

# FPStat 2, niveau 2, module 1

## Introduction à la décision statistique

Rédacteurs : Francis COLIN, Catherine COLLET, Fabrice DESSAINT, Vincent GINOT, Christelle HENNEQUET, Kiên KIÊU (animateur responsable), François LAURENS, Annick MOISAN (animatrice), Pierre MONTPIED, Catherine RAVEL, Brigitte SCHAEFFER, Pierre WAVRESKY

### 1 Objectifs

Les objectifs sont à la fois :

- prioritairement, d’assurer les **bases de l’inférence statistique** : faire assimiler la démarche intellectuelle de la prise de décision (inférence) statistique ainsi que les grands axes de la problématique statistique (estimation par intervalles de confiance, tests d’hypothèses), afin de pouvoir suivre avec profit les autres modules de niveau 2 qui utilisent ces notions (module 2-3 : *modèle linéaire* et les modules 4 : *planification expérimentale* et 5 : *modélisation de l’interaction* qui lui font suite) ;
- mais aussi de donner aux stagiaires quelques exemples *opérationnels* de méthodes de la statistique *nonparamétrique* classique (vue autant comme technique intéressante à pratiquer en soi, que comme illustration conceptuellement simple des principes de l’inférence statistique), afin de pouvoir reconnaître les cas où ce type de méthodes se révélerait préférable, et consulter alors un statisticien ou la littérature (le livre de P. Sprent, *Pratique des Statistiques Nonparamétriques*, INRA Éditions, 1992, a été traduit en français pour compléter ce module).

### 2 Programme

Le programme comprend quatre parties :

*1. Introduction à l’inférence statistique.* Les notions de variable et de distributions sont introduites en s’appuyant sur des exemples. On parcourt ensuite les représentations graphiques et les résumés numériques (paramètres) de distribution standard. La seconde partie de ce chapitre est dédiée à la problématique de l’estimation. On y présente les estimateurs empiriques les plus courants.

*2. Intervalles de confiance.* La notion d’intervalle de confiance est ici décorquée sur le problème de la position d’un échantillon, avec les trois intervalles de confiance de la moyenne ou de la médiane les plus classiques (associés aux tests du signe, des signes et rangs de Wilcoxon, et de Student). Elle permet

d'aborder ensuite de façon progressive la notion de test d'hypothèse, en ayant déjà rencontré les notions «annexes» (mais importantes) de procédures uni- et bi-latérales, et d'approximations asymptotiques.

3. *Tests d'hypothèses.* La méthode du signe sert de transition entre la construction d'intervalles de confiance et le test d'hypothèse. On présente les notions de base des tests (région de rejet, règle de décision, risques d'erreurs, puissance). Deux autres méthodes de test (test de Wilcoxon et test de Student) sont aussi présentées (sans faire référence aux méthodes correspondantes de construction d'intervalles de confiance). La notion de test d'hypothèses uni- ou bi-latérales suit de même la présentation des intervalles de confiance uni- ou bi-latéraux faite précédemment. La notion de « $P$ -variables» conclut cette partie.

4. *Exemples d'applications de tests.* Ce chapitre est à la carte. Son objectif est double : consolider les acquis des chapitres précédents et donner une certaine autonomie aux stagiaires en leur apprenant à se servir des fiches aide-mémoires qui leur ont été fournies. Il est composé d'une suite de problèmes dans laquelle le formateur et les stagiaires piochent suivant leurs intérêts. Le traitement d'un problème suit le plan suivant :

- formalisation statistique du problème
- exploration graphique des données
- choix d'une méthode de test
- mise en oeuvre du test.

Les problèmes posés portent essentiellement sur la comparaison de deux échantillons. Les tests à mettre en oeuvre ne figurent donc pas parmi ceux abordés dans les chapitres précédents mais sont décrits dans les fiches aide-mémoire (fournies aux stagiaires). Pour l'exploration graphique et la mise en oeuvre de test, on s'appuie sur les logiciels SAS et/ou Splus.

### 3 Principes d'organisation pratique

Les stagiaires auront une copie des transparents, des fiches aide-mémoire et un glossaire-index. Des exercices nombreux ont été prévus pour aider l'assimilation des notions.

Une session de ce module peut se dérouler, selon les circonstances (niveau initial des stagiaires, disponibilités), sur une période de deux (minimum strict) à trois journées.

Les *connaissances requises initialement* des stagiaires ont été volontairement réduites ; on supposera néanmoins qu'ils ont déjà reçu une initiation aux calculs de probabilités les plus élémentaires, et connaissent les notations mathématiques de base (comme les variables avec des indices et la notation  $\Sigma$  pour les sommes, à peu près inévitables en statistique)...

Pour les aider à s'en assurer, les candidats à une session du module voudront bien remplir un *questionnaire de positionnement des connaissances*, destiné également à permettre au formateur de constituer des groupes de stagiaires homogènes, d'ajuster l'intensité de la session en fonction du groupe, et éventuellement de conseiller à certains une préparation préliminaire (lectures personnelles ou stages de niveau plus élémentaire).

## 4 Conception du module

Pour la conception de ce module, nous nous sommes largement appuyés sur le matériel pédagogique déjà réalisé pour FPStat 1 dont les rédacteurs sont

Jean-Pierre LEY	Bernard AUROUSSEAU	Claire CHABANET
Guy FOUILLOUX	Patrick GASQUI	Henriette GOYEAU
François LEFÈVRE	Marianne LEFORT-BUSON	Jeanne-Marie MEMBRÉ
Annick MOISAN	Jean-Claude PIERRAT	François-Xavier OURY
Colette ROUX	Elisabeth de TURCKHEIM	Florence VOLTAIRE.

Les objectifs et le programme ont peu changé par rapport à la version précédente. Les modifications les plus notables sont :

- dans le chapitre 1, la notion de variable est explicitée.
- le chapitre 4 a une orientation très “travaux pratiques” et s’appuie beaucoup sur les fiches aide-mémoire.
- A la fois au niveau du chapitre 4 et des fiches aide-mémoire, on aborde la mise en oeuvre informatique des méthodes présentées. Les logiciels utilisés sont SAS et Splus.

Enfin, pour préparation de ce module, nous avons bénéficié d’aide et de l’appui de Jacques BADIA et Jean-Claude LAVERGNE (responsables FPStat 2). Maïthé FILIPPI a assuré une partie de la saisie des transparents et du cours. Nous les en remercions chaleureusement.