# TD 3 — Régressions sur EViews : lecture des résultats, tests t & F

Pierre Beaucoral

2025-09-15

## Introduction

Ce document reprend le **contenu de TD 3** sous forme de **cours**. On revoit les éléments fondamentaux de la **régression linéaire**, la **décomposition ANOVA**, les indicateurs de **qualité d'ajustement** ( $R^2$, $\bar{R}^2$ ), puis les **tests t** et **F**. Une section explique l'**importance économique** des variables et la lecture des **coefficients standardisés**. Les **questions / réponses** du TD sont fournies à la fin.

## Rappels sur la régression linéaire

La régression linéaire sert à **quantifier et tester la relation entre une variable expliquée (Y)** et **une ou plusieurs variables explicatives (X)**.

En pratique, elle permet de:

- **Décrire** : mesurer la force et le sens d'un lien (ex. hausse du revenu → hausse de la consommation).
- **Prédire** : estimer la valeur attendue de Y pour de nouvelles valeurs de X.
- **Expliquer** : isoler l'effet propre de chaque facteur en contrôlant les autres.
- **Tester** : vérifier des hypothèses (par exemple $H_0 : \beta_j = 0$ ) grâce aux tests t ou F.

  En résumé, c'est un outil central pour **analyser et interpréter des données**, évaluer l'importance relative des déterminants d'un phénomène et faire des **prévisions fondées**.

### Modèle simple et interprétation de $\beta_0$ et $\beta_1$

On suppose : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, N$, où $\varepsilon_i$ est centré et non corrélé aux régresseurs.
**Objectif MCO (OLS)** : minimiser $\sum_i (Y_i - \beta_0 - \beta_1 X_i)^2$ .

**Interprétation** :
- $\beta_0$ est l'**ordonnée à l'origine** (valeur de (Y) quand (X=0)).
- $\beta_1$ est la **pente** (variation moyenne de (Y) quand (X) augmente d'une unité).

#### Illustration graphique

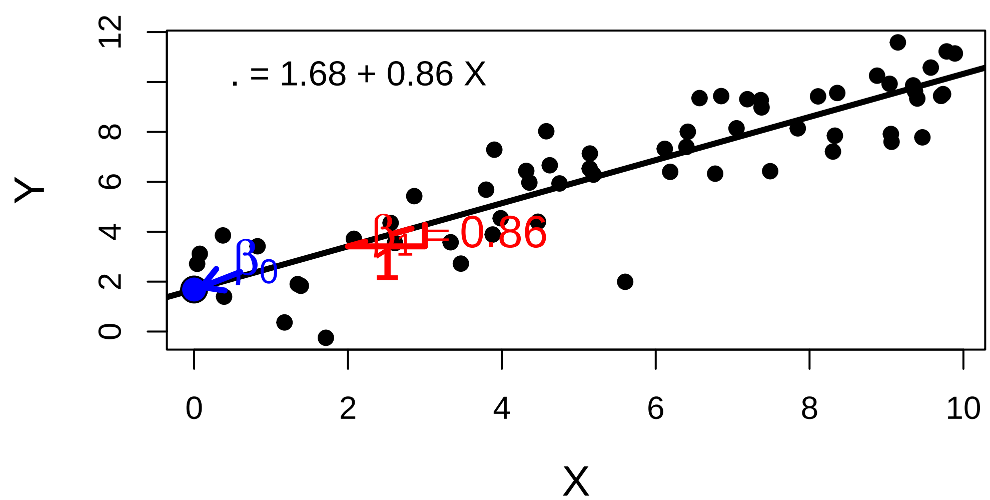


Figure 1: (interception) et pente illustrée par un triangle rectangle (Δx = 1, Δy = ).

### Forme matricielle et estimateur OLS

En multiple : $Y = X\beta + \varepsilon$ et $\hat{\beta} = (X'X)^{-1}X'Y$ .
Les **valeurs ajustées** sont $\hat{Y} = X\hat{\beta}$ et les **résidus** $\hat{\varepsilon} = Y - \hat{Y}$ .

La **forme matricielle** de la régression linéaire : $Y = X\beta + \varepsilon$ n'est pas juste un "raccourci d'écriture", elle a plusieurs implications importantes :

#### Écriture compacte et générale

- Elle **englobe en une seule équation** le modèle avec plusieurs variables explicatives et plusieurs observations.
- Que l'on ait 2 ou 200 régresseurs, la notation reste la même.

- Le vecteur $Y$ contient toutes les observations de la variable dépendante, la matrice $X$ toutes les observations de toutes les variables explicatives (y compris une colonne de 1 pour l'intercept).

#### Estimation par l'algèbre linéaire

- L'estimateur OLS $\hat{\beta} = (X'X)^{-1}X'Y$ se déduit **directement** des règles de dérivation matricielle (minimisation de la somme des carrés).
- Cette formule montre les **conditions d'existence** : la matrice $X'X$ doit être **inversible** $\rightarrow$ donc les colonnes de $X$ (les variables explicatives) doivent être **linéairement indépendantes** (pas de multicolinéarité parfaite).

#### Propriétés géométriques

- Le vecteur des valeurs ajustées $\hat{Y} = X\hat{\beta}$ est la **projection orthogonale** de $Y$ sur l'espace engendré par les colonnes de $X$.
- Les résidus $\hat{\varepsilon} = Y - \hat{Y}$ sont **orthogonaux** à cet espace :
  - $X'\hat{\varepsilon} = 0$ .
  - $\rightarrow$ Les variables explicatives ne sont pas corrélées aux résidus.

---

Extension à d'autres modèles

- Cette écriture facilite les **généralisations** : régression multiple, modèles de panel, régressions pondérées, moindres carrés généralisés, etc.
- Elle permet d'utiliser directement les outils de l'algèbre linéaire (décomposition en valeurs propres, moindres carrés ordinaires ou généralisés).

## Décomposition ANOVA et qualité d'ajustement

L'**ANOVA** (pour *ANalysis Of VAriance*, ou **analyse de la variance**) est une méthode statistique qui décompose la variabilité totale d'une variable en plusieurs composantes, afin de comparer et de tester les effets de différents facteurs.

Dans le cadre de la **régression linéaire**, l'ANOVA sert à expliquer d'où vient la variance observée de la variable dépendante $Y$ :

## ANOVA : SCT = SCE + SCR

La somme des carrés **totale** (SCT) se décompose en somme des carrés **expliquée** (SCE) et somme des carrés des **résidus** (SCR) : $\sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i (Y_i - \hat{Y}_i)^2$.

### Schéma visuel

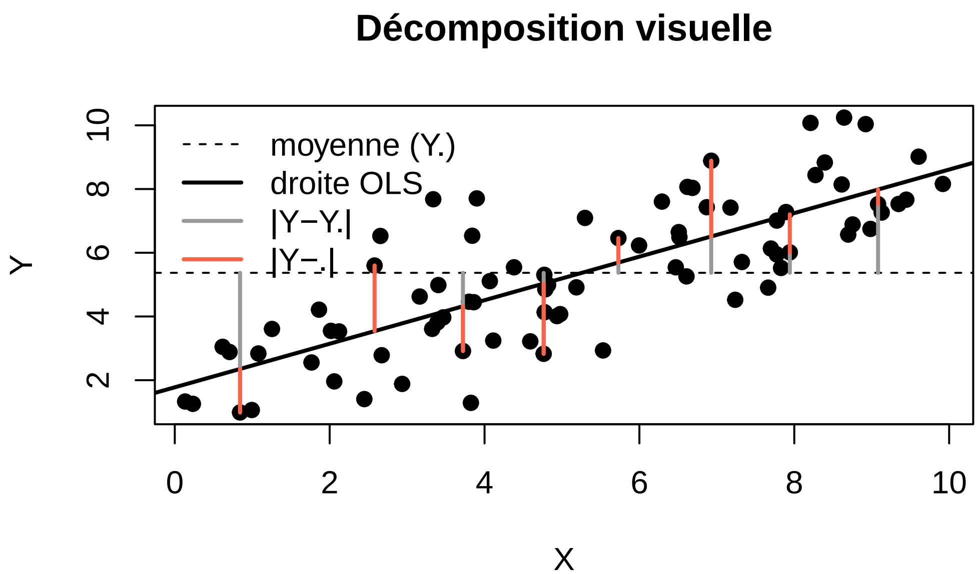


Figure 2: Décomposition ANOVA : SCT = SCE + SCR.

## $R^2$ et $\bar{R}^2$

$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}}, \quad \bar{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-p}$ .

- $R^2$ mesure la part de variance expliquée.
- $\bar{R}^2$ pénalise l'ajout de régresseurs superflus.
- **Attention** : on ne cherche pas à **maximiser** $R^2$ en ajoutant des variables sans justification.

# Tests d'hypothèses

## Utiliser un test d'hypothèse avec une table de valeurs critiques

Pour utiliser un **test d'hypothèse** avec une **table de valeurs critiques** (table t de Student, table F, table du Khi-deux…), on suit toujours la même logique en 4 étapes :

---

### Formuler les hypothèses

- **Hypothèse nulle** $H_0$ : ce qu'on veut tester (ex. $\beta = 0$, « les moyennes sont égales »).
- **Hypothèse alternative** $H_1$ : ce qu'on conclut si $H_0$ est rejetée (ex. $\beta \neq 0$).

Préciser si le test est :

- **bilatéral** : on rejette si la statistique est trop grande en valeur absolue.
- **unilatéral** : on rejette seulement dans une queue.

---

### Choisir le niveau de risque

Fixer le **niveau de signification** $\alpha$, par exemple 5 % $0,05$.
Cela correspond au **risque d'erreur de type I** (rejeter $H_0$ alors qu'elle est vraie).

---

### Calculer la statistique de test

À partir de vos données :

- pour un **test t** : $t = \frac{\hat{\beta}-\beta_0}{\widehat{se}(\hat{\beta})}$
- pour un **test F** : $F = \frac{(\text{SCE}/q)}{(\text{SCR}/(N-p))}$

… ou la statistique adaptée au test choisi.

---

#### Comparer à la table

1. Chercher dans la **table de la loi correspondante** (t, F, …) la valeur critique :
   - connaître les **degrés de liberté** (ex. $N - p$ pour $t$, $q$ et $N - p$ pour $F$) ;
   - choisir la colonne de $\alpha$ (ou $\alpha/2$ pour un test bilatéral).
2. **Décision** :
   - **bilatéral** : rejeter $H_0$) si |statistique| $> t^*_{\alpha/2}$.
   - **unilatéral à droite** : rejeter $H_0$ si statistique $> t^*_{\alpha})ou(F > F^*_{\alpha}$.

### Test t (significativité individuelle)

On teste typiquement $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$.
Statistique : $t = \hat{\beta}_j/\widehat{se}(\hat{\beta}_j)$ .
**Décision bilatérale** au seuil $\alpha$ : rejeter (H_0) si $|t| > t^*_{\alpha/2,\,\nu}$.

#### Visualisation du test t (bilatéral)

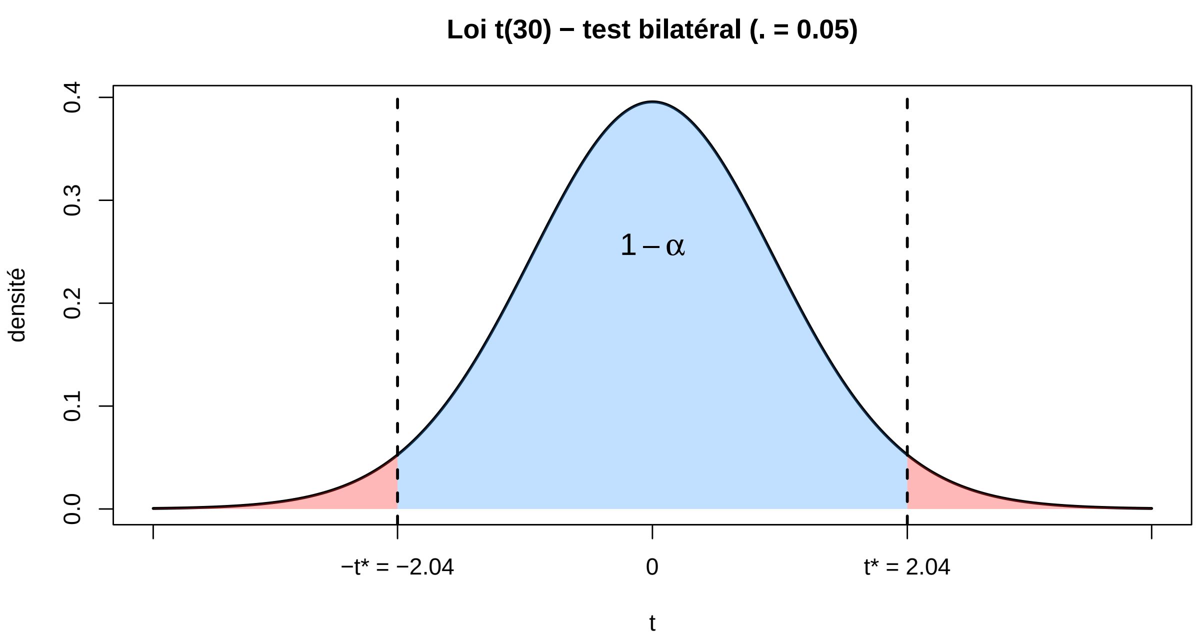


Figure 3: Test bilatéral :  /2 décalés vers l'extérieur, -t* et t* en graduations de l'axe X.

### Test F (significativité conjointe)

On teste $H_0$ : **un ensemble** de coefficients (= 0) (excluant la constante).
Sous $H_0$\$, $F \sim F(q,\ N-p)$ (queue **droite**).

#### Visualisation du test F

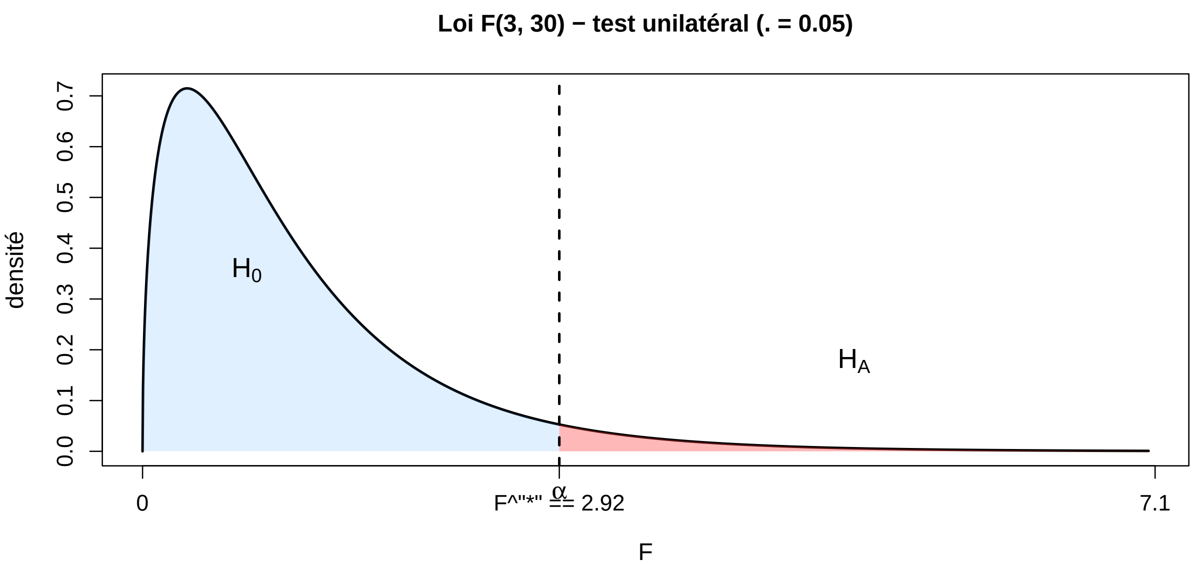


Figure 4: Loi F(3,30) — test unilatéral (queue droite).

## Importance économique et coefficients standardisés

Une variable peut être **statistiquement significative** mais **économiquement peu pertinente**. Pour évaluer l'**importance économique** :

1. **Comparer les coefficients standardisés** (EViews : `View → Coefficient diagnostics → Scaled coefficient`).
   → effet en **écarts-types** (comparaison **relative** entre variables).

2. **Interpréter l'ordre de grandeur** selon la **forme fonctionnelle** :

   - **Log–lin** (Y en log, X en niveau) : 1 unité de (X)   **%** de variation de (Y).
   - **Log–log** : élasticité (1 % de (X) → ( ) % de (Y)).
   - **Lin–log** : 1 % de (X) → variation absolue de (Y).

**Exemple du TD** : la variable **intercontinentale** a le plus grand **coef. standardisé** (~0,42). Comme (Y) est en log et (X) en niveau, devenir **intercontinental** est associé à **+10,5 %** de passagers (ceteris paribus).

## Questions – Réponses (TD2)

> Les questions sont celles de `module2.docx` (Module 2). Les réponses décrivent la **procédure EViews** et l'interprétation.

---

**Question : Importez la base de données sur les compagnies aériennes.**

Afficher la réponse

Menu **File → Open → Foreign Data as Workfile** puis sélectionner le fichier.

---

**Question : Créez le logarithme du nombre de passagers (passagers). Quelle est l'utilité de cette transformation ?**

Afficher la réponse

Commande :

```
genr logpassagers = log(passagers)
```

**Utilités** : stabiliser les variances, réduire l'influence des valeurs extrêmes, faciliter une **lecture en %** et souvent **linéariser** la relation.

---

**Question : Estimez l'équation suivante par les MCO. Dans quelle mesure cette équation peut-elle être considérée comme linéaire ?**

Afficher la réponse

**Object → New Object → Equation → LS (Least Squares)**.
Une équation est **linéaire en paramètres** si les coefficients sont à la **puissance 1** et s'additionnent (les logs n'empêchent pas la linéarité en paramètres).

---

**Question : Distinguez les variables dépendantes, indépendantes, d'intérêt et de contrôle.**

Afficher la réponse

- **Dépendante** : `logpassagers`.
- **Indépendantes** (ex.) : `ratio`, `public`, `low_cost`, `age`, `intercontinental`, croissance trafic **destination** (2010–2013), croissance trafic **pays d'origine** (2010–2013).
- **D'intérêt** : `ratio` (ou sa déclinaison en mortels / non mortels).
- **Contrôle** : les autres explicatives listées.

---

**Question : D'après le $R^2$ de l'estimation, l'équation a-t-elle un pouvoir explicatif correct ?**

Afficher la réponse

Lire **R-squared** et **Adjusted R-squared**. Par ex., un $R^2$ 0,39 indique un **pouvoir explicatif modéré** ; $\bar{R}^2$ pénalise les variables superflues.

---

**Question : Le nombre d'accidents par passagers est-il significativement différent de zéro ?**

Afficher la réponse

Regarder **t-Statistic** et **Prob.** (p-value). En bilatéral, si $p <$ ( ) (ou |t| > t*), on **rejette** (H_0: =0). (Ex. : (|t| 3,22 > 1{,}658)).

---

**Question : Distinguer accidents mortels et non mortels et réestimer l'équation.**

Afficher la réponse

Créer deux variables (ou utiliser celles existantes) et réestimer en remplaçant `ratio` par **mortels** et **non mortels** par passager.

---

**Question : Ces variables sont-elles individuellement et conjointement significatives ?**

Afficher la réponse

- **Individuellement** : lire **t-Statistic / Prob.**.
- **Conjointement** : **View → Coefficient Diagnostics → Wald test** (ou **test F** entre modèles emboîtés) pour tester que **les deux coefficients = 0**.

---

**Question : Quelle variable semble la plus importante d'un point de vue économique ? Comment interpréter le coefficient obtenu ?**

Afficher la réponse

Procédure : **View → Coefficient diagnostics → Scaled coefficient**.
Résultat : **intercontinental** a le **plus grand coef. standardisé** (~0,42).
Interprétation (log–lin) : devenir intercontinental est associé à **+10,5 %** de passagers, toutes choses égales par ailleurs.