

## Random samples generation with Stata from continuous and discrete distributions

G. Aguilera-Venegas<sup>1</sup>, J.L. Galán-García<sup>1</sup>,  
M.Á. Galán-García<sup>1</sup>, Y. Padilla-Domínguez<sup>1</sup>,  
P. Rodríguez-Cielos<sup>1</sup>, R. Rodríguez-Cielos<sup>2</sup>

<sup>1</sup>University of Málaga, Spain

<sup>2</sup>University of Madrid, Spain

Spanish Stata Users Group meeting

October 19. Madrid, Spain

# Contents

- 1 Generating random samples from Statistical Distributions
  - Authors' Background
  - Random sample generation using Stata
- 2 Pros and cons of current functions and commands
- 3 Our approach
  - Our commands
  - Comparisons
  - Examples
- 4 Conclusions

## Authors' Background

- Random samples generators using CAS (Computer Algebra Systems)
  - Derive
  - Maxima
- A very important application of generating random samples: Simulations
  - Accelerated Time Simulations (ATS)
    - Traffic control (GRAM, ATISMART, ATISMART+)
    - Baggage handling (ATISBAT)
    - In progress: ATS in biological and medical applications

# Random sample generation using Stata

- Build-in Stata functions
  - `runiform`, `rnormal`, `rbeta`, `rgamma`, `rchi2`, `rt`, `rbinomial`, `rhipergeometric`, `rnbinomial`, `rpoisson`, ...
- Users' contributions
  - `rndwei`, `rndexp`, `rndivg`, `rndlog`, `rndlgn`, `rndf`, `rndchi`, `rndt`, `rndnbx`, `rndbb`, `rndpoi`, ...
  - `rsample`

# Pros and cons of current functions and commands

- Pros
  - Stata functions are fast
  - `rsample` works for generic distributions
  - `rsample` optionally plots the generated sample
- Cons
  - Stata functions only for specific distributions
  - Stata functions do not plot the generated sample
  - `rsample` very slow when the size is high
  - `rsample` needs the user to introduce suitable limits
  - The size in `rsample` cannot be easily changed

## Our commands

- Include new distributions not considered in Stata functions
- Are fast even for high sizes
- Work with suitable limits automatically computed
- Can easily change the size of the sample
- Optionally plot the generated sample
- Optionally compute the Median Squared Error
- Display time spent in the generation
- `scauchy`, `sexponential`, `slognormal`, `snormal`,  
`spareto`, `sweibull`, `sbinomial`, `sdiscreteuniform`
- Other continuous and discrete distributions in progress

## Comparisons

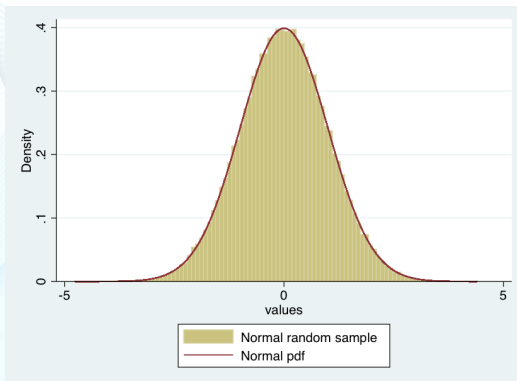
Distribution	Command	Time	Error	Plot
Normal(0,1)	rnormal	1.150e-07	1.030e-06	No
	snormal	1.360e-07	9.772e-07	Yes
	rsample	.00044102	.00001524	Yes
Pareto(8,1)	rpareto	Not available in Stata functions		
	spareto	1.090e-07	9.739e-07	Yes
	rsample	.00044182	.00029966	Yes

# Examples

- `snormal 10000000`
- `snormal 100000, pl(1)`



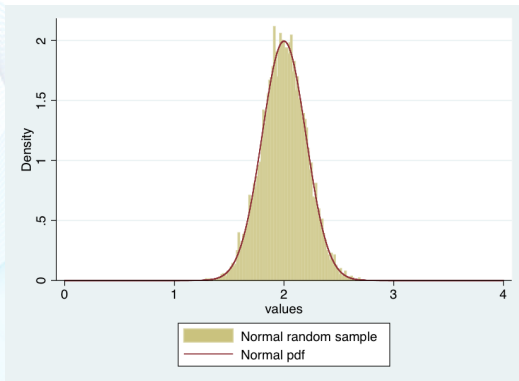
# Examples



## Examples

- `snormal 10000000`
- `snormal 100000, pl(1)`
- `snormal 100000, mse(1)`
- `snormal 10000, m(2) s(0.2) le(0) ri(4) mse(1) pl(1) nr(10)`

# Examples



## Conclusions

- New commands for random numbers generation from continuous and discrete distributions
- Same time order in computation as build-in stata functions
- Computation of media squared error (optionally)
- Display mean time spend (optionally specifying the number of iterations)
- Plot the generated random sample (optionally)
- Computation of suitable limits automatically (user can change them)
- Improve the time, error and default bounds regarding `rsample`

## Random samples generation with Stata from continuous and discrete distributions

G. Aguilera-Venegas<sup>1</sup>, J.L. Galán-García<sup>1</sup>,  
M.Á. Galán-García<sup>1</sup>, Y. Padilla-Domínguez<sup>1</sup>,  
P. Rodríguez-Cielos<sup>1</sup>, R. Rodríguez-Cielos<sup>2</sup>

<sup>1</sup>University of Málaga, Spain

<sup>2</sup>University of Madrid, Spain

Spanish Stata Users Group meeting

October 19. Madrid, Spain