

Économétrie — TD 3

Regressions MCO (EViews)

Régression linéaire : rappel & pratique

Modèle de régression linéaire simple

Une régression consiste à expliquer les variations d'une variable dépendante Y par celles d'une ou plusieurs variables indépendantes X .

On suppose la relation (droite de régression) :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, N$$

où ε_i est centré et non corrélé aux régresseurs.

Objectif MCO (OLS). Estimer β_0, β_1 en minimisant la somme des carrés :

$$\min \sum_{i=0}^N \varepsilon_i^2 = \min \sum_{i=0}^N (Y_i - \beta_0 - \beta_1 X_i)^2.$$

Tip

β_1 Représente ici la magnitude de “l’effet” de la variable X_1 sur Y . ε_i Représente la partie non expliquée de la relation (ou terme d’erreur)

Terme d'erreur

L'introduction du terme d'erreur recouvre deux grands types d'erreurs :

Erreur de spécification :

Les variables introduites ne sont pas suffisantes pour expliquer toutes les variations de Y

Erreur de mesure :

La variable expliquée (Y) est mesurée de manière imparfaite (bruitée)

Prédiction de Y

Une fois estimée, la relation s'écrit:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i$$

Avec $\hat{\varepsilon}_i$, terme d'erreur estimé aussi appelé résidu, on peut prédire Y:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Graphique 1 — Nuage de points + droite OLS

```
set.seed(42)
N <- 60
x <- sort(runif(N, 0, 10))
y <- 2 + 0.8*x + rnorm(N, sd = 1.5)
df <- data.frame(x, y)
mod <- lm(y ~ x, data = df)

plot(df$x, df$y, pch = 19, xlab = "X", ylab = "Y", cex.lab = 1.4)
abline(mod, lwd = 3)

b0 <- coef(mod)[1]
b1 <- coef(mod)[2]

## ---- 0 : ordonnée à l'origine ----
points(0, b0, pch = 21, bg = "blue", cex = 1.8)
arrows(0.6, b0 + 0.7, 0.1, b0 + 0.1, length = 0.12,
       col = "blue", lwd = 3)
text(0.8, b0 + 1.1, expression(beta[0]),
     col = "blue", cex = 1.6, font = 2)

## ---- Triangle rectangle pour la pente ----
x0 <- 2 # point de départ en X
y0 <- b0 + b1*x0 # point sur la droite
```

```

# Base de 1 en X et hauteur correspondante en Y = b1
x1 <- x0 + 1
y1 <- b0 + b1*x1

# Triangle
segments(x0, y0, x1, y0, col="red", lwd=3) # base (Δx = 1)
segments(x1, y0, x1, y1, col="red", lwd=3) # hauteur (Δy = 1)
segments(x0, y0, x1, y1, col="red", lwd=3, lty=2) # hypoténuse

# Étiquettes
text((x0+x1)/2, y0 - 0.7, "1", col="red", cex=1.4, font=2) # Δx
text(x1 + 0.4, (y0 + y1)/2,
     bquote(beta[1] == .(round(b1,2))),
     col="red", cex=1.4, font=2) # Δy = 1

legend("topleft",
      legend = sprintf("Ŷ = %.2f + %.2f X", b0, b1),
      bty = "n", cex = 1.2)

```

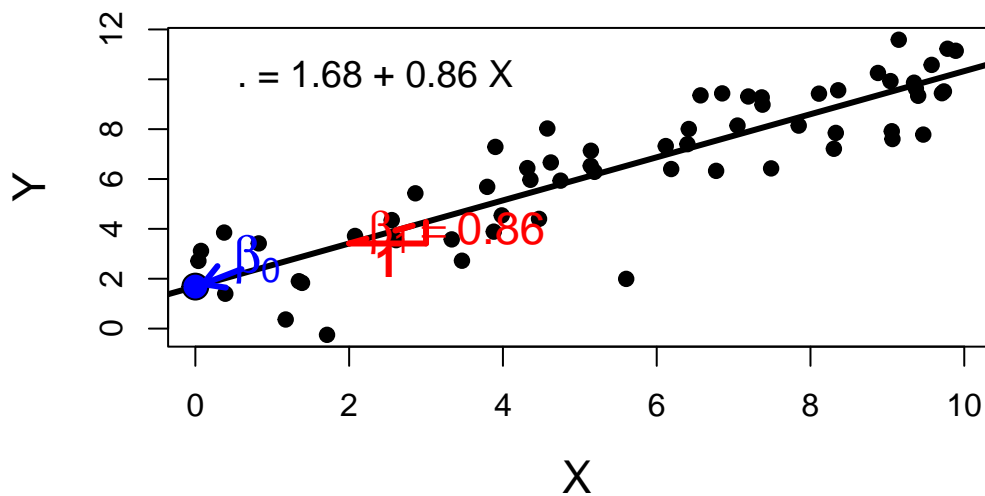


Figure 1: et illustrée par un triangle rectangle ($\Delta x = 1$, $\Delta y = 0.86$).

Forme matricielle (régression multiple)

En multiple : $Y = X\beta + \varepsilon$, $\hat{\beta} = (X'X)^{-1}X'Y$.

Exemple avec deux régresseurs

```
set.seed(123)
N <- 120
x1 <- runif(N, 0, 10)
x2 <- rnorm(N, 5, 2)
y <- 1.5 + 0.6*x1 - 0.3*x2 + rnorm(N, sd = 1)
d <- data.frame(y, x1, x2)
m <- lm(y ~ x1 + x2, data = d)
summary(m)
```

Call:

```
lm(formula = y ~ x1 + x2, data = d)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.78604	-0.62855	-0.05144	0.66068	2.06691

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.87289	0.26031	7.195	6.43e-11 ***
x1	0.61541	0.03088	19.931	< 2e-16 ***
x2	-0.37545	0.04565	-8.224	3.07e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9416 on 117 degrees of freedom

Multiple R-squared: 0.7792, Adjusted R-squared: 0.7755

F-statistic: 206.5 on 2 and 117 DF, p-value: < 2.2e-16

Note

Ceci est un exemple de tableau de régression, sur Eviews, le tableau sera similaire, mais un peu différent, celui-là a été fait dans R pour l'exemple.

Procédure sur EViews

Estimation

- Faire Object → New Object → Equation
 - Autre méthode :
 - 1. Sélectionner les variables en débutant par la variable dépendante (Y)
 - 2. Faire Open → as Equation
 - La fenêtre ouverte a deux onglets :
Specification : Entrer la spécification choisie
Options :
 - Cet onglet sert pour la correction de la matrice de variance-covariance
 - Nous ignorons pour le moment cet onglet
-

Estimation

- Equation specification : Permet d'entrer l'équation estimée
Il faut mettre d'abord la variable expliquée (Y) puis les variables explicatives (X1 ; X2, . . .) : Y X1 X2 . . . c
c sert à spécifier l'introduction d'une constante
Nota : Si la deuxième méthode est utilisée, l'équation est déjà spécifiée mais peut être modifiée
 - Estimation Settings :
Method : Permet de choisir l'estimateur (MCO [LS] par défaut)
Sample : Permet de choisir l'échantillon retenu
-

Commandes post-estimations

- Les coefficients estimés sont conservés dans l'objet `c`
 - Les résidus estimés de la dernière équation sont stockés dans "resid"
 - Name :
 - Permet de conserver la régression dans un workfile
 - View →representation :
 - Permet de visualiser la ligne de commande effectuée, l'équation théorique et l'équation avec les valeurs estimées des coefficients
 - View →estimation output :
 - Permet de visualiser les résultats bruts de la régression.
-

Commandes post-estimations

- View →actual, fitted, residual :
 - actual : valeur de la variable dépendante utilisée dans la régression,
 - fitted : valeurs de la variable dépendante prédites par la régression en appliquant les coefficients de la régression sur les variables explicatives,
 - residual (actual-fitted) : indication sur les erreurs de prévision éventuelles, bornes à 5%.
- Freeze :

Permet de conserver les résultats.

Commandes post-estimations

- Il est possible de vouloir conserver plusieurs éléments de l'équation estimée
 - Ex : Pour calculer des points de retournement ou pour certains tests il faut conserver les R^2 , la SCR, . . .
- Pour ce faire, il faut généralement de créer un objet (scalaire, matrice) qui puisse accueillir ces nouveaux éléments
 - Exemples :

- Scalaire : `scalar nom=nomequation.operation`
- Matrice : `matrix nom=nomequation.operation`
- Ex : `scalar rsq=eq1.@r2`
- Ex : `matrix coefficients=eq1.@coefs`

💡 Tip

L'opération commence par `.@` en général

Commandes post-estimations

Quelques éléments disponibles (non exhaustif):

Élément	Opération	Type d'objet
R^2	<code>@r2</code>	<code>scalar</code>
R^2 ajusté	<code>@rbar2</code>	<code>scalar</code>
SCR	<code>@ssr</code>	<code>scalar</code>
Coefficient pour la i -ème variable	<code>c(i)</code>	<code>scalar</code>
t-stat pour la i -ème variable	<code>@tstats(i)</code>	<code>scalar</code>
Matrice de variance-covariance	<code>@coefcov</code>	<code>matrix</code>
Matrice des coefficients	<code>@coefs</code>	<code>matrix</code>
Matrice des t-stat	<code>@tstats</code>	<code>matrix</code>

Une liste plus complète est disponible dans Users Guide II page 16

Le coefficient de détermination : Le R^2

- Le pouvoir explicatif du modèle

L'économétrie cherche à expliquer les variations de Y . Ceci est la variabilité totale (SCT pour somme des carrés totale) et est donnée par : $SCT = \sum_{i=1}^N (y_i - \bar{y})^2 = SCE + SCR$

Cette variabilité se décompose en :

- Variabilité expliquée : SCE (pour somme des carrés expliquée)

- Variabilité non expliquée : SCR (pour somme des carrés des résidus)

Le coefficient de détermination : le R^2

- Le coefficient de détermination mesure le pouvoir explicatif du modèle et se calcule comme suit :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \quad \text{avec} \quad \begin{cases} SCR = \sum_{i=1}^N \hat{\varepsilon}_i^2 \\ SCT = \sum_{i=1}^N (y_i - \bar{y})^2 \end{cases}$$

- Ce coefficient mesure la qualité de l'ajustement de la régression en indiquant le pourcentage de la variance totale expliquée par le modèle :
 - Si ($R^2 \rightarrow 1$) : le modèle est **très explicatif**.
 - Si ($R^2 \rightarrow 0$) : le modèle est **peu explicatif**.
-

Le coefficient de détermination : le R^2

- Il faut en réalité faire attention avec le (R^2) :
 - Le (R^2) augmente mécaniquement avec l'ajout de variables explicatives.
 - Il faut par conséquent privilégier une version ajustée du nombre de degrés de liberté, le (R^2) ajusté :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{N-1}{N-p}$$

- (N) : nombre d'observations
 - (p) : nombre de variables explicatives (sans la constante)
 - Le (R^2) n'est pas un objectif en soi, il ne faut pas chercher à le maximiser.
-

Significativité statistique

La significativité simple

- **Objectif** : déterminer si le coefficient estimé est précis.
- Pour cela, on fait un **test de Student** à partir de :
 - la valeur estimée du coefficient ($\hat{\beta}_j$),
 - et la valeur estimée de son écart-type ($\hat{\sigma}_\beta$).

Rappel : l'écart-type mesure la dispersion d'une série autour de sa moyenne.

- La statistique de test est la suivante :

$$t_{\beta_j} = \frac{\hat{\beta}_j - \beta_{th}}{\hat{\sigma}_\beta}$$

- Les hypothèses testées sont :
 - $H_0 : \beta_j = \beta_{th}$
 - $H_1 : \beta_j \neq \beta_{th}$
-

La significativité simple

- Le test consiste souvent à savoir si le paramètre est significativement différent de 0 ($\beta_{th} = 0$).
- La statistique de test devient donc :

$$t_{\beta_j} = \frac{\hat{\beta}_j}{\hat{\sigma}_\beta}$$

- Les hypothèses testées deviennent :
 - $H_0 : \beta_j = 0$
 - $H_1 : \beta_j \neq 0$
-

La significativité simple

- La statistique de test calculée t_{β_j} est comparée à la statistique théorique t_α tabulée pour un risque de première espèce α .
- Remarque : il s'agit en général d'un **test bilatéral**.

```
alpha <- 0.05
df     <- 30
xlim   <- c(-4, 4)
tcrit  <- qt(1 - alpha/2, df = df)

xx <- seq(xlim[1], xlim[2], length.out = 2000)
yy <- dt(xx, df = df)

plot(xx, yy, type = "l", lwd = 2,
      xlab = "t", ylab = "densité",
      main = sprintf("Loi t(%d) - test bilatéral ( = %.2f)", df, alpha),
      xaxt = "n") # on dessine l'axe X nous-mêmes

shade_region <- function(x_from, x_to, col){
  xseq <- seq(x_from, x_to, length.out = 500)
  yseq <- dt(xseq, df = df)
  polygon(c(xseq, rev(xseq)), c(yseq, rep(0, length(yseq))),
          col = col, border = NA)
}

# Colorier les zones
shade_region(-tcrit, tcrit, col = rgb(0.2, 0.6, 1, 0.3)) # zone centrale
shade_region(xlim[1], -tcrit, col = rgb(1, 0.2, 0.2, 0.35)) # queue gauche
shade_region(tcrit, xlim[2], col = rgb(1, 0.2, 0.2, 0.35)) # queue droite

# Traits verticaux
abline(v = c(-tcrit, tcrit), lwd = 2, lty = 2)

# Axe X avec -t* et t* comme graduations
axis(1,
      at = c(xlim[1], -tcrit, 0, tcrit, xlim[2]),
      labels = c("", sprintf("-t* = %.2f", -tcrit), "0", sprintf("t* = %.2f", tcrit), ""),
      tick = TRUE)

# Étiquettes
text(0, max(yy)*0.65, expression(1 - alpha), cex = 1.4)
```

```
# /2 décalés sous l'axe X, vers l'extérieur
text(-tcrit, -0.015, expression(alpha/2), cex = 1.2, pos = 1, offset = 1)
text( tcrit, -0.015, expression(alpha/2), cex = 1.2, pos = 1, offset = 1)

legend("topright",
      legend = c("densité t(df)",
                  "région d'acceptation (1-)",
                  "régions de rejet (/2)"),
      lty = c(1, NA, NA), lwd = c(2, NA, NA),
      pch = c(NA, 15, 15),
      pt.cex = 2,
      col = c("black", rgb(0.2,0.6,1,0.3), rgb(1,0.2,0.2,0.35)),
      bty = "n", cex = 0.9)
```

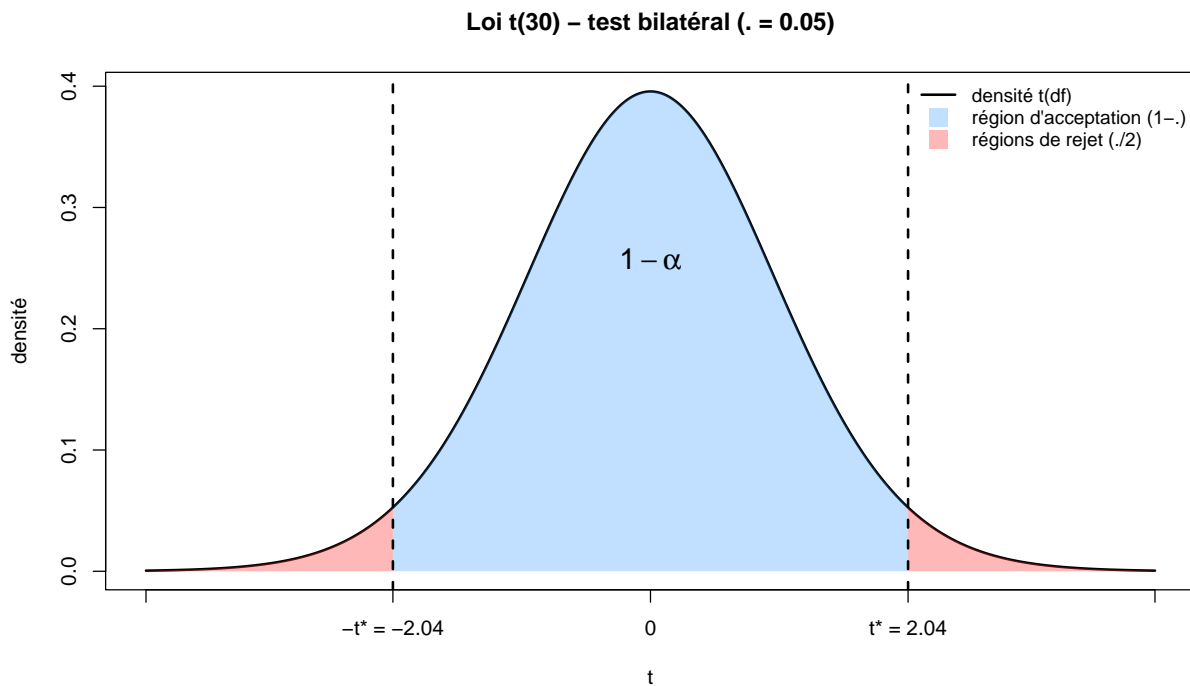


Figure 2: Test bilatéral : $\alpha/2$ décalés vers l'extérieur, $-t^*$ et t^* en graduations de l'axe X.

La significativité simple — procédure

1. **Calculer** la statistique de Student (Coef / SE) : $t_{\beta_j} = \hat{\beta}_j / \sigma \hat{\beta}_j$.

2. **Choisir** un niveau de risque de première espèce α .
3. **Déterminer** la valeur critique tabulée $t_{\alpha/2, \nu}$ pour un **test bilatéral**, avec $\nu = N - p$ (ddl : nb d'observations N moins nb de paramètres p) .
4. **Conclure** sur la significativité selon la règle de décision :
 - si $|t| < t_{\alpha/2, \nu} (\Rightarrow)$ non-rejet de H_0
 - si $|t| \geq t_{\alpha/2, \nu} (\Rightarrow)$ rejet de (H_0)

Tableau de décision test bilatéral

Décision / Réalité	H_0 vraie	H_0 fausse
Rejeter H_0	Erreur α	Décision correcte
Ne pas rejeter H_0	Décision correcte	Erreur β

Caution

On parle de rejet ou non rejet d'une hypothèse, pas d'acceptation.

La significativité conjointe

Tip

Dans un modèle, nous pouvons nous intéresser à déterminer si nos différentes variables ont un effet significatif sur notre variable Y , dépendamment les unes des autres. C'est à dire, est-ce que mes variables sont significatives *conjointement* (X_1 significative ET X_N ...)

Dans ce cadre, les simples tests de Student ne sont pas suffisants, pour tester plusieurs restrictions, il faut recourir à d'autres tests:

Test de Fisher dans le cas des modèles linéaires

Tests de Wald, du log de vraisemblance ou du multiplicateurs de Lagrange dans les cas plus complexes

La significativité conjointe le test de Fisher (F-test)

- Le **F-test** permet de tester la significativité conjointe de plusieurs paramètres, voire la significativité globale d'un modèle linéaire. La statistique de test est la suivante :

$$F = \frac{SCR_r - SCR_{nr}}{SCR_{nr}} \frac{N - p}{q}$$

où :

- q : nombre de restrictions testées (sans la constante),
 - p : nombre de paramètres dans le modèle non restreint (avec la constante),
 - N : nombre d'observations.
 - SCR_r : somme des carrés des résidus du modèle **restreint** (les paramètres imposés sont fixés),
 - SCR_{nr} : somme des carrés des résidus du modèle **non restreint** (modèle usuel non contraint).
-

La significativité conjointe — F-test (unilatéral)

- Le test de Fisher est **unilatéral** (rejet dans la **queue droite**).
- Sous (H_0), la statistique suit une loi de **Fisher-Snedecor** : $F \sim F(q, N - p)$, où q = nb de restrictions testées et $N - p$ = ddl résiduels du modèle non restreint.
- Les logiciels (EViews, R, etc.) donnent directement (F), la p-value et la table ANOVA.

```
alpha <- 0.05
q      <- 3
df2    <- 30
Fcrit  <- qf(1 - alpha, df1 = q, df2 = df2)

xmax <- qf(0.999, df1 = q, df2 = df2)
xx   <- seq(0, xmax, length.out = 2000)
yy   <- df(xx, df1 = q, df2 = df2)

plot(xx, yy, type = "l", lwd = 2,
      xlab = "F", ylab = "densité",
      main = sprintf("Loi F(%d, %d) - test unilatéral ( = %.02f)", q, df2, alpha),
      cex.lab = 1.3, cex.axis = 1.2, xaxt = "n")
```

```

shade_region <- function(x_from, col){
  xseq <- seq(x_from, xmax, length.out = 600)
  yseq <- df(xseq, df1 = q, df2 = df2)
  polygon(c(xseq, rev(xseq)), c(yseq, rep(0, length(yseq))),
          col = col, border = NA)
}

# Zones
polygon(c(0, xx[xx <= Fcrit], Fcrit),
        c(0, yy[xx <= Fcrit], 0),
        col = rgb(0.2, 0.6, 1, 0.15), border = NA)
shade_region(Fcrit, col = rgb(1, 0.2, 0.2, 0.35))

abline(v = Fcrit, lwd = 2, lty = 2)

# Axe X plus lisible avec F* bien marqué
axis(1,
     at = c(0, Fcrit, round(xmax,1)),
     labels = c("0",
                bquote(F~"*" == .(round(Fcrit,2))),
                round(xmax,1)),
     cex.axis = 1.2)

# Étiquettes
text(mean(c(0,Fcrit))*0.5, max(yy)*0.5, expression(H[0]), cex = 1.4)
text((Fcrit + xmax)/2, max(yy)*0.25, expression(H[A]), cex = 1.4)
text(Fcrit, par("usr")[3] - 0.02, expression(alpha),
     xpd = NA, pos = 1, cex = 1.3)

legend("topright",
      legend = c("densité F(q, N-p)",
                 "région H0 (non rejet)",
                 "région de rejet ( )"),
      lty = c(1, NA, NA), lwd = c(2, NA, NA),
      pch = c(NA, 15, 15), pt.cex = 2,
      col = c("black", rgb(0.2,0.6,1,0.15), rgb(1,0.2,0.2,0.35)),
      bty = "n", cex = 1)

```

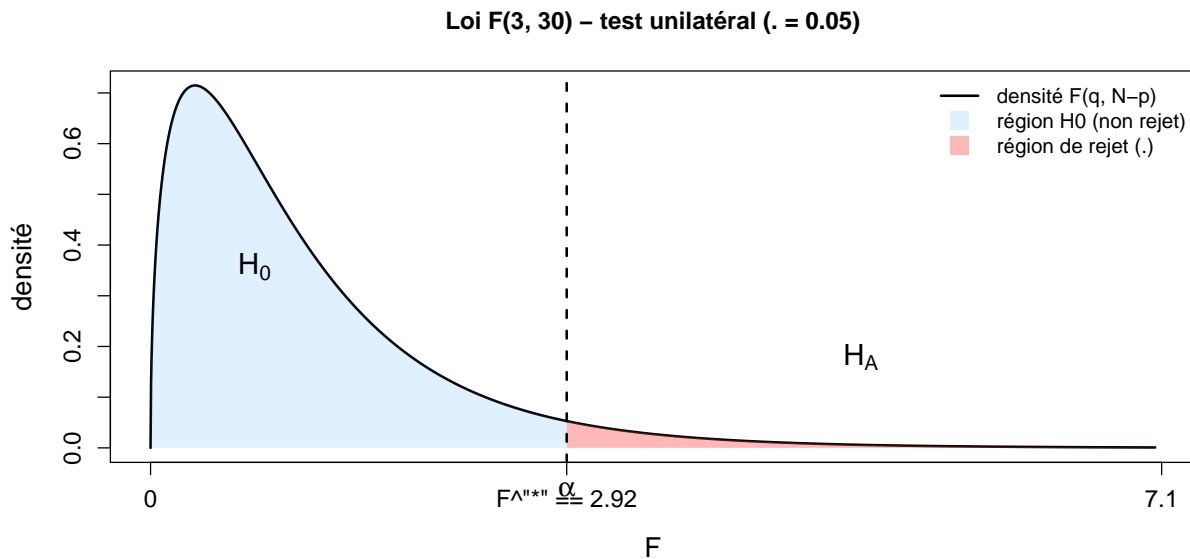


Figure 3: Loi F(3,30) — test unilatéral : étiquettes lisibles.

La significativité conjointe — hypothèses usuelles du F-test

- On teste généralement la contrainte selon laquelle **tous les coefficients (hors constante) sont nuls**.
 - H_0 : tous les coefficients du modèle sont égaux à 0 (sauf l'intercept), c.-à-d. $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$
 - H_1 : **au moins un** coefficient est différent de 0.
- Dans ce cas, le **modèle contraint** est le modèle **avec seule la constante**. **Règle de décision** :
 si $F > F_{\text{table}}$ (au niveau α) et ddl $(q, N - p) \Rightarrow$ **rejet de H_0** .
- **Interprétation** :
 - **Non-rejet de H_0** \Rightarrow pas de relation linéaire significative entre la variable expliquée et l'ensemble des variables explicatives.
 - Autrement dit, la **SCE** (somme des carrés expliquée) n'est pas significativement différente de 0 ; la variabilité de (Y) demeure essentiellement **aléatoire**.

La significativité conjointe — F-Test : procédure EViews

- Procédure à suivre :

1. Régresser le modèle **non contraint** et relever la SCR.
2. Régresser le modèle **contraint** et relever la SCR.
3. Calculer la statistique de Fisher.
4. Comparer la valeur obtenue à la valeur théorique (table de Fisher).

La significativité conjointe — F-Test : Exemple de commandes EViews

```
equation eqnr Y X1 X2 X3 X4 X5 c
scalar scnr = eqnr.@ssr
equation eqr Y X1 X3 c
scalar scrr = eqr.@ssr
scalar F = ((scrr - scnr) / scnr) * ((129 - 5) / 3)
```

Ici :

- **eqnr** : estimation du modèle **non restreint** (toutes les variables).
- **eqr** : estimation du modèle **restreint**.
- **scnr** et **scrr** : sommes des carrés des résidus respectivement non restreint et restreint.
- **F** : statistique de Fisher calculée manuellement.

La significativité conjointe — Wald-test

- La procédure selon le **Wald-test** est pré-enregistrée dans EViews :

1. **Ouvrir** les résultats de l'estimation.
2. **Menu** : View → Coefficient diagnostic → Wald test.
3. **Saisir les contraintes** de la forme :

`c(numéro_coef1) = 0`

`c(numéro_coef2) = 0`

par exemple :

`c(3) = 0`

`c(5) = 0`

Significativité économique

! Important

Une variable peut avoir une très grande significativité statistique mais une faible significativité économique. Ici nous ne nous intéressons pas à la “robustesse” de l'estimation de l'effet, mais à sa “magnitude”

L'interprétation du coefficient estimé β dépend de la manière dont **Y** (variable expliquée) et **X** (variable explicative) sont exprimées : en **niveau** ou en **logarithme**.

Significativité économique - un tableau récapitulatif :

Variable expliquée (Y)	Variable explicative (X)	Interprétation du coefficient β
Niveau	Niveau	Une augmentation de 1 unité de X entraîne une variation moyenne de β unités de Y.

Variable expliquée (Y)	Variable explicative (X)	Interprétation du coefficient β
Niveau	Logarithme	Une augmentation de 1 % de X entraîne une variation moyenne de $\beta/100$ unités de Y.
Logarithme	Niveau	Une augmentation de 1 unité de X entraîne une variation moyenne de $\beta \times 100$ % de Y.
Logarithme	Logarithme	Une augmentation de 1 % de X entraîne une variation moyenne de β % de Y.

Questions – Réponses TD2 (Module 2)

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

Afficher la réponse

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

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

Afficher la réponse

Commande : `genr logpassagers = log(passagers)`

Cette transformation :

- réduit l'impact des valeurs extrêmes en **compressant l'échelle**,
- rapproche la distribution d'une **loi normale**,
- **stabilise les variances**,
- permet une **interprétation en pourcentage** : une variation d'une unité du log une variation d'environ 100 % de la variable d'origine,

- aide à **linéariser les relations** et donc facilite l'usage de la régression linéaire.
-

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

Afficher la réponse

Menu **Object** → **New Object** → **Equation**, choisir « Linear ».

Même si la variable dépendante est en logarithme, l'équation reste **linéaire** car les variables explicatives apparaissent en **première puissance** et la relation est **additive**.

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

Afficher la réponse

- **Variable dépendante (expliquée)** : logpass (log du nombre de passagers).
 - **Variables explicatives** : Ratio, croissance annuelle du trafic aérien de la destination principale (2010-2013), public, low cost, age, intercontinental, croissance annuelle du trafic aérien du pays d'origine (2010-2013).
 - **Variable d'intérêt** : Ratio.
 - **Variables de contrôle** : toutes les autres variables explicatives listées ci-dessus.
-

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

Afficher la réponse

Le R^2 obtenu est 0,39 : le modèle explique environ **40 % de la variabilité** du nombre de passagers.

pouvoir explicatif modéré, le modèle reste relativement peu explicatif.

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

Afficher la réponse

Test de Student bilatéral :

- $|t_{\text{calculé}}| = 3,22 > t_{\alpha/2, N-p} = 1,658$.
* Avec $N - p = 94 - 8 = 86$.

Rejet de (H_0) : le coefficient du nombre d'accidents par passagers est **significativement différent de 0**.

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

Afficher la réponse

Créer les variables :

`genr Dummy_fatal = fatal>=1 genr Dummy_non_fatal = non_fatal>=1`

Puis relancer la régression en remplaçant `ratio` par les deux nouvelles variables.

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

Afficher la réponse

- **Individuellement :**
 - Accidents mortels : $|t| = 1,713 > 1,658$ significatif.
 - Accidents non mortels : $|t| = 1,671 > 1,658$ significatif.
 - **Conjointement (test de Fisher) :**
 - $F_{\text{calculé}} = 6,792 > F_{\text{table}} = 3,07$ **rejet de H_0** ,
 - donc les deux variables sont **conjointement significatives**.
-

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

Afficher la réponse

Procédure :

Dans EViews → View → Coefficient diagnostics → Scaled coefficient.

Le **coefficient standardisé** indique de combien d'écarts-types Y varie quand X varie d'un écart-type.

Cela permet de comparer directement l'importance relative des variables.

Résultat :

La variable **intercontinentale** a le plus grand coefficient standardisé ($\sim 0,42$).

Comme Y est en logarithme et X en niveau, cela signifie qu'une compagnie qui devient **intercontinentale** augmente en moyenne le nombre de passagers d'environ **10 %**, toutes choses égales par ailleurs.

C'est la variable la plus importante d'un point de vue économique.