Corrigé de l'exercice 5 TD3

Soit X_i la variable aléatoire qui vaut 1 si la ième personne interrogée est décidée à modifier ses habitudes de transport avec la probabilité p. La variable X_i suit une loi de $Be(p)$.

D'après le théorème central limite appliqué à l'échantillon $(X_1, ..., X_n)$, la variable aléatoire $Z = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ converge vers une v.a. de loi normale centrée, réduite $N(0,1)$ dès que $n > 30$, $np > 5$ et $n(1-p) > 5$.

Posons $0.90 = P(-t < Z < t)$ où le quantile $t = 1.645$ est lu dans la table de la loi normale et, ainsi, l'intervalle de confiance pour p, est donné par

$$P\left(\bar{X}_n - 1.645\sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + 1.645\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$$

En substituant, de part et d'autre de l'inégalité, le paramètre p par son estimation $\bar{x}_n$ on obtient finalement une fourchette d'estimation asymptotique pour p, à partir de l'échantillon des observations $(x_1, x_2, ..., x_n)$:

$$\bar{x}_n - 1.645\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \leq p \leq \bar{x}_n + 1.645\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}.$$

Pour $n = 16$ et $n = 25$, la condition $n > 30$ n'est pas vérifiée, on ne peut donc pas construire l'intervalle de confiance par cette méthode.

Pour $n = 40$, $\bar{x}_n = 0,2$, on obtient donc
$0.096 < p < 0.3040$. Vérifions à postériori les conditions d'application du TCL: $n * 0.096 = 3.84 < 5$, on ne peut donc pas construire l'intervalle de confiance par cette méthode.

Pour $n = 100$, on obtient donc
$0.1342 < p < 0.2658$. Comme $n * 0.1342 > 5$ et $n * (1 - 0.2658) > 5$, l'intervalle de confiance est bien validé.

Pour $n = 1000$, on obtient donc
$0.1792 < p < 0.2208$. Comme $n * 0.1792 > 5$ et $n * (1 - 0.2208) > 5$, l'intervalle de confiance est bien validé.

Bonnes révisions!