Université Abou berkBelkaid Tlemcen
Faculté des sciences
Département de mathématiques

2019/2020
Cours : statistiques
MI

TD N°2

Exercice N°1

Sur un échantillon de la population française, on a noté pour chaque personne, la couleur des yeux et celle des cheveux (naturelle).

Cheuveux / yeux	Noirs	bruns	blonds	roux	ni.
Marrons	152	247	83	11	493
Verts-gris	?	114	37	?	?
Bleus	36	102	?	10	?
n.j	261		247	29	N=1000

1) Quel est l'effectif des personnes ayant des cheveux noirs et des yeux verts-gris.
2) Quel est l'effectif des personnes ayant des cheveux bruns
3) Combien de personnes ont des cheveux roux et des yeux verts-gris.
4) Combien de personnes ont les yeux verts-gris
5) Finir le tableau avec les effectifs manquants
6) Calculerf_{12}, f_{24}, $f_{.3}$, $f_{3.}$ que représente ces fréquences
7) Les variables X : les yeux et Y : les cheveux sont elles indépendantes ??

Exercice N°2

Le mur d'une habitation est constitue par une paroi en béton et une couche de polystyrène d'épaisseur variable x (en cm).
On a mesuré, pour une même épaisseur de béton, la résistance thermique y (en m^2 / watt) de ce mur pour différentes valeurs de x. On a obtenu les résultats suivant :

Epaisseur xi	2	4	6	8	10	12	15	20
Résistance yi	0.83	1.34	1.63	2.29	2.44	2.93	4.06	4.48

a) Déterminer une équation de la droite de régression de cette série de y en fonction de x.
b) représenter cette droite avec le nuage de points.
c) Calculer le coefficient de corrélation, Que peut-on en déduire ?
d) Quelle résistance thermique peut-on espérer obtenir avec une couche de polystyrène de 18 centimètres d'épaisseur ?

Exercice N°3

Sur un échantillon de vingt individus x, appartenant à une même tranche d'âge, on a étudié les caractères taille **ti** en mètres et pointure des chaussures **pi**
Les résultats obtenus sont les suivants :

$$\sum_{i=1}^{20} t_i = 34.28, \quad \sum_{i=1}^{20} p_i = 848, \quad \sum_{i=1}^{20} t_i * p_i = 1445.18$$

$$\sum_{i=1}^{20} t_i^2 = 58.8614, \quad \sum_{i=1}^{20} p_i^2 = 35996$$

Dans ce qui suit, tous les résultats numériques seront donnés à 10^{-2} prés

1. Calculer le coefficient de corrélation linéaire de la série statistique en les variables **t** et **p.** Que peut-on en déduire ?
2. Déterminer par la méthode des moindres carres la droite de régression de p en t permettant d'estimer la pointure d'un individu en fonction de sa taille.
3. Quelle pointure peut-on estimer pour un individu mesurant 1.83 mètres ?

Exercice N°4

Dans un village où on a N petites fermes, on compte le nombre de poule par petite ferme (X) et le nombre d'œufs récolté (Y), on a les résultats suivants :

X \ Y	[2,10[	[10, 18[	[18, 26[	[26, 34[
[2,6[	0	2	2	5
[6,10[	6	5	0	0
[10,14[	13	12	15	0

1. Déterminer « N » le nombre de fermes où on a fait l'étude.
2. Donner le tableau marginal de X et de Y
3. Calculer le nombre de poules moyen par ferme, et le nombre d'œufs moyen par ferme.
4. X et Y sont-elles des variables indépendantes ?
5. Calculer la covariance.

« On n'est pas toute sa vie à l'école, vient l'âge ou il faut mettre en pratique »
« Le travail paie dans le futur, la paresse elle paie comptant »
« Vise toujours la réussite et oublie le succès »

2019/2020

Corrigé du TD N°2

MI

Exo1

1) On cherche $n_{21} = ??$

$n_{\cdot j} = \sum_{i=1}^{3} n_{ij} \Rightarrow n_{\cdot 1} = \sum_{i=1}^{3} n_{i1}$
$= n_{11} + n_{21} + n_{31}$

donc $261 = 152 + n_{21} + 36 \Rightarrow \boxed{n_{21} = 73}$

2) On cherche $n_{\cdot 2} = ??$

1ère méthode : $n_{\cdot j} = \sum_{i=1}^{3} n_{ij} \Rightarrow n_{\cdot 2} = \sum_{i=1}^{3} n_{i2}$

$\Rightarrow n_{\cdot 2} = n_{12} + n_{22} + n_{32} = 247 + 114 + 102 = \boxed{463}$

2ème méthode $N = \sum_{j=1}^{4} n_{\cdot j} = N = n_{\cdot 1} + n_{\cdot 2} + n_{\cdot 3} + n_{\cdot 4}$

$\Rightarrow n_{\cdot 2} = 1000 - (261 + 247 + 29) = \boxed{463}$

3) On cherche $n_{24} = ??$

$n_{\cdot j} = \sum_{i=1}^{3} n_{ij} \Rightarrow n_{\cdot 4} = n_{14} + n_{24} + n_{34}$

$\Rightarrow n_{24} = 29 - (11 + 10) = \boxed{8}$

4) On cherche $n_{2\cdot} = ??$

$n_{i\cdot} = \sum_{j=1}^{4} n_{ij} \Rightarrow n_{2\cdot} = \sum_{j=1}^{4} n_{2j}$
$= n_{21} + n_{22} + n_{23} + n_{24}$
$= 73 + 114 + 37 + 8 = \boxed{232}$

5) On cherche $n_{3\cdot} = ??$ et le $n_{33} = ??$

$N = \sum_{i=1}^{3} n_{i\cdot} \Rightarrow N = n_{1\cdot} + n_{2\cdot} + n_{3\cdot}$

$\Rightarrow \boxed{n_{3\cdot} = N - n_{1\cdot} - n_{2\cdot}}$

$\Rightarrow n_{3\cdot} = 1000 - 493 - 232 = \boxed{275}$

On a : $n_{i\cdot} = \sum_{j=1}^{4} n_{ij} \Rightarrow n_{3\cdot} = \sum_{j=1}^{4} n_{3j}$

$$\Rightarrow n_{33} = n_{3\cdot} - (n_{31} + n_{32} + n_{34})$$
$$= 275 - (36 + 102 + 10) = \boxed{127}$$

ou bien

$n_{\cdot j} = \sum_{i=1}^{3} n_{ij} \Rightarrow n_{\cdot 3} = \sum_{i=1}^{3} n_{i3}$

$$\Rightarrow n_{33} = n_{\cdot 3} - (n_{13} + n_{23})$$
$$= 247 - (83 - 37) = \boxed{127}$$

6) $\boxed{f_{ij} = \frac{n_{ij}}{N}}$ $\boxed{f_{i\cdot} = \frac{n_{i\cdot}}{N}}$ $\boxed{f_{\cdot j} = \frac{n_{\cdot j}}{N}}$

$f_{12} = \frac{n_{12}}{N} = \frac{247}{1000} = 0,247$ la proportion des personnes ayant cheveux Bruns et yeux marrons

$f_{24} = \frac{n_{24}}{N} = \frac{8}{1000} = 0,008$ la proportion des personnes ayant cheveux roux et yeux verts-gris.

$f_{3\cdot} = \frac{n_{3\cdot}}{N} = \frac{275}{1000} = 0,275$ la proportion des personnes ayant yeux Bleus.

$f_{\cdot 3} = \frac{n_{\cdot 3}}{N} = \frac{247}{1000} = 0,247$ la proportion des personnes ayant des cheveux Blonds.

7) $\boxed{X \text{ et } Y \text{ indep} \Leftrightarrow N \times n_{ij} = n_{i\cdot} \times n_{\cdot j} \quad \forall i,j}$

On Pose $i = 1$ $j = 4$

$$\left.\begin{array}{l} N\, n_{14} = 1000 \times 11 = 11\,000 \\ n_{1\cdot} \times n_{\cdot 4} = 493 \times 29 = 14297 \end{array}\right\} \Rightarrow N \times n_{14} \neq n_{1\cdot} \times n_{\cdot 4}$$

donc $\exists i,j$ tq $N \times n_{ij} \neq n_{i\cdot} \times n_{\cdot j}$

cl : X et Y ne sont pas independants.

Exo2 :

x_i	2	4	6	8	10	12	15	20	Σ 77
y_i	0,83	1,34	1,63	2,29	2,44	2,93	4,06	4,48	20
$x_i \times y_i$	1,66	5,36	9,78	18,32	24,4	35,16	60,9	89,6	245,18
x_i^2	4	16	36	64	100	144	225	400	989
y_i^2	0,6889	1,7956	2,6569	5,2441	5,9536	8,5849	16,4836	20,0704	61,478

1) Droite de régression Δ : $y = ax + b$

avec $a = \frac{\text{cov}(x,y)}{V(x)}$ $\quad b = \bar{y} - a\bar{x}$

$\bar{X} = \frac{1}{N} \sum_i x_i = \frac{77}{8} = \boxed{9,625}$

$\bar{Y} = \frac{1}{N} \sum y_i = \frac{20}{8} = \boxed{2,5}$

$\overline{X^2} = \frac{1}{N} \sum x_i^2 = \frac{989}{8} = 123,625$

$\overline{Y^2} = \frac{1}{N} \sum_i y_i^2 = \frac{61,478}{8} = 7,6847$

$\overline{XY} = \frac{1}{N} \sum_i x_i y_i = \frac{245,18}{8} = 30,6475$

$\text{cov}(X,Y) = \overline{XY} - \bar{X} \cdot \bar{Y} = 30,6475 - (2,5)(9,625)$

$\text{cov}(X,Y) = \boxed{6,585}$

$V(X) = \overline{X^2} - (\bar{X})^2 = 123,625 - (9,625)^2$

$V(X) = 30,984 \longrightarrow \sigma_X = \sqrt{V(X)} = 5,566$

$V(Y) = \overline{Y^2} - (\bar{Y})^2 = 7,6847 - (2,5)^2$

$V(Y) = 1,4347 \longrightarrow \sigma_Y = \sqrt{V(Y)} = 1,197$

Conclusion

$$a = \frac{Cov(X,Y)}{V(X)} = \frac{6,585}{30,984} = 0,212$$

$$b = \overline{Y} - a\overline{X} = 2,5 - (0,212)(9,625) = 0,459$$

$$\boxed{(D) : y = 0,212\, x + 0,459}$$

2) coefficient de correlation $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \ \sigma_Y}$

donc $\rho_{XY} = \frac{6,585}{5,566 \times 1,197} = 0,988$

$|\rho_{XY}| = 0,988 > 0,85 \Rightarrow$ la relation linéaire est acceptée

3) Si $x = 18$ cm $\Rightarrow$

$y = 0,212\,(18) + 0,459 = 4,275$ m²/Watt

Exo3 :

1) $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \ \sigma_Y} = \frac{Cov(t,P)}{\sigma_t \ \sigma_P}$

$$\overline{t} = \frac{1}{N}\sum_{i=1}^{20} t_i = \frac{34,28}{20} = 1,714$$

$$\overline{P} = \frac{1}{N}\sum_i P_i = \frac{848}{20} = 42,4$$

$$\overline{t^2} = \frac{1}{N}\sum_i t_i^2 = \frac{58,8614}{20} = 2,943$$

$$\overline{P^2} = \frac{1}{N}\sum_i P_i^2 = \frac{35996}{20} = 1799,8$$

$$\overline{tP} = \frac{1}{N}\sum_i t_i \times P_i = \frac{1445,18}{20} = 72,259$$

$cov(t,P) = \overline{tP} - \overline{t}.\overline{P} = 72,259 - (1,714)(42,4)$

$= -0,4146 = -0,414$

$V(t) = \overline{t^2} - (\overline{t})^2 = 2,943 - (1,714)^2 = 0,005$

$V(P) = \overline{P^2} - (\overline{P})^2 = 1799,8 - (42,4)^2 = 2,04$

$$\rho_{tP} = \frac{cov(t,P)}{\sigma_t \, \sigma_P} = \frac{-0,414}{\sqrt{0,005} \times \sqrt{2,04}} = \frac{-0,414}{0,1009}$$

$$\boxed{\rho_{tP} = -0,41}$$

$|\rho_{tP}| = 0,41 < 0,85 \Rightarrow$ la droite obtenue par régression est non acceptée

2) la Droite (Δ): $p = at + b$

avec $a = \dfrac{cov(P,t)}{V(t)}$ $\qquad b = \overline{P} - a\overline{t}$

$$a = \frac{cov(P,t)}{V(t)} = \frac{-0,414}{0,005} = -82,8$$

$b = \overline{P} - a\overline{t} = 42,4 - (-82,8)(1,714) = 184,32$

$$\boxed{(\Delta) : p = -82,8\,t + 184,32}$$

3) On ne peut pas estimer, vu que la droite de régression est refusée (non acceptée)

Rmq Techniquement le calcul peut se faire, mais il sera faux (ie le résultat ne peux pas être pris en considération)

Exo4:

1) $N = \sum_i \sum_j n_{ij} = 60$

2) $\overline{X} = \frac{1}{N} \sum_i n_{i\cdot} \times x_i = \frac{1}{60}(9\times 4 + 11\times 8 + 40\times 12) = 10,06$

$\overline{Y} = \frac{1}{N} \sum_j n_{\cdot j}\, y_j = \frac{1}{60}(6\times 19 + 14 \times 19 + 22\times 17 + 5\times 30)$

$= 15,06$

3) $\boxed{X \text{ et } Y \text{ indep} \iff \forall i,j \quad N\times n_{ij} = n_{i\cdot} \times n_{\cdot j}}$

On Pose $i = 3$, $j = 4$

$$\left.\begin{aligned} N\times n_{ij} &= 60 \times n_{34} = 60\times 0 = 0 \\ n_{ij}\times n_{\cdot j} &= n_{3\cdot} \times n_{\cdot 4} = 40\times 5 = 200 \end{aligned}\right\} \quad 0 \neq 200$$

$\Rightarrow \exists i,j$ tq $N n_{ij} \neq n_{i\cdot} \times n_{\cdot j}$

donc X et Y ne sont pas indep

5) $cov(X,Y) = \overline{XY} - \overline{X}.\overline{Y}$

$\overline{XY} = \frac{1}{N} \sum_i \sum_j n_{ij}\, x_i\, y_j$

$$= \frac{1}{N}\left[\begin{aligned} & n_{11}\, x_1\, y_1 + n_{12}\, x_1 y_2 + n_{13}\, x_1\, y_3 + n_{14} x_1 y_4 \\ & + \vdots \\ & + n_{31}\, x_3\, y_1 + n_{32}\, x_3 y_2 + n_{33}\, x_3 y_3 + n_{34}\, x_3 y_4 \end{aligned}\right]$$

$$= \frac{1}{60}\left(0 + 112 + 176 + 60 + 288 + 560 + 0 + 0 + 936 + 2016 + 3960 + 0\right)$$

$$= \frac{8108}{60} = 135,133$$

cl : $cov(X,Y) = -16,37$