Indications TD 4 « suite » :

Exercice 3 : À partir de cet exercice, nous allons manipuler des tableaux statistiques à double entrée de type général. Par conséquent, nous allons également utiliser les formules adaptées à cette situation :

$$\overline{X}_M = \frac{1}{N}\sum_{i=1}^{k} n_{i.} X_i \; ; \; \overline{Y}_M = \frac{1}{N}\sum_{j=1}^{\ell} n_{.j} Y_j .$$

$$Var_M(X) = \left[\frac{1}{N}\sum n_{i.} X_i^2\right] - \overline{X}_M^2 \; ; \; Var_M(Y) = \left[\frac{1}{N}\sum n_{.j} Y_j^2\right] - \left(\overline{Y}_M\right)^2 .$$

$$Cov(X;Y) = \left[\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{\ell} n_{ij} X_i Y_j\right] - \overline{X}_M \overline{Y}_M .$$

Pour simplifier les calculs, on utilise ce type de tableaux :

Y \ X	2	6	8	12	14	$\sum_j n_{ij} X_i Y_j$	$n_{.j}$	$n_{.j} Y_j$	$n_{.j} Y_j^2$
7	9 / 126	7 / 294	1 / 56	0 / 0	0 / 0	476	17	119	
11	2 / 44	27 / 1782	3 / 264	0 / 0	1 / 154	2244	33	363	
12	2 / 48	4 / 288	15 / 1440	4 / 576	2 / 336	2688	27	324	
15	0 / 0	1 / 90	4 / 480	17 / 3060	1 / 210	3840	23	345	
$\sum_i n_{ij} X_i Y_j$	218	2454	2240	3636	700	9248	X	X	X
$n_{i.}$	13	39	23	21	4	X	N = 100	1151	13889.
$n_{i.} X_i$	26	234	184	252	56	X	752	X	X
$n_{i.} X_i^2$	52	1404	1472	3024	784	X	6736	X	X

Remarque : Lorsque les totaux sont égaux, les résultats sont justes.

$$\overline{X}_M = \frac{752}{100} = 7,52 \quad et \quad \overline{Y}_M = \frac{1151}{100} = 11,51 .$$

$$Var_M(X) = 10,81 \quad et \quad Var_M(Y) = 6,41 .$$

$$\sigma_M(X) = 3,29 \quad et \quad \sigma_M(Y) = 2,53 .$$

$$cov(X;Y) = \frac{9248}{100} - (7,52)\times(11,51) = 5,9248 .$$

Coefficient de corrélation : $\rho = \frac{cov(X,Y)}{\sigma_n(X)\,\sigma_n(Y)} \simeq 0,71$

③ $\begin{cases} a = \frac{cov(X,Y)}{Var(X)} = 0,548. \\ b = \overline{Y} - a\overline{X} = 7,389. \end{cases}$ Ainsi, la droite d'ajustement a pour équation :

$$Y = 0,548X + 7,389.$$

④ Pour estimer, il suffit de remplacer X par 15 dans l'équation.

<u>Exercice 4</u> : On commence par préparer un tableau statistique qui permet de répondre aux questions. Le tableau se construit de la même manière qu'à l'exo 3.

① $\overline{X}_n = 4,857$; $\overline{Y}_n = 2,3077$.

$Var_n(X) = 4,408$ $\quad Var_n(Y) = 0,6746$.

$\sigma_n(X) = 2,099$ $\quad \sigma_n(Y) \simeq 0,8213$.

② Lorsque $Y = 4$, la moyenne conditionnelle de X est :

$$\overline{X}_4 = \frac{1}{n_{.4}} \sum_{i=1}^{i=4} n_{i4} X_i = \frac{34}{7} \simeq 4,857.$$

$$Var_4(X) = \frac{1}{n_{.4}} \sum n_{4.} X_i^2 - \overline{X}_4^2 \simeq 4,408.$$

$$\sigma_4(X) \simeq 2,099.$$

$$cov(X,Y) = \left[\frac{1}{N} \sum n_{ij} X_i Y_j\right] - \overline{X}_n \overline{Y}_n = 0.$$

Les deux variables sont donc indépendantes.