Computational Statistics

Autumn 2017

Problem Set 1 (Confidence Intervals and Testing of Hypotheses)

Question 1.

Generate a sample of size $n = 10$ from the exponential distribution with mean 4. Calculate the mean of the generated sample. Repeat this process 100 times and in each case record the sample mean. Plot the histogram of these sample means. Does this plot resemble to normal distribution? If the size of the sample is changed from $n = 10$ to $n = 150$, what is your observation. Can you think of a theoretical result that supports this observation.

Question 2.

Generate a sample of size $n = 50$ from a standard normal distribution. Calculate the 95% confidence interval for the mean. Does the mean of the sample lie in this confidence interval? Repeat the previous steps 100 times and see for yourself how many times the true mean lies in the confidence interval. Does your experiment agree with the concept of confidence interval?

Question 3.

It is assumed, when constructing the t-test, that the measurements are normally distributed. In this exercise we examine the robustness of the test with respect to divergence from the assumption. You are required to compute the significance level of a two-sided t-test of $H_0 : \mu = 4$ versus $H_1 : \mu \neq 4$. Assume there are n = 20 observations and use a t-test with a nominal 5% significance level. Consider the case where X ~ Exponential (1/4).

Question 4.

Assume that you are interested in testing $H_0 : \mu = 20$ versus $H_1 : \mu \neq 20$ with a significance level of 5% using the t-test. Let the sample average, of a sample of size n = 55, be equal to $\bar{x} = 22.7$ and the sample standard deviation be equal to s = 5.4.

1. Do you reject the null hypothesis?
2. Use the same information. Only now you are interested in a significance level of 1%. Do you reject the null hypothesis?
3. Use the information the presentation of the exercise. But now you are interested in testing $H_0 : \mu = 20$ versus $H_1 : \mu \neq 20$ (with a significance level of 5%). Do you reject the null hypothesis?

Question 5 .

The sample average in one sub-sample is $\bar{x}_a = 124.3$ and the sample standard deviation is $s_a = 13.4$. The sample average in the second subsample is $\bar{x}_b = 80.5$ and the sample standard deviation is $s_b = 16.7$. The size of the first sub-sample is $n_a = 15$ and this is also the size of the second sub-sample. We are interested in the estimation of the ratio of variances $Var(X_a)/Var(X_b)$.

1. Compute the estimate of parameter of interest.
2. Construct a confidence interval, with a confidence level of 95%, to the value of the parameter of interest.
3. It is discovered that the size of each of the sub-samples is actually equal to 150, and no to 15 (but the values of the other quantities are unchanged). What is the corrected estimate? What is the corrected confidence interval?

Question 6.

Twenty observations are used in order to construct a confidence interval for the expectation. In one case, the construction is based on the Normal approximation of the sample average and in the other case it is constructed under the assumption that the observations are normally distributed. Assume that in reality the measurement is distributed Exponential (1/4).

1. Compute, via simulation, the actual confidence level for the first case of a confidence interval with a nominal confidence level of 95%.
2. Compute, via simulation, the actual confidence level for the second case of a confidence interval with a nominal confidence level of 95%.
3. Which of the two approaches would you prefer?

Question 7.

A certain amount of natural gas is produced with each barrel of crude oil. This gas escapes from the oil near the top of the well pipe. In an attempt to estimate the amount of natural gas available from wells in Kuwait these data are obtained on X, the number of cubic feet of gas obtained per barrel of crude oil :

290 610 790 670 770 420 600 350 800 920 410 810 620 560 550

610 510 390 480 630 470 380 550 570 730 680 530 650 1000 720

Find a 99% confidence interval for the average volume of natural gas produced per barrel of crude oil by wells in Kuwait.

Question 8.

A study of report writing by engineers is conducted. A scale that measures the intelligibility of engineers' English is devised. This scale called an "index of confusion", is devised so that low scores indicate high readability. These data are obtained on articles randomly selected from engineering journals and from unpublished reports written in 1979 :

Journals	Unpublished Reports
1.79 1.75 1.67 1.65	2.39 2.51 2.86 2.14
1.87 1.74 1.94	2.56 2.29 2.49
1.62 2.06 1.33	2.36 2.58 2.33
1.96 1.69 1.70	2.62 2.41 1.94

Find a 90% confidence interval on $\mu_1 - \mu_2$.