IUT Dijon-Auxerre
GEA 2ème année, 2015-2016

Corrigé du devoir à la Maison de Mathématiques

Problème : Une rumeur affirme que le distributeur de café situé dans la gare de la ville de Triches (au sud de Belluno en Italie), délivre des cafés dont le volume est inférieur à celui annoncé, c'est-à-dire 10cl. On se propose de réaliser une étude statistique pour infirmer ou confirmer cette rumeur.

Partie A :

On s'intéresse dans un premier temps au volume moyen m de café versé dans un gobelet. Pour cela, on a prélevé un échantillon de 50 cafés issus de ce distributeur et mesuré la quantité de café qu'ils contenaient. On a observé une moyenne de 9,7cl et un écart-type de 1cl.

1. Déterminons un intervalle centré en la moyenne observée contenant m avec probabilité 0,95 puis 0,99. Il s'agit de déterminer un intervalle de confiance pour la moyenne, aux seuils de 5% puis 1%, dans le cas où le « vrai » écart-type est inconnu et l'échantillon est de grande taille (ici $n = 50 \geq 30$).

   Pour cela, on utilise que :
$$Z := \frac{\overline{X}_n - m}{s/\sqrt{n}} \sim \mathcal{N}(0;1),$$
   où m et s sont respectivement la « vraie » moyenne et le « vrai » écart-type dans la population, $n = 50$ la taille de l'échantillon aléatoire observé et $\overline{X}_n$ la moyenne observée dans cet échantillon (vue comme une variable aléatoire). On notera $\overline{X}$ la valeur observée de cette moyenne.

   Pour $\alpha = 0,05$ puis $\alpha = 0,01$, on cherche un intervalle de la forme $I_{1-\alpha} = [\overline{X} - l_\alpha; \overline{X} + l_\alpha]$ tel que :
$$\mathbf{P}\left[m \in I_\alpha\right] = 1 - \alpha.$$
   On a :
$$\begin{aligned}
\mathbf{P}\left[m \in I_{1-\alpha}\right] = 1 - \alpha &\iff \mathbf{P}\left[\overline{X}_n - l_\alpha \leq m \leq \overline{X}_n + l_\alpha\right] = 1 - \alpha \\
&\iff \mathbf{P}\left[\overline{X}_n - l_\alpha \geq m \text{ ou } m \geq \overline{X}_n + l_\alpha\right] = \alpha \\
&\iff \mathbf{P}\left[\overline{X}_n - m \geq l_\alpha \text{ ou } \overline{X}_n - m \leq -l_\alpha\right] = \alpha \\
&\iff \mathbf{P}\left[\frac{\overline{X}_n - m}{s/\sqrt{n}} \geq \frac{\sqrt{n}}{s} l_\alpha \text{ ou } \frac{\overline{X}_n - m}{s/\sqrt{n}} \leq -\frac{\sqrt{n}}{s} l_\alpha\right] = \alpha \\
&\iff \mathbf{P}\left[Z \geq \frac{\sqrt{n}}{s} l_\alpha \text{ ou } Z \leq -\frac{\sqrt{n}}{s} l_\alpha\right] = \alpha.
\end{aligned}$$
   Puisque Z suit la loi $\mathcal{N}(0;1)$, la table de la loi normale centrée réduite dite « de l'écart-réduit », donne la valeur de t_α pour laquelle on a :
$$\mathbf{P}\left[Z \geq t_\alpha \text{ ou } Z \leq -t_\alpha\right] = \alpha.$$
   Par identification, on obtient que :
$$\frac{\sqrt{n}}{s} l_\alpha = t_\alpha,$$
   soit encore
$$l_\alpha = t_\alpha \frac{s}{\sqrt{n}}.$$

Ne connaissant pas la valeur du « vrai » écart-type s, on l'approxime en utilisant l'estimateur sans biais de l'écart-type :

$$\sqrt{\frac{n}{n-1}}\sigma_n,$$

où σ_n est l'écart-type observé dans l'échantillon (vue comme une variable aléatoire). En notant σ, la valeur de l'écart-type observée dans l'échantillon, on obtient que :

$$l_\alpha = t_\alpha \frac{s}{\sqrt{n}} \simeq t_\alpha \frac{\sigma}{\sqrt{n-1}}$$

puis que

$$I_{1-\alpha} = [\overline{X} - t_\alpha \frac{\sigma}{\sqrt{n-1}}; \overline{X} + t_\alpha \frac{\sigma}{\sqrt{n-1}}].$$

<u>Applications numériques :</u> On a $n = 50$, $\overline{X} = 9,7$, $\sigma = 1$, $t_{0,05} = 1,96$, $t_{0,01} = 2,576$. On obtient donc :

$$I_{0,95} = \left[\overline{X} - t_{0,05}\frac{\sigma}{\sqrt{n-1}}; \overline{X} + t_{0,05}\frac{\sigma}{\sqrt{n-1}}\right] \simeq [9,42; 9,98]$$

et

$$I_{0,99} = \left[\overline{X} - t_{0,01}\frac{\sigma}{\sqrt{n-1}}; \overline{X} + t_{0,01}\frac{\sigma}{\sqrt{n-1}}\right] \simeq [9,332; 10,068]\,.$$

2. L'intervalle de confiance pour la moyenne au seuil de 5% ne contient pas la valeur annoncée (10cl) alors que celui au seuil de 1% la contient. On peut donc affirmer au risque de 5% que la quantité de café moyenne distribuée est inférieure à celle annoncée mais on ne peut pas faire la même affirmation au risque de 1%. Autrement dit, il y a au moins de 95% de chances pour que la quantité de café moyenne distribuée soit inférieure à celle annoncée mais il y a au moins 1% de chances pour que ce ne soit pas le cas. Pour pouvoir affirmer au risque de 1% que la quantité de café moyenne distribuée est inférieure à celle annoncée, il faudrait disposer d'un échantillon de plus grande taille.
3. Pour confirmer la rumeur au risque de 1%, il faut que $IC_{0,99}$ soit contenu dans l'intervalle $[0; 10[$. En supposant que l'on observe les mêmes moyenne et écart-type, il faut et suffit pour cela que la taille n de l'échantillon satisfasse :

$$\overline{X} + t_{0,01}\frac{\sigma}{\sqrt{n-1}} < 10$$

c-à-d

$$9,7 + \frac{2,576}{\sqrt{n-1}} < 10$$

c-à-d

$$\frac{2,576}{\sqrt{n-1}} < 0,3$$

c-à-d

$$\frac{2,576}{0,3} < \sqrt{n-1}$$

c-à-d

$$n > \left(\frac{2,576}{0,3}\right)^2 + 1 \simeq 74,73.$$

Ainsi, il faudrait un échantillon de taille au moins 75.

<u>Partie B :</u> On s'intéresse, cette fois, à la proportion p de gobelets contenant un volume de café inférieur à celui annoncé. On a prélevé 200 gobelets et observé que 116 d'entre eux contenaient strictement moins de 10cl. On cherche à savoir si plus de la moitié des cafés distribués ont un volume inférieure à celui annoncé.

1. Déterminons un intervalle centré en la proportion observée contenant p avec probabilité 0,9 puis 0,99. Puisque l'échantillon est de taille suffisante (ici $n = 200 \geq 30$), on sait que :

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

où $\overline{p}_n$ est la variable aléatoire (v.a.) représentant la proportion observée dans l'échantillon. On notera $\overline{p}$ la valeur prise par cette v.a. ; on a ici $\overline{p} = \frac{116}{200} = 0,58$.

On doit déterminer un intervalle de la forme $I_{1-\alpha} = [\overline{p} - l_\alpha ; \overline{p} + l_\alpha]$ contenant p avec probabilité $1-\alpha$, pour $\alpha = 0,1$ puis $\alpha = 0,01$. Pour cela, on écrit que :

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

Puisque Z suit la loi $\mathcal{N}(0;1)$, la table de la loi normale centrée réduite, dite « de l'écart-réduit », donne la valeur de t_α pour laquelle on a :

$$\mathbf{P}\left[Z \geq t_\alpha \text{ ou } Z \leq -t_\alpha\right] = \alpha.$$

Par identification, on obtient que :

$$\sqrt{\frac{n}{\overline{p}_n(1-\overline{p}_n)}}\, l_\alpha = t_\alpha,$$

soit encore

$$l_\alpha = t_\alpha \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}}.$$

Ainsi,

$$I_{1-\alpha} = [\overline{p} - t_\alpha \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}} ; \overline{p} + t_\alpha \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}}].$$

Applications numériques : On a $n = 200$, $\overline{p} = \frac{116}{200} = 0,58$, $t_{0,10} = 1,645$, $t_{0,01} = 2,576$. On obtient donc :

$$I_{0,90} = [\overline{p} - t_{0,10} \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}} ; \overline{p} + t_{0,10} \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}}] \simeq [0,52 ; 0,64]$$

et

$$I_{0,99} = [\overline{p} - t_{0,01} \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}} ; \overline{p} + t_{0,01} \sqrt{\frac{\overline{p}_n(1-\overline{p}_n)}{n}}] \simeq [0,49 ; 0,67].$$

2. On peut conclure que la « vraie » proportion p de cafés dont la contenance est inférieure à celle annoncée est contenue dans l'intervalle :
   — $[0,52; 0,64]$ avec probabilité 0,9 ;
   — $[0,49; 0,67]$ avec probabilité 0,99.
   Ainsi, on peut affirmer au risque de 10% que plus de la moitié des cafés ont un volume inférieur à celui annoncé mais on ne peut pas faire la même affirmation au risque de 1%.

Partie C : On dispose cette fois d'un échantillon de seulement cinq cafés. Pour $i = 1, \ldots, 5$, on note Y_i la variable aléatoire valant 1 si le $i$ème café a un volume strictement inférieur à 10cl et 0 sinon. On suppose que les volumes de cafés distribués durant différentes utilisations de la machine sont indépendants. On suppose également qu'en observant la totalité des cafés distribués par la machine exactement la moitié d'entre eux ont un volume strictement inférieur à 10cl.

1. Pour $i = 1, \ldots, 5$, $Y_i \sim \mathcal{B}er(0,5)$. Les Y_i étant supposées indépendantes, $N \sim \mathcal{B}in(5; 0,5)$ de cafés d'un échantillon de 5 cafés dont le volume est inférieur à 10cl.

2. La proportion de cafés ayant un volume strictement inférieur à 10cl dans un échantillon de cinq cafés est simplement $\frac{N}{5}$. Celle-ci dépasse 58% si et seulement si $N \geq 5 \times 0,58 = 2,9$. On a vu que $N \sim \mathcal{B}in(5; 0,5)$.

Ainsi, la probabilité pour que, dans un échantillon de 5 cafés, au moins 58% des cafés aient un volume strictement inférieur à 10cl est :

$$\begin{aligned}
\mathbf{P}\left[N \geq 2,9\right] &= \mathbf{P}\left[N = 3\right] + \mathbf{P}\left[N = 4\right] + \mathbf{P}\left[N = 5\right] \\
&= \mathbf{C}_5^3 \times 0,5^3 \times 0,5^2 + \mathbf{C}_5^4 \times 0,5^4 \times 0,5^1 + \mathbf{C}_5^5 \times 0,5^5 \times 0,5^0 \\
&= 0,5^5 \times \left(\frac{5!}{3!2!} + \frac{5!}{4!1!} + \frac{5!}{5!0!}\right) \\
&= \frac{1}{32}\left(10 + 5 + 1\right) = \frac{1}{2}.
\end{aligned}$$

Un échatillon de si petite taille ne permet pas de détecter un comportement anormal de la machine.