DUT GEA 2^{e} année
Applications professionnelles statistiques, S4 2017-2018
TD GCF 24

Corrigé du Contrôle Continu n° 1

Exercice 1 Pour estimer dans une population la proportion p des individus se déclarant végétarien, on interroge au hasard 80 personnes de cette population. On observe que 18 personnes ont répondu qu'elles sont végétariennes.

1. Déterminons un intervalle centré en la proportion observée $\overline{f}$ contenant p avec probabilité 0,95. Puisque l'échantillon est de taille suffisante (ici $n = 80 \geq 30$), on sait que :

$$Z := \frac{\overline{f}_n - p}{\sqrt{\overline{f}_n(1-\overline{f}_n)/n}} \sim \mathcal{N}(0;1)$$

où $\overline{f}_n$ est la variable aléatoire (v.a.) représentant la proportion observée dans l'échantillon. On notera $\overline{f}$ la valeur prise par cette v.a. ; on a ici $\overline{f} = \frac{18}{80} = 0,225$.
On doit déterminer un intervalle de la forme $I_{0,05} = [\overline{f} - l_{0,05}; \overline{f} + l_{0,05}]$ contenant p avec probabilité $0,95$. Pour cela, on écrit que :

$$\begin{aligned}
&\mathbf{P}\left[p \in I_{0,05}\right] = 0,95 \\
&\iff \mathbf{P}\left[\overline{f}_n - l_{0,05} \leq p \leq \overline{f}_n + l_{0,05}\right] = 0,95 \\
&\iff \mathbf{P}\left[\overline{f}_n - l_{0,05} \geq p \text{ ou } p \geq \overline{f}_n + l_{0,05}\right] = 0,05 \\
&\iff \mathbf{P}\left[\frac{\overline{f}_n - p}{\sqrt{\overline{f}_n(1-\overline{f}_n)/n}} \geq \sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}l_{0,05} \text{ ou } \frac{\overline{f}_n - p}{\sqrt{\overline{f}_n(1-\overline{f}_n)/n}} \leq -\sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}l_{0,05}\right] = 0,05 \\
&\iff \mathbf{P}\left[Z \geq \sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}l_{0,05} \text{ ou } Z \leq -\sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}l_{0,05}\right] = 0,05.
\end{aligned}$$

Puisque Z suit la loi $\mathcal{N}(0;1)$, la table de la loi normale centrée réduite, dite « de l'écart-réduit », donne que

$$\mathbf{P}\left[Z \geq t_{0,05} \text{ ou } Z \leq -t_{0,05}\right] = 0,05$$

pour $t_{0,05} = 1,960$
Par identification, on obtient que :

$$\sqrt{\frac{n}{\overline{f}(1-\overline{f})}}l_{0,05} = t_{0,05} = 1,960,$$

soit encore

$$l_{0,05} = t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}} = 1,960\sqrt{\frac{0,225(1-0,225}{80}} \simeq 0,092.$$

Ainsi,

$$\begin{aligned}
I_{0,05} &= \left[\overline{f} - t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}}; \overline{f} + t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}}\right] \\
&\simeq [0,225 - 0,092; 0,225 - 0,092] = [0,133; 0,317].
\end{aligned}$$

2. On souhaite, qu'avec la même fréquence observée l'intervalle de confiance au seuil de 5%, soit de longueur au plus 0,05 c'est-à-dire que $2l_{0,05} \leq 0,05$. Pour cela, il faut et suffit que :

$$2l_{0,05} = 2t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}} \leq 0,05$$

soit

$$2 \times 1,960\sqrt{\frac{0,225(1-0,225)}{n}} \leq 0,05$$

ou encore

$$3,920\sqrt{\frac{0,225(1-0,225)}{n}} \leq 0,05$$

c'est-à-dire

$$3,920\frac{\sqrt{0,225(1-0,225)}}{0,05} \leq \sqrt{n}$$

c'est-à-dire

$$n \geq \left(3,920\frac{\sqrt{0,225(1-0,225)}}{0,05}\right)^2 \simeq 1071,8.$$

Il faut interroger au moins 1072 individus.

Exercice 2

Une expérience d'échantillonnage consiste à simuler le prélèvement de 100 échantillons de 400 graines de tournesol. La variable aléatoire "masse" d'une graine de tournesol est supposée normale de moyenne $51mg$ et d'écart-type $5mg$. On suppose les masses des graines 2 à 2 indépendantes.

1. Notons X_i, $i = 1, \ldots, 400$, la masse de la i^{e} graine de tournesol d'un échantillon de 400 graines. La variable aléatoire "masse moyenne d'une graine de tournesol d'un échantillon de 400 graines" s'écrit :

$$\overline{X} = \frac{X_1 + \cdots + X_{400}}{400} = \frac{X_1}{400} + \cdots + \frac{X_{400}}{400}.$$

Puisque l'on sait que, pour tout i, $X_i \sim \mathcal{N}(51;5)$ et les X_i sont deux à deux indépendantes, on en déduit que $\overline{X}$ suit une loi normale de moyenne :

$$\mathbf{E}[\overline{X}] = \mathbf{E}\left[\frac{X_1}{400} + \cdots + \frac{X_{400}}{400}\right] = \frac{\mathbf{E}[X_1]}{400} + \cdots + \frac{\mathbf{E}[X_{400}]}{400} = \frac{51}{400} + \cdots + \frac{51}{400} = 51$$

et de variance

$$\mathbf{V}[\overline{X}] = \mathbf{V}\left[\frac{X_1}{400} + \cdots + \frac{X_{400}}{400}\right] = \frac{\mathbf{V}[X_1]}{400^2} + \cdots + \frac{\mathbf{V}[X_{400}]}{400^2} = \frac{5^2}{400^2} + \cdots + \frac{5^2}{400^2} = \frac{5^2}{400}$$

c'est-à-dire d'écart-type

$$\sigma = \sqrt{\mathbf{V}[\overline{X}]} = \frac{5}{20} = \frac{1}{4}.$$

Pour résumer :

$$\overline{X} \sim \mathcal{N}\left(51; \frac{1}{4}\right).$$

2. On a donc que :

$$\frac{\overline{X} - 51}{\frac{1}{4}} \sim \mathcal{N}(0;1).$$

La probabilité pour que la masse moyenne dans un échantillon soit inférieure à $50,6mg$ est :

$$\begin{aligned}
\mathbf{P}\left[\overline{X} \leq 50,6\right] &= \mathbf{P}\left[\frac{\overline{X} - 51}{\frac{1}{4}} \leq \frac{50,6-51}{\frac{1}{4}}\right] \\
&= \mathbf{P}\left[Z \leq -1,6\right] \\
&= 1 - \mathbf{P}\left[Z \leq 1,6\right] \\
&\simeq 1 - 0,9452 = 0,0548.
\end{aligned}$$

Parmi les 100 échantillons, on peut s'attendre à ce que $100 \times 0,0548 \simeq 5$ présentent une masse moyenne inférieure à $50,6mg$.