LAB 3 ASSIGNMENT

NORMAL & SAMPLING DISTRIBUTIONS, CENTRAL LIMIT THEOREM

Examining a Filling Process

1)

(a) Assume that the mean is set at 130 ml. Enter the value of σ as 7, then 14, and eventually 21 ml. After each entry, carefully examine the shape of the corresponding density curve. Briefly describe the change in the appearance of the density curve as σ increases from 7 to 14 to 21. What does the increase in the standard deviation value mean to the filling process? Does the percentage of underfilled boxes (boxes containing less than 130 ml) change?

The increase in the standard deviation value means that the graph's spread will be larger. As σ increases from 7 to 14 to 21 the probability density function curves' shape has altered, but its central location has not. Long, flat bell-shaped curves are produced by large values of standard deviation σ whereas small values of σ result in thinner, sharper bell-shaped curves. Specifically, the peak is lowered for bigger σ, the left and right tails become longer while the graph is still symmetrical.

The percentage of underfilled boxes (boxes containing less than 130 ml) will become bigger and bigger as the value of σ changes from 7, to 14, and eventually to 21 since the variance become larger. Percentage of underfilled and overfilled increases but they are all at 50%

The area under the tails is bigger which means that more data will be distributed near the mean's value. The percentage of underfilled increase but the ratio between the overfilled and underfilled will be same.

Now set the standard deviation to 7 ml and change the mean. Enter the value of μ as 130, then 137, and eventually 144 ml. The corresponding density curve changes. The change in the value of μ affect the filling process? Does the percentage of underfilled boxes increase or decrease?

Since the probability density function is a normal distribution, it is symmetric about the mean value μ and has a "bell-shaped" curve. The probability density functions of normal distributions with μ = 130, then 137, and eventually 144 will with σ remains the same, which illustrates the fact that as the mean value μ is changed, the shape of the density function remains unaltered while the location of the density function changes. Specifically, the max height (~0.013) and range stays constant while the curve shift to the right because the value of μ increases along the x axis move the graph's peak to the right. For example, the graph shift rights by exactly 7 units each time μ changes from 130 to 137, and then from 137 to 144 ml.

The area under the tails for the distance from the end of the left tail to 130 becomes smaller as μ shifts from 130 o 144, which means that the percentage of underfilled boxes decreases.

2)

PART	PARAMETERS	PROBLEM	ANSWER
(a)	$\mu = 137$ and $\sigma = 14$	Percentage of underfilled	0.3085
	$\mu = 144$ and $\sigma = 14$	Percentage of underfilled	0.1587
(b)	$\mu = 137$ and $\sigma = 21$	Percentage of underfilled	0.3694
(c)	$\mu = 137$ and $\sigma = 14$	Within 1 standard deviation	0.6827
		Within 2 standard deviations	0.9545
(d)	$\mu = 130$ and $\sigma = 7$	Amount exceeded by 95%	118.486
		Amount exceeded by 99%	113.7156

(a) With a value of m as 135 and s as 14, 30.85% of the boxes will now contain less than 130ml and be underfilled. When m as 144 and s is 14, then 15.87% of the boxes will be underfilled.

(b) When the mean fill amount is 137 ml, and standard deviation is equal to 21, then 36.94% of the boxes will be

underfilled.

(c) For the fill amount to be within 1 standard deviation of the mean, the fill amount must be within 120 and 150 ml, which corresponds to 68.27% of the boxes. The fill amount within 2 standard deviations of the mean (105-165 ml) is 95.45%.

(d) The fill amount exceeded by 95% of the boxes is 118. .486 ml, and the fill amount exceeded by 99% of the boxes is 113.7156 ml.

.

3) Open the worksheet Data. The worksheet contains three random samples of size 25, 50, and 400 of (ordered) fill amounts expected to follow a normal distribution with a mean of 137 ml and a standard deviation of 14 ml.

(a) The relative frequencies for the number of underfilled boxes in each sample are below:

Size	Frequency	Relative Frequency
25	7	0.28
50	15	0.3
400	118	0.295

With the exception of size 25, which is 0.0285 below, the findings of the relative frequency calculation shown in the above table are in line with the theoretical prediction in 2a. The sample size 50's relative frequency is only 0.085 lower than the predicted value. Size 25 does not stay consistent because it contains only 25 samples which is not large enough.

The sample with size of 50 is the one that produces the value closest to the predicted value.

(b)

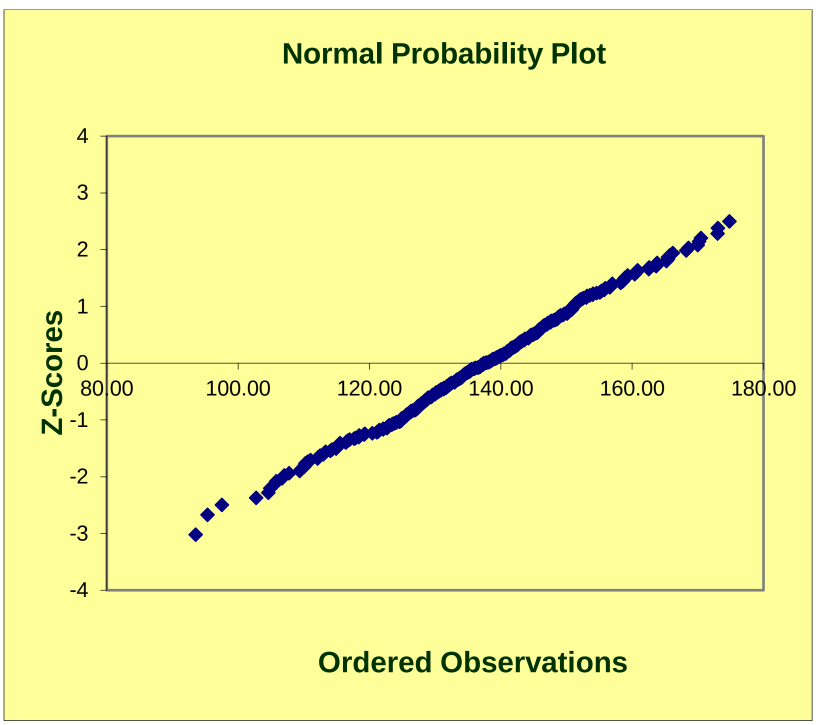
Size	K	Within K sd	Frequency	Relative Frequency
25	1	123-151	21	0.840
	2	109-165	25	1.000
	3	95-179	25	1.000
50	1	123-151	38	0.760
	2	109-165	48	0.960
	3	95-179	50	1.000
400	1	123-151	283	0.708
	2	109-165	374	0.935

Size	K	Within K sd	Frequency	Relative Frequency
	3	95-179	397	0.993

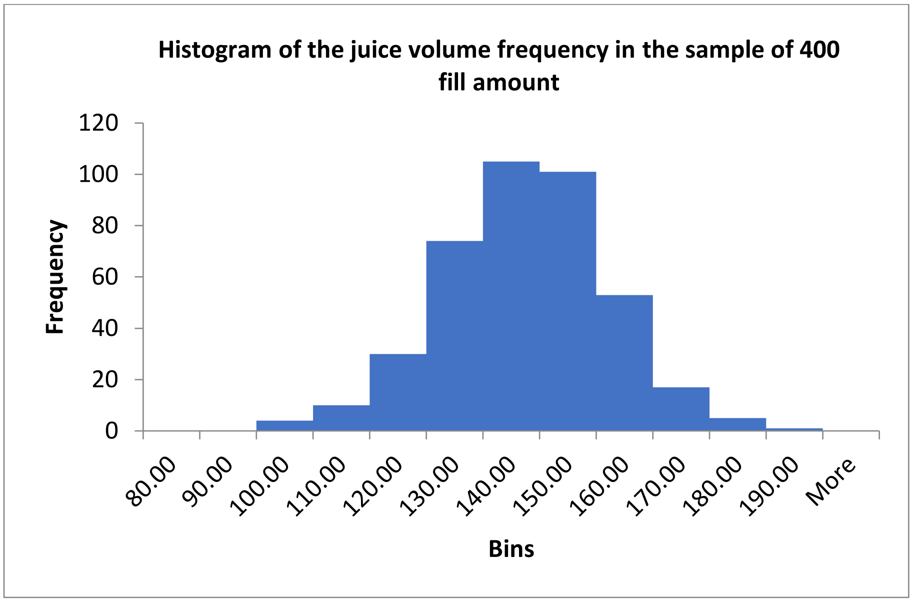
The Empirical Rule states that approximately 65%,95% and 99.7% of the observation lie within 1,2 and 3 standard deviations respectively.

According to the data obtained in the relative frequency column, the sample of size 400 produces results which are the most consistent with the values predicted by the rule. All samples with 2 standard deviations of the mean follow the empirical rule of being above 95% except the sample of 400 which is only 93.5% (with 1.5% difference) and lastly, all sample with n=25 follow the empirical rule with all 3 standard deviation of the mean above 99.7%. except the sample n=400 giving 99.25% with 0.45% difference.

Overall, the n=25 and 50 samples are compatible with the empirical rule, however the n=400 sample is not. The best results are obtained from samples of 50 since each standard deviation interval's values are close together and greater than those of the empirical rule.



The data points in the normal probability plot are consistent with a sample from a normal distribution as the points lie close to a straight line. The statistical rule that stipulates that for a normal distribution, practically all observed data will fall within three standard deviations of the mean or average. The more the points deviate from this line, the higher the indication of departure from normality. There is no strong indication that the data do not follow the normal distribution although the data points don't really line up at the two ends, especially at the very beginning. This means that the smallest and largest values are not as extreme as expected in normality.



The histogram is unimodal and symmetrical, so it follows a normal distribution.

4.

(a) Does the Central Limit Theorem apply for a random sample of 25? Why or why not? Assuming it does apply, what is the probability that the average volume in a random sample of 25 boxes does not exceed 130?

According to the central limit theorem, even if the population distribution is not normal, the sampling distribution of a sample mean is about normal if the sample size is large enough. In general, we consider "sufficiently large" to be 30 or larger. Since n= 25 < 30 so the Central Limit Theorem does not apply for a random sample of 25.

Assuming it does apply, the probability that the average volume in a random sample of 25 boxes does not exceed 130 is 0.0062 using Left – Interval Probability with mean = 137 and standard deviation= 14^2/25 = 2.8.

4)(b) The histogram is somewhat symmetrical, but it is still not quite the bell shape. Since the data simulates from a normal distribution population, the histogram should look exactly like a normal graph which is a bell-shaped curve that is symmetric about μ = 137. However, this is not the case for the histogram below as the sample size n = 25 is not large enough. The histogram in Figure 4b has a shape that doesn't represent a normal distribution. The data for the mean concentrates between 135 to 142 in the main body of the graph.

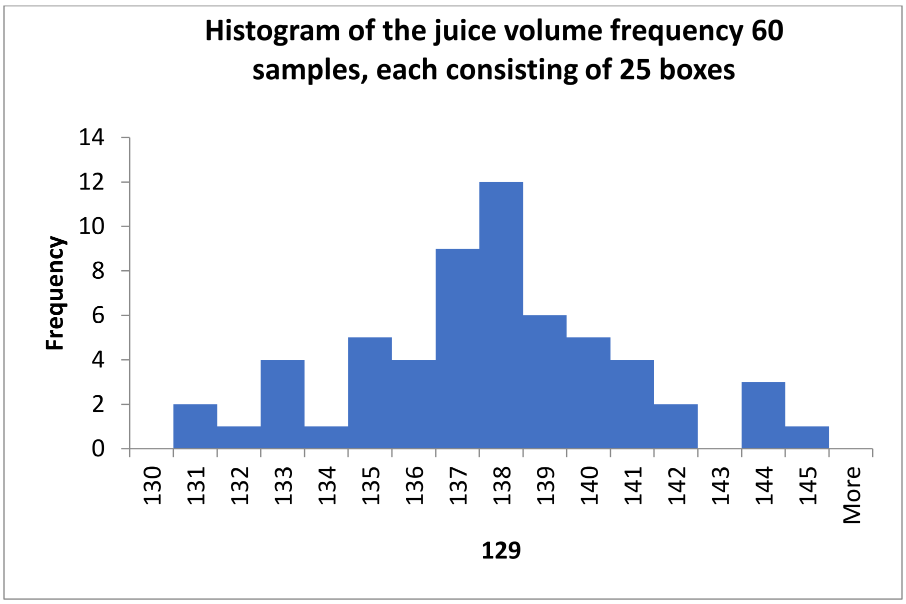


Figure 4b

4.(c) The variable AVERAGE's summary statistics are likewise shown on the worksheet Simulation. Using the functionality, the sample means is 137.3265 and standard deviations is 3.1194 for the 60 samples. For comparison, the standard deviation 3.1194 is a larger than the theory standard deviation 2.8. Similarly, the sample's mean 137.3265 is a bit larger than the theory mean 137. The average and standard deviation are very close but not identical since this is the simulation only for a sample of 60 boxes.

5)

(a) The Central Limit Theorem applies for a random sample of 400 because n= 400 > 30, which indicates that the random sample size is large enough.

Using the value in the cell Left – Interval Probability with mean = 137 and standard deviation= 14^2/400 = 0.7, the obtained probability that the average volume in a random sample of 400 boxes does not exceed 130 is 0). The sample is consistent with the Center Limit Theory, which states the 99.7% of the sample observation 3 sd. Since 3sd = 0.7*3 = 2.1, 99.7% of the sample lie within 137±2.1 (between 139.1 and 134.9) and only 0.3% outside that range.

(b Describe the shape of the histogram, comparing to the histogram obtained in Question 4, part (b).

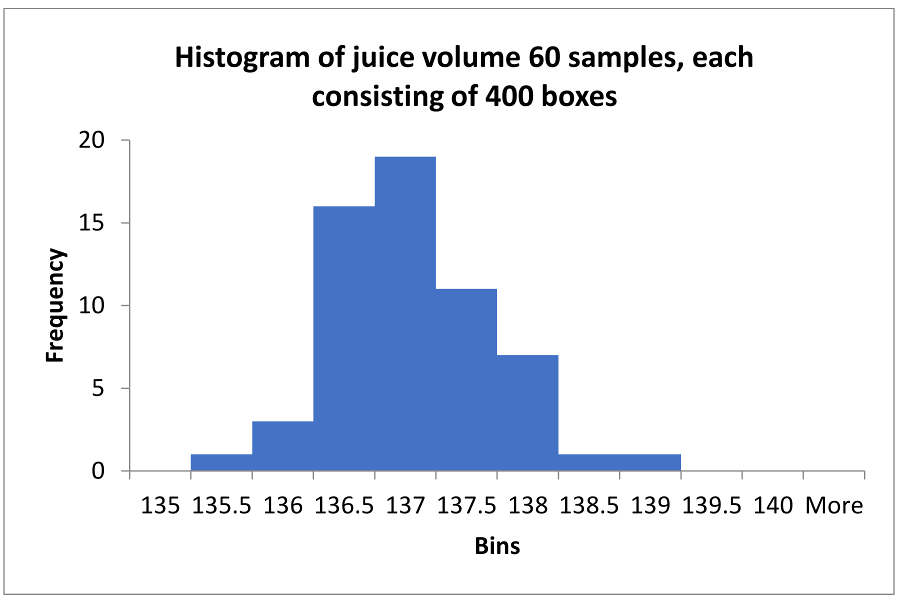


Figure 5b

The histogram looks right skewed, however, the mode=mean=137. The right tail is longer than the left tail and data concentrates from 135.5 to 139.5.

The spread in the histogram figure 5b is smaller than the spread in the histogram figure 4b.

The peak is about 19 at the mean volume 137, which is higher than the peak value obtained from the histogram 4b at around 13.

(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.

In question 5c, the mean and standard deviation of the sample means for the 60 samples are 136.8163 and 0.6320, respectively. The theoretical value for mean and standard deviation are 137 and 0.7 (=14^2/400= 0.7) which is close but not identical to the sample mean and standard deviation.

Refer to question 4, the actual standard deviation is 3.1194; the theoretical standard deviation is 2.8. The sample mean of 137.3265 is also slightly higher than the theoretical mean of 137.

In conclusion, the difference between the theory and the simulation value is smaller in 5c compared to 4c because the sample size is bigger. In 5c, the sample mean and standard deviation are even closer to the theoretical values, compared to the previous sample and theoretical means and standard deviations in 4c. The much larger sample size makes the sampling distribution of the mean of theobservations much more normal than the smaller sample sizes. The increasing sample sizes makes the sampling distribution increasingly normal.