

Feuille de TD 4 – Statistique descriptive bivariée

Exercice 1

Les résultats d'une étude statistique sur l'âge et le poids de 142 enfants sont les suivants :

Poids X (Kg)	Age Y (années)		
	3-4	4-5	5-6
10-15	19	7	1
15-20	32	21	12
20-25	3	18	28
25-30	0	0	1

1. Déterminer les distributions et les fréquences marginales correspondantes.
2. Calculer la moyenne et la variance marginale de X en termes d'effectif et de fréquence marginale.
3. Calculer la moyennes et la variance marginale de Y en termes d'effectif et de fréquence marginale.
4. En déduire l'écart-type marginal de X et de Y .

Exercice 2

La répartition de 290 entreprises selon le Chiffre d'affaires (Y) et les dépenses publicitaires (X) au 1^{er} janvier 2011 est donnée par le tableau ci-dessous :

$X \backslash Y$	[15,25[[25,35[[35,45[[45,55[[55,65[[65,75[[75,85[Total
[1,3[6	8	2	0	0	0	0	16
[3-5[8	10	14	12	0	0	0	44
[5-8[0	4	18	20	18	2	0	62
[8-9[0	4	12	16	20	12	4	68
[9-11[0	0	4	8	16	16	8	52
11 et plus	0	0	0	0	14	16	18	48
Total	14	26	50	56	68	46	30	290

1. Calculer les fréquences conditionnelles de Y selon X .
2. Calculer les fréquences conditionnelles de X selon Y .
3. Calculer les moyennes et les variances marginales.
4. Calculer les moyennes et variances conditionnelles de X selon Y .

Exercice 3

Une étudiante en sociologie veut analyser, dans le cadre d'un projet de fin de session, s'il existe une relation linéaire entre la densité de population dans les régions métropolitaines et le taux de criminalité correspondant dans ces régions. Le taux de criminalité (Y) est indiqué en nombre de crimes par 10000 habitants et la densité de population (X) est mesurée en milliers d'habitants par km^2 .

Region	1	2	3	4	5	6	7	8	9	10	11	12
x_i	7,7	5,8	11,5	2,1	3,7	3,6	7,5	4,2	3,8	10,3	8,6	7,2
y_i	12	9	15	4	4	2	10	3	5	11	10	11

$$\sum x_i = 76; \quad \sum y_i = 96$$

$$\sum x_i^2 = 576,46; \quad \sum y_i^2 = 962; \quad \sum x_i y_i = 732,6$$

1. Définir la population et l'échantillon. Identifier la variable à expliquer et la variable variable explicative. Pour chaque variable calculer la moyenne observée et la variance observée. Donner également la valeur de la covariance et du coefficient de corrélation. Commenter.
2. Faire un graphique de Y en fonction de X . Que peut-on en conclure ? Une modélisation linéaire est-elle pertinente ?
3. Écrire l'équation de la droite de régression obtenue en utilisant les valeurs estimées des coefficients β_0 et β_1 (notés $\hat{\beta}_0$ et $\hat{\beta}_1$ respectivement).
4. Estimer le taux de criminalité le plus plausible pour une densité de population de 7500 habitants par km^2 .

Exercice 4

On a relevé pour différents pays le PIB par habitant en 2004 X (en dollars) et le taux brut de scolarisation des moins de 24 ans la même année Y (en pourcentage). Les résultats

sont les suivants

Pays	PIB X	Taux de scolarisation Y
Pays en développement	4775	63
Pays les plus pauvres	1350	45
Pays arabes	5680	62
Asie de l'Est et Pacifique	5872	69
Amérique latine et Caraïbes	7964	81
Asie du Sud	3072	56
Afrique Sub-saharienne	1942	50
Europe centrale, orientale et CEI	8802	83

$$\sum x_i = 39457; \quad \sum y_i = 509$$
$$\sum x_i^2 = 245474957; \quad \sum y_i^2 = 33685; \quad \sum x_i y_i = 2763685$$

1. Définir la population et l'échantillon étudiée. On cherche à expliquer le taux de scolarisation en fonction du PIB. Identifier la variable à expliquer et la variable explicative. Pour chaque variable calculer la moyenne observée et la variance observée.
2. Expliquer l'objectif de la régression linéaire simple et préciser ses conditions d'application. Donner l'équation du modèle théorique.
3. Donner l'équation de la droite avec les valeurs estimées des coefficients inconnus β_0 et β_1 .

Pr. Toufik Chaayra

t.chaayra@gmail.com