

# Logiciels Statistiques

*Programmation & Logiciels  
Statistiques*

Cours 7

# Autres logiciels statistiques

# Logiciels libres

- **ADaMSoft** : c'est un logiciel libre open source développé en Java et qui peut fonctionner sur toute plateforme où Java est disponible.
- Les outils suivants sont disponibles sous ADaMSoft :
  - ✓ Réseaux de Neurones
  - ✓ Graphiques
  - ✓ Algorithmes de Data Mining

# Logiciels libres

- ✓ Régression linéaire
- ✓ Régression Logistique
- ✓ Méthodes de classification
- ✓ Arbres de décision
- ✓ Analyse discriminante
- ✓ ACP
- ✓ Analyse des correspondances

# Logiciels libres

- ADaMSoft peut accéder à des données de type
- Texte
- Excel
- ODBC
- MySQL
- Postgresql
- Oracle

# Logiciels libres

- **ADMB** : c'est un logiciel libre open source de modélisation non linéaire développé en C++
- ADMB implémente :
  - ✓ Diverses méthodes de Monte-Carlo par Chaînes de Markov, ce qui le rend utile dans le cadre d'analyses bayésiennes
  - ✓ Les modèles à effets aléatoires
- Il est particulièrement utilisé en statistique environnementale

# Logiciels libres

- **BFL** (Bayesian Filtering Library) : C'est une bibliothèque C++ open source dédiée à l'estimation recursive bayésienne
- BFL implémente :
  - ✓ Le filtre de Kalman
  - ✓ Les filtres particulaires
  - ✓ Les méthodes de Monte-Carlo séquentielles
  - ✓ Les filtres de mélange

# Logiciels libres

- **Chronux** : C'est un paquet open source dédié à l'analyse des séries temporelles neurobiologiques. Il est composé de routines MATLAB et de diverses librairies C
- Chronux implémente :
  - ✓ Les méthodes de régression locale
  - ✓ Diverses méthodes de lissage
  - ✓ L'analyse spectrale des séries temporelles
  - ✓ Divers outils de visualisation des séries chronologiques



# Logiciels libres

- **Dap** : C'est logiciel gratuit d'analyse statistique et de visualisation des données dans la lignée de SAS mais qui nécessite une bonne maîtrise du langage C
- Dap implémente :
  - ✓ Les méthodes de statistique descriptive univariée
  - ✓ L'analyse de variance

# Logiciels libres

- ✓ La régression linéaire
- ✓ La régression logistique
- ✓ Diverses méthodes de statistique non-paramétrique
- ✓ Divers outils graphiques de visualisation
- Dap a été spécialement conçu pour traiter des tableaux de données importants, pouvant excéder la mémoire de l'ordinateur

# Logiciels libres

- **Fityk** : C'est un logiciel de régression et d'analyse des données. Il est très utilisé en chromatographie et en spectroscopie.
- Fityk implémente notamment divers algorithmes de moindres carrés pondérés pour la régression non linéaire :
  - ✓ L'algorithme de Levenberg-Marquardt
  - ✓ L'algorithme de Nelder-Mead
  - ✓ Les algorithmes génétiques

# Logiciels libres

- **Gretl** (Gnu Regression, Econometrics and Times-series Library) : C'est un paquet open-source dédié principalement à l'économétrie.
- Gretl dispose d'une interface graphique qui peut être utilisée avec X12-ARIMA, R, Octave et Ox.
- Gretl est programmé en C.
- Il supporte différentes langues, dont le Français.

# Logiciels libres

- Gretl peut accéder à des données au format :
  - ✓ Ascii
  - ✓ Csv
  - ✓ Eviews
  - ✓ Excel
  - ✓ Octave
  - ✓ SAS xport
  - ✓ SPSS
  - ✓ Stata

# Logiciels libres

- **Incanter** : c'est un logiciel gratuit, de type R pour l'analyse statistique des données, basé sur le langage de programmation Clojure (dialecte de LISP) et sur Java.
- Il implémente notamment divers outils de visualisation des données et les fonctions statistiques usuelles.

# Logiciels libres

- **JAGS** (Just Another Gibbs Sampler) : C'est un programme dédié à l'analyse des modèles hiérarchiques bayésiens par MCMC
- Il est écrit en C++ et peut donc être utilisé sur différentes plateformes, contrairement à WinBUGS et OpenBUGS
- JAGS peut interagir avec R à l'aide du paquet `rjags`

# Logiciels libres

- **JHepWork** : c'est un logiciel interactif dédié au calcul scientifique, à l'analyse des données et à leur visualisation. Il est programmé en JAVA et peut donc être utilisé sur diverses plateformes.
- JHepWork implémente :
  - ✓ Des outils de visualisation 2D et 3D des données, fonctions et histogrammes



# Logiciels libres

- ✓ Calculs scientifiques avec la syntaxe Matlab ou Octave
- ✓ Des générateurs de nombres aléatoires et des méthodes d'échantillonnage
- ✓ Nuages de points, diagrammes de contours
- ✓ Réseaux de neurones
- ✓ Régression linéaire
- ✓ Diverses méthodes de classification, dont CAH

# Logiciels libres

- ✓ Graphiques EPS, PDF, Postscript
- JHepWork est parfois appelé Jwork
- Il fonctionne sous Windows, Mac, Linux et Android
- Sous Android, il est appelé AWork

# Logiciels libres

- **JMulTi** : c'est un logiciel interactif dédié à l'analyse économétrique, en particulier des séries temporelles univariées et multivariées
- Il dispose d'une interface graphique Java
- Il implémente notamment la technique d'Impulse Response Analysis et d'autres techniques avancées non disponibles sous d'autres logiciels

# Logiciels libres

- **Octave** : c'est un langage de haut niveau, principalement dédié au calcul scientifique, qui est en grande partie compatible avec Matlab
- Il implémente diverses techniques statistiques de base mais est surtout utile pour implémenter de nouvelles méthodes

# Logiciels libres

- **Mondrian** : C'est un logiciel dédié principalement à la visualisation des données.
- Les graphiques de Mondrian sont pleinement dynamiques : un point sélectionné sur un graphique est sélectionné sur tous les autres
- Mondrian implémente :
  - ✓ Les nuages de points
  - ✓ Les histogrammes

# Logiciels libres

- ✓ Les diagrammes en barres
- ✓ Les boxplots
- ✓ L'ACP
- ✓ Les techniques de Multidimensional Scaling
  - Mondrian peut accéder à des données au format délimité par des tabulations ou des virgules et lire des données R
  - Mondrian peut interagir avec R

# Logiciels libres

- **OpenBUGS** : C'est un logiciel dédié à l'analyse bayésienne de modèles statistiques complexes par les méthodes de MCMC
- C'est une version open source du logiciel WinBUGS
- Il fonctionne sous Windows et Linux et peut être appelé sous R et S+

# Logiciels libres

- **OpenEpi** : C'est un logiciel gratuit, open source composé de différents programmes dédiés à l'épidémiologie, la biostatistique, la santé publique et la médecine.
- Il a été développé en JavaScript et HTML et peut fonctionner dans un Browser.
- Il peut être utilisé directement sur le site OpenEpi ou être téléchargé et utilisé en local.



# Logiciels libres

- La version disponible sous Windows est appelée Winpepi (freeware)
- OpenEpi implémente :
  - ✓ Le calcul d'intervalles de confiance pour des proportions, taux, taux de mortalité, moyennes, médianes, fractiles
  - ✓ Analyse de tables de contingences 2x2
  - ✓ Analyses cas-contrôle

# Logiciels libres

- ✓ Tests de tendances pour des données de comptage
- ✓ Test de Student et Anova
- ✓ Courbes ROC
- ✓ Calcul de tailles d'échantillon pour des proportions, enquêtes, cohortes, essais cliniques et comparaison de moyennes
- ✓ Calcul de puissances pour des proportions et comparaison de moyennes

# Logiciels libres

- ✓ Générateurs de nombres aléatoires
- OpenEpi offre des fonctionnalités qui ne sont pas disponibles sous d'autres logiciels pour l'analyse épidémiologique, notamment différents tests sur les taux de risque et les odds-ratios.

# Logiciels libres

- **OpenMx** : c'est un programme open source pour la modélisation avancée par équations structurelles.
- Il est implémenté sous la forme d'un paquet R.
- OpenMx fonctionne sur différentes plateformes : Linux, Mac et Windows.

# Logiciels libres

- **Orange** : C'est un logiciel de data mining et d'apprentissage statistique.
- Il dispose d'une interface graphique pour l'analyse exploratoire des données et leur visualisation et interagit avec Python
- Il est programmé en C++ et Python
- Il fonctionne sous Linux, Mac et windows

# Logiciels libres

- **Ploticus** : c'est un logiciel gratuit, open source, pour la visualisation des données.
- Il fonctionne sous Unix, Solaris, Linux, Mac et Win32.
- Il implémente les graphiques usuels de statistique ainsi que les diagrammes de Venn et les arbres.

# Logiciels libres

- **PSPP** : c'est un logiciel gratuit pour l'analyse de données obtenues par échantillonnage / sondage, conçu comme alternative libre à SPSS.
- Il dispose d'une interface graphique et d'une ligne de commande.
- Il est écrit en C et basé sur la GNU Scientific Library.

# Logiciels libres

- PSPP implémente :
  - ✓ Calculs de fréquences
  - ✓ Comparaison de moyennes
  - ✓ Analyse de la variance
  - ✓ Régression linéaire
  - ✓ Outils de fiabilité
  - ✓ Tests non paramétriques
  - ✓ Analyse en Facteurs



# Logiciels libres

- Ses possibilités graphiques sont limitées mais incluent :
  - ✓ Histogrammes
  - ✓ Diagrammes en secteurs (camemberts)
- PSPP peut importer des données au format Gnumeric, OpenDocument, Excel, CSV, ASCII et BDD Postgres.
- Il peut exporter les données au format SPSS.

# Logiciels libres

- **RapidMiner** : C'est un environnement dédié à l'apprentissage statistique, le data mining, le text mining, la prévision et l'analyse de marché.
- RapidMiner implémente diverses procédures de data mining et d'apprentissage statistique, parmi lesquelles :
- ✓ Lecture et transformation des données (Extract Transform Load – ETL)

# Logiciels libres

- ✓ Prétraitement des données et leur visualisation
- ✓ Modélisation
- ✓ Evaluation des modèles
- ✓ Déploiement des modèles
- Il est écrit en Java et fait appel à des outils [Weka](#) et [R](#).
- RapidMiner permet de créer un arbre des opérations à appliquer aux données, comme SAS Enterprise Miner.

# Logiciels libres

- RapidMiner peut être appelé à partir d'autres programmes et dispose également d'une ligne de commande.
- Il est très utilisé en bioinformatique, text mining, multimédia mining et data mining distribué.

# Logiciels libres

- **Revolution R** : Version de R dédiée aux données de grande dimension, basée sur le calcul multiprocesseur et multicore.
- Revolution R est développé par une société privée : Revolution Analytics.
- Il permet d'utiliser R pour analyser des données excédant la mémoire de l'ordinateur, notamment via le package RevoScaleR.

# Logiciels libres

- **Salstat** : c'est un logiciel gratuit pour l'analyse statistique des données, l'accent étant mis sur la facilité d'utilisation.
- Il dispose d'une interface graphique et d'une ligne de commande.
- Salstat implémente entre autres :
  - ✓ Les méthodes de statistique descriptive
  - ✓ Les tests paramétriques et non-paramétriques

# Logiciels libres

- Salstat est programmé en Python et fonctionne sur toute plateforme supportant Python, NumPy, SciPy et wxPython.
- Il est notamment utilisé en climatologie.

# Logiciels libres

- **Scilab** : c'est un logiciel et langage matriciel de programmation open source multiplateforme, dédié au calcul scientifique et qui peut être notamment utilisé pour la Statistique et le Traitement du signal.
- Tout code Matlab peut être converti en Scilab.
- Scilab est disponible sous plusieurs langues, dont le Français.



# Logiciels libres

- **Scipy** : c'est une bibliothèque open source implémentant divers algorithmes mathématiques sous Python.
- Elle peut notamment être utilisée pour le Traitement du Signal et des images.
- Elle peut interagir avec R via le package Rpy.
- Les modules stats et cluster implémentent diverses fonctions de statistique descriptive et de classification.

# Logiciels libres

- **Scikit-learn** : c'est une bibliothèque open source sous Python, dédiée à l'apprentissage statistique.
- Elle implémente notamment :
  - ✓ Diverses méthodes de classification
  - ✓ La régression linéaire
  - ✓ Diverses méthodes d'analyse discriminante
  - ✓ Les SVM (Support Vector Machines)

# Logiciels libres

- ✓ La régression logistique
- ✓ L'analyse discriminante par méthode de Bayes naïve
- ✓ Les k-moyennes
- Elle est conçue pour interagir avec Numpy et Scipy.

# Logiciels libres

- **Shogun** : C'est une boîte à outils gratuite et open source programmée sous C++ et dédiée à l'apprentissage statistique.
- Outre l'Analyse Discriminante Linéaire, Shogun implémente notamment diverses méthodes à noyaux comme les SVM pour la régression et l'analyse discriminante.
- Il implémente également les modèles de Chaînes de Markov Cachées.

# Logiciels libres

- Shogun peut interagir avec Matlab, Octave, Python, R, Java, Ruby, Lua et C#.
- Les noyaux actuellement disponibles pour les données numériques incluent :
  - ✓ Le noyau linéaire
  - ✓ Le noyau gaussien
  - ✓ Les noyaux polynômiaux
  - ✓ Les noyaux sigmoïdes

# Logiciels libres

- **SimFiT** : c'est un paquet Windows open source dédié à la simulation, la régression, la visualisation des données et à la Statistique.
- Il peut être utilisé sous Linux via WINE.
- Il est notamment utilisé en Epidémiologie et en Médecine.

# Logiciels libres

- **SOCR** (Statistics Online Computational Resource) : c'est une suite d'outils online dédiés à l'apprentissage de la Statistique et de la Théorie des Probabilités.
- Elle est développée en HTML, XML et Java.
- SOCR implémente notamment divers outils d'Analyse Exploratoire des Données.

# Logiciels libres

- **SOFA Statistics** : c'est un logiciel statistique open source mettant l'accent sur la facilité d'utilisation et la qualité des sorties.
- Il dispose d'une interface graphique et peut importer des données directement de MySQL, PostgreSQL, SQLite, MS Access, Microsoft SQL Server et CUBRID.
- Il peut également accéder à des données aux formats CSV, Excel, OpenOffice, Gnumeric et Google Docs.



# Logiciels libres

- SOFA Statistics implémente notamment :
  - ✓ Divers tests statistiques : Student, Wilcoxon, Mann-Whitney, Pearson, Kruskal-Wallis, Spearman, Pearson...etc.
  - ✓ Divers outils de statistique descriptive : moyenne, quantiles, médiane, écart-type.
  - ✓ Divers graphiques dynamiques : camemberts, histogrammes, nuages de points...etc.

# Logiciels libres

- SOFA Statistics fonctionne sous Windows, Mac et Linux.
- Il est écrit en Python et fait appel au module statistique de Scipy.
- Les analyses et rapports d'analyse peuvent être automatisés à l'aide de scripts en Python, produits directement par SOFA ou écrits manuellement.

# Logiciels libres

- **Weka** : c'est une suite logicielle gratuite dédiée au data mining et à l'apprentissage statistique, programmée en Java.
- Weka implémente notamment des méthodes de :
  - ✓ Prétraitement des données
  - ✓ Classification
  - ✓ Analyse discriminante

# Logiciels libres

- ✓ Régression
- ✓ Visualisation des données
- ✓ Courbes ROC
- Weka peut accéder aux bases de données SQL ou importer des données au format texte ou CSV.

# Domaine public

- **Demetra+** : c'est un logiciel dédié à l'analyse de saisonnalité développé par Eurostat pour la commission européenne.
- Il est programmé en C++ et Java.
- Il implémente notamment les méthodes de désaisonnalisation de séries temporelles TramoSeats et X12-ARIMA.

# Domaine public

- **CSPro** : c'est un logiciel statistique du domaine public développé par le US Census Bureau et ICF International.
- Il est principalement dédié à la saisie, l'édition, la tabulation et la dissémination des données de recensement ou de sondage.
- Il fonctionne sous Windows.

# Domaine public

- **Epi Info** : c'est un logiciel statistique du domaine public dédié à l'épidémiologie.
- Il est développé à Atlanta (USA).
- Il fonctionne sous Windows et est programmé en Visual C# 3.5.
- Epi Info implémente notamment :
  - ✓ Des méthodes de création de questionnaires
  - ✓ Une interface de saisie des données

# Domaine public

- ✓ Test de Student
- ✓ ANOVA
- ✓ Tests non-paramétriques
- ✓ Tableaux croisés
- ✓ Estimation d'odds-ratios, risques, différences de risques
- ✓ Régression logistique
- ✓ Analyse de Survie (Kaplan-Meier, Modèle de Cox)
- ✓ Analyse de données de sondage à plan complexe



# Domaine public

- **X-12 ARIMA** : c'est un logiciel de désaisonnalisation développé par le US Census Bureau.
- Il peut être utilisé avec d'autres logiciels statistiques, comme **Gretl** ou **Eviews**, qui disposent d'une interface graphique.
- Il succède à X-11 ARIMA.

# Freeware

- **BV4.1** : c'est un programme facile d'utilisation pour la décomposition des séries temporelles et l'ajustement saisonnier.
- Il est développé par l'Office Fédéral de Statistique Allemand.
- C'est un freeware pour une utilisation non commerciale.
- Il supporte les formats Excel, Access, SQL Server et CSV et fonctionne sous Windows.

# Freeware

- **GeoDA** : c'est un logiciel gratuit dédié à l'analyse spatiale des données, à la géovisualisation, à la modélisation spatiale et à l'analyse des corrélations spatiales.
- Il fonctionne sous Windows, Mac et Linux.
- GeoDA implémente également la régression linéaire simple.

# Freeware

- **WinBUGS** : c'est un logiciel de statistique bayésienne basé sur les méthodes de Monte Carlo par Chaînes de Markov (MCMC).
- Il fonctionne sous Windows mais peut être utilisé sous Linux via WINE.
- La dernière version de WinBUGS date de 2007.
- Il n'est plus développé, l'effort portant actuellement sur OpenBUGS.

# Freeware

- **IDAMS** : c'est un logiciel propriétaire développé par l'UNESCO pour le traitement et l'analyse des données numériques.
- Il implémente notamment diverses méthodes de Statistique Exploratoire et d'analyse de séries temporelles.
- Une version spécifique à Windows, WinIDAMS a été publiée en 2006.

# Add-In Excel

- **Analyse-it** : c'est un add-in pour Excel dédié à l'analyse statistique.
- Analyse-it implémente notamment :
  - ✓ Méthodes de Statistique Descriptive
  - ✓ ANOVA
  - ✓ Tests de Mann-Whitney, Wilcoxon, Khi-Deux
  - ✓ Régression linéaire
  - ✓ Régression polynômiale

# Add-In Excel

- **NumXL** : c'est un add-in pour Excel dédié à l'économétrie et à l'analyse des séries temporelles.
- Il implémente diverses techniques d'analyse des séries temporelles, notamment par modélisation linéaire ou non-linéaire et divers tests statistiques.
- Entre autres, NumXL implémente :

# Add-In Excel

- ✓ Calcul des fonctions d'autocorrélation et d'autocorrélation partielle
- ✓ Fonctions de répartitions empiriques, estimation non-paramétrique de la densité
- ✓ Tests de Normalité
- ✓ Modèles linéaires généralisés
- ✓ Modèles ARIMA
- ✓ Modèles ARCH et GARCH



# Add-In Excel

- **SigmaXL** : c'est un add-in Excel implémentant divers outils de la méthodologie Six-Sigma.
- Il implémente notamment :
  - ✓ Diagrammes de Pareto
  - ✓ Histogrammes
  - ✓ Boxplots
  - ✓ Nuages de points
  - ✓ Méthodes de Statistique descriptive

# Add-In Excel

- ✓ Tests de Student
- ✓ ANOVA
- ✓ Tests d'égalité des variances
- ✓ Tests de régression
- ✓ Test du Khi-Deux
- ✓ Tests non-paramétriques

# Add-In Excel

- D'autres add-in Excel sont : [Xlfit](#), dédié notamment à la régression linéaire et non-linéaire, [SPC XL](#), [Stats Helper](#) et [StatBox](#), ce dernier implémentant entre autres la régression linéaire, la régression PLS, la régression logistique, l'ACP et diverses méthodes de classification.

# Logiciels commerciaux

- **Eviews** : Logiciel d'économétrie dédié notamment à l'analyse des séries temporelles.
- Eviews implémente diverses techniques d'estimation et de prévision des séries temporelles ainsi que des techniques d'analyse des données longitudinales et des données de Panel.

# Logiciels commerciaux

- **GAUSS** : c'est un langage de programmation matriciel dédié aux Mathématiques et à la Statistique.
- Il est très utilisé en économétrie.
- Gauss implémente entre autres :
  - ✓ Les méthodes de Statistique descriptive
  - ✓ La régression linéaire
  - ✓ La régression non-linéaire

# Logiciels commerciaux

- ✓ Les modèles de choix discret
- ✓ Diverses méthodes d'estimation par maximum de vraisemblance
- ✓ L'analyse des données de Panel
- ✓ Les modèles ARIMA, VARIMA, VARMAX, VARMA, ARIMAX
- ✓ Les modèles log-linéaires
- ✓ Graphiques 2D et 3D

# Logiciels commerciaux

- **IBM SPSS Statistics** : c'est un logiciel de Statistique, très utilisé en Sciences Sociales.
- Il implémente notamment :
  - ✓ Statistique descriptive univariée et bivariée
  - ✓ Tests de Student
  - ✓ Anova
  - ✓ Tests de corrélation
  - ✓ Tests non-paramétriques

# Logiciels commerciaux

- ✓ Régression linéaire et logistique
- ✓ Analyse en Facteurs
- ✓ Classification (k-moyennes, CAH)
- ✓ Analyse discriminante
- Il dispose d'une interface graphique et de menus
- Il peut lire les données aux formats texte, Excel, BDD et ceux d'autres logiciels statistiques
- Il peut interagir avec des BDD relationnelles via ODBC et SQL



# Logiciels commerciaux

- **IBM SPSS Modeler** : c'est un logiciel complet de data mining et de text mining orienté vers la modélisation prédictive.
- Il implémente notamment :
  - ✓ Diverses méthodes de classification
  - ✓ Les réseaux bayésiens
  - ✓ Le modèle de Cox
  - ✓ L'ACP, l'Analyse en Facteurs

# Logiciels commerciaux

- ✓ La régression logistique
- ✓ Les cartes de Kohonen
- ✓ Les SVM (Support Vector Machines)
- ✓ Les réseaux de neurones
- ✓ La régression linéaire
- ✓ Les modèles linéaires généralisés
- ✓ Divers outils d'analyse de séries temporelles

# Logiciels commerciaux

- **JMP** (« jump ») : c'est un programme d'analyse statistique dynamique des données.
- Un point sélectionné sur un graphique est mis en évidence dans les tables et sur les autres graphiques.
- Il est développé par SAS.
- JMP implémente divers outils statistiques dont des outils de développement de plans d'expériences et de contrôle de la qualité.

# Logiciels commerciaux

- JMP peut accéder à des données au format texte, Excel, SAS et ODBC.
- Il dispose d'un langage de programmation, JSL qui permet, entre autres, de soumettre et de récupérer les résultats d'analyses effectuées avec R.
- JSL permet également de soumettre du code SAS.

# Logiciels commerciaux

- **MATLAB** : c'est un langage matriciel de calcul scientifique, particulièrement adapté à la visualisation 2D et 3D des données.
- En plus des fonctions disponibles en standard sous MATLAB, la boîte à outils « Statistics Toolbox » implémente diverses techniques statistiques, dont l'ACP, la régression linéaire, la classification, l'ANOVA, les tests d'hypothèses, les plans d'expériences.

# Logiciels commerciaux

- **MedCalc** : c'est un logiciel statistique dédié aux sciences biomédicales.
- Il peut accéder à des données aux formats Excel et CSV, entre autres.
- MedCalc implémente notamment :
  - ✓ Les courbes ROC
  - ✓ L'estimateur de Kaplan-Meier et les graphiques associés

# Logiciels commerciaux

- **Minitab** : c'est un logiciel statistique, très utilisé en conjonction avec diverses méthodologies d'amélioration des procédés.
- Il est disponible en plusieurs langues, dont le Français.
- Minitab implémente notamment :
  - ✓ Régression linéaire
  - ✓ Calculs de puissance et de taille d'échantillon

# Logiciels commerciaux

- ✓ Graphiques 2D
- ✓ Analyse en facteurs, ACP, AFC
- ✓ Outils de lissage et de prévision de séries temporelles
- ✓ Outils de contrôle statistique des procédés
- ✓ ANOVA



# Logiciels commerciaux

- **NCSS** : c'est un logiciel statistique généraliste.
- Il peut accéder aux données aux formats texte, Excel, BDD et aux données issues de la plupart des autres logiciels statistiques.
- NCSS implémente notamment :
  - ✓ ANOVA
  - ✓ Régression linéaire et non-linéaire
  - ✓ Plans d'expériences

# Logiciels commerciaux

- ✓ Outils de prévision de séries temporelles
- ✓ Modèles linéaires généralisés
- ✓ Outils de méta-analyse
- ✓ Outils de contrôle de la qualité
- ✓ Modèles mixtes
- ✓ Outils d'analyse de Survie
- ✓ Tests de Student

# Logiciels commerciaux

- **Q research software** : c'est un logiciel d'analyse des données d'études de marché, très utilisé en Marketing.
- Il peut accéder aux données aux formats CSV et SPSS.
- Q implémente notamment :
  - ✓ Régression linéaire
  - ✓ Régression logistique

# Logiciels commerciaux

- ✓ Arbres de régression
- ✓ Arbres de classification
- ✓ Modèles de choix
- ✓ ACP
- ✓ Analyse en classes latentes
- ✓ Modèles de mélange
- ✓ Classification par K-moyennes
- ✓ AFC, ACM

# Logiciels commerciaux

- **RATS** : c'est logiciel d'économétrie et d'analyse des séries temporelles.
- RATS implémente notamment :
  - ✓ Régression linéaire, dont stepwise
  - ✓ Régression non-linéaire
  - ✓ Estimation par maximum de vraisemblance
  - ✓ Modèle ARIMA
  - ✓ Modèles économétriques de grande dimension

# Logiciels commerciaux

- ✓ Analyse spectrale des séries temporelles
- ✓ Réseaux de neurones
- ✓ Filtre de Kalman
- ✓ Modèles ARCH et GARCH
- ✓ Modèles autorégressifs vectoriels VAR
- ✓ Estimation par méthode des moments généralisée GMM

# Logiciels commerciaux

- **SHAZAM** : c'est un logiciel de statistique et d'économétrie.
- SHAZAM peut accéder en standard à des données au format texte, CSV ou Excel.
- L'édition Professionnel dispose d'une connectivité étendue, notamment vers MS Access et divers DBMS.
- SHAZAM implémente notamment :

# Logiciels commerciaux

- ✓ Régression linéaire
- ✓ Outils d'analyse des séries temporelles
- ✓ Tests d'hypothèses et intervalles de confiance
- ✓ ACP et Analyse en Facteurs
- ✓ Simulations de Monte-Carlo
- ✓ Régression non-linéaire



# Logiciels commerciaux

- **SPAD** : c'est un logiciel d'analyse des données et de data mining orienté vers l'analyse predictive.
- Il peut accéder aux données aux formats texte, CSV, Excel, SPSS et sous diverses BDD.
- SPAD implémente notamment :
  - ✓ Statistique descriptive univariée et bivariée
  - ✓ ACP, ACM, AFC

# Logiciels commerciaux

- ✓ CAH
- ✓ ANOVA
- ✓ Régression linéaire
- ✓ Régression logistique
- ✓ Régression PLS
- ✓ Analyse discriminante et scoring
- ✓ Réseaux de neurones
- SPAD peut interagir avec Python et R

# Logiciels commerciaux

- **Stata** : c'est un logiciel statistique généraliste, très utilisé dans le monde de la Recherche.
- Il peut accéder aux données aux formats texte, CSV, Excel et peut lire et écrire au format SAS xport.
- Stata implémente la plupart des techniques statistiques disponibles sous les autres logiciels.

# Logiciels commerciaux

- **Statgraphics** : c'est un logiciel statistique généraliste au fonctionnalités avancées.
- Il se distingue notamment par sa version Mobile qui peut être utilisée sur Pocket PC et Windows Mobile.
- Il existe également une version Online qui peut être utilisée à partir d'un browser et qui permet d'effectuer des analyses statistiques sur un poste distant.

# Logiciels commerciaux

- **STATISTICA** : c'est un logiciel statistique complet implémentant divers outils de régression, classification, prévision et d'analyse exploratoire des données.
- STATISTICA peut interagir avec R, ce qui étend ses possibilités et dispose de fonctionnalités graphiques avancées.

# Logiciels commerciaux

- **Spotfire S+** : c'est une implémentation commerciale du langage S.
- S+ est très analogue à R, si ce n'est qu'il dispose d'une interface graphique complète et d'un support commercial.
- Les versions récentes sont compatibles avec les packages R.
- Les boîtes à outils dédiés aux essais cliniques et à l'analyse de Survie justifient souvent son achat.

# Logiciels commerciaux

- **Unistat** : c'est un logiciel statistique pouvant être également utilisé comme add-in Excel.
- Unistat implémente notamment :
  - ✓ Les graphiques statistiques usuels
  - ✓ Tests de Student, de Fisher, de Levene
  - ✓ Tests d'ajustement, de normalité
  - ✓ Tests non-paramétriques
  - ✓ ANOVA

# Logiciels commerciaux

- ✓ Régression linéaire, dont stepwise
- ✓ Régression non-linéaire
- ✓ Régression logistique
- ✓ Modèle de Cox
- ✓ PCA, CAH, MDS
- ✓ Analyse discriminante linéaire
- ✓ Outils d'Analyse de Survie, dont Kaplan-Meier
- ✓ Outils de Contrôle de la Qualité



# Logiciels commerciaux

- **WPS** (World Programming System) : c'est un logiciel statistique analogue à SAS et pouvant exécuter des programmes SAS.
- Il peut accéder aux données aux formats texte, Excel, Access, SAS xport, SPSS, XML, et à diverses BDD.
- Il fonctionne, sous Windows, Linux, Mac, Solaris.

# Comment choisir ?

- Cela dépend essentiellement des facteurs suivants :
  - ✓ Le niveau d'analyse requis
  - ✓ Le budget dont on dispose
  - ✓ Le système d'exploitation utilisé
  - ✓ Le temps qu'on est prêt à consacrer à l'apprentissage du logiciel
  - ✓ Des logiciels utilisés autour de vous

# Références

- Wikipedia
- Sites internet des différents éditeurs ou développeurs des logiciels