Ex - Estimation par intervalle de confiance

On s'intéresse au taux de glucose dans une population de 768 patients atteints de diabète. On note m le taux moyen de glucose dans cette population et à σ son écart type.

1. Donner une estimation de m à l'aide de l'échantillon fourni en suivant ce lien, en précisant la taille de l'échantillon donné et l'estimateur choisi. Avec l'échantillon de taille 40 dont la réalisation est fournie, on obtient avec l'estimateur de moyenne empirique (non biaisé) l'estimation $\overline{x} = 124{,}35$.
2. Donner une estimation de m par intervalle de confiance au niveau 95% et 99%, en donnant les valeurs numériques des calculs intermédiaires.

✓

On utilise la formule du cours :

$$I_{conf}(\overline{X}) = \left[\overline{X} - u_{\alpha/2}\frac{S}{\sqrt{n}}\ ;\ \overline{X} + u_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

avec $\overline{x} = 124{,}35$, $s = 30{,}0994505$ et $n = 40$.
Pour $\alpha = 0.05$, on obtient : $I_{conf} = [115{,}022255; 133{,}677745]$
Pour $\alpha = 0.01$, on obtient : $I_{conf} = [112{,}0912651; 136{,}6087349]$

3. Le fichier "diabetes.csv" contient les données des 768 patients. Convertir les données pour pouvoir les afficher avec le tableur et donner la valeur réelle de m. Quel niveau de confiance avait-on besoin de prendre, a posteriori, pour que l'intervalle de confiance de la question précédente contienne bien la valeur recherchée ?

✓

Dans la population totale, on trouve $m = 120{,}8945313$. Pour que I_conf ne contienne pas cette valeur, il faudrait que $\overline{X} - u_{\alpha/2}\frac{S}{\sqrt{n}} > m$ soit $(\overline{X} - m)\frac{\sqrt{n}}{S} > u_{\alpha/2}$ soit $u_{\alpha/2} < 0.726$ soit $\alpha \geqslant 0,467795957$.