

Séance 10 – Méthodes statistiques multivariées explicatives

Honoka OYAMADA

1. Questions de cours

1. La corrélation dans un espace à n dimensions généralise la corrélation simple en mesurant la part de variance expliquée d'une variable par plusieurs autres simultanément. Elle repose sur la covariance et la géométrie des vecteurs dans \mathbb{R}^n .
2. Une régression multiple vise à expliquer une variable quantitative par plusieurs variables explicatives, en estimant un hyperplan des moindres carrés.
3. L'ANOVA à double entrée sert à comparer des moyennes selon deux facteurs qualitatifs et leur interaction.
4. L'analyse canonique est très générale mais difficilement interprétable, ce qui limite son usage opérationnel.
5. Les méthodes explicatives sont utiles pour modéliser, prédire et comprendre les déterminants spatiaux en géographie.

2. Régression sur les températures

Les corrélations et statistiques descriptives montrent une forte dépendance de la température de janvier à la latitude. La régression OLS confirme cet effet, avec une latitude significative et un R^2 élevé. L'altitude joue un rôle secondaire, tandis que la longitude est peu significative.

3. Régression géomarketing

La régression multiple appliquée au chiffre d'affaires montre que la surface du magasin et le potentiel de population sont les variables les plus explicatives. Le modèle avec variables sélectionnées présente un bon compromis entre explicativité et parcimonie. Le modèle avec toutes les variables améliore légèrement le R^2 mais au prix d'une multicolinéarité accrue.

4. Conclusion

Ces analyses illustrent l'intérêt des méthodes explicatives pour comprendre des phénomènes géographiques et économiques complexes, tout en soulignant l'importance de la sélection des variables.