f3-ex1 : (1) On peut utiliser la formule de calcul de χ^2 du cours

$$\chi^2(X,Y) = N\sum_{i,j}\frac{\left(f_{x_i,y_j} - f_{x_i}\times f_{y_j}\right)^2}{f_{x_i}\times f_{y_j}}$$

qu'on peut reécrire $\chi^2 = \frac{1}{N}\sum_{i,j}\frac{\left(N\times n_{x_i,y_j} - n_{x_i}\times n_{y_j}\right)^2}{n_{x_i}\times n_{y_j}}$. On calcule $N = 63+42+110+195 = 410$, $n_{x_1} = 63+42+0+0 = 105$, $n_{x_2} = 195$, $n_{x_3} = 110$, $n_{y_1} = 63$, $n_{y_2} = 42$, $n_{y_3} = 110$, $n_{y_4} = 195$, $n_{x_1,y_1} = 63$, $n_{x_1,y_2} = 42$,... puis

$$\begin{aligned}\chi^2 &= \frac{1}{410}\frac{410\times 63 - 105\times 63}{105\times 63} + \frac{1}{410}\frac{410\times 42 - 105\times 42}{105\times 42} + \dots\\ &\overset{\text{(calculatrice)}}{=} 820\end{aligned}$$

(2) On peut aussi observer que le caractère X est déterminé par Y : $(Y = y_1) \Rightarrow (X = x_1)$, $(Y = y_2) \Rightarrow (X = x_1)$, $(Y = y_3) \Rightarrow (X = x_2)$, $(Y = y_4) \Rightarrow (X = x_3)$. Le cours affirme alors $\chi^2 = N\times\min(m-1, n-1) = 410\times\min(2,3) = 820$ (m = nbre de lignes et n = nbre de colonnes)

Commentaire : $\chi^2(X,Y)$ est maximal ce qui équivaut à ce que l'un des caractères est déterminé par l'autre, ici X est déterminé par Y.

f3-ex2 : Notons L le caractère Longévité et M le caractère marque. On calcule

(1) les fréquences des évènements "M = A" et "M = B" : $\mathrm{f}_{M=\mathrm{A}} = \frac{1}{2} = \mathrm{f}_{M=\mathrm{B}}$

(2) les moyennes de L conditionné aux valeurs de M et la moyenne de L :

$\overline{L|M=\mathrm{A}} = \frac{1}{20}(81.3 + 73.3 + 76 + \cdots) \overset{\text{(calculatrice)}}{=} 70.45$

$\overline{L|M=\mathrm{B}} = \frac{1}{20}(69.5 + 65.3 + 72 + 64 + \cdots) = 66.97$

$\overline{L} = \mathrm{f}_{M=\mathrm{A}}\times\overline{L|M=\mathrm{A}} + \mathrm{f}_{M=\mathrm{B}}\times\overline{L|M=\mathrm{B}} = 68.71$

(3) la variance de L :

$$\begin{aligned}\sigma^2(L) &= \overline{L^2} - \overline{L}^2 = \frac{1}{40}\left(69.5^2 + 65.3^2 + 72^2 + \cdots\right) - (68.71)^2\\ &\overset{\text{(calculatrice)}}{\approx} 30.8\end{aligned}$$

(4) la variance inter-groupe $\mathrm{f}_{M=\mathrm{A}}\times\left(\overline{L|M=\mathrm{A}}\right)^2 + \mathrm{f}_{M=\mathrm{B}}\times\left(\overline{L|M=\mathrm{B}}\right)^2 - \left(\overline{L}\right)^2 = 3.03$

d'où le rapport de corrélation $\eta^2(L|M) = \frac{3.03}{30.8} \approx 0.1$. Ce rapport n'est pas très loin de 0 donc n'indique pas clairement une liaison entre les caractères L et M. Pour autant les caractères L et M ne sont pas indépendants : par exemple les variances conditionnelles diffèrent nettement :

$$\begin{aligned}\sigma^2(L|M=\mathrm{A}) &= \overline{L^2|M=\mathrm{A}} - \overline{L|M=\mathrm{A}}^2\\ &= \frac{1}{20}\left(81.3^2 + 73.3^2 + 76^2 + \cdots\right) - \left(\frac{1}{20}(81.3 + 73.3 + 76 + \cdots)\right)^2\\ &\approx 47.8\end{aligned}$$

$$\sigma^2(L|M=\mathrm{B}) \approx 7.7$$

autre arguement : les boîtes à moustache de L sachant M = A et de L sachant M = B sont très différentes, cf `http://math.unice.fr/~dehon/Ens/L1aes/09-10/1-impr.pdf#page=15`

Commentaire : le rapport de corrélation permet de conclure à une liaison dans certains cas mais ne permet jamais de conclure à une indépendance.

int-stat ex.c-d-e : (voir le sujet sur la page `http://math.unice.fr/~dehon/Ens/L1aes`)

La fréquence de la filière Langues parmis les étudiants de Master vaut $\mathrm{f}_{\text{Langues}|\text{Master}} = \frac{110}{240+110+135} \approx 0.227$

La fréquence de la filière Langues parmis la population totale vaut $\mathrm{f}_{\text{Langues}} = \frac{265+110+20}{1664} \approx 0.237$

Ces fréquences s'interprêtent comme la probabilité qu'un étudiant pris au hasard étudie les langues sachant qu'il est en Master pour la première, comme la probabilité qu'un étudiant pris au hasard étudie les langues pour la seconde.

L'information "l'étudiant est inscrit en master" fait passer la probabilité d'étudier les langues de 0.237 à 0.227 (elle est pratiquement inchangée) donc multiplie la première par $\frac{0.227}{0.237} \approx 0.96 = 1 - 4\%$. Autrement dit la probabilité d'étudier les langues est diminuée de 4%. Le rapport 0.96 est proche de 1 donc montre qu'on a pratiquement l'indépendance entre les deux évènements "étudier les langues" et "être inscrit en Master".

On a $\mathrm{f}_{\text{Philosophie}|\text{Doctorat}} = \frac{42}{92+20+42} \approx 0.273$; $\mathrm{f}_{\text{Philosophie}} = \frac{160+135+42}{1664} \approx 0.203$. Cette fois le quotient $\frac{0.273}{0.203}$ vaut environ 1.34 ; autrement dit la probabilité d'étudier la philosophie est augmentée de 34% lorsqu'on apprend que l'étudiant est inscrit en doctorat. Ceci indique clairement une liaison entre les deux évènements.