LAB 3 ASSIGNMENT

NORMAL & SAMPLING DISTRIBUTIONS, CENTRAL LIMIT THEOREM

In this lab assignment, you will explore important properties of the normal distribution and the sampling distribution of a sample mean in the context of quality control and computer simulation. In particular, you will use the *Random Number Generation* tool to demonstrate the validity of the Central Limit Theorem.

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

Engines Getting Stats & Spurious

Experts tend to agree the optimal engine temperature is around 190°F to 225°F (or 88°C to 107°C). If an engine overheats for too long, or too often, it could potentially break down, requiring repair or replacement. Failure to reach the minimum temperature results in decreased engine power and efficiency. Thus, a manufacturing company needs to build engines that maintain this acceptable range. Typically, an engine would be tested by recording the temperature of the engine after being given sufficient time to warm up. Engines with temperatures recorded outside the temperature range above are considered “unacceptable”.

The necessary Excel file is available in the Data link located in the Lab 3 tab display in the Labs section on eClass. The data are not to be printed in your submission. The Excel file consists of five worksheets: *Normal Density*, *Normal Probabilities*, *Normal Probability Plot*, *Data*, and *Simulation*. The first three are instructional templates; the last two contain data related to the above situation. To eliminate the possibility of mistakenly deleting charts and formulas in the first three worksheets, some cells and interactive charts are protected and, therefore, read-only. **Unless otherwise stated, numerical answers should be to four decimal places.**

1. Open the worksheet *Normal Density*. This worksheet contains a graph of the density function for temperature. The graph shares your screen with the values of μ and σ. This will enable you to change those values and immediately see the change in the density function without scrolling your screen or changing worksheets. Use the interactive graph to answer the following questions.

   (a) Assume that σ = 5.5°F and that the mean temperature changes. Enter the value of μ as 205, then 210, and finally 215. After each entry, carefully examine the shape of the corresponding density curve. You are not supposed to print the density curves. Describe briefly how the shape of the corresponding density curve changes. How does the change in the value of μ affect the density curve of temperature? Does the proportion of manufactured engines that are unacceptable (recorded temperature values below 190°F or above 225°F) increase or decrease? Explain briefly. Do not print the density curves.

(b) Now assume that the mean of temperature is set at 210°F. Enter the value of σ as 4.75, then 5.5, and eventually 7.25. After each entry, carefully examine the shape of the corresponding density curve. You are not supposed to print the density curves. Describe briefly the change in the appearance of the density curve as σ increases from 4.75 to 7.25. How does the change in the value of σ affect the density curve of temperature? Does the proportion of manufactured engines that are unacceptable (temperature values below 190°F or above 225°F) increase or decrease? Explain briefly.

2. In this part, you will determine how the changes in the mean and standard deviation affect the proportion of engines that are unacceptable. Assume in this question that temperature follows a normal distribution. Answer the questions below using either the template *Normal Probabilities* or the appropriate built-in functions in Excel.

   (a) Suppose $\mu = 208$ and $\sigma = 7.25$. What proportion of engines are unacceptable (below 190°F or above 225°F)? What proportion of engines are unacceptable when the value of μ is 215 with the same value of σ? You do not have to print out the values displayed on the worksheet, just copy the numerical values from the screen to your report.

   (b) Now assume $\mu = 208$ and $\sigma = 4.75$. What proportion of engines are unacceptable?

   (c) What is the proportion of engines with the temperature value within 1, 2, and 3 standard deviation of the mean of 208? Assume here that $\sigma = 6.25$.

   (d) Assume $\mu = 208$ and $\sigma = 6.25$. What is the temperature value exceeded by 90% of engines? What would be the temperature value exceeded by 95% of engines?

   The results to the above question with 4 decimal places for parts (a) – (d) should be reported in the form of the table like the one below. The table is also included in the worksheet *Data* in the data file. You can paste the table into your report. **Express all proportions in decimal form.**

PART	PARAMETERS	PROBLEM	ANSWER
(a)	$\mu = 208$ and $\sigma = 7.25$	Proportion of unacceptable	
	$\mu = 215$ and $\sigma = 7.25$	Proportion of unacceptable	
(b)	$\mu = 208$ and $\sigma = 4.75$	Proportion of unacceptable	
(c)	$\mu = 208$ and $\sigma = 6.25$	Within 1 standard deviation	
		Within 2 standard deviations	
		Within 3 standard deviations	
(d)	$\mu = 208$ and $\sigma = 6.25$	Temp exceeded by 90%	
		Temp exceeded by 95%	

3. Open the worksheet *Data*. The worksheet contains three random samples of size 20, 60, and 400 of (ordered) temperatures expected to follow a normal distribution with a mean of 208°F and a standard deviation of 7.25°F. In parts (a) and (b), you will use the COUNTIF function described in the *Lab 2 Instructions.*

   (a) Count the number of unacceptable engines (temperature below 190 or above 225) in the sample. Obtain the relative frequencies for each sample. Are your results consistent with the theoretical prediction obtained in Question 2, part (a)? Which sample produces the value closest to the predicted value?

   (b) Use the COUNTIF function to determine the number of observations in each of the three samples within 1, 2, and 3 standard deviations of the mean, inclusively. Also, obtain the relative frequencies. Compare your results with the Empirical Rule. Which sample produces results which are the most consistent with the values predicted by the rule?

(c) Use the worksheet *Normal Probability Plot* to verify the assumption of normality for the sample of 400. The template automatically displays the normal probability plot once the number of observations $n = 400$ is entered and the data are pasted from the worksheet *Data*. Place the plot in your report. Describe the pattern displayed in the plot. Is there any strong indication that the data do not follow the normal distribution? Explain briefly. (**Hint:** You may also find a histogram to describe the shape of the distribution.)

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 2016 or a newer version should be used in this part (older versions may produce different sequences of random numbers for the same seed).

4. Open the worksheet *Simulation*. The worksheet allows you to simulate the temperature measurements for 60 random samples of engines. Use the *Random Number Generation* feature (Normal, seed 2023) to generate 60 samples of size $n = 20$ from the normal distribution with $\mu = 208$ and $\sigma = 5.5$. This corresponds to selecting 60 samples, each consisting of 20 engines. The data will be entered in the form of 60 columns, each consisting of 20 rows into the range B10:BI29. In other words, the range contains the temperature measurements for 1200 engines.

   Once the data are entered, the values of the variables AVERAGE are automatically displayed in row 411.

   (a) What are the mean and standard deviation for the distribution of the sample mean predicted by theory? Does the Central Limit Theorem apply for a random sample of 20? Why or why not?

   (b) The variable AVERAGE in row 411 shows the average temperature for each sample. Obtain a histogram of the sample means using the following bins: 203.00, 204.00, 205.00, ..., 211.00. The format of your histogram should be the same as the format from previous labs and the *Lab 1 Instructions*. Describe the shape of the histogram (modality, skewness, outliers).

   (c) The worksheet *Simulation* also displays *Summary Statistics* for the variable AVERAGE. Use the feature to obtain the mean and standard deviation of the sample means for the 60 samples. Compare them with the values predicted by theory. Should the values be identical? Explain briefly.

5. Now we repeat all parts of Question 4 with $n = 400$. First, clear the range B10:BI29 in the worksheet *Simulation*. Use the *Random Number Generation* feature (Normal, seed 2023) to generate 60 samples of size $n = 400$ from the normal distribution with $\mu = 208$ and $\sigma = 5.5$. The data will be entered into the range B10:BI409 in the form of 60 columns, each consisting of 400 rows.

   (a) What are the mean and standard deviation for the distribution of the sample mean predicted by theory? Does the Central Limit Theorem apply for a random sample of 400? Why or why not?

   (b) Obtain a histogram of the sample means using the following bins: 207.00, 207.25, 207.50, ..., 209.00. Describe the shape of the histogram (modality, skewness, outliers). Compare to the histogram obtained in Question 4, part (b).

   (c) Obtain the mean and standard deviation of the sample means for the 60 samples. Compare them with the values predicted by theory and relate to the similar comparison in Question 4, part (c). What do you conclude? State your findings briefly.

LAB 3 ASSIGNMENT MARKING SCHEMA

Question 1 (8)

(a) Description of the change in the density function as μ increases: 2 points
Effect of the change in μ on the proportion of unacceptable engines: 2 points
(b) Description of the change in the density function as σ increases: 2 points
Effect of the change in σ on the proportion of unacceptable engines: 2 points

Question 2 (16)

PART	PARAMETERS	MARKS
(a)	$\mu = 208$ and $\sigma = 7.25$	2
	$\mu = 215$ and $\sigma = 7.25$	2
(b)	$\mu = 208$ and $\sigma = 4.75$	2
(c)	$\mu = 208$ and $\sigma = 6.25$	2
		2
		2
(d)	$\mu = 208$ and $\sigma = 6.25$	2
		2

Question 3 (40)

(a) Counts and relative frequency of unacceptable engines: 2 points each (6 points total)
Discussion of consistency with the theoretical prediction: 2 points
Sample that produces the value closest to the predicted value: 2 points
(b) Counts and relative frequencies in each sample: 2 points each (18 points total)
Discussion of consistency with the Empirical Rule: 2 points
Sample that produces the value closest to the predicted value: 2 points
(c) Normality plot: 4 points
Pattern in the normal probability plot: 2 points
Discussion of normality: 2 points

Question 4 (22)

(a) Mean and standard deviation predicted by theory: 2 points each (4 points total)
Application of Central Limit Theorem: 2 points
(b) Correctly formatted histogram of the sample means: 6 points
Shape of the histogram: 3 points
(c) Mean of the averages: 2 points
Standard deviation of the averages: 2 points
Comparison with values predicted by theory: 3 points

Question 5 (26)

(a) Mean and standard deviation predicted by theory: 2 points each (4 points total)
Application of Central Limit Theorem: 2 points
(b) Correctly formatted histogram of the sample means: 6 points
Shape of the histogram: 3 points
Histogram comparison: 2 points
(c) Mean of the averages: 2 points
Standard deviation of the averages: 2 points
Comparison with values predicted by theory: 3 points
Comparison with Question 4 values: 2 points

TOTAL = 112