

## Exercices de Biostatistique

### Fiche 3 : Distribution bivariée, régression linéaire

#### Exercice 1

Le tableau ci-dessous donne le classement de 100 individus suivant les deux caractères masse corporelle et taille

X / Y	[40,45[	[45,50[	[50,55[	[55,60[
[150,155[	20	9	1	0
[155,160[	2	18	4	1
[160,165[	0	5	12	6
[165,170[	0	1	7	14

- 1) En utilisant les centres de classes, dessiner le nuage des points.
- 2) Déterminer les lois marginales et déterminer les moyennes et variances de ces lois.
- 3) Déterminer les lois conditionnelles ainsi que leurs moyennes.
- 4) Calculer les rapports de corrélation.
- 5) Calculer le coefficient de corrélation linéaire. Commenter.

#### Exercice 2

Les performances réalisées par 10 coureurs à pied sur un semi-marathon et un marathon (les temps sont donnés en minutes) sont données dans le tableau suivant :

Numéro i du coureur	1	2	3	4	5	6	7	8	9	10
Temps $X_i$ au semi-marathon	68	70	76	80	90	96	104	110	125	138
Temps $Y_i$ au marathon	145	150	170	185	200	220	250	285	320	345

1. Représenter le nuage des points  $(X_i, Y_i)$  dans un repère orthogonal.
2. Calculer la moyenne, la variance et l'écart-type de la série des temps  $(X_i)$  réalisés au semi-marathon, puis de la série des temps  $(Y_i)$  réalisés au marathon
3. Calculer la covariance de  $X$  et  $Y$  et le coefficient de corrélation linéaire
4. Déterminer l'équation de la droite de régression de  $Y$  sur  $X$  par la méthode des moindres carrés. Représenter cette droite sur le graphique de la question a).
5. Estimer le temps mis sur un marathon par un coureur ayant réalisé 1h 56 min au semi-marathon.

#### Exercice 3

Considérons l'étude suivante sur l'effet d'un facteur de risque (qui peut être le tabac) sur la survenue d'une maladie (le rhume). En hiver, une population d'environ 1000 sujets est tirée au sort parmi la population générale. On dénombre alors les fumeurs et les enrhumés. Les résultats sont résumés dans le tableau suivant :

Calculer  $RR$  et  $OR$ .

Soit maintenant l'enquête cas-témoins réalisée à partir d'un échantillon de 60 enrhumés et de 90 non-enrhumés tirés au sort de la population générale. En respectant des

	enrhumé : oui	enrhumé : non
fumeur : oui	30	300
fumeur : non	30	600

proportions de fumeurs et des non-fumeurs chez les enrhumés et les non-enrhumés identiques au tableau précédent, nous obtenons le tableau suivant :

	enrhumé :oui	enrhumé : non
fumeur : oui	30	30
fumeur : non	30	60

Calculer de nouveau le  $RR$  et l' $OR$ . Quelle conclusion vous pouvez en tirer.

#### **Exercice 4**

Quinze élèves, désignés par les lettres de A à O ont été classés une première fois par une épreuve de français, une seconde fois par une épreuve de mathématiques. Calculer le coefficient décrivant la corrélation entre ces deux classements.

Elèves	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Fran.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Math.	9	3	1	11	2	5	8	13	4	10	7	14	15	6	12

Pensez vous que les élèves ont le même niveau en mathématique qu'en français ?

#### **Exercice 5**

On demande à deux experts de donner leurs avis sur un groupe de 300 candidats pour des postes de gardiens de la paix. Les résultats d'un test psychotechnique sont regroupés dans le tableau suivant :

évaluateur 1 \ évaluateur 2	pas de problème	légers problèmes	beaucoup de problèmes	<b>total</b>
pas de problème	150	20	30	200
legers problèmes	10	30	20	50
beaucoup de problèmes	0	10	30	40
<b>total</b>	160	60	80	300

Peut-on dire que le jugement des experts est concordant ? (utiliser un indice simple)