E.S.S.T.I.N. **Jeudi 11 mai 2017**
UNIVERSITE DE LORRAINE

3e année
Durée : 1 heure

Eléments de correction
DEVOIR SURVEILLE n°1

Statistiques

Une entreprise fabrique des comprimés effervescents. Le cahier des charges indique que chaque comprimé doit contenir 1625 mg de bicarbonate de sodium.

Afin de contrôler la fabrication, l'ingénieur qualité a prélevé un échantillon de 150 comprimés et a mesuré la quantité de bicarbonate de sodium pour chacun d'eux. Les résultats sont résumés dans le tableau suivant :

X	$[1610;1615[$	$[1615;1620[$	$[1620;1625[$	$[1625;1630[$	$[1630;1635[$	$\sum_{i=1}^{5} n_i\, x_{i\,cc} = 243\,820$
Effectif	7	8	42	75	18	$\sum_{i=1}^{5} n_i\, x_{i\,cc}^2 = 396\,324\,537{,}5$

1. *Dans le cadre de cette étude, préciser le nom et la nature du caractère, l'individu, l'échantillon et la population.*

Nom : Masse de bicarbonate de sodium

Nature : Quantitative continue

Individu : 1 comprimé - Echantillon : 150 comprimés - Population : Tous les comprimés produits

2. *Représenter graphiquement les résultats du tableau.*

Histogramme des effectifs (ou des fréquences) ou diagramme des fréquences cumulées

3. *A partir des résultats fournis, calculer l'étendue, la moyenne et l'écart type de l'échantillon.*

$$E = 1635 - 1610 = 25 \text{ mg}$$

$$\bar{x} = \frac{\sum_{i=1}^{5} n_i\, x_{i\,cc}}{150} = 1625{,}47 \text{ mg} \qquad s = \sqrt{\frac{\sum_{i=1}^{5} n_i\, x_{i\,cc}^2}{150} - \bar{x}^2} = 4{,}66 \text{ mg}$$

4. *Etablir le tableau des fréquences cumulées afin de calculer la médiane par interpolation linéaire. Interpréter le résultat obtenu.*

X	1610	1615	1620	1625	1630	1635
Fréquence cumulée	0	0,05	0,10	0,38	0,88	1

$$Me = 1625 + (1630 - 1625)\frac{0{,}50 - 0{,}38}{0{,}88 - 0{,}38} = 1626{,}2 \text{ mg}$$

Il y a 50 % des mesures inférieures à la médiane et 50 % des mesures supérieures à la médiane. La valeur médiane est légèrement supérieure à la valeur moyenne et proche de la valeur du cahier des charges.

5. *Discuter le nombre de classes choisi par l'analyste pour la mise en classes des données brutes.*

L'analyste a choisi de répartir les données brutes en 5 classes de hauteur 5 mg. Le nombre de classes choisi est faible en regard des préconisations pour une taille d'échantillon de 150 et des effectifs importants des classes centrales ce qui engendrera un manque de précision pour les calculs.

$$k = \sqrt{150} \simeq 12{,}2 \qquad \text{ou} \qquad k = 1 + 3{,}322 \log(150) \simeq 8{,}2$$

6. *Expliquer la raison pour laquelle, on peut assimiler l'échantillon prélevé à un échantillon non exhaustif.*

L'échantillon est prélevé sans remise dans une population de très grande taille (n << N). Dans ce cas, les prélèvements peuvent être considérés indépendants comme dans le cas d'un échantillon prélevé avec remise.

7. *Donner les estimations ponctuelles sans biais de la moyenne et de la variance de la population.*

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{5} n_i\, x_{i\,cc}}{150} = 1625{,}47 \text{ mg} \qquad \hat{\sigma}^2 = s^{*2} = \frac{n}{n-1}s^2 = 21{,}845$$

8. *En utilisant les résultats du cours, justifier le choix de la formule de l'estimation de la variance de la population (<u>ne pas faire de démonstrations</u>).*

La formule pour le calcul de l'estimation de la variance a été choisie car la moyenne de la population est inconnue et on souhaitait un estimateur à biais nul.

On appelle $\bar{X}$ la variable aléatoire qui, à tout échantillon de taille n, associe la quantité moyenne de bicarbonate de sodium de cet échantillon.

9. *Préciser la loi suivie par $\bar{X}$ ainsi que ses paramètres.*

$T = \dfrac{\bar{X} - \mu}{s^*/\sqrt{n}}$ suit une loi de Student à n – 1 degrés de liberté car la variance de la population est inconnue (le théorème central limite ne s'applique pas).

10. *La loi de $\bar{X}$ peut être approximée par une autre loi. Préciser le type de cette loi ainsi que ses paramètres.*

Comme la taille de l'échantillon est grande, la loi de Student peut être approximée par une loi normale. On peut donc considérer que $U = \dfrac{\bar{X} - \mu}{s^*/\sqrt{n}}$ suit une loi normale centrée réduite.

11. *Déterminer, en utilisant la loi approximée, un intervalle de confiance bilatéral à 95 % pour la moyenne de la population. Interpréter le résultat.*

La variance de la population n'est pas connue et l'échantillon est de grande taille d'où :

$$\bar{x} - u_{1-\alpha/2}\frac{s^*}{\sqrt{n}} \le \mu \le \bar{x} + u_{1-\alpha/2}\frac{s^*}{\sqrt{n}} \quad \text{avec} \quad u_{1-\alpha/2} = 1{,}96 \quad \text{d'où} \quad 1624{,}7 \text{ mg} \le \mu \le 1626{,}2 \text{ mg}$$

12. *Quelle devrait être la taille minimum de l'échantillon à prélever pour estimer, avec un niveau de confiance de 95 %, la quantité moyenne de bicarbonate de sodium de la population à $\pm 0{,}5$ mg près ?*

$$k = u_{1-\alpha/2}\frac{s^*}{\sqrt{n}} = 0{,}5 \quad \text{avec} \quad u_{1-\alpha/2} = 1{,}96 \qquad \text{d'où} \qquad n = \left(\frac{u_{1-\alpha/2}\, s^*}{0{,}5}\right)^2 = 335{,}7 \simeq 336$$

13. Sur la base des résultats présentés dans le tableau, le fabricant peut-il conclure que le cahier des charges est en moyenne respecté ? Détailler la démarche utilisée.

1. H0 μ = 1625 mg H1 $\mu \neq$ 1625 mg
2. $\alpha = 0{,}05$
3. Echantillon grand – Variance population inconnue – Approximation de la loi de Student par la loi normale
4. $U = \dfrac{\overline{X} - 1625}{s^* / \sqrt{n}}$ / H0 $\mu = 1625$ mg $\rightarrow$ LNCR
5. $\underline{\alpha = 0{,}05}$ $u_{1-\alpha/2} = 1{,}96$
6. $u_{Ech} = 1{,}22$
7. $\underline{\alpha = 0{,}05}$ $-u_{1-\alpha/2} < u_{Ech} < u_{1-\alpha/2}$ Non refus de H0

On peut donc considérer que l'échantillon n'est pas en contradiction avec l'hypothèse selon laquelle la quantité moyenne de bicarbonate de soude de la population est égale à 1625 mg ($\alpha = 0{,}05$) et conclure que le cahier des charges est en moyenne respecté.

14. Quels sont les avantages et les inconvénients de ne pas effectuer les analyses précédentes sur les données brutes ?

<u>Avantages</u> : calculs plus rapides, graphiques plus lisibles mais attention à bien choisir les paramètres de la mise en classes.

<u>Inconvénients</u> : perte d'information due à l'hypothèse d'une répartition des données uniforme dans chaque classe ce qui engendre un manque de précisions dans les résultats surtout si le nombre de classes choisi est petit.