

# **Pédagogie statistique des sondages électoraux**

**Benoît Riandey (INED)**

# Objets de la communication

- Confusion de résultats troublante !
- Prudence, médias !
- Estimations statistiques ou dires d'experts politologues
- Estimations de premier ou second tour
- Taille efficace d'échantillon et intervalles de confiance
- L'indicateur de qualité de l'échantillon
- Les vraies questions et leurs modes de réponse :  
estimation ou inférences ?

Conclusion :

appel aux instituts et aux universitaires

# Confusion de résultats troublante !

- Lundi 12 mars, IFOP : croisement des courbes !  
le candidat S en tête au premier tour (28, 5%).

écart réduit de 2 points au second tour

- Mardi 13 mars, SOFRES : le candidat H  
confirme son avance (30 %).

écart accru d'un point au second tour

# Médias, Prudence !

## Exemple

L'excellent « rolling » de l'IFOP

*(en français : enquête en continue)*

±300q/jour puis moyenne sur trois jours  
une estimation chaque jour

(d'égale précision que les concurrentes)

Bravo !

# médias Prudence !

## Exemple

*L'enquête en continue de l'IFOP*  
±300q/jour puis moyenne sur trois

Ajout de Paris-Match :  $\Delta_{j,j-1} = p_j - p_{j-1}$   
= comparer deux échantillons de 300q

$$IT = \pm\sqrt{2} \times 5,3\% = 7,5 \% \text{ (pour } p = 0,3 \text{ et } n_j = n_{j-1} = 300)$$

$$IT = \pm\sqrt{2} \times 5,8\% = 8, \% \text{ (pour } p = 0,3 \text{ et } n_j = n_{j-1} = 300)$$

Consternation

et **Bonnet d'âne** statistique pour Paris-Match !

# QUESTION

**Où peut-on lire des choses  
un peu consistantes sur la précision  
des sondages électoraux français ?**

*Que propose-t-on à nos étudiants ?*

# Pédagogie positive

## Exemple

*Le Monde* 16 janvier 2012

Vers un match à quatre ?

Comparaison des scores des  
4 candidats B,H,LP,S

Pour 4 instituts

IFOP, CSA, OpinionWay, BVA  
en novembre, décembre , janvier

Une **Pédagogie positive des médias**  
*le Comparateur des sondages de la présidentielle*  
*(=10.000 enquêtés)*

Suivi graphique de toutes les estimations de 8 instituts  
par candidat

de novembre 2011 à au second tour

<http://tempsreel.nouvelobs.com/sondage-presidentielle-2012/20120207.OBS0739/infographie-le-comparateur-des-sondages-de-la-presidentielle.html>





# Estimation statistique

ou

dire d'expert ?

- Estimation statistique :
- choisir une sous-population « *les certains de leur choix* »
- Une pondération « *calage socio-démo et sur le dernier vote* »
- Distribution de la variable intention pour cette population  
avec cette pondération et s'y tenir

# Estimation statistique ou dire d'expert ?

Sinon : Dire d'experts

En particulier

- « Faire une estimation candidat par candidat »
- Faire exception pour un candidat
- « Tenir compte d'une certaine marge d'appréciation » ...
- Demeurer dans la « fourchette » **anti-statistique**

# Estimations post-stratifiées

Estimation pondérée par calage socio-démographique  
et sur le dernier vote

= mesure des transferts d'intentions entre  
les candidats du vote antérieur et ceux du  
vote présent ...

... en tenant compte des entrées (**facile**)  
et sorties (**bien difficile**) du corps électoral

# Estimations statistiques ou dires d'expert ?

Sinon : Dire d'experts

Pourquoi pas !

mais il faudrait le dire !

*et le justifier ?*

Attention :

La fourchette statistique ne s'applique pas  
aux dires d'experts !

# Estimations statistiques ou dires d'expert ?

Qui parle ?

Sondeur ou politologue ?

Ne confondons pas les rôles

# Estimations post-stratifiées

- Premier tour : estimation délicate
  - référence ancienne
  - beaucoup d'entrants (de mœurs politiques inconnues)  
Risque de biais
- Second tour ... dans la foulée :
  - corps électoral identique
  - référence très corrélée (55 % aux 2 candidats + consignes de report)

D'Où 1000 enquêtés second tour = 2000 enquêtés de premier tour

# Taille efficace d'échantillon

*(exemple second tour équilibré)*

$(p=0,5)$

- Combien d'enquêtés ? 1000
- Combien d'intentions exprimées ? 700
- Combien de 18-29 ans ? 150

# Bien jouer avec les intervalles de confiance

Dussaix et Grosbras

## précision d'une proportion

Tableau 2.2. – Précision de l'estimation d'une proportion calculée à partir d'un échantillon

Proportion taille observée d'échan- tillon n	P										
	5 % ou 95 %	8 % ou 92 %	10 % ou 90 %	15 % ou 85 %	20 % ou 80 %	25 % ou 75 %	30 % ou 70 %	35 % ou 65 %	40 % ou 60 %	50 %	
100					8	8,6	9,2	9,6	9,8	10	
150				5,7	6,4	6,9	7,3	7,6	7,8	8	
200			4,3	5,1	5,7	6,1	6,5	6,8	6,9	7,1	
250	2,8	3,4	3,8	4,5	5	5,4	5,8	6	6,2	6,3	
300	2,5	3,1	3,5	4,2	4,6	5	5,3	5,6	5,7	5,8	
350	2,3	2,9	3,2	3,8	4,2	4,6	4,9	5,1	5,2	5,3	
400	2,2	2,7	3	3,6	4	4,3	4,6	4,8	4,9	5	
500	2	2,4	2,7	3,2	3,6	3,9	4,1	4,3	4,4	5	
600	1,8	2,2	2,4	3	3,3	3,5	3,8	3,9	4	4,1	
700	1,7	2,1	2,3	2,7	3	3,3	3,5	3,5	3,7	3,8	
800	1,5	1,9	2,1	2,5	2,8	3	3,2	3,3	3,4	3,5	
900	1,5	1,8	2	2,4	2,7	2,9	3	3,1	3,2	3,3	
1 000	1,4	1,7	1,8	2,3	2,5	2,7	2,9	3	3	3,1	
1 500	1,2	1,4	1,5	1,9	2,1	2,3	2,4	2,5	2,6	2,6	
2 000	1	1,2	1,3	1,6	1,8	2	2,1	2,2	2,2	2,3	
3 000	0,8	1	1,1	1,3	1,4	1,5	1,6	1,7	1,8	1,8	
5 000	0,6	0,8	0,8	1	1,1	1,2	1,3	1,4	1,4	1,4	
10 000	0,4	0,5	0,6	0,7	0,8	0,9	0,9	1	1	1	

# Taille efficace d'échantillon

*(exemple second tour équilibré)*

$(p=0,5)$

- Combien d'enquêtés ? 1000  $> \pm 3,1 \%$
- Combien d'intentions exprimées ? 700  $> \pm 3,8 \%$
- Combien 18-29 ans ? 150  $> \pm 8 \%$

Calcul immédiat.

A quand et chez qui la présentation d'un simple clic ?

# L'indicateur de qualité d'un échantillon

- La distribution de la pondération des enquêtés
- Le coefficient de variation de la pondération :

écart-type/moyenne

# Les vraies questions

- Le candidat potentiel A serait-il mieux placé que le candidat B ?
- Le candidat A progresse-t-il ?
- L'avance du candidat A sur le candidat B est-elle significative ?
- **Les courbes se sont-elles croisées ?**
- Les deux enquêtes plaçant les intentions pour le candidat A à f1 et f2 sont-elles compatibles ? Si oui, avec quel degré de certitude ? Sinon, pourquoi de tels écarts ?
- L'enquête A a-t-elle correctement annoncé le résultat de l'élection ?
- Les enquêtes par Internet donnent-elles des résultats identiques, différents, meilleurs, moins fiables que les enquêtes téléphoniques ?

# Les vraies réponses ?

- Dans les intervalles de confiance ?

# Les vraies réponses ?

- Dans les intervalles de confiance ? NON
- Dans la validation des inférences !

# Les vraies réponses ?

- Dans les intervalles de confiance ? NON
- Dans la validation des inférences !
- Avec quel degré de certitude ?

# Les vraies réponses

## Avec quel degré de certitude ?

- 99,999 comme le boson de Higgs ?
- 95 % comme les statisticiens ?

# Les vraies réponses

## Avec quel degré de certitude ?

- 99,999 comme le boson de Higgs ?
- 95 % comme les statisticiens ?
- C'est Ça la réponse !  
et c'est de la statistique (de niveau 2)

# Quel degré de certitude ? exemples

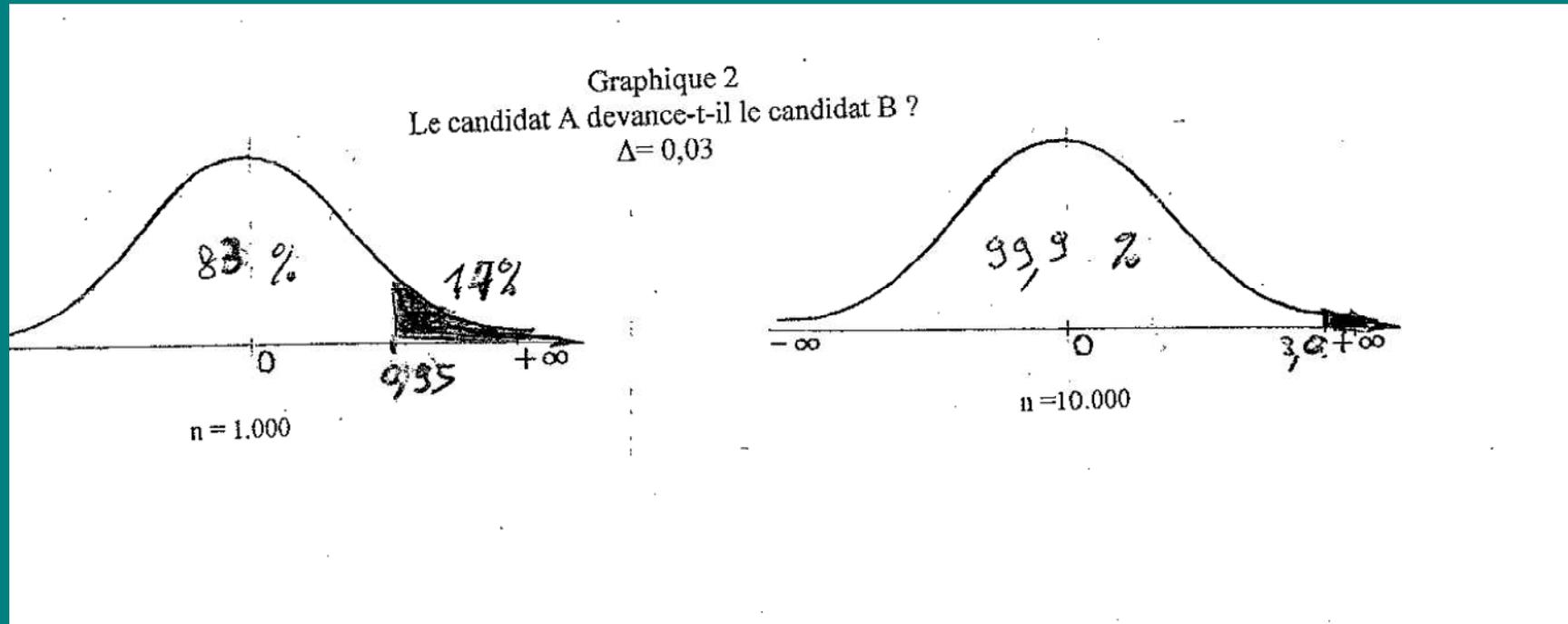
Le candidat A devance-t-il le candidat B ?

$$V_{\Delta} = p_A (1-p_A)/n + p_B (1-p_B)/n + 2 p_A p_B/n$$

$$\{V (p_A - p_B) = V (p_A) + V (p_B) - 2 \text{cov} (p_A , p_B)\}$$

# Avec quelle certitude ?

## Exemple d'un second tour



# Quid des universitaires français ?

- Les universitaires français se sont-ils livrés à cette pédagogie élémentaire ?
- A quand un fichier pédagogique d'une enquête d'intentions de vote pour les étudiants de sciences-Po ?  
(ou les autres)

# Quid des universitaires français ?

- Les universitaires français se sont-ils livrés à cette pédagogie élémentaire ?
- Non, sans doute, faute d'accéder aux fichiers de ces enquêtes.
- A quand une thèse de statistique appliquée synthétisant une campagne électorale ?

# Références

- Chiche J. et Riandey B. « Qualité des enquêtes politiques » (à paraître)
- La commission des sondages face aux élections présidentielles et législatives de 2012
- Dussaix AM et Grosbras JM (1993) « Les sondages : principes et méthodes »
- Riandey B. (2012) « Inférence statistique et sondages électoraux », APMEP (à paraître)