

Parcours TDMM-Enseignement de statistique

INSA-Departement INFO

- 1 La statistique outil du quotidien
- 2 Connaissances statistiques acquises en TDMM
- 3 Situations abordées en 4ème année

- "Les intentions de votes des rennais..."
- "La campagne de pub a-t-elle été efficace..."
- "Les scientifiques prévoient une augmentation de la Température de 3 degrés..."
- "Le nombre d'expériences est-il suffisant pour la précision souhaitée..."
- "La construction d'un barrage est-elle à l'origine de la pollution de la rivière...?"
- "L'introduction de l'euro a-t-il causé une hausse de l'inflation...?"
- ETC.....!!

Principaux buts.

- Résumer l'information (statistique descriptive...).
- Analyser l'information (modélisation, tests...).
- Préparer l'acquisition de l'information (plans d'expériences).

Des débouchés très diversifiés.

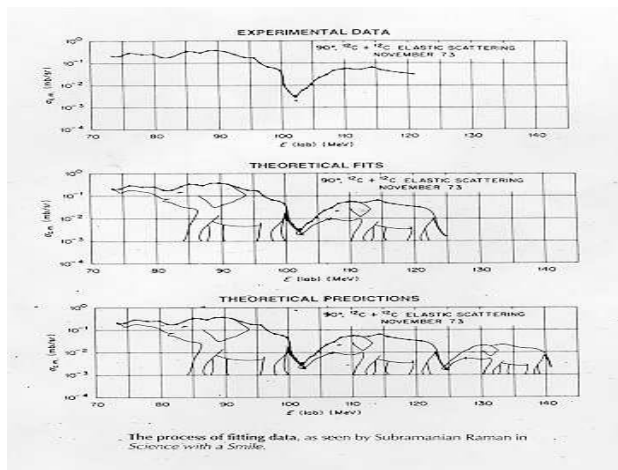
- Instituts de sondages, marketing, droit...
- Informatique, industrie, recherche...
- Banque, consulting, assurance...
- Climatologie, santé, biologie...

Sites.

Société française de statistique: <http://www.sfds.asso.fr/>

- **Description du métier de statisticien en France**
<http://www.phosphore.com/metier/191/nom/Statisticien>
 - **Aux Etats-Unis** <http://www.bls.gov/oco/ocos045.htm>
 - **Offres d'emplois en France**
<http://www.optioncarriere.com/emploi-statisticien.html>
 - **Savoir faire très utilisé**
 - **Mais souvent mal maîtrisé, mal interprété!**
- ⇒ **Connaissances et methodologies sont requises!**

Etape 1: L'information.



Choix de l'information.

- Quelles variables choisir?
- Transformer les données peut aider à améliorer la compréhension d'un phénomène?
- Quelle information passée sélectionner pour prédire le futur?

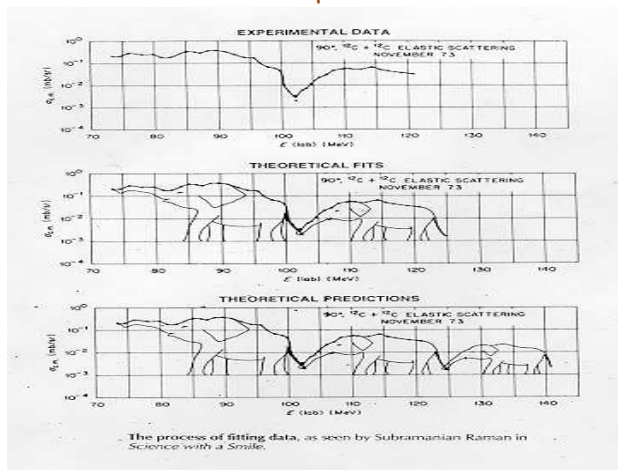
Choix de l'information: Exemples

- ⇒ Expliquer le pouvoir d'achat par les salaires et l'inflation?
- ⇒ Quelle information passée du cours boursier utiliser pour prédire le cours de demain?
- ⇒ Critères à prendre en compte pour identifier un individu à risque pour la grippe?

Choix de l'information.

- Nature de l'information (par exemple: déterministe/aléatoire).
 - Donner des outils "objectifs" (= basés sur les données) pour résoudre ces problèmes...etc
- + Connaissance basique du problème sous jacent → Interaction avec des milieux professionnels divers.

Etape 2: Modélisation statistique.



Choix du modèle: Examen des données.

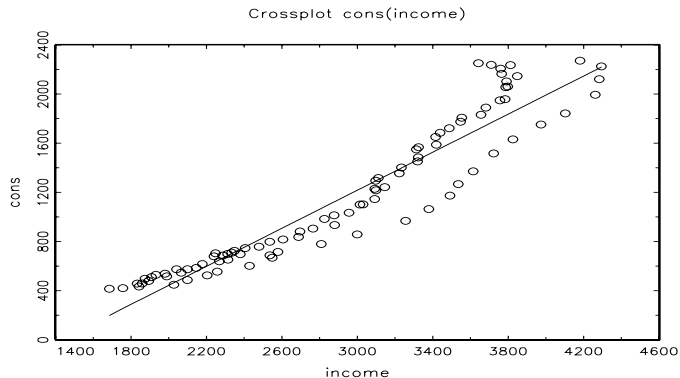


Figure: Consommation en fonction du revenu en Allemagne.

On peut penser à une relation du type $y = ax + b$.

Choix du modèle: Examen des données.

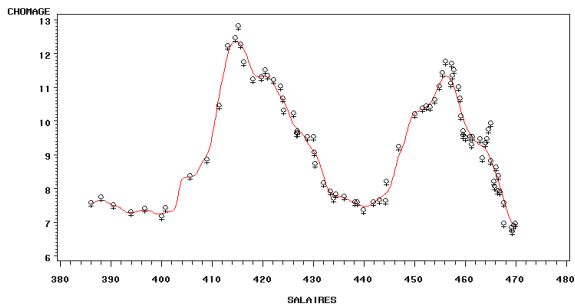
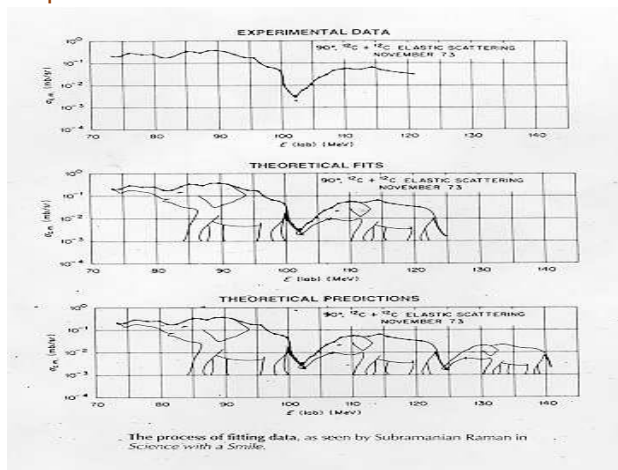


Figure: Taux de chômage en fonction des salaires réels au Canada.

D'autres outils plus objectifs pour le choix de modèles existent...

Etape 3: La prévion.



Sources de la pŕev́ision.

- Choix de l'information exploitée
- Choix de la modélisation

Connaissance du problème sous jacent!

Présentation usuelle de prévisions.

- "Le gouvernement table sur 0.5% de croissance"
- "Ce patient a une probabilité d'avoir deux accidents vasculaires l'année prochaine"

Le statisticien a-t-il une boule magique?

Présentation plus rigoureuse de prévisions.

Votre expérience des prévisions météo (indices de confiance)

- ⇒ La croissance sera entre 0.3 et 0.6 à 95%
- ⇒ La probabilité d'un patient d'avoir deux accidents vasculaires est situé dans l'intervalle ... avec une probabilité de ...

Prédiction: Illustration.

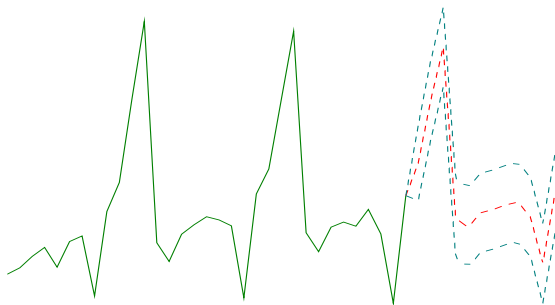


Figure: Prédiction de ventes de champagne.

Aide à la décision.

- Un médicament est-il efficace? (remboursement de la Sécu).
 - Un fruit contient-il trop de pesticides? (commercialisation, autorisation d'importation...)
 - Un champ a-t-il subi une pollution? (justice, ...)
 - Les étudiants issus du département INFO ont-ils un salaire plus élevé que les autres? (Question **vous** concernant!!)
- ⇒ On mène des expériences, (par exemple un sondage...)
- ⇒ Aléa (des informations récoltées...) à "contrôler"!!

Aide à la décision.

- La plupart des tests reposent sur une modélisation.
- On tire de l'information à partir des données.
- Ces informations permettent de prendre une décision avec un niveau de confiance.

Exemple actuel.

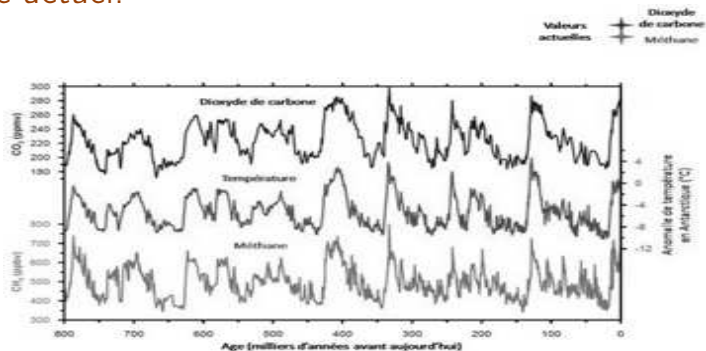


Figure: Taux de concentration du dioxyde de carbone et de méthane et température de l'air au cours des 800 000 dernières années.

Le CO_2 cause-t-il la température ou l'inverse?

Identifier les fraudeurs.

- Dans la ville de "wonderland" on a décidé de taxer les véhicules ayant un moteur en "V", plus polluants. Cette ville voudrait identifier les auteurs de fausses déclarations efficacement!

Banque.

- Une banque veut disposer d'un outil efficace pour repérer les personnes qui sont défailtantes pour le remboursement de leur prêt.

Environnement.

- On veut éliminer les éléments OGM d'un champs à 99%.
Quelle zone traiter pour arriver à ce but efficacement?

Améliorer les logiciels existants!!

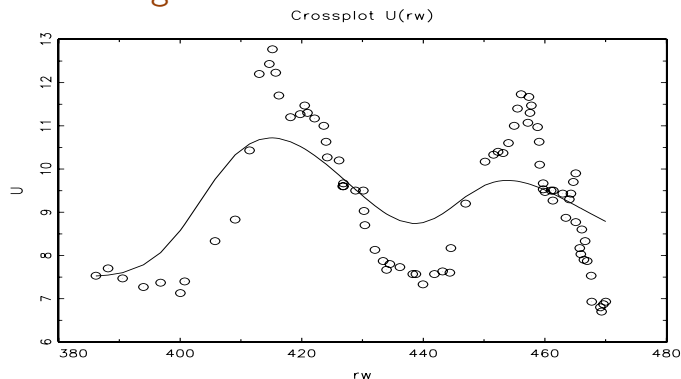


Figure: Taux de chômage en fonction des salaires réels au Canada. Regression en utilisant le logiciel JMulti.

LE RESULTAT NE NOUS SATISFAIT PAS!!

Pratique ludique de la statistique

Analyser la recherche d'un mot clé en utilisant "google insights"