

## LAB 3 ASSIGNMENT

### SAMPLING DISTRIBUTIONS, CENTRAL LIMIT THEOREM

In this lab assignment, you will explore important properties of sampling distributions in the context of a quality control process and computer simulation. In particular, you will use the *Random Number Generation* tool to demonstrate the validity of the Central Limit Theorem. You will also have an opportunity to review some features of the Poisson distribution.

As in the previous lab assignment, you will be given a choice to use either the built-in statistical functions in Excel for normal and Poisson distributions or the instructional templates (interactive worksheets) that are included in the associated data file. The Excel functions for Poisson and normal distributions were discussed in the *Lab 2 Instructions*.

#### Flaws in Plastic Panels

A company makes plastic panels used in automobiles. In order to minimize the number of defective panels, a strict quality control process is being used in the company. Random samples of panels in the manufacturing process are taken every hour and the data provided by quality control inspectors is instantly analyzed against the target parameters. A quality inspector inspects panels to check for any slight flaws. The number of these flaws in a panel is often modelled by a Poisson distribution with a parameter  $\lambda$ . We will use Excel to simulate the outcomes of the quality control process.

The workbook *lab3.xls* available at <http://www.stat.ualberta.ca/statslabs/index.htm> (click *Stat 235* link, and *Data* for *Lab 3*) contains the templates and the data to be used in the lab assignment. The workbook consists of three worksheets: *Poisson Probabilities*, *Normal Probabilities*, and *Simulation*.

The worksheet *Poisson Probabilities* enables you to calculate Poisson probabilities and cumulative Poisson probabilities. The worksheet *Normal Probabilities* allows you to calculate normal probabilities provided the distribution parameters are entered. In order to eliminate the possibility of mistakenly deleting charts and formulas in the worksheets, some cells and interactive charts are protected and, therefore, read-only.

1. Open the worksheet *Poisson Probabilities*. The worksheet contains a graph of the Poisson probability distribution function. The top of the worksheet contains the value of  $\lambda$ . Any change to the distribution parameter is immediately reflected in the probability distribution function. In this part, you will see how various choices of  $\lambda$  affect the shape of the Poisson distribution.

Enter the value of  $\lambda$  as 0.5, then 1, 2, 3, 4, 5, and finally 6. After each entry, carefully examine the shape of the corresponding density curve. You are not supposed to print the probability distribution functions.

Describe briefly the change in the appearance of the probability distribution function as  $\lambda$  increases from 0.5 to 6. What does the increase mean to the panel manufacturing process?

2. In this question, you can use either the built-in Excel Poisson function or the worksheet *Poisson Probabilities*. Suppose that the number of flaws in a panel has a Poisson distribution with parameter  $\lambda = 0.5$ . This implies that the expected number of flaws per panel is 0.5. Answer the following questions:
  - (a) What is the probability that there are no flaws in a randomly selected panel? In other words, what is the percentage of panels in perfect condition?
  - (b) Panels with two or more flaws are considered defective and scrapped by the company. Use the template to determine the percentage of panels that need to be scrapped.
  - (c) What is the probability that a random sample of ten panels does not contain any defective panel (at most one flaw)?

3. In order to check on the manufacturing process, random samples of several panels are taken repeatedly. According to the Central Limit Theorem, if a random sample of size  $n$  is drawn from a population with mean  $\mu$  and variance  $\sigma^2$ , the sample mean approximately follows a normal distribution with mean  $\mu$  and variance  $\sigma^2/n$  if the sample size is sufficiently large.

Assume  $\lambda = 0.5$  and use either the Excel NORMDIST function or the worksheet *Normal Probabilities* to answer the following questions:

- (a) Can we safely assume that the sampling distribution of the average number of flaws in a random sample of 10 panels follows approximately a normal distribution? Why or why not? Assuming it does apply, what is the probability that the average number of flaws in a random sample of 10 panels does not exceed 0.30?
- (b) Can we safely assume that the sampling distribution of the average number of flaws in a random sample of 30 panels follows approximately a normal distribution? Why or why not? Assuming it does apply, what is the probability that the average number of flaws in a random sample of 30 panels does not exceed 0.30?

In Questions 4 and 5, you must use Excel in a Windows environment to obtain data (different data may be produced by the *Random Number Generator* in Excel on MAC OS). Excel 2010 or higher version should be used in this part (earlier versions may produce different sequence of random numbers for the same seed).

Now suppose that the quality inspector is going to take a random sample of several panels to estimate the average number of flaws. How likely will his estimate be close to the true value? Which estimate is more likely to be better: one based on a sample of 10 panels or one based on a sample of 30 panels? We will answer the above questions by simulation.

4. Open the worksheet *Simulation*. The worksheet allows you to simulate the outcomes of the quality control process for 60 random samples of panels. Use the *Random Number Generation* feature (Poisson, seed 1000) to generate 60 samples of size  $n = 10$  from the Poisson distribution with the parameter  $\lambda = 0.5$ . This corresponds to selecting 60 samples, each consisting of 10 panels from the production run. The data will be entered in the form of 60 columns, each consisting of 10 rows into the range B9:BI18. In other words, the range contains the outcome of quality testing for 600 panels.

Once the data are entered, the values of the variables AVERAGE and COUNT are automatically displayed in rows 40 and 42, respectively.

- (a) Use the COUNTIF function to determine the percentage of perfect panels (no flaws) among the 600 panels. Compare the value with the probability obtained in part (a) of Question 2. Should the values be identical? Explain briefly. The COUNTIF function was discussed in the *Lab 2 Instructions*.
- (b) The variable COUNT in row 42 counts the number of defective panels in each sample. Use the values to determine the percentage of samples of ten panels containing no defective panels (at most one flaw)? Compare the value with the probability obtained in part (c) of Question 2. Should the values be identical?
- (c) The variable AVERAGE in row 40 shows the average number of flaws for each sample. Obtain and print a histogram of the average number of flaws using the following bins: 0.0, 0.1, ..., 1.2. The format of your histogram should be the same as the format of the sample histogram in the *Lab 1 Instructions* (labels, title, no gaps between bars). Describe the shape of the histogram.
- (d) The worksheet *Simulation* also displays *Summary Statistics* for the variable AVERAGE. Use the feature to obtain the mean and standard deviation of the average number of flaws for the 60 samples. Compare them with the values predicted by theory. Should the values be identical? Explain briefly.

5. Now we repeat parts (c) and (d) of Question 4 with  $n = 30$ . First, clear the range B9:BI18 in the worksheet *Simulation*. Use the *Random Number Generation* feature (Poisson, seed 2000) to generate 60 samples of size  $n = 30$  from the Poisson distribution with the parameter  $\lambda = 0.5$ . The data will be entered into the range B9:BI38 in the form of 60 columns, each consisting of 30 rows.
- (a) Obtain and print a histogram of the average number of flaws using the following bins: 0.3, 0.4, ..., 0.8. Describe the shape of the histogram, comparing to the histogram obtained in part (c) of Question 4.
- (b) Obtain the mean and standard deviation of the average number of flaws for the 60 samples. Compare them with the values predicted by theory and relate to the similar comparison in part (d) of Question 4. What do you conclude? State your findings briefly.

### LAB 3 ASSIGNMENT MARKING SCHEMA

Proper Header and appearance: 10 points

#### Question 1 (5)

Change in the appearance of the probability distribution function: 3 points  
 Implication to panel manufacturing process: 2 points

#### Question 2 (9)

- (a) Percentage of panels in perfect condition: 3 points  
 (b) Percentage of panels that need to be scrapped: 3 points  
 (c) Percentage of samples of ten panels with no defective panel: 3 points

#### Question 3 (10)

- (a) Application of Central Limit Theorem: 2 points  
     Probability that the average does not exceed 0.30 ( $n = 10$ ): 3 points  
 (b) Application of Central Limit Theorem: 2 points  
     Probability that the average does not exceed 0.30 ( $n = 30$ ): 3 points

#### Question 4 (24)

- (a) Percentage of panels in perfect condition: 2 points  
     Comparison with the probability: 2 points  
 (b) Percentage of samples of ten panels with no defective panel: 2 points  
     Comparison with the probability: 2 points  
 (c) Correctly formatted histogram of the average number of flaws: 6 points  
     Shape of the histogram: 3 points  
 (d) Mean of the averages: 2 points  
     Standard deviation of the averages: 2 points  
     Comparison with values predicted by theory: 3 points

#### Question 5 (16)

- (a) Correctly formatted histogram of the average number of flaws: 6 points  
     Shape of the histogram: 3 points  
 (b) Mean of the averages: 2 points  
     Standard deviation of the averages: 2 points  
     Comparison with values predicted by theory: 3 points

**TOTAL = 74 POINTS**