

## **SONDAGES D'INTENTION DE VOTE : L'ESTIMATION DES «MARGES D'ERREUR»**

**Léo Gerville-Réache**

Université de Bordeaux 2, CNRS, UMR 5251, Bordeaux, F-33000, France – [leo.gerville@u-bordeaux2.fr](mailto:leo.gerville@u-bordeaux2.fr)

# Motivation

---

- ▶ « *En dehors des sondages pré-électoraux, pour lesquels on a, le jour de l'élection, la mesure de l'erreur commise, beaucoup d'enquêtes produisent des statistiques dont on ne connaîtra jamais vraiment la pertinence.* » Pascal Ardilly (2006).

# Ne pas confondre : erreurs et marges d'erreur

---

- ▶ L'erreur est l'écart entre l'estimation et la « vraie valeur »

$$\hat{p} - p$$











- ▶ La marge d'erreur est le demi intervalle de confiance pour la « vraie valeur », calculé à partir des données.
  - ▶ Pour un échantillon aléatoire simple, la marge d'erreur (au niveau de confiance de 0,95) sur une proportion  $p$  est calculée par la formule approximative suivante:

$$1,96\sqrt{\hat{p}(1-\hat{p})/n}.$$

- ▶ Telle que :  $\mathbf{P}\left\{|\hat{p} - p| \leq 1,96\sqrt{\hat{p}(1-\hat{p})/n}\right\} \approx 0,95.$
- ▶ Quelle relation y-a-t-il entre erreurs et marges d'erreur?

# Passage des erreurs aux marges d'erreur

- ▶ Exemple d'un sondage d'intention de vote : Ipsos, le 20 Avril 2012

Candidat	Sondage	Vote	Ecart
Eva Joly	2,0	2,3	-0,3 
François Bayrou	10,0	9,1	0,9 
François Hollande	29,0	28,6	0,4 
Jacques Cheminade	0,5	0,3	0,3 
Jean-Luc Mélenchon	14,0	11,1	2,9 
Marine Le Pen	16,0	17,9	-1,9 
Nathalie Arthaud	< 0,5	0,6	0,3 
Nicolas Dupont-Aignan	1,5	1,8	-0,3 
Nicolas Sarkozy	25,5	27,2	-1,7 
Philippe Poutou	1,5	1,2	0,4 

- ▶ Question : quelles marges d'erreur produisent de telles d'erreurs?

# Notre approche : La « référence à la méthode aléatoire »

---

- ▶ Pour la commission du Sénat (2011) « *Art. 2. – La première publication ou la première diffusion de tout sondage, tel que défini à l'article 1<sup>er</sup>, est accompagnée des indications suivantes, :[...]*
- ▶ **Les marges d'erreur des résultats publiés ou diffusés, le cas échéant par référence à la méthode aléatoire. »**

# Notre approche : La « référence à la méthode aléatoire »

---

- ▶ Pour L'ISO 20252 (2006) « *Pour les études quantitatives, les informations détaillées minimales suivantes doivent être documentées dans le rapport d'étude. [...];*
  - ▶ *les procédures de pondération, le cas échéant ;*
  - ▶ *les procédures d'estimation et d'imputation, le cas échéant ;*
  - ▶ ***la fiabilité des conclusions ;***

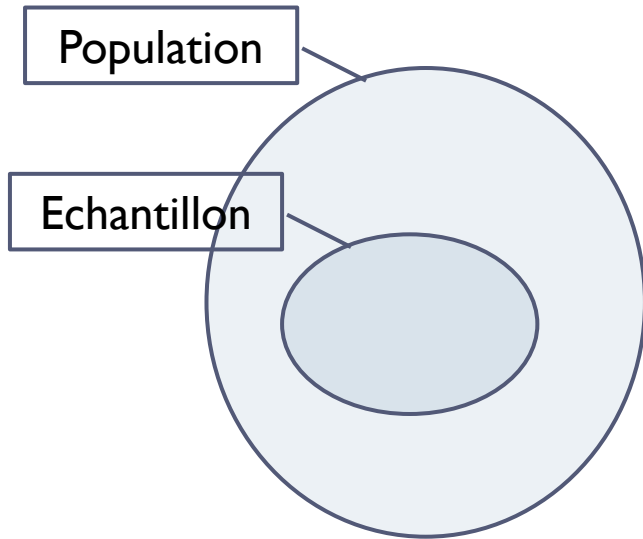
# Notre approche : La « référence à la méthode aléatoire »

---

- ▶ Pour IPSOS (2012) (<http://www.ipsos.fr/faq>) « Les lois statistiques qui permettent de la déterminer ne valent théoriquement que pour les sondages aléatoires.

**En pratique, on considère cependant que la marge d'erreur des sondages par quotas est égale ou inférieure à celle des sondages aléatoires.»**

## «Méthode aléatoire»...



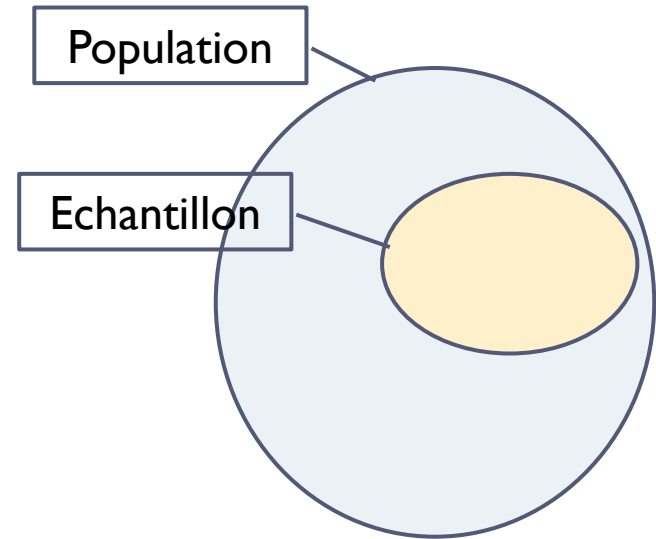
### ▶ Echantillon aléatoire simple

- ▶ sans non-réponse
- ▶ sans indécis
- ▶ sans mensonge

### ▶ Marges d'erreur :

$$Me = 1,96\sqrt{\hat{p}_k(1-\hat{p}_k)/n}$$

## «Pratique»...



### ▶ Echantillon par quotas

- ▶ non-réponse
  - ▶ indécis
  - ▶ mensonge
- |                   |
|-------------------|
| Redressement      |
| Affinage          |
| Analyse politique |

### ▶ Marges d'erreur :

**Comment faire « référence à la méthode aléatoire »?**



# Approche 1: échantillon aléatoire de « taille équivalente »

---

- ▶ Cette approche consiste à estimer la taille de l'échantillon aléatoire simple qui produirait des écarts tels que ceux observés entre le sondage d'un institut et le vote.
- ▶ Pour un échantillon aléatoire simple l'erreur normalisée de chaque candidat,

$$e_k = (\hat{p}_k - p_k) / \sqrt{p_k(1-p_k)}$$

suit approximativement une loi normale avec:  $E(e_k) = 0$  et  $V(e_k) = 1/n$

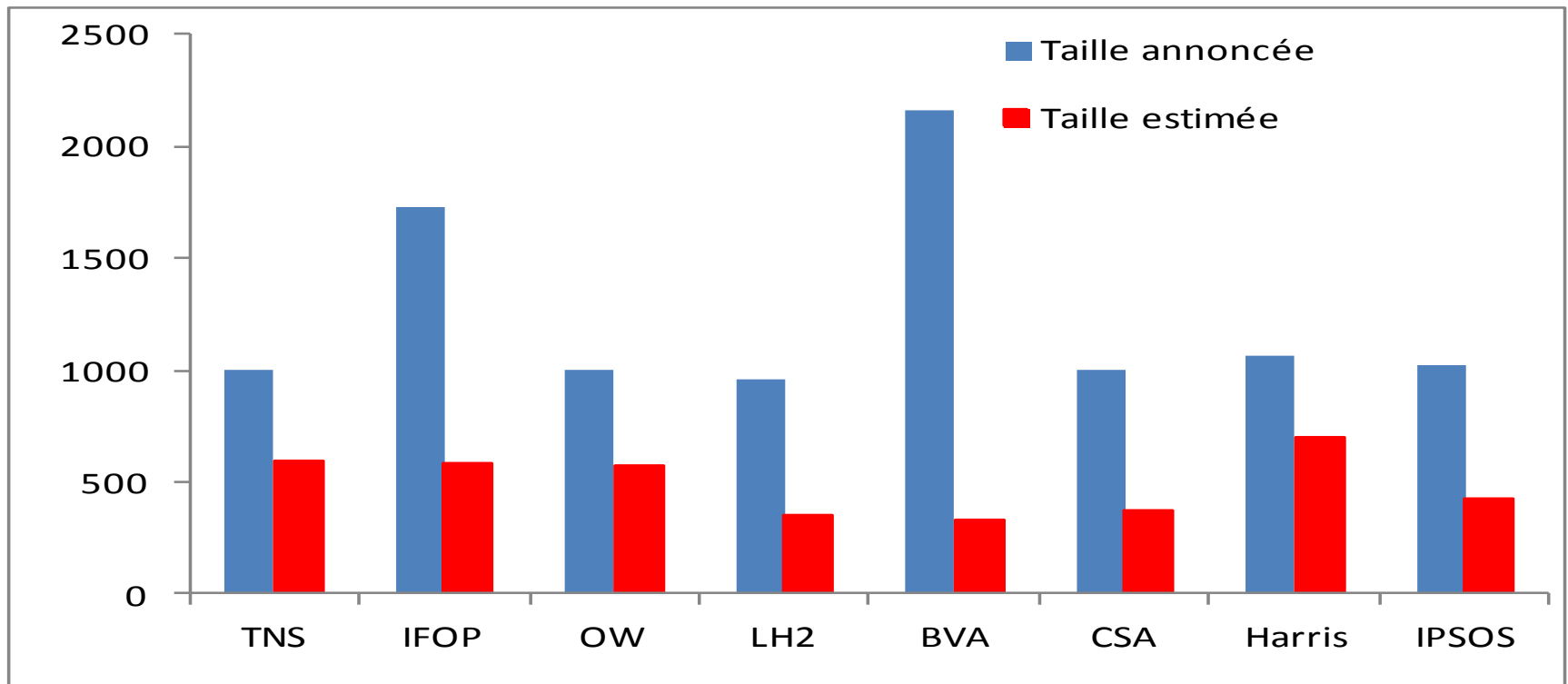
- ▶ Un estimateur consistant de la « taille équivalente » est donné par la formule suivante :

- ▶  $\hat{n} = 1 / S_n^2$  où  $S_n^2$  est la variance empirique des  $e_k$

- ▶ Aussi, la marge d'erreur peut être estimée par :

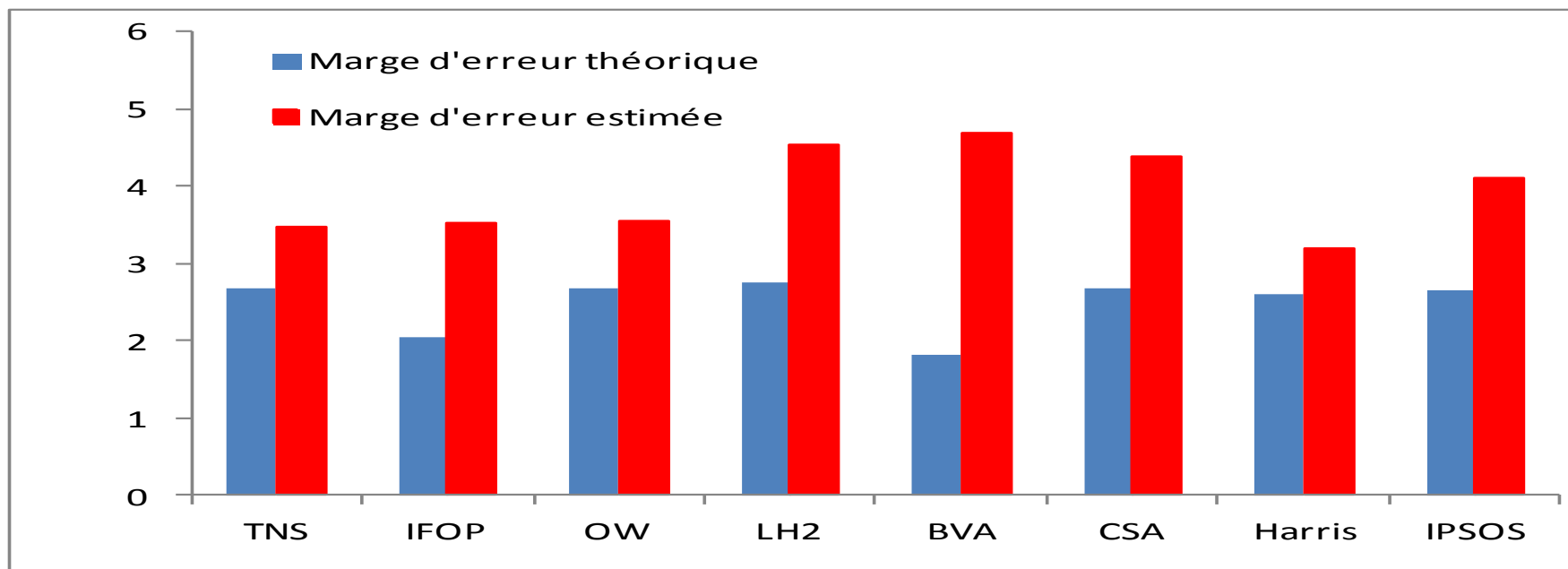
$$1,96\sqrt{\hat{p}_k(1-\hat{p}_k)/\hat{n}}$$

# Application au dernier sondage de chacun des 8 instituts (validé par la commission des sondages)



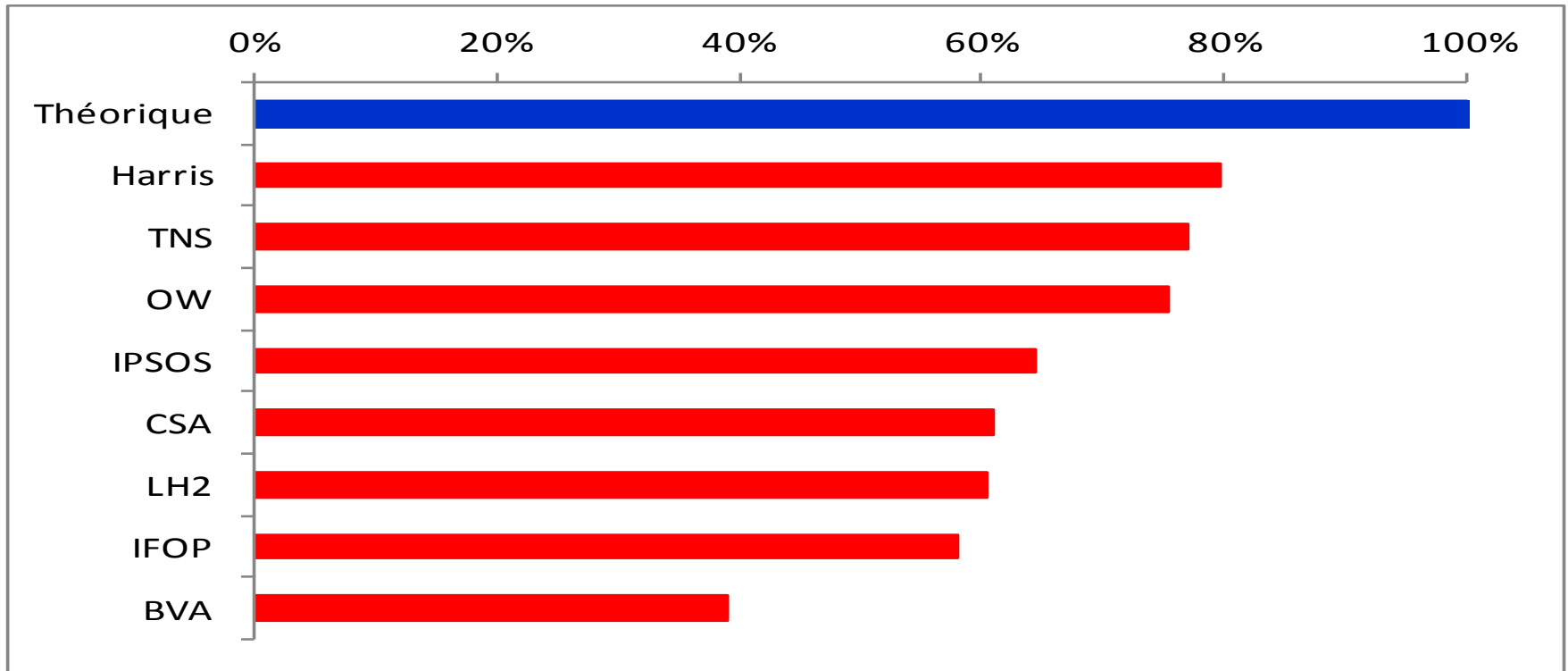
- ▶ Par exemple, BVA déclare interroger 2167 personnes. Les écarts aux votes sont équivalents à ceux d'un sondage aléatoire simple sur 328 personnes, IC[57;625].
- ▶ Les instituts sont loin de faire aussi bien que l'aléatoire simple...

# Marges d'erreur des instituts en 2012 pour le dernier sondage ( à p=25% )



- ▶ Pour un candidat dont l'intention de vote est de 25% :
  - ▶ L'analyse du dernier sondage de chaque institut montre des marges d'erreur entre 3,2 et 4,7 points.
  - ▶ Les marges « théoriques » sont entre 1,8 et 2,7 points.

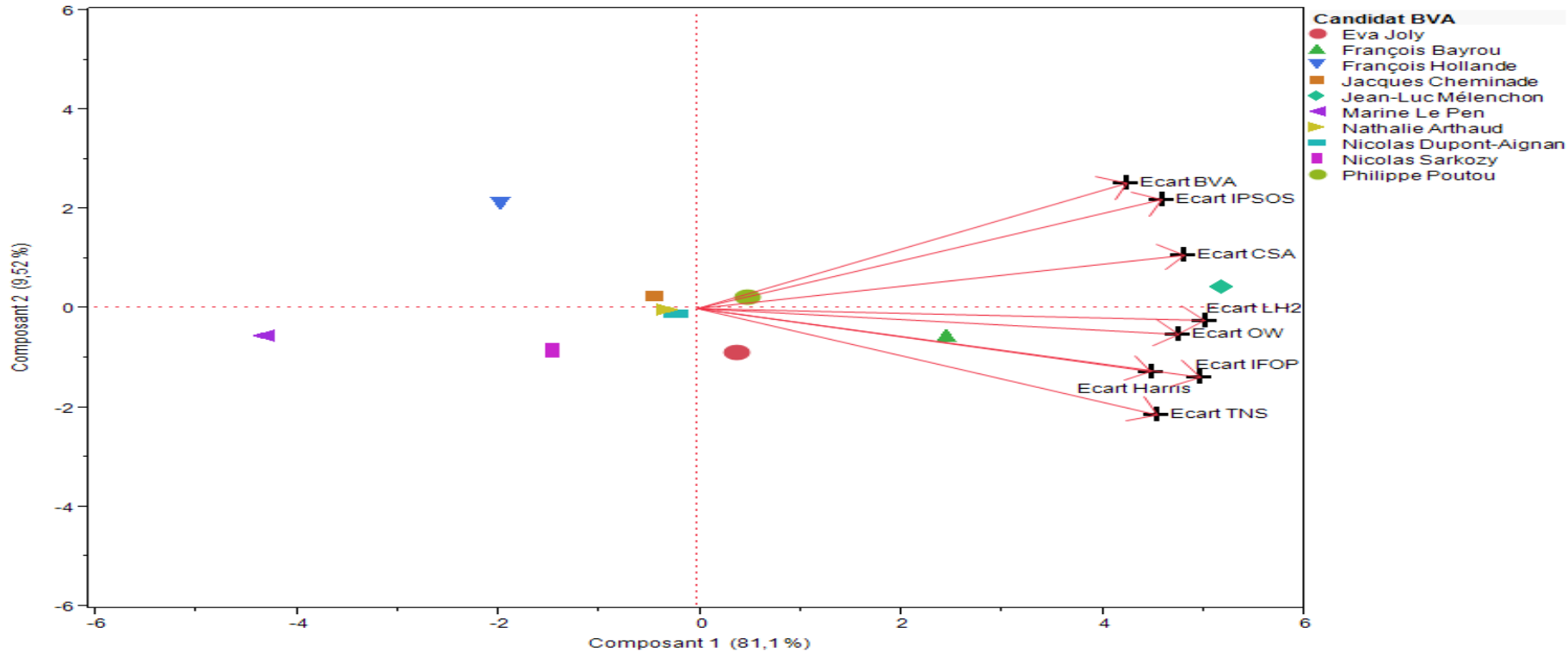
# Classement : Efficacité des instituts en 2012 sur le dernier sondage (rapport des marges)



L'efficacité des instituts ne dépasse pas les 80%.

BVA a une efficacité inférieure à 40%.

# Remarque : Des biais sont-ils à l'origine de l'inefficacité des instituts de sondage ?



- ▶ Une ACP sur les erreurs du dernier sondage de chaque institut montre une très forte corrélation entre instituts...!

# Les biais participent aux erreurs et donc aux marges d'erreur

---

► Pour Pascal Ardilly (Commission du Sénat 2011 )

*« Dans tous les cas, il est probablement nécessaire de réfléchir au contenu des encadrés méthodologiques accompagnant la diffusion des résultats, dont l'obligation serait bienvenue, [...] rappelant l'existence d'erreurs autres que la variance d'échantillonnage (la diffusion de cette seule composante pouvant faire illusion et tromper les utilisateurs sur la nature de l'erreur totale). »*

**ERREUR TOTALE = ERREUR D'ÉCHANTILLONNAGE + ERREUR  
D'OBSERVATION + ERREURS DUES AU DÉFAUT  
DE COUVERTURE ET A LA NON-RÉPONSE**

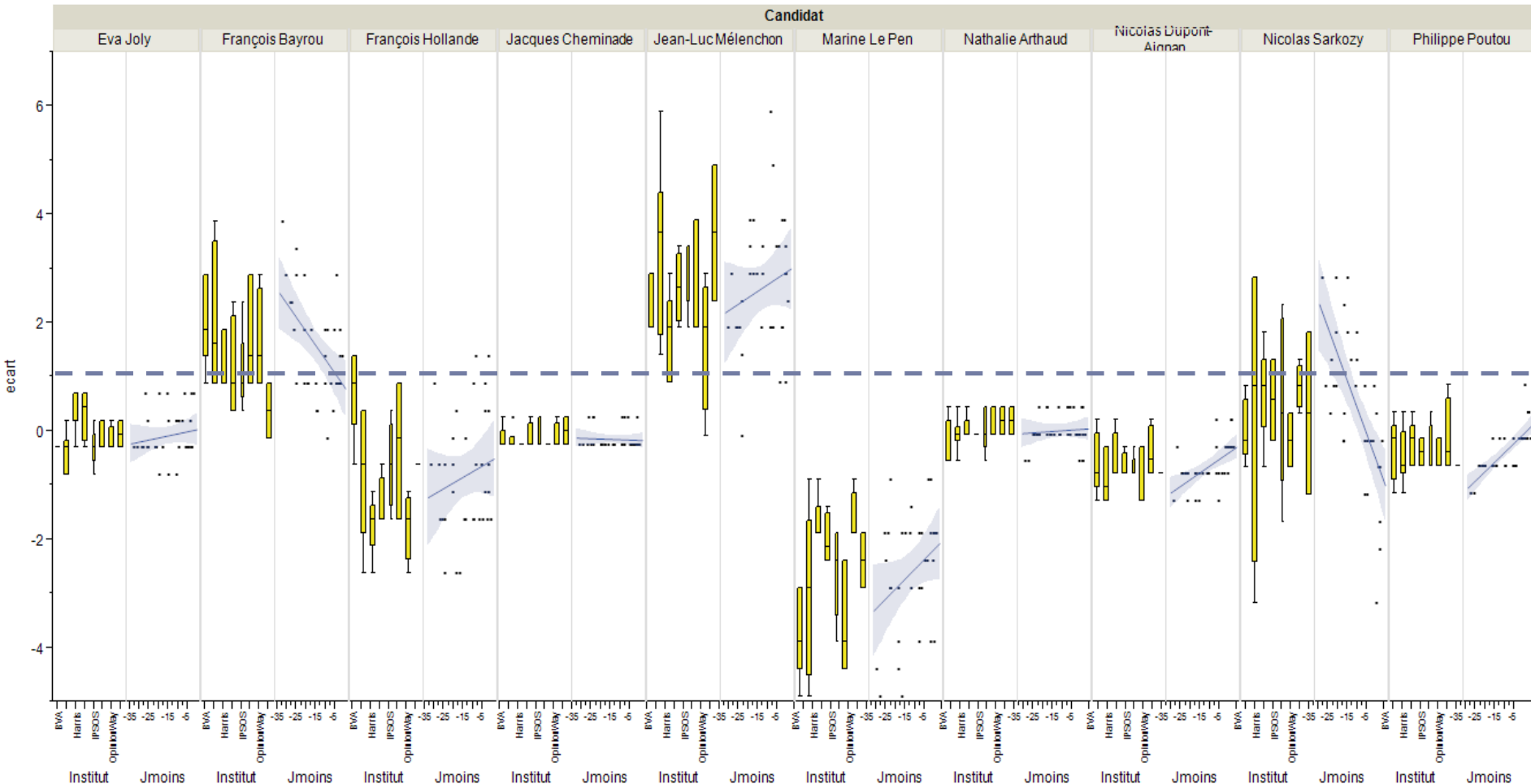
# Approche 2 : Estimation et intégration des biais

---

- ▶ Pour l'estimation des biais, nous avons analysé par ANOVA sur les erreurs, l'ensemble des sondages réalisés depuis le 20 Mars 2012 (liste officielle des candidats). Soit 35 sondages .
- ▶ Pour chaque candidat, une ANOVA incluant les facteurs Institut, Date et Institut\*Date a été réalisée.
- ▶ Dans ce cadre, **la constante de chaque modèle** est une estimation du biais de sondage moyen (de l'ensemble des instituts) associé au candidat.

Ensemble des sondages (hors rolling) réalisés par les 8 instituts autorisés, à partir du 20 Mars (liste officielle des candidats). Soit 35 sondages.

## Ecarts entre les sondages et les résultats du vote (1<sup>er</sup> tour 2012).





## Significativité des ANOVA pour chaque candidat

Facteurs (institut , date , institut \*date); Réponse : Ecart au réel

Candidat	ANOVA	Constante (Biais)	Valeur	Effet Institut	Effet Date	Interaction
Eva Joly	S	NS	0,1	S	NS	NS
François Bayrou	NS	NS	0,5	-	-	-
François Hollande	NS	NS	-0,6	-	-	-
Jacques Cheminade	NS	S	-0,2	-	-	-
Jean-Luc Mélenchon	NS	S	3,3	-	-	-
Marine Le Pen	S	S	-2,0	S	S	NS
Nathalie Arthaud	NS	NS	0,1	-	-	-
Nicolas Dupont-Aignan	NS	S	-0,3	-	-	-
Nicolas Sarkozy	S	S	-1,1	NS	S	S
Philippe Poutou	S	NS	0,1	NS	S	NS

- ▶ On observe une surestimation significative de plus de 3 points pour JLM et une sous-estimation de 2 points pour MLP.
- ▶ Le biais maximal de chaque institut sur l'ensemble des candidats est également estimé (noté  $\hat{B}_{IMax}$ )

# Formule de la marge d'erreur calculée à partir d'un estimateur biaisé

---

- ▶ Soit  $p_k$ , une proportion dans la population, soit  $\hat{p}_k$  un estimateur de cette proportion tel que :

$$E(\hat{p}_k) = p_k + B \quad V(\hat{p}_k) = \frac{(p_k + B)(1 - p_k - B)}{n}$$

- ▶ On montre dans le cas aléatoire simple que :

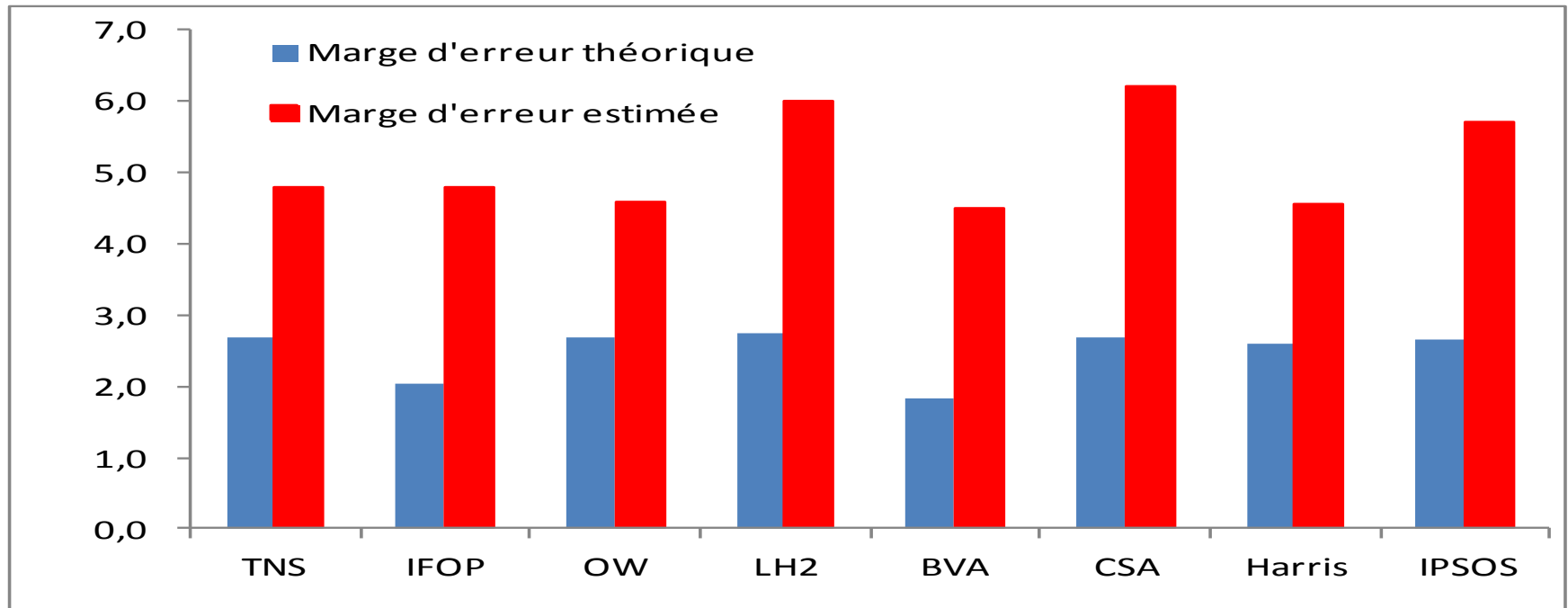
$$M_E \approx z \sqrt{\hat{p}_k (1 - \hat{p}_k) / n}$$

- ▶ où, en notant  $\Phi$  la fonction de répartition de la loi normale standard,  $z$  est tel que :

$$\Phi\left(\left|\hat{B}_{IMax}\right| / \sqrt{\frac{\hat{p}_k (1 - \hat{p}_k)}{n}} + z\right) - \Phi\left(\left|\hat{B}_{IMax}\right| / \sqrt{\frac{\hat{p}_k (1 - \hat{p}_k)}{n}} - z\right) = 1 - \alpha$$

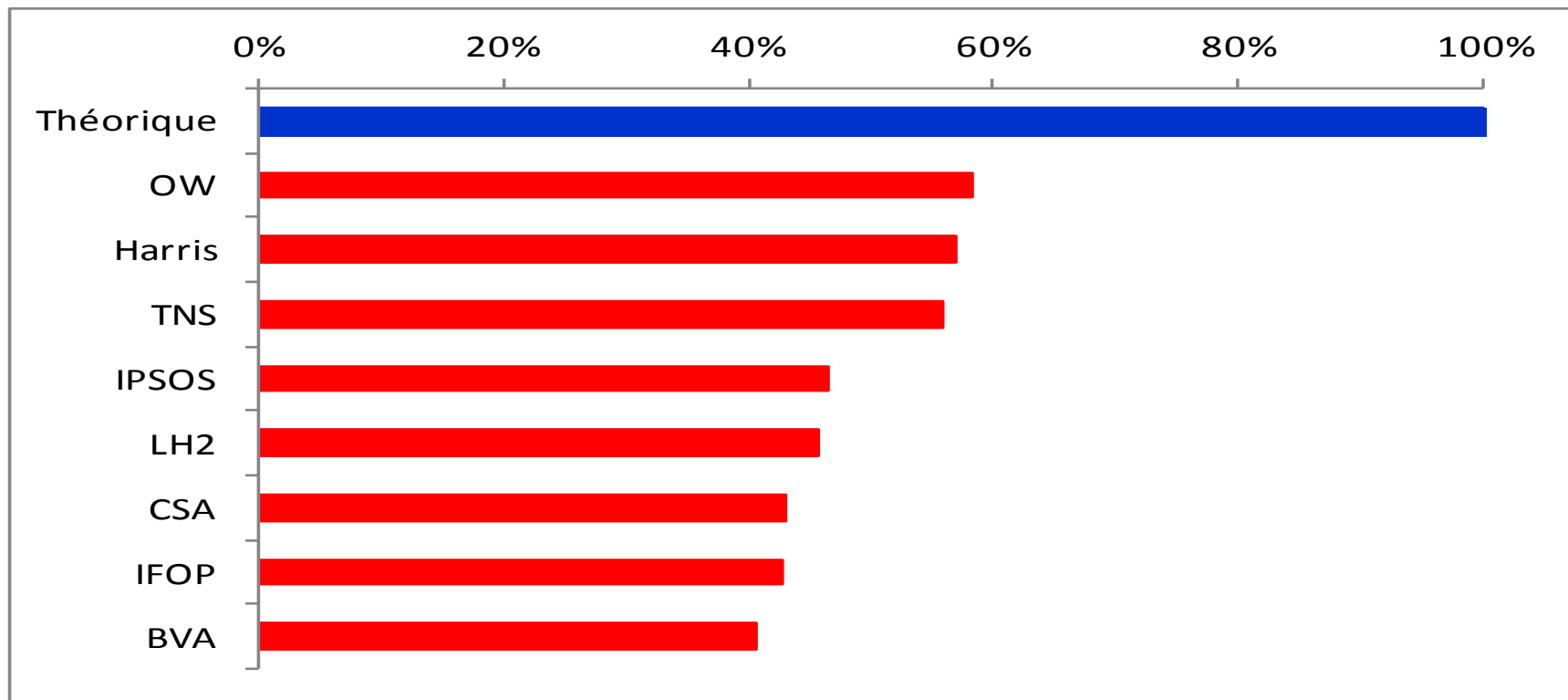
- ▶ Le biais augmente la marge d'erreur.

# Marges d'erreur des instituts en 2012 pour les sondages du dernier mois( p=25%)



- ▶ Pour un candidat dont l'intention de vote est de 25% :
  - ▶ L'analyse de l'ensemble des sondages du dernier mois montre des marges d'erreur entre 4,5 et 6,2 points.
  - ▶ Les marges « théoriques » sont entre 1,8 et 2,7 points.

# Classement : Efficacité des instituts en 2012 pour les sondages du dernier mois (rapport des marges)



L'efficacité des instituts ne dépasse pas les 60%

BVA a une efficacité inférieure à 40%

# Vers une charte graphique de la publication d'un sondage... un exemple



## Les intentions de vote au premier tour de l'élection présidentielle

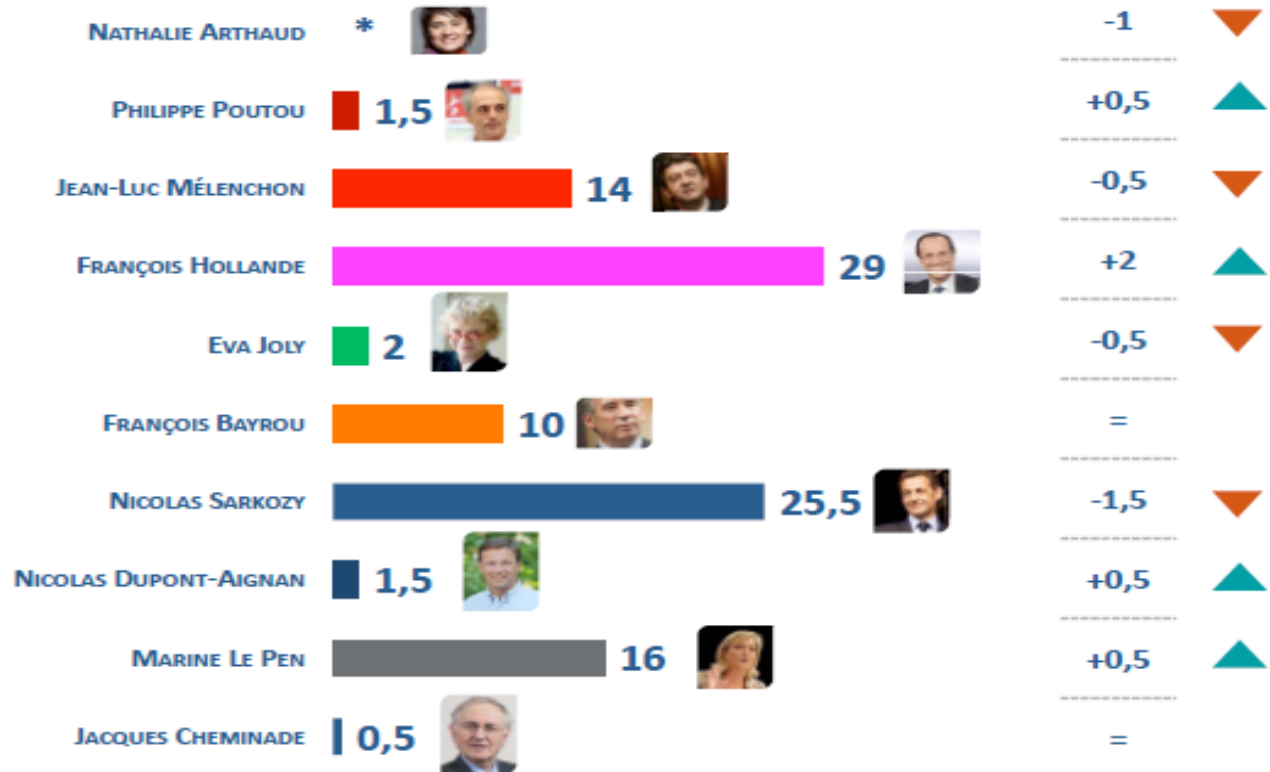


Quel est le candidat pour lequel il y a le plus de chance que vous votiez dimanche prochain, lors du premier tour de l'élection présidentielle ?

Base : inscrits, certains d'aller voter, exprimés

18-19 avril

Evolution 13-14 avril



16% des personnes interrogées et certaines d'aller voter n'ont pas exprimé d'intention de vote

\* Score inférieur à 0,5 %

# Vers une charte graphique de la publication d'un sondage...

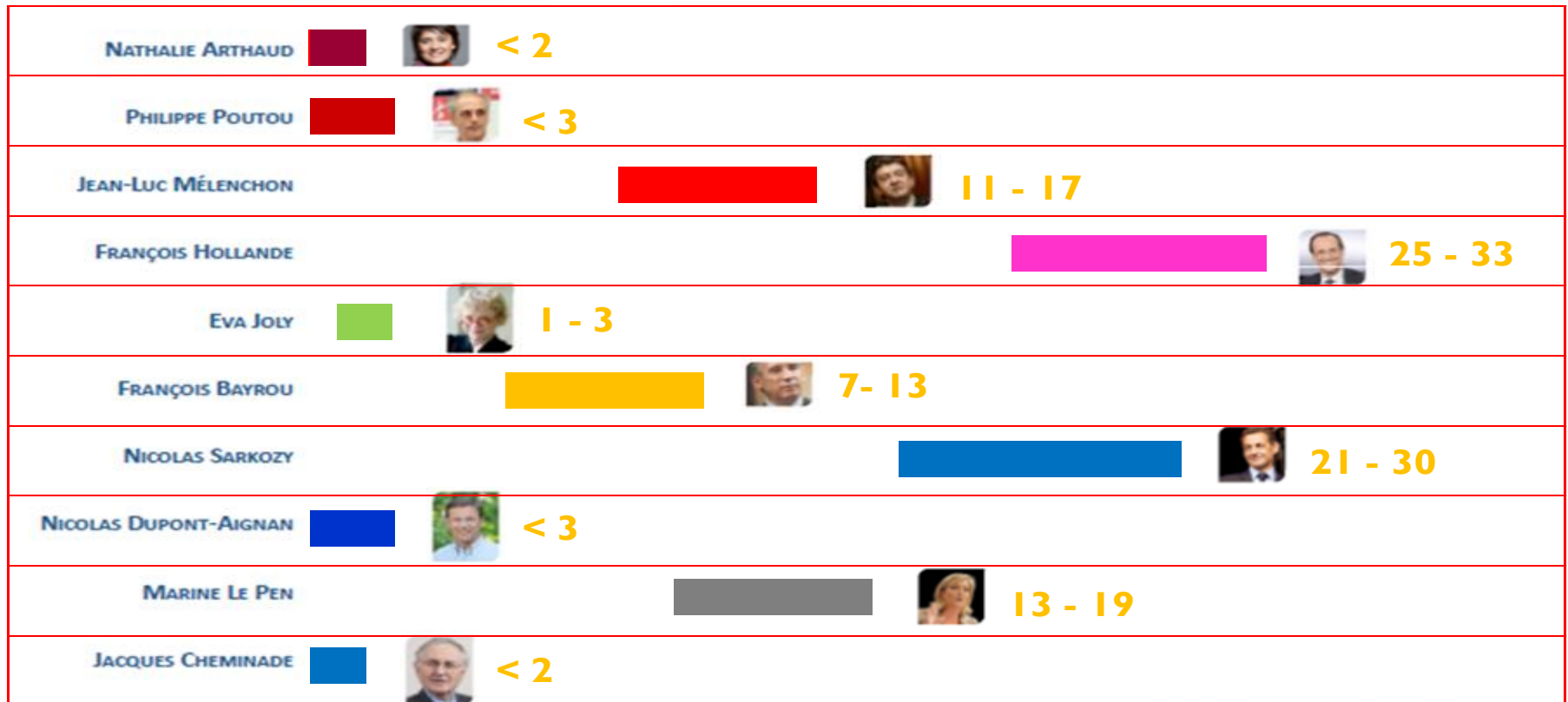


## Les intentions de vote au premier tour de l'élection présidentielle



Estimation au 20 Avril 2012

Base : inscrits, certains d'aller voter, exprimés



# Pour conclure : que faire en 2017...

---

- ▶ Les approches proposées permettent d'estimer (certes a posteriori) les marges d'erreur de chaque institut par référence à la méthode aléatoire.
- ▶ En 2012, les mesures d'intention de votes du premier tour avaient des marges d'erreur bien supérieures à celles sous-entendues par les instituts de sondage.
- ▶ Pour 2017, des marges d'erreur crédibles sont celles observés, pour chaque institut, en 2012.
- ▶ Pour 2017, une charte graphique, statistiquement honnête et compréhensible par tous, est possible et nécessaire.

# Bibliographie

---

- ▶ [1] Ardilly P. (2006), Les techniques de sondage, Edition TECHNIP.
- ▶ [2] Gerville-Réache L., Couallier V. et Paris N. (2011) : Echantillon représentatif (d'une population finie) : définition statistique et propriétés. Preprint HAL-00655566, 11p.
- ▶ [3] Kruskal W., Mosteller F. (1979) Representative Sampling, III: The Current Statistical Literature. International Statistical Review Vol. 47, No. 3, pp. 245-265
- ▶ [4] Neyman J. (1934) On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Journal of the Royal Statistical Society, Vol. 97, No. 4, pp. 558-625
- ▶ [5] Norme ISO 20252 (2006) Études de marché, études sociales et d'opinion - Vocabulaire et exigences de service.
- ▶ [6] Portelli H., Sueur JP. (2010), Rapport d'information du Sénat n°54 sur les sondages.
- ▶ [7] <http://www.sondages-en-france.fr/sondages/Elections/>