

Allier statistique et informatique : quelles filières après le bac ?

Frédérique Letué(*)

Note de la rédaction :

Cet article contient de nombreux liens qui l'enrichissent et en précisent certains aspects. Pour la commodité de celles et ceux qui souhaiteraient explorer ces liens, une version électronique de l'article est à leur disposition dans le sommaire en ligne du numéro⁽¹⁾. Elle facilitera aussi la diffusion de l'article aux élèves et à leurs parents : ils y trouveront des informations propres à faire naître des vocations. Les vidéos auxquelles l'article fait allusion dans ses différents développements sont regroupées ici : <http://metiers.sfds.asso.fr/web/>. Elle confèrent à l'article une dimension humaine, au-delà de toute formation technique et professionnelle.

Le but de cet article est de présenter un panorama des filières, accessibles après le bac, qui allient des enseignements de statistique et d'informatique dans l'enseignement supérieur public en France métropolitaine.

Dans un premier temps, nous présentons les différentes étapes de la démarche statistique en montrant dans quelle mesure elles s'appuient sur les mathématiques, l'informatique, voire d'autres disciplines. Dans une deuxième partie, nous présentons les différentes formations post-bac jusqu'à bac + 3, en insistant volontairement sur les moins connues d'entre elles : DUT Statistique et Informatique Décisionnelle, licences professionnelles de statistique, parcours des licences de mathématiques préparant aux masters de statistique, licences de Mathématique et Informatique appliquées aux Sciences Humaines et Sociales (anciennes licences MASS), en finissant par un rapide aperçu sur quelques écoles d'ingénieurs spécialisées en statistique, auxquelles on peut accéder après des classes préparatoires aux grandes écoles.

Cette présentation s'appuie sur une revue diffusée par l'ONISEP⁽²⁾ et la Société Française de Statistique⁽³⁾ intitulée « les métiers de la statistique⁽⁴⁾ ». Elle fait notamment appel à des vidéos⁽⁵⁾ diffusées sur le site de la SFdS⁽⁶⁾, présentant des interviews filmées d'anciens étudiants, réalisées à l'occasion de la préparation de cette revue. Elle présente également des cartes des différentes formations, avec des liens sur leurs sites.

(*) Maître de conférence en statistique au Département STID de Grenoble, membre du bureau du groupe Enseignement de la SfdS, Frederique.Letue@iut2.upmf-grenoble.fr

(1) <http://www.apmep.asso.fr/-Les-sommaires-et-articles->

(2) <http://www.onisep.fr/>

(3) <http://www.sfds.asso.fr/>

(4) <http://www.sfds.asso.fr/images/zoom-statistique-2011.pdf>

(5) <http://metiers.sfds.asso.fr/web/>

(6) <http://www.sfds.asso.fr/>

Allier statistique et informatique

La sélection des formations ci-dessous a été essentiellement faite selon la part de l'enseignement de probabilités et de statistique. Dans la plupart de ces formations, on notera la présence d'un enseignement important d'informatique accompagnant ceux de statistique. Pour bien comprendre comment a été pensée cette articulation statistique-informatique, il faut bien avoir en tête la démarche statistique :

- La statistique consiste à répondre à une **question posée par d'autres sciences** (biologie, médecine, économie, psychologie, etc) ou par la société, à l'aide d'un jeu de données. La première qualité d'un statisticien est donc sans doute de bien savoir dialoguer avec d'autres sciences pour bien comprendre la **nature des données** présentes dans le jeu de données et la **question posée**.
- Les données sur lesquelles va travailler le statisticien ont été **collectées et saisies** informatiquement. Ces deux étapes peuvent être séparées (questionnaires ou relevés expérimentaux ou de terrain sur papier, puis saisie informatique), mais sont de plus en plus souvent simultanées (formulaires on-line, logiciels de statistique, etc) afin de limiter au maximum les erreurs de saisie. Le statisticien doit donc savoir comment a été réalisée cette collecte, être conscient des biais induits par la collecte elle-même. Pour cela, il doit être formé à la création, la gestion, la maintenance, l'interrogation des **bases de données**, via des logiciels (Oracle, MySQL, ...). Parfois, un simple **tableur** sera suffisant pour créer une base de données simple. D'autres fois en revanche, il faudra faire appel à des systèmes de gestion de bases de données spécialisés dans un domaine d'application : ERP (*entreprise resource planning*, destiné à la gestion des données d'entreprises), SIG (systèmes d'information géographique, dédié à la gestion des données spatialisées).
- Une fois les données récoltées, le premier travail du statisticien est le **nettoyage** des données. Le statisticien doit en effet savoir quelle fiabilité accorder à ses données.
- L'analyse statistique proprement dite commence par une étape de **statistique descriptive ou exploratoire**. Le statisticien doit avoir une idée de la direction dans laquelle il va pouvoir chercher la réponse à sa question. Il construit alors des tableaux, des graphiques, des indicateurs lui permettant de résumer ses données, sans recours aux modèles probabilistes. La statistique exploratoire multidimensionnelle, aussi appelée **analyse de données**, fait appel à des notions d'**algèbre linéaire**.
- Vient ensuite l'étape de **statistique inférentielle** qui va lui permettre de répondre aux questions initiales, de tirer des conclusions et de prendre des décisions. Celle-ci est essentiellement basée sur la **statistique mathématique**, dont l'objet est de construire des estimateurs et des tests statistiques, dont on sait prouver mathématiquement, à travers des **théorèmes**, les propriétés. Elle s'appuie sur des **modèles probabilistes** et nécessite des

notions d'**analyse**. Le statisticien est parfois amené à illustrer certains de ces résultats par des **simulations**, faisant alors appel à l'**analyse numérique** (pour des problèmes d'optimisation mathématique, par exemple) ou à l'**algorithmique** et à la **programmation** (pour des problèmes d'optimisation numérique).

- Pour appliquer les deux étapes de statistique (descriptive et mathématique) à son jeu de données, le statisticien utilise des **logiciels spécialisés de statistique**, que ce soit des logiciels du marché (SAS, SPSS, Statistica, Sphinx, Stata, ...) ou des logiciels libres comme R.
- En outre, le statisticien doit également être capable à l'issue d'une étude de **présenter ses résultats** aux commanditaires de l'étude. Ils doit donc être capable de **rédiger** un rapport écrit ou un article scientifique, de présenter un exposé oral, aussi bien en français qu'en anglais, voire de concevoir des posters. La maîtrise des **logiciels de bureautique** usuels, des **langues française et anglaise** est donc également indispensable.

Des filières post-bac alliant statistique et informatique

Dans cette partie, nous présentons les filières post-bac reposant sur une part importante de statistique et d'informatique, en montrant comment chacune d'entre elle se positionne par rapport à la démarche statistique présentée ci-dessus.

Le DUT Statistique et Informatique Décisionnelle

Les douze départements Statistique et Informatique Décisionnelle de France accueillent, sur dossier, des bacheliers issus des filières S, ES, L spécialité « mathématiques », STMG spécialité « gestion des systèmes d'information », STI2D spécialité « systèmes d'information et numérique », STL. Certains départements (Paris, Grenoble, Carcassonne) accueillent également en année spéciale des étudiants à bac+2, qui souhaitent se spécialiser en préparant le DUT en un an. Enfin, le département STID Lyon propose un DUT en alternance la deuxième année.

Les compétences visées s'acquièrent en suivant un Programme Pédagogique national (PPN⁽⁷⁾), commun aux douze départements STID. La formation en deux ans est organisée en quatre unités d'enseignement :

- la statistique et les outils mathématiques
- l'informatique des bases de données et en particulier, l'informatique décisionnelle
- des disciplines dites « d'ouverture »: économie, gestion, droit, expression et communication, anglais
- des modules professionnels : projet professionnel personnalisé (PPP), projet tuteuré, stage, ...

L'interaction entre la statistique et les outils informatiques est permanente au cours

(7) http://cache.media.enseignementsup-recherche.gouv.fr/file/24/23/5/PPN_STID_255235.pdf

de la formation, que ce soit lors des TP de statistique qui allient statistique, programmation, maîtrise de logiciels spécialisés, ou lors des projets tuteurés et des stages, pendant lesquels les étudiants doivent mettre en œuvre les techniques acquises dans leurs cours sur des situations réelles.

Le DUT STID a pour ambition de former des techniciens capables de

- créer, gérer, maintenir, interroger des bases de données
- réaliser des analyses statistiques de jeux de données, extraire et résumer l'information pertinente
- exploiter des systèmes d'information décisionnelle, créer des indicateurs et des tableaux de bord, faire du reporting, ...

En cela, ils abordent toutes les étapes de la démarche statistique, en privilégiant les aspects pratiques (pratique des bases de données, de la programmation, de la statistique appliquée à des jeux de données réels issus de différents domaines). En particulier, ils sont experts dans l'utilisation du **tableur**, aussi bien en tant qu'outil purement informatique (par exemple pour gérer des bases de données simples, trier, filtrer des données, créer des macros en VBA, des tableaux de bord, ...) qu'en tant qu'outil statistique (pour recueillir des données via des masques de saisie, créer des tableaux croisés dynamiques, des graphiques, ...). Cet outil, bien que limité sur le plan de l'analyse statistique, est incontournable, car utilisé en interne dans la quasi-totalité des organisations (entreprises, administrations, associations, ...). Les diplômés STID sont également des experts de certains **logiciels statistiques**. Certains de ces outils sont incontournables dans certains domaines d'application : par exemple, pour commercialiser un médicament aux USA, une firme pharmaceutique doit être capable de fournir les résultats de ses études sous un certain format lié à un logiciel précis (SAS en l'occurrence).

Les débouchés de la filière sont très nombreux et variés, que ce soit en terme d'insertion professionnelle, de poursuites d'études courtes ou longues,

- en licences professionnelles (bac+3)
- en licences généralistes (bac+3), suivies d'un master en statistique ou en informatique (bac+5)
- en écoles d'ingénieurs (bac+5)

que dans les domaines d'application :

- santé : essais cliniques, études médicales, pharmaceutiques et vétérinaires, dans les hôpitaux, laboratoires pharmaceutiques et leurs sous-traitants (CRO), organismes paramédicaux (caisses d'assurance-maladie, mutuelles), associations d'aide aux patients, ...)
- enquêtes, sondages et marketing : dans les banques, assurances, instituts de sondage, laboratoire de recherche en sciences humaines et sociales, ...
- qualité et fiabilité : tous types d'industrie
- environnement : organismes de surveillance de la pollution de l'air ou de l'eau, organismes de recherche en environnement, en transport, ...
- services publics ou semi-publics : collectivités territoriales, rectorats, universités, observatoires, ...

Les métiers auxquels peuvent prétendre les diplômés du DUT STID s'intitulent, par exemple, « chargé d'études statistiques », « développeur statistique », « data-manager » (gestionnaire de données), « chargé d'études décisionnelles – développeur décisionnel ». Des exemples de parcours de diplômés STID et d'offres d'emplois ou de stages sont disponibles sur la page de STID France⁽⁸⁾.

Interviews de diplômés STID sur le site de la SFdS :

- Sonia Mahé, responsable outils de pilotage et statistiques
- Aude Guiraudou, chargée d'études statistiques à la SNCF
- Guillaume Main, consultant en web analytique et statistique
- Martina Gramondo, consultante formatrice chez SAS Institute
- Myriam Balmy, chargée d'études statistiques dans une société de services

Carte des départements STID⁽⁹⁾

Les licences professionnelles de statistique

Les licences professionnelles sont des formations à vocation professionnalisante, ouvertes aux étudiants ayant acquis un bac+2. Certaines de ces licences proposent des spécialités en statistique et informatique, en général appliquées à un domaine particulier : on recense en effet des licences professionnelles spécialisées

- en santé : Carcassonne (bioinformatique), Grenoble, Lisieux, Metz, Paris, Vannes,
- en marketing, en enquêtes et sondages : Grenoble, Lyon, Metz, Niort, Vannes, Toulouse
- en informatique décisionnelle : Aix-Marseille, Nice, Paris
- en systèmes d'information géographique : Carcassonne, Grenoble
- en qualité : Pau

Un certain nombre de ces licences professionnelles sont en outre ouvertes en alternance, permettant aux étudiants de passer une partie de leur année en études, et l'autre en tant qu'employé d'une entreprise.

Par rapport à un DUT STID, ces licences professionnelles représentent une véritable spécialisation dans un domaine particulier. On retrouve dans les programmes de ces formations la statistique et l'informatique des bases de données, mais en reprenant des modèles et des exemples spécifiques au domaine d'application. En outre, une unité d'enseignement dans le domaine d'application lui-même est souvent présente, dans laquelle un certain nombre d'enseignements sont assurés par des professionnels du domaine.

Les métiers auxquels mènent ces licences professionnelles sont à peu de choses près les mêmes que ceux d'un DUT STID, à ceci près que la spécialisation acquise dans cette troisième année augmente très certainement les opportunités d'emploi, ainsi que le type de contrat (CDI plutôt que CDD) et le salaire d'embauche. Certaines

(8) <http://www.stid-france.com/>

(9) <https://mapsengine.google.com/map/viewer?mid=zPkq9sfpkYTQ.k7H8hjrW1C9c>

offres d'emploi demandent des bacs +2 ou 3, mais d'autres recherchent explicitement des profils à bac +3 ou plus.

Interview d'une diplômée d'une licence professionnelle santé sur le site de la SFdS :

- Céline Colin, biostatisticienne à l'Inserm

Carte des licences professionnelles en statistique et informatique⁽¹⁰⁾

Les licences généralistes

Les licences de mathématiques

Contrairement à ce qu'on pourrait attendre, la statistique est très peu enseignée dans les licences généralistes de mathématiques. Dans les filières de mathématiques dites « pures », les probabilités sont introduites en L2 ou L3 et la statistique en L3 ou en master. Cependant, quelques universités ont aménagé, au sein de leurs licences de mathématiques, des parcours menant préférentiellement aux masters de mathématiques appliquées et/ou de probabilités et statistique. Le point de vue adopté dans ces filières est très mathématique, très formaliste et souvent très éloigné des données. Dans la démarche statistique présentée dans la première partie, c'est surtout la partie statistique mathématique et son cadre probabiliste qui sont développés. Si application à des données il y a, c'est plutôt sur des données simulées que sur des données réelles.

Les étudiants ayant suivi ces licences pourront néanmoins s'orienter vers des masters de statistique par la suite et les solides bases mathématiques qu'ils y auront acquises seront évidemment un grand atout, notamment pour s'orienter vers des masters recherche en statistique, puis en thèse, notamment en statistique mathématique.

Interview de diplômés de licences de mathématiques sur le site de la SFdS :

- Séverine Demeyer, doctorante en statistique
- Michel Bobbia, ingénieur d'études statistiques chez Air Normand
- Delphine Grancher, ingénieure statisticienne au CNRS
- Stéphane Amarsy, chef d'entreprise

Les licences de mathématiques et informatique appliquées aux sciences humaines et sociales (MIASHS, anciennement MASS)

Les étudiants qui aiment les mathématiques, mais aimeraient garder un contact avec l'aléa plus tôt dans leurs cursus, peuvent s'orienter vers les licences MIASHS (mathématiques et informatique appliquées aux sciences humaines et sociales, anciennement MASS pour mathématiques appliquées aux sciences sociales). Dans celles-ci, en effet, la présence des probabilités et statistique est beaucoup plus forte que dans les licences de mathématiques : elles sont abordées dès la deuxième année (parfois même en première année) et représentent la majorité des enseignements de mathématiques.

(10) <https://mapsengine.google.com/map/viewer?mid=zPkq9sfpkYTQ.kpRYXK3coBXg>

Ceci est motivé par les débouchés visés et par les poursuites d'études possibles : les licences MIASHS forment des statisticiens qui peuvent directement travailler ou également poursuivre des études en statistique, mais aussi des économistes, sociologues, chercheurs en sciences sociales, qui sont des grands utilisateurs de statistique. Parfois, une préparation aux concours de la fonction publique, notamment celui de professeurs des écoles, est proposée. Ils peuvent en outre également prétendre aux masters de statistique, au même titre que ceux issus des licences de mathématiques. Leur culture de mathématique sera sans doute moindre, mais ils auront acquis un savoir-faire en matière de démarche statistique et un recul sur le type de données qu'ils traitent, dont les autres manquent souvent.

Par rapport au contenu en statistique et en informatique, les licences MIASHS s'avèrent être des formations intermédiaires entre les licences de mathématiques et les DUT STID : elles proposent une solide formation de mathématique, où la formalisation n'est pas en reste, mais sont d'emblée orientées vers les applications en sciences sociales. Ainsi l'application à des jeux de données réelles est abordée, et les logiciels de statistique sont utilisés lors de projets ou de TP de statistique. Les étudiants de licence MIASHS n'ont pas autant de technique statistique que des DUT STID, mais seront en revanche mieux armés pour aborder ensuite des masters de statistique où la modélisation, qui nécessite de l'abstraction, est importante. Leur atout indéniable par rapport aux autres formations est la présence d'au moins une discipline des sciences sociales qui leur apporte une vraie connaissance des données traitées : c'est, rappelons-le, le premier point de la démarche statistique.

Interview d'une diplômée de licence MASS sur le site de la SFdS :

- Rebecca Nielbien, économètre-statisticienne, la Société Générale

Carte des licences à forte composante statistique⁽¹¹⁾

Les écoles d'ingénieurs

La plupart des écoles d'ingénieurs généralistes, accessibles après des classes préparatoires aux grandes écoles, offrent des possibilités de s'orienter vers des options de statistique. Cependant, certaines écoles sont vraiment spécialisées dans ce domaine et méritent d'être citées ici.

ENSAE :

Cette école a pour particularité d'accueillir notamment les élèves administrateurs de l'INSEE avec statut de fonctionnaires et des élèves non fonctionnaires se destinant à des emplois de statisticiens ou d'économistes. Elle est accessible pour le concours mathématiques sur le programme de Maths spé MP*, pour le concours économie et sciences sociales sur le programme des khâgnes BL et pour le concours économie et mathématiques sur le programme des classes préparatoires économiques et

(11) Cette carte, qui se veut la plus exhaustive possible, peut révéler quelques oublis. Que les collègues responsables des formations omises n'en prennent pas ombrage et les signalent à l'auteur : Frederique.Letue@iut2.upmf-grenoble.fr, qui se fera un plaisir de les ajouter sur la carte !

commerciales, filière scientifique. En outre, elle est également ouverte sur titre après certaines filières universitaires (M1 en mathématiques, MASS ou économie).

Interview d'une diplômée de l'ENSAE sur le site de la SFdS :

- Odile Rouschmeyer, ingénieure en informatique décisionnelle

ENSAI :

Petite sœur de l'ENSAE, l'ENSAI accueille également des élèves fonctionnaires et des non-fonctionnaires. Elle est accessible pour le concours mathématiques sur le programme de Maths spéciales MP* ou MP, pour le concours économie et sciences sociales sur le programme des khâgnes BL et pour le concours économie et mathématiques sur le programme des classes préparatoires économiques et commerciales, filière scientifique. En outre, elle est également ouverte sur titre après certaines filières universitaires (M1 en mathématiques, MASS ou économie) et de DUT STID ou informatique.

Interview de diplômés de l'ENSAI sur le site de la SFdS :

- Guillaume Chauvet, attaché principal à l'Insee
- Bibbi Ndiaye, consultant indépendant pour une banque

ISUP :

École de formation en statistique, composante de l'Université Pierre et Marie Curie (Paris 6), elle est accessible sur concours sur le programme de mathématiques spéciales MP ou PSI. Elle admet également sur titres des étudiants issus de licences de mathématiques ou MIASHS.

Interview d'un diplômé de l'ISUP sur le site de la SFdS :

- Marc Bagarry, responsable actuariat chez Groupama

Le réseau Polytech

Le réseau Polytech est un réseau d'écoles d'ingénieurs universitaires. Ces écoles sont accessibles sur concours Polytech. Elles proposent également elles-mêmes des classes préparatoires à ce concours (prépa intégrées). Polytech'Lille propose une filière à forte composante statistique : Informatique et Statistique. Polytech'Clermont-Ferrand et Polytech'Lyon proposent des filières de mathématiques appliquées et de modélisation.

INSA

Le groupe INSA est un groupement d'écoles d'ingénieurs accessibles après un bac scientifique sur dossier. Deux INSA proposent ensuite des filières à forte composante statistique :

- INSA Toulouse : génie mathématique et modélisation
- INSA Rouen : génie mathématique.

Interview d'une diplômée de l'INSA Toulouse sur le site de la SFdS :

- Eleonore Gravier, ingénieure biostatisticienne

Conclusion

Au final, il est tout à fait possible pour un bachelier scientifique, mais également un bachelier L spécialité « mathématiques », STMG spécialité « gestion des systèmes d'information », STI2D spécialité « systèmes d'information et numérique », STL, de trouver une formation juste après bac avec une forte part d'enseignements de probabilités ou de statistique. Les enseignants de mathématiques de l'enseignement secondaire doivent en être conscients, car la statistique constitue un moyen d'exercer les mathématiques en entreprise. Les lycéens doivent comprendre que des études de mathématiques ne mènent pas qu'aux métiers de l'enseignement mais offrent des parcours passionnants dans d'autres secteurs d'activités. De plus, ce secteur est en pleine expansion, notamment à l'heure du « big data⁽¹²⁾ », où les masses de données stockées par les entreprises et sur internet requièrent de plus en plus de statistique pour être traitées. Les domaines d'activité⁽¹³⁾ et les métiers⁽¹⁴⁾ y sont variés. Pour vous en convaincre, vous pouvez également visionner la conférence Horizons de la statistique⁽¹⁵⁾, conférence de clôture de l'année mondiale de la statistique.

(12) Voir par exemple les articles parus récemment :

- sur le site de l'Étudiant
- sur le site de Regard sur le numérique
- sur le site des Échos
- sur le site du Monde
- etc

(13) http://www.sfds.asso.fr/54-Statisticien_un_metier_varie

(14) http://www.sfds.asso.fr/55-Quelques_metiers_pour_les_statisticiens

(15) <http://horizons.sfds.asso.fr/>