa05\_c,

c(unique(Obama$CDa05\_c)),

c("oppose", "neutral", "favor", "oppose", "favor", NA))

Obama$CDd01\_c = recodeVar(Obama$CDd01\_c,

c(unique(Obama$CDd01\_c)),

c("favor", "oppose", "oppose", "favor", NA))

Obama$MA04\_c = recodeVar(

Obama$MA04\_c,

c(unique(Obama$MA04\_c)),

c(

"moderate",

"liberal",

"liberal",

"conservative",

"conservative",

NA

)

)

Obama$WA03\_c = recodeVar(

Obama$WA03\_c,

c(unique(Obama$WA03\_c)),

c(

"Some education",

"Lot of education",

"Some education",

"Lot of education",

"No education",

"No education",

"Some education",

"Lot of education",

"Some education",

NA

)

)

## Factor-isation

Obama$RD01\_c = as.factor(Obama$RD01\_c)

Obama$RD01\_c = relevel(Obama$RD01\_c, "kerry")

Obama$CDb01\_c <- as.factor(Obama$CDb01\_c)

Obama$CDb01\_c <-

relevel(Obama$CDb01\_c, "withdraw as soon as possible")

Obama$CDd01\_c <- as.factor(Obama$CDd01\_c)

Obama$CDd01\_c <- relevel(Obama$CDd01\_c, "favor")

Obama$MA04\_c <- as.factor(Obama$MA04\_c)

Obama$MA04\_c <- relevel(Obama$MA04\_c, "liberal")

Obama$WA03\_c <- as.factor(Obama$WA03\_c)

Obama$WA03\_c <- relevel(Obama$WA03\_c, "No education")

################################################################################

## Régression de Age

summary(reg\_age <- lm(Obama$RCa10 ~ Obama$WA02\_c))

## Pas siginificatif écoomiquement et statistiquement

## Régression de Educ

unique(Obama$WA03\_c)

reg\_educ = lm(Obama$RCa10 ~ Obama$WA03\_c)

summary(reg\_educ)

## OKKKKKK

## Régression de Sexe

reg\_femme <- lm(Obama$RCa10 ~ Obama$WA01\_c)

summary(reg\_femme)

## Statistiquement significatif et Economiquement non

## Régression de Ethnie

summary(Obama$WC03\_c)

hist(Obama$WC03\_c)

table(Obama$WC03\_c)

unique(Obama$WC03\_c)

Obama$WC03\_c = as.factor(Obama$WC03\_c)

reg\_race = lm(RCa10 ~ WC03\_c, Obama)

summary(reg\_race)

## Parfaitement significatif

# test = subset(Obama,WC03\_c != "no answer" & WC03\_c != "don't know" )

# #test$WC03\_c = as.factor(test\_res$WC03\_c)

# test$WC03\_c = recodeVar(test$WC03\_c, c(unique(test$WC03\_c)),c("0","1","0","0","0","0","0"))

# unique(test$WC03\_c)

# reg\_test = lm(test$RCa10 ~ test$WC03\_c )

# summary(reg\_test)

## Pas significatif pour les autres modalités que W/B

## Régression de party id

unique(Obama$MA01\_c)

reg\_id = lm(RCa10 ~ MA01\_c, Obama)

summary(reg\_id)

## Très significatif

##Régression de Last vote

reg\_lastvote = lm(RCa10 ~ RD01\_c , Obama)

summary(reg\_lastvote)

## Très significatif

##Régression de Favor negotiating

unique(Obama$CDa05\_c)

reg\_negotiation = lm(RCa10 ~ CDa05\_c, Obama)

summary(reg\_negotiation)

##Significatif Stat et Eco sauf pour neutre, mais logique car neutre => pas d'effet

##Régression de Withdraw Troops

table(Obama$CDb01\_c)

reg\_withdraw = lm(RCa10 ~ CDb01\_c, Obama)

summary(reg\_withdraw)

## Parfait !

##Régression de Citizenship

table(Obama$CDd01\_c)

summary(lm(RCa10 ~ CDd01\_c, Obama))

## Résultat pas significatif (pas d'impact ?)

##Régression de people cons lib

table(Obama$MA04\_c)

reg\_ppl\_cons\_lib = lm(RCa10 ~ MA04\_c, Obama)

summary(reg\_ppl\_cons\_lib)

## Parfait et logique

##Régression de taxe

table(Obama$CBb01\_c)

test = subset(Obama, Obama$CBb01\_c != "none of these")

summary(lm(RCa10 ~ CBb01\_c, test))

## Bizarre mais peut s'expliquer par le fait que Obama veut absolument sortir de la crise

##Régression de salaire

table(Obama$WA04\_c)

#Obama$WA04\_c = recodeVar(Obama$WA04\_c, c(unique(Obama$WA04\_c)),c("less","more","more","more","more","less","less",NA,"more","less"))

#Obama$WA04\_c = as.factor(Obama$WA04\_c)

#Obama$WA04\_c = relevel(Obama$WA04\_c, "less")

summary(lm(RCa10 ~ WA04\_c, Obama))

## Pas fou

##Regression par bloc !

## Bloc 1 (Socio-démographiques) : WA01\_c + WA02\_C + WA03\_c + WA04\_c + WC03\_c

## Bloc 2 (Fixe) : RD01\_c + MA04\_c + MA01\_c

## Bloc 3 (Fluide) : CDb01\_c + CDa05\_c + CBb01\_c + SB01\_c + CA02\_c + ABo12\_c

6 Des tests peuvent être conjointement signicatifs sans l'être individuellement ⇒ Problème de multicolinéarité

## Regression logit

reg1 = lm(RCa10 ~ WA01\_c + WA02\_c + WA03\_c + WC03\_c,test)

reg1 = glm(RCa10 ~ WA01\_c + WA02\_c + WA03\_c + WC03\_c,data = test,family = binomial(link="logit"))

**Introduction (1page)**

En 2008, le jeune sénateur démocrate de l’Illinois, Barack Obama, n’ayant que dix ans d’expérience politique, remporte l’élection américaine face à John McCain, un vétéran du parti républicain, devenant ainsi le premier président de couleur des Etats-Unis d’Amérique. Dans un système électoral traditionaliste, ne comportant que deux partis ayant les moyens de leurs ambitions et un principe de grand électeur ne représentant pas toujours très bien l’avis populaire, Obama se distingue en ne s’associant non pas à des slogans à visée politique, mais plutôt à visée émotionnelle tels que “ Yes We Can ” ou “ HOPE “.

Ce revirement par rapport aux stratégies classiques s’explique notamment par le contexte économique et géopolitique de l’époque. En effet, en 2008 les Etats-Unis sont au coeur de la crise économique des “subprimes” tout en s’enlisant dans des guerres à la fois contre le terrorisme en Afghanistan et en Irak, depuis respectivement sept et cinq ans. C’est en réponse à ces sentiments d’insécurité et de colère qu’Obama décide de se montrer non pas en tant que candidat démocrate, mais en tant que candidat de l’espoir.

A l'aide de nos connaissances statistiques et économétriques, nous nous sommes ainsi intéressés aux déterminants de cette élection singulière, et plus particulièrement à la véracité de l’idée selon laquelle Obama a réussi à s’affranchir des déterminants électoraux classiques, tels que la socio-démographie ou l’appartenance politique, pour se concentrer sur les sentiments et les opinions.

En se basant sur les données issus de l’enquête téléphonique de la National Annenberg Election Survey, une enquête universitaire de grande envergure réalisée avant les élections, notre travail s’est naturellement découpé en trois grandes parties :

* Nous avons tout d’abord trié, classifié et recodé ces résultats à l’aide de statistiques descriptives et de régression linéaire simple, afin d’obtenir une analyse à la fois économiquement et statistiquement significative.
* Puis nous avons analysé les effets de ces variables dans une régression linéaire multiple, nous permettant de mieux comprendre les déterminants de cette élection.
* Enfin nous avons procédé à une analyse de modèle par bloc en divisant nos variables en trois catégories : *Socio-démographie* , *Idéologie* et *Opinions* . Cette analyse repose sur des tests statistiques, et a pour objectif d'expliciter mathématiquement l’importance relative de chacun de ces blocs et donc leur rôle dans le résultat de l’élection

**I. Choix de Variables (3 pages)**

La table de données à laquelle nous avons accès, issue de l’enquête téléphonique de la National Annenberg Election Survey, est une base en coupe transversale avec 3737 observations et 66 variables (c'est-à-dire 3737 individus qui ont répondu à 66 questions de l’enquête).

Notre objectif étant d’expliquer comment Obama a réussi à convaincre les électeurs, nous avons choisi comme variable dépendante la variable RCa\_10, qui correspond à la réponse à la question *“ Pensez-vous voter pour Obama ? ”*. Afin d’étudier cette variable, nous avons supposé que les individus ont suivi leurs croyances et ont voté pour Obama lorsqu’ils ont répondu oui. Cette variable est nominale, et admet trois modalités différentes: *“Yes”*, *“No”*, et *“Don’t know”*, ainsi que 842 valeurs manquantes. 182 personnes dans l’enquête ont répondu “*Don´t know*” , ce qui est un chiffre relativement plus petit que le nombre d’individus qui ont répondu “*Yes*” ou “*No*” (2713). De plus, on a considéré que cette valeur n’est pas pertinente à notre étude puisque notre objectif est de différencier les individus qui sont convaincus par Obama et ceux qui ne le sont pas. On a donc décidé de transformer cette valeur en “*No*” afin d’obtenir une variable indicatrice et de déterminer la probabilité des individus de voter pour Obama à la fin de notre étude grâce à une régression linéaire multiple.

Pour sélectionner nos variables indépendantes, nous avons tout d’abord commencé par choisir de manière intuitive les variables qui pourraient avoir une corrélation avec notre variable dépendante telles que le salaire de l’individu ou le parti avec lequel il s’identifie le plus.

Ensuite, nous avons regroupé les variables en trois blocs différents: le bloc *Socio-démographie* regroupant les caractéristiques éponymes, le bloc *Idéologie* représentant l’appartenance idéologique et le bloc *Opinion* regroupant celles des électeurs sur Obama et différents sujets politiques. Nous avons alors étudié de manière qualitative chaque variable en effectuant des régressions linéaires simples et des statistiques descriptives. Nous avons ainsi pu vérifier la significativité économique et statistique de chaque variable prise séparément. En effet, même si les estimations faites en régression linéaire simple risquent d’être biaisées par des corrélations avec d’autres variables, en vérifiant la significativité de chaque variable de manière individuelle on évite la pollution de notre modèle final et un problème de multicolinéarité de nos variables. Cette procédure nous a amené à :

- Supprimer les variables qui présentaient beaucoup de valeurs manquantes et/ou une significativité économique et statistique trop faibles.

- Recoder les variables ordinales et nominales , selon les résultats de la régression et d’une analyse descriptive, afin de créer des modalités composites pertinentes.

- Transformer les nombreuses réponses indiquant que l’individu ne peut/veut pas répondre à la question en valeurs manquantes.

Bloc Socio-démographie:

WA01\_c (Sex): variable indicatrice avec valeur de référence *“femme”.*

WA02\_c (Age): variable discrète que nous avons traité comme variable continue.

WA03\_c (Education): variable ordinale avec neuf catégories que l’on a regroupé en trois, afin d’augmenter la significativité statistique de la variable et de faciliter sa lecture.

WA04\_c (Salary): variable ordinale avec neuf catégories que l’on a regroupé en trois catégories (low, medium and high income) afin d’en augmenter la significativité statistique et de faciliter sa lecture. Cette variable présente beaucoup de valeurs manquantes (1662), son ajout peut donc poser un problème de biais de sélection. Nous avons cependant considéré que cette variable est trop fondamentale pour être supprimée. Nous discuterons tout de même de sa pertinence et du possible biais ainsi amené. à la fin du mémoire.

WC03\_c (Race): variable nominale avec sept catégories. En partant de l’idée que nous souhaitions différencier les individus appartenant à une minorité ethnique des blancs, nous avons remarqué que les individus ayant répondu “black, african american, black hispanic” avaient une probabilité de voter pour Obama bien plus haute que le reste des minorités. Nous avons donc regroupé les modalités en trois catégories: *“black, african american, black hispanic”*, *“white or white hispanic”* et *“other”*.

Bloc Idéologie:

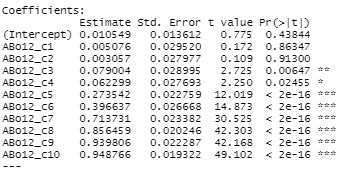
RD01\_c (Last-vote): variable nominale représentant le vote de l’individu aux élections en 2004.

MA04\_c (Conservative-liberal): variable ordinale avec 5 catégories (*“very liberal”*, *“somewhat liberal”*, *“moderate”*, *“somewhat conservative”*, *“very conservative”*). Elle indique pour chaque individu s’il est plutôt conservateur ou libéral. Nous avons choisi de regrouper ces catégories en trois : “*liberal*”, “*moderate*”, “*conservative*”, car les différences de probabilité de voter pour Obama entre “*very*” et “*somewhat*” pour les libéraux et pour les conservateurs sont négligeables.

MA01\_c (Id-party): variable nominale qui indique le parti auquel le candidat s'identifie le plus. Elle peut prendre les valeurs suivantes; “*republican*”, “*democrat*”, “*independent*”, “*other*”. Cette variable présente des paramètres estimés, économiquement et statistiquement significatifs.

Bloc Opinion:

ABo12\_c (Obama-share-values): variable ordinale qui indique, dans une échelle de 0 à 10, si l’individu partage ses valeurs avec Obama. Nous avons d’abord décidé de la traiter comme une variable ordinale, et de traiter chaque niveau comme une variable indicatrice pour étudier la différence de probabilités estimées pour Obama entre chaque niveau.



*Résultat de la régression linéaire de RCa10 sur ABo12\_c*

Nous avons pu distinguer trois groupes de niveaux; de 0 à 4, de 5 à 6, et de 7 à 10 que l’on a ainsi regroupé en : *“No”*, “*Yes*”, “*Somewhat*”. En effet, entre chacun de ces groupes nous avons remarqué une différence de probabilité estimée par le modèle de voter pour Obama bien supérieure que pour les reste des sous-niveaux.

CBb01\_c (Taxes): variable nominale qui regroupe trois différents avis possibles par rapport aux taxes: “*cut taxes*”, “*keep them as they are*”, “*raise taxes if necessary*”.

CDb01\_c (Withdraw Iraq Troops): variable nominale avec trois avis possibles pour chaque individu: *“withdraw as soon as possible”* , *“keep troops until stable government”* , *“set withdrawal deadline ”*

ABo05\_c (Obama-leader): variable ordinale qui indique, sur une échelle de 0 à 10, si l’individu pense que Obama est un bon leader. Nous avons traité cette variable comme une variable continue car la différence de probabilité estimée lorsque l’on augmente la valeur de la variable de 1 est toujours environ la même.

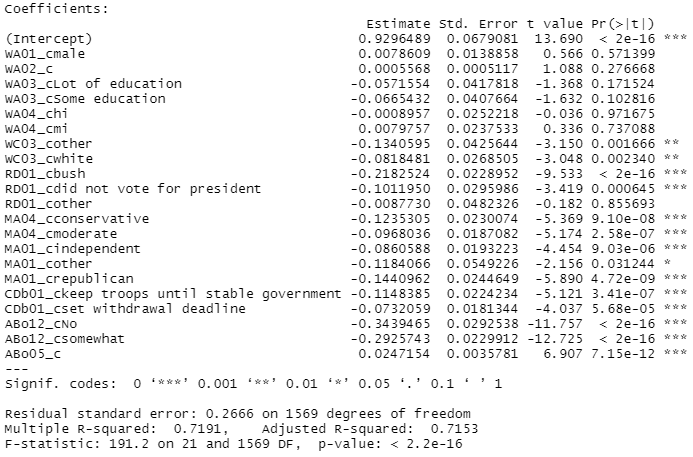
( optionnel, c’est des variables que l’on a pas ajouté finalement)

Variables avec trop de valeurs manquantes que l’on a supprimé:

Trust in white/black people SB01\_c/SB02\_c: Ces deux variables sont corrélées

Country going in right direction

**II. Analyse du modèle complet (3,15 pages)**

*****Résultat de notre régression linéaire effectuée sous R.*

**A. Vérification des hypothèses de Gauss-Markov**

Ce modèle reposant sur une enquête téléphonique, nous sommes bien conscients des limites de l'échantillonnage, en effet il est beaucoup plus facile de raccrocher et donc de ne pas répondre à certaines questions, que de quitter une enquête ayant lieu de vive voix. C’est pour cela que nous avons décidé d’effectuer un test sur la distribution conditionnel des valeurs manquantes de la variable dépendante RCa10. Pour le faire, nous avons créé une nouvelle variable RCa10\_na qui est une indicatrice valant 1 si RCa10 présente une valeur manquante, 0 sinon. Puis nous avons refait la régression complète de notre modèle en remplaçant RCa10 par RCa10\_na.

Nos résultats sont ici très probants[[1]](#footnote-0), en effet quasiment tous les paramètres sont non statistiquement significatifs et les peu étant significatifs ont un impact économique proche de zéro. Étant donné que la significativité statistique représente la probabilité de rejeter l’hypothèse nulle à tort, et qu’ici dans le cas où l'estimateur est tout de même significatif sa valeur est très proche de zéro, nous pouvons affirmer que, prises ensembles, les variables n’impactent pas la probabilité d’avoir une valeur manquante. Ainsi on peut en conclure que les valeurs manquantes de RCa10 ne sont pas distribuées en fonction d’une des autres variables de notre modèle, ce qui limite donc le biais induit par l'échantillonnage de cette enquête.

La régression linéaire de RCa10 sur notre modèle complet (ie. celui contenant toutes nos variables) présente un de 0.7191 et un ajusté de 0.7153, ces deux valeurs étant très proches cela nous indique que notre modèle n’est pas mal spécifié, de plus ces valeurs étant élevées, cela implique que notre modèle explique une grande part de la variance de RCa10. Ainsi nous pouvons tirer des conclusions de nos paramètres estimés, tout du moins en fonction du sens et de la grandeur relative de leur impact. Nous ne pouvons pas garantir que ce modèle est non biaisé étant donné la complexité de nos données, mais au vu de la répartition aléatoire des valeurs manquantes de notre variable dépendante , du fait que l’espérance empirique de nos résidus est quasiment nulle (E[û] = 9\*10-8) et de l’absence de colinéarité parfaite dans nos variables[[2]](#footnote-1), les résultats obtenus paraissent satisfaisants pour notre étude.

**B. Analyse du modèle par paramètres**

Nous allons à présent analyser les paramètres obtenus lors de la régression linéaire de RCa10 sur le modèle complet.

Afin d’alléger les interprétations, nous précisons ici qu'à partir de maintenant nous raisonnerons entièrement de manière probabiliste et dans le cadre de notre modèle, de plus toutes ces interprétations se feront par rapport à notre groupe de référence. Concernant les notations, “p.p” signifie points de pourcentage.

Groupe de référence :

WA01\_c (Sex) : Le paramètre valant 0.008, ainsi le fait d’être un homme augmente la probabilité de voter pour Obama de 0.8 p.p par rapport à une femme. Cette valeur très faible se remarque aussi dans la p-value qui est de 0.57, ce qui nous permet d’affirmer que le sexe n’a pas eu un effet important sur la probabilité de voter pour Obama.

WA02\_c (Age) : Le paramètre valant 0.0005, ainsi chaque année de vie supplémentaire augmente de 0.05 p.p la probabilité de voter pour Obama. Cette valeur se retrouve ici aussi dans la p-value qui est de 0.28, ce qui nous permet d’affirmer que l'âge n’a pas non plus eu un fort effet sur l’élection. *(En considérant deux électeurs de 20 et 60 ans, leur différence d’âge n’implique qu'une hausse de 2 p.p de la probabilité de voter pour Obama)*

WA03\_c (Education) : Les paramètres pour les deux modalités valant -0.057 et -0.066, le fait d’avoir fait beaucoup (resp. un peu) d’études fait baisser la probabilité de voter pour Obama par rapport aux individus n’ayant pas fait d’études de 5,7 p.p (resp. 6,6%). La p-value valant 0.17 (resp. 0.10), nous remarquons que celle-ci s’améliore par rapport aux précédentes, ce qui coïncident avec un effet plus fort. Cependant le modèle ne nous fournit pas de preuve que l’éducation ait pu être déterminante.

WA04\_c (Salary) : Les paramètres valant -0.0009 et 0.008, le fait d’avoir un haut (resp. moyen) salaire fait baisser (resp. augmenter) la probabilité de voter pour Obama par rapport aux individus ayant un faible salaire de 0.09 p.p (resp. 0.8 p.p). Les deux p-value étant supérieures à 0.75, on peut affirmer que le salaire n’a pas eu d’impact significatif sur l’élection.

WC03\_c (Race) : Les paramètres valant -0.08 et -0.13 , le fait d’être blanc (resp. d’une autre ethnie) fait baisser la probabilité de voter pour Obama par rapport aux individus noirs/afro-américains de 8 p.p (resp. 13 p.p). Les deux p-value étant inférieures à 0.003, nous pouvons considérer que l’ethnie a eu un réel impact. Cependant le fait que l’effet de la modalité *“white”* soit inférieur à celui de *“other”* nous montre que les électeurs ne se sont pas uniquement concentrés sur sa couleur de peau.

RD01\_c (Last vote) ; Les paramètres valant -0.22, -0.10, et -0.009 le fait d'avoir voté Bush[[3]](#footnote-2), (resp. de ne pas avoir voté/d’avoir voté pour un autre candidat) fait baisser la probabilité de voter pour Obama par rapport aux individus ayant voté Kerry[[4]](#footnote-3) de 22 p.p (resp. 10 p.p /0.09 p.p). Les p-value des deux premières modalités sont très statistiquement significatives tandis que celle de la troisième ne l’est pas du tout. Ainsi on retrouve ici une continuité chez les électeurs qui ont tendance à voter pour des candidats affiliés aux mêmes partis. De plus, Obama a aussi réussi à rattacher à sa cause les indécis comme le montrent les deux autres paramètres.

MA04\_c (conservative-liberal): Les paramètres valant -0.124 pour la modalité “*conservative*” et -0.097 pour “*moderate*”. Les conservateurs (resp. modérés) ont donc une probabilité estimée de voter pour Obama inférieure aux libéraux de 12.4 P.P (resp. 9.7). De plus, les deux paramètres ont une p-valeur inférieure à 0.001, les paramètres sont donc très statistiquement significatifs.

MA01\_c (id-party): Les paramètres ont pour valeurs ; -0.086 pour “*independent*”, -0.118 pour “*other*” et -0.144 pour “*republican*”. La probabilité de voter pour Obama pour les individus proches d’aucun parti est inférieure de 8.6 p.p (resp. 11.8 p.p pour les individus s’identifiant avec un parti autre que le parti Démocrate et le parti Républicain, et 14.4 p.p pour les individus s’identifiant avec le parti Républicain) par rapport aux individus qui s’identifient avec le parti Démocrate. Les paramètres estimés de “*independent*” et “*republican*” sont statistiquement significatifs avec des p-valeurs inférieures à 0.001. Le paramètre estimé de “*other*” a une p-valeur comprise entre 0.01 et 0.001. Il est donc statistiquement significatif au seuil de 1%.

ABo12\_c (Obama-share-values): Les paramètres valant -0.344 pour “*no*” et -0.293 pour “*somewhat*”, le fait de ne pas partager les valeurs de Obama, (resp. de partager un peu les valeurs de Obama) fait baisser la probabilité de voter pour Obama par rapport aux individus qui partagent les valeurs de Obama de 34.4 p.p (resp. 29.3 p.p). Les deux paramètres ont une p-valeur inférieure à 0.001, les paramètres sont donc statistiquement significatifs au seuil de 0.1%.

CDb01\_c (Withdraw Iraq Troops): Les paramètres valant -0.073 pour “*set withdrawal deadline*” et -0.115 pour “*keep troops until stable government*”, le fait de vouloir fixer une date limite pour retirer les forces de l’armée, (resp. de vouloir laisser l’armée jusqu’à l’arrivée d’un gouvernement stable) fait baisser la probabilité de voter pour Obama par rapport aux individus qui veulent retirer l’armée immédiatement de 7.3 p.p (resp. 11.5 p.p). Les deux paramètres ont une p-valeur inférieure à 0.001, les paramètres sont donc statistiquement significatifs au seuil de 0.1%.

ABo05\_c (Obama-leader): Le paramètre vaut 0.025.m Ainsi, augmenter la note selon laquelle Obama est un bon leader d’un point augmente la probabilité de voter pour Obama de 0.025 p.p. Le paramètre a une p-valeur inférieure à 0.001, le paramètre est donc statistiquement significatif au seuil de 0.1%.

**III. Analyse du modèle par bloc (4pages)**

Nous nous intéressons ici à l’importance relative de chacun des blocs précédemment définis au travers d’une régression par bloc. En effet, même si quasiment chacune de nos variables sont à la fois statistiquement et économiquement significatives, il serait intéressant de les comparer en fonction de leur bloc d’appartenance, afin de définir leur importance relative. Afin de mener cette étude nous utiliserons majoritairement deux moyens statistiques, l’analyse de variance et le critère d’information d’Akaike, l’AIC.

Ces deux moyens seront appliqués de façon combinatoire, ie. pour chaque possibilité de combinaison de bloc, qui sont en les notant respectivement 1, 2 & 3 : { 1, 2, 3, 1&2, 1&3, 2&3, 1&2&3 }. Ces tests seront de plus effectués sur un même panel cylindré, c'est-à-dire l'intersection des panels sans valeur manquante pour chaque combinaison de bloc, ce qui est équivalent au panel sans valeur manquante associé à la régression complète.

*Analyse de la Variance :*

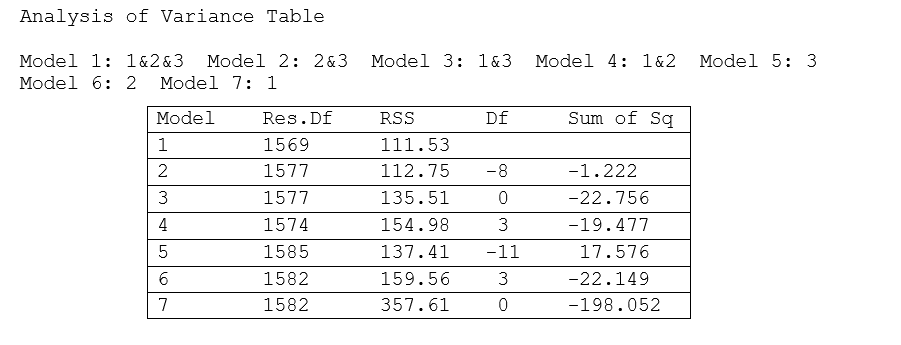
L’analyse de la variance explique la différence de la somme des carrés expliqués d’un modèle à un autre. Cette somme représentant la qualité d’ajustement de nos modèles par rapport aux données réelles, plus celle-ci est élevée et donc proche de la somme des carrés totaux, plus le modèle s’ajuste aux données et est donc efficace. Ces sommes de carrés sont plus souvent usité sous la forme du le coefficient de *“goodness-of-fit”* qui est calculé comme la , ce coefficient appartenant à , nous indique la part de la variance expliquée par le modèle .

Concernant les degrés de libertés, nous avons un nombre conséquent d’observations dans le panel cylindré , 1591, et l’amplitude des degrés de libertés associés à chaque modèle est de 16. Ainsi l’impact des degrés de libertés qui participent au calcul du ajusté est ici négligeable ( ).

Concernant les hypothèses nécessaires à la réalisation d’une analyse de la variance, nous nous retrouvons ici dans l’impossibilité de tester celles d'homoscédasticité de normalité, en effet étant donné que nous disposons de 15 variables à plusieurs modalités chacunes, la réalisation de groupe correspondant à toutes les combinaisons possibles de variable nous amène presque sûrement à des groupes vides ou à un individu. Nous avons tout de même essayé d’effectuer sous R des tests de Bartlett et de Levene, mais ceux-ci ne nous ont retourné que des erreurs.

Voici le résultat de la commande *“Anova( )”* de R qui nous permet d’effectuer automatiquement un test comparatif d’analyse de la variance pour nos différents modèles.

On note ici les blocs de cette façon : 1 = *Socio-démographie*, 2 = *Idéologie*, 3 = *Opinion*



Dans cette table nous allons donc nous intéresser à la valeur de la colonne *“Sum of Sq.”*  qui correspond à la différence de la somme des carrés expliqués de la ligne (n-1) - ceux de la ligne n. Plusieurs résultats sont ici intéressants :

* Ligne 2 : Ici le retrait du bloc socio-démographique par rapport au modèle complet

n’enlève que 1.22 à la somme des carrés expliqués ce qui est très faible, on peut donc en conclure que le bloc 1 n’explique pas une grande part de la variance relativement au modèle complet.

* Ligne 5 : Ici le remplacement des blocs socio-démographique et idéologique par

celui des opinions nous donne un résultat surprenant, en effet ici la différence est positive, cela signifie que le bloc 3 explique plus la variance de vote que les blocs 1 & 2 réunit.

* Ligne 6 : Relativement au boc 3 le bloc 2 fait perdre -22.149 à la somme des carrés

expliqués ce qui montre une moins bonne capacité d’explication du bloc 2 par rapport au bloc 3.

* Ligne 7 :On remarque ici que par rapport au bloc 2 le bloc 1 amène une perte de

198.052 dans la somme des carrés expliqués et donc de - (198.052 + 22.149) = -220,201 par rapport au bloc 3, ce qui indique un très faible niveau de variance expliqué du bloc 1 par rapport aux deux blocs précédents.

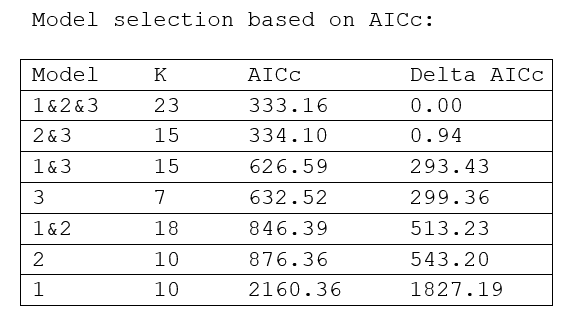
De ces résultats quantitatifs nous pouvons tirer une conclusion qualitative, tout d’abord le bloc Opinion est le bloc qui explique le mieux le modèle, au point où ce bloc suffit à lui tout seul à fournir une meilleure régression que les blocs *Socio-démographie* et *Idéologie* conjoints. Nous retrouvons ici le résultat attendu dans l’introduction, l’élection d’Obama a été le plus fortement déterminée par les opinions. De plus, nous pouvons aussi remarquer que relativement au bloc *Opinion* , le bloc *Idéologie* explique une part certes plus faible de la variance mais qui reste cependant conséquente, ainsi même si Obama s’est fortement appuyé sur les opinions pour son élection, nous ne pouvons tout de même pas dire qu’il a réussi à s’affranchir totalement des considérations idéologiques. Enfin nous pouvons remarquer que les caractéristiques socio-démographiques ont eu un impact très faible voir quasiment nul sur l’élection d’Obama, cela nous montre que son électorat est très disparate, et qu’il a su toucher par ses prises de positions toutes les classes sociales, toutes les ethnies et tous les âges.

*Critère AIC :*

Le critère AIC, est un critère de qualité d’un modèle ne reposant pas seulement sur la qualité d’ajustement mais aussi sur le critère de parcimonie[[5]](#footnote-4). En effet, selon la formule des carrés expliqués, il est mécaniquement possible d’augmenter le ,autant que l’on le souhaite, en ajoutant des variables aux modèles peu importe leur explicativité, le AIC permet donc de pénaliser le nombre de variables d’un modèle et se calcule selon la formule suivante :

*avec k le nombre de variable, et L le maximum de la fonction de vraisemblance du modèle.*

Ainsi un modèle sera meilleur selon le critère AIC, plus son AIC sera faible.



Voici le résultat de la commande *“AIC( )”* de R qui nous retourne un tableau où les modèles sont triés par AIC croissant, ie. les meilleurs modèles sont en haut du tableau, et les moins bons en bas. La colonne K quant à elle représente le nombre de variables, et donc la pénalisation de chaque modèle.

Ici aussi plusieurs résultats sont pertinents pour notre analyse.

* La différence d’AIC entre les deux premières lignes nous montre que malgré l’ajout

de 8 variables sur 15 déjà présentes, le bloc 1 apporte tout de même une baisse de l’AIC, ce qui peut donc s’expliquer par une hausse conséquente de la vraisemblance.

* La différence d’AIC entre les lignes deux et trois nous montre que relativement au

bloc 3, sa combinaison avec le bloc 2 est fortement préférable à celle avec le bloc 1, et ceci pour un nombre de variables identique (ie. la baisse d’AIC est amenée seulement par une hausse de la vraisemblance) .

* La différence d’AIC entre les lignes quatre et cinq nous apporte une information très

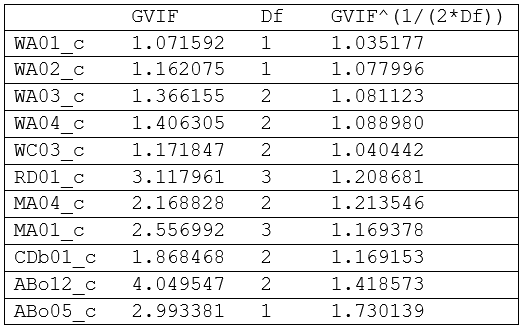
significative, en effet selon le critère AIC, il serait préférable de se contenter d’un modèle ne comportant que le bloc 3 plutôt qu’un modèle composé des blocs 1&2.

Ainsi, de ces observations quantitatives nous pouvons tirer des conclusions qualitatives sur les modèles à préférer, et surtout sur la qualité amenée par chaque bloc. Nous retrouvons évidemment les conclusions de l’analyse de variance, notamment sur l'ordre d’importance des blocs, qui sont dans l’ordre décroissant : *Opinion* , *Idéologie* , *Socio-démographie*. Cependant ici nous pouvons apporter plus de nuances et de précision à ces conclusions. Malgré son nombre de variables élevé et sa faible quantité de carrés expliqués, le bloc *Socio-démographie* reste pertinent dans notre analyse. Il n’a certes pas été déterminant dans l’élection d’Obama mais il permet tout de même d’augmenter la vraisemblance de nos modèles statistiques, sûrement grâce à un effet de contrôle. De plus, le bloc *Idéologie* apparaît ici comme un bloc important dans l’analyse de cette élection, même s' il ne l’est pas autant que celui des opinions. Enfin, ce dernier bloc, *Opinion*, renforce encore ici sa position de dominance en termes de qualité et d’impact.

De ces analyses statistiques , il ressort une version plus nuancée de notre supposition initiale, selon laquelle Obama aurait pu s’affranchir des déterminants électoraux classiques au profit des opinions. En effet, même si nous avons pu observer à quel point les opinions avaient été importantes pour son accession au poste de président, cette accession à tout de même été fortement influencée par les principes idéologiques des citoyens des Etats-Unis. De plus, même si les caractéristiques socio-démographiques n’ont certainement pas joué un rôle très important, elles restent tout de même significatives, et importantes dans les études statistiques en leur qualité de variable de contrôle.

**ANNEXE**

1. Analyse de la multi colinéarité à l’aide de la commande “*vif(reg)”*

****

Ici les résultats s'interprètent comme ceci : l’absence de multicolinéarité parfaite dans nos variables est vérifié si la valeur de la colonne 4 : est proche de 1 ce qui est le cas ici. (Dans la littérature associé il n'existe pas de consensus sur une valeur seuil à partir de laquelle cette hypothèses n’est plus vérifié, cependant tous les seuils proposés restent supérieur à 2 ce qui rend les résultats obtenus satisfaisants)

1. Le tableau de la régression étant très grand, nous avons décidé de le mettre en annexe. [↑](#footnote-ref-0)
2. Cette vérification a été effectuée à l’aide la commande *“vif(reg)”* de R, dont le résultat est en annexe. [↑](#footnote-ref-1)
3. Georges W. Bush, candidat républicain et vainqueur de l’élection présidentielle de 2004. [↑](#footnote-ref-2)
4. John F. Kerry, candidat démocrate à l’élection présidentielle de 2004. [↑](#footnote-ref-3)
5. Le critère de parcimonie, ou *"rasoir d'Ockham''*, est un principe philosophique fondamental en science qui prescrit de préférer les modèles d’hypothèses les plus simples lorsque l’on cherche à inférer sur des caractéristiques réelles. [↑](#footnote-ref-4)