

UE STA 108 : Enquêtes et sondages

Travail pratique obligatoire

Les données : résultats des élections présidentielles 2007 :

La population étudiée est celle des bureaux de votes des élections présidentielles de 2007. Les données sont les résultats au premier et au deuxième tour sur 65466 bureaux. Ces données longtemps conservées au ministère de l'intérieur sont maintenant téléchargeables à cette adresse : <http://www.nosdonnees.fr/dataset/resultats-des-elections-presidentielles-2007>

Nous vous les avons préparées aux formats SPSS, Excel et SAS (programme de création des données)

Le contenu des fichiers est identique : pour chacun des 65466 bureaux, il y a des informations géographiques, les résultats au premier et au deuxième tour, ainsi qu'une variable de stratification politique. On étudiera le pourcentage d'exprimés sur les candidats du deuxième tour (Ségolène Royal et Nicolas Sarkozy)

On connaît le résultat, l'objectif est donc d'étudier la variabilité des estimations selon la méthode d'échantillonnage. La valeur exacte dans ce fichier diffère très légèrement des résultats publiés par le ministère de l'intérieur, car il ne contient pas les résultats des français à l'étranger.

	<i>Fichier</i>		<i>Ministère</i>	
<i>Nicolas Sarkozy</i>	46.96%	16 509 734	46.94%	16 790 440
<i>Ségolène Royal</i>	53.04%	18 648 644	53.06%	18 983 138
		35 158 378		35 773 578

La liste des variables

Variable	Libellé	Unité	Tour
D	Departements regroupés en 9 régions		
C	Code Commune		
NumBur	Numéro de bureau dans la commune		
nTour	numéro de tour		
ELECTION	Code election		
Inscrits	inscrits	nombre d'inscrits	1
Votants	votants	nombre de voix	1
Exprime	Exprimes	nombre de voix	1
BAYR	Bayrou	nombre de voix	1
BESA	Besancenot	nombre de voix	1
BOVE	Bové	nombre de voix	1
BUFF	Buffet	nombre de voix	1
LAGU	Laguiller	nombre de voix	1

LEPE	Le Pen	nombre de voix	1
NIHO	Nihous	nombre de voix	1
ROYA	Royal	nombre de voix	1
SARK	Sarkozy	nombre de voix	1
SCHI	Schivardi	nombre de voix	1
VILL	De Villiers	nombre de voix	1
VOYN	Voynet	nombre de voix	1
Inscrits2	inscrits	nombre de voix	2
Votants2	votants	nombre de voix	2
Exprime2	exprimés	nombre de voix	2
ROYA2	Royal au 2ème tour	nombre de voix	2
SARK2	Sarkozy ou deuxième tour	nombre de voix	2
abstention	abstention au 1er tour	nombre de voix	1
abstention2	abstention au 2ème tour	nombre de voix	1
LIBGEO	Nom de la commune		
REG	code region INSEE		
DEP	Code département		
TAU2010	Tranche d'unité urbaine 2010		
POP_MUN_2008	Population municipale (INSEE RP 2008)		
Pctabstention	Abstention	Pourcentage	1
PctBAYR	Bayrou	Pourcentage	1
PctBESA	Besancenot	Pourcentage	1
PctBOVE	Bové	Pourcentage	1
PctBUFF	Buffet	Pourcentage	1
PctLAGU	Laguiller	Pourcentage	1
PctLEPE	Le Pen	Pourcentage	1
PctNIHO	Nihous	Pourcentage	1
PctROYA	Royal	Pourcentage	1
PctSARK	Sarkozy	Pourcentage	1
PctSCHI	Schivardi	Pourcentage	1
PctVILL	De Villiers	Pourcentage	1
PctVOYN	Voynet	Pourcentage	1
Pctabstention2	Abstention	Pourcentage	2
PctSARK2	Royal	Pourcentage	2
PctROYA2	Sarkozy	Pourcentage	2
IDENT	Identifiant unique de bureau		
Strate	Stratification politique (6 classes)		

Les soirées électorales

Lors des soirées électorales, les instituts (TNS, IPSOS,..) constituent un échantillon de bureaux fermant avant 20 heures, puis extrapolent les résultats nationaux sur la base de cet échantillon. Pour la présidentielle 2007, l'échantillon de TNS contenait 220 bureaux de vote. L'estimation se faisait avec un estimateur par la régression (sur les résultats du premier tour), après avoir calé l'échantillon (sur les résultats du premier tour aussi).

Voici les estimations de 20h pour le deuxième tour 2007 de TNS Sofres :

Estimation en voix (niveau national)

tns sofres Unilog a logic company

Imprimer Recharge 5 sec

20:03:55 0 1 2 3 4 Total E

Nb bureaux 0 0 12 0 160 172

Echantillon NATIONAL										
Tendance	Scénarios	INS/EXP	Résultat brut	Corrélation	Ecart	Indice	Brut.Ecart	Brut.Indice	Incertitude	
Abstention	ABSTENTION P07.1	INS	15.40	0.94	14.5	14.5	0.9	0.9	0.22	
Abstention	ABS+ 2BAY+ 30LP P07.1	INS	15.40	0.93	14.4	14.7	1.0	0.7	0.23	
Abstention	ABSTENTION auto P07.1	INS	15.40	0.94	14.5	14.5	0.9	0.9	0.22	
Abstention	auto tout	INS	15.40	0.94	14.6	14.6	0.8	0.8	0.20	
Blancs nuls	BLANCS+NULS P07.1	INS	3.74	0.36	3.7	3.5	0.1	0.2	0.13	
Blancs nuls	BN auto P07.1	INS	3.74	0.36	3.7	3.5	0.1	0.2	0.13	
SARKOZY	SARKOZY P07.1	EXP	52.89	0.90	53.8	54.5	-0.9	-1.6	0.69	
SARKOZY	SARKOZY+ 4bay P07.1	EXP	52.89	0.87	54.1	54.6	-1.2	-1.7	0.75	
SARKOZY	Sarkozy Compo P07.1	EXP	52.89	0.98	53.3	53.3	-0.4	-0.4	0.43	
SARKOZY	auto P07.1	EXP	52.89	0.98	53.1	53.1	-0.2	-0.2	0.38	
ROYAL	ROYAL P07.1	EXP	47.11	0.91	47.2	47.3	-0.1	-0.2	0.71	
ROYAL	ROYAL+ 9exg P07.1	EXP	47.11	0.96	46.8	46.7	0.3	0.4	0.44	
ROYAL	ROYAL+exg+ 4bay P07.1	EXP	47.11	0.97	47.1	47.1	0	0	0.40	
ROYAL	ROYAL composite P07.1	EXP	47.11	0.98	46.9	46.8	0.3	0.3	0.41	
ROYAL	auto P07.1	EXP	47.11	0.98	46.9	46.9	0.2	0.2	0.38	

Exercice :

Etude de distributions d'échantillonnage selon la méthode : On effectuera des tirages d'échantillons (50 tirages) de n= 220 et 500 unités selon des plans : aléatoires simples sans remise, systématiques, stratifiés à allocation proportionnelle (sur la variable **Strate**). Les résultats des 50 tirages seront présentés sous forme statistique (moyenne, écart-type, histogramme) et commentés, en précisant les algorithmes utilisés.

Soirée électorale : On bâtera un estimateur par la régression pour les scores de Ségolène Royal et Nicolas Sarkozy : on fera la régression sur les données du premier tour. Une seule estimation sera faite, à partir d'un échantillon de n = 220 stratifié à allocation proportionnelle sur la variable **Strate** (on ne tiendra pas compte de la stratification pour ne pas complexifier le calcul de l'estimateur)

La date de remise du travail qui peut s'effectuer à deux, est fixée au 10 février 2014 pour la première session ou au 21 avril 2014 pour la deuxième session. Le travail sera remis par courrier électronique à gilbert.saporta@cnam.fr. Le rapport devra être rédigé et consister en un **fichier unique** au format .doc ou .pdf

La note finale de STA108 sera la moyenne arithmétique équipondérée de la note d'examen (1^{ère} ou 2^{ème} session) et de la note de projet.