# Evolution proba/stat 1A

13 avril 2017

Equipe pédagogique -> ‘science des données’

# Etat de l’art

- TC S1 : **Modélisation de l'information et de l'aléatoire**
  - Probabilités appliquées
- TC S2 : **Aide à la décision**
  - Principes et méthodes statistiques

# Constat (5 points)

- Changement de programme en prépa
- Hétérogénéité des niveaux (DUT vs MP)
- Introduction de R mal cadrée (vs python)

- Emergence de la ‘science des données’ comme un secteur d’activité porteur->peu visible en filière 2A.
- Faible place faite à la démarche expérimentale et scientifique (en général)

# Objectif : Evolution du TC

- Donner des bases de la ‘science des données’ et de la démarche « scientifique »
- Formation (réelle) à R et à Rstudio
- Mieux prendre en compte l’hétérogénéité des niveaux (soutien, stage de rentrée)
- Support en ligne (type wiki)
- A coût constant (supprimer les permanences?)

# Sciences des données

La science des données est l'extraction de connaissance d'ensembles de données.

Elle emploie des techniques et des théories tirées de la statistique, la théorie de l'information et la technologie de l'information, notamment le traitement de signal, des modèles probabilistes, l'apprentissage automatique, la programmation informatique, l'ingénierie de données, la visualisation, la modélisation d'incertitude, le stockage de données, la compression de données et le calcul à haute performance.

# Ce qui ce fait ailleurs

- Toulouse INSA : Spécialité MA: data science & simulation
  - wikistat.fr
- Telecom ParisTech : filière big data/data scientist/ML
- Ensae ParisTech : filière data science
- Telecom Nancy : Ingénierie des masses de données

# Proposition générale : Fusion des probas - stats

- Objectifs S1 : Langage R, rappel bases de proba, simulation, échantillonnage, variabilité, biais, corrélation, significativité, puissance.

- But: Simuler une expérience, comparer deux méthodes (tests), comprendre les notions de ‘faux positifs/faux négatifs’, d’erreur statistique, etc, programmer en R.

# Proposition générale : Fusion proba-stats

- Objectifs S2 : Conditionnement, régression, modèles gaussiens, vraisemblance, classification, visualisation et analyse des données, bootstrap/validation croisée.

- But: Introduction à l'analyse des données avec R. Résoudre un problème de régression/classification. Comparaison de méthodes.

# Propositions additionnelles

- Ajouter Rstudio dans un stage de rentrée : Rmarkdown, création de projets, version des codes, notebooks.
- Principes de bonnes pratiques : développement, sauvegarde, reporting, reproductibilité.
- Wiki
- Trouver un projet/challenge fil rouge motivant

# Proposition encore plus générale

- Revoir la place de la science des données dans les filières (= vrai secteur d'embauche)

-> Identifier un parcours ingénieur dans l'ecole?

Grenoble INP ensimag

# 2A

- IF : PSAF, IPD, FDAS, SIA
- MMIS : MPA, FDAS, SIA
- ISI : PIEP, FDAS