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

Corrigé du Contrôle Continu no 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. Puisqu'une proportion est nécessairement positive, pour qu'avec la même fréquence observée l'intervalle de confiance au seuil de 5% soit inclus dans $[0;\ 0,25]$, il faut et suffit que :

$$\overline{f} + t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}} \leq 0,25$$

soit

$$0,225 + 1,960\sqrt{\frac{0,225(1-0,225)}{n}} \leq 0,25$$

ou encore

$$1,960\sqrt{\frac{0,225(1-0,225)}{n}} \leq 0,25 - 0,225 = 0,025$$

c'est-à-dire

$$1,960\frac{\sqrt{0,225(1-0,225)}}{0,025} \leq \sqrt{n}$$

c'est-à-dire

$$n \geq \left(1,960\frac{\sqrt{0,225(1-0,225)}}{0,025}\right)^2 \simeq 1071,8.$$

Il faut interroger au moins 1072 individus.

Exercice 2

Une expérience d'échantillonnage consiste à simuler le prélèvement de 200 échantillons de 900 grains de café. La variable aléatoire "masse" d'un grain de café est supposée normale de moyenne $114mg$ et d'écart-type $20mg$. On suppose les masses des graines 2 à 2 indépendantes.

1. Notons X_i, $i = 1, \ldots, 900$, la masse du i^{e} grain de café d'un échantillon de 900 grains. La variable aléatoire "masse moyenne d'un grain de café d'un échantillon de 900 grains" s'écrit :

$$\overline{X} = \frac{X_1 + \cdots + X_{900}}{900} = \frac{X_1}{900} + \cdots + \frac{X_{900}}{900}.$$

Puisque l'on sait que, pour tout i, $X_i \sim \mathcal{N}(114; 20)$ et les X_i sont deux à deux indépendantes, on en déduit que $\overline{X}$ suit une loi normale de moyenne :

$$\mathbf{E}\left[\overline{X}\right] = \mathbf{E}\left[\frac{X_1}{900} + \cdots + \frac{X_{900}}{900}\right] = \frac{\mathbf{E}[X_1]}{900} + \cdots + \frac{\mathbf{E}[X_{900}]}{900} = \frac{114}{900} + \cdots + \frac{114}{900} = 114$$

et de variance

$$\mathbf{V}[\overline{X}] = \mathbf{V}\left[\frac{X_1}{900} + \cdots + \frac{X_{900}}{900}\right] = \frac{\mathbf{V}[X_1]}{900^2} + \cdots + \frac{\mathbf{V}[X_{900}]}{900^2} = \frac{21^2}{900^2} + \cdots + \frac{21^2}{900^2} = \frac{21^2}{900}$$

c'est-à-dire d'écart-type

$$\sigma = \sqrt{\mathbf{V}[\overline{X}]} = \frac{21}{30} = 0,7.$$

Pour résumer :

$$\overline{X} \sim \mathcal{N}\left(114; 0,7\right).$$

2. On a donc que :

$$\frac{\overline{X} - 114}{0,7} \sim \mathcal{N}\left(0; 1\right).$$

La probabilité pour que la masse moyenne dans un échantillon soit inférieure à $100mg$ est :

$$\begin{aligned}\mathbf{P}\left[\overline{X} \leq 100\right] &= \mathbf{P}\left[\frac{\overline{X} - 114}{0,7} \leq \frac{100 - 114}{0,7}\right] \\ &= \mathbf{P}\left[Z \leq -20\right] \\ &= 1 - \mathbf{P}\left[Z \leq 20\right].\end{aligned}$$

Puisque $20 \gg 4,5$, cette probabilité est inférieure à 0,000003. Parmi les 200 échantillons, on peut s'attendre à ce qu'aucun ne présente une masse moyenne inférieure à $100mg$.