STATS 151 A1

GROUP #95

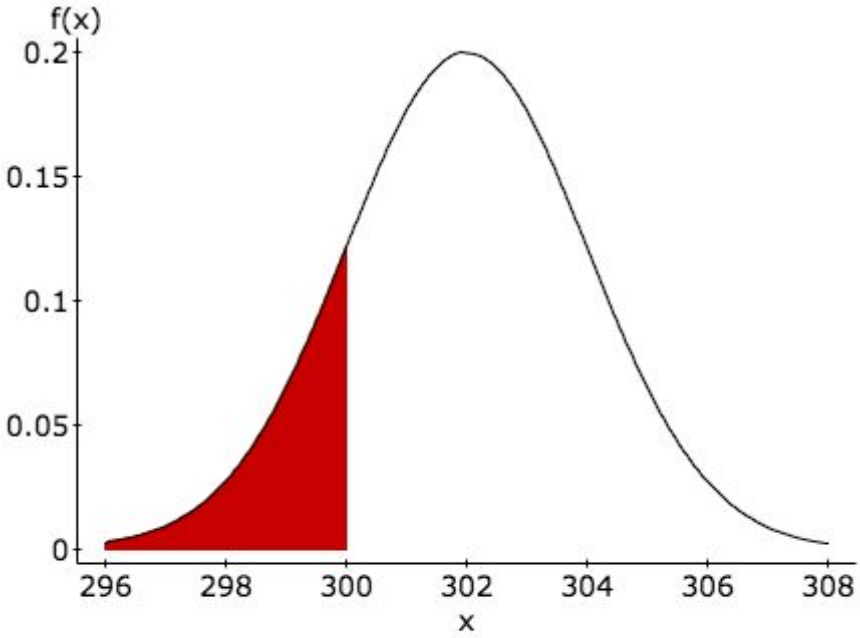
LAB 2

MULYK, Jasmine

FONTAINE, Emme

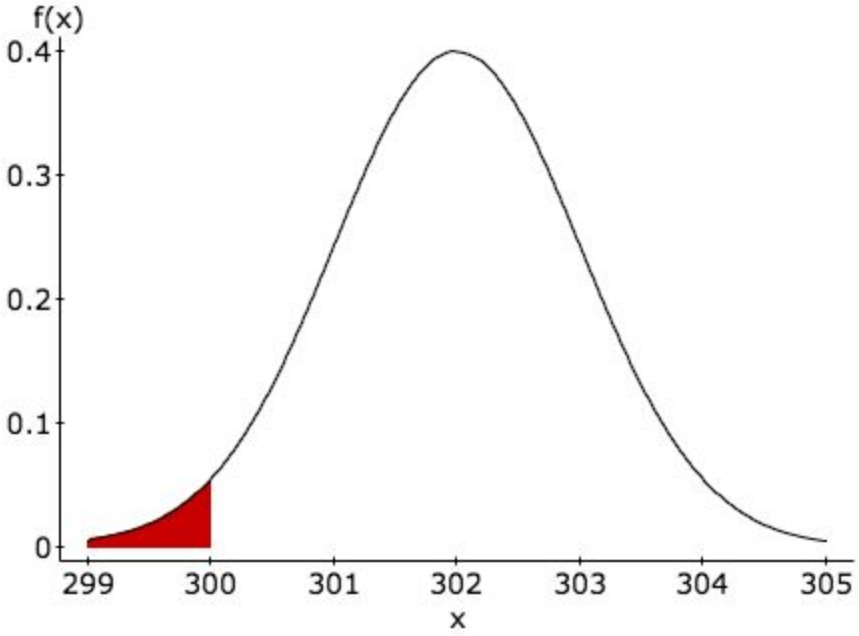
1a) Since the mean is the same, the standard deviation has no effect on the graph

b)



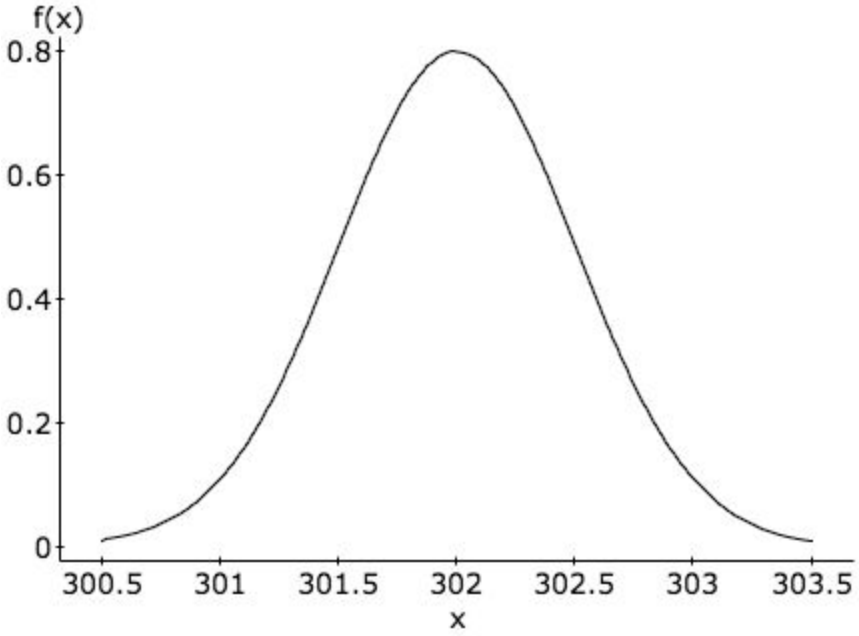
Normal Distribution
Mean:302 Std. Dev.:2
P(X≤300) = 0.15865525

The percentage of underfilled bottles containing less than 300ml when the mean is 300 ml and the standard deviation is 2ml is 15.87%



Normal Distribution
Mean:302 Std. Dev.:1
P(X≤300) = 0.02275013

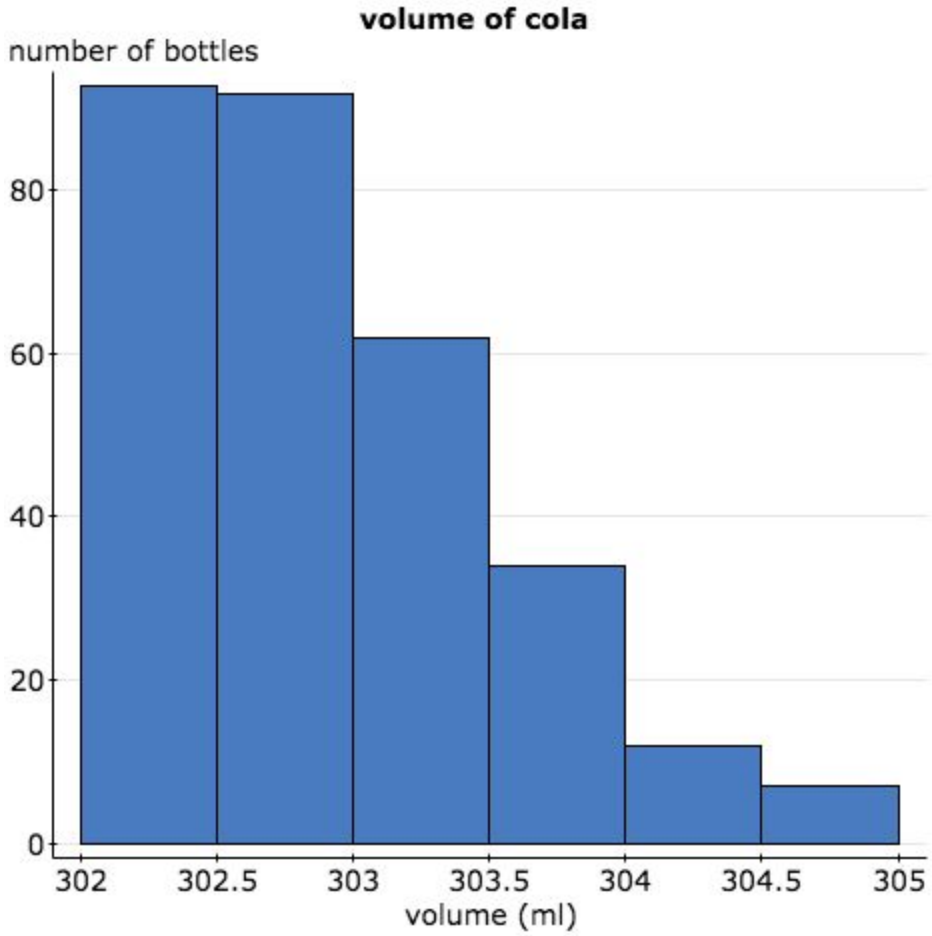
When the standard deviation is 1ml and the mean is 302ml, the percentage of underfilled bottles is 2.28%



Normal Distribution
Mean:302 Std. Dev.:0.5
P(X≤300) = 0.00003167

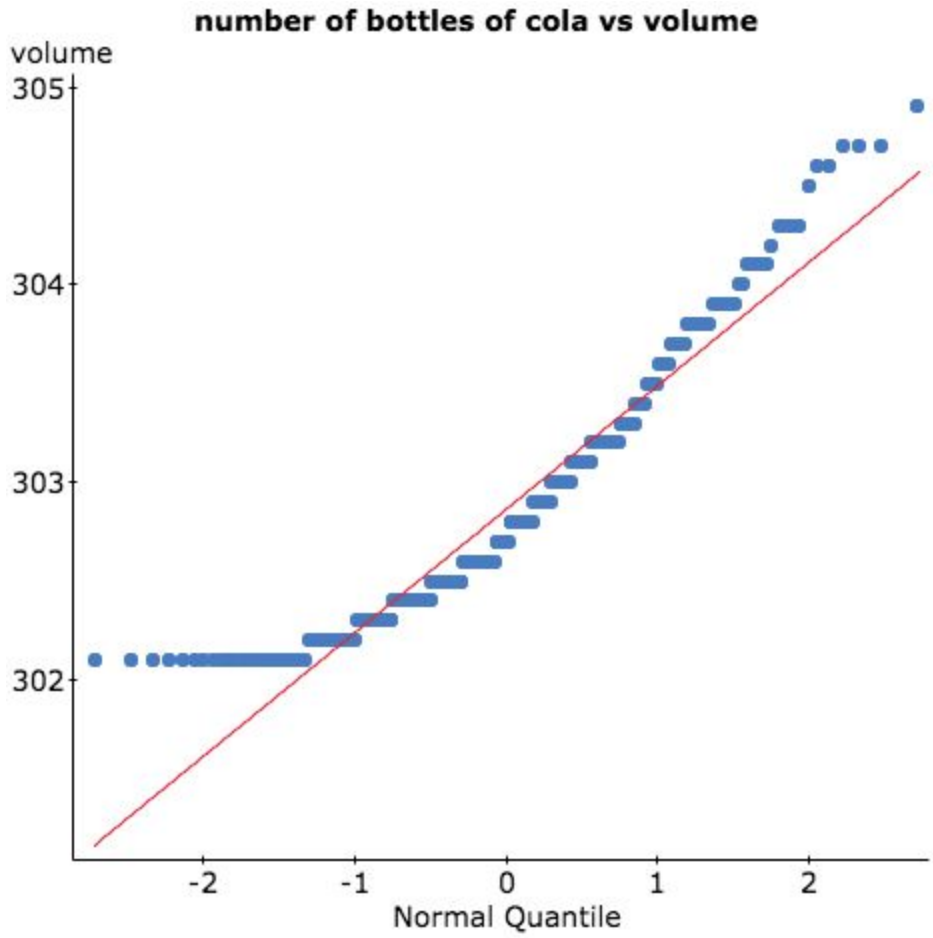
When the standard deviation is 0.5 ml and the mean is 302 ml, the percentage of underfilled bottles is 0. 003%

2a)



b) This graph is unimodal, has an outlier of 305 ml and is right skewed. The histogram supports the claim of the company that the bottles are overfilled.

c)



The shape of this graph supports the conclusion of the company that the bottles are overfilled. This is because the plots are arranged in a way that almost makes a straight line. The plots are wider on the tail and becomes more narrow towards the middle making us come to the conclusion that the bottles are overfilled.

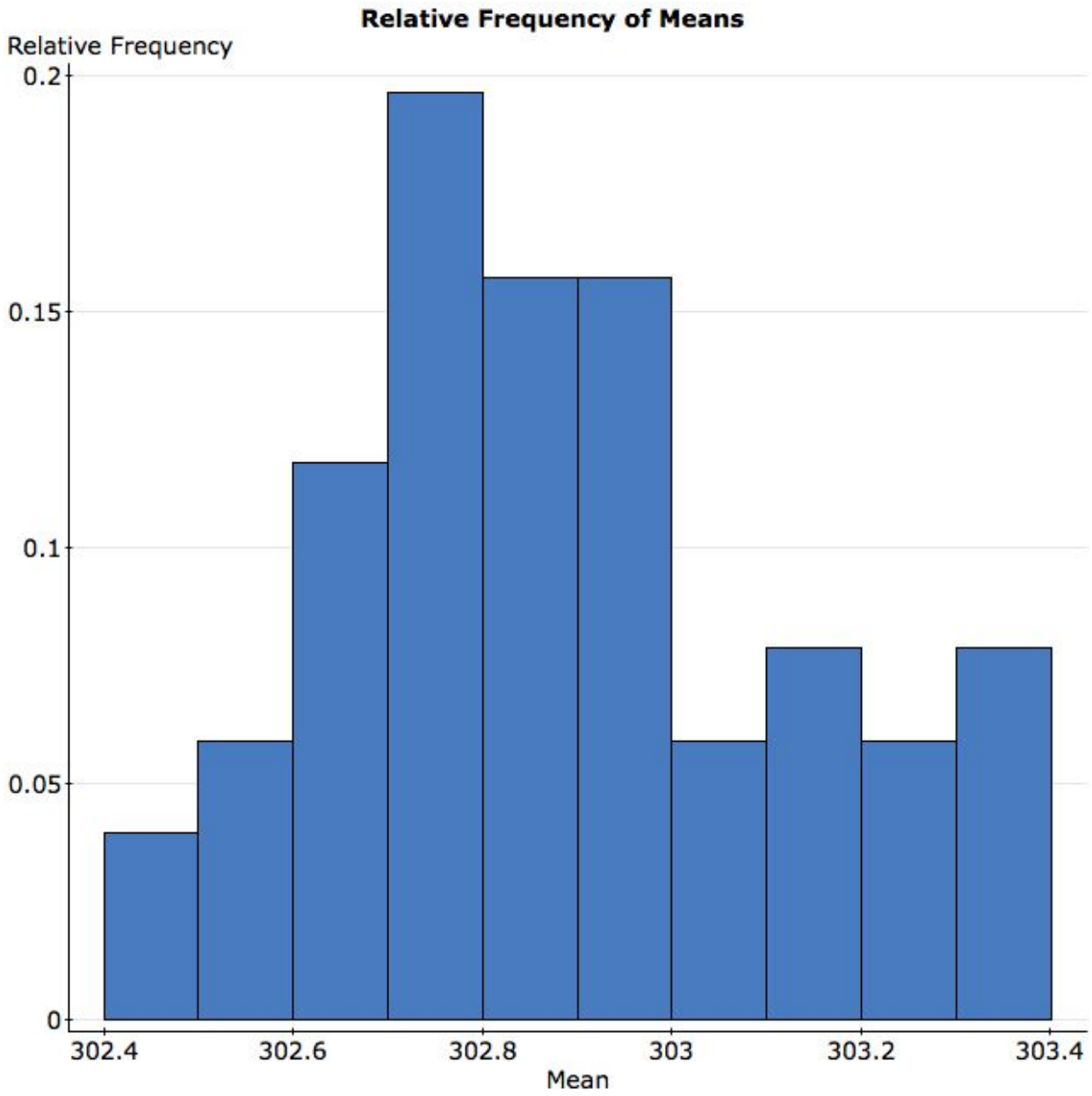
d)

Summary statistics:

Column	n	Mean	Variance	Std. dev.	Std. err.	Median	Range	Min	Max	Q1	Q3
volume	300	302.86567	0.39236243	0.6263884	0.036164551	302.7	2.8	302.1	304.9	302.4	303.2

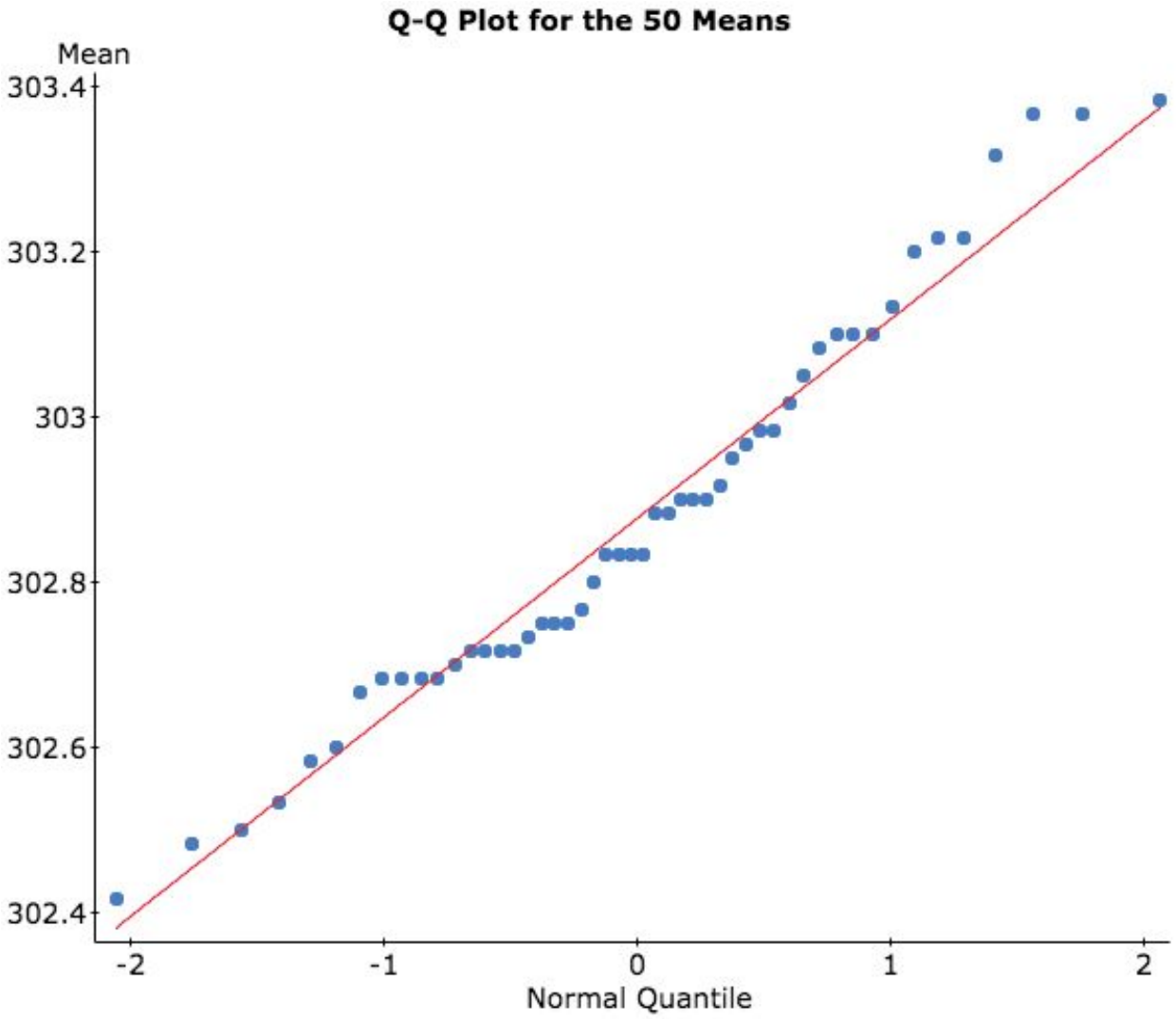
Since the relationship between the three quartiles are relatively the same, it supports the claim made in part b) in which the company overfills the bottles. The mean is greater than the median making this graph right skewed.

3 a)



b) The data shown in the histogram above is not normally distributed, the shape of this graph confirms that it is not normally distributed as it is not symmetrical about the mean of this data set. The data in question 2 is right-skewed and has a larger spread going from 302 mL to 305 mL, whereas in this histogram it is slightly right-skewed due to the greater amounts of samples collected on the right side and, has a smaller spread of only 1 mL.

c)



This plot proves the conclusions in part c), the points of the plot start on the left side, and curving more towards the right and then ending on the left side once more, demonstrating that it is right-skewed. Although it is not largely skewed, the data nonetheless has too many deviations from the centre line to have been produced by a normal distribution. The Q-Q plot in question 2c) is skewed as well and has heavier tails than the plot in 3c), shown by the more compressed line.

d)

Summary statistics:

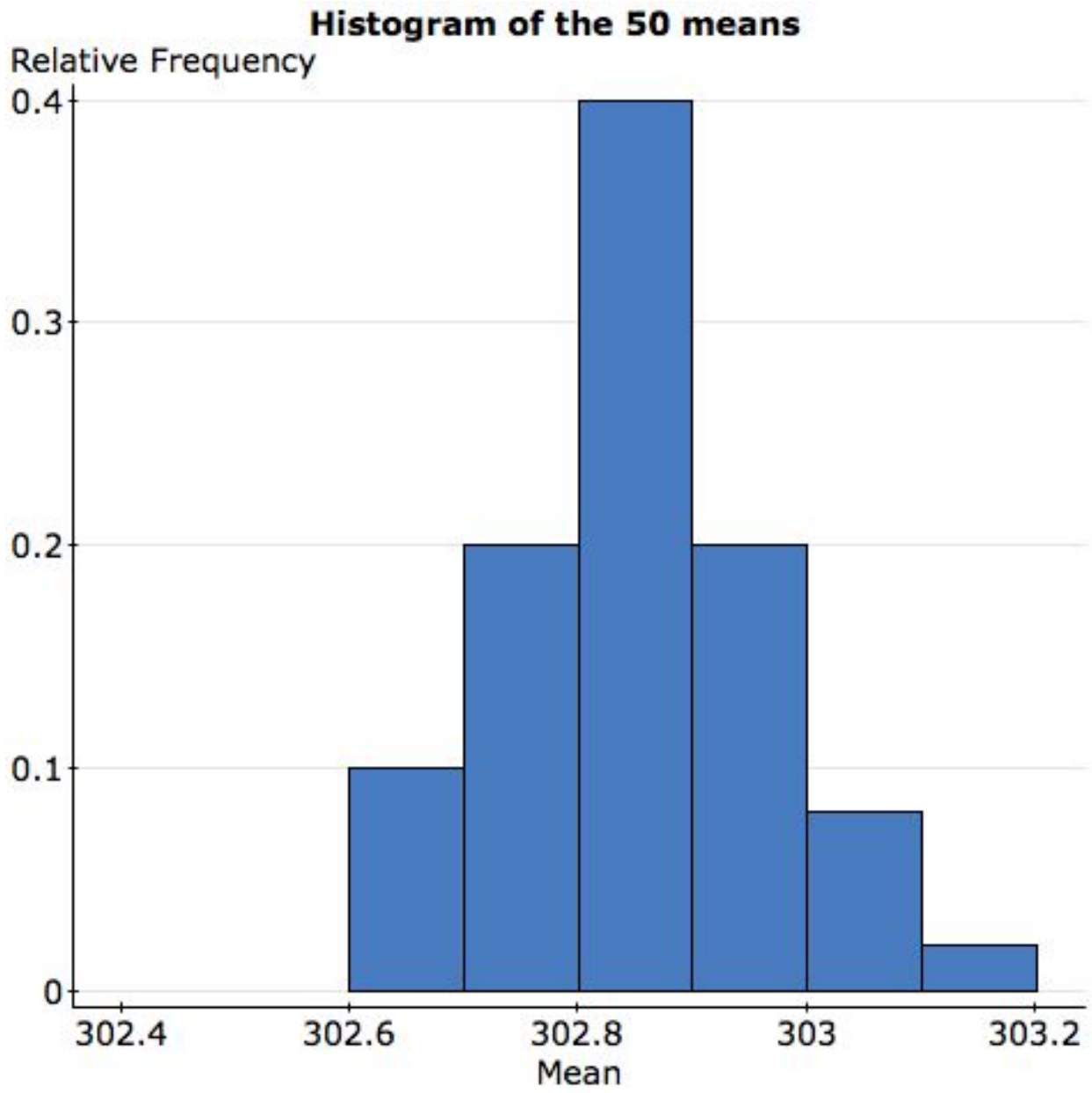
Column	Std. dev.	Mean
Mean	0.23764736	302.87743

The mean of the population is equal to the mean of the sampling distribution which is 302.87767. The standard deviation of the sampling distribution is equal to:

(standard deviation)/ n = (0.24005400)/(6) = 0.040009

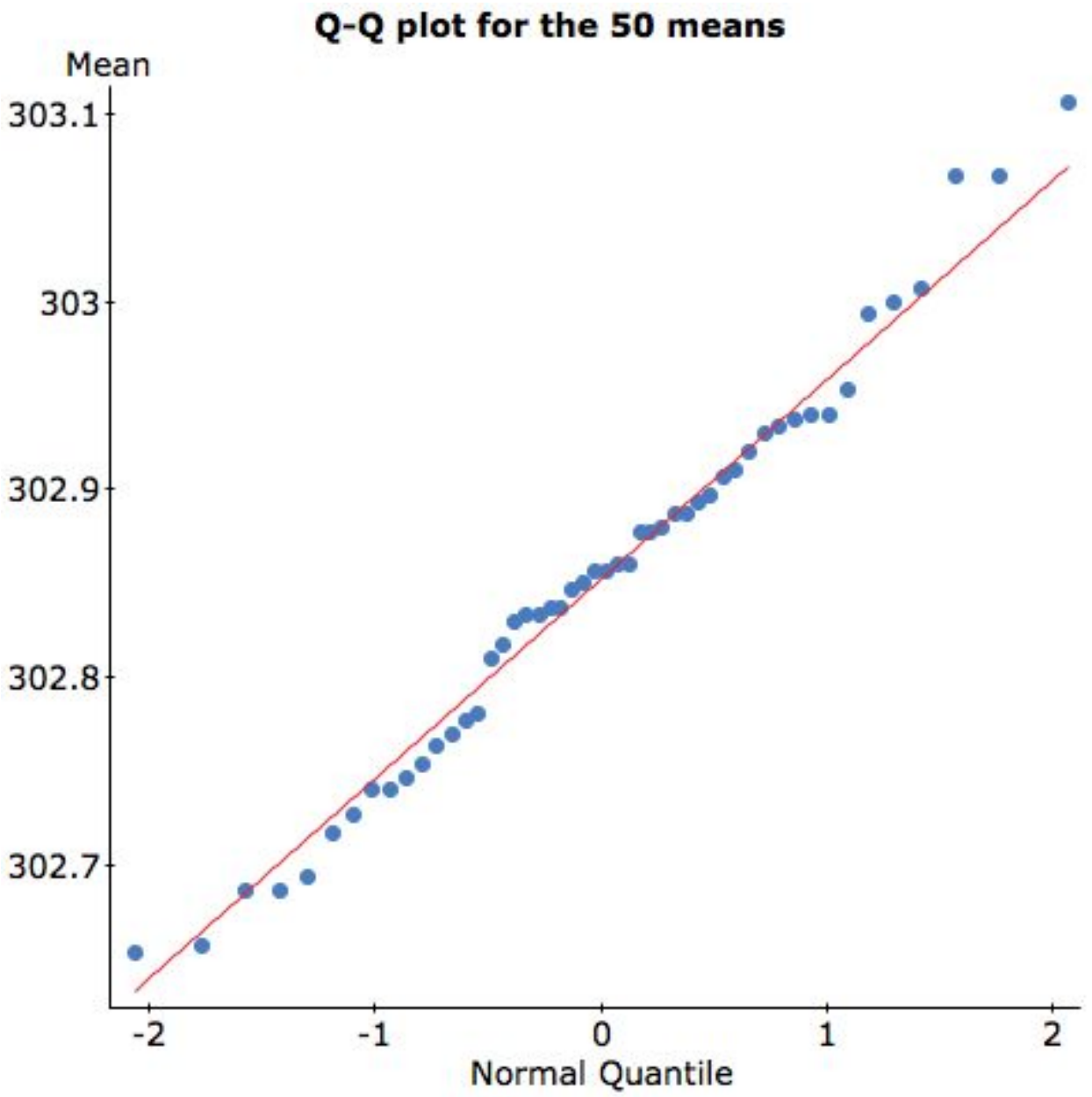
This represents the spread of the 6 bottles in each sample.

4 a)



b) The data shown in this histogram is more or less normally distributed, as the data is spread close to symmetrically on either side. In comparison to the graph in question 2, this histogram has virtually no skewness, because is right-skewed and it also has a smaller spread considering the data ranges from 302.6 mL to 303.2 mL whereas in question 2 it ranges from 302 mL to 305 mL. The histogram in question 3 has a somewhat larger spread (only by 0.4 mL) and despite the fact that it is slightly left-skewed, the skewness is more alike to the graph in question 4 as these graphs are more normally distributed compared to the others.

c)



This plot demonstrates that the points on the graph are mostly centralized along the 45° line and where they are dispersed they do not stray far from the line, meaning that the sample means come from a relatively normal distribution. Compared to the plot in question 3, it is clear that this data set has a more normal distribution and since the plot in question 3 has a smaller sample size it is easier for it to be more skewed.

d)

Summary statistics:

Column	Std. dev.	Mean

Mean	0.23764736	302.87743

The standard deviation for the data set in question 3 (n=6) is higher than the standard deviation for this data set (n=30). This concludes that the numbers from the data in this question are closer to the average and have less variation than those of question 3. The mean of the population is equal to the mean of the standard deviation which is 302.885247. The standard deviation of the sampling distribution is equal to:

(standard deviation)/ n = (0.1064492)/(30) = 0.00355

The theory of sampling distributions predicts the mean by use of the Central Limit Theorem which states that when n is sufficiently large, the sampling mean is well estimated by the normal curve, even when the population distribution is not normal, therefore the samples of n=30 are a more precise estimate of the population mean since the original data in question 2 was right-skewed.