

Tests et Analyse de Variance en Régression OLS

GSF-6053

Hiver 2025

Table des matières

| | | |
|----------|---|-----------|
| 1 | Analyse de Variance | 2 |
| 1.1 | Coefficient de Détermination R^2 | 3 |
| 2 | Tests d'Hypothèses | 4 |
| 2.1 | Tests de Significativité | 5 |
| 2.1.1 | Le Test de Student | 6 |
| 2.1.2 | Le Test de Significativité Conjointe F | 6 |
| 2.2 | Les Contraintes Linéaires | 7 |
| 2.3 | Le Test de Wald | 9 |
| 2.4 | La Statistique LR | 9 |
| 2.5 | Le Multiplicateur de Lagrange (LM) | 10 |
| 2.5.1 | Lien entre les Statistiques F, Wald, LR et LM | 12 |
| 3 | Bibliographie | 13 |
| | Bibliographie | 13 |

1 Analyse de Variance

[Gujarati et Porter](#) : Chapitres 3 et 7. [Wooldridge](#) chapitres 2, 3, 4.

Dans une régression OLS avec terme constant, la variation dans la variable expliquée (Y) peut être décomposée pour analyser la variance.

$$Y_t = \hat{Y}_t + \hat{u}_t$$

Où \hat{Y}_t est la valeur ajustée.

Donc, la variation de Y résulte en deux composantes :

- une variation due à la partie expliquée par les régresseurs du modèle (\hat{Y}_t),
- une variation induite par la partie de Y non expliquée par la régression (\hat{u}_t).

La variation totale de Y par rapport à sa moyenne est donnée par la somme des carrés totaux :

$$SS_{\text{tot}} = \sum_{t=1}^T (Y_t - \bar{Y})^2$$

La variation de la partie expliquée, la somme des carrés expliqués par la régression est donnée par :

$$SS_{\text{reg}} = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$$

Finalement, les résidus sont nuls en moyenne, donc la variation dans les résidus est :

$$SS_{\text{err}} = \sum_{t=1}^T \hat{u}_t^2$$

On voit que :

$$SS_{\text{tot}} = SS_{\text{reg}} + SS_{\text{err}}$$

c'est-à-dire que la variance totale est divisée en sa composante expliquée



et non expliquée.

$$SS_{\text{tot}} = \sum_{t=1}^T (Y_t - \bar{Y})^2 = \sum_{t=1}^T (\hat{Y}_t + \hat{u}_t - \bar{Y})^2$$

Développant l'expression :

$$SS_{\text{tot}} = \sum_{t=1}^T \left[\hat{Y}_t^2 + 2\hat{Y}_t\hat{u}_t + \hat{u}_t^2 - 2\hat{Y}_t\bar{Y} - 2\hat{u}_t\bar{Y} + \bar{Y}^2 \right]$$

Simplifiant et regroupant les termes :

$$SS_{\text{tot}} = \sum_{t=1}^T \left(\hat{Y}_t^2 + \bar{Y}^2 - 2\hat{Y}_t\bar{Y} \right) + 2 \sum_{t=1}^T \hat{Y}_t\hat{u}_t + \sum_{t=1}^T \hat{u}_t^2 - 2\bar{Y} \sum_{t=1}^T \hat{u}_t$$

Étant donné que les résidus sont orthogonaux au sous-espace engendré par les colonnes de X et que la somme des résidus est nulle, les deux derniers termes s'annulent :

$$SS_{\text{tot}} = SS_{\text{reg}} + SS_{\text{err}}$$

1.1 Coefficient de Détermination R^2

[Gujarati et Porter](#) section 7.8. [Wooldridge](#) : 2.3, 3.2h.

Le coefficient de détermination indique le pouvoir explicatif de la régression, c'est la proportion des variations de Y expliquées par les régresseurs du modèle par rapport aux variations totales. Donc,

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}} = 1 - \frac{SS_{\text{err}}}{SS_{\text{tot}}}$$

Clairement,

$$0 \leq R^2 \leq 1$$

On utilise cette mesure pour évaluer la performance d'un modèle et la comparaison de modèles. Par contre, pour que la mesure soit valide, il faut



comparer les modèles sur les mêmes variables expliquées. Vous pouvez par contre évaluer la contribution marginale d'un régresseur à l'équation en regardant combien la mesure de R^2 augmente avec l'ajout de celui-ci.

Une faiblesse de cette mesure de la qualité de l'ajustement du modèle est qu'elle augmente nécessairement avec le nombre de régresseurs, peu importe la pertinence de ceux-ci. Un modèle avec plus de régresseurs n'étant pas nécessairement un meilleur modèle pour expliquer les variables endogènes, il faut introduire une mesure ajustée pour ce facteur, le R^2 ajusté.

$$R_{\text{adj}}^2 = 1 - \left(\frac{SS_{\text{err}}/(n - K)}{SS_{\text{tot}}/(n - 1)} \right)$$

Où K est le nombre de régresseurs.

On peut montrer que :

$$R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{n - 1}{n - K}$$

Le R^2 ajusté tient donc compte du nombre de régresseurs et le modèle sera pénalisé si l'on ajoute un régresseur sans pouvoir explicatif.

2 Tests d'Hypothèses

[Gujarati et Porter](#) : Chapitres 3, 5 et 7. [Wooldridge](#) : chapitre 4.

Un test d'hypothèse sur les paramètres d'un modèle économétrique requiert les étapes suivantes :

1. Écrire l'hypothèse nulle (H_0) et l'alternative (H_1). La plupart du temps, H_0 est l'hypothèse que l'on veut rejeter. Par exemple, qu'un paramètre est non significatif ou égal à une certaine valeur. L'hypothèse alternative peut être unilatérale (\leq, \geq) ou bilatérale (\neq).
2. Définir la statistique de test (Student, F, Wald, LR, LM, etc.) et déterminer si possible sa loi de distribution sous H_0 .
3. Choisir le niveau de significativité du test α . On fixe donc la possibilité d'erreur de type I, soit la probabilité de rejeter H_0 lorsque celle-ci est



vraie. Généralement, on fixe le niveau à 5 % ou 1 % et plus rarement à 10 %.

4. Déterminer la règle de décision du test avec la région critique de confiance CR_α . La plupart du temps, cela demande de connaître la distribution de la statistique sous H_0 . Lorsque la valeur calculée pour la statistique de test se trouve dans la région critique, on rejette H_0 en faveur de H_1 au seuil de confiance α .
5. Utiliser les valeurs obtenues de la régression pour calculer la statistique de test.
6. Appliquer la règle de décision vue en 4.

Note sur la p-value ([Gujarati et Porter](#), p.835 ; [Wooldridge](#) : appendices C-6d et 4.2d.)

Parfois, on regarde la p-value au lieu de comparer CR_α à la valeur de la statistique.

p-value : C'est le plus petit niveau auquel on peut rejeter l'hypothèse nulle. En d'autres mots, c'est la probabilité d'avoir un événement aussi extrême que celui observé en assumant que H_0 est vraie. Plus la p-value est faible, plus la probabilité que l'événement observé fasse partie de la distribution est faible, étant donné l'hypothèse nulle. Donc, plus il y a de chances que l'hypothèse nulle soit rejetée.

2.1 Tests de Significativité

Il y a deux types de significativité :

- La significativité individuelle d'un paramètre. On regarde si l'effet d'un régresseur est statistiquement non nul. Pour ce faire, on utilise le test de Student.
- La significativité conjointe des paramètres du modèle. On teste ici si le modèle est spécifié de telle sorte qu'au moins un des régresseurs a un effet statistiquement non nul. Pour effectuer ce test, on utilise le test F.



2.1.1 Le Test de Student

[Gujarati et Porter](#) 5.7 et 8.3; [Wooldridge](#) : chapitre 4.

Pour chaque variable explicative, l'hypothèse nulle est la non-significativité du coefficient de régression et l'hypothèse alternative l'inverse :

$$H_0 : \beta_k = 0 \quad \text{contre} \quad H_1 : \beta_k \neq 0$$

La statistique de Student pour un test est :

$$t = \frac{|\hat{\beta}_k - \beta_0|}{SE_{\hat{\beta}_k}}$$

Ici, β_0 est la valeur du coefficient sous la nulle, soit 0. $SE_{\hat{\beta}_k}$ est l'écart type associé à l'estimation de $\hat{\beta}_k$.

La règle de décision du test est de rejeter H_0 si la statistique t est plus grande que le point critique associé à la distribution sous H_0 pour un niveau de confiance de α :

$$t = \frac{|\hat{\beta}_k - \beta_0|}{SE_{\hat{\beta}_k}} > t_{\alpha/2, n-K}$$

On peut aussi faire des tests de Student avec d'autres valeurs d'intérêt sous la nulle que 0. On peut faire des tests unilatéraux. Lorsque l'on fait un test unilatéral, il faut attention si c'est un test à gauche ou à droite afin d'ajuster la règle de décision en conséquence.

2.1.2 Le Test de Significativité Conjointe F

[Gujarati et Porter](#), 8.4; [Wooldridge](#) : 4.5.

Le test de significativité conjointe du modèle teste la pertinence de celui-ci. Notez cependant que la constante n'est pas incluse dans ce test.

$$H_0 : \beta_2 = 0 \text{ et } \beta_3 = 0 \text{ et } \dots \text{ et } \beta_K = 0$$

Contre l'alternative qu'au moins un des coefficients estimés est pertinent pour expliquer les variations dans Y :



$$H_1 : \beta_2 \neq 0 \text{ et/ou } \beta_3 \neq 0 \text{ et/ou } \dots \text{ et/ou } \beta_K \neq 0$$

La statistique de test est basée sur la somme des carrés des résidus du modèle non contraint ($\hat{u}'\hat{u}$) et du modèle contraint ($\hat{u}'_0\hat{u}_0$) par les restrictions sous la nulle, soit dans ce cas-ci un modèle avec constante uniquement.

$$F = \frac{(\hat{u}'_0\hat{u}_0 - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/(T - K)}$$

où K est le nombre de régresseurs et q le nombre de contraintes sous l'hypothèse nulle.

L'intuition de la statistique de Fisher est de comparer la différence entre les résidus au carré dans le modèle contraint ($\hat{u}'_0\hat{u}_0$) et non contraint ($\hat{u}'\hat{u}$). Plus cette différence est grande, plus les résidus du modèle contraint sont grands par rapport au modèle non contraint et plus la contrainte « coûte cher » à appliquer en termes de performance du modèle. Plus cette différence entre les résidus contraints et non contraints est grande, plus la statistique F est grande. On rejette l'hypothèse nulle passé un certain « coût critique ».

On rejette H_0 à un seuil de α si :

$$F = \frac{(\hat{u}'_0\hat{u}_0 - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/(T - K)} > F_{(q, T-K; \alpha)}$$

2.2 Les Contraintes Linéaires

[Gujarati et Porter](#) 8.9 et appendice 8A2. [Wooldridge](#) E.4a, 17.1c, 5.2a.

On peut tester beaucoup d'autres hypothèses dans une régression que la significativité individuelle ou conjointe des paramètres estimés. On peut adapter la statistique de Student et la statistique de Fisher pour d'autres hypothèses, mais d'autres types de tests sont aussi disponibles. Nous verrons le Wald, le LR et le LM. Dans tous les cas, on teste les contraintes linéaires du type général suivant :

$$H_0 : R\beta = r$$



$$R = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & -2 & \dots & 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{K-1} \end{bmatrix}, \quad r = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$$

Attention : encore une fois, ici la clé est de regarder les dimensions de ces matrices.

Le multiplicateur de Lagrange (Lagrangien) est défini comme un vecteur λ dans un problème d'optimisation :

$$\mathcal{L}^* = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) + \lambda'(R\beta - r)$$

En prenant les conditions de premier ordre du Lagrangien présenté ci-dessus, nous trouvons que :

$$C\hat{\lambda} = -2(R\hat{\beta} - r)$$

Où $C = [R(X'X)^{-1}R']^{-1}$ (nous assumons que $R(X'X)^{-1}R'$ est carré et de plein rang inversible).

$$\sigma^2 = \frac{u_c' u_c}{T}$$

Ainsi, on trouve que les $\hat{\beta}_c$ s'expriment en fonction de l'estimé non contraint $\hat{\beta}$:

$$\hat{\beta}_c = \hat{\beta} - (X'X)^{-1}R'C(R\hat{\beta} - r)$$

La somme des résidus contraints s'exprime également en fonction des résidus non contraints :

$$\hat{u}_c' \hat{u}_c = \hat{u}' \hat{u} + (R\hat{\beta} - r)' C (R\hat{\beta} - r)$$

On voit donc que l'on peut exprimer les estimés contraints en fonction des estimés non contraints.



2.3 Le Test de Wald

Wooldridge E-4a

Le test de Wald est une forme quadratique basée sur la distance entre $(R\hat{\beta} - r)$ et zéro, c'est-à-dire si l'estimé non contraint vérifie la contrainte. On rejette l'hypothèse nulle si la statistique est suffisamment grande.

$$H_0 : R\hat{\beta} = r \quad \text{contre} \quad H_1 : R\hat{\beta} \neq r$$

$$W = (R\hat{\beta} - r)'[V(R\hat{\beta} - r)]^{-1}(R\hat{\beta} - r)$$

Où $V(R\hat{\beta} - r)$ est la variance entre les deux termes. Ici, la variance est une pondération par l'échelle, car on ne travaille pas en unité de 1.

En décomposant la variance, on peut démontrer que le Wald s'écrit :

$$W = \frac{(R\hat{\beta} - r)'CR(R\hat{\beta} - r)}{\sigma^2}$$

Sous l'hypothèse nulle, W suit aussi une loi du khi carré (χ^2) avec q degrés de liberté asymptotiquement.

2.4 La Statistique LR

Wooldridge présente la statistique au chapitre 17. L'application est plus avancée que la forme générale que nous avons en ce moment, mais la base est la même.

Intuition : comparer la valeur de la vraisemblance au maximum contraint et non contraint. Si ces deux valeurs sont proches l'une de l'autre, cela implique que la contrainte est aisément satisfaite par les observations et non coûteuse en termes de maximisation de la vraisemblance.

Calculons $\hat{\mathcal{L}} = \mathcal{L}(\hat{\beta}, \hat{\sigma}^2)$ et $\hat{\mathcal{L}}_c = \mathcal{L}(\hat{\beta}_c, \hat{\sigma}_c^2)$. Avec l'hypothèse de normalité, on a que :

$$\mathcal{L} = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)$$



$$\hat{\mathcal{L}} = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \hat{u}'\hat{u}$$

et donc

$$\hat{\mathcal{L}} = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\hat{\sigma}^2) - \frac{T}{2}$$

Et

$$\hat{\mathcal{L}}_c = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\hat{\sigma}_c^2)$$

On ne veut pas spécifiquement dériver les estimateurs des modèles contraints et non contraints, mais plutôt comparer la valeur de la vraisemblance $\hat{\mathcal{L}}$ et $\hat{\mathcal{L}}_c$. La statistique

$$LR = 2[\hat{\mathcal{L}} - \hat{\mathcal{L}}_c] = T \ln \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_c^2} \right)$$

Cette statistique sous aussi une loi du khi carré (χ^2) avec q degrés de liberté asymptotiquement sous l'hypothèse nulle.

2.5 Le Multiplicateur de Lagrange (LM)

[Wooldridge](#) 5.2a pour le cas plus simple des tests de q exclusions. Dans ce cas, vous verrez que le test revient à faire une régression des résidus de la régression du modèle contraint sur tous les régresseurs. Il s'agit d'une régression auxiliaire. La statistique de test sera : $T \cdot R^2$. Cette statistique suit alors une loi du khi carré (χ^2) avec q degrés de liberté.

Intuition de ce test : vérifier si le score (la dérivée première du Lagrangien en fonction des paramètres) est proche de zéro évalué en $\hat{\beta}_c$. Si oui, la contrainte n'est pas très coûteuse en termes d'optimisation et il est probable que les paramètres prennent les valeurs définies par l'hypothèse nulle.

$$LM = \left(\frac{\partial \mathcal{L}}{\partial \beta} \Big|_{\hat{\beta}_c} \right)' \left[V \left(\frac{\partial \mathcal{L}}{\partial \beta} \Big|_{\hat{\beta}_c} \right) \right]^{-1} \left(\frac{\partial \mathcal{L}}{\partial \beta} \Big|_{\hat{\beta}_c} \right)$$

Sachant que (nous avons fait cette dérivation plus haut) :



$$\left. \frac{\partial \mathcal{L}}{\partial \beta} \right|_{\hat{\beta}_c} = -\frac{1}{\sigma^2} X' \hat{u}_c$$

Le score s'écrit en fonction des résidus estimés :

$$\left. \frac{\partial \mathcal{L}}{\partial \beta} \right|_{\hat{\beta}_c} = -\frac{1}{\sigma^2} X' \hat{u}_c$$

Remarquez qu'ici la statistique s'obtient avec seulement les estimés contraints.

On peut démontrer que :

$$LM = \frac{(R\hat{\beta} - r)'CR(R\hat{\beta} - r)}{\sigma^2}$$

Et donc

$$LM = \frac{(R\hat{\beta} - r)'CR(R\hat{\beta} - r)}{\sigma^2}$$

En remplaçant r par $R\hat{\beta}_c$, on a :

$$LM = \frac{(R\hat{\beta} - R\hat{\beta}_c)'CR(R\hat{\beta} - R\hat{\beta}_c)}{\sigma^2} = \frac{(R(\hat{\beta} - \hat{\beta}_c))'CR(R(\hat{\beta} - \hat{\beta}_c))}{\sigma^2}$$

On peut aussi réécrire cette statistique comme une fonction des résidus contraints et non contraints :

$$LM = \frac{\hat{u}_c' \hat{u}_c - \hat{u}' \hat{u}}{\hat{u}' \hat{u}}$$

Cette statistique suit une loi du khi carré (χ^2) avec q degrés de liberté asymptotiquement sous l'hypothèse nulle.

Le LM exploite le fait que la maximisation du log vraisemblance sous la contrainte de l'hypothèse nulle revient à maximiser sans contrainte la fonction Lagrangienne associée. Dans ce contexte, le vecteur de Lagrange s'interprète aussi comme un « shadow price » soit le prix à payer pour une violation de cette contrainte. Si la contrainte est satisfaite par ces observations, alors le prix à payer sera nul ou faible, sinon il sera élevé. Le test est alors basé sur le fait que si la contrainte sous H_0 est respectée par les données, le vecteur



de Lagrange devrait être nul.

Dans le livre de [Wooldridge](#), on vous donne la forme spécifique pour un test LM en particulier qui vise à tester l'exclusion possible de q variables dans le modèle. Le test se base sur une régression auxiliaire, une méthode que nous reverrons en section trois.

Ces trois tests rejettent l'hypothèse nulle à droite, c'est-à-dire si la valeur calculée de la statistique de test est plus grande que le point critique de la distribution.

2.5.1 Lien entre les Statistiques F, Wald, LR et LM

En remarquant que toutes ces statistiques se calculent à partir de la somme des résidus au carré, il est possible d'exprimer les trois statistiques présentées comme une transformation de la statistique F dans le cas de contraintes linéaires et du modèle linéaire simple.

Avec

$$F = \frac{\hat{u}'_0 \hat{u}_0 - \hat{u}' \hat{u}}{q} \cdot \frac{T - K}{\hat{u}' \hat{u}} > F_{(q, T-K; \alpha)}$$

On rejette H_0 si la statistique de test est plus grande que le point critique associé.

$$Wald = \frac{\hat{u}'_c \hat{u}_c - \hat{u}' \hat{u}}{\hat{u}' \hat{u}} = \frac{Tq}{T - K} F$$

$$LR = T \ln \left(\frac{\hat{u}' \hat{u}}{\hat{u}'_c \hat{u}_c} \right) = T \ln \left[F \frac{T - K}{q} + 1 \right]$$

$$LM = \frac{\hat{u}'_c \hat{u}_c - \hat{u}' \hat{u}}{\hat{u}' \hat{u}}$$

Rejette si la statistique de test est plus grande que le point critique d'une loi du khi carré (χ^2) avec q degrés de liberté.

Il faut faire attention, car bien que toutes ces statistiques soient maintenant exprimées en fonction des résidus contraints et non contraints, elles ne donnent pas la même valeur. L'inférence pourrait donc potentiellement être différente.



3 Bibliographie

Références

- [1] Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill.
- [2] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- [3] Greene, W. H. (2012). *Econometric Analysis*. Pearson.

