Econométrie - L3 - Formules

Modèle de régression linéaire SIMPLE / MULTIPLE

✓ Spécification du modèle :

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + ... + \beta_k X_{kt} + \varepsilon_t$$

$$\underbrace{\mathbf{y}}_{(T,1)} = \underbrace{\mathbf{X}}_{(T,k+1)} \underbrace{\boldsymbol{\beta}}_{(k+1,1)} + \underbrace{\boldsymbol{\varepsilon}}_{(T,1)}$$

✓ Estimation par MCO : intuition, définition

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\mathbf{ArgMin}} \sum_{t=1}^{T} e_t^2 \equiv \underset{\boldsymbol{\beta}}{\mathbf{ArgMin}}\ \mathbf{e'e}$$

✓ Hypothèses MCO

✓ Estimateurs MCO, $\hat{\beta}$:
— définition

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

— propriétés à distance finie + propriétés asymptotiques
— matrice de variance covariance, $\mathbb{V}(\hat{\beta}) \equiv \Sigma_{\hat{\boldsymbol{\beta}}} = \sigma_\varepsilon^2(\mathbf{X'X})^{-1}$
— distribution, $\hat{\beta} \sim N(\mathbb{E}(\hat{\beta}), \mathbb{V}(\hat{\beta}))$

✓ Qualité d'ajustement et Analyse de la variance : R^2, $\bar{R}^2$ (définition, interprétation)

$$R^2 = \frac{\mathbb{V}(\hat{Y})}{\mathbb{V}(Y)} = 1 - \frac{\mathbb{V}(e)}{\mathbb{V}(Y)}$$

$$\bar{R}^2 = 1 - \frac{T-1}{T-k-1}(1-R^2)$$

✓ Estimateur sans biais, $\hat{\sigma}_\varepsilon^2$, de la variance des termes d'erreur, $\hat{\sigma}_\varepsilon^2$.

Intervalle de confiance & Tests statistiques

✓ Intervalle de confiance sur β_i

$$IC(\beta_i, 1-p) = [\hat{\beta}_i \pm t(T-k-1)_{p/2} \times s_{\hat{\beta}_i}]$$

✓ Niveau de risque *versus* niveau de confiance
✓ Erreur de type I, Erreur de type II, Puissance du test

✓ Etapes à suivre :
— Test d'hypothèses
— Niveau de risque
— Statistique de test
— Décision & Conclusion

✓ **Tests sur un coefficient**

Test	H_0	H_1	Décision Rejet de H_0 si
Two-tailed test	$\beta = \beta^*$	$\beta \neq \beta^*$	$\lvert t_{stat}\rvert > \lvert t_{p/2}(df)\rvert$
One-tailed test (upper)	$\beta = \beta^*$	$\beta > \beta^*$	$t_{stat} > t_{1-p}(df)$
One-tailed test (lower)	$\beta = \beta^*$	$\beta < \beta^*$	$t_{stat} < t_p(df)$

β^* est la valeur de β sous H_0, $t_p(df)$ est le quantile d'ordre p d'une distribution de Student avec df degrés de liberté.

✓ **Tests sur plusieurs coefficients**

— **Cas A** :
Test d'hypothèses

$$H_0 : \text{Restrictions linéaires}$$

Statistique de test

$$F_{stat} = \frac{(e_c'e_c - e_{nc}'e_{nc})/c}{e_{nc}'e_{nc}/(T-k-1)} \sim F(c, T-k-1)$$

avec
— $e_c'e_c$ la somme des carrés des résidus du **modèle contraint**
— $e_{nc}'e_{nc}$ la somme des carrés des résidus du **modèle non-contraint**
— T est le nombre d'observations
— k est le nombre de variables explicatives dans le modèle non-contraint
— c est le nombre de restrictions
Décision & Conclusion

— **Cas B** :
Test d'hypothèses

$$H_0 : \underset{(c,k+1)}{R}\ \underset{(k+1,1)}{\beta} = \underset{(c,1)}{r}$$

avec
c le nombre de restrictions
$k+1$ le nombre de paramètres à estimer (constante comprise)

Statistique de test

$$F_{stat} = \frac{(R\hat{\beta} - r)'[\hat{\sigma}_\varepsilon^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{c} \sim F(c, T-k-1)$$

Décision & Conclusion

— **Cas spécifique : Test de significativité globale**
Test d'hypothèses
Statistique de test

$$F_{stat} = \frac{R^2/k}{(1-R^2)/(T-k-1)} \sim F(k, T-k-1)$$

Décision & Conclusion

Prévisions et Intervalles de confiance

✓ En supposant que la relation engendrant la variable expliquée reste identique et que les valeurs des variables explicatives sont connues en $T+h$, on a :

$$\hat{Y}_{T+h} = \hat{\alpha} + \hat{\beta}_1 X_{1,T+h} + \hat{\beta}_2 X_{2,T+h} + ... + \hat{\beta}_k X_{k,T+h}$$

✓ **L'erreur de prévision** :

$$e_{T+h} = Y_{T+h} - \hat{Y}_{T+h} = \varepsilon_{T+h} + X^{(T+h)}(\beta - \hat{\beta})$$

avec $X^{(T+h)}$ le vecteur contenant les valeurs des variables explicatives à la date $T+h$ et dont le premier élément est 1 : $X^{(T+h)} = (1\ X_{1,T+h}\ X_{2,T+h}\ ...\ X_{k,T+h})$

✓ **Intervalle de confiance** pour Y_{T+h} autour de la prévision $\hat{Y}_{T+h}$

$$IC(Y_{T+h}, 1-p) = [\hat{Y}_{T+h} \pm t(T-k-1)_{p/2} s_{e_{T+h}}]$$

Variables indicatrices / dummy

Tests de diagnostic

✓ **Test de mauvaise spécification (hétéroscédasticité et autocorrélation dans les termes d'erreur)**
— Sources de l'autocorrélation et de l'hétéroscédasticité
— Quelles sont les conséquences de la présence des erreurs autocorrélées et/ou hétéroscédastiques ?
— Comment détecter l'autocorrélation et l'hétéroscédasticité ?

1. **Autocorrélation**
   — Test de Durbin-Watson
   Test d'hypothèses
   Statistique de test
   Décision & Conclusion

   — Test de Breusch-Godfrey (optionnel)

2. **Hétéroscédasticité**
   — Test de Goldfeld-Quandt
   Test d'hypothèses
   Statistique de test

$$GQ_{stat} = \frac{max(\hat{\sigma}^2_{\varepsilon_1}, \hat{\sigma}^2_{\varepsilon_2})}{min(\hat{\sigma}^2_{\varepsilon_1}, \hat{\sigma}^2_{\varepsilon_2})}$$

   avec $\hat{\sigma}^2_{\varepsilon_i}$ la variance résiduelle de la période i

   Décision & Conclusion

   — Test de Breusch-Pagan (optionnel)
   — Test de White (optionnel)

— Comment corriger de l'autocorrélation et de l'hétéroscédasticité ?

✓ **Tests de stabilité structurelle (intuition)**

— Test de Chow
Test d'hypothèses
Statistique de test

$$F_{stat} = \frac{(SSR - (SSR_1 + SSR_2))/(k+1)}{(SSR_1 + SSR_2)/(T - 2(k+1))}$$

avec
— SSR la somme des carrés des résidus sur l'ensemble de la période
— SSR_1 la somme des carrés des résidus sur la première période
— SSR_2 la somme des carrés des résidus sur la deuxième période
— T le nombre d'observations sur l'ensemble de la période

Décision & Conclusion