

Introduction à l'économie appliquée

Chapitre I L'analyse empirique

où l'on définit l'objet de l'économie empirique, et deux moyens que sont l'économétrie et l'économie expérimentale.

Trois temps de l'économie

L'Économie se développe en trois temps,

- l'économie empirique,
- l'économie positive et,
- l'économie normative.

Les faits

Tout commence par des faits.

«Si nous considérons les phénomènes naturels et sociaux, nous voyons que ces phénomènes ne représentent nullement un assemblage désordonné de faits qu'on ne peut ni comprendre ni prévoir. Au contraire, il suffit d'étudier partout les choses d'un peu près, pour apercevoir une certaine régularité dans les phénomènes ? La première tâche de la science consiste à découvrir cette régularité, à la retrouver parmi le chaos des phénomènes» N. Bouhkarine, La théorie du matérialisme historique. Manuel populaire de sociologie marxiste

Les faits repris par l'analyse économique

- ❑ L'analyse commence par la collecte de données, comprenant la définition de variables pertinentes pour l'objet étudié
- ❑ Elle peut se poursuivre par des prédictions qualitatives (ou théoriques)
- ❑ Elle peut s'achever par le test de la validité d'une théorie (c'est l'économétrie ou l'économie expérimentale),

Économétrie et Economie expérimentale

Définition

L'économétrie est une science sociale qui mobilise les concepts de l'économie théorique, les mathématiques et les statistiques pour analyser empiriquement (avec des données) un phénomène économique

Définition

L'économie expérimentale consiste à expérimenter les comportements économiques individuels et/ou collectifs, et à analyser statistiquement les résultats. Elle collecte des données à partir d'expériences et utilise les méthodes de l'économétrie pour les analyser.



La méthodologie entre Econométrie et Economie expérimentale est assez similaire :

- Choix de la théorie à mobiliser / qualification de l'hypothèse à tester
- Spécification du modèle, avec une forme mathématique simple
- Spécification du modèle sous forme statistique et économétrique / Discussion sur les variables testées expérimentalement
- Collecte des données (éventuellement par l'expérience)
- Estimation des liens entre ces données
- Tests d'hypothèses dérivées du modèle

Corrélation et causalité

Les études appliquées vont établir des corrélations entre des variables, et si possible des liens de causalité

Définition

En probabilités et en statistique, la corrélation entre plusieurs variables aléatoires ou statistiques est une notion de liaison qui contredit leur indépendance.

Cette corrélation est très souvent réduite à la corrélation linéaire entre variables quantitatives, c'est-à-dire l'ajustement d'une variable par rapport à l'autre par une relation affine obtenue par régression linéaire.

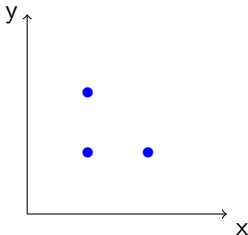
Le fait que deux variables soient « fortement corrélées » ne démontre pas qu'il y ait une relation de causalité entre l'une et l'autre. Le contre-exemple le plus typique est celui où elles sont en fait liées par une causalité commune. Cette confusion est connue sous l'expression *Cum hoc ergo propter hoc*.

Corrélation exemple simple

Soient deux variables X et Y qui prennent chacune deux valeurs, $X_0 < X_1$ et $Y_0 < Y_1$ et dont la distribution jointe est

	Y_1	Y_0
X_1	0	0,4
X_0	0,3	0,3

- Ces deux variables ne sont pas corrélées. On peut s'en rendre compte assez facilement en les représentant sur un graphique à deux dimensions où la variable horizontale est X et la variable verticale Y



Exemple liminaire - Production tablettes chocolat

Supposons qu'une entreprise de production de chocolat recrute un économètre pour déterminer l'effet du prix de la tablette de chocolat sur la demande.

- (i) Théorie : l'analyse de la demande sur un marché d'un bien homogène
- (ii) Modèle $q = b_1 + b_2p$ (on pense a priori $b_2 < 0$. Avez-vous un a priori sur b_1 ?)
- (iii) Modèle économétrique : le modèle n'est pas déterministe, parce qu'il y a d'autres facteurs qui peuvent affecter la demande de tablettes (sa qualité, les phénomènes de mode, etc...)

Exemple liminaire - Production tablettes chocolat

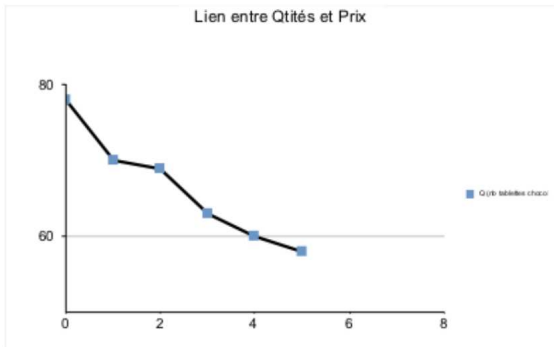
Prenons un exemple de données entre q et p : q est l'estimation de la demande pour p donné.

p	q (nb. tablettes de chocolats)
0	78
1	70
2	69
3	63
4	60
5	58

- ▶ ; Dans la réalité, même si en tendance, il y aurait une certaine linéarité entre prix et quantités, en joignant bout à bout les données, le lien linéaire n'est pas parfait
- ▶ ; En général, on collecte soit des séries temporelles de variables macroéconomiques, soit des séries individuelles, collectées sur un ensemble d'individus statistiques à un point donné dans le temps.

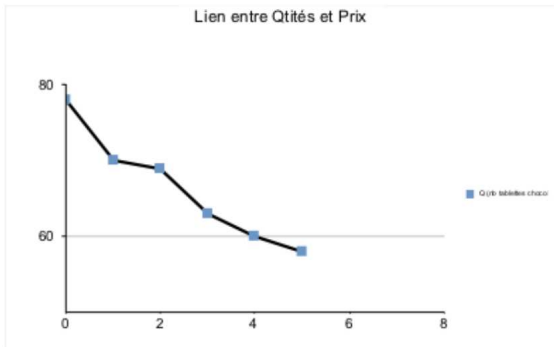
Exemple linéaire - Estimation

(faire le graphique et estimer la relation linéaire)



Exemple linéaire - Estimation

(faire le graphique et estimer la relation linéaire)



$$b_1 \approx 76,05 \quad b_2 \approx -3,88$$

► ; Il y a une vraie question, pourquoi doit-on penser que la relation est linéaire. Même si en tendance il y aurait une certaine linéarité entre les prix et les quantités, en joignant bout à bout les données, le lien n'est pas parfait.

Exemple linéaire - Estimation avec Excel

p	q (nb. tablettes)
0	78
1	70
2	69
3	63
4	60
5	58

Pour "calculer" dans excel l'ajustement linéaire, en représentant la droite de régression, le tableur propose de calculer sa pente et son origine. En utilisant les fonctions pente et origine, qui donnent respectivement la pente et l'abscisse à l'origine de la droite

Pente, ici : -3,885714286

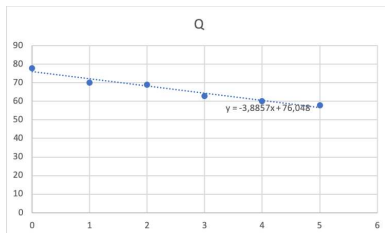
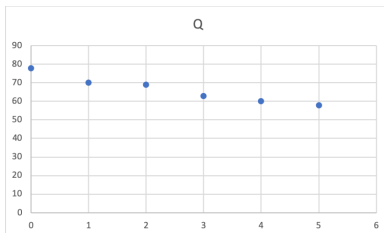
Ordonnée origine, ici : 76,04761905

La droite de régression est donc, suivant cette méthode simple :

$$y = 76,05 - 3,86 x$$

Exemple linéaire - Estimation avec Excel

On pourra écouter la méthode pour représenter le nuage de point, correspondant aux données, et pour représenter le nuage de point avec « la droite de tendance » à l'adresse : https://www.youtube.com/watch?v=6hm0_tfbUSc



Après avoir sélectionné les données dans le tableau

À gauche insertion graphique
« nuage de points »

À droite insertion graphique
« nuage de points » PUIS clic droit
sur l'un des points pour faire apparaître la droite de tendance

Exemple : effet de l'éducation sur le salaire

Dans leur manuel, Colin Cameron et Pravin Trivedi donnent l'exemple de l'évaluation des rendements de l'éducation. On cherche à évaluer l'effet d'une année d'éducation supplémentaire sur le salaire qu'un individu obtient sur le marché du travail. Pour cela, il est courant d'écrire le log du salaire comme une fonction linéaire du nombre d'années d'éducation et d'un certain nombre de facteurs observables ayant une influence potentielle sur le salaire, par exemple le nombre d'années d'expérience sur le marché du travail, le fait d'être une femme, etc. Dans ce cas, le modèle peut alors s'écrire :

$$\log w_i = \alpha_0 + \alpha_1 \text{education}_i + \alpha_2 \text{experience}_i + \alpha_3 \text{femme}_i + u_i \quad \log w_i = \alpha_0 + \alpha_1 \text{ed}$$

avec w_i le salaire de l'individu i , education_i le nombre d'années d'éducation de l'individu i , experience_i le nombre d'années d'expérience sur le marché du travail de l'individu i , femme_i une variable indicatrice valant 1 si i est une femme et 0 sinon et u_i une variable aléatoire représentant l'ensemble des variables non observées dans les données pouvant expliquer le salaire de l'individu i ³⁴.

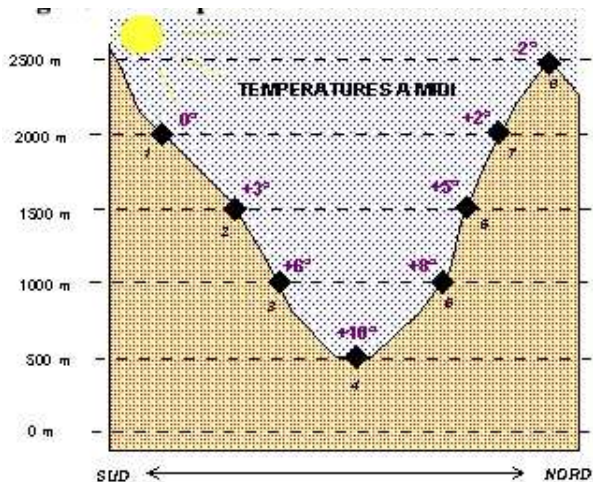
Variété des exemples en économie

- effet causal de la taille des classes sur les performances scolaires des élèves
- effet du nombre de policiers sur la criminalité
- effet des institutions sur le développement actuel des pays
- effet des lois autorisant le travail le dimanche sur la participation religieuse
- évaluer l'estime de soi en fonction du niveau de consommation de cannabis, de l'âge et du sexe

Comment poser le problème

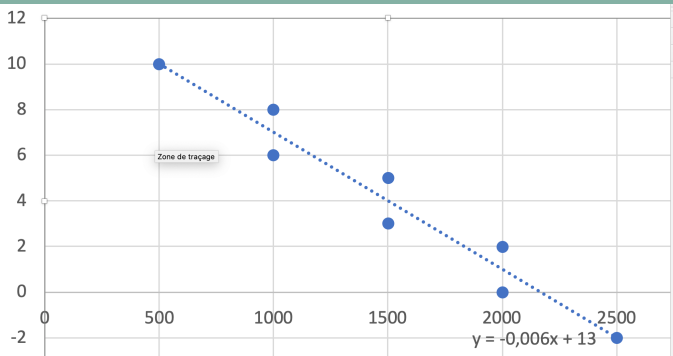
Cherchons par exemple le lien entre l'altitude (X) et température (Y) à l'intérieur d'une région de taille suffisamment petite.

Prenons les données imaginaires d'une donnée alpine



Analyse du pb : série excel et droite

x_i	y_i
2000	0
1500	3
1000	6
500	10
1000	8
1500	5
2000	2
2500	-2

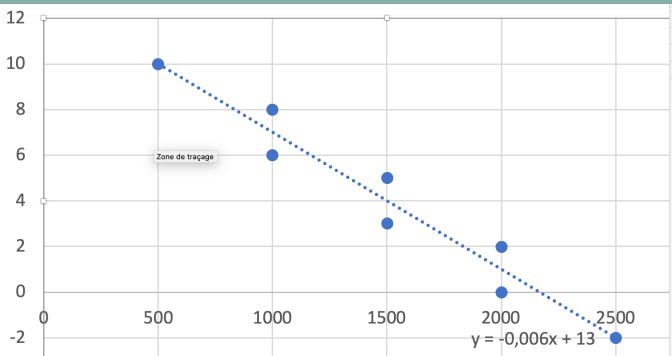


Maintenant que vous savez régresser, comment interpréter le résultat

$$y = -0,006x + 13$$

Analyse du pb : série excel et droite

x_i	y_i
2000	0
1500	3
1000	6
500	10
1000	8
1500	5
2000	2
2500	-2



Maintenant que vous savez régresser, comment interpréter le résultat

$$y = -0,006x + 13$$

Tout simple :

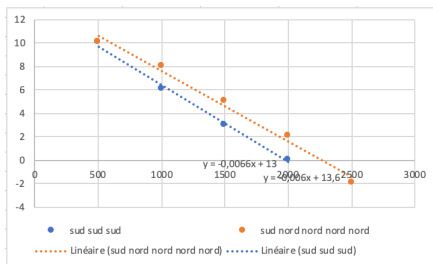
Quand vous montez de 1000 mètres, vous perdez en moyenne 6 degrés de température.

Retour sur l'exemple précédent

Dans cet exemple, on voit bien qu'apparaît en fait deux droites de régression, et il semble au premier coup d'oeil que les données versant S et versant N sont différentes. Il s'agit donc de différencier le nuage de point précédent en deux nuages de points. (Voir le tuto très facile de François MASSON <https://www.youtube.com/watch?v=bndr1B7UUbk>)

Retour sur l'exemple précédent

Dans cet exemple, on voit bien qu'apparaît en fait deux droites de régression, et il semble au premier coup d'oeil que les données versant S et versant N sont différentes. Il s'agit donc de différencier le nuage de point précédent en deux nuages de points. (Voir le tuto très facile de François MASSON <https://www.youtube.com/watch?v=bndr1B7UUbk>)



On distingue alors deux droites de régression

$$y = 13,6 - 0,006x \quad y = 13 - 0,0066x$$

Collecter les données

Différents types de données :

- ❑ les séries temporelles, cad des variables collectées à des intervalles réguliers, tels les données macro, le PIB, le chômage, le CAC 40
- ❑ Les séries individuelles, collectées sur un ensemble d'individus statistiques, à un point donné dans le temps.

Où trouver des données accessibles gratuitement :

- ❑ <https://www.data.gouv.fr/fr/>
- ❑ sinon, répertoriés dans <https://www.stat4decision.com/fr/10-sites-de-referance-open-data/> les sites, www.data.gov , ouvert.canada.ca , data.gov.uk , data.europa.eu , opendata.cityofnewyork.us , datasf.org , opendata.paris.fr , data.sncf.com , donneespubliques.meteofrance.fr

Un autre exemple : l'IFI en 2018



Télécharger le fichier ificom2018.xlsx sur le site du cours,

Quels sont les traitements qui peuvent être faits.

En particulier, on régressera l'impôt versé sur le capital moyen déclaré,

et on interprétera les nombres obtenus

Un dernier exemple : l'aide aux communes

Télécharger le fichier `CG71AideCommune.csv` sur le site du cours,
Quels sont les traitements qui peuvent être faits.