

Statistiques : Enoncé du TP 3 Fourchettes de sondages

On considère une population importante où une élection oppose deux candidats X et Y . La proportion p des personnes qui voteront pour X est inconnue. On prélève au hasard dans cette population un échantillon de taille n de personnes que l'on interroge et on calcule la fréquence f d'opinions favorables à X dans cet échantillon. Le principe du sondage est d'estimer, à partir de la fréquence f observée, la valeur inconnue p .

1^e étape : fluctuation des sondages de taille $n=100$.

1. Ouvrir un nouveau classeur qu'on enregistrera sous le nom **Sondages**. Dans la cellule A1, taper la formule `=ENT(ALEA()+0,4)`. Cette formule permet de simuler un nombre aléatoire de loi de Bernoulli $B(1,p)$ avec $p=0,4$. Pour le comprendre, rechercher en tapant le mot **ENT** dans l'aide en ligne (? en haut de l'écran) le sens de cette fonction.
2. Recopier cette formule par glissement jusqu'en cellule **A100** puis calculer en cellule **A101** la somme des nombres de cette colonne . On considère que la colonne **A** contient à présent les résultats d'un sondage auprès d'un échantillon de 100 personnes, le 1 signifiant que la personne votera pour X et le 0 qu'elle votera pour Y. Calculer en cellule **A102** la fréquence f d'opinions favorables à X.
3. Pour étudier les *fluctuations d'échantillonnage*, c'est-à-dire les variations de la fréquence f lorsqu'on change d'échantillon, on va renouveler 50 fois ce sondage. Recopier les formules de la colonne **A** par glissement en colonnes **B** à **AX** afin de simuler 50 sondages de taille $n=100$. En **A103**, taper Moyenne : et en **B103**, taper la formule `=SOMME(A102 :AX102)/50`.
4. On va regrouper les diverses valeurs de f obtenues en 14 classes d'amplitude 0,025. En cellule **A105**, entrer la valeur 0,25 et remplir les cellules **A106** à **A118** en série de pas 0,025 (la dernière valeur sera 0,575). Indiquer en face, en colonne **B**, les centres des classes en recopiant la formule `=A105+0,0125`. Sélectionner les cellules **C105** à **C118**, puis inscrire dans la barre de formules `=FREQUENCE(A102 :AX102 ; A105 :A118)` puis appuyer simultanément sur les touches **CTRL**, **MAJ**, **ENTREE**. Cela permet l'affichage dans les cellules **C105** à **C118** des effectifs (et non des fréquences !!) de chacune des 14 classes.
5. A l'aide de l'assistant graphique, réaliser l'histogramme correspondant (on donnera un titre au graphique, on vérifiera que les classes sont étiquetées par leurs centres, on effacera le fond gris du graphique). Faire plusieurs fois **F9** pour voir les fluctuations d'échantillonnage.
6. En **A120**, entrer la formule `=(NB.SI(A102 :AX102 ; ">=0,3")-NB.SI(A102 :AX102 ; ">=0,5"))*2` puis en **B120** taper % entre 0,3 et 0,5. Faire **F9** pour de nouvelles simulations.

2^e étape : fluctuation des sondages de taille $n=1000$.

Sur une nouvelle feuille de calcul (nommer la première $n=100$ et celle-ci $n=1000$), on refait l'étude précédente pour $n=1000$ cette fois.

1. Dans la cellule A1, taper la formule `=ENT(ALEA()+0,4)` et la recopier par glissement jusqu'en cellule **A1000** puis calculer en cellule **A1001** la somme des nombres de cette colonne et en cellule **A1002** la fréquence f d'opinions favorables à X.
2. Recopier les formules de la colonne **A** par glissement en colonnes **B** à **AX** afin de simuler 50 sondages de taille $n=1000$. En **A1003**, taper Moyenne : et en **B1003**, taper la formule `=SOMME(A1002 :AX1002)/50`.
3. Marquer les cellules **A105** à **C118** de la feuille $n=100$ et les recopier dans la plage **A1005** à **C1018** de la feuille $n=1000$. Enfin réaliser l'histogramme correspondant à l'aide de l'assistant graphique. Faire plusieurs fois **F9** pour voir les fluctuations d'échantillonnage.
4. En **A1020**, entrer la formule `=(NB.SI(A1002 :AX1002 ; ">=0,37")-NB.SI(A1002 :AX1002 ; ">=0,43"))*2` puis en **B1020** taper % entre 0,37 et 0,43. Faire **F9** pour de nouvelles simulations.

3° étape : Seuils de confiance

On dit qu'un sondage est faux (ou qu'il se trompe) s'il donne un intervalle de confiance (ou fourchette) ne contenant pas la proportion p à estimer. Pour voir dans quelle mesure un sondage peut se tromper, on va visualiser les fourchettes fournies par 10 sondages et calculer la proportion de ceux qui se trompent selon le seuil de confiance choisi.

1. Sur une nouvelle feuille qu'on nommera, entrer en **A1** le nombre 0,48 (valeur de p pour une élection serrée) et en **A2** taper 1,645 (valeur permettant de calculer la fourchette au seuil de 90%). Dans la cellule **B1**, entrer la formule $=ENT(ALEA()+\$A\$1)$ et la recopier jusqu'en **B100**. En **A102** taper inf, en **A103** taper sup et en **A104** taper p^{\wedge} . Calculer en **B102** la borne inférieure de la fourchette en tapant $=B104-\$A\$2*RACINE(B104*(1-B104)/100)$. Calculer en cellule B103 la borne supérieure et en cellule **B104**, $=SOMME(B1 :B100)/100$ pour estimer la proportion cherchée. Enfin en **B105**, entrer la formule $=ET(\$A\$1>=B102 ;\$A\$1<=B103)$.
2. Recopier les cellules **B1** à **B105** sur les colonnes **C** à **K** afin de simuler 10 sondages de chacun 100 personnes. A l'aide de l'assistant graphique, sélectionner le type de graphique **boursier**, la plage de données **B102 :K104** puis cocher la série **en ligne**. Effacer la légende, ajouter un titre au graphique et ajuster l'axe des ordonnées. Appuyer sur la touche **F9** pour simuler 10 nouveaux sondages et étudier comment évoluent les fourchettes.
3. En **A106**, taper **FAUX** et en **B106** entrer la formule $=NB.SI(B105 :K105 ;FAUX)$ et en **C106** taper **sur 10**. Ceci permet de calculer le nombre de sondages « faux » au seuil de confiance de 90%.
4. Remplacer dans la cellule **A2** la valeur 1,645 par 1,96 et observer, en reproduisant les simulations au moyen de la touche **F9**, le nombre de sondages faux au seuil de 95%. Remplacer dans la cellule **A2** la valeur 1,96 par 2,58 et observer, en reproduisant les simulations au moyen de la touche **F9**, le nombre de sondages faux au seuil de 99%.