Modélisation Statistique
Chapitre 3: Extensions de la régression

Mohamed Essaied Hamrita

ISMAI, Université Kairouan. Tunisie
mhamrita@gmail.com
http://hamrita.e-monsite.com/

Avril 2016

Plan du chapitre

Applications

Les **variables muettes (ou indicatrice ou dummies)** sont des variables explicatives de type **qualitatif** et qui sont représentées sous forme de code.

Applications

Les **variables muettes (ou indicatrice ou dummies)** sont des variables explicatives de type **qualitatif** et qui sont représentées sous forme de code.

Exemple:

Y a-t-il une différence entre le salaire moyen des hommes et celui des femmes?

Considérons le modèle suivant:

$$S_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_i + \varepsilon_i, \qquad i = 1, \dots, n \tag{1}$$

avec:

S_i: le salaire, X_i = nombre d'années d'éducation et D_i=1 si l'individu i est un homme et 0 sinon.

D_i est une variable indicatrice (ou binaire). On dit que l'individu de référence (catégorie omise) est la femme.

Interprétation du paramètre β_2

Calculons l'espérance des salaires des hommes et celle des femmes.

$$E\left(S_i \mid D_i = 1\right) = \beta_0 + \beta_1 X_i + \beta_2$$

$$E\left(S_i \mid D_i = 0\right) = \beta_0 + \beta_1 X_i$$

D'où $\beta_2 = E\left(S_i \mid D_i = 1\right) - E\left(S_i \mid D_i = 0\right) \Longrightarrow \beta_2$ indique l'écart moyen entre le salaire des hommes et celui des femmes toutes choses étant égales par ailleurs.

Pour tester s'il y a une différence significative entre le salaire moyen des hommes et celui des femmes, on applique le test de student.

Il s'agit de tester $H_0 : \beta_2 = 0$.

Modèle semi-logarithmique

Considérons la modélisation du log du salaire

$$s_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \varepsilon_i$$

où $s_i = \ln S_i$. Soient s_i^H et s_i^F le log des salaires des hommes et des femmes respectivement. D'où

$$s_i^H - s_i^F = \ln\left(\frac{S_i^H}{S_i^F}\right) = \beta_2 \implies \frac{S_i^H}{S_i^F} - 1 = e^{\beta_2} - 1$$

$\Longrightarrow$ le salaire moyen des hommes est de $e^{\beta_2} - 1$ plus proportionnelle que celui des femmes.
Exemple: $e^{\beta_2} - 1 = 1.25\% \Longrightarrow$ le salaire moyen des hommes est de plus 1.25% que celui des femmes.
Remarque: Si β_2 est proche de zéro, alors $e^{\beta_2} - 1 \simeq \beta_2$ et dans ce cas, on peut interpréter β_2 comme la différence proportionnelle entre les deux catégories.

Autres applications

Différence des pentes:

Considérons le modèle suivant:

$$S_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + \varepsilon_i$$

où $D_i = 1$ si homme et 0 sinon. Donc

$$S_i^H = \beta_0 + (\beta_1 + \beta_2) X_i + \varepsilon_i, \quad \text{et} \quad S_i^F = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Afin d'interpréter β_2, prenons la dérivée de S_i par raport à X_i

$$\frac{\partial S_i^H}{\partial X_i} = \beta_1 + \beta_2 \quad \text{et} \quad \frac{\partial S_i^F}{\partial X_i} = \beta_1$$

Donc β_2 représente la différence des pentes entre le modèle des hommes et celui des femmes.
Si $\beta_2 > 0$, celà veut dire que le nombre d'années d'expériences a un effet positif sur le salaire moyen plus important chez les hommes que chez les femmes.

Autres applications

Changement structurel Considérons le modèle suivant:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 D_t + \varepsilon_t$$

où $D_t = 1$ si $t > t^*$ et $D_t = 0$ sinon. D'où

$$Y_t = \begin{cases} \beta_0 + \beta_1 X_t + \varepsilon_t & \text{si } t \leq t^* \\ (\beta_0 + \beta_2) + \beta_1 X_t + \varepsilon_t & \text{si } t > t^* \end{cases}$$

Si β_2 est statistiquement significatif, alors, il existe un changement structurel à la pèriode t^*.

Autres applications

Effets saisonniers On considère le modèle suivant:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 D_{1t} + \beta_3 D_{2t} + \beta_4 D_{3t} + \varepsilon_t$$

où $D_{it} = 1$ si l'observation correspond au trimestre i et $D_{it} = 0$ sinon.

Autres applications

Effets saisonniers On considère le modèle suivant:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 D_{1t} + \beta_3 D_{2t} + \beta_4 D_{3t} + \varepsilon_t$$

où $D_{it} = 1$ si l'observation correspond au trimestre i et $D_{it} = 0$ sinon.

Pour tester l'abscence d'effets saisonniers, on peut effectuer le test de Ficher:
$H_0 : \ \beta_2 = \beta_3 = \beta_4 = 0.$

Modèles dynamiques

Considérons le modèle **dynamique** suivant:

$$Y_t = \beta_0 + \beta_1 X_t + \gamma Y_{t-1} + \varepsilon_t$$

avec $|\gamma| < 1$.

- Les observations à une période donnée sont liées à celles pour des pèriodes antérieures.
- Par exemple, Y_t représente la consommation:
  La consommation (Y_t) d'aujourd'hui est explicitement liée à la consommation passée (Y_{t-1}).

Modèle dynamique général

Le modèle linéaire général exprime la variable dépendente en fonction d'une combinaison linéaire des valeurs actuelles et antérieures de la variable dépendante et des variables explicatives: $Y_t = f(Y_j, X_{k,t-q_i})$. Ce modèle est noté $ADL(p, q_1, q_2, \ldots, q_k)$ (Autoregressive Distributed Lag model) c-à-d modèle autoregressive à retards échelonnés. Ce modèle est définit comme suit:

$$\Phi(L)Y_t = \alpha + \gamma_1(L)X_{1t} + \ldots + \gamma_k(L)X_{kt} + \varepsilon_t$$

avec $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$
$\gamma_i(L) = \gamma_{0i} + \gamma_{1i} L + \ldots + \gamma_{q_i i} L^{q_i}$.
où L est l'opérateur retard (Lag) qui vérifie:
$LX_t = X_{t-1}$ et $L^j X_t = X_{t-j}$.

Exemple: modèle ADL$(1,1)$ $p = 1,\ k = 1,\ q_1 = 1$.

$$y_t = \underbrace{\alpha + \phi_1 y_{t-1}}_{\text{partie autorégressive}} + \underbrace{\gamma_0 X_t + \gamma_1 X_{t-1}}_{\text{partie retard échelonnée}} + \varepsilon_t$$

Cas particuliers

- $p = q_1 = \ldots = q_k = 0$: C'est le modèle statique.
- $\gamma_i(L) = 0$: modèle autorégressive.
- modèle statique à erreurs autoregrégressive:

$$Y_t = \alpha^* + \sum_{j=1}^{k} \gamma_j X_{jt} + u_t, \qquad u_t = \Phi(L)^{-1} \varepsilon_t$$

Multiplicateurs

- Effet (multiplicateur) court terme:
  L'effet de court terme d'une variation de X_t sur Y_t est mesurée par:
  $$ECT = \frac{\partial Y_t}{\partial X_t} = \gamma_0.$$

Multiplicateurs

- Effet (multiplicateur) court terme:
  L'effet de court terme d'une variation de X_t sur Y_t est mesurée par:
  $$ECT = \frac{\partial Y_t}{\partial X_t} = \gamma_0.$$

- Effets (multiplicateurs) intermédiaires (dynamiques):
  On peut exprimer la variable Y en fonction des observations courantes et celles retardées de Y et de X pour déterminer la forme réduite du modèle:
  $$\Phi(L)Y_t = \alpha + \gamma(L)X_t + \varepsilon_t$$
  $$\Longrightarrow Y_t = \Phi^{-1}(L)\alpha + \Phi^{-1}(L)\gamma(L)X_t + \Phi^{-1}(L)\varepsilon_t$$
  $$\frac{\partial Y_t}{\partial X_t} = \gamma_0, \quad \frac{\partial Y_t}{\partial X_{t+1}} = \gamma_1 + \phi_1\gamma_0, \ldots$$
  Ces valeurs représentent les effets dynamiques. L'effet intermédiaire de n premières périodes est définie par la somme partielle des n premiers effets dynamiques, soit:
  $$EI = \sum_{j=1}^{n} \frac{\partial Y_{t+j}}{\partial X_t}$$

Multiplicateurs

- Effet (multiplicateur) de long terme:
  L'effet de long terme d'une variation de X_t sur Y_t est la somme infinie des effets dynamiques:

$$ELT = \sum_{j=1}^{\infty} \frac{\partial Y_{t+j}}{\partial X_t} = \frac{\gamma(1)}{\Phi(1)}.$$

Multiplicateurs

- Effet (multiplicateur) de long terme:
  L'effet de long terme d'une variation de X_t sur Y_t est la somme infinie des effets dynamiques:

$$ELT = \sum_{j=1}^{\infty} \frac{\partial Y_{t+j}}{\partial X_t} = \frac{\gamma(1)}{\Phi(1)}.$$

- Retard (délai) moyen:
  On définit le retard moyen par :

$$\overline{R} = \frac{A'(1)}{A(1)} \text{ avec } A(L) = \frac{\gamma(L)}{\Phi(L)}$$

Exemple 1

Soit le modèle $ADL(1,1)$ suivant:

$$y_t = \alpha + \phi y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + \varepsilon$$
$$\iff \underbrace{(1-\phi L)}_{\Phi(L)} y_t = \alpha + \underbrace{(\gamma_0 + \gamma_1 L)}_{\gamma(L)} x_t + \varepsilon_t$$

$$\iff y_t = \frac{\alpha}{1-\phi L} + \frac{\gamma_0 + \gamma_1 L}{1-\phi L} x_t + \frac{\varepsilon_t}{1-\phi L}$$

avec

$$\frac{1}{1-\phi L} = 1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \ldots + \phi^j L^j + \ldots$$

$$\frac{\gamma_0 + \gamma_1 L}{1-\phi L} = \gamma_0 + (\gamma_1 + \gamma_0 \phi) L + (\gamma_1 \phi + \gamma_0 \phi^2) L^2 + \ldots + (\gamma_1 \phi^{j-1} - \gamma_0 \phi^j) L^j + \ldots$$

D'où,

$$\frac{\partial y_t}{\partial x_t} = \gamma_0, \quad \frac{\partial y_{t+1}}{\partial x_t} = \gamma_1 + \gamma_0 \phi, \ \ldots, \ \frac{\partial y_{t+j}}{\partial x_t} = \gamma_1 \phi^{j-1} - \gamma_0 \phi^j$$

Exemple 1

L'effet de court terme de x sur y est: γ_0.
L'effet de long terme de x sur y est la somme infinie (si elle existe) des effets dynamiques:

$$ELT = \sum_{j=1}^{\infty} \left(\gamma_1 \phi^{j-1} - \gamma_0 \phi^{j}\right) = \frac{\gamma(1)}{\Phi(1)} = \frac{\gamma_0 + \gamma_1 \phi}{1 - \phi}$$

Exemple 2

Soit le modèle: $Y_t = \text{const} \ + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2} + u_t$, où Y_t indique le niveau de consommation pendant l'année t et X_t désigne le niveau de revenu.

L'effet de court terme est égale à 0.4. Ceci veut dire que: si le revenu augmente de 1 DT, le consommateur augmente ses dépenses de 400 millimes pendant l'année courante.

L'effet de long terme est égale à 0.4+0.3+0.2=0.9. Suite à une augmentation du revenu de 1 DT, le consommateur va augmenter son niveau de consommation de 400 millimes pendant l'année courante, de 300 millimes de plus pendant l'année qui suit et de 200 millimes de plus pendant l'année suivante.

Le retard moyen est:

$$\overline{R} = \frac{0.3 + 2 \times 0.2}{0.4 + 0.3 + 0.2} = 0.7777.$$