Introduction to Probability, Statistics and Data Handling	statgraphics 19 centurion
AGH UST ESA LAB 2	**Confidence Intervals Hypotheses testing**

CONFIDENCE INTERVALS

1. Solve the problem.
2. Copy the file `Height_data1.txt` from my webpage:

[open the course webpage]

The file `Height_data1.txt` contains 100 measurements of women's height. Chose randomly more than 20 values.

3. Calculate the sample mean and the sample standard deviation. Assume the population standard deviation is the value obtained from the whole data set $\sigma = 8.34$. Fit the normal distribution.
4. Construct a 90%, 95%, 99% confidence intervals (CI) for the mean for your sample. Which table of distribution you should use?
   From the main menu: ***Describe->Numerical Data->One-Variable Analysis***, highlight ***Col_0x***, and press ►, tick options from ***Tables and Graphs*** window, chose ***Confidence Intervals.***
5. Answer the questions:
   a) Assuming the population mean is $\mu = 161.02$. With the CIs from the group, count how many intervals include the value μ? Divide this number by the total numbers of CIs from the group. Is it close to the Confidence Level (CL) 90%, 95%, 99%?
   b) Suppose we had generated 100 confidence intervals. What do you think would happen to the percent of confidence intervals that contained the population mean?
   c) When we construct a 90% confidence interval, we say that we are 90% confident that the true population mean lies within the confidence interval. Explain what we mean by this phrase.
   d) Some students think that a 90% confidence interval contains 90% of the data. Use the list of data given (the heights of women) and count how many of the data values lie within the confidence interval that you generated based on that data. How many of the 100 data values lie within your confidence interval? What percent is this? Is this percent close to 90%? Does it make sense to count data values that lie in a confidence interval?

HYPOTHESES TESTING

2. Basing on your sample calculate the sample mean. Assuming the significance level $\alpha = 0.05$, can it be concluded that the population mean is 161.02 cm with the alternative hypothesis that population mean is lower or higher (your result) than 161.02 cm?

   In ***One-Variable Analysis/Tabular Options*** choose ***Hypothesis Tests***. Set the alternative hypothesis and $\alpha = 0.05$. Compare the result with the direct calculation.

3. Basing on your sample and one of your colleague's sample, at $\alpha = 0.05$ level, can it be concluded that both of you got the same mean height value?

Use: ***Compare->Two Samples->Two Sample Comparison****,* set columns as *Sample 1, Sample 2.* In ***Tabular Options*** choose ***Analysis Summary-> Summary Statistics-> Comparison of Means*** and ***Comparison of Standard Deviations***.

4. Another group provided the sample with the sample mean $\bar{x} = 171$ cm. Let's assume that the standard deviation is the known population standard deviation. Calculate the p-value. Interpret the result.

5. In one year, 240 of the 560 technical graduates did not solve a certain task in the preliminary exam in mathematics. By contrast, 380 candidates did not solve this task in 1040 of the students of general education. At the level of relevance α=0.05, verify the hypothesis of equal ability to solve this by graduates of both types of schools.

   This is a problem testing equality of structure indicators in two populations – otherwise – the problem of testing proportionality (probability of success) in two binomial distributions. Using CLT for the asymptotic test of hypothesis H_0: $p_1=p_2$ vs H_1: $p_1 \neq p_2$. ***Compare→Hypothesis Tests –> Binomial Proportion***

6. Employers want to know which days of the week employees are absent in a five-day work week. Most employers would like to believe that employees are absent equally during the week. Suppose a random sample of 60 managers were asked on which day of the week they had the highest number of employee absences. The results were distributed as in **Table**. For the population of employees, do the days for the highest number of absences occur with equal frequencies during a five-day work week? Test at a 5% significance level.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Number of Absences	15	12	9	9	15