Балтийский государственный технический университет «Военмех» им. Д.Ф.Устинова

Кафедра прикладной математики и информатики (И7).

Математическая статистика и случайные процессы.

Лабораторная работа №5.

Проверка статистических гипотез о числовых значениях нормальных распределений в пакетах Statgraphics и Mathcad.

Вариант 12.

Выполнил

Романов Н.

группа И383

Проверил

Гнидин В.В.

Санкт-Петербург

2011

**Задание 1.**

Смоделировать нормальную выборку с параметрами Mx (месяц рождения) и Dx (номер фамилии по списку) объемом 50 элементов. Смоделировать равномерную выборку и засорить ей нормальную. Смоделировать вторую нормальную выборку с параметрами My = Mx + 1.5, Dy = Dx + 3 и объемом 100 элементов. Проверить гипотезы о числовых значениях мат.ожидания и дисперсии.

**Решение в пакете Statgraphics:**

Нормальная выборка:

**Probability Distributions**

Distribution: Normal

Parameters: Mean Std. dev.

Dist. 1 8 3,5

**The StatAdvisor**

*This procedure allows you to analyze any of 24 probability distributions. Currently, the normal distribution has been selected. You can create various plots, compute tail areas and critical values, and generate random numbers from the selected distribution. Up to five sets of parameters can be specified by pressing the alternate mouse button and selecting Analysis Options.*

**Cumulative Distribution**

Distribution: Normal

Lower Tail Area (<)

Variable Dist. 1

0 0,0111354

Probability Density

Variable Dist. 1

0 0,00836293

Upper Tail Area (>)

Variable Dist. 1

0 0,988865

**The StatAdvisor**

*This pane evaluates the cumulative normal distribution. It will calculate the tail areas for up to 5 critical values of the distribution. It will also calculate the probability density or mass function. For example, the output indicates that, for the first distribution specified, the probability of obtaining a value less than 0,0 is 0,0111354. Also, the probability of obtaining a value greater than 0,0 is 0,988865. The height of the probability density function at 0,0 is 0,00836293.*

**Inverse CDF**

Distribution: Normal

CDF Dist. 1

0,01 -0,142233

0,1 3,51456

0,5 8

0,9 12,4854

0,99 16,1422

**The StatAdvisor**

*This pane finds critical values for the normal distribution. You may specify up to 5 five tail areas. The critical value is defined as the largest value for the normal distribution such that the probability of not exceeding that value does not exceed the area specified. For example, the output indicates that, for the first distribution specified, -0,142233 is the largest value such that the probability of not exceeding -0,142233 is less than or equal to 0,01.*

**Random Numbers**

To generate random numbers from the selected distribution,

use the save button on the analysis toolbar.

Random numbers to be generated: 50

**The StatAdvisor**

*This pane allows you to specify the number of observations desired in a random sample from the normal distribution. You set the number of observations by pressing the alternate mouse button and selecting Pane Options. After setting the size, press the Save Results button on the analysis toolbar. This allows you to save random samples from the specified distribution in columns of the current data file. Every time you select Save Results, a new random sample will be generated.*



Равномерная выборка:

**Probability Distributions**

Distribution: Uniform

Parameters: Lower limit Upper limit

Dist. 1 0 0,5

**The StatAdvisor**

*This procedure allows you to analyze any of 24 probability distributions. Currently, the uniform distribution has been selected. You can create various plots, compute tail areas and critical values, and generate random numbers from the selected distribution. Up to five sets of parameters can be specified by pressing the alternate mouse button and selecting Analysis Options.*

**Cumulative Distribution**

Distribution: Uniform

Lower Tail Area (<)

Variable Dist. 1

0 0,0

Probability Density

Variable Dist. 1

0 2,0

Upper Tail Area (>)

Variable Dist. 1

0 1,0

**The StatAdvisor**

*This pane evaluates the cumulative uniform distribution. It will calculate the tail areas for up to 5 critical values of the distribution. It will also calculate the probability density or mass function. For example, the output indicates that, for the first distribution specified, the probability of obtaining a value less than 0,0 is 0,0. Also, the probability of obtaining a value greater than 0,0 is 1,0. The height of the probability density function at 0,0 is 2,0.*

**Inverse CDF**

Distribution: Uniform

CDF Dist. 1 Dist. 2 Dist. 3 Dist. 4 Dist. 5

0,01 0,005

0,1 0,05

0,5 0,25

0,9 0,45

0,99 0,495

**The StatAdvisor**

*This pane finds critical values for the uniform distribution. You may specify up to 5 five tail areas. The critical value is defined as the largest value for the uniform distribution such that the probability of not exceeding that value does not exceed the area specified. For example, the output indicates that, for the first distribution specified, 0,005 is the largest value such that the probability of not exceeding 0,005 is less than or equal to 0,01.*

**Random Numbers**

To generate random numbers from the selected distribution,

use the save button on the analysis toolbar.

Random numbers to be generated: 50

**The StatAdvisor**

*This pane allows you to specify the number of observations desired in a random sample from the uniform distribution. You set the number of observations by pressing the alternate mouse button and selecting Pane Options. After setting the size, press the Save Results button on the analysis toolbar. This allows you to save random samples from the specified distribution in columns of the current data file. Every time you select Save Results, a new random sample will be generated.*



Засоренная выборка:

**Analysis Summary**

Data variable: normun

50 values ranging from -0,744798 to 15,1846

**The StatAdvisor**

*This procedure is designed to summarize a single sample of data. It will calculate various statistics and graphs. Also included in the procedure are confidence intervals and hypothesis tests. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.*

**Summary Statistics for normun**

Count = 50

Average = 8,59768

Median = 9,03327

Mode =

Geometric mean =

Variance = 10,8516

Standard deviation = 3,29418

Standard error = 0,465868

Minimum = -0,744798

Maximum = 15,1846

Range = 15,9294

Lower quartile = 6,62122

Upper quartile = 11,4116

Interquartile range = 4,79037

Skewness = -0,509037

Stnd. skewness = -1,46946

Kurtosis = 0,431256

Stnd. kurtosis = 0,622465

Coeff. of variation = 38,3148%

Sum = 429,884

**The StatAdvisor**

*This table shows summary statistics for normun. It includes measures of central tendency, measures of variability, and measures of shape. Of particular interest here are the standardized skewness and standardized kurtosis, which can be used to determine whether the sample comes from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate any statistical test regarding the standard deviation. In this case, the standardized skewness value is within the range expected for data from a normal distribution. The standardized kurtosis value is within the range expected for data from a normal distribution.*

**Percentiles for normun**

1,0% = -0,744798

5,0% = 2,64614

10,0% = 4,75182

25,0% = 6,62122

50,0% = 9,03327

75,0% = 11,4116

90,0% = 12,5713

95,0% = 13,5899

99,0% = 15,1846

**The StatAdvisor**

*This pane shows sample percentiles for normun. The percentiles are values below which specific percentages of the data are found. You can see the percentiles graphically by selecting Quantile Plot from the list of Graphical Options.*

**Frequency Tabulation for normun**

--------------------------------------------------------------------------------

Lower Upper Relative Cumulative Cum. Rel.

Class Limit Limit Midpoint Frequency Frequency Frequency Frequency

--------------------------------------------------------------------------------

at or below -2,0 0 0,0000 0 0,0000

1 -2,0 0,571429 -0,714286 1 0,0200 1 0,0200

2 0,571429 3,14286 1,85714 2 0,0400 3 0,0600

3 3,14286 5,71429 4,42857 4 0,0800 7 0,1400

4 5,71429 8,28571 7,0 16 0,3200 23 0,4600

5 8,28571 10,8571 9,57143 12 0,2400 35 0,7000

6 10,8571 13,4286 12,1429 12 0,2400 47 0,9400

7 13,4286 16,0 14,7143 3 0,0600 50 1,0000

above 16,0 0 0,0000 50 1,0000

--------------------------------------------------------------------------------

Mean = 8,59768 Standard deviation = 3,29418

**The StatAdvisor**

*This option performs a frequency tabulation by dividing the range of normun into equal width intervals and counting the number of data values in each interval. The frequencies show the number of data values in each interval, while the relative frequencies show the proportions in each interval. You can change the definition of the intervals by pressing the alternate mouse button and selecting Pane Options. You can see the results of the tabulation graphically by selecting Frequency Histogram from the list of Graphical Options.*

**Stem-and-Leaf Display for normun: unit = 1,0 1|2 represents 12,0**

LO|-0,744798

2 0|1

4 0|23

9 0|45555

21 0|666666667777

(13) 0|8888999999999

16 1|00111111

8 1|2222233

1 1|5

**The StatAdvisor**

*This display shows a frequency tabulation for normun. The range of the data has been divided into 8 intervals (called stems), each represented by a row of the table. The stems are labeled using one or more leading digits for the data values falling within that interval. On each row, the individual data values are represented by a digit (called a leaf) to the right of the vertical line. This results in a histogram of the data from which you can recover at least two significant digits for each data value. If there are any points lying far away from most of the others (called outside points), they are placed on separate high and low stems. In this case, there is one outside point. Outside points are illustrated graphically on the box-and-whisker plot, which you can access via the list of Graphical Options. The leftmost column of numbers are depths, which give cumulative counts from the top and bottom of the table, stopping at the row which contains the median.*

**Confidence Intervals for normun**

95,0% confidence interval for mean: 8,59768 +/- 0,936198 [7,66148;9,53388]

95,0% confidence interval for standard deviation: [2,75174;4,10499]

**The StatAdvisor**

*This pane displays 95,0% confidence intervals for the mean and standard deviation of normun. The classical interpretation of these intervals is that, in repeated sampling, these intervals will contain the true mean or standard deviation of the population from which the data come 95,0% of the time. In practical terms, we can state with 95,0% confidence that the true mean normun is somewhere between 7,66148 and 9,53388, while the true standard deviation is somewhere between 2,75174 and 4,10499.*

*Both intervals assume that the population from which the sample comes can be represented by a normal distribution. While the confidence interval for the mean is quite robust and not very sensitive to violations of this assumption, the confidence interval for the standard deviation is quite sensitive. If the data do not come from a normal distribution, the interval for the standard deviation may be incorrect. To check whether the data come from a normal distribution, select Summary Statistics from the list of Tabular Options, or choose Normal Probability Plot from the list of Graphical Options.*

**Hypothesis Tests for normun**

Sample mean = 8,59768

Sample median = 9,03327

t-test

------

Null hypothesis: mean = 8,0

Alternative: not equal

Computed t statistic = 1,28294

P-Value = 0,205548

**Do not reject the null hypothesis for alpha = 0,05.**

sign test

---------

Null hypothesis: median = 8,0

Alternative: not equal

Number of values below hypothesized median: 21

Number of values above hypothesized median: 29

Large sample test statistic = 0,989949 (continuity correction applied)

P-Value = 0,322197

**Do not reject the null hypothesis for alpha = 0,05.**

signed rank test

----------------

Null hypothesis: median = 8,0

Alternative: not equal

**



Проверка гипотезы о равенстве среднеквадратичных отклонений нормальной выборки и засоренной.

**Hypothesis Tests**

----------------

Sample standard deviation = 3,29418

Sample size = 100

95,0% confidence interval for sigma: [2,89231;3,82677]

Null Hypothesis: standard deviation = 3,5

Alternative: not equal

Computed chi-squared statistic = 87,6988

P-Value = 0,43067

Do not reject the null hypothesis for alpha = 0,05.

**The StatAdvisor**

*This analysis shows the results of performing a hypothesis test concerning the standard deviation (sigma) of a normal distribution. The two hypotheses to be tested are:*

*Null hypothesis: sigma = 3,5*

*Alternative hypothesis: sigma <> 3,5*

*Given a sample of 100 observations with a standard deviation of 3,29418, the computed chi-square statistic equals 87,6988. Since the P-value for the test is greater than or equal to 0,05, the null hypothesis cannot be rejected at the 95,0% confidence level. The confidence interval shows that the values of sigma supported by the data fall between 2,89231 and 3,82677.*



Вторая нормальная выборка:

**Analysis Summary**

Data variable: norm1

100 values ranging from -0,322761 to 20,2111

**The StatAdvisor**

*This procedure is designed to summarize a single sample of data. It will calculate various statistics and graphs. Also included in the procedure are confidence intervals and hypothesis tests. Use the Tabular Options and Graphical Options buttons on the analysis toolbar to access these different procedures.*

**Summary Statistics for norm1**

Count = 100

Average = 9,928

Median = 9,94889

Mode =

Geometric mean =

Variance = 17,9275

Standard deviation = 4,23409

Standard error = 0,423409

Minimum = -0,322761

Maximum = 20,2111

Range = 20,5339

Lower quartile = 6,70171

Upper quartile = 12,8795

Interquartile range = 6,17774

Skewness = 0,030019

Stnd. skewness = 0,122552

Kurtosis = -0,24938

Stnd. kurtosis = -0,509046

Coeff. of variation = 42,648%

Sum = 992,8

**The StatAdvisor**

*This table shows summary statistics for norm1. It includes measures of central tendency, measures of variability, and measures of shape. Of particular interest here are the standardized skewness and standardized kurtosis, which can be used to determine whether the sample comes from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate any statistical test regarding the standard deviation. In this case, the standardized skewness value is within the range expected for data from a normal distribution. The standardized kurtosis value is within the range expected for data from a normal distribution.*

**Percentiles for norm1**

1,0% = -0,173475

5,0% = 3,18253

10,0% = 4,4741

25,0% = 6,70171

50,0% = 9,94889

75,0% = 12,8795

90,0% = 15,3824

95,0% = 16,6485

99,0% = 20,153

**The StatAdvisor**

*This pane shows sample percentiles for norm1. The percentiles are values below which specific percentages of the data are found. You can see the percentiles graphically by selecting Quantile Plot from the list of Graphical Options.*

**Frequency Tabulation for norm1**

--------------------------------------------------------------------------------

Lower Upper Relative Cumulative Cum. Rel.

Class Limit Limit Midpoint Frequency Frequency Frequency Frequency

--------------------------------------------------------------------------------

at or below -2,0 0 0,0000 0 0,0000

1 -2,0 1,0 -0,5 2 0,0200 2 0,0200

2 1,0 4,0 2,5 6 0,0600 8 0,0800

3 4,0 7,0 5,5 20 0,2000 28 0,2800

4 7,0 10,0 8,5 23 0,2300 51 0,5100

5 10,0 13,0 11,5 25 0,2500 76 0,7600

6 13,0 16,0 14,5 16 0,1600 92 0,9200

7 16,0 19,0 17,5 6 0,0600 98 0,9800

8 19,0 22,0 20,5 2 0,0200 100 1,0000

above 22,0 0 0,0000 100 1,0000

--------------------------------------------------------------------------------

Mean = 9,928 Standard deviation = 4,23409

**The StatAdvisor**

*This option performs a frequency tabulation by dividing the range of norm1 into equal width intervals and counting the number of data values in each interval. The frequencies show the number of data values in each interval, while the relative frequencies show the proportions in each interval. You can change the definition of the intervals by pressing the alternate mouse button and selecting Pane Options. You can see the results of the tabulation graphically by selecting Frequency Histogram from the list of Graphical Options.*

**Stem-and-Leaf Display for norm1: unit = 1,0 1|2 represents 12,0**

0 0|

6 0|223333

16 0|4444455555

30 0|66666666667777

49 0|8888888889999999999

(16) 1|0000000001111111

35 1|2222222223333333

19 1|444445555

10 1|66666

5 1|8

4 2|00

**The StatAdvisor**

*This display shows a frequency tabulation for norm1. The range of the data has been divided into 11 intervals (called stems), each represented by a row of the table. The stems are labeled using one or more leading digits for the data values falling within that interval. On each row, the individual data values are represented by a digit (called a leaf) to the right of the vertical line. This results in a histogram of the data from which you can recover at least two significant digits for each data value. If there are any points lying far away from most of the others (called outside points), they are placed on separate high and low stems. In this case, there are no outside points. Outside points are illustrated graphically on the box-and-whisker plot, which you can access via the list of Graphical Options. The leftmost column of numbers are depths, which give cumulative counts from the top and bottom of the table, stopping at the row which contains the median.*

**Confidence Intervals for norm1**

------------------------------

95,0% confidence interval for mean: 9,928 +/- 0,840137 [9,08787;10,7681]

95,0% confidence interval for standard deviation: [3,71756;4,91864]

**The StatAdvisor**

*This pane displays 95,0% confidence intervals for the mean and standard deviation of norm1. The classical interpretation of these intervals is that, in repeated sampling, these intervals will contain the true mean or standard deviation of the population from which the data come 95,0% of the time. In practical terms, we can state with 95,0% confidence that the true mean norm1 is somewhere between 9,08787 and 10,7681, while the true standard deviation is somewhere between 3,71756 and 4,91864.*

*Both intervals assume that the population from which the sample comes can be represented by a normal distribution. While the confidence interval for the mean is quite robust and not very sensitive to violations of this assumption, the confidence interval for the standard deviation is quite sensitive. If the data do not come from a normal distribution, the interval for the standard deviation may be incorrect. To check whether the data come from a normal distribution, select Summary Statistics from the list of Tabular Options, or choose Normal Probability Plot from the list of Graphical Options.*









Проверка гипотезы о равенстве мат.ожиданий нормальных выборок:

**Hypothesis Tests**

----------------

Sample means = 8,33233 and 10,5727

Sample standard deviations = 3,29153 and 4,12798

Sample sizes = 50 and 100

95,0% confidence interval for difference between means: -2,24037 +/- 1,32499 [-3,56536;-0,915384]

Null Hypothesis: difference between means = 0,0

Alternative: not equal

Computed t statistic = -3,34136

P-Value = 0,00105553

Reject the null hypothesis for alpha = 0,05.

(Equal variances assumed).

**The StatAdvisor**

*This analysis shows the results of performing a hypothesis test concerning the difference between the means (mu1-mu2) of two samples from normal distributions. The two hypotheses to be tested are:*

*Null hypothesis: mu1-mu2 = 0,0*

*Alternative hypothesis: mu1-mu2 <> 0,0*

*Given one sample of 50 observations with a mean of 8,33233 and a standard deviation of 3,29153 and a second sample of 100 observations with a mean of 10,5727 and a standard deviation of 4,12798, the computed t statistic equals -3,34136. Since the P-value for the test is less than 0,05, the null hypothesis is rejected at the 95,0% confidence level. The confidence interval shows that the values of mu1-mu2 supported by the data fall between -3,56536 and -0,915384.*

*NOTE: in running this test, it has been assumed that the standard deviations of the two samples are equal. You can relax this assumption by pressing the alternate mouse button and selecting Analysis Options.*



Проверка гипотезы о равенстве среднеквадратичных отклонений нормальных выборок:

**Hypothesis Tests**

Sample standard deviations = 3,29153 and 4,12798

Sample sizes = 50 and 100

95,0% confidence interval for ratio of variances: [0,398025;1,05785]

Null Hypothesis: ratio of variances = 1,0

Alternative: not equal

Computed F statistic = 0,6358

P-Value = 0,0805704

Do not reject the null hypothesis for alpha = 0,05.

**The StatAdvisor**

This analysis shows the results of performing a hypothesis test concerning the ratio of the standard deviations (sigma1/sigma2) of two samples from normal distributions. The two hypotheses to be tested are:

Null hypothesis: sigma1/sigma2 = 1,0

Alternative hypothesis: sigma1/sigma2 <> 1,0

Given one sample of 50 observations with a standard deviation of 3,29153 and a second sample of 100 observations with a standard deviation of 4,12798, the computed F statistic equals 0,6358. Since the P-value for the test is greater than or equal to 0,05, the null hypothesis cannot be rejected at the 95,0% confidence level. The confidence interval shows that the values of sigma1/sigma2 supported by the data fall between 0,398025 and 1,05785.

