Professor *Ana Cristina Costa* (ccosta@novaims.unl.pt)

HOMEWORK

LU3: SAMPLING DISTRIBUTIONS

***Suggestion**: Use the form to select the appropriate formula, and solve the exercises using statistical tables. Then, use Excel to find the exact solutions provided here.*

I) In the following multiple-choice questions, choose only one option.

1. The Central Limit Theorem refers to

a) the distribution of the sample mean of X when *n* is small.

b) the distribution of X when *n* is large.

c) the distribution of the sample mean of X when *n* is large

d) the median of X when *n* is large

2. Complete the following statement: As the sample size ________ the variability of the sampling distribution of the sample mean ________.

a) decreases, decreases

b) increases, remains the same

c) decreases, remains the same

d) increases, decreases

3. A bottling company uses a filling machine to fill plastic bottles with a popular cola. The bottles are supposed to contain 300 millilitres. In fact, the contents vary according to a normal distribution with mean 298 ml and standard deviation 3 ml. Which is the distribution used to calculate the probability that the mean contents of the bottles in a nine-pack is less than 295 ml?

a) Normal distribution with mean 298 ml and standard deviation 3 ml.

b) Normal distribution with mean 298 ml and standard deviation 1 ml.

c) Approximately normal distribution with mean 298 ml and standard deviation 1 ml.

d) Student's t-distribution with 8 degrees of freedom.

4. One 16-ounce bottle of an energy drink has an average of 500 mg of caffeine with a standard deviation of 25 mg. In a carton containing 30 bottles, what is the standard deviation of the average amount of caffeine?

a) 0.83

b) 1.20

c) 4.56

d) 25.0

5. One 16-ounce bottle of an energy drink has an average of 400 mg of caffeine with a standard deviation of 20 mg. What is the probability that the average caffeine in a sample of 25 bottles is no more than 395 milligrams?

a) 0.05

b) 0.11

c) 0.16

d) 0.22

6. The duration of the Lisbon's marathon of October 2, 2016, has a Normal distribution with mean 172 minutes and unknown standard deviation. A random sample of 16 individuals was selected and it was observed a sample variance of 9. The probability that the sample mean of the marathon duration to be greater than 174.21 minutes is approximately:

a) 0.99

b) 0.995

c) 0.005

d) 0.01

II) Solve the following problems.

1. In a given year, the average selling price of new apartments in a city was 115.000,00€ with a standard deviation of 25.000,00€. Suppose that it was selected a sample of 100 new apartments for sale.

a) Determine the probability that the sample mean selling price had been more than 110.000,00€.

b) Determine the probability that the sample mean selling price has been situated between 113.000,00€ and 117.000,00€.

c) Determine the probability that the sample mean selling price has been situated between 114.000,00€ and 116.000,00€.

2. Suppose that the diameter of a certain type of pieces is characterized by a normal distribution with standard deviation equal to 4 cm.

a) If a sample of 9 elements was obtained from that population, what is the probability that the sample mean does not differ by more than 2 cm from the population's mean value?

Hint: *Compute* $P(-2 < \bar{X} - \mu < 2)$.

b) Suppose that the statistical control of quality required that the sample mean did not differ by more than 1 cm from the population's mean value, with probability of 0.9. How many pieces should be included in the sample, in order to ensure this level of robustness?

Hint: *Using the statistical tables, find* n *so that* $P(-1 < \bar{X} - \mu < 1) = 0.9$*, and then round the obtained value up to the next integer.*

3. An automobile manufacturer argues that the new model that will be launched next month spends on average 9.7 litres per 100 km, in urban areas, and has unknown standard deviation. Through a sampling scheme, it was estimated that this standard deviation as being equal to 1 litre. Assuming that the consumption follows a normal distribution, what is the probability that, in a random sample of 20 cars, the mean consumption sample to be greater than 10 litres? And less than 8.9 litres?

4. The outcome of a chemical process should be 500 g/ml (supposed population expected value). The outcome was measured in 25 batches, and the obtained standard deviation was 40 g/ml. Assuming that the outcome of the chemical process follows a normal distribution, what is the probability of finding a sample mean greater than 518 g/ml?

5. A manufacturer makes a consignment of 1000 lots of 100 light bulbs. If 5% of the bulbs are usually faulty, in how many lots can be expected to exist:

a) Less than 90 good bulbs?

Hint: *The probability of the number of good bulbs in a sample of n = 100 being less than 90 is equal to the probability of the proportion of good bulbs in a sample being less than 90/100, thus compute* $P(\hat{p} < 0.9)$*. In 1000 lots, there will be* $1000 \times P(\hat{p} < 0.9)$ *lots with less than 90 good bulbs.*

b) 98 or more good bulbs?

6. It is known that 70% of the fines are paid within the time limit in a police station.

a) What is the probability that, in a sample of 35 fines, at least 65% has been paid within the time limit?

b) To increase the probability of the preceding paragraph to 80%, what should be the sample size? (assume a sample of large size so that the CLT can be applied)

Hint: *Using the statistical tables, find* n *so that* $P(\hat{p} > 0.65) = 0.80$.

III) OPTIONAL: Sampling distributions and computer simulation in Excel

Adapted from: Weiers, R. M. (2008). *Introduction to Business Statistics*. Sixth edition, Thomson South-Western.

The following procedures simulate 200 samples from a Poisson distribution with a mean of 4. In the first simulation (A), the size of each sample is n=10. In the second simulation (B), the size of each sample is n=50. Afterwards, the means of the 200 samples are computed and plotted to illustrate the distribution of the sample mean when the sample sizes are small (n=10) and large (n=50). Activate the **Analysis ToolPak** add-in.

A. Simulate 200 samples of size n=10 from a Poisson distribution with a mean of 4.

1. Start with a blank worksheet.
2. Click **Data Analysis**, and then select **Random Number Generation**.
3. Enter **10** into the **Number of Variables** box, which specifies n=10. Enter **200** into the **Number of Random Numbers** box. Within the **Distribution** box, select **Poisson**. Enter **4** into the **Lambda** box. Select **Output Range** and enter **A2** into the dialog box. Click **OK**.

   The 200 simulated samples will be located in **A2:J201**, with each row representing one of the 200 samples.
4. Click on cell **K2** and enter the formula **=AVERAGE(A2:J2)**. Click at the lower right corner of cell **K2** and drag downward to fill the remaining cells in column **K**.

   These values are the 200 sample means. Click on cell **K1** and enter the title for the **K** column as **MEANS**.
5. Determine the mean of the sample means by clicking on cell **L2** and entering the formula **=AVERAGE(K2:K201)**. Click on cell **L1** and enter the title for the **L** column as **Mean of MEANS**. Observe how close the obtained value is of the Poisson's mean value of 4.
6. Determine the standard deviation of the sample means by clicking on cell **M2** and entering the formula **=STDEV.S(K2:K201)**. Click on cell **M1** and enter the title for the **M** column as **Standard deviation of MEANS**.
7. Generate the histogram of the column **MEANS**. The result should be similar to Figure 1.

B. The procedure for generating 200 samples of size n=50, from a Poisson distribution with a mean of 4, will be similar to the steps shown above.

   a. Observe how close the mean of the sample means is of the Poisson's mean value of 4.
   b. Compare the range and spread of the sample means when n=10 and n=50.
   c. The resulting histogram should be identical to the one shown in Figure 2.

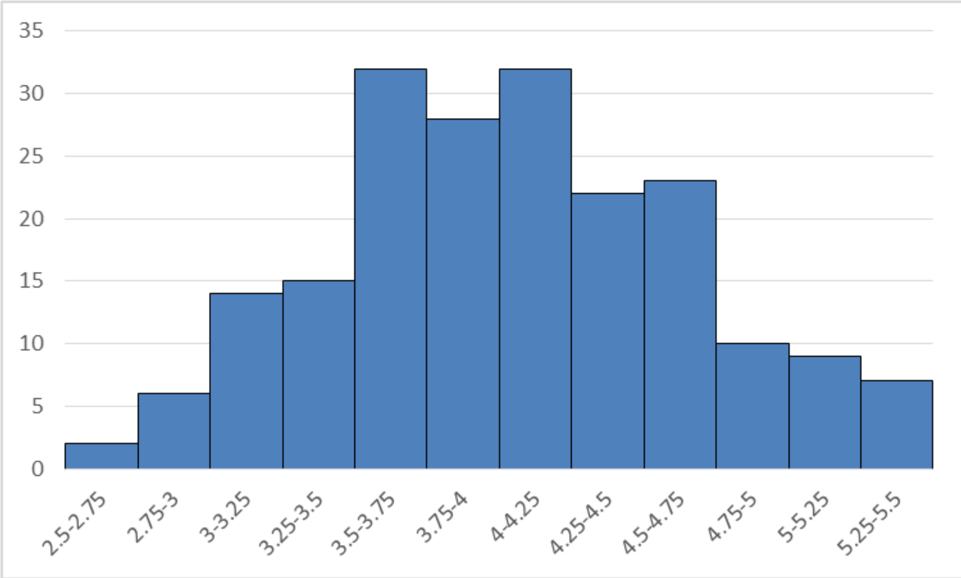

Figure 1: Histogram of the sample means from 200 samples of size n=10 simulated from a Poisson distribution with a mean of 4.

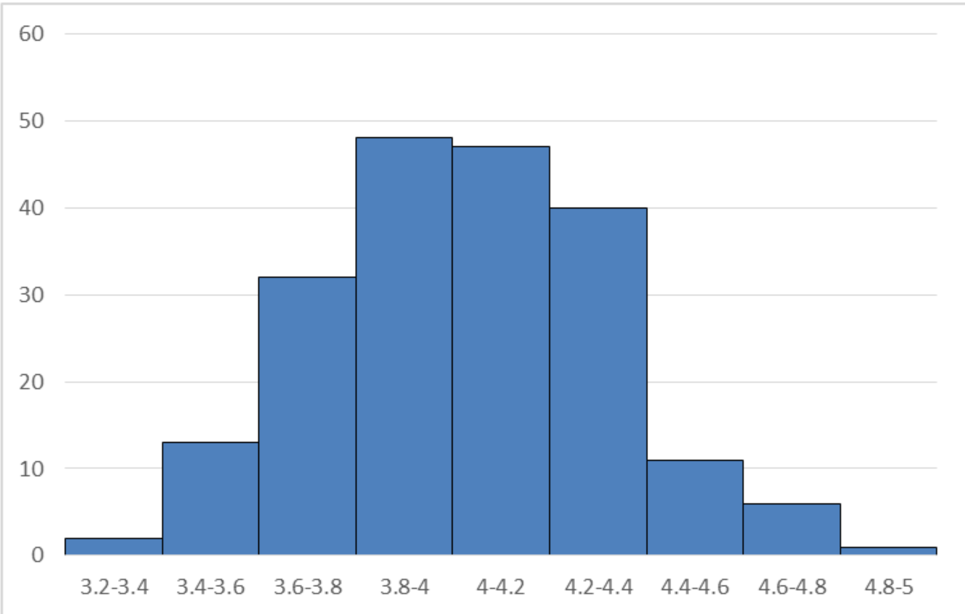

Figure 2: Histogram of the sample means from 200 samples of size n=50 simulated from a Poisson distribution with a mean of 4.

SOLUTIONS

***Note**: The exercises have been solved without rounding the values of intermediate calculations.*

Group I) Multiple choice

1. c
2. d
3. b
4. c
5. b
6. c

Group II) Problems

Question 1)

a) 0.9772

b) 0.5762

c) 0.3108

Question 2)

a) 0.8664

b) 44

Question 3) 0.1; 0.001

Question 4) 0.0169

Question 5)

a) 11

b) 84

Question 6)

a) 0.7422

b) 60