

Elections au CA d'Attac – juin 2006

François Sermier

formateur indépendant

Paris, 1^{er} septembre 2006

Objet du présent document

Lors des élections au conseil d'administration d'Attac, des indices sérieux ont pu laisser penser que des éléments anormaux étaient venus perturber le bon déroulement du processus électoral.

J'ai pris conscience de ces problèmes vers la mi-août par la presse. Je ne suis pas membre de l'association ni ne l'ai été, mais je suis sensible aux thèmes qu'elle entend explorer et aux dimensions touchant à l'éducation du « public », la diffusion d'informations et la mise en débat des grandes options économiques.

Je suis travailleur indépendant, statisticien de formation. J'interviens en formation permanente dans des stages proposant à des salariés en cours d'emploi d'approfondir la production et l'exploitation de données et d'en tirer des analyses pertinentes en utilisant notamment des représentations graphiques compréhensibles et parlantes.

Enfin, je ne suis membre d'aucune organisation politique ; je suis membre de la Société Française de Statistique dont j'ai été plusieurs années secrétaire générale et de l'association Pénombre qui s'intéresse à l'usage du nombre dans le débat public.

J'ai pris connaissance des données ([anatomie.xls](#)) et des différents rapports (experts et synthèse par R. Passet) à partir du site de Michel Husson ([hussonet.free.fr](#)) que je ne connais pas par ailleurs. Comme les experts, je ne connais pas la constitution des « listes » virtuelles qui circulaient avant le vote. Je me suis basé pour le déroulement de la procédure sur le contenu des rapports et la [synthèse](#) rédigée par Franz.

L'objectif de ce document est de proposer une présentation principalement graphique, accessible au non statisticien, des éléments quantifiés du débat sur les « interventions ».

Le constat

Les résultats

La figure 1 présente les pourcentages obtenus par chacun des candidats, repéré par son rang, en regroupant chronologiquement les voix obtenus lors des dépouillements quotidiens des différents lots. Dans les différents rapports, ces chiffres sont appelés « cumul *jj* », où *jj* est le quantième du mois de juin du jour de dépouillement.

D'après l'exposé du déroulement de la procédure, celle-ci a nécessité un temps dépassant ce qui a été prévu, provoquant des difficultés d'organisation du dépouillement des presque 6 000 votes (vote par liste, en principe entre 12 et 24 candidats retenus sur chaque bulletin). La

prolongation du dépouillement a favorisé la diffusion plus ou moins large de résultats partiels. Enfin, à l'exception du vote du 17 juin (vote direct ou par procuration) où les votants étaient susceptibles de connaître un ou plusieurs résultats partiels, l'organisation du dépouillement des votes par correspondance ne permet pas d'établir un lien entre la date du dépouillement et le résultat du vote. Les éventuels « tendances » des votants étant éclatées par ordre alphabétique et donc, *a priori*, brassées entre elles.

L'examen du graphique permet de repérer immédiatement la journée du 17 juin, où la comparaison du vote du jour et du résultat final met en évidence une forte divergence :

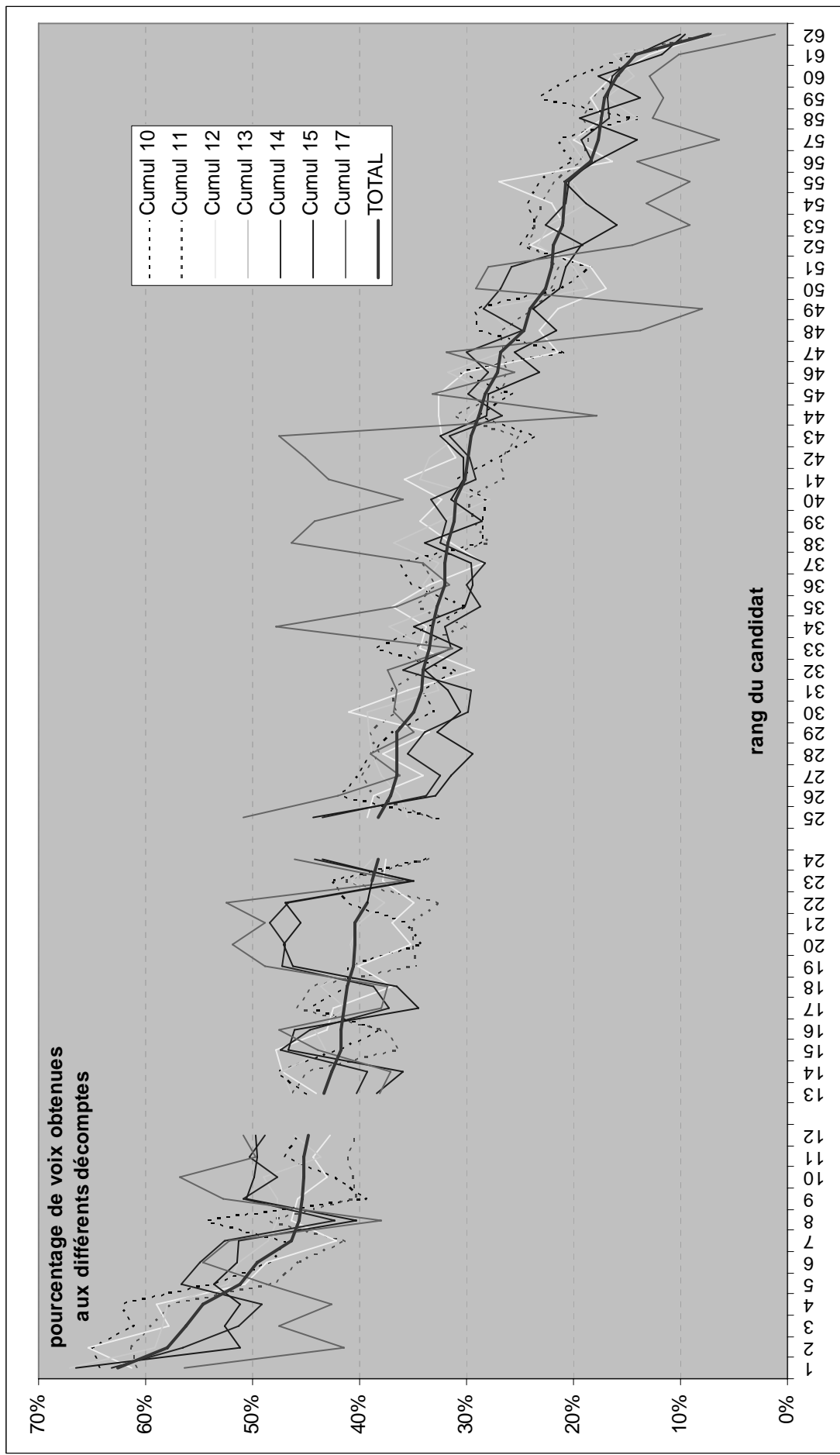
- les candidats extrêmes (rangs 1 à 4 et rangs 48 à 62, hormis 50 et 51) reçoivent un pourcentage de voix très inférieur à leur résultat global,
- à l'inverse, les candidats de rangs 9 à 12, 19 à 25 sauf le 23, 34 et 38 à 43 reçoivent un pourcentage de voix nettement supérieur à leur résultat global.

Il est frappant de constater le caractère groupé de ces divergences : un candidat « rattrapé » par le vote en AG a de grandes chances d'être environné de candidats dont l'évolution est similaire. On note également que les candidats préférés le 17 juin ont des rangs précédant immédiatement les frontières 12 et 24 d'élection au CA ou au CA élargi. Le caractère groupé des divergences est une indication probante de l'existence de listes, au moins implicites.

Soulignons qu'à ce stade du constat, il est exclu de mentionner la possibilité d'une fraude ; on peut juste faire remarquer que le déroulement du dépouillement a introduit une rupture dans l'égalité entre les votants : les participants à l'AG (6% des votants) ayant pu disposer de résultats partiels avant leur vote. A titre d'illustration, imaginons un 21 avril 2002 où une estimation provisoire des résultats aurait été diffusée aux alentours de 16 heures...

Là où les choses commencent à poser problème, c'est que ce constat, poussé à l'extrême pour le vote du 17, se retrouve dans les cumuls des 14 et 15 juin au profit d'un sous-ensemble des candidats ayant eu de meilleurs résultats le 17 (rang 7, 9 à 12, 19 à 25, sauf le 23). ou au détriment de ceux qui auront de mauvais résultats le 17 (candidats de rang 2 à 4, 8, 13, 14, 17, 18, 23). En revanche, les rangs 26 à 31 ont de moins scores les 14 et 15 sans que cela se retrouve le 17 juin et en sens inverse, les rangs 48 à 62, hormis 50 et 51 s'écartent moins de leur moyenne.

Fig. 1 : Pourcentages de voix obtenus par les candidats aux différents cumuls



Éléments d'appréciation statistique

Le constat précédent est « naïf » au sens où il compare les pourcentages à chaque cumul avec le résultat global. On peut essayer de quantifier l'écart. On utilisera un calcul approximatif simple qui a le mérite de fournir une indication de l'ordre de grandeur de l'écart.

Pour le comprendre, nous allons considérer le candidat de rang 6 dont le résultat final est très proche de 50%. Si la répartition des bulletins entre les différents jours de dépouillement a été réalisée aléatoirement, on peut alors assimiler le résultat du vote à la réalisation de 5800 expériences de pile ou face que l'on aurait décidé de réaliser en plusieurs jours (c'est long !) en réalisant 285 jets le 10 juin, 2373 le 11 juin, ... , 364 jets le 17 juin. Bien sûr, les résultats partiels des 7 jours d'expérience ne fourniront pas tous la valeur d'ensemble de 50% et se répartiront (se disperseront) *autour* de cette valeur. La théorie des probabilités nous fournit une mesure de la dispersion autour de la moyenne¹.

En divisant l'écart d'un résultat quotidien au résultat global par cette quantité, on obtient un indicateur standardisé (et sans dimension) lié à la probabilité d'obtenir ledit résultat quotidien. Si cet indicateur est voisin de zéro (en valeur absolue), le résultat obtenu est un résultat courant ne posant aucun problème. Si au contraire la valeur absolue est grande, on est en présence d'un résultat rare, voire exceptionnel. A titre indicatif, « rare » (une chance sur vingt) commence au-delà de plus ou moins 2, « exceptionnel » (trois chances sur mille) au-delà de 3 en valeur absolue. Un résultat de 7 est ahurissant.

On notera que si la présence de valeurs rares est source d'interrogations, la trop grande proximité des résultats à la valeur attendue est aussi problématique. Dans le cas du candidat de rang 6, on obtient par exemple :

cumul	résultat	votants	indicateur
11	45,7%	2 373	-3,7
15	54,9%	857	1,1
17	54,7%	364	2,0
total	49,5%	5 838	

Le résultat du 11 juin (la valeur absolue de l'indicateur supérieure à 3,7) a ainsi une probabilité de l'ordre de deux sur dix mille de se réaliser...

Remarques techniques (facultatives !)

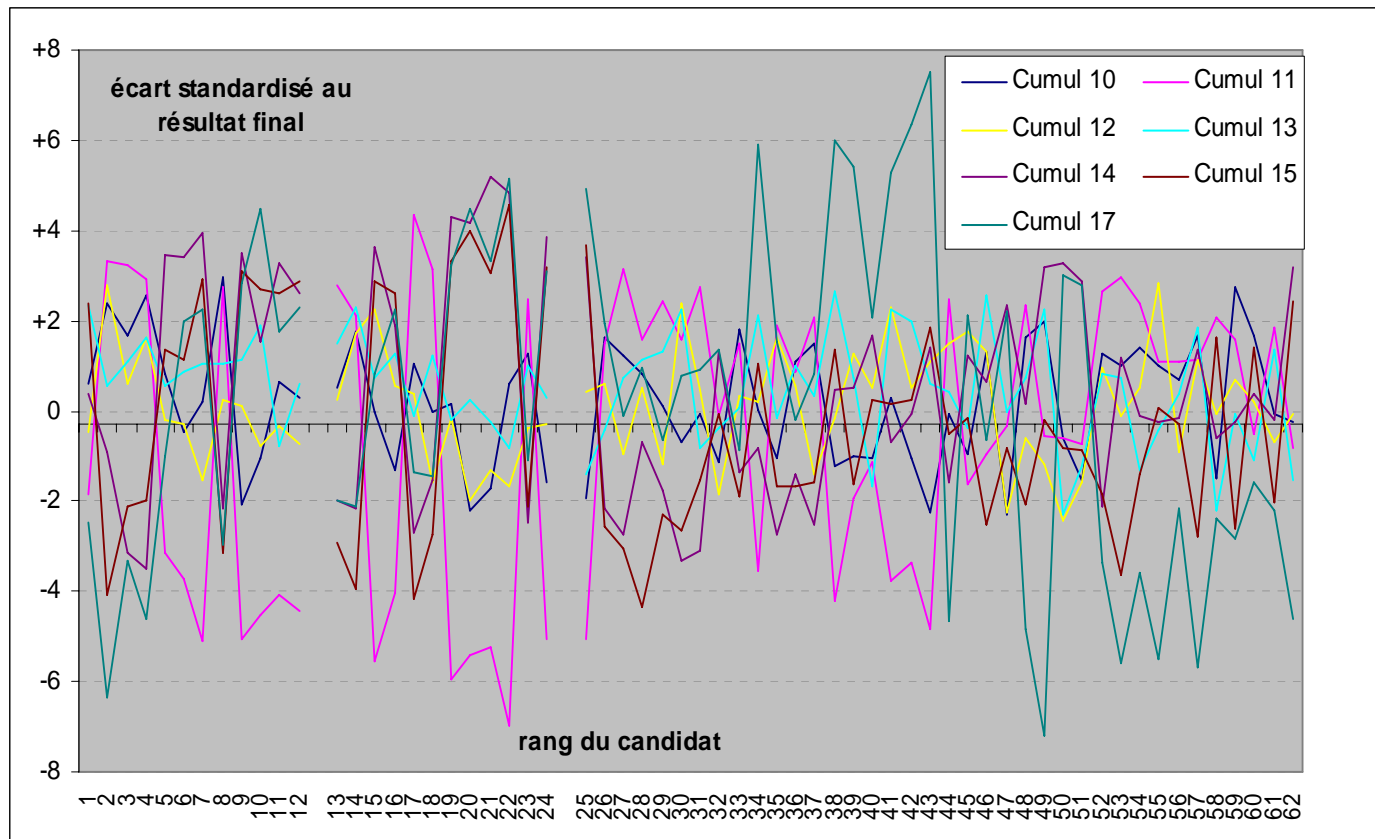
On notera d'abord que les « raretés » ne se cumulent pas : si l'on a un résultat anormalement bas pour une journée, il faut s'attendre à en trouver un anormalement haut. Après tout, il n'y a que 7 jours et le résultat global est la moyenne (pondérée) des 7 résultats intermédiaires.

Le calcul théorique s'applique pour l'exemple de pile ou face décrit. L'étude d'un seul candidat s'y ramène sans trop de problème (le choix du rang 6 n'est justifié que par sa proximité avec 50%). Quand on examine les 62 candidats, le problème est plus complexe... Néanmoins, l'ordre de grandeur reste valide.

¹ On utilise ici la formule bien connue de l'écart-type de l'estimateur de la moyenne d'un échantillon extrait d'une loi binomiale : $\sqrt{\frac{p(1-p)}{n}}$, où p est la proportion dans la population (ou son estimation) et n la taille de l'échantillon.

L'utilisation de la méthode exposée a permis de détecter une erreur matérielle dans le comptage de l'huissier (oubli du chiffre des centaines pour le candidat de rang 4, dans le dépouillement du 10 juin, total reporté de 77 voix au lieu de 177 comptabilisées). Erreur identifiée par un échange de courriers avec M. Husson. Cf. feuille « écarts2 » dans le classeur Excel : jouer avec la case à cocher présente au dessus du graphique.

Fig. 2 : indicateurs d'écart standardisé (voir texte)

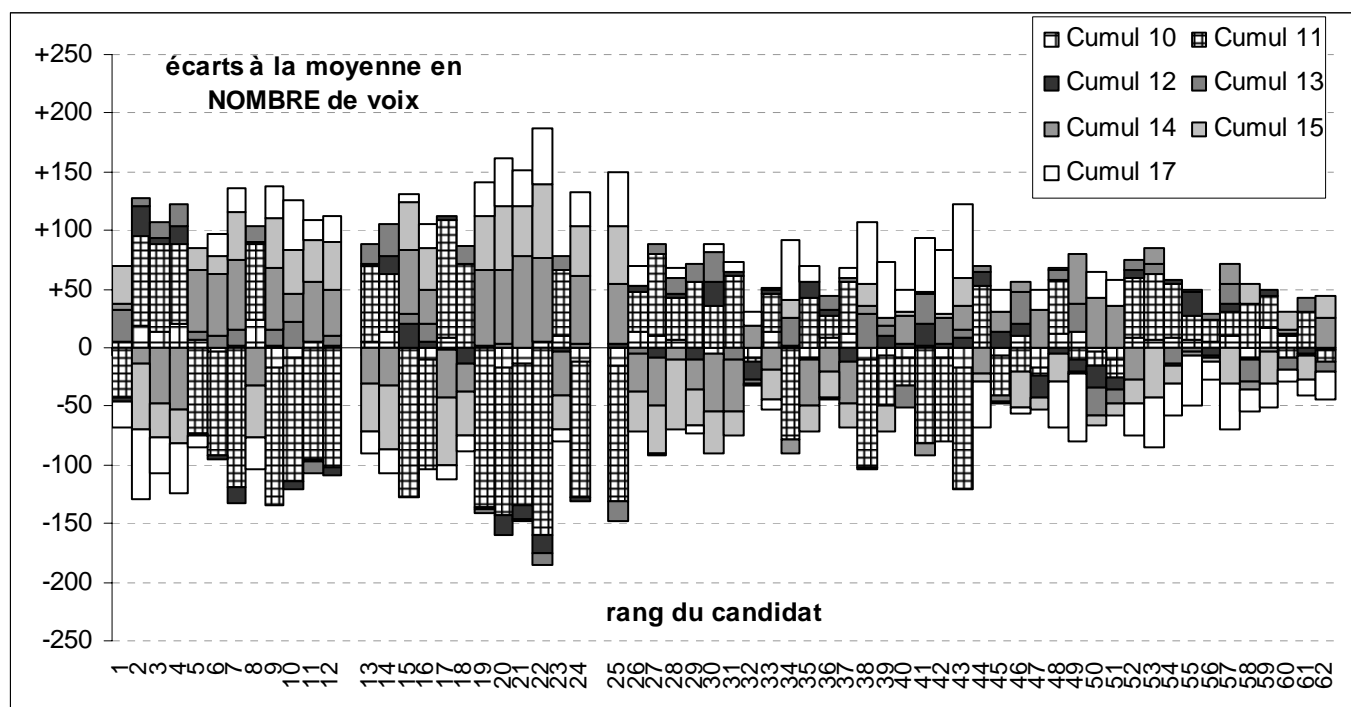


L'examen de la figure 2, hors « cumul 17 » pour lequel les grands écarts sont « explicables », révèle un nombre très important d'écarts « rares » et un grand nombre d'écarts plus qu'« exceptionnels » (cf. par exemple la journée du 11 juin.)

Visualisation en nombre de voix

Après l'appréciation statistique, essayons de visualiser l'ampleur du phénomène en nombre de voix. Dans la figure 3, nous avons représenté pour chaque candidat, l'écart de chaque cumul quotidien au nombre de voix théoriques correspondant au pourcentage final, en cumulant les différents jours de dépouillement. Pour faciliter la lecture, les jours 10 et 11 sont représentés avec une trame quadrillée (plus ou moins serrée) 12, 13, 14 et 15 par un dégradé de gris uni et le 17 juin en blanc.

Fig. 3 : cumul des écarts à la moyenne, en nombre de voix



Dans ce graphique, où, par construction, l'écart total positif et l'écart total négatif s'annulent (la hauteur de chaque colonne au-dessus de l'axe horizontal est la même que la profondeur au-dessous de l'axe), on perçoit très clairement la très forte opposition entre les deux premiers jours de dépouillement et les autres.

La figure 3 fournit également un ordre de grandeur de l'ampleur des manipulations éventuelles. En effet, pour les candidats (de rang 19 à 23) où l'écart entre les deux résultats partiels cumulés (10-11 juin et 12-17 juin) est la plus forte, on trouve une différence d'environ 300 voix (environ 5% des suffrages exprimés). L'obtention d'un tel écart suppose l'introduction de 300 bulletins modifiés en faveur des « bénéficiaires » ou la suppression d'un même nombre de bulletins en faveur des « victimes ». Cette opération est difficile à réaliser puisqu'elle modifie le nombre total de bulletins.

On peut aussi obtenir un résultat similaire en substituant environ 150 bulletins favorables à la liste « amie » à la place de 150 bulletins de la liste « ennemie ». Cette manipulation ne modifie pas le nombre de suffrages exprimés, mais elle suppose que le(s) manipulateur(s) ai(en)t pu opérer un tri entre les deux types de bulletins de vote, préparer un nombre équivalent de bulletins « amis » et réaliser la substitution.

Enfin, une dernière possibilité peut être envisagée, elle consiste à « compléter » des bulletins « ennemis » ne comportant pas 24 candidats en y ajoutant des candidats « amis ». C'est la façon de procéder la plus discrète car elle ne nécessite ni addition, ni soustraction, ni substitution de bulletins de vote. Elle ne peut être repérée que par une étude de l'ensemble des bulletins de vote pour détecter l'éventuelle surreprésentation l'un ou l'autre jour de bulletins panachant les deux listes en présence. Dans ce cadre, il faut rajouter environ 200 à 300 croix au profit d'une vingtaine de candidats, soit un total d'environ 5 000 désignations de candidats frauduleuses.

Une étude approfondie, lot par lot, des conditions matérielles de dépouillement et de conservation des lots est seule susceptible de mettre en évidence les conditions qui auraient rendu possible une ou plusieurs manipulations telles que décrites ici.

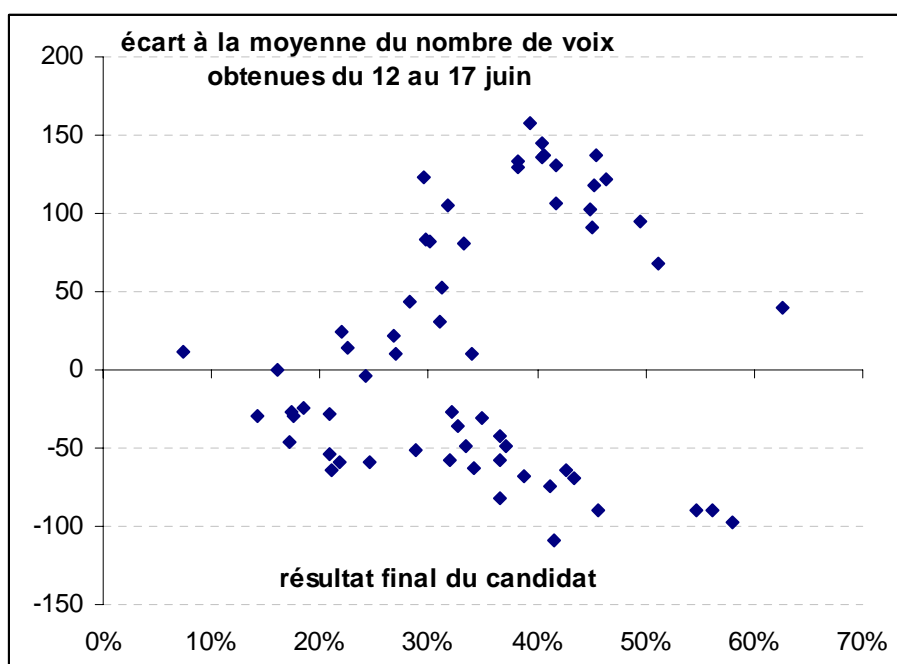
Graphique « discriminant »

On terminera cette partie en constatant que la figure 3 n'est pas éloignée d'un graphique présenté dans l'un des rapports d'experts.

Cherchons à identifier dans la figure 3 les listes éventuelles, c'est-à-dire dans ce cas les groupes de candidats ayant des profils identiques quant à leurs résultats partiels quotidiens. On constate alors que présenter le cumul des deux premiers jours *et* le cumul des cinq autres fait double emploi et alourdit le graphique. En fait, il s'agit là de la remarque déjà émise sur les tailles identiques des colonnes positive et négative de chaque candidat. Supprimons donc l'une des deux et conservons donc uniquement la somme des cumul 12 à cumul 17.

Sur l'axe horizontal, remplaçons simplement le rang du candidat (de 1 à 62) par son résultat final (en voix ou en pourcentage, cela correspond à un changement d'échelle). Il ne reste plus qu'à choisir un graphique de type « nuage de points ». On obtient alors la figure 4.

Fig. 4 : regroupement des candidats en fonction des résultats quotidiens



Au passage des scores au rang près, et compte tenu du changement de direction de l'axe horizontal (les rangs croissants correspondent à des scores décroissants) la figure 4 est voisine de la figure 3. Pour s'en convaincre, il suffit de modifier dans le fichier Excel joint, sur la feuille « écarts3 » de modifier le type du graphique et de le passer alternativement en histogramme ou en nuage de points.

La figure 4 est extrêmement voisine du graphique 2 du rapport Chiche-Rouanet. La succession de raisonnements graphiques « naïfs » retrouve une représentation voisine de celle des candidats dans le plan principal de l'analyse en composantes principales. L'ACP de Chiche-Rouanet confirme sur des bases théoriques solides qu'en dehors du classement des candidats, l'autre fait marquant du dépouillement est l'opposition absolue entre les résultats des deux premiers jours et ceux des jours suivants.

Examen de deux méthodes éventuelles de manipulation

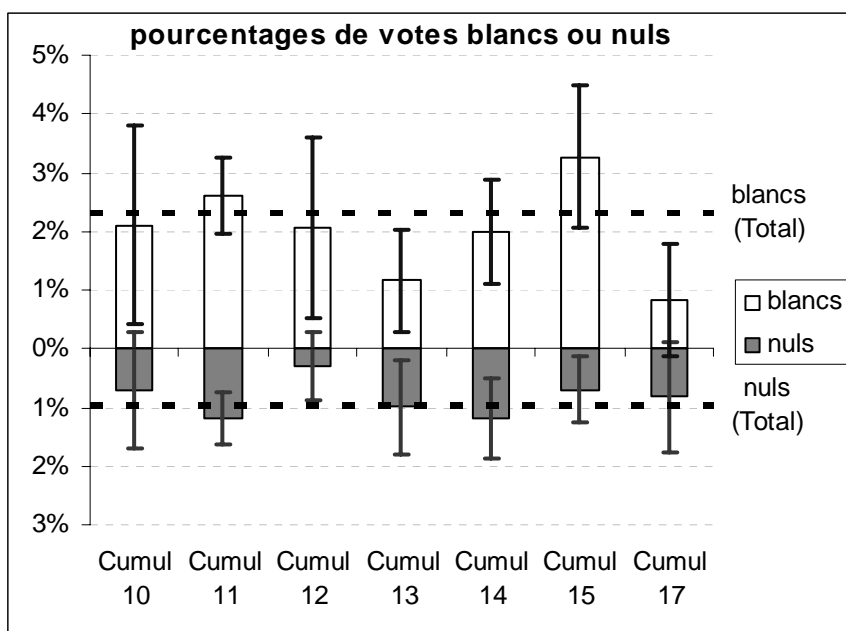
Ayant, comme le rapport de synthèse, conclu sur un constat d'anomalies, constat plus « naïf » que celui des experts, il nous est apparu intéressant d'explorer, dans le même esprit que précédemment, ce que l'on peut obtenir à partir des données disponibles sur les hypothèses de manipulation du dépouillement évoquées plus haut. Nous avons envisagé deux hypothèses de manipulation, hypothèses que l'on peut tester facilement avec les données dont on dispose. Elles se ramènent toutes les deux à une modification physique des bulletins en ajoutant des candidats sélectionnés dans le bulletin².

Ces deux hypothèses sont les suivantes :

- ajout de la désignation de candidats au-delà de la limite de 24 afin de provoquer l'invalidation de bulletins (qui votaient pour la liste « à désavantager »),
- complétion à 24 candidats de bulletins qui ne comportaient 24 candidats désignés, en cochant (?) les candidats « amis ».

Votes blancs ou nuls

Fig. 5 : pourcentages des votes blancs ou nuls



Au final, les pourcentages de votes blancs ou nuls sont de 2,3% et 1,0%. Ils fluctuent lors des dépouillements partiels (les barres d'erreurs matérialisent un intervalle de confiance à 95% en utilisant un indicateur similaire à celui présenté plus haut).

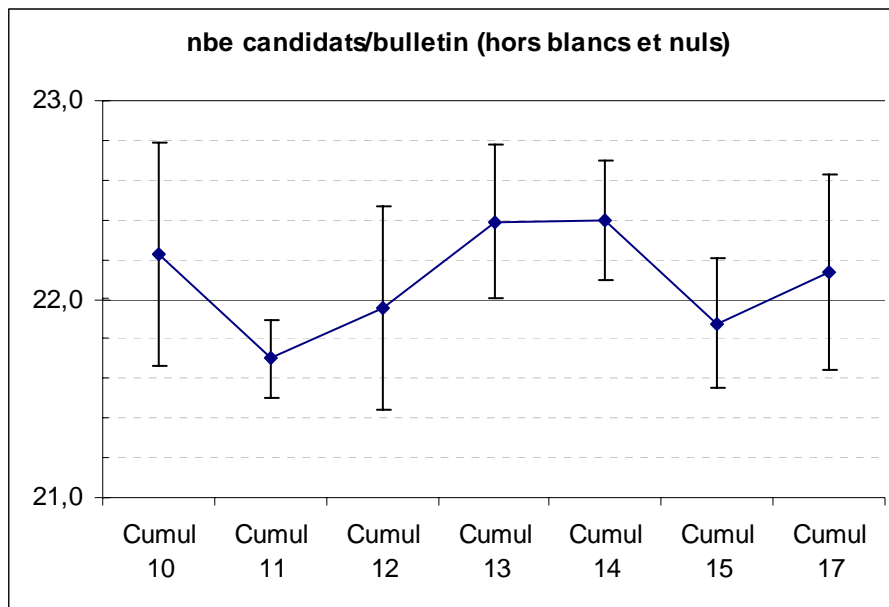
On note un déficit en bulletins blancs le 13 et en votes nuls le 12, sans que le résultat soit vraiment rare et surtout de l'ampleur nécessaire au résultat à obtenir : il y a de l'ordre de 1% d'écart avec un nombre de suffrages exprimés de 600 et 350 bulletins les 13 et 12 juin...

² On notera que je ne dispose pas de l'image d'un bulletin de vote et que mes hypothèses sont éventuellement irréalistes par rapport à la disposition matérielle du bulletin.

Nombre de candidats par bulletin

La figure 6 retrace l'évolution du nombre moyen de candidats par bulletin. Pour un jour de dépouillement donné, il est obtenu en divisant le nombre total obtenus par l'ensemble des candidats par le nombre de bulletins de vote exprimés (hors blancs et nuls). Pour l'ensemble du vote, on obtient en moyenne 22 candidats désignés par bulletin de vote.

Fig. 6 : nombre de candidats par bulletin



On retrouve dans le graphique l'opposition entre le 11 (2 300 bulletins, avec un écart de -0,3 candidat par bulletin) d'une part et les 13 et 14 juin (environ 1 500 bulletins, avec en moyenne 0,4 candidat en plus par bulletin). Les indicateurs standardisés d'écart³ sont supérieurs à 2 (-2,9, +2,1 et +2,7) indiquant un effet significatif.

Cependant, l'ordre de grandeur de l'effet détecté est insuffisant pour obtenir le résultat constaté : entre 600 et 1 000 additions de candidats selon que l'on prend la référence à la moyenne d'ensemble (+0,4 par rapport au 22 de la moyenne d'ensemble) ou au dépouillement du 11 juin (+0,4 -(-0,3)).

En conclusion de cette partie, on ne détecte pas de variation significative du nombre de bulletins blancs ou nuls entre les différentes journées de dépouillement. En revanche, une modification physique lors de dépouillement de lots du 13 et 14 juin par addition de candidats sur des bulletins ne comportant pas initialement le nombre maximum de candidats est envisageable ; elle ne contribue cependant que pour environ 20% à l'amplitude du constat de manipulation.

L'analyse demande bien sûr à être détaillée au niveau des lots.

³ Les indicateurs sont calculés en utilisant l'estimation de la moyenne dans un échantillon distribué selon une loi de Poisson. La formule de la dispersion ($\sqrt{\frac{m}{n}}$, où m est le nombre moyen de candidats par bulletin et n la taille de l'échantillon) et l'utilisation de l'approximation normale sont certainement erronées du fait de la présence de la troncature de la distribution à 24 (inclus), les valeurs supérieures à 24 entraînant la nullité du bulletin. Néanmoins, on peut retenir l'ordre de grandeur.