SERIE DE TD N° 2 ***en BIOSTATISTIQUE 2019/2020***

Exercices sur les Séries statistiques doubles

et ajustements linéaires

Exercice n° 01 :

Montrer la formule de König–Huyghens (vue dans le cours, diapo 4, cours 3) suivante :

$$\sigma^2 = \frac{\sum_{i=1}^{p} n_i x_i^2}{\sum_{i=1}^{p} n_i} - \bar{X}^2$$

Solution :

$$\sigma^2 = \frac{\sum_{i=1}^{p} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{p} n_i} = \frac{\sum_{i=1}^{p} n_i (x_i^2 - 2 x_i * \bar{x} + \bar{x}^2)}{\sum_{i=1}^{p} n_i} = \frac{\sum_{i=1}^{p} n_i x_i^2}{\sum_{i=1}^{p} n_i} - 2\bar{X} \frac{\sum n_i x_i}{\sum n_i} + \bar{X}^2 \frac{\sum n_i}{\sum n_i}$$

$$\sigma^2 = \frac{\sum_{i=1}^{p} n_i x_i^2}{\sum_{i=1}^{p} n_i} - 2\bar{X} * \bar{X} + \bar{X}^2 = \frac{\sum_{i=1}^{p} n_i x_i^2}{\sum_{i=1}^{p} n_i} - \bar{X}^2$$

Exercice n° 02 :

Dans une petite entreprise de 20 employés on a collecté les données concernant le sexe, l'absentéisme et l'âge, dans le tableau suivant, où X représente le sexe Mâle (M) ou Femelle (F), Y représente le nombre de jours d'absentéisme en un mois et Z l'âge en années.

Dans ce tableau, on a utilisé les conventions suivantes :

Employé(e)	Sexe X		Absentéisme Y				Âge Z			
	M	F	0	1	2	3	[20;30[	[30;40[	[40;50[	[50;60[
1	X		X				X			
2		X		X			X			
3	X				X			X		
4	X		X					X		
5		X		X					X	
6	X			X					X	
7	X		X					X		
8	X		X				X			
9		X		X				X		
10		X				X				X
11	X		X					X		
12	X		X					X		
13	X			X				X		
14		X	X				X			
15	X				X		X			
16		X		X				X		
17	X		X						X	
18	X		X							X
19		X			X		X			
20	X		X					X		

1) Construire le tableau de contingence des couples (X ; Y) et (X ; Z)
2) Construire les distributions marginales de X, Y et Z.
3) Calculer la moyenne et la variance de Z.
4) Calculer Cov (Y;Z)

Solution :

1) Les tableaux de contingence des couples (X ; Y), (X ; Z) et (Y ; Z) sont les suivants :

X \ Y	0	1	2	3	Σ
M	9	2	2	0	13
F	1	4	1	1	7
Σ	10	6	3	1	20

X \ Z	[20;30[	[30;40[	[40;50[	[50;60[	Σ
M	3	7	2	1	13
F	3	2	1	1	7
Σ	6	9	3	2	20

2) Les distributions marginales sont les suivantes :
   Variable X : {(M ; 13) , (F ; 7)}
   Variable Y : {(0 ; 10) , (1 ; 6) , (2 ; 3) ,(3 ; 1)}
   Variable Z : {([20 ;30[; 6) , ([30 ;40[; 9) , ([4 ;50[; 3) , ([50 ;60[; 2)}
3) La moyenne de Z est $\bar{z}$ = 35.5 Variance de Z est σ_z^2 = 84.75

Y \ Z	[20;30[	[30;40	[40;50[	[50;60[	Σ
0	3	5	1	1	10
1	1	3	2	0	6
2	2	1	0	0	3
3	0	0	0	1	1
Σ	6	9	3	2	20

4) Covariance (Y ; Z) = 1.125

Exercice n° 03 :

Dans les années 1950, une fuite de déchets radioactifs issue d'une zone de stockage près d'une ville en Amérique du nord, s'est répandue dans une rivière.

On a calculé pour 9 régions situées en aval un indice d'exposition (fonction de la distance à la ville) d'une part et la mortalité par cancer (nombre de décès annuels par cancer) d'autre part ; les données sont les suivantes :

Région	1	2	3	4	5	6	7	8	9
Indice d'exposition x_i	8.3	5.2	3.0	2.6	11.5	1.5	1.75	1.5	2.0
Mortalité cancer y_i	210	170	150	160	220	155	150	149	130

1) Dans cette étude quelle est la variable indépendante et la variable dépendante ?
2) Dessiner le nuage de points de cette série double.
3) Calculer les moyennes et les écart-types de variables X et Y.

4) Calculer la Cov (X ; Y) et le coefficient de corrélation linéaire. Interpréter le résultat.
5) Trouver les droites de régression Dy(x) et Dx(y).
6) Estimer la mortalité lorsque x = 5 et x = 0.

Solution :

1) La variable indépendante est x (indice d'exposition) et la variable dépendante est y (mortalité par cancer).
2) Le nuage de points est ci-après :
3) $\bar{x} = 4.15$ et $\bar{y} = 166$
   $\sigma_x = 3.341656276$ et $\sigma_y = 28.13064758$
4) Covariance (X ; Y) = 89.044444….4
   Coefficient de corrélation linéaire r = 0.947251545
5) Dy(x) : y = A + B * x où : A = 132.907 et B = 7.974
   Dx(y) : x = A' + B'*y où : A'= – 14.529 et B'=0.11252457
6) $\widehat{y(5)} = 172.778$ et $\widehat{y(0)} = 132.907$

Exercice n° 04 :

Dans le tableau de contingence suivant sont représentées les données d'une étude concernant l'influence de la durée d'exposition au soleil d'une feuille sur le nombre de stomates aérifères. X représente le nombre de jours d'exposition au soleil et Y le nombre de stomates aérifères au millimètre carré.

X jrs \ Ysto	6	15	39	62	85	Total
2	26	11	3	0	0	40
6	16	14	4	4	0	38
8	5	10	12	8	1	36
24	3	0	12	15	16	46
52	0	0	2	18	20	40
Total	50	35	33	45	37	200

1) Quelle est le nombre de feuilles (taille de l'échantillon) concernées par cette étude statistique? Rép : 200
2) Quel est le nombre de feuilles exposées au soleil 2 jours et le nombre de celles exposées 24 jours ? Rép : 40 Rép : 46
3) Quel est le nombre de feuilles ayant eu 15 stomates aérifères et celui des feuilles avec moins de 62 stomates aérifères ? Rép : 35 Rép : 118
4) Dans cette étude statistique, déterminer la variable statistique indépendante et la variable statistique dépendante. X : indépendante Y : dépendante
5) Représenter la distribution marginale de X et celle de Y. dernière colonne pour X et dernière ligne pour Y.
6) En utilisant la calculatrice, il faut calculer :

a) Les moyennes $\bar{x}$ et $\bar{y}$ et les variances de X et de Y
b) La Covariance (X ; Y).

7) Trouver la distribution conditionnelle des fréquences de Y pour x = 2.
8) Trouver la distribution conditionnelle des fréquences de Y pour x = 24.
9) Trouver la distribution conditionnelle des fréquences de Y pour x = 52.
10) Parmi les feuilles exposées au soleil deux jours, quel est le pourcentage de celles avec au moins 39 stomates aérifères ?
11) Parmi les feuilles exposées au soleil 24 jours, quel est le pourcentage de celles avec au moins 39 stomates aérifères ?
12) Parmi les feuilles exposées au soleil 52 jours, quel est le pourcentage de celles avec au moins 39 stomates aérifères ?
13) Est-il légitime de penser que l'exposition au soleil influence le nombre de stomates aérifères ? Si oui dans quel sens ?

Solution :

6) $\bar{x} = 18.9$ $\bar{y} = 40.235$ $\sigma_x = 18.31$ $\sigma_y = 29.7$ Cov (X ;Y) = 399.96

X \ Y	6	15	39	62	85	Total
7) 2	26 65%	11 27.5%	3 7.5%	0	0	40 100%
6	16	14	4	4	0	38
8	5	10	12	8	1	36
8) 24	3 6.5%	0	12 26.1%	15 32.6%	16 34.8%	46 100%
9) 52	0	0	2 5%	18 45%	20 50%	40 100%
Total	50	35	33	45	37	200

10) 7.5 % 11) 93.5 % 12) 100 % 13) Oui : on remarque que les 2 variables augmentent dans le même sens.

Exercice n° 05 :

On a administré une drogue à 40 rats à différentes doses x et on a mesuré la performance y de ces rats sous l'influence de la drogue dans une certaine tâche et on a obtenu les résultats représentés par le tableau suivant :

x_i	1	1	2	2	3	3	4	4
y_i	8	24	8	12	12	24	12	24
n_i	9	1	3	7	3	7	2	8

On fait le changement de variable Z = $\frac{2\,Log\,x}{Log\,2} - 3$; utiliser une calculatrice pour :

1) Construire le tableau représentant la nouvelle série double (Z ; Y)
2) Calculer les moyennes et les variances de variables Z et Y.
3) Calculer la Cov (Z ; Y).
4) Trouver la droite de régression de Y en Z.
5) Utiliser cette régression pour estimer la performance y pour la dose x = 5.

Solution :

1) Le Nouveau tableau est le suivant :

x_i	1	1	2	2	3	3	4	4
z_i	–3	–3	–1	–1	0.169925001	0.169925001	1	1
y_i	8	24	8	12	12	24	12	24
n_i	9	1	3	7	3	7	2	8

2) $\bar{z} = -0.707518749$ $\bar{y} = 15.6$ $\sigma_z = 1.50221032$ $\sigma_z^2 = 2.256635845$
$\sigma_y = 7.031358332$ $\sigma_y^2 = 49.44$

3) Cov (z ; y) = 7.403910001

4) Dy(z) = A + B∗z où : A = 17.92133384 B = 3.280950277 r = 0.70095665

5) $y(\widehat{x = 5}) = y(z = 1.\widehat{643}85619) = 23.31474426$

Exercice n° 06 :

1) Montrer que la droite de régression linéaire de y en x, obtenue par la méthode des moindres carrés, passe par le point ($\bar{x}$; $\bar{y}$).

2) Montrer que la droite de régression linéaire de x en y, obtenue par la méthode des moindres carrés, passe par le point ($\bar{y}$; $\bar{x}$).

Solution :

1) En remplaçant dans l'équation y = a ∗ x + b de régression de y en x, l'inconnue x par $\bar{x}$, on a : $\frac{Cov\,(x\,;y)}{\sigma_x^2}\ \bar{x} + (\bar{y} - \frac{Cov\,(x\,;y)}{\sigma_x^2}\,\bar{x}) = \bar{y} \Rightarrow \bar{y} = \bar{y}$

2) En remplaçant dans l'équation x = a' ∗ y + b' de régression de x en y, l'inconnue y par $\bar{y}$, on a : $\frac{Cov\,(y\,;x)}{\sigma_y^2}\ \bar{y} + (\bar{x} - \frac{Cov\,(y\,;x)}{\sigma_y^2}\,\bar{y}) = \bar{x} \Rightarrow \bar{x} = \bar{x}$