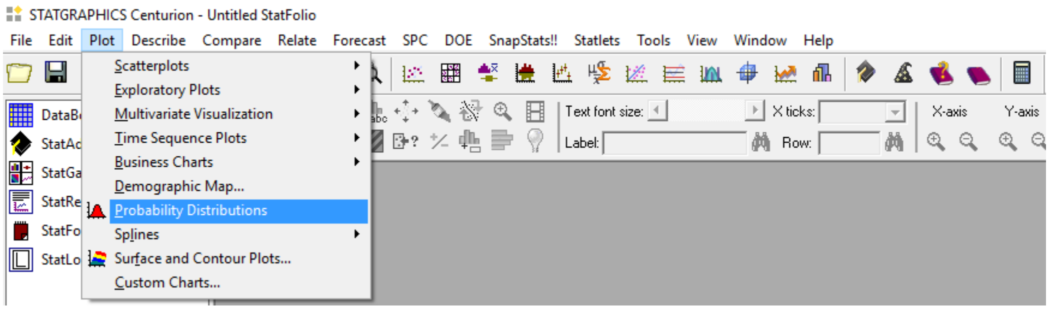
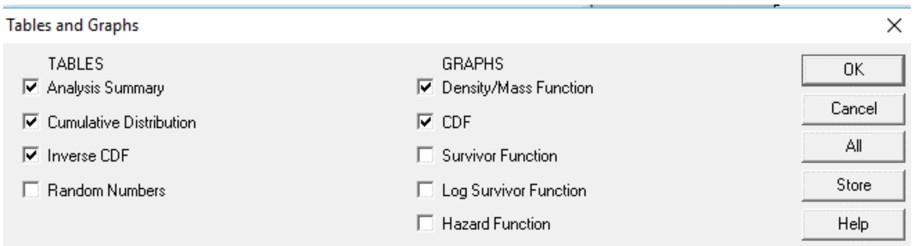
# Normal distribution

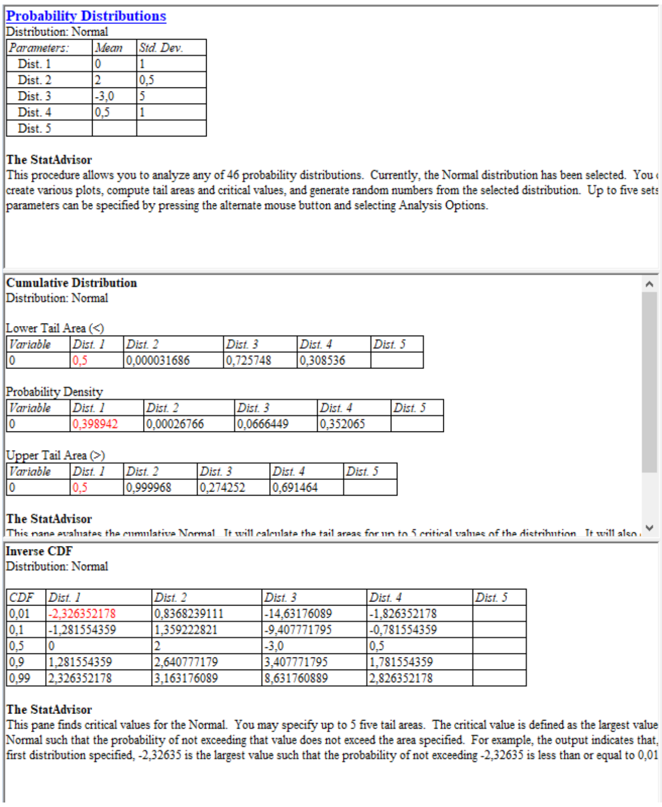
1. **Plot the normal probability distribution:**
   a) From the main menu choose ***Plot/Probability*** distribution: ***Normal***,




   b) Input ***Mean*** and ***Std. Dev***, choose a few values,
   c) Choose items from ***Tables and Graphs*** window and click OK:




   d) Analyse the results:



**Probability Distributions**
Distribution: Normal

| *Parameters:* | *Mean* | *Std. Dev.* |
|---|---|---|
| Dist. 1 | 0 | 1 |
| Dist. 2 | 2 | 0,5 |
| Dist. 3 | -3,0 | 5 |
| Dist. 4 | 0,5 | 1 |
| Dist. 5 | | |

**The StatAdvisor**
This procedure allows you to analyze any of 46 probability distributions. Currently, the Normal distribution has been selected. You c create various plots, compute tail areas and critical values, and generate random numbers from the selected distribution. Up to five sets parameters can be specified by pressing the alternate mouse button and selecting Analysis Options.

**Cumulative Distribution**
Distribution: Normal

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,5 | 0,000031686 | 0,725748 | 0,308536 | |

Probability Density

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,398942 | 0,00026766 | 0,0666449 | 0,352065 | |

Upper Tail Area (>)

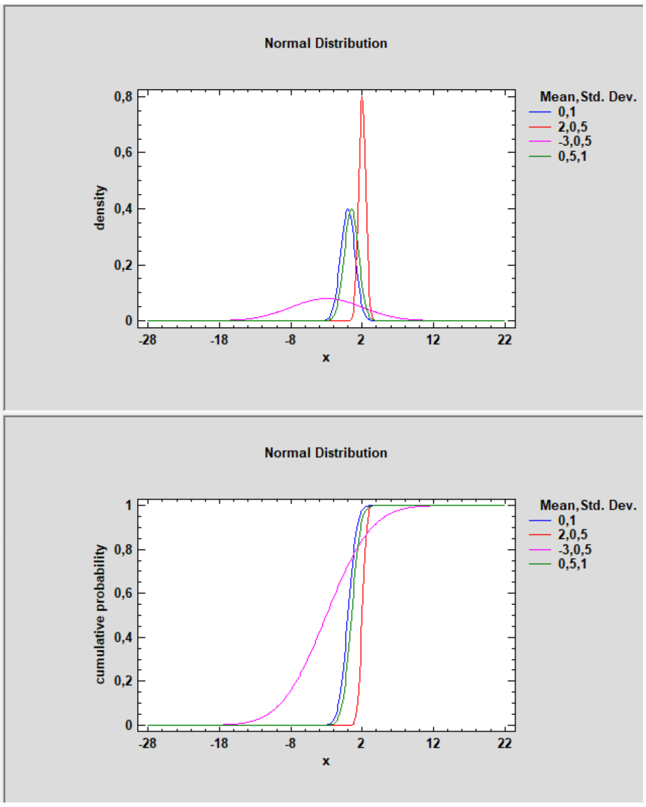
| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,5 | 0,999968 | 0,274252 | 0,691464 | |

**The StatAdvisor**
This pane evaluates the cumulative Normal. It will calculate the tail areas for up to 5 critical values of the distribution. It will also

**Inverse CDF**
Distribution: Normal

| *CDF* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0,01 | -2,326352178 | 0,8368239111 | -14,63176089 | -1,826352178 | |
| 0,1 | -1,281554359 | 1,359222821 | -9,407771795 | -0,781554359 | |
| 0,5 | 0 | 2 | -3,0 | 0,5 | |
| 0,9 | 1,281554359 | 2,640777179 | 3,407771795 | 1,781554359 | |
| 0,99 | 2,326352178 | 3,163176089 | 8,631760889 | 2,826352178 | |

**The StatAdvisor**
This pane finds critical values for the Normal. You may specify up to 5 five tail areas. The critical value is defined as the largest value Normal such that the probability of not exceeding that value does not exceed the area specified. For example, the output indicates that, first distribution specified, -2,32635 is the largest value such that the probability of not exceeding -2,32635 is less than or equal to 0,01




2. Using Statgraphics solve the problem:

The final exam scores in a statistics class were normally distributed with a mean of 63 and a standard deviation of five.

a) Find the probability that a randomly selected student scored more than 65 on the exam.

b) Find the probability that a randomly selected student scored less than 85.

c) Find the 90th percentile (that is, find the score k that has 90% of the scores below k and 10% of the scores above k).

d) Find the 70th percentile (that is, find the score k such that 70% of scores are below k and 30% of the scores are above k).

Copy or write down your results. Show them to the teacher.

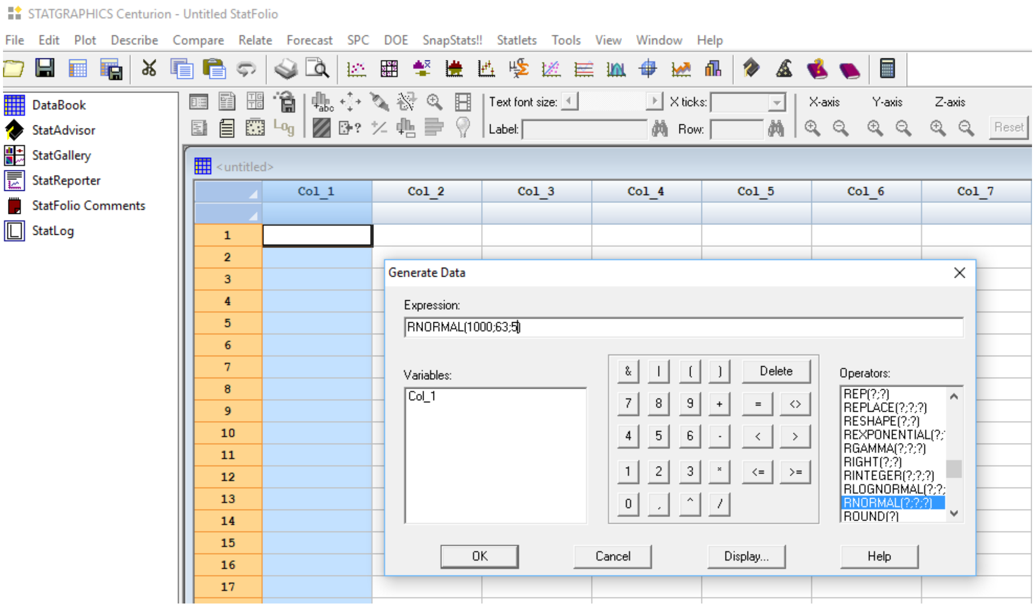
**Hint:** use ***Inverse CDF*** Table. Right-click on the table and chose ***Pane*** option. Write appropriate values (be sure you what you are doing) into the window "***Inverse CDF Option***".

3. Transform the normal distribution from the previous problem into the standard normal distribution $N(0,1)$ and answer the same question. Compare the results.

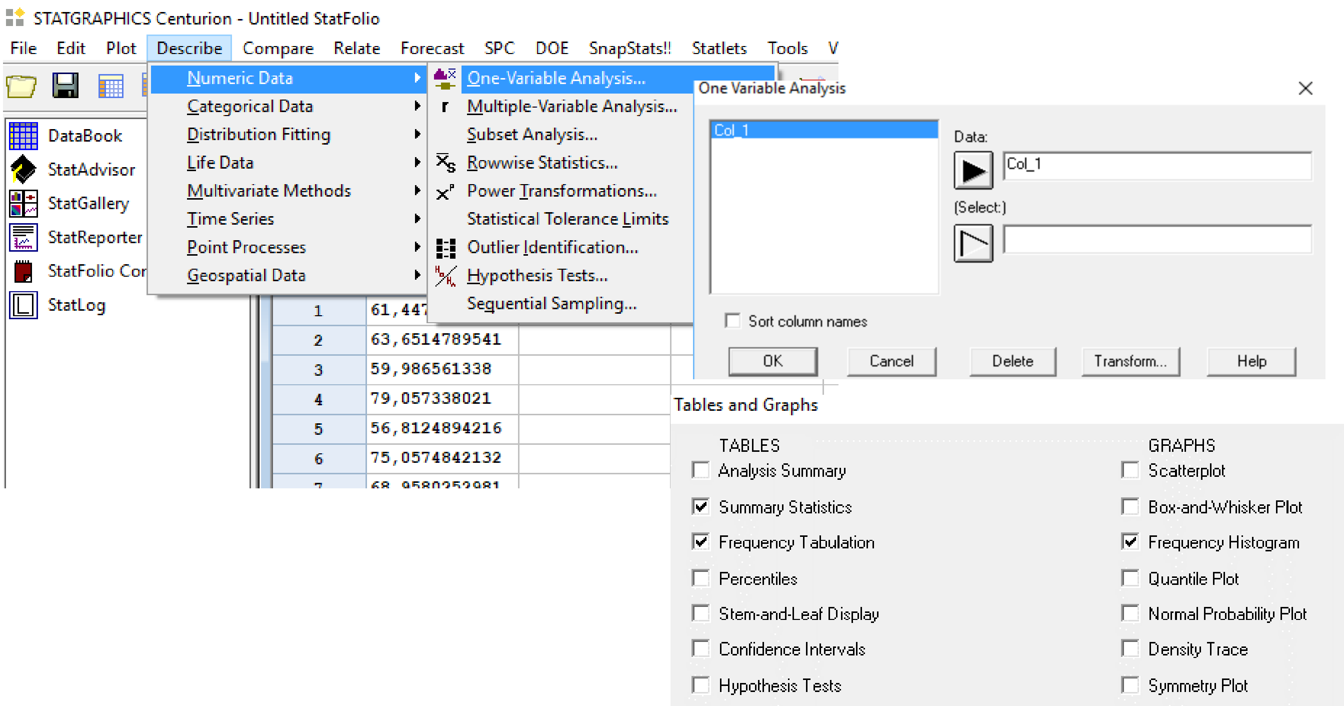
   **Hint:** calculate the parameters of a standardised variable (from a known formula). Plot two normal distribution functions for two random variables. Then analyse both distributions in the same table.

## 3. Generation of data (histograms) with a normal distribution

a) From ***DataBook*** click on ***Col_1*** and chose ***Generate Data***. Scroll down ***Operators*** window to see ***RNORMAL(?;?;?)***. Doble click on it, replace the "?" with: numbers of points to generate, mean, standard deviation. Press ***OK***. You will have numbers in ***Col_1***.



b) From the main menu: ***Describe->Numerical Data->One-Variable Analysis***, highlight ***Col_1***, and press ►, tick options from ***Tables and Graphs*** window




c) Analyse the results.

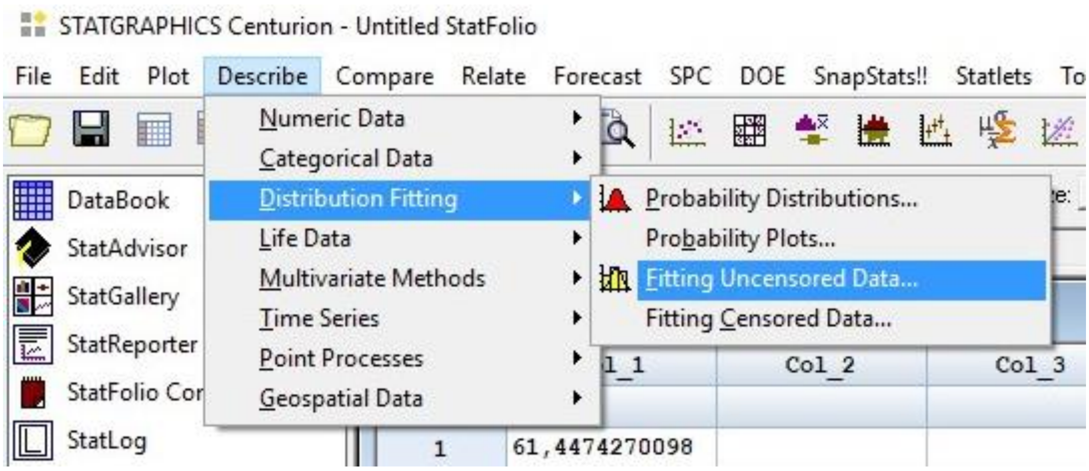
d) You can change the binning of histogram: right-click on the plot, chose ***Pane*** option, change value in ***Frequency Plot Options***. Note the difference in ***Frequency Tabulation window*** when choosing lower or higher ***Number of Classes***

**4. Fit your data (histogram) with a normal distribution**

***a)*** From the main menu: ***Describe->Distribution Fitting->Fitting Uncensored Data***

b) From ***Distribution Fitting Option*** window tick ***Normal***.

c) Analyse the results.



| Introduction to Probability, Statistics and Data Handling | statgraphics 18 |
|---|---|
| AGH UST ESA LAB 2a | Everything (almost) is normal |

I. Binomial distribution can be approximated by the normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$.

A multiple-choice test has 15 questions, each of which has: i) five choices, ii) two choices. An unprepared student taking the test answers each of the questions completely randomly by choosing an arbitrary answer. Suppose $X$ denotes the number of answers that the student gets right. The student passes the exam if the number of correct answers is at least 8. Calculate the probability of his/her success.

Solve the above problem using:

a) Binomial distribution;

b) Normal distribution.

a) From the main menu choose **Plot/Probability distribution**: ***Binomial,*** input parameters for both options: i) five choices, ii) two choices.

b) Calculate the probability from the `Cumulative Distribution Panel`:

**Cumulative Distribution**
Distribution: Binomial

Lower Tail Area (<)

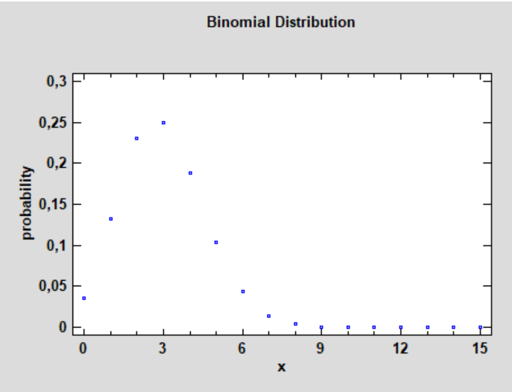
| Variable | Dist. 1 |
|---|---|
| 1 | 0,0351844 |
| 7 | 0,981941 |
| 8 | 0,995761 |
| 9 | 0,999215 |

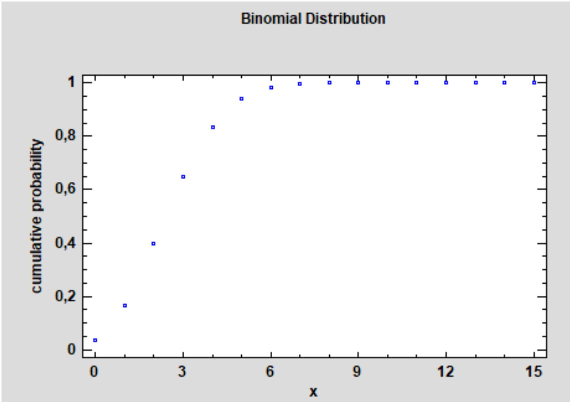
Probability Mass (=)

| Variable | Dist. 1 |
|---|---|
| 1 | 0,131941 |
| 7 | 0,0138191 |
| 8 | 0,00345476 |
| 9 | 0,00067176 |

Upper Tail Area (>)

| Variable | Dist. 1 |
|---|---|
| 1 | 0,832874 |
| 7 | 0,00423949 |
| 8 | 0,000784728 |
| 9 | 0,000112968 |




c') Copy the plot Binomial Distribution to StatGallery:

- *RightClick on the* `Plot ->Copy PanetoStatGallery`
- In *StatGallery* Window *RightClick Paste*
- Choose Overlay option to superpose plots

c) Plot the normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$.

Copy plots to StatGallery with *Overlay* option.

Compare and comment the results. Is the approximation correct?

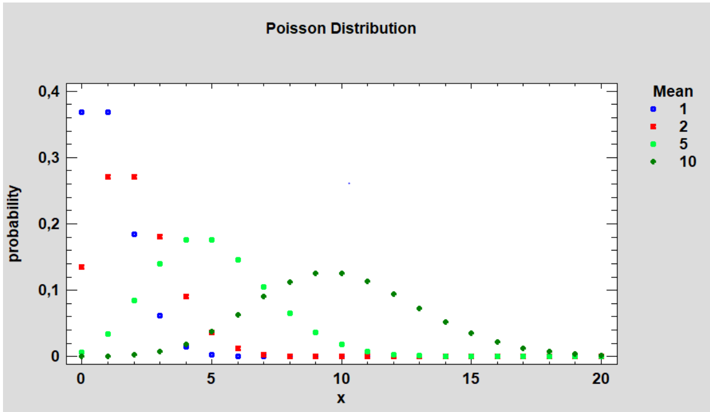
In order to get the best approximation, add 0.5 to x or subtract 0.5 from x (use x + 0.5 or x – 0.5). The number 0.5 is called the continuity correction factor.

II. The **Poisson probability distribution** gives the probability of a number of events occurring in a **fixed interval** of time or space if these events happen with a known average rate and independently of the time since the last event (number of failures, guests at the hotel, fish caught, etc). The Poisson distribution is given by the function: $\boldsymbol{f(n;\nu = n \cdot p) = \frac{\nu^n}{n!}\, e^{-\nu}}$ ; $n$ stands for the number of occurrence, $\nu$ is a mean value,

Plot a few Poisson distributions with $\boldsymbol{\nu} = \{1, 2, 5, 10\}$ and compare the shapes :

a) in the limit of large $n$ and very small $p$ (rare events) **binomial distribution becomes Poisson distribution,**

b) if $n$ is large then it can treated as a continuous RV following the **normal distribution.**




III. t-Student distribution.

a) Plot t-Student distributions for $\boldsymbol{n} = \{1, 3, 5, 10, 30\}$ on the same plot. Scale the x-axis to (-5,5) with step 2 (right click axis on the plot and adjust Graphics Options).

Copy plot to StatGallery (see the description to task I).

b) Plot a $\mathcal{N}(0,1)$ distribution and enlarge the line Thickness. Scale x-axis as for Student distribution. Copy it to StatGalery overlying on Student distribution and compare.