

# ECON-S307 Séminaire d'économie appliquée - Prix

Séance d'introduction

Owen Brown

February 6, 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:

- de l'objectif général du projet
- des outils mobilisables pour ce projet
- du travail final à rendre

# Structure du projet

Chaque groupe doit rédiger un projet de 4 pages / membre (page de garde, appendix et références exclues - taille de police 12). Le projet doit prendre la forme suivante :

- 0. Page de garde** - titre, noms, matricules
- 1. Revue de littérature** - synthèse de la recherche relative à votre bien
- 2. Discussion de la source de données** - justification, fiabilité, méthode de récolte
- 3. Analyse descriptive des données** - sommaire des variables, graphiques, tables
- 4. Méthode, résultats, discussion** - spécification de la régression hédonique, estimation de modèle, discussion des résultats, liens contextuels
- 5. Références & Appendix**

Le travail comprendra en préambule une table de matières ainsi qu'une table de graphiques.

# Remise & défense de projet

- Votre fichier devra être remis sur l'UV sous format PDF.
- En indiquant votre numéro de groupe et vos noms (de famille) par ordre alphabétique, vous nommerez votre le fichier de votre travail :

“[GRP\_] - Séminaire d'économie appliquée (Prix)\_NOM\_1-NOM\_2-NOM\_3.pdf”.

- Vous devrez également soumettre votre jeu de données ainsi que votre code. **Il est impératif que le code que vous communiquez produise exactement les mêmes résultats que votre travail.**
- Votre code devra être commenté. Chaque étape clé devra être clairement indiquée et brièvement décrite.

“[GRP\_] - Code\_NOM\_1-NOM\_2-NOM\_3.R”

“[GRP\_] - Dataset\_NOM\_1-NOM\_2-NOM\_3.csv”

# Dates du projet

- **1er mars - 23h59** : Remettre une note dans laquelle vous expliquez quel bien vous souhaitez analyser et expliquez comment vous construirez un jeu de données (1 page max).

[GRP\_] - Note 1 (PrixD) NOM\_1-NOM\_2-NOM\_3.pdf

- **15 mars - 23h59** : Remettre une note dans laquelle vous expliquez votre avancée de récolte de données - Méthode employée, source, quantité de données, objectifs (2 pages max).

[GRP\_] - Note 2 (PrixD) NOM\_1-NOM\_2-NOM\_3.pdf

- **30 avril - 23h59** : Remettre votre travail sur l'UV

- Votre rapport final - format PDF

# Des questions ?

# 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  $\beta$ .

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:

```
1 dta %>% select(c("prix","puissance","poids")) %>% summary
```

	prix	puissance	poids
Min.	:2172	Min. :-54.7	Min. :-308.1
1st Qu.	:3030	1st Qu.:149.3	1st Qu.: 380.8
Median	:3200	Median :199.6	Median : 542.5
Mean	:3206	Mean :199.5	Mean : 544.6
3rd Qu.	:3381	3rd Qu.:249.9	3rd Qu.: 710.6
Max.	:4211	Max. :439.7	Max. :1446.7

```
1 dta %>% select(c("prix","puissance","poids")) %>% dim
```

```
[1] 5000     3
```

# Testons notre modèle

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

```
1 reg <- dta %>% lm(formula= prix ~poids+ puissance-1)
2 reg %>% tidy %>% kbl() %>% kable_styling(font_size = 35)
```

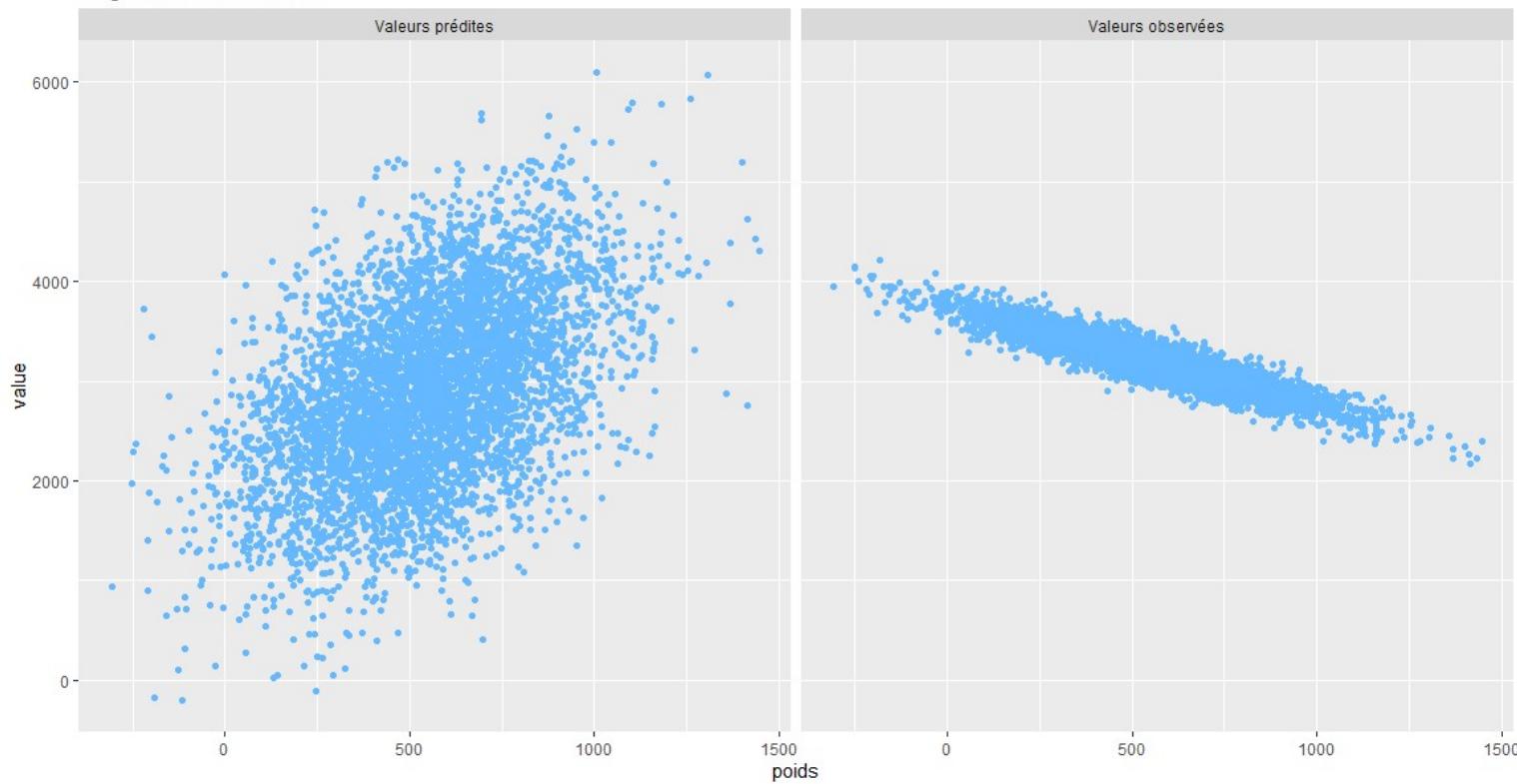
term	estimate	std.error	statistic	p.value
poids	1.41737	0.0501524	28.26127	0
puissance	10.68017	0.1408273	75.83878	0

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

## Régression VS. Réalité



# Revoir le modèle - I

- Ne faudrait-il pas un intercept ?

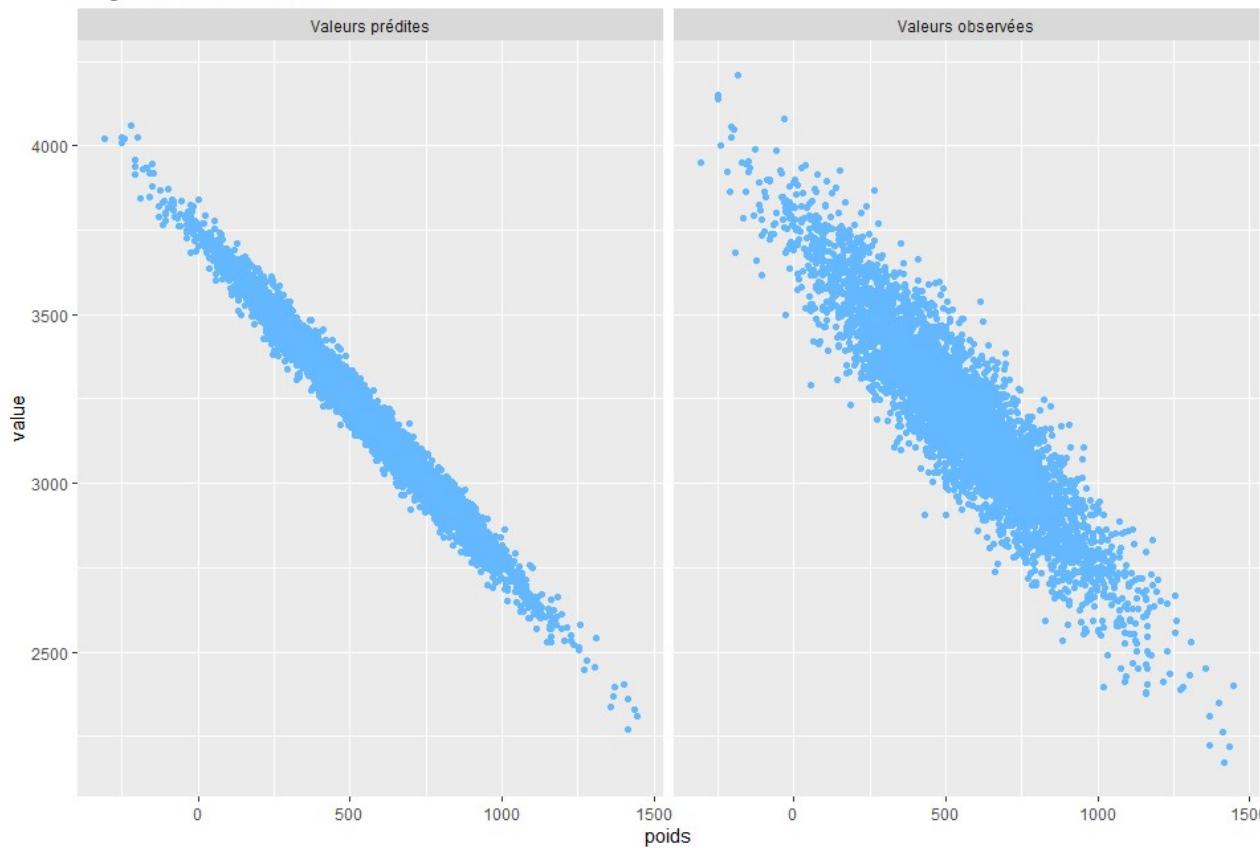
$$Prixi = \alpha_0 + \beta_1 \cdot Puissance_i + \beta_2 \cdot Poids_i$$

```
1 reg2 <- dta %>% lm(formula= prix ~poids+ puissance)
2 reg2 %>% tidy %>% kbl() %>% kable_styling(font_size = 35)
```

term	estimate	std.error	statistic	p.value
(Intercept)	3648.7080863	4.8986699	744.83648	0
poids	-0.9992213	0.0057432	-173.98399	0
puissance	0.5064512	0.0190694	26.55832	0

# Revoir le modèle - II

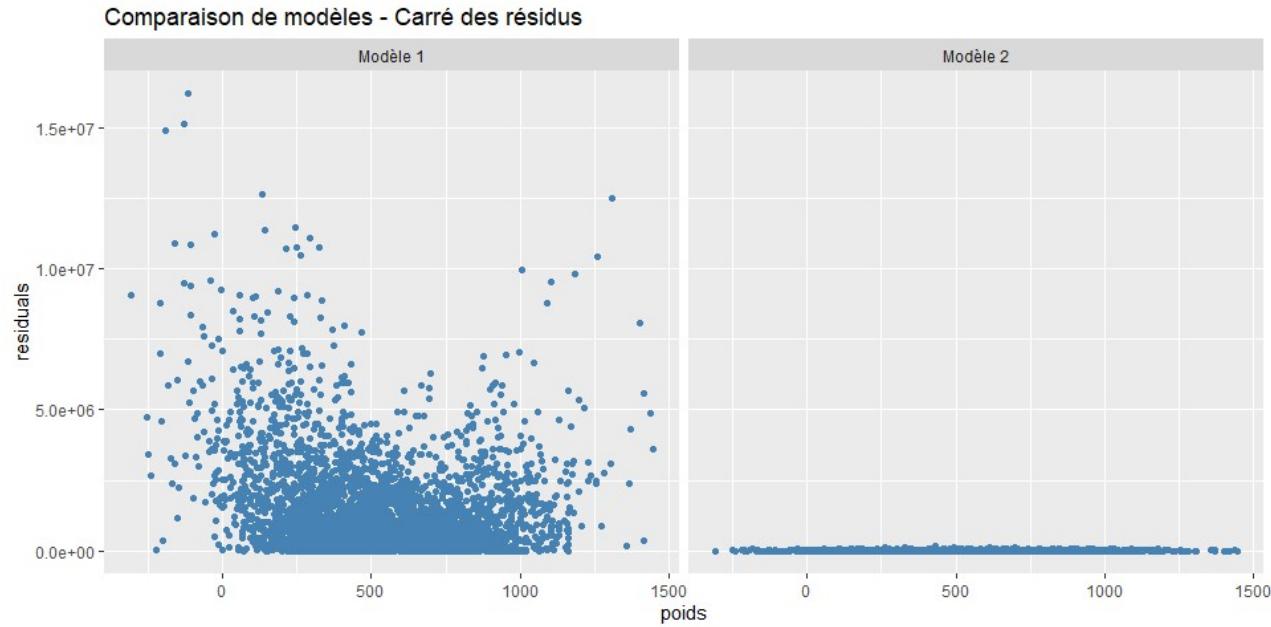
## Régression VS. Réalité



# Problème résolu ?

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

Comparons l'hétéroscédasticité ( $\sigma^2$ ) 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 obtenez un jeu de données réelles.**
- Rassembler des données peut être couteux. Toutefois avec internet, vous avez à disposition 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 patriarchal, 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 Tropismes : [Le banquet des empouses](#)

```
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-8">
      <div>
        <div class="cck-line-body">
          <div class="cck-plr cck-ptb">
            <div>
              <div class="fiche-livre col-md-12">
                <h3>Le banquet des empouses</h3>
                <div class="ficheLivreImageTexte">
                  <div class="imageIntroLivre" style="float:left;">...</div>
                  <div class="auteurEditeurDateLivre" style="float:left;">
                    <span>LabelAuteurLivre</span>
                    <span>AuteurLivre</span>Olga Tokarczuk</span>
                    <br>
                    <span>LabelEditeurLivre</span>Ed.</span>
                    <span>EditeurLivre</span>Le livre de poche</span>
                    <br>
                    <span>labelDatePublicationLivre</span>Date de publication :</span>
                    <span>datePublicationLivre</span>04/09/2025</span>
                    <br>
                    <span>labelEditeurLivre</span>
                    <span>EditeurLivre</span>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="detailsLivre" style="float:left;">...</div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
```

# Scaper une page - I

- À quelques exceptions près, tous les sites web suivent un structure par “balise”
- Avec [R](#) ou [Python](#) et la librairie *Selenium*, il est possible de simuler une navigation internet et d'enregister 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 !

# Jeu de donnée

Après la récolte automatisée ainsi que le nettoyage des données, nous obtenons un jeu tel que suit :

```
1 phonesdf %>% print(n=28)

# A tibble: 28 × 14
  Produit price `Qualité de la caméra` `Année d'introduction` Marque Garantie
  <chr>   <dbl> <fct>                <chr>           <fct>   <chr>
1 969443    1309 Excellent             2025            Apple   2 ans
2 969442    1329 Excellent             2025            Apple   2 ans
3 960968     373 Très bien            2025            Samsung 2 ans
4 969450    1479 Excellent             2025            Apple   2 ans
5 957545     499 Excellent             2025            OnePlus 2 ans
6 960966     443 Très bien            2025            Samsung 2 ans
7 969451    1469 Excellent             2025            Apple   2 ans
8 969260     739 Excellent             2025            Google  2 ans
9 968966    1479 Excellent             2025            Apple   2 ans
10 935188    659 Très bien            2023            Apple   2 ans
11 969423    969 Très bien            2025            Apple   2 ans
```

# Notre régression hédonique - I

Sur base de ce qui a été récolté, nous pouvons évaluer la contribution des diverses caractéristiques des biens à l'aide de la régression hédonique suivante:

$$\begin{aligned} Prix_{i,m} = & \alpha_m + \beta_1 \cdot Poids_i \\ & + \beta_2 \cdot megapixels_i \\ & + \beta_3 \cdot age_i + \beta_4 \cdot RAM_i + \beta_5 \cdot stockage_i \\ & + \beta_6 \cdot Qualité_Camera_i + \varepsilon_{i,m} \end{aligned}$$

# Notre régression hédonique - II

```

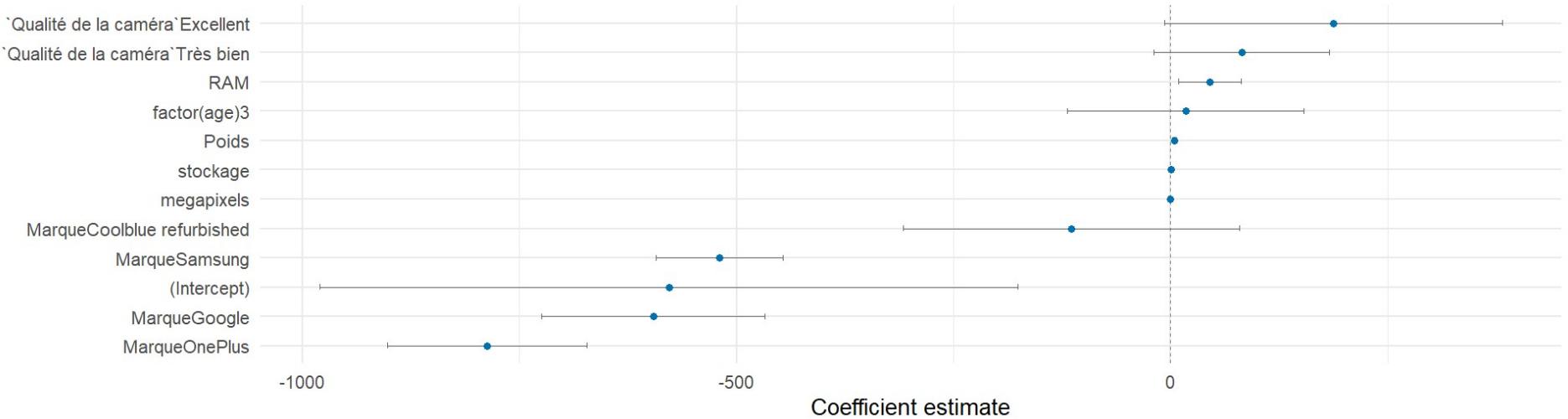
1 preg2 <- linear_reg() %>% set_engine("lm") %>%
2   fit( price~ Poids + megapixels + factor(age) + RAM +
3       stockage + (`Qualité de la caméra`)+(Marque), data = phonesdf)
4
5 preg2 %>% tidy %>% kbl() %>% kable_styling(full_width = F,font_size = 20) %>%
6   row_spec(c(10:12),background = "steelblue1")

```

term	estimate	std.error	statistic	p.value
(Intercept)	-577.940655	205.2751457	-2.8154439	0.0124385
Poids	4.610544	0.8158018	5.6515488	0.0000361
megapixels	-0.518293	0.7165147	-0.7233529	0.4799054
factor(age)3	17.623707	69.5644673	0.2533435	0.8032292
RAM	45.423687	18.3874036	2.4703698	0.0251261
stockage	1.032592	0.1340227	7.7046002	0.0000009
`Qualité de la caméra` Excellent	187.495546	99.3178400	1.8878335	0.0773217
`Qualité de la caméra` Très bien	81.992351	51.7327882	1.5849204	0.1325480
MarqueCoolblue refurbished	-114.307030	98.6721110	-1.1584533	0.2636785
MarqueGoogle	-595.655406	65.6758668	-9.0696238	0.0000001
MarqueOnePlus	-787.606439	58.5517744	-13.4514530	0.0000000
MarqueSamsung	-519.670553	37.2459179	-13.9524163	0.0000000

# Notre régression hédonique - III

Coefficient plot with 95% confidence intervals



Coefficients de régression hédonique

On observe une nette différence entre les prix de Samsung / Google / OnePlus et ceux de Apple !

# Notre régression hédonique - IV

Discuter ses résultats

- Le signe des coefficients
- La forme fonctionnelle
- L'expression des variables explicatives (quantitative vs qualitative)
- Pourquoi observe-t-on de tels résultats ? -> Recherche contextuelle !
- Comparer les marchés !
  - Une autre enseigne établit-elle des prix différents ?

# Remarques

- Ceci est une méthode de récolte de données. Vous êtes libres de récolter des données autrement. Des sites recensent des jeux de données existantes.
- Votre travail devra être rigoureusement référencé (sites, articles scientifiques, actualité). Tout plagiat sera sévèrement et indiscutablement sanctionné.
- Vous êtes libres d'utiliser le logiciel statistique que vous préférez... mais choisissez-le bien !
- Une séance pourra être organisée pour introduire l'utilisation de R ainsi que les bonnes pratiques d'usage ainsi qu'une guidance pour d'éventuelles questions.
- Mobilisez vos connaissances d'autres cours ! cf. *Méthode et modélisation économique & Introduction à l'économétrie*

# Prochaines échéances

- **1er mars - 23h59** : Remettre une note (1 page max) dans laquelle vous expliquez quel bien vous souhaitez analyser et expliquez comment vous construirez un jeu de données.

*[GRP\_] - Note 1 (Prix)\_NOM\_1-NOM\_2-NOM\_3.pdf*

- **15 mars - 23h59** : Remettre une note dans laquelle vous expliquez votre avancée de récolte de données - Méthode employée, source, quantité de données, objectifs (2 pages max)

*[GRP\_] - Note 2 (Prix)\_NOM\_1-NOM\_2-NOM\_3.pdf*

**Bon travail !**

# Liens utiles

Certains sites recensent des bases de données, notamment:

- Kaggle:<https://www.kaggle.com>
- Google Dataset Search:<https://datasetsearch.research.google.com>
- UC Irvine Machine learning repository:<https://archive.ics.uci.edu/>

/!\ Si vous exploitez une base de données que vous n'avez pas constitué vous-mêmes, il est exigé de référencer dûment les auteurs et autrices du jeu de données.

- [Tutoriel vidéo d'installation R & Rstudio](https://www.youtube.com/watch?v=hhqhyexDrsg): <https://www.youtube.com/watch?v=hhqhyexDrsg>
- [Tutoriel d'installation Pycharm](https://www.jetbrains.com/help/pycharm/installation-guide.html#toolbox):<https://www.jetbrains.com/help/pycharm/installation-guide.html#toolbox>
- [Tutoriel vidéo d'utilisation Selenium \(Webscraping\)](https://www.youtube.com/watch?v=N8OceGZGjA):<https://www.youtube.com/watch?v=N8OceGZGjA>

# Bibliographie

Wooldridge, Jeffrey M. 2015. *Introduction à l'économétrie: Une Approche Moderne*. Louvain-La-Neuve, Belgium: De Boeck superieur.