Introduction

Table of contents

# 1. **Historique de Stata**

L’entreprise **Stata Corp Lp** a été fondée par William Gould. Il est toujours à la tête de l’entreprise.

* Première version (sous MS Dos) en janvier 1985.
* Première version Windows en 1995 (Stata 4).
* La version la plus récente est la 17 [printemps 2021].
* Le rythme de changement de version est de 2 ans. La version 18 devrait donc sortir au printemps 2023. Attention tout de même, il s’agit plutôt de grosses mises à jour, dont l’acquisition systématique s’avère généralement dispensable.

|  |
| --- |
| Version 17 |
| [Présentation](https://www.stata.com/new-in-stata/)  A noter:   * Amélioration de la vitesse d’exécution (je confirme) * Suite **collect** pour gérer et générer les outputs des commandes statistiques. Rapidement testé pour des régressions, et cela me semble plutôt prometteur [rapide présentation](https://mthevenin.github.io/stata_programmation/tableaux/intro.html) |

# 2. **Stata à l’Ined**

Version 17 Windows SE déployée au printemps 2022.  
Version 17 SE sous Linux (serveur margaux).

Petit rappel pour l’Ined: Quel que soit le logiciel choisi sous environnement Windows (SAS v9.4 ou Stata) vous avez accès à une version de l’autre application sous Linux via le serveur Margaux (SAS Studio ou STATA v17 SE).

**Version SE: Standard Edition**

* Nombre d’observations: 2.14 milliards
* Nombre de variables dans les bases: 32 767
* Nombre de variables RHS (limite nombre de colonnes des matrices: 10998)

Il existe également des versions dites MP (Multiple Process), dont la plus puissante gère 120000 variables et 20 milliards d’observations).

# 3. **Les ressources**

## 3.1 Les ressources externes

* Les manuels édités par STATA. Certains sont disponibles au GED [http://www.stata.com/bookstore/books-on-stata]
* \***Stata Journal**, la revue trimestrielle éditée par Stata [http://www.stata-journal.com ]. Depuis 2015 accès total aux versions PDF (via portail GED maintenant). L’accès et le sommaire des 4 derniers numéro est disponible ici: [Lien](https://mthevenin.github.io/stata_fr/Stata_journal.html)
* Une chaîne Youtube [http://www.youtube.com/user/statacorp/videos?view=0] qui propose des tutoriels via les manips par les boites de dialogue.
* Un forum particulièrement actif (<https://www.statalist.org/>)
* Auto-promo: le support que j’ai mis en place à l’été 2022 et qui héberge cette formation introductive [Lien](https://mthevenin.github.io/stata_fr/)

## 3.2 Les ressources internes

* Le manuel de Stata est directement intégré au logiciel. Pour y accéder à partir du menu : help => PDF documentation. ***Il est vraiement de très bonne qualité***.
* Pour une instruction, on peut obtenir son aide en tapant dans la fenêtre *command* de l’environnement principal **help nom\_commande**.
  + Pour les commandes internes, un lien permet d’accéder à l’entrée du manuel officiel, plus riche.
  + Dans les fenêtre de dialogue, on peut accéder directement au fichier d’aide en cliquant sur le point d’interrogation.

help generate

L’aide comprend généralement:

* Le nom de la commande
* Le lien vers l’entrée du manuel si commande officielle. Pour les commandes externes, un lien vers le sitee support peut-être proposé.
* La description de la syntaxe
* Le détail des options
* Des exemples avec éventuellement une exécution directe (voir avec help tw)
* Les informations sauvegardées temporairement (jusqu’à l’exécution d’une autre commande) => *stored results*.
* L’instruction findit nom\_command permet d’obtenir la liste de toutes les ressources disponibles d’une commande, internes ou externes. Exemple: findit fitstat

|  |
| --- |
| Warning: compatibilité entre les versions |
| Des problèmes de compatibilité entre les versions de Stata se sont posés. Cela affecte la lecture des bases entre les différentes versions du logiciel. C’est le cas entre la version 14 et les versions antérieures (dont la version 13 qui date de 2012).  Version 14 (idem 15,16): refonte total du système d’encodage vers le standard UTF8. Les versions inférieures ne peuvent pas ouvrir les bases au format actuel, qui doivent être enregistrées avec une commande particulière (saveold) sous Stata 14 à 17. Les accents ne sont pas lu après cette conversion, et ce qui nécessite d’exécuter un programme pour les convertir (au secours!!!!).  A l’Ined, depuis le déploiement de la version 14 et le passage à la version 15 Linux, il n’y a plus de problème. Attention toutefois pour les collaborations avec l’extérieur. |

# 4. **Les users packages**

Stata fonctionne avec un système hybride de commandes officielles ou usines et de commandes externes (idem R, Python…).  
Commandes externes ou *user package*:

* Il s’agit d’instructions créees par les utilisateurs
* Ces instructions couvrent le data management et l’analyse statistique (descriptive, modélisation, graphiques)
* Même syntaxe que les instructions officielles
* La liste (triée par année) de ces instructions se trouve à cette adresse [http://ideas.repec.org/s/boc/bocode.html]. Elle n’est pas exhaustive, seulement celles sur le dépôt de Stata. Les dépôts dans les gits, en particulier *Github* tendent à se développer. Tout comme R, les dépôts usines ne sont pas forcément les plus à jour.

**Installation directe d’une commande externe**  
Pour télécharger un package, on peut utiliser la ligne de commande:

ssc install nom\_commande, replace

Il est possible d’installer un package après avoir consulté les fichiers avec la commande findit (on peut donc consulter l’aide en amont)

Les programmes des packages externes se trouve dans un répertoire spécifique **plus/ado**, souvent installé dans la racine du lecteur C:. On peut déplacer ce répertoire, et c’est conseillé si on dispose, comme à l’Ined, d’une sauvergarde régulière sur un lecteur. On peut vérifier la localisation de ce répertoire avec la commande sysdir, et plus généralement la localisation des répertoires d’utilisation de Stata.

sysdir

STATA: C:\PROGRA~1\Stata17\  
 BASE: C:\PROGRA~1\Stata17\ado\base\  
 SITE: C:\PROGRA~1\Stata17\ado\site\  
 PLUS: D:\D\ado\plus\  
PERSONAL: D:\D\ado\personal\  
OLDPLACE: c:\ado\

**Exercice**

Ouvrir la base d’exemple *auto*  
Dans la fenêtre *Command*: **sysuse auto, clear**

* [1] Accéder à l’aide (pdf) de l’instruction tabulate. Faire un tableau croisé entre la variable *foreign* et *rep78*
* [2] Installer le package externe findname, exécuter findname t\* et chercher les variables de type chaine de caractère (string)

|  |
| --- |
| profile.do |
| Paramétrer vos sessions avec un fichier ***profile.do*** [[Lien](https://mthevenin.github.io/stata_programmation/profile/profile1.html)]  Sur ce chapitre introductif: déplacer le répertoire où sont stocker les commandes externes dans un lecteur faisant l’objet de sauvergardes régulières sur un réseau. C’est une opération rarement effectué, dans ce fichier, on utilise donc un certains nombre de commandes qu’on aura pas besoin de rechercher le nom, dans le manuel ou sur internet par exemples. |