Université Paul Valéry
Arts - Lettres - Langues
Sciences humaines & sociales

L3 AES–MISASHS
Statistique-Inférence - E511XS4
Examen Décembre 2014

NOM PRÉNOM :
N° Carte d'étudiant :
Signature :

- **Durée** : 2h. Calculatrice autorisée. Téléphone portable interdit.
- Les 5 exercices sont indépendants.

Exercice 1: Dans une population, on suppose que 8% des personnes sont vaccinées contre la grippe.

1. Exprimer par une variable aléatoire notée V le fait qu'une personne prise au hasard dans la population soit vaccinée ou pas contre la grippe; quelle est sa loi?

   $V =$ *"Etre vacciné (ou pas) contre la grippe"*

   $V \sim Bern(p = 0.08)$

2. Donner la loi exacte de la variable aléatoire, notée NV, du nombre de personnes vaccinées parmi n personnes prises au hasard dans la population.

$$NV \sim Binom(n; p = 0.08)$$

3. En supposant n assez grand, quelle est la loi approchée de NV?

   Pour n assez grand,

$$NV \sim N(n * 0.08; n * 0.08 * 0.92)$$

4. Donner l'expression de l'intervalle de dispersion de NV pour un échantillon de taille n (choisir un risque $\alpha = 1.6\%$).

$$ID_{98.4\%}(NV) = \left[n * p_0 - l_{\alpha/2}\sqrt{n * p_0 * (1 - p_0)}; n * p_0 + l_{\alpha/2}\sqrt{n * p_0 * (1 - p_0)}\right]$$

   avec $p_0 = 0.08$

5. Calculer les bornes de cet intervalle lorsque la taille de l'échantillon est égale à 520.

   $l_{\alpha/2} = l_{0.8\%} = 2.4089$ *d'où*

$$ID_{98.4\%}(NV) = [26.7; 56.5]$$

Exercice 2: Vers la fin des années 2000, 6% de la population d'un pays était atteint d'une maladie grave. Après une campagne de prévention effectuée dans le pays, on pense que ce pourcentage a baissé. Pour s'en assurer, on a relevé le nombre de personnes atteintes de la maladie dans un échantillon de 1000 individus.

1. Indiquer les hypothèses du test à mettre en oeuvre.

   Hypothèses:

   $H_0 : p = p_0$ *contre* $H_1 : p < p_0$

   avec $p=$*"proba d'être malade (après la campagne de prévention)" et* $p_0 = 0.06$.

2. Calculer la zone de rejet de ce test (on prendra $\alpha = 4\%$).

   Rejet de H_0 *si* $\bar{X} < p_0 - l_{0.04} * \sqrt{\frac{p_0*(1-p_0)}{n}} = 0.06 - 1.7507 * \sqrt{\frac{0.06*0.94}{1000}}$
   donc Rejet de H_0 *si* $\bar{X} < 0.0469$

3. Sachant que sur les 1000 individus de l'échantillon, 43 sont atteints de la maladie, quelle est la décision à prendre concernant ce test statistique?

   $\bar{x} = \frac{43}{1000} = 0.043 < 0.0469$ *donc rejet de* H_0 *("baisse significative de la probabilité d'être malade").*

4. En choisissant un risque $\alpha = 3\%$, construire l'intervalle de confiance de la proportion de personnes atteintes par la maladie dans la population.

$$\begin{aligned} IC_{97\%}(p) &= \left[\bar{x} - l_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}; \bar{x} + l_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right] \\ &= \left[0.043 - l_{0.015}\sqrt{\frac{0.043 * 0.957}{1000}}; 0.043 + l_{0.015}\sqrt{\frac{0.043 * 0.957}{1000}}\right] \\ &= [0.029; 0.057] \end{aligned}$$

car $l_{0.015} = 2.1701$

Exercice 3: On se propose de modéliser le temps d'attente chez un médecin par une loi exponentielle $\mathcal{E}xp(\lambda)$ de paramètre $\lambda = 30$ min. Les valeurs de cette loi de probabilité sont résumées dans le tableau suivant :

Tps d'attente	< 10	$[10, 20[$	$[20, 30[$	$[30, 40[$	≥ 40
Probabilité	0.2835	0.2031	0.1455	0.1043	0.2636

Pour vérifier cette hypothèse de modélisation, on a relevé le temps d'attente de 80 patients chez ce médecin. Les résultats sont résumés dans le tableau suivant :

Tps d'attente	< 10	$[10, 20[$	$[20, 30[$	$[30, 40[$	≥ 40
Effectifs	11	8	25	20	16

1. Ecrire précisément les hypothèses du test à mettre en oeuvre.

   Hypothèses:

   H_0: *"X suit une loi exponentielle* $\mathcal{E}xp(\lambda = 30)$*"*

   contre

   H_1: *"X ne suit pas une loi exponentielle* $\mathcal{E}xp(\lambda = 30)$*"*

2. Donner alors l'expression de la statistique du test ainsi que sa loi sous l'hypothèse nulle.

$$\chi^2 = \sum_{k=1}^{5} \frac{(n_k - \tilde{n}_k)^2}{\tilde{n}_k}$$

Sous H_0, χ^2 suit une loi du Khi2 à 5-1=4 ddl.

3. Donner la zone de rejet du test (on prendra $\alpha = 5\%$).

Rejet de H_0 si $\chi^2 > l_4(95\%) = 9.4877$

4. Faire les calculs puis conclure (par une phrase).

Eff. théoriques: $\tilde{n}_1 = n * 0.2835 = 80 * 0.2835 = 22.68$
$\tilde{n}_2 = 16.25$, $\tilde{n}_3 = 11.64$, $\tilde{n}_4 = 8.34$ *et* $\tilde{n}_5 = 21.09$

Eff. observés: $n_1 = 11$, $n_2 = 8$,..., $n_5 = 16$

$$\begin{aligned} \chi^2 &= \frac{(11-22.68)^2}{22.68} + \frac{(8-16.25)^2}{16.25} + ... \\ &= 6.03 + 4.15 + ... \\ &= 10.18 + ... \end{aligned}$$

D'où $\chi^2 > 9.4877$: rejet de H_0 donc X ne suit pas une loi $\mathcal{E}xp(\lambda = 30)$.

Exercice 4: Un test de réaction a été effectué par des personnes provenant de 3 pays différents. Pour chaque personne, on a mesuré son temps de réaction en centièmes de secondes. Les résultats du test sont résumés dans le tableau ci-dessous :

	Nbre personnes	Tps de réaction moyen	Variance
Pays P_1	120	46.1	147.0
Pays P_2	100	50.0	114.1
Pays P_3	80	45.7	129.8

On souhaite alors comparer les résultats du test dans ces 3 pays.

1. Indiquer par une phrase les deux hypothèses du test à mettre en oeuvre.

H_0: "Les temps de réaction moyens (espérances) sont identiques dans les 3 pays"

contre

H_1: "Les temps de réaction moyens ne sont pas identiques dans les 3 pays"

2. Donner l'expression de la statistique du test et sa loi sous l'hypothèse nulle.

   Stat. du test: $T = (n-3) * \frac{INTER}{INTRA}$ *avec* $n = 300$.

   Sous H_0, T *suit une loi du Khi2 à 3-1=2 ddl.*

3. Donner la zone de rejet du test (on prendra $\alpha = 5\%$).

   Rejet de H_0 *si* $T > l_2(95\%) = 5.9915$

4. Faire les calculs et conclure.

   $\bar{y} = \frac{120*46.1+100*50+80*45.7}{300} = 47.29$
   $INTRA = 120*147+100*114.1+80*129.8 = 39434$
   $INTER = 120*(46.1-47.29)^2+100*(50-47.29)^2+80*(45.7-47.29)^2 = 1106.59$
   $t = (300-3)*\frac{1106.59}{39434} = 8.33 > 5.9915$ *donc rejet de* H_0 *("Les temps de réaction moyens ne sont pas identiques dans les 3 pays").*

<u>Exercice 5</u>: On dispose pour 600 personnes de l'observation des deux variables $X =$ "Salaire (en euros)" et $Y =$ "Epargne (en euros)". Les observations de ces deux variables sont notées $(x_1, y_1, x_2, y_2, ..., x_{600}, y_{600})$. On cherche à mettre en évidence l'existence d'une liaison linéaire entre X et Y.

1. Indiquer par une phrase les deux hypothèses du test à mettre en oeuvre.

   H_0*: "Absence de lien linéaire entre* X *et* Y*"*
   contre
   H_1*: "Existence d'un lien linéaire entre* X *et* Y*"*

2. Indiquer par une phrase en quoi consiste l'erreur de $1re$ espèce dans ce contexte.

   Affirmer qu'il y existe un lien linéaire entre X *et* Y *alors que ce n'est pas le cas.*

3. Donner l'expression de la statistique du test et sa loi sous l'hypothèse nulle.

   Stat. du test: $\hat{\rho}(X,Y) = \frac{cov(X,Y)}{\hat{\sigma}_X \hat{\sigma}_Y}$.

   Sous H_0, $\hat{\rho} * \sqrt{n-1} \sim N(0,1)$

4. Donner l'expression de la zone de rejet avec $\alpha = 2\%$.

   Rejet de H_0 *si* $|\hat{\rho}| > \frac{l_{\alpha/2}}{\sqrt{n-1}} = \frac{2.3263}{\sqrt{599}} = 0.095$.