

Modélisation statistique

#2.b Invariance aux transformations linéaires des variables explicatives

Dr. Léo Belzile
HEC Montréal

Transformations linéaires des variables explicatives

Considérez le log du nombre quotidien de locations de Bixi en fonction de la température en degrés Celcius et Fahrenheit.

Supposez que l'effet de la température sur le log du nombre de locations est

$$\text{logutilisateur} = \alpha_0 + \alpha_1 \text{celcius} + \varepsilon.$$

- ✚ Interprétation de α_1 : *l'augmentation moyenne du log du nombre de locations par jour quand la température augmente de 1 °C.*

Le modèle en Fahrenheit est

$$\text{logutilisateur} = \gamma_0 + \gamma_1 \text{fahrenheit} + \varepsilon.$$

Sortie SAS

Paramètre	Estimation	Erreur type	Valeur du test t	Pr > t
Constante	8.844327052	0.02819099	313.73	<.0001
celcius	0.048566261	0.00135205	35.92	<.0001

Paramètre	Estimation	Erreur type	Valeur du test t	Pr > t
Constante	7.980926861	0.05132678	155.49	<.0001
farenheit	0.026981256	0.00075114	35.92	<.0001

Les deux unités de mesures sont reliées par la transformation linéaire

$$1.8\text{celcius} + 32 = \text{farenheit}.$$

et on trouve que $\alpha_0 = \gamma_0 + 32\gamma_1$ et $\alpha_1 = 1.8\gamma_1$.

Unicité de la solution

Les paramètres du modèle postulé avec les deux prédicteurs,

$$\text{logutilisateur} = \beta_0 + \beta_c \text{celcius} + \beta_f \text{fahrenheit} + \varepsilon,$$

ne sont pas **identifiables**, puisque toute combinaison linéaire des deux solutions donne les mêmes valeurs ajustées.

Pour $k \in \mathbb{R}$, $\beta_0 = k\alpha_0 + (1 - k)\gamma_0$, $\beta_1 = k\alpha_1$ et $\beta_2 = (1 - k)\gamma_1$ sont équivalents.

Le rang de \mathbf{x} est 2, mais la matrice de plan d'expérience a 3 colonnes.

- + $\mathbf{x}^\top \mathbf{x}$ n'est pas inversible.
- + la solution de l'équation normale n'est donc **pas unique**.

Colinéarité

Paramètre	Estimation		Erreur type	Valeur du test t	Pr > t
Constante	8.844327052	B	0.02819099	313.73	<.0001
celcius	0.048566261	B	0.00135205	35.92	<.0001
fahrenheit	0.000000000	B	.	.	.

SAS inclut un avertissement si les données sont *exactement* colinéaires.

Note: The $X'X$ matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.