DUT TC 2 **ULCO - Dunkerque**

Probabilités et Statistiques Inférentielles Novembre 2016 - IE 2, Semestre 1

Durée de l'épreuve : 1h30 Tous documents autorisés

Exercice 1 *Correction* : $\boxed{\text{6pts}}$

1. $\boxed{\text{1,5pt}}$ On suppose que les personnes ont bien été interrogées indépendamment. Ainsi, on a un schéma de Bernoulli : une personne interrogée va au cinéma chaque mois -> SUCCÈS, sinon ÉCHEC. Et donc X suit une loi binomiale $\mathcal{B}(100, p)$. La distribution de probabilité est donnée par :

$$P(X = k) = \binom{100}{k} p^k (1-p)^{100-k},\ k = 0, \dots, 100.$$

$\boxed{\text{0,5pt}}$ Comme $n \geq 20$, si $np > 5$ et $n(1-p) > 5$ (conditions à vérifier lors de l'application numérique), on peut approcher cette loi par la loi normale

$$\mathcal{N}(np, \sqrt{np(1-p)}) = \mathcal{N}(100p, \sqrt{100p(1-p)}).$$

$\boxed{\text{0,5pt}}$ Donc F suit approximativement la loi

$$\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{100}}\right).$$

2. $\boxed{\text{0,5pt}}$ D'après le cours,

$$IC = \left[f - Z_\alpha\sqrt{\frac{p(1-p)}{100}}, f + Z_\alpha\sqrt{\frac{p(1-p)}{100}}\right].$$

3. (a) $\boxed{\text{1pt}}$ Pour $f = 0,1$ et $1 - \alpha = 90\% \Leftrightarrow Z_\alpha = 1,645$, $IC = [0,05; 0,15]$.

   (b) $\boxed{\text{1pt}}$ Pour $f = 0,1$ et $1 - \alpha = 95\% \Leftrightarrow Z_\alpha = 1,96$, $IC = [0,04; 0,16]$.

   (c) $\boxed{\text{1pt}}$ Pour $f = 0,1$ et $1 - \alpha = 98\% \Leftrightarrow Z_\alpha = 2,326$, $IC = [0,03; 0,17]$.

Exercice 2 *Correction* : $\boxed{\text{5pts}}$

1. On travaille sur un échantillon de taille 200.

   (a) $\boxed{\text{1pt}}$ On s'intéresse à la statistique $\overline{X}$, moyenne des poids maximaux sur tout échantillon de taille 200. Si $X \rightsquigarrow \mathcal{N}(58; 3)$, $\overline{X} \rightsquigarrow \mathcal{N}\left(58; \frac{3}{\sqrt{200}}\right)$.

   $\boxed{\text{0,5pt}}$ On cherche par conséquent, en posant $T = \frac{\overline{X} - 58}{\frac{3}{\sqrt{200}}}$, la valeur a telle que

$$P(-a \leq T \leq a) = 0,99 \Leftrightarrow 2\Pi(a) - 1 = 0,99 \Leftrightarrow \Pi(a) = 0,995 \Leftrightarrow a = 2,575.$$

   $\boxed{\text{1pt}}$ Donc

$$P\left(-2,575 \leq \frac{\overline{X} - 58}{\frac{3}{\sqrt{200}}} \leq 2,575\right) = 0,01 \Leftrightarrow P(57,45 \leq \overline{X} \leq 58,55) = 0,99.$$

   L'intervalle recherché est donc $IC = [57,45; 58,55]$.

(b) $\boxed{\text{0,5pt}}$ Le poids moyen constaté sur l'échantillon précédent est conforme aux attentes car $57,7 \in IC$.

2. $\boxed{\text{2pts}}$ On cherche b tel que

$$P(\overline{X} > b) = 0,97 \Leftrightarrow P\left(\frac{\overline{X} - 58}{\frac{3}{\sqrt{200}}} > \frac{b - 58}{\frac{3}{\sqrt{200}}}\right) = 0,97 \Leftrightarrow P(T > \frac{b - 58}{\frac{3}{\sqrt{200}}}) = 0,97$$

$$\Leftrightarrow 1 - P\left(T \leq \frac{b - 58}{\frac{3}{\sqrt{200}}}\right) = 0,97 \Leftrightarrow P\left(T \leq \frac{b - 58}{\frac{3}{\sqrt{200}}}\right) = 0,03 \Leftrightarrow \frac{b - 58}{\frac{3}{\sqrt{200}}} = -1,88 \Leftrightarrow b = 57,6.$$

Exercice 3 *Correction* : $\boxed{\text{9pts}}$

1. (a) $\boxed{\text{1pt}}$ La moyenne et l'écart-type se calculent aisément à l'aide du tableau ci-dessous :

x_i	n_i	$n_i x_i$	$n_i x_i^2$
250	3	750	187500
251	2	502	126002
252	2	504	127008
253	2	506	128018
254	2	508	129032
255	3	765	195075
256	1	256	65536
257	1	257	66049
261	1	261	68121
TOTAL	17	4309	1092341

Ainsi, $\overline{x}_e = \dfrac{4309}{17} \simeq 253,47$ et $\sigma_e^2 = \dfrac{1092341}{17} - (253,47)^2 = 8,31$.

(b) $\boxed{\text{0,5pt}}$ Une estimation ponctuelle $\hat{m}$ de la moyenne m de la production est donnée par $\overline{x}_e = 253,47$.

(c) $\boxed{\text{1pt}}$ Une estimation ponctuelle $\hat{\sigma}^2$ de la variance σ^2 de la production est donnée par $\hat{\sigma}^2 = \dfrac{n}{n-1}\sigma_e^2 = 8,83$.

2. $\boxed{\text{1pt}}$ La population est normale, sa variance est inconnue, on utilise une loi de Student-Fisher. Soit T la variable aléatoire définie par $T = \dfrac{\overline{X} - m}{\frac{\hat{\sigma}}{\sqrt{n}}}$ (où $\overline{X}$ est la variable aléatoire moyenne de l'échantillon aléatoire de taille n), qui est distribuée selon la loi de Student-Fisher t_ν à $\nu = n - 1 = 16$ degrés de liberté.
$\boxed{\text{0,5pt}}$ Grâce à la table de la loi de Student-Fisher, on récupère pour $\nu = 16$, la valeur $t_{\nu,\alpha}$ vérifiant $P(-t_{\nu,\alpha} < T < t_{\nu,\alpha}) = 0,95$ soit $t_{\nu,\alpha} = 2,12$.
$\boxed{\text{1,5pt}}$ On obtient ainsi l'intervalle de confiance aléatoire au niveau 95% demandé :

$$\left[\overline{X} - 2,12 \times \frac{\hat{\sigma}}{\sqrt{17}}; \overline{X} + 2,12 \times \frac{\hat{\sigma}}{\sqrt{17}}\right].$$

On a vu que $\overline{x} = 253,47$g et $\hat{\sigma} = \sqrt{8,83} = 2,97$g d'où une estimation de m par intervalle de confiance au niveau 95% à savoir $IC = [251,94; 254,99]$.
La masse moyenne de la production a 95 chances sur 100 de se trouver dans cet intervalle.

3. (a) $\boxed{\text{1pt}}$ Si X est de loi normale , la variable T définie par $T = \dfrac{\overline{X} - m}{\frac{\hat{\sigma}}{\sqrt{n}}}$ où $\overline{X}$ est la moyenne de l'échantillon aléatoire simple de taille n de la variable aléatoire X, est distribuée selon la loi de Student à $n-1$ ddl (cf. question précédente). Ce qui fait l'intérêt de la variable de Student T, c'est qu'elle ne dépend pas de l'écart-type de la variable mère X. On utilise ce résultat à chaque fois que σ est inconnu. Si X est distribuée selon un loi normale, de moyenne m et d'écart-type σ inconnus, l'intervalle de confiance aléatoire de la moyenne m au niveau 95% est :

$$\left[\overline{X} - t_{0,975;n-1} \times \frac{\hat{\sigma}}{\sqrt{n}}; \overline{X} + t_{0,975;n-1} \times \frac{\hat{\sigma}}{\sqrt{n}}\right].$$

Cet intervalle a pour amplitude $2 \times t_{0,975;n-1} \times \dfrac{\hat{\sigma}}{\sqrt{n}}$. Le problème posé par le responsable de fabrication correspond à la résolution de l'inéquation :

$$2 \times t_{0,975;n-1} \times \frac{\hat{\sigma}}{\sqrt{n}} \leq 1.$$

Or $\hat{\sigma}$, n et $t_{0,975;n-1}$ dépendent de n. En l'absence de données supplémentaires, on ne peut pas résoudre cette inéquation.

(b) $\boxed{\text{2,5pts}}$ X est de loi normale et on connaît maintenant sa variance σ. Sous ces conditions, $\overline{X}$ est distribuée selon $\mathcal{N}\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$ et $T = \dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$ selon la loi normale $\mathcal{N}(0,1)$. L'intervalle de confiance aléatoire de la moyenne m au niveau 95% est :

$$\left[\overline{X} - Z_{0,975} \times \frac{\sigma}{\sqrt{n}}; \overline{X} + Z_{0,975} \times \frac{\sigma}{\sqrt{n}}\right].$$

Cet intervalle a pour amplitude $2 \times Z_{0,975} \times \dfrac{\sigma}{\sqrt{n}}$. Pour obtenir un intervalle de confiance d'amplitude inférieure ou égale à 1, il suffit de choisir n tel que

$$2 \times Z_{0,975} \times \frac{\sigma}{\sqrt{n}} \leq 1 \Leftrightarrow n \geq (2 \times Z_{0,975})^2 \sigma^2.$$

Application numérique :

$$n \geq (2 \times 1,96)^2 \times 6,25 \Leftrightarrow n \geq 96,04.$$

Un échantillon doit compter au minimum 97 camemberts pour que l'intervalle de confiance à 95% correspondant ait une amplitude inférieure à 1.

4. HORS-PROGRAMME.
On considère le test unilatéral droit :

$$H_0 : \text{“}p = 0,15\text{”} \text{ contre } H_1 : \text{“}p > 0,15\text{”} .$$

- Seuil de signification : $\alpha = 5\%$.
- Conditions d'application : grand échantillon, de taille $n = 200$, $np_0 \geq 5$ et $n(1-p_0) \geq 5$.
- Variable de décision : on utilise pour ce test la statistique F qui prend pour valeur la proportion des camemberts de l'échantillon pesant plus de 257g.

Sous l'hypothèse H_0, $F = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ est distribuée selon la loi binomiale $\mathcal{B}(n, p_0)$ avec $p_0 = 0,15$. On a $E(F) = p_0$ et $V(F) = \sqrt{\dfrac{p_0(1-p_0)}{n}}$. Les conditions de normalité étant vérifiées, sous l'hypothèse H_0, la variable F est approximativement distribuée selon la loi normale $\mathcal{N}\left(p_0, \sqrt{\dfrac{p_0(1-p_0)}{n}}\right) = \mathcal{N}(0,15; 0,025)$.

On prend pour variable de décision Z définie par $Z = \dfrac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ soit ici $\dfrac{F - 0,15}{\sqrt{\frac{0,15\times 0,85}{200}}}$. Sous l'hypothèse H_0, Z est approximativement distribuée selon la loi normale $\mathcal{N}(0, 1)$.

Le test est unilatéral à droite, Z_α est définie à l'aide de $P(F \leq f_\alpha) = P(Z \leq Z_\alpha) = 0,95$.

On lit dans la table de la fonction de répartition de la loi normale centrée réduite

$$Z_\alpha = 7_{0,95} = 1,645 \Leftrightarrow f_\alpha = Z_\alpha \times \sqrt{\frac{0,15 \times 0,85}{200}} + 0,15 = 0,192.$$

Soit f_{obs} la valeur de la variable F observée sur l'échantillon.

- Si $f_{obs} > f_\alpha$, on rejette l'hypothèse nulle H_0.
- Si $f_{obs} < f_\alpha$, on ne rejette pas l'hypothèse nulle H_0.

Calculons la valeur observée : $f_{obs} = \dfrac{40}{200} = 0,2$. Comme $(f_{obs} = 0,2) > (f_\alpha = 0,192)$, on rejette l'hypothèse H_0.

Conclusion, au seuil de signification 5%, et au vu de l'échantillon, on rejette l'hypothèse nulle H_0 : plus de 15% des camemberts de la production ont une masse supérieure à 257g. Le reponsable de production n'a pas tort.