E.S.S.T.I.N. **Mercredi 15 juin 2016**
UNIVERSITE DE LORRAINE
3e année
Durée : 2 heures

Eléments de correction
DEVOIR SURVEILLE n°2

Statistiques

Barème indicatif sur 20 : Exercice 1 : 10 points Exercice 2 : 6 points Exercice 3 : 4 points

EXERCICE 1

Un fabricant de tissu essaye une nouvelle machine. Il fabrique des échantillons de 10 mètres et compte le nombre de défauts par échantillon. Après avoir examiné 126 échantillons, il a trouvé les résultats suivants :

Nombre de défauts : X	*0*	*1*	*2*	*3*	*4*
Nombre d'échantillons	*44*	*49*	*24*	*7*	*2*

1. Préciser le nom, la nature et les valeurs possibles de la variable X.

X : nombre de défauts par échantillon de 10 mètres

Variable quantitative discrète

Valeurs possibles : de 0 à l'infini

2. Préciser l'individu, l'échantillon et la population associés à cette étude.

Individu : 1 échantillon de 10 mètres

Echantillon : 126 échantillons de 10 mètres

Population : tous les échantillons de 10 mètres que peut produire la machine

3. Calculer la moyenne arithmétique et la variance de X.

$$\bar{x} = 1 \text{ défaut / échantillon} \qquad s^2 = 0{,}905$$

4. Justifier le choix d'une loi de Poisson de paramètre μ pour modéliser la variable X.

La variable X est discrète, ses valeurs possibles ne sont pas finies et sa moyenne arithmétique est proche de sa variance donc le modèle choisi peut être une loi de Poisson de paramètre μ.

5. Ecrire alors la loi de probabilité de X, préciser son espérance et sa variance.

$$\Pr(X = x) = \frac{e^{-\mu}\mu^x}{x\,!} \qquad E(X) = \mu \qquad V(X) = \mu$$

6. Quelle est la borne inférieure de la variance de tout estimateur ponctuel de μ construit à partir d'un échantillon indépendant de taille n ?

$$V(\hat{\mu}) \geq \frac{1}{I_n(\mu)}$$

$I_n(\mu) = n.\,I_1(\mu)$ car le domaine de définition de X ne dépend pas de μ

$$I_1(\mu) = E\left\{\left[\frac{d}{d\mu}\ln\frac{e^{-\mu}\mu^x}{x\,!}\right]^2\right\} = E\left\{\left[\frac{x-\mu}{\mu}\right]^2\right\} = \frac{\mu}{\mu^2} = \frac{1}{\mu}$$

d'où $\quad V(\hat{\mu}) \geq \dfrac{\mu}{n}$

7. La moyenne arithmétique $\bar{X}$ est-elle un estimateur efficace pour μ ?

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\text{Biais}(\bar{X}) = E(\bar{X}) - \mu = \frac{1}{n}\sum_{i=1}^{n} E(X_i) - \mu = \mu - \mu = 0$$

$$V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\mu}{n} \qquad \text{car } X_i \text{ indépendants}$$

Le biais est nul et la variance est égale à la borne inférieure donc $\bar{X}$ est un estimateur efficace pour μ.

8. *Tester l'ajustement de la distribution expérimentale à une loi de Poisson dont on précisera le paramètre.*

H0 $X \rightarrow P\left(\hat{\mu} = \bar{x} = 1\right)$

H1 $X \rightarrow$ Autre loi

Calcul des probabilités théoriques puis des effectifs théoriques / H0 vraie.

Regroupement des deux dernières classes pour respecter la condition : effectifs théoriques supérieurs ou égaux à 5.

Calcul des 4 termes d'écarts (après regroupement) :

$$\chi^2_{Ech} = \sum_{i=1}^{4} \frac{\left(n_{o\,i} - n_{t\,i}\right)^2}{n_{t\,i}} = 0,42$$

α = 5 % $\chi^2_{Ech} \leq \chi^2_{0,95\,;\,4-1-1=2} = 5,99$ Acceptation de H0

On peut affirmer avec une grande confiance que le nombre de défauts par échantillon de 10 mètres de tissu suit la loi de Poisson de paramètre 1.

EXERCICE 2

Une usine fabrique des chargeurs pour ordinateur portable conçus pour délivrer une tension de sortie égale à 20 V.

On suppose que la loi normale d'écart type 2 V est un bon modèle pour la distribution de cette variable.

Pour vérifier que la tension de sortie n'est pas trop importante, on prélève de façon aléatoire un échantillon de n chargeurs et on met en œuvre un test d'hypothèses.

1. *Détailler toutes les étapes du test utilisé en choisissant un risque de première espèce de 5 %. Faire un schéma.*

H0 μ = 20 V Tension de sortie conforme

H1 μ > 20 V Tension de sortie trop élevée

X suit une loi normale, variance population connue

Pour α = 0,05, si $\bar{x}_{Ech} > \bar{x}_c = 20 + u_{0,95}\frac{\sigma}{\sqrt{n}} = 20 + 1,645\frac{2}{\sqrt{n}}$ on refuse H0 (tension de sortie trop élevée), sinon on accepte (tension de sortie conforme).

Schéma de $\bar{X}$ suivant la loi normale centrée sur 20 V, délimitant les zones d'acceptation et de refus de H0 et représentant le risque α.

2. *On suppose que la taille de l'échantillon prélevé est égale à 16. La tension de sortie moyenne mesurée sur cet échantillon est égale à 21,25 V. Quel est le seuil descriptif du test ?*

$n = 16$ $\bar{x}_{Ech} = 21,25$ V

$$\alpha_P = \Pr\left(\bar{X} > 21,25\right) = 1 - \Pr\left(\bar{X} \leq 21,25\right) = 1 - \Pr\left(U = \frac{\bar{X} - 20}{\sigma / \sqrt{n}} \leq \frac{21,25 - 20}{2 / \sqrt{16}}\right)$$

$$\alpha_P = 1 - \Pr\left(U \leq 2,5\right) = 1 - 0,9938 = 0,0062 = 0,62\ \%$$

3. *Etablir la règle de décision en fonction de la valeur de ce seuil descriptif et conclure.*

<u>Règle de décision</u> :

Si $\alpha_P < \alpha$ Refus de H0

Si $\alpha_P \geq \alpha$ Acceptation de H0

$n = 16$ $\bar{x}_{Ech} = 21,25$ V $\alpha = 0,05$

$\alpha_P = 0,0062 < \alpha$ Refus de H0

On conclut avec une grande confiance (seuil descriptif très faible) que la tension de sortie est trop élevée.

4. *Quelle devrait être la taille minimum de l'échantillon à prélever pour que la puissance du test soit égale à 0,95 lorsque l'on choisit pour alternative 21 V ? Faire un schéma.*

H0 $\mu = 20\text{ V} = \mu_0$ Tension de sortie conforme

H1 $\mu = 21\text{ V} = \mu_1$ Tension de sortie trop élevée

Puissance du test $= 0,95 = 1 - \beta$

$\bar{x}_c = 20 + u_{1-\alpha}\frac{\sigma}{\sqrt{n}}$ et $\bar{x}_c = 21 - u_{1-\beta}\frac{\sigma}{\sqrt{n}}$

$$\text{d'où}\quad n = \frac{\sigma^2\left(u_{1-\alpha} + u_{1-\beta}\right)^2}{\left(21-20\right)^2} = \frac{\sigma^2\left(1,645 + 1,645\right)^2}{\left(21-20\right)^2} = 43,3 \simeq 44$$

et $\bar{x}_c = 20,5$ V

Schéma de $\bar{X}$ suivant la loi normale centrée sur 20 V et la loi normale centrée sur 21 V, visualisant les zones d'acceptation et de refus de H0 et de H1 délimitées par $\bar{x}_c$ et représentant les risques α et β.

EXERCICE 3

Sur la production d'une journée de deux machines fabriquant une même pièce, on a prélevé deux échantillons indépendants. Le tableau suivant présente les résultats de ce contrôle qualité :

	Machine 1	*Machine 2*
Nombre de pièces contrôlées	*100*	*120*
Nombre de pièces avec défaut	*12*	*16*

1. *Peut-on conclure au risque de 5 % que l'écart observé entre les deux machines est significatif ? Détailler la démarche utilisée.*

$n_1 = 100 \qquad f_1 = \frac{12}{100} = 0{,}120 \qquad n_2 = 120 \qquad f_2 = \frac{16}{120} = 0{,}133$

H0 $\quad p_1 = p_2$

H1 $\quad p_1 \neq p_2$

$n_1 \,.\, f_1 \geq 5 \qquad$ et $\qquad n_2 \,.\, f_2 \geq 5$

$$\hat{P}_1 - \hat{P}_2 \diagup_{H0} \to N\left(0\,;\sqrt{\hat{P}\left(1-\hat{P}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right)$$

$$u_{Ech} = \frac{f_1 - f_2}{\sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = -0{,}295 \qquad \text{avec} \qquad \hat{p} = \frac{n_1\,.\,f_1 + n_2\,.\,f_2}{n_1 + n_2}$$

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

$-1{,}96 \leq u_{Ech} \leq 1{,}96 \qquad$ Acceptation de H0

On ne peut pas conclure que les proportions de pièces avec défaut produites par les deux machines sont différentes (α = 5 %). L'écart est donc non significatif.

Un test du Khi-Deux de comparaison des deux proportions mène à la même conclusion.

2. *Quelle est l'estimation ponctuelle de la différence des deux proportions de pièces avec défaut ?*

$$\hat{P}_1 \to N\left(f_1\,;\sqrt{\frac{f_1\left(1-f_1\right)}{n_1}}\right) \qquad \hat{P}_2 \to N\left(f_2\,;\sqrt{\frac{f_2\left(1-f_2\right)}{n_2}}\right) \qquad \text{avec}$$

$$E\left(\hat{P}_1\right) = f_1 \qquad E\left(\hat{P}_2\right) = f_2 \qquad V\left(\hat{P}_1\right) = \frac{f_1\left(1-f_1\right)}{n_1} \qquad V\left(\hat{P}_2\right) = \frac{f_2\left(1-f_2\right)}{n_2}$$

$$\text{d'où} \qquad \hat{P}_1 - \hat{P}_2 \to N\left(f_1 - f_2\,;\sqrt{\frac{f_1\left(1-f_1\right)}{n_1}+\frac{f_2\left(1-f_2\right)}{n_2}}\right)$$

$$E\left(\hat{P}_1 - \hat{P}_2\right) = E\left(\hat{P}_1\right) - E\left(\hat{P}_2\right) = f_1 - f_2 = -0{,}013$$

3. *Quel est l'écart type de la différence de ces deux proportions ?*

$$\sqrt{V\left(\hat{P}_1 - \hat{P}_2\right)} = \sqrt{V\left(\hat{P}_1\right) + V\left(\hat{P}_2\right)} = \sqrt{\frac{f_1\left(1-f_1\right)}{n_1}+\frac{f_2\left(1-f_2\right)}{n_2}} = 0{,}0449$$

car les deux variables aléatoires sont indépendantes.

4. *Déterminer l'intervalle de confiance à un niveau de confiance de 95 % associé à la différence des deux proportions. Conclure.*

$$\hat{P}_1 - \hat{P}_2 \to N\left(f_1 - f_2\,;\sqrt{\frac{f_1\left(1-f_1\right)}{n_1}+\frac{f_2\left(1-f_2\right)}{n_2}}\right) \qquad \text{d'où}$$

$$f_1 - f_2 - u_{1-\alpha/2}\sqrt{\frac{f_1\left(1-f_1\right)}{n_1}+\frac{f_2\left(1-f_2\right)}{n_2}} \leq p_1 - p_2 \leq f_1 - f_2 + u_{1-\alpha/2}\sqrt{\frac{f_1\left(1-f_1\right)}{n_1}+\frac{f_2\left(1-f_2\right)}{n_2}}$$

$-0{,}101 \leq p_1 - p_2 \leq 0{,}074$

La différence entre les deux proportions n'est pas significative (α = 5 %) car la valeur 0 est comprise dans l'intervalle de confiance à 95 %.

5. *Comparer les conclusions des questions 1. et 4.*

Les deux conclusions sont identiques pour une même valeur de α.