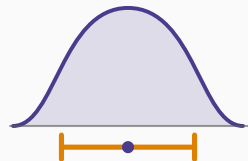


Statistique inférentielle

Chapitre 21 : Tester un coefficient



Jaouad Madkour

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

La pente est une variable aléatoire

Le test de significativité

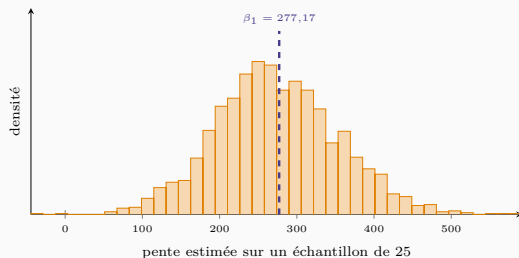
La sortie d'un logiciel

Ce qu'il faut retenir

La pente est une variable
aléatoire

Sa distribution d'échantillonnage

2 000 échantillons : la pente estimée est une variable aléatoire



La boucle est bouclée

Moyenne des 2 000 pentes : 273.89.

Écart-type : 79.56.

L'estimateur est sans biais : la cloche est centrée sur β_1 , exactement comme $\bar{X}$ l'était sur m .

Et il est asymptotiquement normal, par le théorème central limite — ce qui autorise le même intervalle qu'au chapitre 10 :

$$\hat{\beta}_1 \pm t_{\alpha/2} \hat{\sigma}(\hat{\beta}_1)$$

Sur la population entière : $\hat{\sigma}(\hat{\beta}_1) = 15,06$ et $t = 18,41$. La relation entre ancienneté et salaire est hors de doute.

Le parallèle complet

Chapitre 3

$\bar{X}$ estime m

$$E(\bar{X}) = m$$

$$\sigma(\bar{X}) = \sigma/\sqrt{n}$$

loi normale par le TCL

Chapitre 21

$\hat{\beta}_1$ estime β_1

$$E(\hat{\beta}_1) = \beta_1$$

$$\sigma(\hat{\beta}_1) = \sigma_\varepsilon / \sqrt{\sum (x_i - \bar{x})^2}$$

loi normale par le TCL

La conséquence

Tout ce que les chapitres 10 et 16 ont fait pour la moyenne s'applique à la pente **sans une formule nouvelle**.

Il suffit de remplacer $\bar{x}$ par $\hat{\beta}_1$ et l'erreur-type par la sienne.

Deux estimations emboîtées

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_i e_i^2}{n-2} \qquad \hat{\sigma}(\hat{\beta}_1) = \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum_i (x_i - \bar{x})^2}}$$

Pourquoi $n - 2$

La règle du chapitre 4 : **un degré de liberté par paramètre estimé.**

Ici deux, $\hat{\beta}_0$ et $\hat{\beta}_1$. Le dénominateur était $n - 1$ pour une variance simple; il sera $n - k - 1$ pour une régression à k variables explicatives. Le principe ne change jamais.

Le test de significativité

La variable explicative sert-elle à quelque chose

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

Sous H_0 , l'ancienneté n'a aucun effet sur le salaire : la droite vraie est horizontale.

La statistique

$$t = \frac{\hat{\beta}_1 - 0}{\hat{\sigma}(\hat{\beta}_1)} \sim \text{Student à } n - 2 \text{ degrés de liberté}$$

C'est exactement la forme du chapitre 13 : *(observé — prévu par H_0) divisé par l'erreur-type.*

Sur nos cinq cents salariés

$$\hat{\sigma}_\varepsilon = 2008,34 \quad \hat{\sigma}(\hat{\beta}_1) = 15,06$$

$$t = \frac{277,17}{15,06} = 18,41 \quad \text{ddl} = 498 \quad p < 0,0001$$

On rejette H_0 : l'ancienneté explique significativement le salaire.

L'intervalle de confiance de la pente

$$277,17 \pm 1,96 \times 15,06 = [247,65 ; 306,68]$$

Chaque année d'ancienneté vaut entre 248 et 307 dirhams de salaire mensuel.

Ce qui est utilisable dans un rapport

« Une année d'ancienneté supplémentaire s'accompagne d'une hausse de salaire estimée à 277 dirhams, avec une marge de ± 30 dirhams au seuil de 95 %. »

C'est l'intervalle, non la p -value, qui permet de **négoier une grille**. Le test dit seulement qu'il y a bien un effet à discuter.

Significatif ne veut toujours pas dire tout expliquer

$t = 18,41$ est écrasant. Et pourtant $R^2 = 0,405$: l'ancienneté n'explique que 40 % de la dispersion des salaires.

Un coefficient très significatif dans un modèle médiocre reste un coefficient dans un modèle médiocre. Les deux indicateurs répondent à deux questions différentes.

La sortie d'un logiciel

Le tableau standard

	Coefficient	Erreur-type	t	p
Constante	4 347,23	177,18	24,54	< 0,0001
Ancienneté	277,17	15,06	18,41	< 0,0001

Et les indicateurs globaux

$$R^2 = 0,405 \quad \hat{\sigma}_\varepsilon = 2\,008,34 \quad F(1; 498) = 338,8$$

Ils disent la même chose

$$18,41^2 = 338,8$$

En régression simple, le F global **est** le carré du t de la pente : une seule variable, un seul test.

La distinction ne prend son sens qu'en régression multiple, où le F teste toutes les variables ensemble et chaque t une variable à la fois.

Ce que le cours suivant ajoutera

Plusieurs variables explicatives, la multicollinéarité, les variables qualitatives, l'hétéroscédasticité, l'autocorrélation, l'endogénéité.

Aucun de ces sujets n'exige un outil statistique nouveau. Ils appliquent tous ce cours à des situations où l'une des quatre hypothèses du chapitre 20 tombe.

Ce qu'il faut retenir

Six points

1. $\hat{\beta}_1$ est une variable aléatoire, sans biais et asymptotiquement normale – comme $\overline{X}$.
2. $\hat{\sigma}_\varepsilon^2 = \sum e_i^2 / (n - 2)$: deux degrés de liberté perdus, deux paramètres estimés.
3. Test de significativité : $t = \hat{\beta}_1 / \hat{\sigma}(\hat{\beta}_1)$, loi de Student à $n - 2$ ddl.
4. Sur nos données : $t = 18,41$, $p < 0,0001$, et $IC_{95\%} = [247,65 ; 306,68]$.
5. C'est l'**intervalle**, non la p -value, qui sert à décider d'une grille salariale.
6. En régression simple, $F = t^2$: les deux tests n'en font qu'un.

La suite

Tout le cours a été conduit à la main, pour que chaque nombre soit vérifiable.

Le chapitre 22 refait l'ensemble en quelques lignes de R et de Python – et montre que les logiciels n'affichent rien d'autre que ce que nous venons de calculer.