Introduction to Stata

Jan Zwierzchowski
15th September 2017
ERFIN Workshop, SGH

What is Stata?

- A general purpose statistical software package created originally in 1985,

- Created by scientists for scientists,

- Most of Stata users work in research,

- Stata's capabilities include: data management, statistical analysis, graphics, simulations and custom programming.

- A new release is usually released every two years, the newest Stata 15 has been released in June 2017.

Why Stata?

- Stata offers a great range of capabilities for a reasonable price,

- Stata licence is perpetual,

- Stata combines the extensibility of open-source packages with attributes of commercial software such as software verification, technical support and professional documentation,

- Stata can be positioned somewhere in between free software packages such as R and expensive, commercial programmes such as SAS, SPSS, etc.

Users support

- Online and stationary trainings,
- Statalist,
- Stata Journal,
- Stata Users Group Meetings,
- Online contact with StataCorp employees.

Stata Journal

- Stata journal is a quarterly publication containing articles about statistics, data analysis, teaching methods, and effective use of Stata's language,

- Main focus on newly released user-created commands,

- Impact factor equal to 1-2.2 during recent years.

Stata Users Groups Meetings

- Organized on a regular basis in many countries,

- Provide a platform for live meetings with other users and StataCorp employees,

- Next meeting in Poland will take place on the 27th November 2017 at the SGH.

User's interface

- Stata was always focused on its commands line,

- Since Stata 8 there has been a graphical interface implemented,

- Every command executed from graphical interface is automatically translated into Stata scripting language,

- Enables replicable analysis.

Do-files

- Scripts created by users which enable for replicable data managment and analysis,

- Contain Stata commands and commands from Stata and Mata programming languages,

- Together with datafiles provide a full technical documentation of any given analysis,

- Stata contains a do-file editor, which is a text editor with some debugging capabilities.

Graphs in Stata
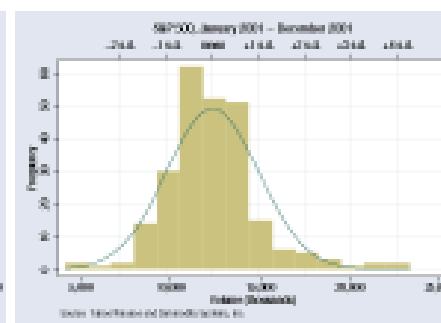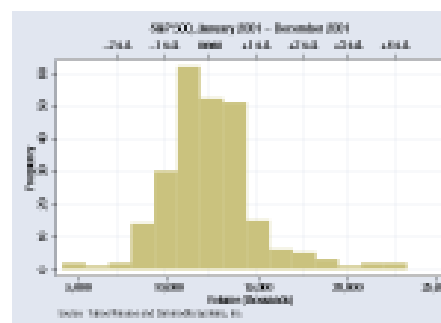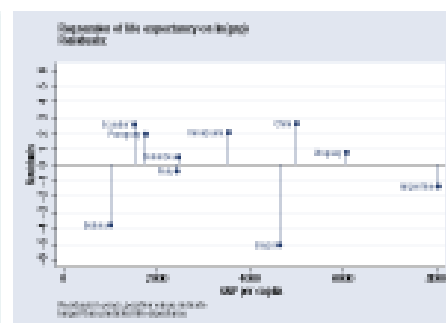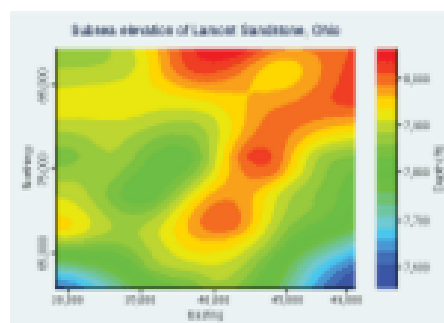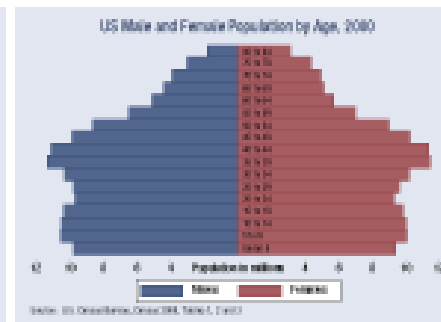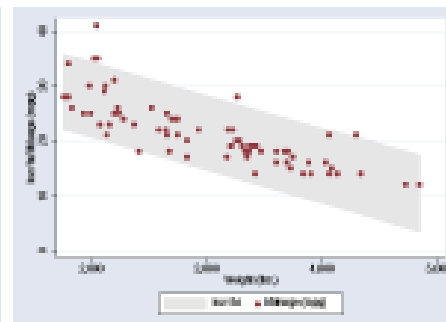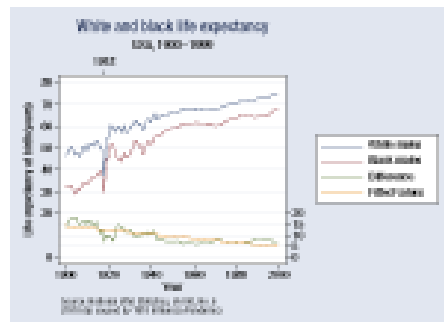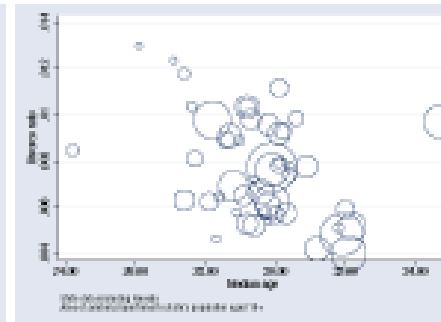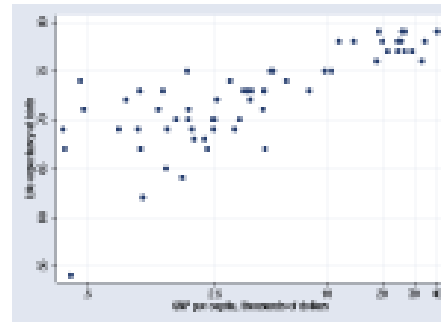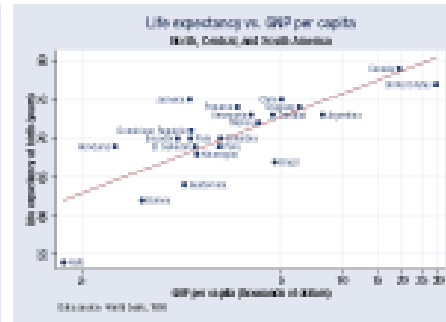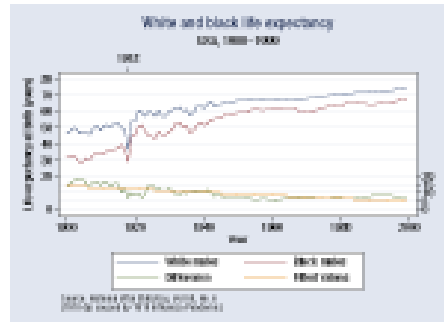
Extensibility

- Stata allows for users written commands to be downloaded from the Internet,

- Users written commands become integrated part of Stata upon download,

- Many of user-written commands becomes part of a subsequent official release after checking, certification, and documentation.

Mata

- Mata is a full-blown matrix programming language integrated in Stata,

- Development environment that can produce compiled and optimized code,

- Includes special features for processing panel data,

- Performs operations on real or complex matrices,

- Provides complete support for object-oriented programming,

- Fully integrated with every aspect of Stata.

Teaching experience with Stata

- Stata is considered a students friendly package,

- During a course „Survey Methods and Statistical Analysis" in SGH we ask students to replicate an analysis from a choosen scientific paper based on open-source datafiles,

- The vast majority of students are capable of learning Stata adequatly within a few weeks in order to produce a „do file" which replicates the analysis from the choosen paper.

Research experience with Stata – AAI

Active Ageing Index

The Active Ageing Index (AAI) is a tool to measure the untapped potential of older people for active and healthy ageing across countries. It measures the level to which older people live independent lives, participate in paid employment and social activities as well as their capacity to actively age.

Employment

Participation in Society

Independent, Healthy and Secure Living

Capacity and Enabling Environment for Active Ageing

(Zaidi et al., 2013, UNECE/European Commission, 2015, p. 13).
"Active Ageing Index project. http://www1.unece.org/stat/platform/display/AAI/Active+Ageing+Index+Home"

AAI Results for Poland
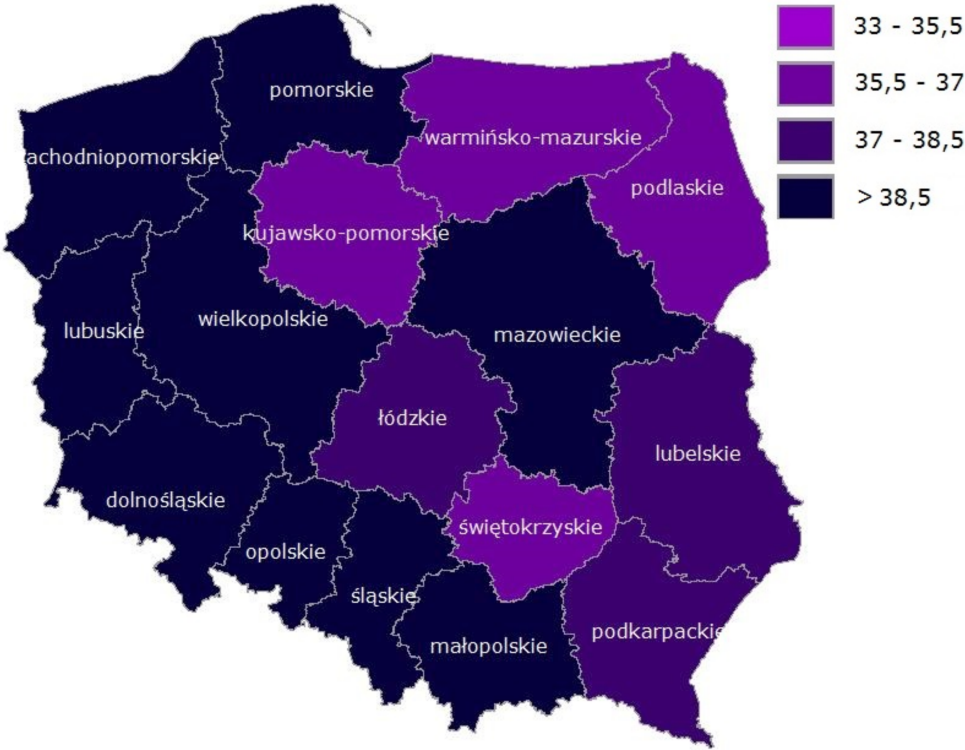
	Województwo	AAI value
		Wartość
1	Dolnośląskie	39,0
2	Kujawsko-pomorskie	34,6
3	Lubelskie	37,5
4	Lubuskie	37,3
5	Łódzkie	36,3
6	Małopolskie	37,4
7	Mazowieckie	39,7
8	Opolskie	38,9
9	Podkarpackie	36,0
10	Podlaskie	34,7
11	Pomorskie	40,2
12	Śląskie	37,6
13	Świętokrzyskie	35,3
14	Warmińsko-mazurskie	33,9
15	Wielkopolskie	37,6
16	Zachodniopomorskie	38,9

AAI assumptions

AAI sensitivity and uncertainty analysis using STATA

Stata was used to run a simulations with thousands of iterations. In each iteration a set of assumptions was drawn from a predefined space of assumptions (set of basic indicators, weighting system, aggregation method):

· Within the sensitivity analysis a variability of results with respect to each assumption was assessed,

· Within the uncertainty analysis ranks of each item were analysed throughout the simulation.

AAI sensitivity and uncertainty analysis using STATA - results

	Variables	Weights	Aggregation
Dolnośląskie	.3341765	.7252939	.5586839
Kujawsko-p~e	.3642183	.8789467	.3971966
Lubelskie	.4064503	.8701917	.3018201
Lubuskie	.3228377	.829451	.450149
Łódzkie	.7300843	.6149125	.4959096
Małopolskie	.4600388	.485167	.7689636
Mazowieckie	.5512358	.8552115	.4031298
Opolskie	.3714355	.8315435	.3766961
Podkarpackie	.3571339	.447581	.8565587
Podlaskie	.3330798	.4936016	.8279633
Pomorskie	.5403852	.6103893	.8283702
Śląskie	.5050216	.8566216	.3161211
Świętokrzy~e	.3448737	.737404	.5377086
Warmińsko-~e	.4017751	.8947249	.3000196
Wielkopols~e	.3805239	.7769633	.5247375
Zachodniop~e	.6778368	.6402881	.714327
Total	.4562804	.7117795	.651048

	5_centyl	mediana	95_centyl
Dolnośląskie	3	5	7
Kujawsko-p~e	11	13	14
Lubelskie	9	12	15
Lubuskie	6	8	13
Łódzkie	8.5	10	11
Małopolskie	8	9	11
Mazowieckie	2	5	6
Opolskie	2	4	7
Podkarpackie	12	15	16
Podlaskie	8	12	15
Pomorskie	1	1	3
Śląskie	2	3	6
Świętokrzy~e	11	14	16
Warmińsko-~e	13	15	16
Wielkopols~e	6	7	9
Zachodniop~e	1	2	3