Examen (Thème Statistiques)

Exercice 1 (Nul n'est censé ignorer la loi normale) Dans une université, une promotion de L1 ne doit pas dépasser 200 étudiants. En se basant sur le constat que seulement un candidat accepté sur trois viendra effectivement à la rentrée, la politique de l'université est d'accepter systématiquement 500 étudiants. Les décisions des étudiants sont indépendantes.

1. Sur 500 candidats acceptés, quelle est la loi de la variable X correspondant au nombre d'étudiants effectivement présents à la rentrée ?
2. En utilisant l'approximation par la loi normale, et la table de sa fonction de répartition, estimer la probabilité qu'il y ait plus de 200 étudiants présents à la rentrée.

1. Chaque étudiant accepté a une probabilité $\frac{1}{3}$ d'être effectivement présent à la rentrée. Les décisions des étudiants étant supposées indépendantes les unes des autres, la loi de X est binomiale : $X \sim \mathcal{B}(500, 1/3)$.
2. Nous sommes dans le cadre d'application du théorème central limite : puisque $\mathbb{E}[X] = 500/3$ et $\mathbb{V}[X] = 1000/9$, nous faisons l'approximation : $X \approx \mathcal{N}(500/3, 1000, 9)$, donc

$$\mathbb{P}(X > 200) = \mathbb{P}\left(\frac{X - 500/3}{\sqrt{1000/9}} > \frac{200 - 500/3}{\sqrt{1000/9}}\right) \approx 1 - \Phi(3.16) \approx 8.10^{-4} \, .$$

   Si le modèle est bon, il y a donc très peu de risques d'être en sureffectif à la rentrée.

Exercice 2 (Les raisins de la collecte) On a mesuré le poids de raisin produit par souche sur 10 souches prises au hasard dans la vigne. On a obtenu les résultats suivants (en kg) :

$$2.4 \quad 3.4 \quad 3.6 \quad 4.1 \quad 4.3 \quad 4.7 \quad 5.4 \quad 5.9 \quad 6.5 \quad 6.9$$

On modélise le poids de raisin produit par une souche de cette vigne par une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$.

1. Donner la formule et le calcul des estimation ponctuelles non biaisées de la moyenne et de la variance de la population dont sont extraites les souches.
2. Donner un intervalle de confiance de niveau 0.95 pour μ. *(Utilisez la table de Student)*
3. Donner un intervalle de confiance de niveau 0.95 pour σ^2. *(Utilisez la table du khi deux)*
4. On suppose désormais que l'écart-type des productions par souche est connu $\sigma = 1.4$. Donner un intervalle de confiance de niveau 0.95 pour μ. *(Utilisez la table de loi normale)*
5. Quel nombre de souches au minimum devrait-on observer pour estimer μ au niveau de confiance 0.99 avec une précision de plus ou moins 500 grammes ?

1. L'échantillon est de taille $n = 10$. Il a pour moyenne empirique $\bar{x}_n = 4.72$ kg et de variance empirique corrigée $s_n'^2 \approx 2.08$ kg.
2. La population étant supposée suivre une loi normale $\mathcal{N}(\mu, \sigma^2)$, l'espérance μ a pour intervalle de confiance au risque α :

$$\left[\bar{x}_n - t_{n-1,\alpha}\frac{s'_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,\alpha}\frac{s'_n}{\sqrt{n}}\right] \ ,$$

où $t_{n-1,\alpha}$ est la valeur telle que $\mathbb{P}(|X| > t_{n-1,\alpha}) = \alpha$ avec $X \sim \mathrm{St}(n-1)$ une loi de Student à $n-1=9$ degrés de liberté. Comme $\alpha = 0.05$ on lit dans la table correspondante que $t_{9,0.05} = 2.262$, ce qui correspond en R à $\mathrm{qt}(1-0.025, 9)$. On obtient ainsi un intervalle de confiance pour μ suivant :

$$[3.69, 5.75] \ ,$$

que l'on retrouve en R avec t.test(x) où x est le vecteur des 10 échantillons. L'intervalle est assez large dans la mesure où l'information expérimentale disponible est faible puisqu'elle ne porte que sur 10 mesures.

3. L'intervalle de confiance au risque α pour la variance σ^2 est :

$$\left[\frac{(n-1)s_n'^2}{z_{n-1,\alpha/2}}, \frac{(n-1)s_n'^2}{z_{n-1,1-\alpha/2}}\right] \ ,$$

où $z_{n,\alpha}$ est la valeur telle que $\mathbb{P}(X > z_{n-1,\alpha}) = \alpha$ avec $X \sim \chi^2_{n-1}$ une loi du χ^2 à $n-1=9$ degrés de liberté. Comme $\alpha = 0.05$ on lit dans la table correspondante que $z_{9,0.025} = 19.02$ et $z_{9,0.975} = 2.70$, ce qui correspond respectivement en R à $\mathrm{qchisq}(1-0.025, 9)$ et $\mathrm{qchisq}(0.025, 9)$. On obtient ainsi un intervalle de confiance pour μ suivant :

$$[0.98, 6.93] \ .$$

4. Lorsque σ est connu l'intervalle de confiance au risque α pour μ est :

$$\left[\bar{x}_n - \frac{\sigma}{\sqrt{n}}u_\alpha, \bar{x}_n + \frac{\sigma}{\sqrt{n}}u_\alpha\right] \ ,$$

où u_α est la valeur telle que $\mathbb{P}(|X| > u_\alpha) = \alpha$ avec $X \sim \mathcal{N}(0,1)$ une loi normale centrée réduite. Comme $\alpha = 0.05$ on lit dans la table correspondante que $u_{0.05} = 1.96$, ce qui correspond en R à $\mathrm{qnorm}(1-0.025)$. On obtient ainsi, avec $\sigma = 1.4$ et $\sqrt{n} = \sqrt{10}$, un intervalle de confiance pour μ suivant :

$$[3.85, 5.59] \ .$$

5. Pour un risque $\alpha = 0.01$, on a $u_{0.01} = 2.57$ et pour atteindre une précision de plus ou moins 500 g dans l'estimation de μ il faut que la demi-largeur de l'intervalle de confiance soit telle que

$$\frac{\sigma}{\sqrt{n}}u_\alpha \leq 0.5 \Leftrightarrow n \geq (2\sigma u_\alpha)^2 \approx 52.01 \ ,$$

donc requiert au minimum un échantillon de 53 souches.