Exercice 1

Un industriel s'interroge sur la proportion d'acheteurs parmi sa clientèle qui ont acheté ou ont l'intention d'acheter le produit A, proportion notée p^A*. En particulier, il souhaiterait construire un intervalle de confiance de cette proportion d'acheteurs* p^A *obtenu à partir d'un échantillon de taille* $n = 500$ *individus issus de la population de taille* $N = 2000000$.

1. *Proposez l'instruction* `R` *ayant permis d'obtenir le résultat ci-dessous correspondant à un intervalle de confiance au niveau de confiance de 90% de* p^A *calculé à partir du jeu de données* $\boldsymbol{y}$ *que l'on note* `y` *en* `R` *(cet intervalle est noté* $[\widetilde{p^A}_{inf}(\boldsymbol{y}), \widetilde{p^A}_{sup}(\boldsymbol{y})]$*):*

   Indication(s) R :

```
> # IC <- (instruction R à fournir dans la rédaction)
> IC
[1] 0.1630267 0.2209733
```

2. *Le produit A est maintenant lancé sur le marché, et il a été alors possible d'évaluer le vrai paramètre* p^A *à* 18.9%*. Pour essayer de faire comprendre à l'un de ses collègues comment il faut interpréter les intervalles de confiance (en particulier le précédent), le concurrent propose l'exercice pédagogique suivant. On construit une urne de taille* $N = 2000000$ *boules dont une proportion* $p^A = 18.9\%$ *sont numérotées 1 (les autres étant numérotées 0). On fait alors* 199 *tirages de 500 boules au hasard au sein de cette urne. Les jeux de données créés sont donc de la même nature que* $\boldsymbol{y}$*. Les* $m = 200$ *jeux de données sont notés* $\boldsymbol{y_{[1]}}$, $\boldsymbol{y_{[2]}}$, $\ldots$, $\boldsymbol{y_{[200]}}$ *(le premier* $\boldsymbol{y_{[1]}}$ *correspondant à* $\boldsymbol{y}$*). Pour chacun de ces jeux de données, on construit un intervalle de confiance au niveau de 90% du paramètre* p^A*. Voici dans l'ordre des tirages quelques uns de ces intervalles :*

```
pInf       pSup
   [1,] 0.1630267 0.2209733
   [2,] 0.1971384 0.2588616
   [3,] 0.2210000 0.2210000
...
[198,] 0.1649122 0.2230878
[199,] 0.1724662 0.2315338
[200,] 0.1573773 0.2146227
```

   Parmi les $m = 200$ *intervalles de confiance, 179 contiennent le vrai paramètre* p^A*, qu'en pensez-vous ? Si l'on construisait une infinité d'intervalles de confiance, combien contiendraient le vrai paramètre* p^A *?*

3. *Complétez sans justification les encadrés ci-dessous :*

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{y_{[1]}}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{y_{[1]}}\right)\right) = \boxed{}$$

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{y_{[2]}}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{y_{[2]}}\right)\right) = \boxed{}$$

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{Y}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{Y}\right)\right) \simeq \boxed{}$$

4. *Complétez sans justification les encadrés ci-dessous :*
   - *si le niveau de confiance avait été de 95% alors*

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{y_{[1]}}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{y_{[1]}}\right)\right) = \boxed{}$$

   - *si le niveau de confiance avait été de 80% alors*

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{y_{[2]}}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{y_{[2]}}\right)\right) = \boxed{}$$

5. *Complétez sans justification les encadrés ci-dessous :*

    - *si le niveau de confiance avait été de 95% alors*

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{y_{[2]}}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{y_{[2]}}\right)\right) = \boxed{}$$

    - *si le niveau de confiance avait été de 80% alors*

$$\mathbb{P}\left(\widetilde{p^A}_{inf}\left(\boldsymbol{y_{[1]}}\right) < p^A < \widetilde{p^A}_{sup}\left(\boldsymbol{y_{[1]}}\right)\right) = \boxed{}$$

Exercice 2 *Avant le premier tour des élections, nous sommes souvent assaillis par de nombreux sondages. Le 13 mars 2012, deux instituts de sondages (IFOP et SOFRES) publient leurs estimations sur les intentions de votes pour deux candidats C1 et C2 :*

- *Sondage IFOP ($n = 1638$):* $\widehat{p^{C1}}\left(\boldsymbol{y^I}\right) = 27\%$ *et* $\widehat{p^{C2}}\left(\boldsymbol{y^I}\right) = 28.5\%$
- *Sondage SOFRES ($n = 1000$):* $\widehat{p^{C1}}\left(\boldsymbol{y^S}\right) = 30\%$ *et* $\widehat{p^{C2}}\left(\boldsymbol{y^S}\right) = 28\%$

1. *A la lumière de ce cours, nous proposons les mêmes résultats présentés à partir des intervalles de confiance à 95% de niveau de confiance :*

    - *Sondage IFOP:* $IC^{C1}(\boldsymbol{y^I}) = [24.85\%, 29.15\%]$ *et* $IC^{C2}(\boldsymbol{y^I}) = [26.31\%, 30.69\%]$
    - *Sondage SOFRES:* $IC^{C1}(\boldsymbol{y^S}) = [27.16\%, 32.84\%]$ *et* $IC^{C2}(\boldsymbol{y^S}) = [25.22\%, 30.78\%]$

    Fournir au choix :

    - *la formule mathématique (générale) permettant d'obtenir l'intervalle de confiance d'une proportion p s'exprimant en fonction de l'estimation $\widehat{p}\left(\boldsymbol{y}\right)$ et de la taille d'échantillon n.*
    - *la vérification à la calculatrice de l'obtention de l'un des intervalles de confiance ci-dessus (détails des calculs à fournir).*
    - *la formule* `R` *d'obtention d'un intervalle de confiance en fonction de* `pEst` *et* `n` *désignant respectivement l'intention de vote pour un candidat et la taille d'échantillon.*

2. *Interpréter via l'approche expérimentale des probabilités les intervalles de confiance obtenus à la question précédente.*

3. *La plupart des commentateurs politiques ont semblé troublés par de tels résultats apparemment contradictoires. A partir de la connaissance acquise dans ce cours et en supposant (de manière un peu abusive) que tous les intervalles de confiances précédents contiennent le vrai paramètre inconnu, pensez-vous qu'on puisse savoir lequel des candidats est en tête au premier tour ? Justifiez très simplement votre réponse en envisageant deux cas de figures bien choisis.*