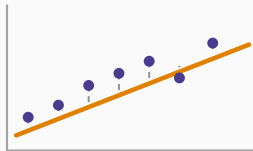


Introduction à l'économétrie

Chapitre 7 : La prévision



Jaouad Madkour

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

Deux questions, pas une

Les deux intervalles

Les pièges de la prévision

Ce qu'il faut retenir

Deux questions, pas une

Elle est immédiate

$$\hat{y}_h = \hat{\beta}_0 + \hat{\beta}_1 x_h$$

On remplace x par la valeur voulue dans la droite estimée.

Sur nos salaires

Pour dix ans d'ancienneté :

$$\hat{y} = 4\,347,23 + 277,17 \times 10 = 7\,118,93 \text{ dirhams}$$

Deux questions se cachent derrière ce nombre

« Quel est le salaire moyen des salariés ayant dix ans d'ancienneté ? »

« Quel sera le salaire de Karim, qui a dix ans d'ancienneté ? »

La réponse ponctuelle est la même : 7 119. La **précision** ne l'est pas du tout.

Pourquoi la seconde est plus incertaine

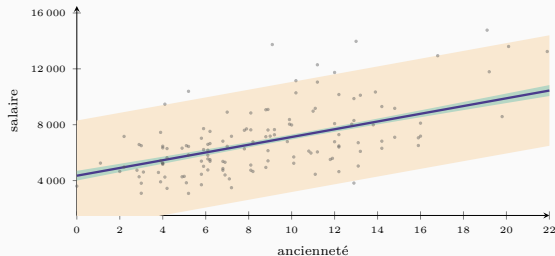
Prévoir une moyenne, c'est ignorer seulement où passe la vraie droite.

Prévoir un individu, c'est ignorer cela **et** son écart personnel à la droite — son ε , que rien ne permet d'anticiper.

Les deux intervalles

La figure décide

deux intervalles, et ils n'ont pas la même largeur



Deux questions différentes

La bande étroite encadre le *salaire moyen* des salariés ayant cette ancienneté. Elle se resserre quand n grandit.

La bande large encadre le salaire d'*un* salarié donné. Elle ne se resserre jamais en dessous de $\pm 1,96 \hat{\sigma}_e$: l'individu garde son aléa propre.

$$\hat{y}_h \pm t_{\alpha/2} \hat{\sigma} \sqrt{\boxed{1} + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Le 1 encadré n'apparaît que pour la prévision individuelle. C'est toute la différence.

Les deux bandes s'**évasent** en s'éloignant de $\bar{x}$: extrapoler coûte cher.

Les deux formules

Elles ne diffèrent que par un terme

Intervalle de confiance de la moyenne :

$$\hat{y}_h \pm t_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

Intervalle de prévision d'une observation :

$$\hat{y}_h \pm t_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\boxed{1} + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

Le 1 encadré est toute la différence

Il représente la variance de l'erreur individuelle, σ_ε^2 .

Et il ne dépend pas de n . Quel que soit le nombre d'observations, l'intervalle individuel ne descendra jamais sous $\pm 1,96 \hat{\sigma}_\varepsilon$. L'aléa personnel ne se dilue pas.

Les deux, chiffrés

$$\hat{\sigma}_\varepsilon = 2\,008, n = 500, x_h = 10 \approx \bar{x}.$$

$$\text{Moyenne : } 7\,119 \pm 1,96 \times 2\,008 \times 0,045 = [6\,943 ; 7\,295]$$

$$\text{Individu : } 7\,119 \pm 1,96 \times 2\,008 \times 1,001 = [3\,179 ; 11\,059]$$

Un écart qui décide de l'usage

Le premier intervalle est étroit : on peut fonder une grille salariale dessus.

Le second est immense : il ne permet **aucune décision individuelle**. Annoncer à Karim que son salaire sera entre trois mille et onze mille dirhams n'est pas une prévision, c'est un aveu.

Les pièges de la prévision

Les bandes s'évasent

Le terme $(x_h - \bar{x})^2$ grandit dès qu'on s'écarte du centre : plus on extrapole, moins on est précis. C'est visible sur la figure.

Et ce n'est que la moitié du problème. **Rien ne garantit que la relation reste linéaire hors du domaine observé** — et là, aucune formule ne prévient.

Un exemple qui a coûté cher

Extrapoler une hausse des prix de l'immobilier observée sur dix ans revient à supposer que le mécanisme qui l'a produite continue de jouer.

Les modèles de risque immobilier d'avant **2007** n'avaient jamais observé de baisse nationale des prix. Leur domaine de validité ne contenait pas l'événement qui allait survenir.

Trois conditions, souvent tues

1. Le modèle est **bien spécifié** — les bonnes variables, la bonne forme.
2. La relation **reste stable** entre la période d'estimation et celle de prévision.
3. Le point à prévoir est **dans le domaine** des données observées.

La deuxième est la plus fragile

En économie, les comportements changent : une réforme fiscale, une crise, un changement de génération modifient la relation elle-même.

C'est la **critique de Lucas** : un modèle estimé sous un régime de politique économique cesse d'être valable si l'on change ce régime — précisément parce que les agents s'adaptent. **Le modèle ne peut donc pas servir à évaluer la politique qui le rendrait faux.**

Ce qu'il faut retenir

Six points

1. La prévision ponctuelle est immédiate : $\hat{y}_h = \hat{\beta}_0 + \hat{\beta}_1 x_h$.
2. Deux questions se cachent derrière : prévoir une **moyenne** ou prévoir un **individu**.
3. Les deux formules ne diffèrent que par un **1** sous la racine — la variance de l'erreur individuelle.
4. Ce **1** ne dépend pas de n : l'intervalle individuel ne se resserre jamais complètement.
5. Sur nos données : [6 943 ; 7 295] contre [3 179 ; 11 059].
6. **Ne jamais extrapoler** hors du domaine observé, et se méfier des changements de régime.

La suite

Une seule variable explicative ne suffit à rien décrire : le salaire dépend aussi du diplôme, de la région, du poste.

La troisième partie généralise tout ce qui précède — et introduit la notion que l'économétrie apporte vraiment à la statistique : le **ceteris paribus**.