CORRIGE PARTIEL BLANC

L3-AES

Décembre 2011

Math et Stat; appliquées

1 EXERCICE-1

1. On note respectivement μ et σ la moyenne et l'écat-type de la population. La moyenne de l'échantillon constitue une bonne estimation de μ, donc la moyenne de la population est estimée à $29.68cm$, et une bonne estimation de l'écart-type de la population est : $s_n = \sqrt{\frac{n}{n-1}}\sigma_{ech}$,donc une estimation de σ est : $s_n = \sqrt{\frac{50}{49}} * 0.06 \simeq 6.06 \times 10^{-2}$
2. On sait que la moyenne d'échantillon suit approximativement la loi normale de moyenne μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$ (approximation valable dès que $n \geq 30$). L'intervalle de confiance au seuil $1-\alpha$ est donné par : $I = \left]\overline{x}_n - t\frac{s_n}{\sqrt{n}}; \overline{x}_n + t\frac{s_n}{\sqrt{n}}\right[$ avec $2F(t) - 1 = 1 - \alpha$, Z suivant la loi $\mathcal{N}(0;1)$;à 95%, on sait que $t = 1.96$, ce qui donne : $I = \left]29.68 - 1.96\frac{6.06 \times 10^{-2}}{\sqrt{50}} \; ; \; 29.68 + 1.96\frac{6.06 \times 10^{-2}}{\sqrt{50}}\right[$: soit $I =]29.66; 29.70[$. Il y a une probabilité de 95% pour que cet intervalle contienne la longueur moyenne d'une feuille.

2 EXERCICE-2

Si l'on note X_i la variable aléatoire aléatoire versée à l'assuré i, $i \in [|1;120|]$, on a : $m = E(X_i) = 50$ et $\sigma = \sigma(X_i) = 30$. La somme totale X versée par la compagnie d'assurance est la variable aléatoire définie par : $X = \sum_{i=1}^{i=120} X_i$, peut être assimilée à une loi normale en vertue du théorème central limite, comme somme de variables aléatoires indépendantes et de même loi (iid). On a : $E(X) = 120m = 120 * 50 = 6000$ et $\sigma(X) = \sigma\sqrt{n} = 30\sqrt{120} = 60\sqrt{30} = 328.63$; la loi de X est assimilable à $\mathcal{N}(6000\,;328.63)$, et celle de Z à $\mathcal{N}(0\,;1)$, en notant $Z = \frac{X-6000}{328.63}$. On a donc : $P(X \leq 6500) = P\left(Z \leq \frac{6500-6000}{328.63}\right) = P(Z \leq 1.52) \simeq \boxed{0.9357}$.

3 EXERCICE-3

Dans le tableau de gauche, figurent les effectifs calculés (théoriques) sous l'hypothèse H_0, c'est à dire les effectifs attendus s'il y avait absence de liaison entre la nationalité et la réponse et la perception . Il suffit en fait de calculer ceux correspondant aux deux cases manquantes ; on trouve par exemple pour les français ayant répondu non (B) : $\frac{1539 * 905}{6133} \simeq 227.10$ et ensuite pour la contribution au Khi^2 : $\frac{(302-227.10)^2}{227.10} \simeq 24.70$.

C_I	A	B	C	Total
Chine	**965.41**	**221.64**	**314.95**	1502
France	**989.20**	**227.10**	**322.71**	1539
Inde	**986.62**	**226.51**	**321.87**	1535
USA	**1000.77**	**229.75**	**326.48**	1557
Total	3942	905	1286	6133

Contributions au Khi²	A	B	C	**Total**
Chine	**17.13**	**23.16**	**10.30**	**50.58**
France	**28.26**	**24.70**	**26.40**	**79.36**
Inde	**66.10**	**19.53**	**110.82**	**196.45**
USA	**46.95**	**17.41**	**72.19**	**136.55**
	Khi²			462.94

On a complété le tableau en utilisant la formule $\chi^2_{cal} = \frac{(O_1-C_1)^2}{C_1^2} + \frac{(O_2-C_2)^2}{C_2} + ... + \frac{(O_n-C_n)^2}{C_n}$, ce qui donne en ajoutant les 12 contributions au Khi^2 : $\chi^2_{cal} = 462.94$:

1. Formulation des hypothèses :

   H_0 : absence de liaison entre la nationalité et la réponse et la perception .

   H_1 : l'hypothèse alternative : il y a une liaison significative entre la nationalité et la réponse et la perception
2. Choix du seuil de signification : 5%
3. Calcul des effectifs théoriques ou calculés que nous noterons C_i (avec comme condition que chaque C_i doit être au moins égal à 5) et calcul du Khi^2. $\chi^2 = \sum \dfrac{(O_i - C_i)^2}{C_i}$; on trouve : $\chi^2_{calc} \simeq \boxed{462.94}$
4. Déterminer le ddl (degré de liberté) : $\nu = (c-1) * (l-1)$ où l et c désignent respectivement le nombre de lignes et de colonnes des données de l'échantillon, c'est à dire le nombre de modalités de chaque caractère. Ici $\nu = (4-1)(3-1) = 6$.
5. Lire dans la table $Khi^2{}_{(0.05\ ;\ 6)} \simeq \boxed{12.59}$; cette valeur est dite "critique", celà signifie que sous l'hypothèse d'indépendance, si l'on prenait tous les échantillons de taille 100 de notre population, on aurait une probabilité de 5%, d'obtenir un Khi^2 supérieur à 12.59.
6. Règle de décision :

   Si le Khi^2 calculé est inférieure ou égale au Khi^2 de la table (valeur critique : seuil limite de la région de non rejet de H_0), on ne peut rejeter H_0, par contre si $\chi^2_{calc} > \chi^2_{(2;0.05)}$, on rejettera l'hypothèse d'indépendance statistique des deux caractères.
7. Décision : ici $\chi^2_{calc} > \chi^2_{(0.05;6)}$; on conclut qu'au seuil de 5%, l'hypothèse H_0 d'abscence de liaison entre les caractères est très clairement rejetée, donc on accepte l'hypothèse alternative H_1 ; au niveau de confiance de 5%, l'écart entre les effectifs observés et les effectifs calculés est significatif, il y a un lien statistique avéré entre les nationalités et les réponses.

4 EXERCICE-4

1. La calculatrice donne :
2.

	X	Y
MOY	2,21	0,03
V	1,94	0,52
σ	1,39	0,72

3. cf tableau ci-dessus.
4. La calculatrice donne : $\begin{cases} \widehat{b} = 0.8028 \\ \widehat{a} = -0.3509 \end{cases}$ soit $\widehat{Y} = -0.3509X + 0.8028$
5. $e_8 = y_8 - \widehat{y_8} = 0.49 - (-0.3509 * 2.06 + 0.8028) = 0.4101$
6. $\widehat{a}$ donne une estimation de la variation de Y suite à une augmentation d'une unité de X : si le taux de croissance du PIB augmente de 1 point, on peut estimer que la variation du taux de chômage baisse de 0.35 points.
7. On trouve $R^2 \simeq 0.4625$, $R^2 = \dfrac{SCE}{SCT}$, il mesure le pourcentage de la variance de Y expliquée par le modèle, c'est-à dire expliquée par la variation de X.
8. Estimation ponctuelle de la variation du taux de chômage : $\widehat{Y}(2.52) = -0.3509 * 2.52 + 0.8028 = -0.08$.
9. $SCT = SCE + SCR$; donc $SCR = SCT - SCE = SCT - R^2(SCT) = SCT(1-R^2) = nV(y)(1-R^2) = 17 * 0.52 * (1 - 0.4625) = 4.75$

10. On calcule l'erreur type de a ; on a : $S^2_{\widehat{a}} = \dfrac{S^2_\varepsilon}{\sum\limits_{i=1}^{i=n} x_i^2} = \frac{SCR/(n-2)}{nV(X)} = \frac{4.75/15}{17*1.94} = 9.6 \times 10^{-3}$ ce qui donne :

$S_{\widehat{a}} = \sqrt{9.6 \times 10^{-3}} = 9.798 \times 10^{-2}$et $t_{\widehat{a}} = \dfrac{\widehat{a}}{S_{\widehat{a}}} = \frac{-0.3509}{9.798 \times 10^{-2}} = -3.58$.

$\boxed{\dfrac{\widehat{a} - a}{S_{\widehat{a}}} \text{ suit une loi de Student avec } (n-2) \text{ degrés de liberté, soit ici un ddl de } 15.}$

Nous allons donc tester si X contribue à expliquer Y en testant l'hypothèse nulle $a = 0$ contre l'hypothèse alternative $a \neq 0$:

$\begin{cases} H_0 : a = 0 \; (X \text{ n'influence pas } Y) \\ H_1 \qquad\qquad : a \neq 0 \end{cases}$. Sous l'hypothèse H_0, $t_{\widehat{a}} = \dfrac{\widehat{a}}{S_{\widehat{a}}}$, le ratio de Student, suit une distribution de Student avec $(n-2)$ degrés de liberté ; ici $t_{\widehat{a}} = -3.58$; le seuil de signification est $\alpha = 0.05$; il reste à comparer ce quotient avec la valeur lue dans la table de Student, de $t_{\alpha/2;n-2}$ soit ici : $t_{0.025\,;\,15} = 2.1315$; $|t_{\widehat{a}}| > t_{\alpha/2;;n-2}$, on rejette H_0, au seuil de $100\alpha\%$; $\widehat{a}$ est suffisamment différent de zéro pour affirmer que a est significativement différent de zéro. On en conclut que X est significative et contribue à l'explication de Y.