

DEUG I Sociologie

Statistiques

Régression linéaire

On considère le tableau de données¹, où le taux d'alphabétisation est la proportion de la population sachant lire et écrire, alors que l'indice de fécondité est le nombre d'enfants par femmes.

	Alphabétisation		Fécondité	
	1980	2000	1981	2001
Rwanda	40%	67%	6,9	5,8
Nigéria	33%	64%	6,9	5,8
Côte d'Ivoire	27%	47%	6,7	5,2
Algérie	40%	63%	7,3	3,1
Afrique du Sud	77%	85%	5,1	2,9
Zimbabwe	80%	93%	6,6	4,0
Colombie	85%	92%	3,9	2,6
Afghanistan	18%	47%	6,9	6,0
Inde	41%	56%	5,3	3,2
Indonésie	69%	87%	4,1	2,7
Philippines	89%	95%	5,0	3,5
Sri Lanka	85%	92%	3,4	2,1
Tadjikistan	94%	99%	5,6	2,4
Iran	51%	77%	5,3	2,6
Mali	14%	40%	6,7	7,0
Niger	8%	16%	7,1	7,5
Chine	66%	85%	2,3	1,8
Pakistan	28%	43%	6,3	5,6

- On va étudier la relation entre le taux d'alphabétisation en 2000 (α) et l'indice de fécondité en 2001 (φ) pour les 16 premiers pays de ce tableau (*i.e.* on ne prend pas en compte la Chine et le Pakistan dans l'étude).
 - Déterminer le coefficient de corrélation linéaire $r(\alpha, \varphi)$ entre les deux paramètres. Qu'en pensez-vous ?
 - Calculer l'équation de la droite de régression de φ en fonction de α .
 - Tracer le nuage de points $(\alpha_i, \varphi_i)_{i=1\dots 16}$ en plaçant le taux d'alphabétisation en abscisse et l'indice de fécondité en ordonnée. Tracer aussi la droite de régression.
- Examinons maintenant le cas de la Chine et du Pakistan.
 - Placer les deux points correspondant à ces deux pays sur le graphique du 1.(c).
 - Quel serait l'indice de fécondité prédit pour la Chine en fonction de son taux d'alphabétisation réel (de 85%). Interpréter le résultat.
 - Même question pour le Pakistan.
- Étudions maintenant l'évolution du taux d'alphabétisation entre 1980 (noté x) et 2000 (noté y) pour les 17 premiers pays (on exclus cette fois le Pakistan de l'étude).
 - Déterminer le coefficient de corrélation linéaire $r(x, y)$. Qu'en pensez-vous ?
 - Calculer l'équation de la droite de régression de y en fonction de x .
 - Dessiner le nuage de points et la droite de régression.
 - Faire une prédiction du taux d'alphabétisation du Pakistan en 2000 à partir de son taux d'alphabétisation en 1980.
 - Toujours sur la seule base du taux d'alphabétisation du Pakistan en 1980, faire une prédiction de son indice de fécondité en 2001.
 - Question plus difficile : peut-on utiliser ces résultats pour prédire le taux d'alphabétisation probable du Pakistan en 2010 ? en 2020 ?

¹Après l'empire, essai sur la décomposition du système américain, Emmanuel Todd, 2002