ARTIFICIAL INTELLIGENCE

Lab4. *Estimation*

CAUTION! Most of the data which appear in the tasks below is available in *data_est.csv*

1. To consider a new process of synthetic diamonds production, the weight (in carat) of 12 synthetic diamonds has been examined. The following results are obtained:

   0.46, 0.61, 0.52, 0.48, 0.57, 0.54, 0.47, 0.63, 0.51, 0.49, 0.58, 0.55

   (a) Define population, sample and random variable, whose realizations are measurements.

   (b) Compute the estimate of the mean weight of all synthetic diamonds. 0.534167

   (c) Estimate with 95% confidence the mean weight of all synthetic diamonds (construct your own function and compare the result with the output of respective function available in R). (0.498; 0.570)

   (d) Increase the confidence level and compare the lengths of intervals.

   (e) Compute the estimates of the variance and standard deviation of the weight of all synthetic diamonds.

   (f) Estimate with 95% confidence the standard deviation of the weight of all synthetic diamonds. (0.0393; 0.0943)

   In points (c), (d) and (f) assume normality of the distribution of the weight of synthetic diamonds.

2. The Environmental Protection Agency (EPA) is concerned about the amounts of PCB, a toxic chemical, in the milk of nursing mothers. In a sample of 20 women, the amounts (in parts per million) of PCB were as follows: 16, 0, 0, 2, 3, 6, 8, 2, 5, 0, 12, 10, 5, 7, 2, 3, 8, 17, 9, 1. Assume normality of respective random variable.

   (a) Define population, sample and measurement.

   (b) Compute an estimate of the average amount of PCB in the milk of nursing mothers. 5.8

   (c) Compute an estimate of the variance and standard deviation of the amount of PCB in the milk of nursing mothers. 25.85; 5.08

   (d) Estimate with 95% confidence the average amount of PCB in the milk of nursing mothers. Interpret the result. (3.42; 8.18)

   (e) Estimate with 95% confidence the variance and standard deviation of the amount of PCB in the milk of nursing mothers. Interpret the results. (14.95; 55.16); (3.86; 7.43)

3. To estimate the average nicotine content of a newly marketed cigarette, 15 packs of these cigarettes were randomly chosen, and their nicotine contents are determined as

   1.87, 2.28, 1.77, 2.13, 1.43, 1.64, 2.38, 1.39, 1.94, 2.68, 1.95, 0.86, 1.98, 1.69, 1.15.

   Assume that it is known from past experience that the distribution of nicotine content is normal with standard deviation 0.7 milligrams.

   (a) Estimate with 95% confidence the mean content of nicotine in cigarettes (construct your own function and compare the result with the output of respective function available in R). Interpret the results. (1.45; 2.17)

   (b) How large sample is needed for the length of the 95% confidence interval to be less than or equal to 0.3 milligrams? 84

   (c) Compute the estimate of standard deviation and compare the result with population standard deviation, which is known from past experience.

4. Suppose that if a signal having intensity μ originates at location A, then the intensity recorded at location B is normally distributed with mean μ and standard deviation 3. That is, due to “noise,” the intensity recorded differs from the actual intensity of the signal by an amount that is normal with mean 0 and standard deviation 3. To reduce the error, the same signal is independently recorded 10 times. If the successive recorded values are: 17, 21, 20, 18, 19, 22, 20, 21, 16, 19, determine the point estimate of μ (the actual intensity) and estimate it with 95% of confidence. Interpret the result. 19.3; (17.44; 21.16)

5. To determine the average time span of a phone call made during midday, the telephone company has randomly selected a sample of 1200 such calls. The sample mean of these calls is 4.7 minutes, and the sample standard deviation is 2.2 minutes. Estimate with 95% confidence the mean length of all such calls and the standard deviation. Interpret the results. (4.57; 4.83); (2.11; 2.3)

6. A random sample of 36 General Electric transistors resulted in the following lifetimes:

   1252.20, 1368.87, 1199.63, 1330.89, 1204.93, 1149.69, 1048.62, 1048.40, 1370.70, 1127.37, 1202.72, 1081.12, 1118.75, 1108.68, 1129.23, 1145.34, 1337.95, 1149.66, 1219.52, 1343.69, 1202.62, 977.81, 1321.36, 1192.10, 1261.03, 1342.02, 1009.17, 1208.48, 1136.12, 1191.59, 1258.49, 1122.9, 1237.43, 1309.87, 1179.85, 1201.03

   Estimate with 95% confidence mean lifetime of all General Electric transistors. Interpret the result.

   (1163.69; 1230.19)

7. An engineer wants to fix the sample size required to obtain a given precision in the estimation of the mean setting time of the new cement mixture. From previous experience it is known that the setting time of cement mixture is a normally distributed random variable with variance 25 (h^2). What should be the size of a sample to get 95% confidence, that the estimation error does not exceed 1? 97

8. From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 90% certain that our estimate of the mean weight of a salmon is correct to within ±0.1 pounds, how large a sample is needed? What if we want to be 99% certain? 25; 60

9. Suppose that an estimate is made of the proportion of operating time spent on unscheduled maintenance for computers of a particular brand and model. A random sample of 100 terminal log entries indicated the computers were down 3 times. Estimate with 95% confidence the true proportion of unscheduled downtime. Write your own function to get the estimates and compare the result with the outputs of `binom.test` and `prop.test` (with continuity correction) functions in R. Interpret the result.

   (0; 6.35%), (0.62%; 8.52%), (0.77%; 9.16%)

10. The controls in a brewery need adjustment whenever the proportion p of underfilled cans is 0.015 or greater. There is no way of knowing the true proportion, however. Periodically a sample of 100 cans is selected and the contents are measured. For one sample, 4 underfilled cans were found. Estimate with 95% confidence the true proportion of underfilled cans. Interpret the result.

11. An industrial engineer's assistant made 120 random observations of the upholstery-installer team in an automobile assembly plant. During 24 of the observations the workers were arranging materials beside their work station. Estimate with 90% confidence the true proportion of the installers arranging materials besides work stations. Interpret the result.

12. A random sample of 1000 construction workers revealed that 122 are presently unemployed. Estimate the proportion of all construction workers who are unemployed. Estimate with 90% confidence the true proportion of unemployed workers. Interpret the result.

13. It is interesting the proportion of people who have sight problems in a particular group. How many people should be examined to obtain an estimation error ±0.05 on confidence level 0.98, if:
    (a) nothing is known about proportion p; 541
    (b) it is known from previous experience, that p is 0.3? 455