Random Data

Generate random data sample.
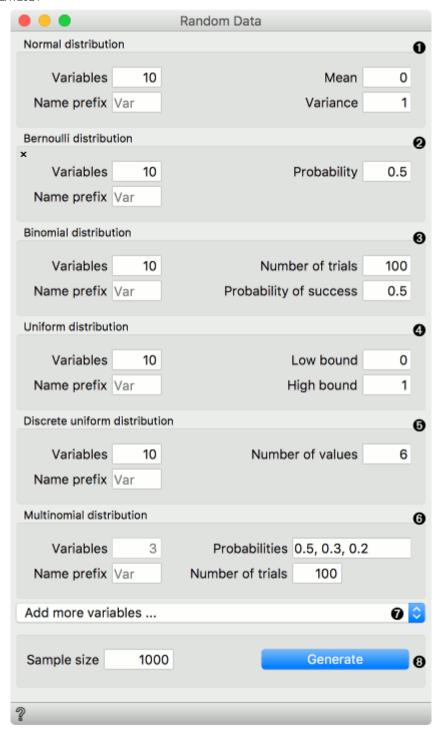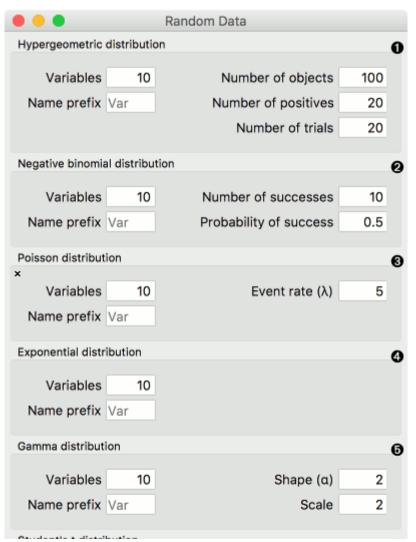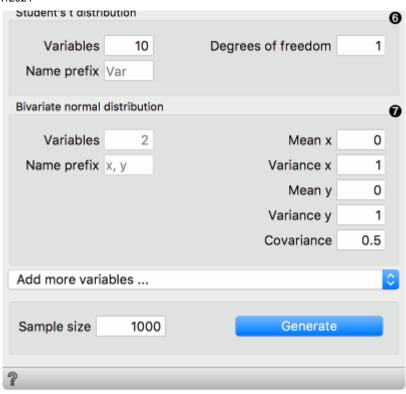
Inputs

- None

Outputs

- Data: randomly generated data

Random Data allows creating random data sets, where variables correspond to the selected distributions. The user can specify the number of rows (samples) and the number of variables for each distribution. Distributions from the Scipy's stats module are used.

Random Data

Normal distribution ❶

| Variables | 10 | Mean | 0 |
| Name prefix | Var | Variance | 1 |

Bernoulli distribution ❷

×

| Variables | 10 | Probability | 0.5 |
| Name prefix | Var | | |

Binomial distribution ❸

| Variables | 10 | Number of trials | 100 |
| Name prefix | Var | Probability of success | 0.5 |

Uniform distribution ❹

| Variables | 10 | Low bound | 0 |
| Name prefix | Var | High bound | 1 |

Discrete uniform distribution ❺

| Variables | 10 | Number of values | 6 |
| Name prefix | Var | | |

Multinomial distribution ❻

| Variables | 3 | Probabilities | 0.5, 0.3, 0.2 |
| Name prefix | Var | Number of trials | 100 |

Add more variables ... ❼

| Sample size | 1000 | Generate | ❽ |

?

1. **Normal**: A normal continuous random variable. Set the number of variables, the mean and the variance.
2. **Bernoulli**: A Bernoulli discrete random variable. Set the number of variables and the probability mass function.
3. **Binomial**: A binomial discrete random variable. Set the number of variables, the number of trials and probability of success.
4. **Uniform**: A uniform continuous random variable. Set the number of variables and the lower and upper bound of the distribution.
5. **Discrete uniform**: A uniform discrete random variable. Set the number of variables and the number of values per variable.
6. **Multinomial**: A multinomial random variable. Set the probabilities and the number of trials. The probabilities should sum to one. The number of probabilities corresponds to the final number of variables generated.
7. *Add more variables…* enables selecting new distributions from the list and with that adding additional variables. Distributions can be removed by pressing an X in the top left corner of each distribution.
8. Define the sample size (i.e. number of rows, default 1000) and press *Generate* to output the data set.
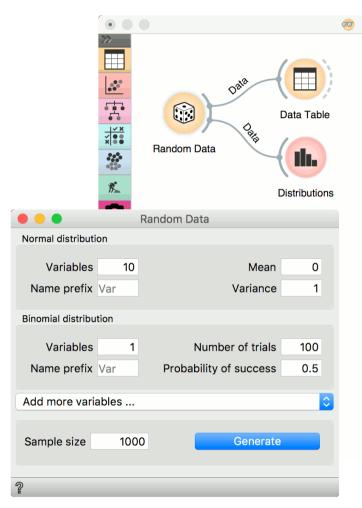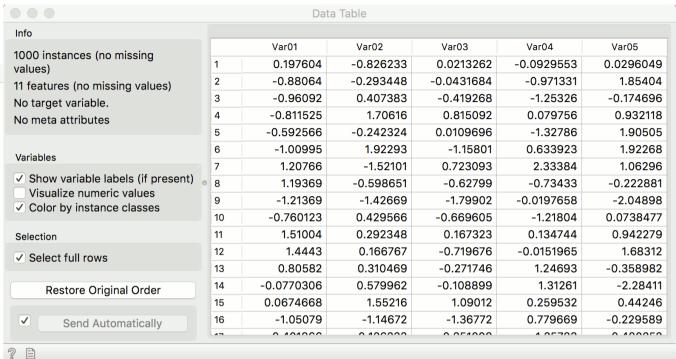
1. **Hypergeometric**: A hypergeometric discrete random variable. Set the number of variables, number of objects, positives and trials.
2. **Negative binomial**: A negative binomial discrete random variable. Set the number of variables, number of successes and the probability of a success.
3. **Poisson**: A Poisson discrete random variable. Set the number of variables and the event rate (expected number of occurrences).
4. **Exponential**: An exponential continuous random variable. Set the number of variables.
5. **Gamma**: A gamma continuous random variable. Set the number of variables, the shape and scale. The larger the scale parameter, the more spread out the distribution.
6. **Student's t**: A Student's t continuous random variable. Set the number of variables and the degrees of freedom.
7. **Bivariate normal**: A multivariate normal random variable where the number of variables is fixed to 2. The number of variables is set to two and cannot be changed. Set the mean and variance of each variable and the covariance matrix of the distribution.

Example

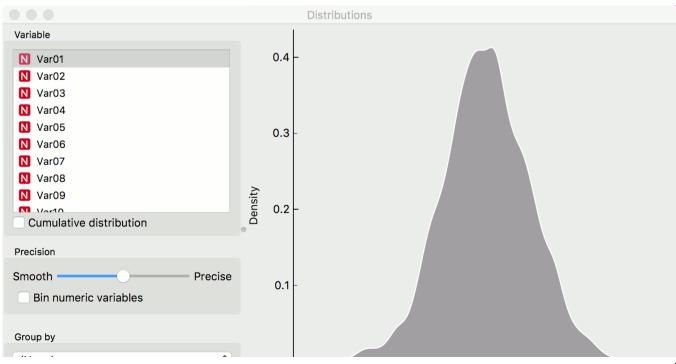We normaly wouldn't create a data set with so many different distributions but rather, for instance, a set of normally distributed variables and perhaps a binary variable, which we will use as the target variable. In this example, we use the default settings, which generate 10 normally distributed variables and a single binomial variable.

We observe the generated data in a **Data Table** and in **Distributions**.

Data Table

Info

1000 instances (no missing values)
11 features (no missing values)
No target variable.
No meta attributes

Variables

☑ Show variable labels (if present)
☐ Visualize numeric values
☑ Color by instance classes

Selection

☑ Select full rows

[Restore Original Order]

☑ [Send Automatically]

	Var01	Var02	Var03	Var04	Var05
1	0.197604	-0.826233	0.0213262	-0.0929553	0.0296049
2	-0.88064	-0.293448	-0.0431684	-0.971331	1.85404
3	-0.96092	0.407383	-0.419268	-1.25326	-0.174696
4	-0.811525	1.70616	0.815092	0.079756	0.932118
5	-0.592566	-0.242324	0.0109696	-1.32786	1.90505
6	-1.00995	1.92293	-1.15801	0.633923	1.92268
7	1.20766	-1.52101	0.723093	2.33384	1.06296
8	1.19369	-0.598651	-0.62799	-0.73433	-0.222881
9	-1.21369	-1.42669	-1.79902	-0.0197658	-2.04898
10	-0.760123	0.429566	-0.669605	-1.21804	0.0738477
11	1.51004	0.292348	0.167323	0.134744	0.942279
12	1.4443	0.166767	-0.719676	-0.0151965	1.68312
13	0.80582	0.310469	-0.271746	1.24693	-0.358982
14	-0.0770306	0.579962	-0.108899	1.31261	-2.28411
15	0.0674668	1.55216	1.09012	0.259532	0.44246
16	-1.05079	-1.14672	-1.36772	0.779669	-0.229589

Random Data

Normal distribution

Variables 10   Mean 0
Name prefix Var   Variance 1

Binomial distribution

Variables 1   Number of trials 100
Name prefix Var   Probability of success 0.5

[Add more variables ...]

Sample size 1000   [Generate]

Random Data

Data Table

Distributions

Distributions

Variable

Ⓝ Var01
Ⓝ Var02
Ⓝ Var03
Ⓝ Var04
Ⓝ Var05
Ⓝ Var06
Ⓝ Var07
Ⓝ Var08
Ⓝ Var09
Ⓝ Var10

☐ Cumulative distribution

Precision

Smooth ———————— Precise
☐ Bin numeric variables

Group by

(None)

☐ Show relative frequencies

Show probabilities:

0

-4   -3   -2   -1   0   1   2   3   4

Var01