SOLUTIONS TO THE LAB 3 ASSIGNMENT

Question 1

(a) In this part, compare the shapes of the normal density curves as the value of the mean increases while the standard deviation is kept at the same level of 5.5°F.

Changing μ without changing σ moves the normal curve to a new location without altering its spread. In our case, increasing the value of μ from 205 to 210 shifts the curve right by 5 units. Similarly, increasing the value of μ from 210 to 215 shifts the normal curve right by another 5 units. The peak is always in line with μ, no matter what its value.

When μ decreases from 210 to 205, there will be a substantial increase in the number of engines below 190 but the number of engines above 225 will decrease. When μ increases from 210 to 215, there will be a substantial decrease in the number of engines above 225 but the number of engines below 190 will decrease. Thus, a movement in either direction away from 210 causes an increase in one proportion of unacceptable engines, but a decrease in the other.

(b) The standard deviation controls the spread of a normal curve. Increasing the standard deviation lowers or flattens the curve (note the lower values on the *y*-axis). Random engines selected from the distribution with σ = 4.75°F are more likely to have temperature values near the mean of 210°F than engines selected from the distribution with σ = 5.5°F. Similarly, random engines selected from the distribution with σ = 5.5°F are more likely to have temperature values near the mean of 210°F than engines selected from the distribution with σ = 7.25°F.

The increase in the standard deviation means that the proportion of manufactured engines that are unacceptable (recorded temperature values below 190°F or above 225°F) increases (more data in the tails). As the standard deviation increases, there is a substantial increase in the number of engines outside the “acceptable” range.

Question 2

The results obtained from the worksheet can be summarized in the following table.

PART	PARAMETERS	PROBLEM	ANSWER
(a)	μ = 208 and σ = 7.25	Proportion of unacceptable	0.0160
	μ = 215 and σ = 7.25	Proportion of unacceptable	0.0842
(b)	μ = 208 and σ = 4.75	Proportion of unacceptable	0.0002
(c)	μ = 208 and σ = 6.25	Within 1 standard deviations	0.6827
		Within 2 standard deviations	0.9545
		Within 3 standard deviations	0.9973
(d)	μ = 208 and σ = 6.25	Temp exceeded by 90%	199.9903
		Temp exceeded by 95%	197.7197

Question 3

(a) The results can be summarized in the following table:

SIZE	FREQUENCY	RELATIVE FREQUENCY
20	1	1/20 = 0.0500
60	3	3/60 = 0.0500
400	7	7/400 = 0.0175

The calculated probability in Question 2, part (a) is 0.0160. Thus, we expect that approximately 1.60% of engines will be unacceptable if μ = 208 and σ = 7.25. Incidentally, the observed relative frequency is closest and, therefore, most consistent with the predicted probability for the sample size of 400. As sample size increases, relative frequency usually gets closer to the theoretical probability, yet the samples here do not fully display that effect (with the first two values being the same), which can happen with random samples.

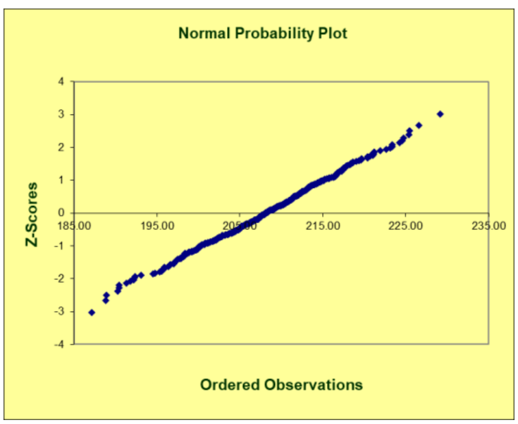
(b) In this part, the students are supposed to determine the number of observations in the samples that fall within 1, 2, and 3 standard deviations of the mean, where $\mu = 208$ and $\sigma = 7.25$. The following table summarizes the results.

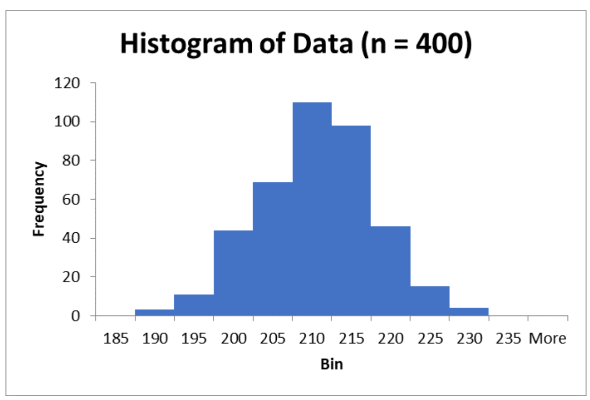
SIZE	***k***	**WITHIN *k* ST. DEV. OF μ = 208 WITH σ = 7.25**	**FREQUENCY**	**RELATIVE FREQUENCY**
20	1	200.75 – 215.25	13	0.6500
	2	193.50 – 222.50	18	0.9000
	3	186.25 – 229.75	20	1.0000
60	1	200.75 – 215.25	35	0.5833
	2	193.50 – 222.50	53	0.8833
	3	186.25 – 229.75	60	1.0000
400	1	200.75 – 215.25	271	0.6775
	2	193.50 – 222.50	377	0.9425
	3	186.25 – 229.75	400	1.0000

To count all cells in the range A2:A21 with the entries between 200.75 and 215.25 (inclusively), you can use the following formula.
=COUNTIF(A2:A21,"<=215.25") - COUNTIF(A2:A21,"<200.75")
Note: This is one of a few possible solutions to find the count. Similar functions will give the remaining counts above.

According to the Empirical Rule, approximately 68%, 95%, and 99.7% of observations in a normal distribution lie within one, two, and three standard deviations of the mean, respectively. For $n = 20$, the results are not that consistent with the Empirical Rule and become considerably worse at $n = 60$, yet the consistency improves when $n = 400$. Incidentally, all 3 sample sizes produce 100% within 3 standard deviations. The sample consisting of 400 observations produces results which are the most consistent with the values predicted by the rule.

(c) The normal probability plot produced by the template is shown below.





As the points on the above normal probability plot generally lie close to a straight line, the plot supports the assumption that the data are normal. (There are points that could be considered outliers at the lower end, however, and there are more points at the lower end not satisfying the overall trend.) The histogram of the data (not required) implies that the distribution is approximately symmetric and it is normal, yet also shows a slightly heavier right tail than the left. Alternatively, the distribution is slightly left-skewed since the skewness value is -0.0802.

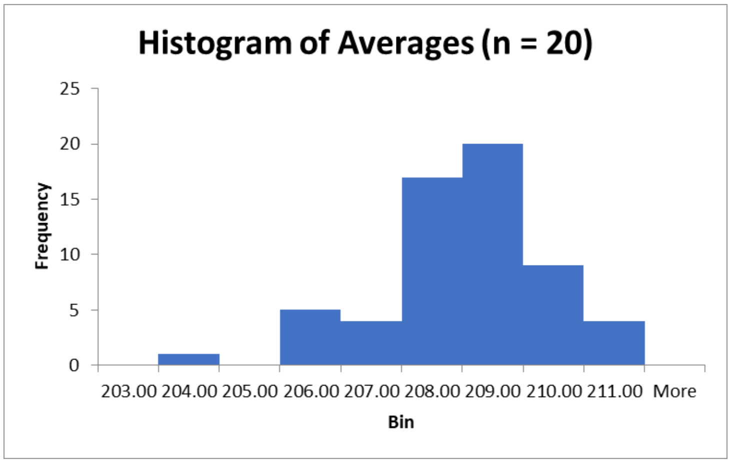
Question 4

According to the Central Limit Theorem, if a random sample of size $n \geq 30$ is drawn from a population with mean μ and variance σ^2, the sample mean approximately follows a normal distribution with mean μ and variance σ^2/n. The larger the sample size, the better the approximation.

(a) The sample mean should follow a normal distribution with the mean $\mu_{\bar{x}} = \mu = 208.0000$ and the variance $\sigma^2/n = (5.5^2)/20 = 1.5125$. Thus, the standard deviation is $\sigma = (1.5125)^{1/2} = 1.2298$.

Since $n = 20 < 30$, the Central Limit Theorem should not apply, but the population distribution of temperature is normal, so the distribution of the sample mean should be normal.

(b) The histogram of the sample means ($n = 20$) is displayed below.



The histogram is unimodal and appears moderately left-skewed. Since the data is generated from normal distribution samples, the distribution of the sample mean should be approximately normal, but this particular simulation does not completely display this idea. The small sample size (and small number of samples) allows this situation to occur. There is one outlier at 203.7645.

(c) The mean of the 60 averages reported by the worksheet is 208.0307 and the standard deviation is 1.3655. Comparing to the values predicted by theory calculated in part (a), the observed value of the mean is very close and the standard deviation is reasonably close to their respective predicted values. The values do not have to be identical because this is a simulation with only 60 samples.

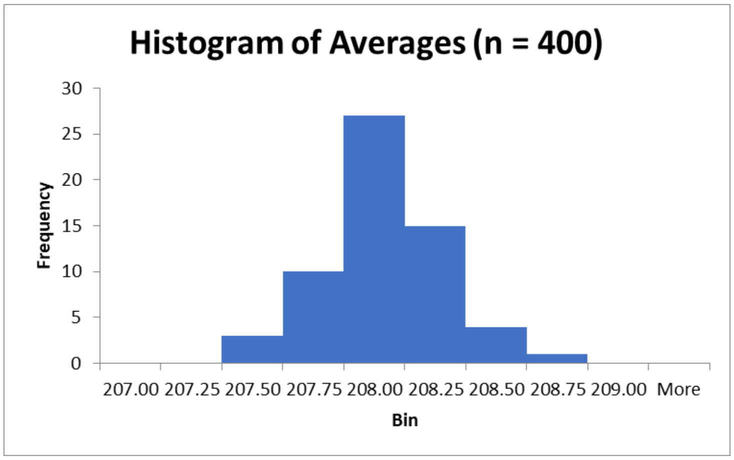
Question 5

(a) The sample mean should follow a normal distribution with the mean $\mu_{\bar{x}} = \mu = 208.0000$ and the variance $\sigma^2/n = (5.5^2)/400 = 0.075625$.

Thus, the standard deviation is $\sigma = (0.075625)^{1/2} = 0.2750$.

Since $n = 400 \geq 30$, the Central Limit Theorem could apply, but the population distribution of temperature is already normal. Either way, the sample mean distribution should be normal.

(b) The histogram of the sample means (n = 400) is displayed below.



The histogram is unimodal and appears approximately symmetric or slightly to moderately right-skewed. Note that skewness is possible due to random simulation. There are no outliers.

The histogram of averages for n = 400 more closely resembles a normal distribution (compared to the histogram in Question 4).

(c) The mean of the 60 averages reported by the worksheet is 207.9278 and the standard deviation is 0.2483. Comparing to the values predicted by theory calculated in part (a), the observed value of the mean is very close and the standard deviation is even closer to their respective predicted values. The values do not have to be identical because this is a simulation with only 60 samples. The mean is slightly farther from the value of 208 compared to the value from Question 4. The standard deviation here is notably closer to its predicted value compared to the difference between values in Question 4. This may not always occur due to the simulation of only 60 samples.

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