Homework 5

Here is the list of problems constituting the fifth homework assignment. First, please, try to find your own solutions and after this effort consult these solutions with the ones presented during tutorials and, finaly, check the solutions that will be posted on the course webpage. Remember that problems can have several different but correct ways of solving them.

Multiple choice questions

1. Statistical inference

   A is about trying to guess how data would look if collected anew;

   B is about producing data on a computer that would look similar to real data;

   C is about inferring from the model how data should look like;

   D is about drawing proper graphs that illustrate raw data;

   E is about inferring from the data about parameters that describe an assumed model for the data.

2. Estimation

   A is a form of statistical inference in which we use computer to evaluate a rough approximation of a certain value that we can not compute exactly;

   B uses a model to predict values of data;

   C uses data to propose reasonable values for parameters of a model;

   D is a procedure of decision making about how to collect data;

   E is another word for testing statistical hypotheses.

3. Confidence intervals

   A tell us what are chances for the sample mean that is computed from collected data to fall into these intervals;

   B always include the true value of a parameter of interest;

   C express our confidence that the proposed model for data is correct;

   D are longer when more data are collected;

   E are intervals that are computed from the data and have a good chance to include the true value of a parameter of interest.

4. When testing statistical hypotheses

   A we find methods that allow us to find out the definite truth about the hypotheses at hand;

   B we never make Type I Errors;

   C we tend to reject null hypothesis more often than not because rejecting null hypothesis is the desireable conclusion of testing;

D the focus is on not to falsely reject the null hypothesis;

E when we reject null hypothesis we know that we are right.

5. Which of the following is correct?

   A increasing the level of confidence associated with a confidence interval gives a narrower interval;

   B increasing the sample size associated with a 95% confidence interval increases the confidence level of the interval;

   C the closer to 50% is the percentage of failures, the narrower is a confidence interval for the percentage calculated from data sample from the process;

   D 95% of all confidence intervals for a process mean are within 2 standard errors of X-bar;

   E an alternative way to implement a significance test is to calculate the corresponding confidence interval and check whether the null hypothesis value is included or not.

6. The difference between $\bar{X}$ and μ is

   A a test of the null hypothesis;

   B that one is a sample statistic and the other is a model parameter;

   C that one is a process characteristic and the other is a population parameter;

   D proportional to the standard error;

   E the standard deviation of the sampling distribution.

Problems

1. In the texbook, pages 146-147 (discussed also in the lecture), we have seen that for the clips data after arrival of a new roll of steel material, the clip gaps have noticeable been smaller. In the lecture we have computed a 95% confidence interval for the mean of clip gaps before arrival of this new batch of material. The computed interval was $[72.2, 75.4]$ and was based on the sample mean of 73.8 and standard deviation 7.3. After the arrival of the new batch, the estimate of the mean of clip gaps is 66.75. Compute the new confidence interval assuming that standard deviation has not changed, i.e. that it is still 7.3.

2. The marketing director of a bank would like to make a 'special offer' ot personal customers whose accounts turn over more than 100,000EUR in a year. He guesses that such customers account for no more than 10% of all the banks customers, and costs the 'special offer' accordingly. The financial controller believes that there are over 20% of such customers and suggests that the cost will be too high. To resolve the issue, the counts department is asked to estimate the percentage of large customers. What sample size is needed to find the interval estimate with the halflength of (i) 1%; (ii) 2%, with (a) 90% confidence; (b) 95% confidence.

3. Following analysis of the 52 cash variances observed over the year, it was concluded that three of the values were exceptional and that the remaining values represented a process in statistical control, with standard deviation 4.15; see page 124, Chapter 4. The estimated mean value, based on the in-control variances, was 0.6. Based on these data, test the null hypothesis that the cash variance process has mean zero. Make a report of the test and its result referring to

   - the null hypothesis;
   - proposed significance level;
   - the standardized test statistic;
   - p-value (observed significance level);
   - the conclusion.

4. A company with a large number of debtors claims (for the purpose of negotiating a bank loan) that the average amount owed is at least 100EUR. An evaluator acting on behalf of the bank looked at a sample of 250 of the company's outstanding accounts and found an average amount owed of 95.53EUR. She found the standard deviation to be 24EUR. Formally test the statistical significance of the claim made by the company. Indicate what you take to be the *null hypothesis*, the *significance level*, the *test statistics*, the *critical value*, and explain the basis for your test in terms of the *sampling distribution* of the test statistic. Calculate the p-value (the observed significance level) for the test and relate its value to the formal conclusion of the test.

5. The catering manager in a hotel suspects that the weight of loaves of bread delivered daily by a bakery is consistently below the nominal weight of 800g. To test this, 10 loaves chosen at random from a day's deliveries are weighed. The mean and standard deviation of the ten weights are 792g and 25g, respectively.

   - Carry out a formal significance test.
   - List the steps involved in this test, from null hypothesis specification to conclusion.
   - Calculate a 95% confidence interval for the average weight of loaves produced by the baker.
   - Comment on the correspondence between the interval, as calculated, and the result of the test.