



Introduction to Stata
Presented by: Binay Shankar
(RA, DOeS, IIT Kanpur)



Official website: - <https://www.stata.com>

According to SatatCorp (2016), Stata is “a complete, integrated statistical software package that provides everything you need for data analysis, data management and graphics”.





Different versions of STATA

- **Stata IC (Intercooled stata)**

This is initial version

Can handle up to 2,048 variables (798 explanatory variables at a time) and 2 billion observations

- **Stata SE (Special Edition)**

Handle up to 32,767 variables (10,998 independent variables) and 2 billion observations.

- **Stata MP (Multiprocessor/multicore)**

Fastest and capable to process huge dataset

Can handle up to 1,20,00 variables (65,532 independent variables) and 10-20 billion observations.





Features	SPSS	SAS	Stata	JMP (SAS)	R	Python
Coding Skills	Basic	Difficult	Basic	Basic	More Difficult	Difficult
User Interface	Point-and-click	Programming	Programming/ point-and-click	Point-and-click	Programming	Programming
Data manipulation	Strong	Very strong	Strong	Strong	Very strong	Strong
Data analysis	Very strong	Very strong	Very strong	Strong	Very strong	Strong
Graphics	Good	Good	Very good	Very good	Excellent	Good
Cost	Expensive	Expensive	Affordable	Expensive	Open Source	Open Source
Released	1968	1972	1985	1989	1995	2008





- Stata is a general-purpose statistical package created in 1985 by StataCorp.
- It can draw various types of statistical graphs, perform regression, time series analysis, handle panel data, treatment effect, item response theory, etc.
- Full form of STATA – **St**atistics and **Data** (syllabic abbreviation)
- First version of STATA i.e. STATA 1.0 was released in 1985.
- Till yet the last version id STATA 17, but it keeps updating





Stata/SE 14.2

File Edit Data Graphics Statistics User Window Help

TOOL BAR

Review Window

Filter commands here

Command _rc

1 use "E:\stata_and_d...

2 log using "E:\stata_...

Result Window

15.1 Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Special Edition

20-student Stata lab perpetual license:
Serial number: 6010029615
Licensed to: IIT Roorkee

Notes:

1. Unicode is supported; see [help unicode_advice](#).
2. Maximum number of variables is set to 5000; see [help set_maxvar](#).

. use "E:\stata_and_data\NSS 73rd round info\nss 73rd round\ (Semi-Round-1)- Block 1-Identification of
> sample enterprises.dta"

. log using "E:\stata_and_data\NSS 73rd round info\nss 73rd round\log.smcl"

name: <unnamed>
log: E:\stata_and_data\NSS 73rd round info\nss 73rd round\log.smcl
log type: smcl
opened on: 22 Jul 2020, 01:45:28

Log Status

log on (smcl)

Command Window

Command

Variable Window

Filter variables here

Name	Label
ENTID	Primary key - unique identifier for an ...
CentreCodeR...	Centre code, Round
Vill_Blkl_Slno	FSU Serial No.
Round	Round
Schedule	Schedule
Sample	Sample
Sector	Sector
State_Region	State-Region
State	State
State_District	State-District
Stratum	Stratum
SubStratumNo	Sub-Stratum-no

Properties Window

Type str30
Format %30s
Value label
Notes

Data

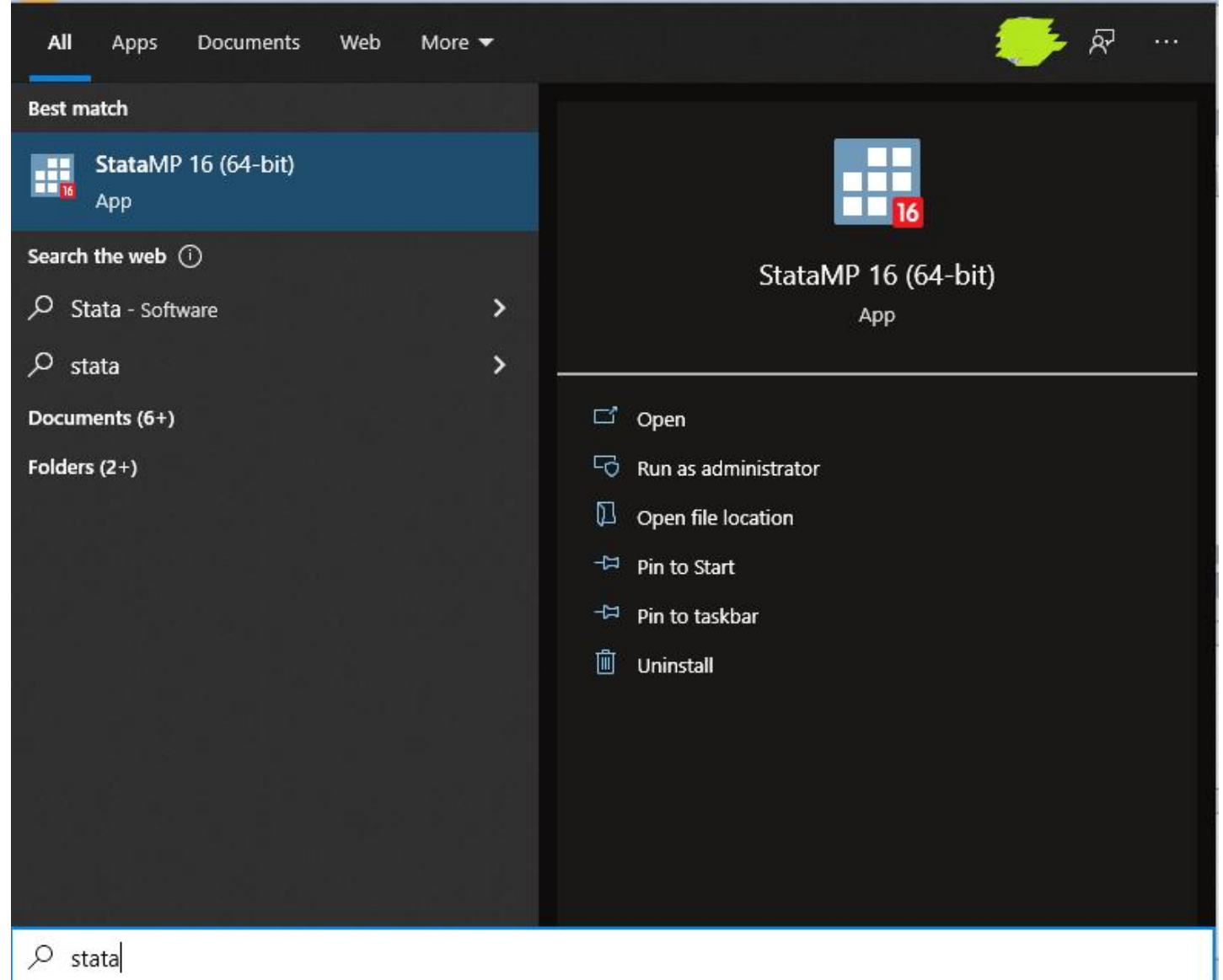
Filename	(Semi-Round-1)- Block 1-Id
Label	
Notes	
Variables	28
Observations	141,744
Size	27.71M
Memory	64M
Sorted by	





Let's Start

Feel free to stop me and ask....





We are running a randomized evaluation in India, with the enumerators sending the data collected daily to you. You have been assigned to do quality checks, known as high-frequency checks.

- a) Our team has obtained data from a partner organization with extra variables that were not asked in our survey, and so we are interested in adding this to our dataset. However, the file that the organization has sent to you, **New Variables.csv**, is not in .dta format. Import this dataset into Stata.
- b) Our main dataset is called **Main Dataset.dta** – open it and then merge in the New Variables dataset to it.
- c) New people have been surveyed today, and their survey responses have been sent to you in the dataset **New Observations.dta**. Add these observations to our main dataset.





Name	Details	Link	Access type
ICPSR	Inter-university Consortium for Political and Social Research	https://www.icpsr.umich.edu/web/pages/	At personal level, one can make id and start access
ICSSR	Indian Council of Social Science Research Data Service	http://www.icssrdataservice.in/	At personal level but require institution affiliation
UK data services	Founded by UKRI – United Kingdom Economic and Research Council	https://ukdataservice.ac.uk/find-data/browse/	At personal level but require institution affiliation





Name	Details	Link	Access type
WHO	World Health organization, Microdata Repository	https://extranet.who.int/ncds/microdata/index.php/catalog/central#r=&collection=&country=&dtype=&from=1999&page=2&ps=&sid=&sk=&sortby=nation&sort_order=&to=2019&topic=&view=s&vk=	At personal level, one can make id and start access
OECD Stats	Organisation for Economic Co-operation and Development	https://stats.oecd.org/	At personal level, one can make id and start access
The DHS program	Demography and Health Survey (US AID)	https://www.dhsprogram.com/	At personal level, one can make id and start access





Name	Details	Link	Access type
DLHS	District Level Household and Facility Survey	http://rchiips.org/index.html	Not open, need to send mail with dd
RBI	Reserve Bank of India	https://rbi.org.in/Scripts/Statistics.aspx	Open Access
EPW	Economics and Political Weekly	https://epwrfits.in/	Not open





Name	Details	Link	Access type
MoSPI NADA	Ministry of Statistics and Programme Implementation, National data Archive	http://microdata.gov.in/nada43/index.php/home	At personal level, one can make id and start access

