Polytech Nancy	**19 avril 2018**
UNIVERSITE DE LORRAINE	
3e année	
Durée : 2 heures	

Eléments de correction
DEVOIR SURVEILLE n°2

Statistiques

EXERCICE 1 *(6 points)*

Dans le but d'étudier les taux de contamination des sols par une substance toxique X, 29 échantillons de sols ont été prélevés. Parmi ces 29 échantillons, 14 échantillons ont été prélevés dans des régions rurales et 15 dans des régions urbaines.

Les taux de X sont mesurés en 0,0001 gramme par kg de sol. Les principaux résultats sont regroupés dans le tableau suivant :

	Régions rurales	*Régions urbaines*
x_{min}	*1,40*	*2,20*
Q1	*1,65*	*8,50*
Médiane	*3,35*	*11,00*
Q3	*5,25*	*12,00*
x_{Max}	*14,00*	*18,00*
Moyenne	*4,46*	*10,54*
Ecart type	*3,72*	*3,54*
Valeurs aberrantes	*11,50 – 14,00*	*2,20 – 18,00*
Valeurs adjacentes	*5,90*	*7,20 – 15,40*

1. *Tracer deux diagrammes en boîte : un pour les régions rurales et un pour les régions urbaines.*

Diagramme en boîtes faisant apparaître les données suivantes : Q1, médiane, Q3, moyenne, valeurs minimum et maximum, valeurs adjacentes et valeurs aberrantes. Adopter la même échelle pour les deux diagrammes afin de les comparer.

2. *Justifier, à l'aide de calculs, l'existence des valeurs aberrantes parmi les données rurales et urbaines. Commenter les résultats.*

Calcul des valeurs pivots : Q1 – 1,5 EIQ et Q3 + 1,5 EIQ

Régions rurales : -3,75 et 10,65 d'où valeurs aberrantes 11,50 et 14,00

Il n'y a pas de valeur aberrante pour les valeurs faibles de dosages. Il y a deux valeurs aberrantes pour les valeurs élevées, l'une proche de la valeur pivot, l'autre plus éloignée. Les deux valeurs aberrantes étant largement supérieures à la valeur adjacente, il est donc nécessaire de vérifier ces deux mesures.

Régions urbaines : 3,25 et 17,25 d'où valeurs aberrantes 2,20 et 18,00

Il y a une valeur aberrante pour les valeurs faibles de dosages et une valeur aberrante pour les valeurs élevées. Les deux valeurs sont relativement proches des valeurs pivots mais très éloignées des valeurs adjacentes. Il est donc également nécessaire de vérifier ces deux mesures.

3. *Commenter et comparer les diagrammes en boîte obtenus.*

Régions rurales

Forme : distribution relativement symétrique si l'on retire les deux valeurs aberrantes, forte asymétrie si l'on conserve les deux valeurs aberrantes.

Tendance centrale : médiane et moyenne assez différentes à cause des deux valeurs aberrantes qui impactent beaucoup le calcul de la moyenne.

Dispersion : très grande dispersion, coefficient de variation égal à 83 %.

Valeurs aberrantes : les valeurs aberrantes ont une influence sur les calculs de la moyenne et de l'écart type, elles sont à vérifier, les résultats seraient beaucoup plus homogènes sans ces deux valeurs.

Régions urbaines

Forme : distribution presque symétrique, moyenne proche de médiane.

Tendance centrale : moyenne proche de médiane, attention aux deux valeurs aberrantes qui impactent le calcul de la moyenne.

Dispersion : dispersion moyenne, coefficient de variation égal à 33 %.

Valeurs aberrantes : les valeurs aberrantes ont une influence sur les calculs de la moyenne et de l'écart type, elles sont à vérifier, les résultats seraient beaucoup plus homogènes sans ces deux valeurs.

Comparaison

On observe un décalage des deux diagrammes en boîtes. Les boîtes sont disjointes si l'on exclut les valeurs aberrantes.

Les taux sont en général beaucoup plus élevés pour les régions urbaines que pour les régions rurales (moyenne et médiane beaucoup plus élevées pour les régions urbaines).

Les EIQ et les écarts types sont proches. Les dispersions sont équivalentes mais il faut vérifier les quatre valeurs aberrantes afin de décider de les conserver ou de les exclure de l'étude.

EXERCICE 2 (14 points)

Dans la fabrication d'une pièce dont plusieurs cotes doivent être contrôlées, le contrôle est effectué par le calcul d'une proportion de pièces non conformes.

Le contrat client-fournisseur stipule que 95 % au moins des pièces livrées doivent être conformes, au niveau de signification de 5 %.

Un échantillon de 220 pièces prélevées de façon aléatoire dans la production présente un nombre de pièces non conformes égal à 14.

1. *Quelle loi de probabilité suit la variable aléatoire X : nombre de pièces non conformes dans un échantillon de taille n ?*

$X \rightarrow B(n\,;p)$ avec n : taille de l'échantillon et p : probabilité qu'une pièce soit non conforme

2. *Approximer cette loi par la loi normale. Déterminer les paramètres de la loi normale. Préciser les conditions de l'approximation.*

$X \rightarrow N\left(n\,p\,;\sqrt{n\,p(1-p)}\right)$ avec $E(X)=n\,p$ et $V(X)=n\,p(1-p)$

n grand et p pas trop éloigné de 0,5 ou $n\,p \geq 5$ et $n(1-p) \geq 5$

3. *Déduire de la question 2. la loi d'échantillonnage de la proportion de pièces non conformes de l'échantillon. Déterminer les paramètres de cette loi.*

$\hat{P} = \frac{X}{n} \rightarrow N\left(p\,;\sqrt{\frac{p(1-p)}{n}}\right)$ avec $E(\hat{P}) = \frac{1}{n}E(X) = p$ et $E(\hat{P}) = \frac{1}{n^2}V(X) = \frac{p(1-p)}{n}$

4. *Déterminer un intervalle de confiance à 95 % pour la proportion de pièces non conformes de la population.*

Intervalle de confiance à 95 % pour la proportion population :

$n\,f = 14 \geq 5$ et $n(1-f) = 206 \geq 5$

$f - u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}} \leq p \leq f + u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}}$ avec

$f = \frac{14}{220}$ et $u_{1-\alpha/2} = u_{0,975} = 1,96$

$0,031 \leq p \leq 0,096$

5. *Peut-on considérer que la proportion de pièces non conformes prévue dans le contrat est respectée ?*

H0 $p = 0,05$ $\rightarrow$ condition respectée

H1 $p > 0,05$ $\rightarrow$ proportion de pièces non conformes trop grande

$n\,0,05 \geq 5$ et $n(1-0,05) \geq 5$

$$u_{Echantillon} = \frac{f - 0,05}{\sqrt{\frac{f(1-f)}{n}}} = 0,93$$

$\alpha = 0,05$ $u_{1-\alpha} = 1,645$

$u_{Echantillon} \leq u_{1-\alpha} = 1,645$ d'où Non refus de H0

On peut conclure, avec une grande confiance, que le contrat est respecté.

6. *Calculer le seuil descriptif du test. Conclure.*

$\alpha_P = \Pr(U > u_{Echantillon} = 0,93) = 1 - \Pr(U \leq u_{Echantillon} = 0,93) = 1 - 0,8238 = 0,1762$

Le seuil descriptif du test étant assez grand, la conclusion de non refus de H0 est prise avec une grande confiance.

7. *Calculer le nombre de pièces à tester pour obtenir une marge d'erreur sur la proportion de pièces non conformes égale à 0,02 avec un risque alpha égal à 0,1.*

$k = u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0,02$ d'où

$n = \left(\frac{u_{1-\alpha/2}}{k}\right)^2 p(1-p) = 321,3 \simeq 322$ avec $u_{1-\alpha/2} = u_{0,95} = 1,645$ et $p = 0,05$

$n\,0,05 \geq 5$ et $n(1-0,05) \geq 5$

A présent, on cherche, par deux méthodes, à déterminer s'il existe une différence entre les proportions de pièces non conformes fabriquées par deux machines.

Les résultats sont présentés dans le tableau suivant :

	Machine A	*Machine B*
Taille d'échantillon	*220*	*220*
Nombre de pièces non conformes	*14*	*18*

Première méthode : test de comparaison de deux proportions

8. *Elaborer le test de comparaison avec un risque de 5 %. Conclure.*

H0 $p_A = p_B$

H1 $p_A \neq p_B$

$n_A f_A \geq 5$ et $n_A\left(1-f_A\right) \geq 5$

$n_B f_B \geq 5$ et $n_B\left(1-f_B\right) \geq 5$

$$u_{Echantillon} = \frac{f_A - f_B}{\sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{n_A}+\frac{1}{n_A}\right)}} = -0,73 \quad \text{avec} \quad \hat{p} = \frac{n_A f_A + n_B f_B}{n_A + n_B}$$

$\alpha = 0,05 \quad u_{1-\alpha/2} = 1,96$

$-u_{1-\alpha/2} \leq u_{Echantillon} \leq u_{1-\alpha/2}$ d'où Non refus de H0

On peut conclure, avec une grande confiance, qu'il n'y a pas de différence entre les proportions de pièces non conformes fabriquées par les deux machines.

Seconde méthode : test du Khi-Deux

9. Elaborer le test du Khi-Deux avec un risque de 5 %. Conclure.

H0 $p_A = p_B = \overline{p}$

H1 $p_A \neq p_B$

Calcul des effectifs théoriques en supposant que H0 est vraie :

n_t / H0	A	B	
Pièces conformes	204	204	408
Pièces non conformes	16	16	32
	220	220	440

$\forall\ i\ j \quad n_{tij} \geq 5$

$$\chi^2_{Echantillon} = \frac{\left(n_{o11}-n_{t11}\right)^2}{n_{t11}} + \frac{\left(n_{o12}-n_{t12}\right)^2}{n_{t12}} + \frac{\left(n_{o21}-n_{t21}\right)^2}{n_{t21}} + \frac{\left(n_{o22}-n_{t22}\right)^2}{n_{t22}} = 0,54$$

$\alpha = 0,05 \quad \chi^2_{1-\alpha\,;\,\nu=1} = 3,84$

$\chi^2_{Echantillon} \leq \chi^2_{1-\alpha\,;\,\nu=1}$ d'où Non refus de H0

On peut conclure, avec une grande confiance, qu'il n'y a pas de différence entre les proportions de pièces non conformes fabriquées par les deux machines.

10. Quelles sont les différences entre les deux méthodes ?

Les deux méthodes amènent à la même conclusion et nécessitent toutes les deux de faire des approximations valides pour des échantillons grands et des proportions ni trop faibles, ni trop grandes.

Les distributions d'échantillonnage utilisées dans les deux tests ne sont pas les mêmes.

La méthode du Khi-Deux permet en plus de comparer plus de deux proportions.