

T. D. n° 4

Régression linéaire simple

Exercice 1. D'après Baillargeon, Probabilités, Statistiques et techniques de régression, Les éditions SMG, 1995.

Nous donnons les couples d'observations suivants :

x_i	18	7	14	31	21	5	11	16	26	29
y_i	55	17	36	85	62	18	33	41	63	87

1. Importer les données.
2. Tracer le diagramme de dispersion des couples $(x_i; y_i)$. À la vue de ce diagramme, pouvons-nous soupçonner une liaison linéaire entre ces deux variables ?
3. Déterminer pour ces observations la droite des moindres carrés, c'est-à-dire donner les coefficients de cette droite.
4. Donner les ordonnées des y_i calculés par la droite des moindres carrés correspondant aux différentes valeurs des x_i .
5. Tracer ensuite la droite sur le même graphique.
6. Quelle est une estimation plausible de Y à $x_i = 21$?
7. Quel est l'écart entre la valeur observée de Y à $x_i = 21$ et la valeur estimée avec la droite des moindres carrés ? Comment appelons-nous cet écart ?
8. Est-ce que la droite des moindres carrés obtenue à la question 3 passe par le point $(\bar{x}; \bar{y})$? Pouvons-nous généraliser cette conclusion à n'importe laquelle droite de régression ?

Exercice 2. D'après Baillargeon, Probabilités, Statistiques et techniques de régression, Les éditions SMG, 1995.

La société de Transport Bertrand veut établir une politique d'entretien des camions de sa flotte. Tous sont de même modèle et utilisés à des transports semblables. La direction de la société est d'avis qu'une liaison statistique entre le coût direct de déplacements (*cents* par *km*) et l'espace de temps écoulé depuis la dernière inspection de ce camion serait utile. Nous avons donc recueilli un certain nombre de données sur ces deux variables. Nous souhaitons utiliser la régression linéaire comme modélisation statistique.

Coût direct	10	18	24	22	27	13	10	24	25	8	16
Nombre de mois	3	7	10	9	11	6	5	8	7	4	6
Coût direct	20	28	22	19	18	26	14	20	26	30	12
Nombre de mois	9	12	8	10	9	11	6	8	10	12	5

1. Quelle variable devrions-nous identifier variable dépendante (Y) et laquelle devrions-nous identifier variable explicative (X) ?

2. Tracer le diagramme de dispersion de ces observations. Est-ce que le nuage de points suggère une forme de liaison particulière ?
3. Calculer l'équation de la droite des moindres carrés.
4. Avec l'équation de la droite des moindres carrés, quelle est l'estimation la plus plausible du coût direct de déplacement pour des camions dont la dernière inspection remonte à 6 mois ?
5. D'après les résultats de cette étude, un délai supplémentaire d'un mois pour l'inspection d'un camion occasionnera-t-il une augmentation ou une diminution du coût direct ? Quelle sera vraisemblablement la valeur de cette variation de coût ?
6. L'équation de la droite des moindres carrés pour les données de la société est : $\hat{y}_i = 1,54941 + 2,26087 x_i$. Calculer la variation qui est expliquée par la droite des moindres carrés.
7. Calculer la variation résiduelle.
8. En déduire la variation totale dans le coût direct de déplacement.
9. Calculer le coefficient R^2 et interpréter le résultat.

Exercice 3. Le taux de criminalité.

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 10 000 habitants et la densité de population (X) est mesurée en milliers d'habitants par km^2 .

Région	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

1. Tracer le diagramme de dispersion de ces observations.
2. Calculer les coefficients de la droite des moindres carrés.
3. À quelle augmentation du taux de criminalité pouvons-nous nous attendre pour une variation unitaire (ici 10 000 habitants par km^2) de la densité de population ?
4. Estimer le taux de criminalité le plus plausible pour une densité de population de 7 500 habitants par km^2 .
5. Calculer la variation qui est expliquée par la droite des moindres carrés.
6. À l'aide des calculs préliminaires, calculer la variation totale du taux de criminalité.
7. Quelle proportion de la variation totale est expliquée par la droite des moindres carrés ?