

Rappels (I): Motivation
Master 2 Recherche SES-IES Analyse de données

Ana Karina Fermin

Université Paris Nanterre

<http://fermin.perso.math.cnrs.fr/>

La statistique dans la presse !

Chef économiste de Google : Hal Varian (2010)



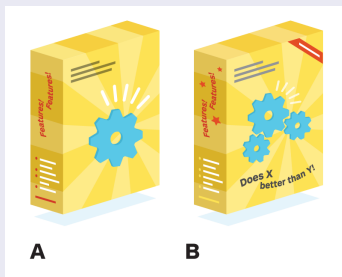
The screenshot shows a news article from Le Figaro. At the top, there is a dark red header with the text "LE FIGARO · fr" in white. Below this, a smaller red box contains the text "LE FLASH ECO". The main navigation path is displayed as "ECONOMIE > CONJONCTURE (ARCHIVES)". The article title is "Le métier de statisticien sera le plus «sexy» de la décennie". Below the title, it says "Par Jean-Pierre Robin | Mis à jour le 18/10/2010 à 11:28 / Publié le 17/10/2010 à 11:27". The main image is a portrait of Hal Varian, a man with glasses, resting his chin on his hand.

La Toile et les moteurs de recherche sont d'extraordinaires pourvoyeurs d'informations et de statistiques. Le tout est de savoir s'en servir.

Motivation I

Packaging A ou packaging B

On demande à des consommateurs s'ils préfèrent, pour un produit de grande consommation qu'on veut relooker, le packaging A ou le packaging B.



Motivation II

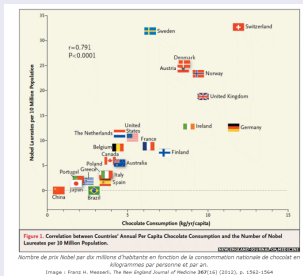
Défaut de Crédit, défaut bancaire, Risque,...



- Données: Profil du client, Historique du crédit du client, ...
- Entrée: Profil du client (profession, montant moyen mensuel d'utilisation de la carte de crédit, revenue,...)
- Sortie: Défaut de crédit

Motivation III

Chocolat et prix Nobel



- Données: 22 pays (The new england journal of medicine)
- Entrée: Consommation de chocolat par habitant
- Sortie: Nombre de lauréats du prix Nobel

Démarche à suivre

En statistique, comme dans la vraie vie, on se pose des questions, et on essaie d'y répondre. Le statisticien cherche à modéliser...

- 1 Visualiser les données et comprendre le problème métier.
- 2 Traduire le problème métier en un problème statistique.
 - Proposer une modélisation mathématique de l'expérience générant ses données.
 - Utiliser une méthode statistique pour proposer une réponse (régression, anova...).
 - Utiliser des outils statistiques pour donner des garanties sur les résultats (intervalles de confiance, tests...).
- 3 Utiliser les résultats pour répondre au problème métier en prenant en compte l'incertitude.

Objectifs du cours

- Présenter les méthodes statistiques de traitement des données comportant des variables qualitatives et quantitatives.
- Traiter à la fois des problèmes de régressions, classification et des problèmes d'apprentissage supervisée avec un accent sur les méthodes dites linéaires.
- Autres types de méthodes ... (si on a le temps)

Évaluation : Un devoir maison (CC) et un examen final.

Remarques importantes

- Ce cours n'est pas un cours de statistiques classique.
- Nous supposons que vous avez déjà une connaissance de certaines méthodes présentées ici.
- Si vous souhaitez des précisions théoriques/méthodologiques à propos d'un certain type d'analyses, nous vous conseillons de voir la doc !

Thèmes abordés dans ce cours

- Introduction : rappels, analyse statistique pour des variables qualitatives et quantitatives. Codage des variables.
Visualisation
- Test d'ajustement ...
- Régression linéaire multiple
- Sélection de modèles
- Régression *logistique* simple. Odds Ratio.
- Régression *logistique* multiple
- Apprentissage supervisée.
- Apprentissage non supervisée (quelques notions)