Correction examen ModIA du 3 février 2022

Jean-Yves TOURNERET

Exercice 1 : Comparaison des réussites de trois écoles (4pts)
Trois écoles ont relevé le nombre d'admis et d'ajournés au brevet dans le tableau suivant

	École 1	École 2	École 3	$N_{k.}$
Admis	50	35	85	170
Ajournés	30	15	35	80
$N_{.l}$	80	50	120	250

On désire tester si les résultats de ces trois écoles sont significativement différents à l'aide d'un test du χ^2 d'homogénéité

1. Exprimer la valeur de la statistique du test du χ^2 d'homogénéité notée I_n en fonction des données du problème (sans chercher à la calculer).
   Le test du χ^2 d'homogénéité rejette l'hypothèse H_0 si
   $$I_n = \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{\left(N_{k,l} - \frac{N_{k,.}N_{.,l}}{n}\right)^2}{\frac{N_{k,.}N_{.,l}}{n}} > S_{K,L,\alpha}$$
   avec $S_{K,L,\alpha} = F^{-1}_{(K-1)(L-1)}(\alpha)$ qui est la fonction de répartition inverse d'une loi du χ^2 à $(K-1)(L-1)$ degrés de liberté. Dans notre exemple, on a
   $$I_n = \frac{\left(50 - \frac{170\times 80}{250}\right)^2}{\frac{170\times 80}{250}} + \frac{\left(35 - \frac{170\times 50}{250}\right)^2}{\frac{170\times 50}{250}} + \frac{\left(85 - \frac{170\times 120}{250}\right)^2}{\frac{170\times 120}{250}} + \frac{\left(30 - \frac{80\times 80}{250}\right)^2}{\frac{80\times 80}{250}} + \frac{\left(13 - \frac{80\times 50}{250}\right)^2}{\frac{80\times 50}{250}} + \frac{\left(35 - \frac{80\times 120}{250}\right)^2}{\frac{80\times 120}{250}}.$$
2. Déterminer la loi asymptotique de I_n et en déduire le seuil du test en fonction du risque de première espèce α et de l'inverse de la fonction de répartition d'une loi du χ^2_K notée F_K^{-1} pour une valeur de K que l'on déterminera.
   La loi asymptotique de I_n est une loi du χ^2 à $(K-1)(L-1)$ degrés de liberté, i.e., $I_n \sim \chi^2_2$. On en déduit
   $$\alpha = P[\text{Rejeter } H_0 | H_0 \text{ vraie}] = P[I_n > S_\alpha | I_n \sim \chi^2_2] = 1 - F_2(S_\alpha),$$
   d'où
   $$S_\alpha = F_2^{-1}(1-\alpha).$$
3. Exprimer la pvaleur de ce test en fonction de I_n et de la fonction de répartition d'une loi du χ^2_K notée F_K pour une valeur de K que l'on précisera.
   La pvaleur est la plus petite valeur de α pour laquelle on rejette H_0, soit
   $$\text{pval} = 1 - F_2(I_n).$$
4. Expliquer comment déterminer si les résultats de ces trois écoles sont significativement différents ou pas.
   Le principe du test du χ^2 d'homogénéité est de rejeter l'hypothèse H_0 (et donc de décider que les trois écoles ont des résultats significativement différents) si $I_n > S_\alpha$.

Exercice 2 : Comparaison du prix de deux produits (4pts)

Les prix en euros d'un même produit dans deux régions Françaises sont indiqués dans le tableau ci-dessous

R 1	$x_1 = 14$	$x_2 = 15.5$	$x_3 = 16.5$	$x_4 = 17$	$x_5 = 18$	$x_6 = 15$	$x_7 = 19.5$	$x_8 = 20.5$	$x_9 = 21$	$x_{10} = 24$
R 2	$y_1 = 17.5$	$y_2 = 16$	$y_3 = 18.5$	$y_4 = 22.5$	$y_5 = 22$	$y_6 = 18$	$y_7 = 20$	$y_8 = 25$	$y_9 = 24$	$y_{10} = 41$

On aimerait déterminer si le prix du produit dans les deux régions sont différents ou pas.

1. Déterminer la valeur de la statistique du test de Wilcoxon W_m associée à ce problème.
   Le test de Wilcoxon rejette l'hypothèse H_0 si $W_m < S_{1,\alpha}$ ou $W_n > S_{2,\alpha}$ où $W_m = \sum_{j=1}^{m} S_j$ et S_j est le rang de Y_j parmi les $n+m$ données réunies $(X_1, ..., X_n, Y_1, ..., Y_m)$. Dans cet exemple, on a les résultats suivants
   - Suite ordonnée
   $$(14, 15, 15.5, 16, 16.5, 17, 17.5, 18, 18, 18.5, 19.5, 20, 20.5, 21, 22, 22.5, 24, 24, 25, 21).$$
   - Données associées
   $$x_1, x_6, x_2, y_2, x_3, x_4, y_1, x_5 = y_6, y_3, x_7, y_7, x_8, x_9, y_5, y_4, x_{10} = y_9, y_8, y_{10}$$
   - rangs du groupe 2
   $$r_2 = 4, r_1 = 7, r_6 = 8.5, r_3 = 10, r_7 = 12, r_5 = 15, r_4 = 16, r_9 = 17.5, r_8 = 19, r_{10} = 20.$$
   - Statistique de Wilcoxon
     La somme des rangs s'écrit
   $$W_m = \sum_{i=1}^{10} r_i = 129.$$
2. Quelle est la loi asymptotique de W_m ? En utilisant cette loi asymptotique, exprimer la p-valeur du test de Wilcoxon en fonction de la fonction de répartition de la loi normale $\mathcal{N}(0,1)$ notée F.
   La loi asymptotique de W_m est une loi normale de moyenne $E[W_m] = \frac{nm}{2} + \frac{m(m+1)}{2} = 105$ et de variance $\mathrm{var}[W_m] = \frac{nm(n+m+1)}{12} = 175$. La $p-$valeur de ce test est définie par
   $$\text{p-val} = 2\left[1 - F\left(\frac{|W_m - E[W_m]| - 0.5}{\sqrt{\mathrm{var}[W_m]}}\right)\right].$$
3. Expliquer le principe de la correction de continuité.
   Sans correction de continuité, la $p-$valeur serait égale à
   $$\text{p-val} = 2\left[1 - F\left(\frac{|W_m - E[W_m]|}{\sqrt{\mathrm{var}[W_m]}}\right)\right].$$
   Le fait de retrancher 0.5 au numérateur permet de compenser le fait qu'on approche une loi discrète par une loi continue.
4. Expliquer comment procéder à l'aide de cette $p-$valeur pour décider si les deux ensembles de prix sont significativement différents avec un risque $\alpha = 0.05$.
   La $p-$valeur est la plus petite valeur de α pour laquelle on rejette H_0. Donc si la $p-$valeur est inférieure à 0.05, on rejette l'hypothèse H_0 avec le risque $\alpha = 0.05$ et donc on décide que si les prix du produit dans les deux régions sont différents. Si la $p-$valeur est supérieure à 0.05, on accepte l'hypothèse H_0 avec le risque $\alpha = 0.05$ et donc on décide que les prix du produit dans les deux régions sont les mêmes.

Exercice 3 : Droite de Henry (2pts)

1. Rappeler ce que représente une droite de Henry pour tester la loi normale.
   Soit $(x_{(1)}, ..., x_{(n)})$ avec $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ la statistique d'ordre de $(x_1, ..., x_n)$. La droite de Henry utilisée pour tester si $(x_1, ..., x_n)$ est issu d'une loi normale représente les n points $(x_{(i)}, t_{(i)})$ avec $t_{(i)} = \phi^{-1}\left(\frac{i}{n}\right)$.
2. Comment reconnait-on que $(x_1, ..., x_n)$ est issu d'une loi normale à l'aide de la droite de Henry ?
   Lorsque $(x_1, ..., x_n)$ est issu d'une loi normale, les n points $(x_{(i)}, t_{(i)})$ sont presque alignés, contrairement au cas où $(x_1, ..., x_n)$ n'est pas issu d'une loi normale.