Stage statistique – probabilités TI graphiques (83 Premium CE & 82 Advanced)	Intervalles de confiance

Énoncé : Lors des élections présidentielles de 2012, une nouvelle polémique sur la fiabilité des sondages est apparue, rappelant celle des présidentielles de 2002.
En effet, de nombreux observateurs se sont interrogés sur les résultats des derniers sondages avant le premier tour des élections et ceux du vote.
Ainsi, considérons le sondage IPSOS pour France Télévisions, France Inter et Le Monde du 20 avril 2012. Pour ce sondage, 1 021 personnes ont été interrogées par la méthode des quotas[1].
Si les résultats du sondage et du vote pour François Hollande, Eva Joly, François Bayrou et les « petits candidats » semblent cohérents, on note des différences importantes pour Nicolas Sarkozy, Marine Le Pen et Jean-Luc Mélenchon :

	Nicolas Sarkozy	Marine Le Pen	Jean-Luc Mélenchon
Sondage IPSOS	25,5 %	16 %	14 %
Vote du 22 avril	27,2 %	17,9 %	11,1 %

À la télévision, le représentant d'IPSOS a justifié ces différences par la « marge d'erreur » des sondages. Sa réponse est-elle acceptable ?
Nous allons déterminer des intervalles de confiance au seuil de confiance de 95 % des résultats de ce sondage.

Partie 1 : Des programmes pour les intervalles de confiance

Rappels : Déterminer un intervalle de confiance, c'est estimer par intervalle une proportion inconnue p à partir d'un échantillon.

- Si f désigne la fréquence observée sur un échantillon de taille n, avec $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors p est élément de l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$ avec un niveau de confiance de plus de 95 %.

- Pour une meilleure précision, sous les mêmes conditions, on estime que p appartient à l'intervalle[2] $\left[f - 1{,}96\frac{\sqrt{f(1-f)}}{\sqrt{n}} ; f + 1{,}96\frac{\sqrt{f(1-f)}}{\sqrt{n}} \right]$ avec un niveau de confiance de plus de 95 %.

a) *En utilisant la première formule, écrire un programme, nommé* **CONFIANC**, *permettant de calculer les extrémités de l'intervalle de confiance et représenter cet intervalle*.

Demander une valeur pour chacune des variables N, F et P.
Placer dans les variables A et B les extrémités de l'intervalle.
Afficher les deux bornes.

```
PROGRAM:CONFIANC
Prompt N,F,P
F–1/√ (N)→A
F+1/√ (N)→B
Disp A,B
```

[1] La méthode des quotas s'appuie sur la répartition connue de la population des électeurs (âge, sexe, situation géographique, catégorie socio-professionnelle, etc.).

[2] Cet intervalle confiance est construit à partir de l'intervalle de fluctuation asymptotique (voir programme).

Pause, pour permettre de noter ces résultats*.
Désactiver toutes les fonctions.
Nettoyer l'écran graphique.
Tracer le segment de (A ; 2) à (B ; 2).
Afficher le point (P ; 2) représenté par un petit carré (choix 2).
Afficher le graphique.

```
Pause
FonctNAff
EffDessin
Ligne(A,2,B,2)
Pt–Aff(P,2,2)
DispGraph
```

* On appuiera sur ENTER pour aller au-delà de l'affichage des deux bornes.

b) *Opérer, de même, avec la deuxième formule. On nommera le programme* **CONFIAN1**.

Pour le programme **CONFIAN1**, remplacer les deuxième et troisième lignes du programme **CONFIANC** par :

```
F–1.96*√ (F*(1–F))/√ (N)→A
F+1.96*√ (F*(1–F))/√ (N)→B
```

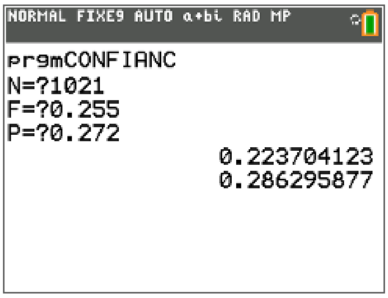
Partie 2 : Une formule simple pour l'intervalle de confiance

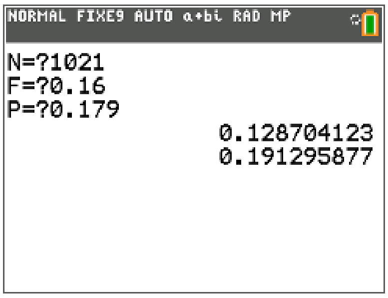
a) *Vérifier que les conditions du sondage vérifient les conditions usuelles de précision.*

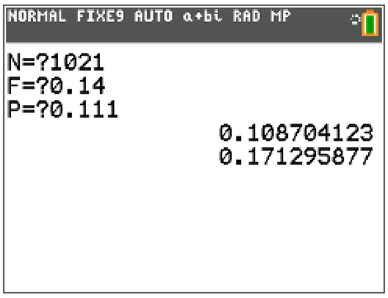
1 021 ≥ 30, dans les trois cas, 0,255 ; 0,16 et 0,14 les produits 1 021 p sont supérieurs à 5.

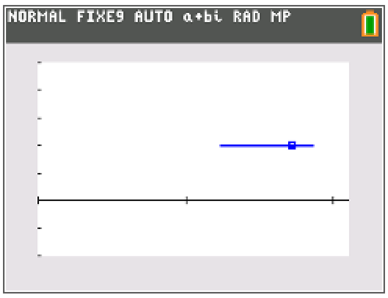
b) *Déterminer, avec le programme* **CONFIANC**, *les encadrements de chacune des proportions concernant les votes des trois candidats cités, à partir des fréquences observées dans le sondage IPSOS.*

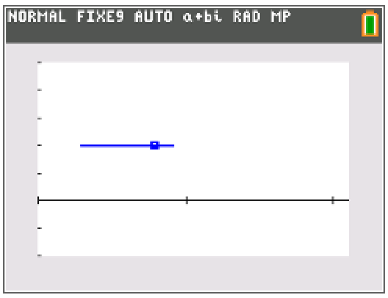
On obtient les écrans ci-dessous.
Pour le graphique, on a choisi une fenêtre d'affichage commune aux trois intervalles :
XMin = 0.1 ; XMax = 0.31 ; Xgrad = 0.1 ; YMin = –2 ; YMax = 5 ; Ygrad = 1.

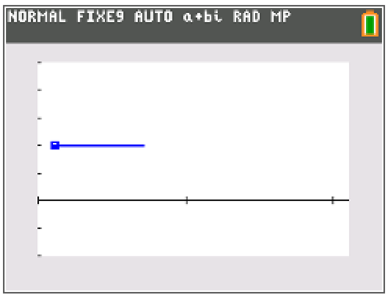












Nicolas Sarkozy | **Marine Le Pen** | **Jean-Luc Mélanchon**

Les intervalles obtenus sont donc peu différents de [0,223 ; 0,287] pour f = 0,255, de [0,128 ; 0,192] pour f = 0,16 et de [0,108 ; 0,172] pour f = 0,14.

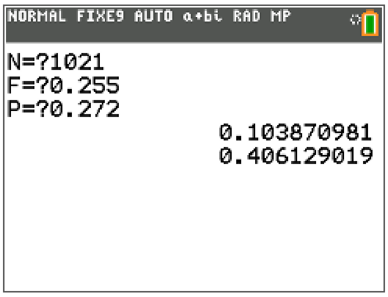
c) *Ces résultats permettent-ils d'affirmer que les « marges d'erreur » justifient les différences entre le sondage et la réalité du vote ?*

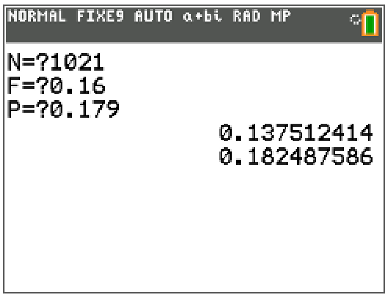
Comme l'indiquent les écrans, on a bien :
0,272 ∈[0,223 ; 0,287] ; 0,179 ∈ [0,128 ; 0,192] ; 0,111 ∈ [0,108 ; 0,172].
Au regard de cette estimation des intervalles de confiance, les résultats du sondage sont donc acceptables au seuil de confiance de plus de 95 %.

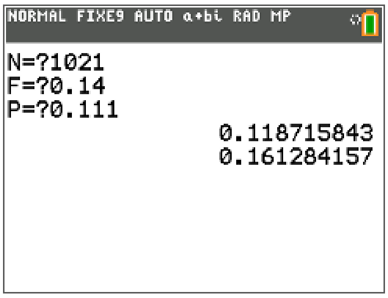
Partie 3 : Une autre formule pour l'intervalle de confiance

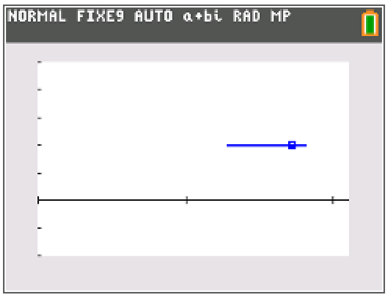
a) *Déterminer, avec le programme* ***confiance1****, les encadrements de chacune des proportions concernant les votes des trois candidats cités, à partir des fréquences observées dans le sondage IPSOS.*

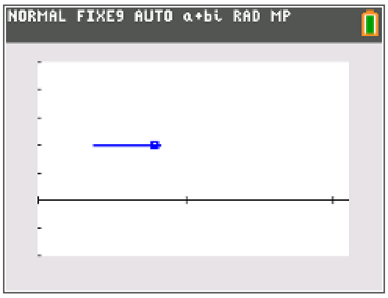
On obtient :

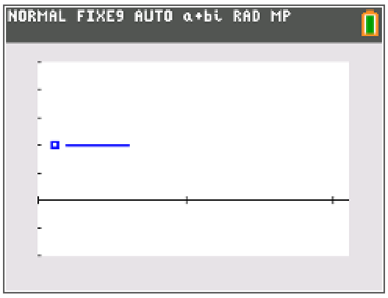












Nicolas Sarkozy | **Marine Le Pen** | **Jean-Luc Mélanchon**

Les intervalles obtenus sont donc peu différents de [0,228 ; 0,282] pour f = 0,255, de [0,137 ; 0,183] pour f = 0,16 et de [0,118 ; 0,162] pour f = 0,14.

b) *Ces résultats permettent-ils d'affirmer que les « marges d'erreur » justifient les différences entre le sondage et la réalité du vote ?*

On a bien : 0,272 $\in$[0,228 ; 0,282] ; 0,179 $\in$ [0,137 ; 0,183]. Mais 0,111 $\notin$ [0,118 ; 0,162].

Au regard de cette estimation des intervalles de confiance, les résultats du sondage sont donc acceptables au seuil de confiance de plus de 95 % pour les résultats de Nicolas Sarkozy et Marine Le Pen, mais pas pour Jean-Luc Mélenchon.

c) *Comment apprécier ces différences ?*

On remarquera – c'est un résultat général – que les intervalles ci-dessus sont contenus dans les intervalles de la première partie, d'où la différence sur l'estimation concernant le sondage de Jean-Luc Mélenchon[3].
Penser que, même dans le cas des intervalles de confiance asymptotiques, 5 % d'entre eux ne contiendront pas la probabilité cherchée[4]…
Enfin, ne pas oublier que les instituts de sondage ne se contentent pas d'établir, par la méthode des quotas, un échantillon le plus représentatif possible de la population des électeurs. Ils corrigent les réponses des personnes sondées en fonction des votes des scrutins du même type des années précédentes. La marge d'erreur peut dépasser de ce fait le simple calcul d'un intervalle de confiance.

[3] La proportion inconnue p appartient à l'intervalle vu en seconde avec une probabilité d'**au moins** 0,95. En revanche, elle appartient à l'intervalle de confiance asymptotique avec une probabilité d'environ 0,95, la valeur approchée étant d'autant plus proche de 0,95 que n est grand.
On remarquera d'autre part que plus l'intervalle est grand, plus il a de chances de contenir la probabilité cherchée…

[4] Noter que tous les résultats que l'on évoque sont valables uniquement pour des échantillons aléatoires…

Compléments

Remarque 1 : on peut automatiser les données de la fenêtre d'affichage en fonction des résultats trouvés. Il suffit pour cela d'ajouter dans le programme, par exemple, **A–0.05→Xmin : B+0.05→Xmax**, etc.

Xmin, **Xmax**, etc. se trouvent dans var **VARIABLES** **1: Fenêtre**.

Remarque 2 : pour faciliter la comparaison entre les deux formules de calcul des intervalles de confiance, on peut afficher, sur un même graphique, pour chaque candidat, les résultats obtenus. (Voir page suivante)

Dans le programme **CONFICOM**, on a regroupé les instructions des programmes **CONFIANC** et **CONFIAN1**.

```
PROGRAM:CONFICOM
EffÉcran
Prompt N,F,P
F-1/√(N)→A
F+1/√(N)→B
F-1.96*√(F*(1-
F))/√(N)→C
F+1.96*√(F*(1-
F))/√(N)→D
Disp A,B,C,D
Pause
FoncNAff
EffDess
Ligne(A,2,B,2)
Ligne(C,3,D,3)
Pt-Aff(P,2,2)
Pt-Aff(P,3,2)
DispGraph
```

On obtient, par exemple, pour Jean-Luc Mélenchon les écrans ci-dessous.