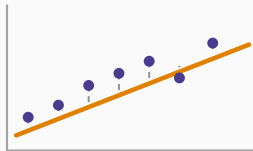


Introduction à l'économétrie

Chapitre 18 : Normalité et points influents



Jaouad Madkour

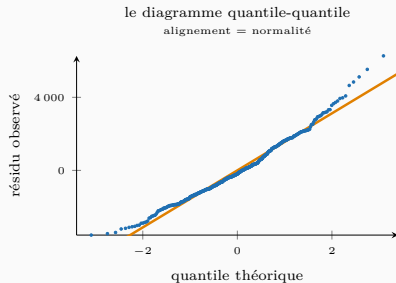
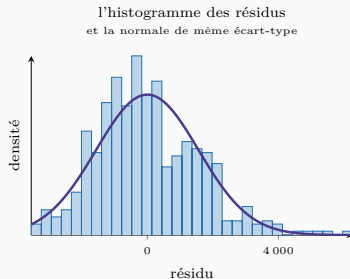
Faculté d'Économie et de Gestion de Tanger · 27 août 2026

La normalité des résidus

Les observations qui pèsent trop

Ce qu'il faut retenir

La normalité des résidus



Les résidus sont étirés à droite : asymétrie 0,56, aplatissement 3,61, Jarque-Bera = 34,1 contre un seuil de 5,99. La normalité est rejetée.

Et cela n'a aucune conséquence ici. Avec cinq cents observations, le théorème central limite rend les coefficients approximativement normaux quelle que soit la loi des erreurs. Tests et intervalles restent valides.

Le diagnostic n'est donc pas inutile pour autant : le décrochage dans le coin supérieur droit signale quelques salaires très élevés, dont le chapitre suivant examinera l'influence.

Il combine asymétrie et aplatissement

$$JB = n \left(\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right) \sim \chi^2_2$$

où S est le coefficient d'asymétrie et K l'aplatissement — les indicateurs de forme du chapitre 9 de la statistique descriptive.

Sur le modèle M2

$$S = 0,563 \quad K = 3,607 \quad JB = 34,1 \quad \text{contre un seuil de } 5,99$$

La normalité est rejetée. Les résidus sont étirés vers la droite.

Pourquoi on peut passer outre

Avec cinq cents observations, le théorème central limite rend $\hat{\beta}_j$ approximativement normal **quelle que soit la loi des erreurs**.

Tests, p -values et intervalles restent valides. **H6 ne servait qu'aux petits échantillons.**

Quand elle importe vraiment

En dessous de trente observations — bilans d'un petit échantillon d'entreprises, séries macro courtes, expériences de laboratoire — la non-normalité invalide les tests.

Il faut alors soit transformer la variable expliquée, soit recourir à des méthodes de rééchantillonnage. **Regardez toujours n avant de décider si le rejet vous concerne.**

Un diagnostic utile, pour une autre raison

Le décrochage en haut à droite du diagramme quantile-quantile signale quelques salaires très supérieurs à ce que le modèle prédit.

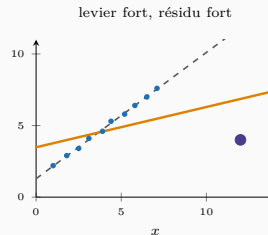
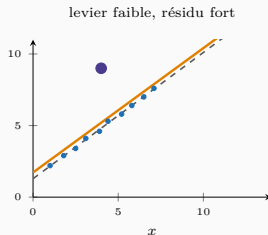
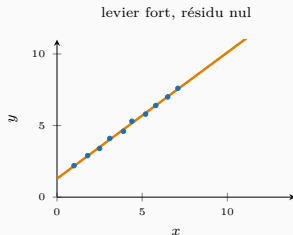
Ce n'est pas un problème de normalité : c'est une poignée d'observations particulières. **Et celles-là méritent qu'on les regarde.**

L'asymétrie appelle souvent le logarithme

Une distribution de résidus étirée à droite accompagne presque toujours une variable expliquée elle-même étirée — salaires, revenus, patrimoines, chiffres d'affaires.

Passer au logarithme corrige simultanément l'asymétrie **et** l'hétéroscédasticité du chapitre 15. C'est l'objet du chapitre 19.

Les observations qui pèsent trop



Trois points ajoutés, trois situations très différentes. La droite en pointillé est celle d'origine.

Levier : le point est-il loin des autres *en x* ? **Résidu** : est-il loin de la droite *en y* ?

Seule la combinaison des deux est dangereuse. Un levier fort sans résidu confirme la droite ; un gros résidu au centre la déplace à peine. C'est le troisième cas qui bascule tout — et c'est le seul qu'un simple examen des résidus ne signale pas comme extraordinaire.

D'où la distance de **Cook**, qui combine les deux, et la règle usuelle : examiner tout point dont la distance dépasse $4/n$. Non pour le supprimer — une observation ne se supprime pas parce qu'elle dérange — mais pour comprendre *pourquoi* il est là.

À ne pas confondre

Notion	Question
Résidu	le point est-il loin de la droite, en y ?
Levier	est-il loin des autres, en x ?
Influence	son retrait changerait-il la droite?

Seule la troisième compte

Un levier fort sans résidu **confirme** la droite. Un gros résidu au centre du nuage la déplace à peine.

C'est leur combinaison qui bascule tout — et c'est le seul cas qu'un simple examen des résidus ne signale pas comme extraordinaire.

Elle combine les deux

$$D_i = \frac{e_i^2}{(k+1) \hat{\sigma}_\varepsilon^2} \cdot \frac{h_i}{(1-h_i)^2}$$

où h_i est le levier de l'observation i . Elle mesure de combien **l'ensemble des prévisions** changerait si l'on retirait ce point.

Le seuil usuel

$$D_i > \frac{4}{n}$$

Sur nos cinq cents salariés : $D_i > 0,008$. C'est un seuil d'**alerte**, non de suppression.

Quatre questions, dans cet ordre

1. Est-ce une **erreur de saisie** ? On la corrige.
2. Appartient-il à la **population étudiée** ? Un dirigeant dans un fichier d'employés n'y a pas sa place.
3. Le modèle est-il **mal spécifié** ? Une non-linéarité produit des points apparemment aberrants.
4. Si rien de tout cela : **c'est une observation, et on la garde.**

La règle déontologique

On ne supprime pas une observation parce qu'elle dérange.

La bonne pratique consiste à estimer le modèle avec et sans, et à publier les deux si les conclusions diffèrent. C'est ce qu'on appelle un test de **robustesse** — et une conclusion qui n'y survit pas n'était pas une conclusion.

Ce qu'il faut retenir

Six points

1. **Jarque-Bera** combine asymétrie et aplatissement : **34,1** ici, la normalité est rejetée.
2. **Sans conséquence** sur cinq cents observations : le théorème central limite suffit.
3. En dessous de trente, elle est **décisive** — regardez toujours n .
4. Une asymétrie des résidus appelle souvent le **logarithme**, qui corrige aussi l'hétéroscédasticité.
5. **Résidu, levier, influence** : seule la troisième compte, et c'est la distance de Cook.
6. On ne supprime pas une observation gênante : on estime **avec et sans**, et l'on publie les deux.

La suite

Trois chapitres ont conclu de la même manière : le logarithme réglerait le problème.

Il est temps de le prendre au sérieux. La sixième partie traite le **choix de la forme** — et l'on y retrouvera l'équation la plus estimée de toute l'économie appliquée.