

Séminaire d'économie appliquée - Prix

Séance d'introduction

Owen Brown

3 février 2026

Objectif de la séance

L'objectif de cette séance sera de revenir sur l'organisation du cours et du matériel que nous emploierons. Nous discuterons:

- du projet (deadline, organisation, évaluation)
- d'outils de traitement statistique (R: son installation et son usage)
- de méthodes de webscrapping (Qu'est-ce que c'est ?)

Du projet

1. L'objectif de ce projet est de vous faire mesurer. Mesurer, c'est central en économie appliquée.
2. Votre objet d'étude sera un bien.
3. Vous devrez mesurer la contribution des caractéristiques d'un bien au prix de ce bien.

→ Ceci peut être réalisé avec des outils économétriques : les régressions hédoniques.

Une régression hédonique décompose le prix en ses constituants caractéristiques, et fournit une estimation de la valeur contributaire de chaque caractéristique.

Rappel - L'économétrie Wooldridge (2015)

L'économétrie est un outil fondé sur des méthodes statistiques ayant pour but :

- d'estimer des relations économiques
- de tester des théories
- de mesurer des effets

Cet outil nous permet de mesurer la contribution d'une caractéristique au prix d'un bien.

Exemple

Imaginons que nous voulions estimer le prix d'une voiture selon ses caractéristiques.

D'intuition, nous dirions que le prix de la voiture dépend de plusieurs facteurs, notamment :

1. Sa puissance (+)
2. Son poids (-)

Exemple

Faisant l'hypothèse que la relation entre le prix et les caractéristiques prend une forme fonctionnelle suivante:

$$Prix_i = \beta_1 \cdot Puissance_i + \beta_2 \cdot Poids_i$$

L'économétrie nous permet d'estimer les coefficients β .

Si notre intuition s'avère correcte, une OLS délivrerait des estimations telles que :

- $\hat{\beta}_1 > 0$
- $\hat{\beta}_2 < 0$

Données

Soit un jeu de données à notre disposition:

| prix | puissance | poids |
|----------|-----------|----------|
| 914.1151 | 109.47007 | 315.6080 |
| 677.6421 | 220.80719 | 630.9535 |
| 906.0181 | 281.33309 | 326.2434 |
| 833.3272 | 24.07267 | 432.9804 |
| 995.8922 | 232.18435 | 408.6543 |
| 492.3984 | 237.95419 | 759.7522 |
| 341.5256 | 156.89450 | 857.2572 |

Testons notre modèle

Une estimation par la méthode des moindres carrés (OLS) nous donne:

Call:

```
lm(formula = prix ~ poids + puissance - 1, data = .)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|--------|
| -3535.8 | -401.0 | 316.1 | 990.6 | 4027.8 |

Coefficients:

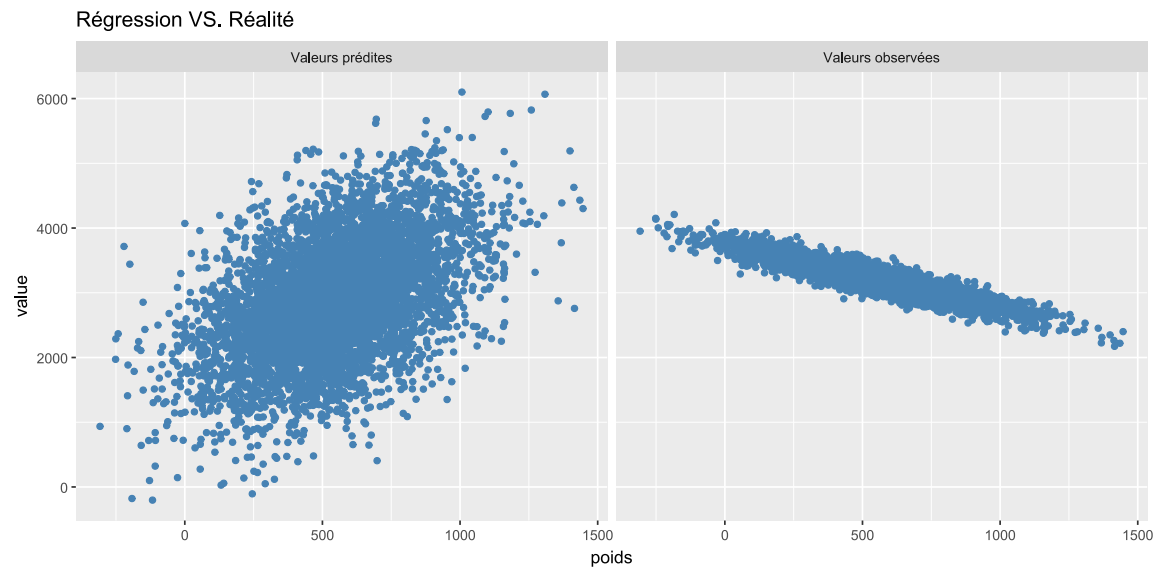
| | Estimate | Std. Error | t value | Pr(> t) |
|-----------|----------|------------|---------|------------|
| poids | 1.41737 | 0.05015 | 28.26 | <2e-16 *** |
| puissance | 10.68017 | 0.14083 | 75.84 | <2e-16 *** |

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

Les deux variables ont un effet positif sur le prix. Les p-valeurs sont significatives à plus de 99% de confiance !

Réfléchir à sa modélisation

- Nos estimations contredisent notre hypothèse ($\hat{\beta}_{poids} > 0$).
Est-ce que nos résultats ont du sens ?
- Représentons nos données



Revoir le modèle - I

- Ne faudrait-il pas un intercept ?

$$Prix_i = \alpha_0 + \beta_1 \cdot Puissance_i + \beta_2 \cdot Poids_i$$

Call:

```
lm(formula = prix ~ poids + puissance, data = .)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|--------|
| -400.33 | -68.09 | 0.88 | 67.04 | 361.14 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|------------|------------|---------|----------|-----|
| (Intercept) | 3.649e+03 | 4.899e+00 | 744.84 | <2e-16 | *** |
| poids | -9.992e-01 | 5.743e-03 | -173.98 | <2e-16 | *** |
| puissance | 5.065e-01 | 1.907e-02 | 26.56 | <2e-16 | *** |
| --- | | | | | |

Revoir le modèle - II

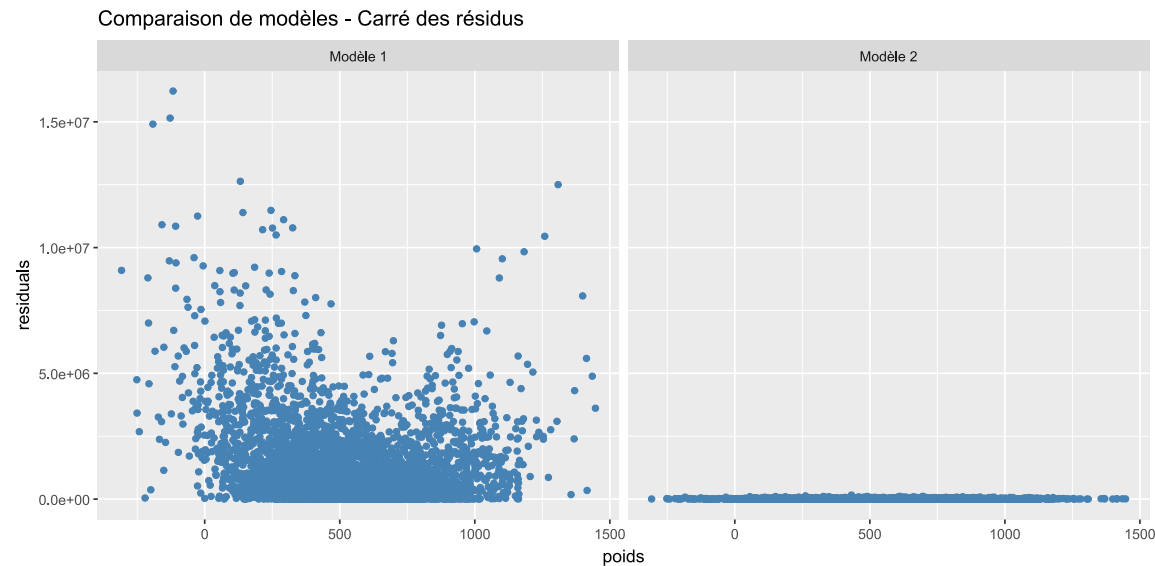


Beaucoup plus cohérent !

Problème résolu ?

Qu'avons nous comme outil pour évaluer s'il y a une potentielle erreur de mesure ?

Comparons l'hétéroscédasticité (σ^2) des deux modèles:



Notre second modèle semble beaucoup plus adapté.

Remarques

- Une *p-valeur* significative n'implique pas automatiquement que votre régression est correcte !
- Soyez toujours critiques de votre modélisation :
 - Est-ce que mes coefficients ont du sens ?
 - Ai-je à disposition un outil pour tester la véracité de mes résultats ?
- Codez ! Appliquer la théorie économétrique est un moyen précieux de rendre cette discipline plus claire.

À vous !

- Pour le projet il vous sera demandé de constituer une base de données. Vous pouvez récolter des données où vous le désirez, **tant que vous créez un jeu de données réelles.**
- Rassembler des données peut être couteux. Toutefois avec internet, vous avez à dispositions une masse considérable de données. Mais comment les réunir ?
- Le webscraping !

Le webscraping

1. Le *webscraping* est une méthode d'extraction de données de site web
2. Cette méthode exploite la structure des pages pour récupérer automatiquement de l'information
3. Elle s'avère très efficace pour constituer rapidement un jeu important de données !
4. En revanche, parce qu'un grand nombre de données en récolté, un travail de nettoyage de données est souvent nécessaire

La structure des pages ?

Les sites web que nous voyons sont en réalité une interprétation de code HTML.

Le banquet des empouses



Olga Tokarczuk

Ed. Le livre de poche

Date de publication : 04/09/2025

traduit du polonais par Maryla Laurent

En septembre 1912, Wojnicz arrive au sanatorium de Göbersdorf dans les montagnes de Basse-Silésie, afin de soigner sa tuberculose. Au cours de son séjour, le jeune homme évolue dans un milieu patriarcal, ce qui le rend mal à l'aise. Quand une mort violente a lieu et que des rumeurs de sorcellerie circulent, Wojnicz cherche à en savoir plus.

■ Acheter chez Tropicismes : [Le banquet des empouses](#)

```
Q Search HTML
<div class="item-page" itemscope="" itemtype="https://schema.org/Article">
  <meta itemprop="inLanguage" content="fr-FR">
  <div itemprop="articleBody">
    <div id="cck18398" class="cck-f100 cck-pad-0">
      <div>
        <div class="cck-line-body">
          <div class="cck-plr cck-ptb">
            <div>
              <div class="fiche-livre col-md-12">
                <div class="fiche-livreImageText">
                  <div class="imageIntroLivre" style="float:left;"></div>
                  <div class="auteurEditeurDateLivre" style="float:left;">
                    <span class="labelAuteurLivre"></span>
                    <span class="AuteurLivre">Olga Tokarczuk</span>
                    <br>
                    <span class="labelEditeurLivre">Ed. </span>
                    <span class="EditeurLivre">Le livre de poche</span>
                    <br>
                    <span class="labelDatePublicationLivre">Date de publication :</span>
                    <span class="datePublicationLivre">04/09/2025</span>
                    <br>
                    <span class="labelEditeurLivre"></span>
                    <span class="EditeurLivre">traduit du polonais par Maryla Laurent</span>
                    <br>
                  </div>
                  <div class="contenuLivre" style="float:left;">
                    <p></p>
                  </div>
                  <div class="clr"></div>
                </div>
              <div class="detailLivre"></div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
```

Scraper une page - I

- À quelques exceptions près, tous les sites web suivent une structure par “balise”
- Avec [R](#) ou [Python](#) et la librairie *Selenium*, il est possible de simuler une navigation internet et d’enregistrer de l’information contenu dans un code *HTML*.

Scrapper une page - II

Par étape :

1. Trouver une page type qui contient l'information sur le bien d'intérêt
2. Comprendre la structure du site
3. Localiser les balises d'information
4. Automatiser la navigation
5. Récolter les données et les stocker dans un fichier lisible par votre programme statistique

Un exemple peut-être ?

Le prix des téléphones

Quels sont les éléments constitutifs des prix de téléphone ?

- Le stockage
- La marque
- L'autonomie
- La performance

Créons un jeu de données !

| X | Variable | Value | model | | | | | | | |
|------|-----------------------------|-----------|-------|--------|----|-----|-----|----|------|--|
| 1 0 | price | € 1.309,- | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 2 1 | Format téléphone portable | Moyen | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 3 2 | Qualité de la caméra | Excellent | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 4 3 | Catégorie de vitesse | Excellent | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 5 4 | Qualité écran | Excellent | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 6 5 | Capacité totale de stockage | 256 Go | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 7 6 | Vitesse de chargement | Moyen | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 8 7 | Année d'introduction | 2025 | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 9 8 | Diagonale de l'écran | 6,3 p | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |
| 10 9 | Produit | 969443 | Apple | iPhone | 17 | Pro | 256 | Go | Bleu | |

References

Wooldridge, Jeffrey M. 2015. *Introduction à l'économétrie: une approche moderne*.
Louvain-La-Neuve, Belgium: De Boeck superieur.