STATS 151 X01

Group #21

Lab 2

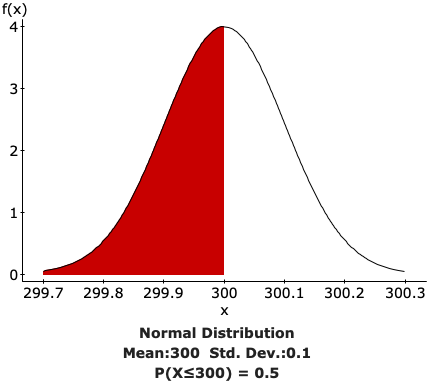
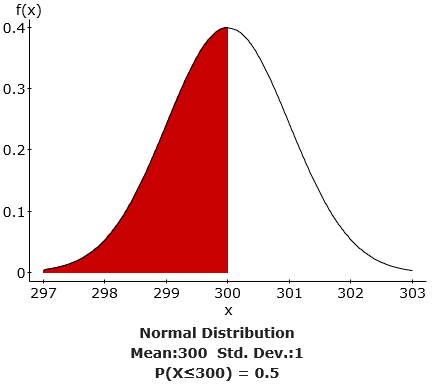
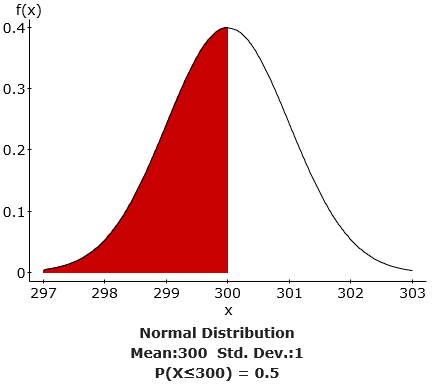
**Joseph, Robert**

Kim, Jamie

Chen,Xinhao 

1

a) When the mean is at 300 the percentage of the bottles containing less than 300ml does not change when the standard deviation increases or decreases. In general, when the standard deviation increases the probability increases and when sigma decreases the probability decreases, the greater the standard deviation, the more underfilled and overfilled the bottles are but the percentage of underfilled and overfilled bottles remains the same because the distribution is normal and the upper bound for underfilled bottles is the same as the mean, which is also the lower bound for overfilled bottles. So, no matter the standard deviation, the proportion of overfilled to underfilled would always be 1 is to 1, meaning the percentage of underfilled bottles is always 50%, regardless of the standard deviation.



b) The percentage of underfilled bottles (the bottles containing less than 300 ml) in this case is 15.865525%

A picture containing screenshot

Description automatically generated

The percentage of underfilled bottles if σ were 1 ml would be 2.275013%

A screenshot of a cell phone

Description automatically generated

The percentage of underfilled bottles if σ were 0.5 ml would be 0.003167%.

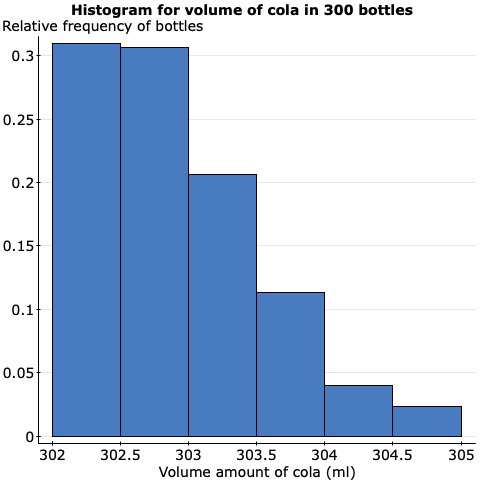
A close up of a logo

Description automatically generated

In general, this implies that the decrease in the percentage of underfilled bottles is directly proportional to the decrease in standard deviation. Therefore, the lower the standard deviation, the lower the percentage of underfilled bottles.

2)

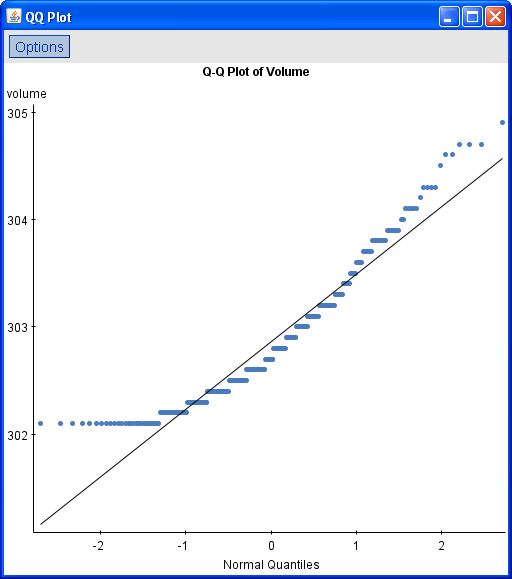
a)



b)

The histogram is beginning with the bin 302 and the width of each bars is 0.5.  The histogram is right skewed and unimodal which means that the mean is greater than the median. The lower quartile which has a relative frequency of 0.25 is in the range 302 to 302.5. The median lies in the range 302.5 to 303 and the upper quartile is in the range 303 to 303.5. The spread is about 3 ml. The bottles are overfilled which is between 0.2-0.3. The company's claim that are bottles are slightly overfilled is correct because almost all the bottles are either 302 ml or greater when the bottles are supposed to have 300 ml of cola in them. Since the sample size is large, we can assume the same is true for all the bottles so we can safely assume that the company’s claims are correct.

c)



The QQ-plot agrees with our histogram and shows that our data is not normal but highly skewed to the right. We can see this because for standard deviation of -1 to -2, the data is very far from the normal distribution plot, similarly for standard deviation of 1 and 2 the plot for the sample is far from the normal distribution line suggesting that the data is skewed and doesn’t have a normal distribution. If the data was normally distributed, most of the plot would have been close to the normal distribution line which clearly isn’t the case here. Also, no plot is less than 302 so the company's claims that the bottles are slightly overfilled are confirmed again. Hence, this plot is consistent with our conclusion in part (b).

d)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Column** | **n** | **Mean** | **Variance** | **Std. Dev.** | **Std. Err.** | **Median** | **Range** | **Min** | **Max** | **Q1** | **Q3** |
| volume | 300 | 302.86566 | 0.39236242 | 0.62638843 | 0.036164552 | 302.7 | 2.8 | 302.1 | 304.9 | 302.4 | 303.2 |

Mean is the skewed direction of Median, which is smaller than mean in the part of skewed distribution and Median does not lies farther from the Q3 than Q1.

The mean is greater than the median with a difference of about 0.17 ml which suggests that the data is right skewed.

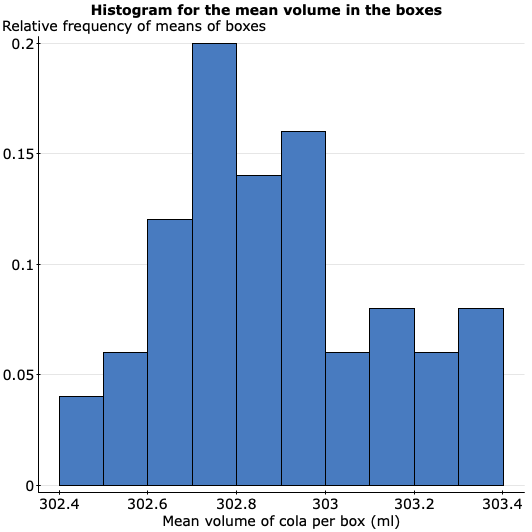
Q3 - Q2 = 303.2 - 302.7 = 0.5

Q2 - Q1 = 302.7 - 302.4 = 0.3

So (Q3 - Q2) > (Q2 - Q1) which means that the data is right skewed and not normally distributed.

This is consistent with conclusions in part (b) as we said that the lower quartile is in the range 302 to 302.5. The median lies in the range 302.5 to 303 and the upper quartile is in the range 303 to 303.5. The minimum value is 302.1 and hence reaffirms the company's claim that all the bottles are slightly overfilled. The standard error is a kind of error of distribution of sample mean, which could estimate population mean and the observed standard error is 0.036164552 which is a classical deviation in 300 randomly selected cases

3)a)



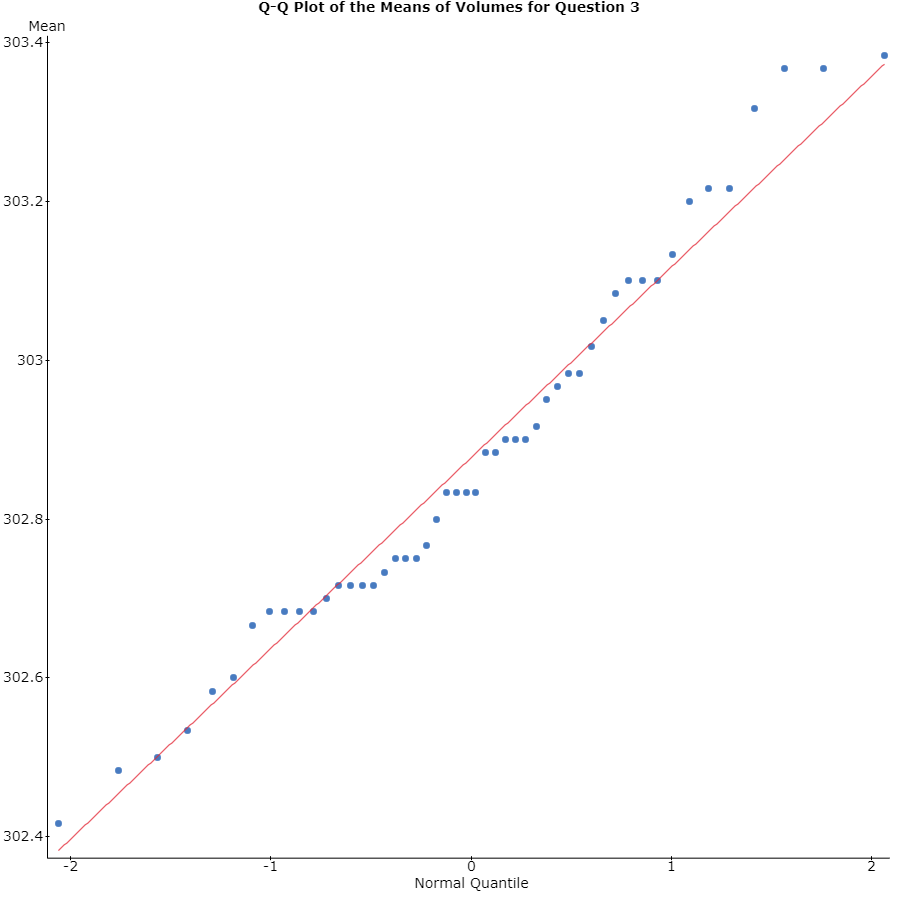
b)

The graph in part a) appears to be unimodal and fairly close to being symmetrical. No, it is not normally distributed as it is not symmetric

The histogram appears to be slightly right skewed and The histogram of the individual observations in 2 was clearly right skewed while the histogram of the means is just slightly right skewed and is unsymmetrical

The spread was greater for the histogram for the individual observations since the range was 3mL (305 - 302) while the histogram of the means has less spread as the range is 1mL (303.4 - 302.4

c)



Yes, the comparison between the Histogram and the Q-Q plot confirms that the data is evenly distributed or normal.

 Comparing the two Q-Q plots, the plot in question 2 has a large portion of the data between quantiles 0 and passed -2 making the plot right skewed unlike the plot in question 3 which is evenly distributed.

d)

**Summary statistics:**

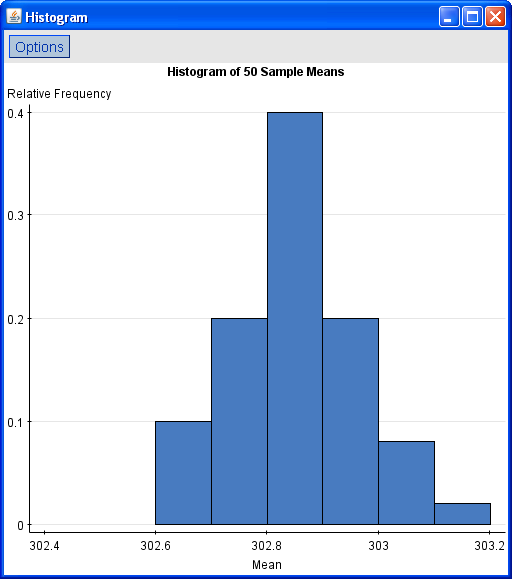
|  |  |  |
| --- | --- | --- |
| **Column** | **Mean** | **Std. dev.** |
| Mean | 302.87767 | 0.24005408 |

The theory of sampling distribution predicts that the mean of the means is the mean hence it is should be 302.86567 as in 2 d). But we obtained a slightly higher mean of 302.87767.

It also predicts that the standard deviation of the means is σ/√n hence this should be 0.6263884/√6 = 0.2557219935 which is slightly higher than the standard deviation that we obtained in 3d) (0.24005462).

The standard deviation here shows the standard deviation from the mean of mean calculated for 50 boxes each with 6 bottles in them. So, if packs of 6 bottles are taken, the expected standard deviation would be 0.24 between the packs but not necessarily individual bottles.

4)a)

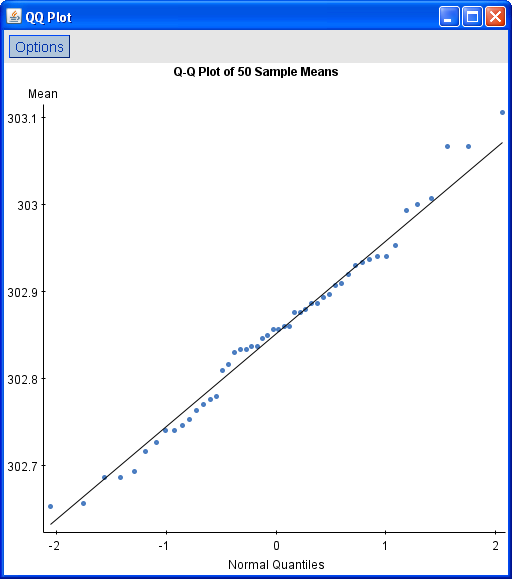


b) The histogram is unimodal and fairly symmetric with a very slight right skew. The range is from 302.6 to 303.2. The data appears to be normally distributed as the histogram is bell shaped and symmetric about the mean.

In comparison to the histogram in Question 2 part (a), we can see that this graph has a normal distribution while the other graph has a distribution skewed to the right. The spread of the histogram in Question 2 is about 3 ml while the spread in this histogram is 0.6.

In comparison to the histogram in Question 3 part (a); while this graph has a normal distribution, the graph in Question 3 is also fairly normal with the right side a little higher but it is still relatively normal. While the spread in this histogram is 0.6, the spread of the histogram in Question 3 is about 1 ml.

c)



The distribution in this Q-Q plot is distributed normally as the points closely follow the line as compared to the Q-Q plot in question 3), in which the points go towards the left of the x axis, between 303.2 and 303.4 on the y axis which shows much more right-skewness as compared to the one in 4b).

d)

**Summary statistics:**

|  |  |  |
| --- | --- | --- |
| **Column** | **Mean** | **Std. dev.** |
| Mean | 302.85247 | 0.1064492 |

The mean in Question 3 is 302.87767 while the mean in Question 4 is 302.85247. The difference between both means is 0.0252. Since, Question 3 has a bigger mean, this means it also has a larger distribution.

The theory of sampling distribution says that the mean of the means is the mean, The predicted standard deviation of the sample mean for n=30 is 0.1144 ( σ/√n) than the predicted standard deviation of n=6  standard deviation of the sample mean for n = 6 of 0.25557 to the a

The predicted standard deviation of the sample mean of 0.1064

According to the sampling distribution theory as the sample size increases, the standard deviation should decrease as the spread decreases around the mean.  When comparing the standard deviation from 3d) and 4d), the theory is proven because 4) has a larger sample size (n=30) and a smaller standard deviation than 3) with a sample size of n=6.