Law of Large Numbers (LLN)

1. State the law of large numbers (see 5.1-5.4 notes). (p124)

LLN says the limit of the sample mean, as the sample size approaches infinity, is the expected value of X (which is LaTeX: μ).

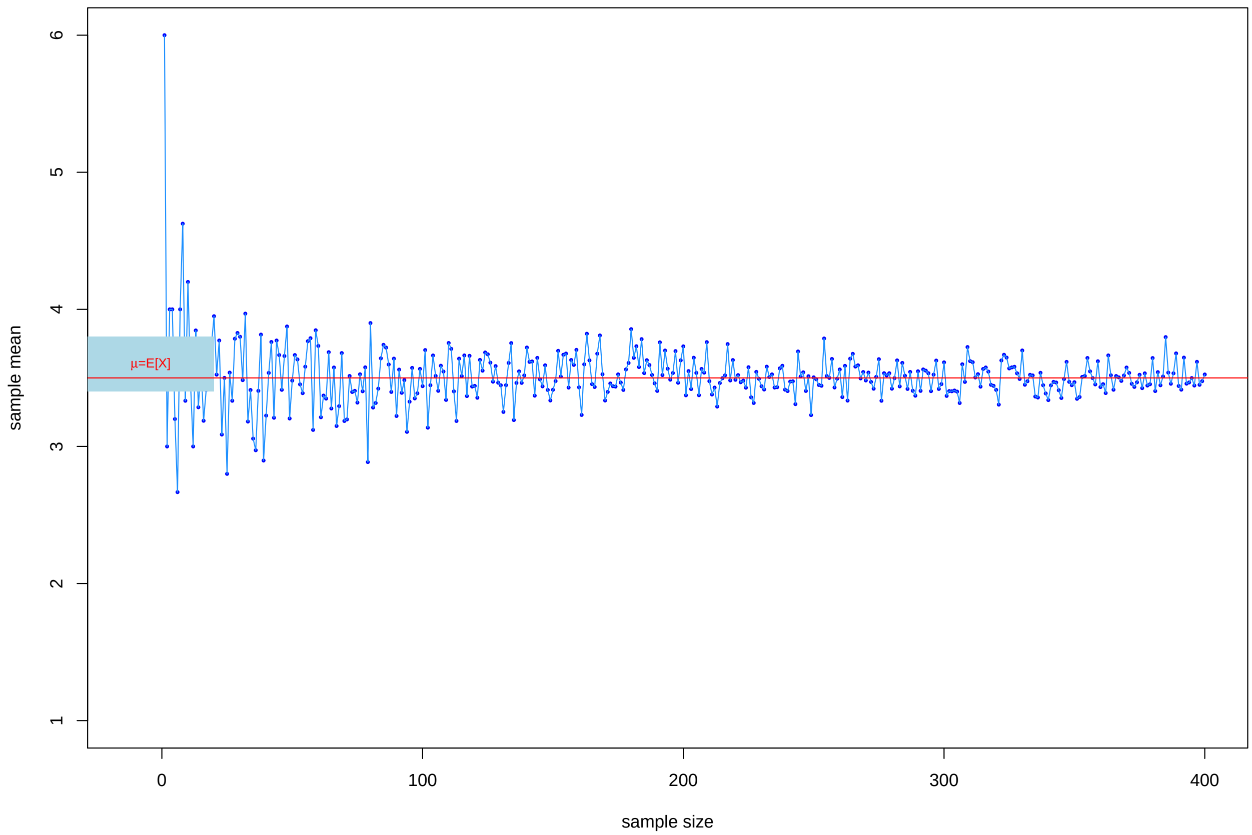
$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} x_i}{n} = \lim_{n\to\infty} \overline{x} = E[x] = \mu$$

Let's observe the LLN for a few distributions.

1a) What is the expected value of a fair 6-sided die? Include the plot in your lab sheet.

Because this is a normal distribution, by definition, the expected value is the mean.

$E[X] = \frac{1*1+1*2+1*3+1*4+1*5+1*6}{6} = 3.5$



1b) In your own words, describe what the "for" loop above does.

The for loop goes from 1 to 400 (N is 400), so the for loop executes 400 times.

It sets x to a random sample of values from 1 to 6, of size i (i depends on the iteration of the for loop)

It then stores the mean of this sample in the ith index of the means array.

Thus, in the end, the means array stores 400 sample means, ranging from size of 1 to 400, in indexes 1 to 400.

2. The exponential distribution with rate $\lambda = 0.25$.

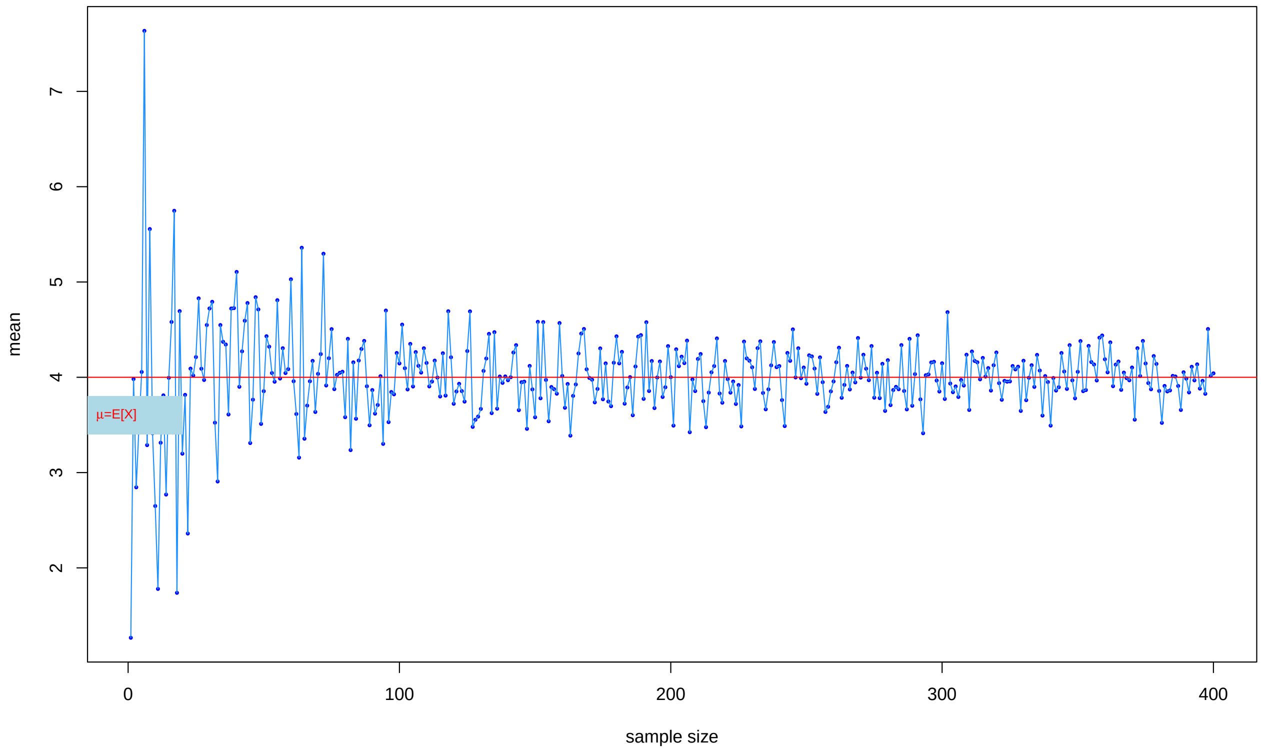
2a) What is the expected value X $\sim$ Exp($\lambda = 0.25$)?

$f(x) = \lambda e^{-\lambda x}$

$f(x) = 0.25e^{-0.25x}$

By definition, for an exponential distribution, $E[X] = \frac{1}{\lambda} = \frac{1}{0.25} = 4$

2b) Add a horizontal line at the expected value of X to your plot, and label the line with the text $\mu = E[X]$.



2c) In your own word, describe what this plot says. Use complete sentences.

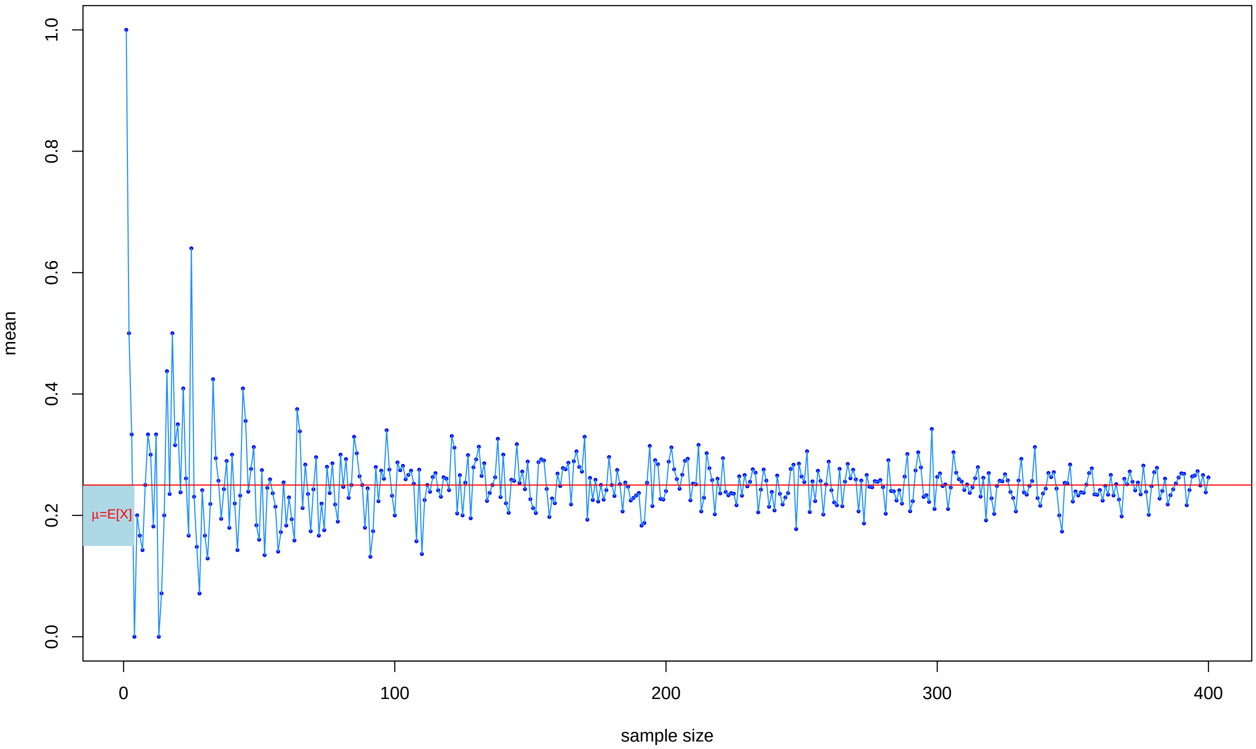
This plot states for the exponential distribution X $\sim$ Exp($\lambda = 0.25$), that the expected value will be 4. This plot displays the law of large numbers in effect, because we can see as the sample size gets larger, the actual values of the plot gets closer and closer to 4 which is the expected value (the mean).

3. The Poisson distribution with $\lambda = 0.25$.

3a) What is the expected value X $\sim$ Poisson($\lambda = 0.25$)?

By definition of the Poisson distribution, $E[X] = \lambda = 0.25$

3b) Adapt the code from the exponential distribution take samples of size 1 to 400 from the Poisson distribution, i.e. X $\sim$ Poisson(λ = 0.25), and graph the result. Include your code in this lab write up. In your own words, describe the plot and how it relates to LLN. Use complete sentences.



Expected Value of Poisson:

$E[x] = \lambda = 0.25$

This plot states for the Poisson distribution X $\sim$ Poisson(λ = 0.25), that the expected value will be 0.25. This is to be expected, as the expected value for a Poisson distribution is defined as λ. This plot displays the law of large numbers in effect, because we can see that as the sample size gets larger, the actual values of the plot gets closer and closer to 0.25 which is the expected value.

Central Limit Theorem (CLT)

CLT Part 1: Sampling distribution of the sample mean when the population is Normal.

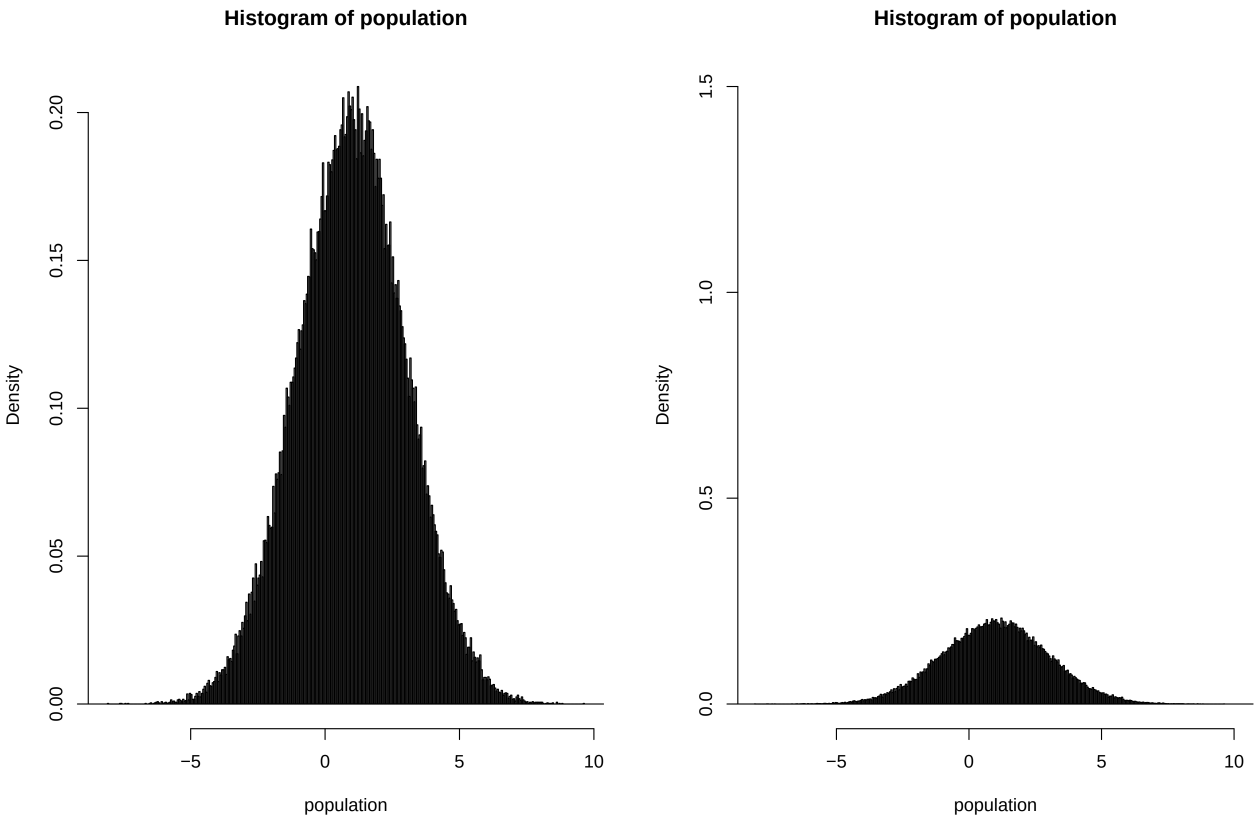
1. Sampling distribution of the sample mean when the population is Normal and n = 5.

1a) Report the population mean, standard deviation, the shape of the population distribution and include its graph.

```
## [1] "populationMean is  0.995511833370105"
```

```
## [1] "populationStdDev is  2.00703643399632"
```

The shape of the population distribution appears to be unimodal and normally distributed.



1b) The population size is 100,00. How many samples of size five are there?

samples $= \frac{\text{population size}}{\text{sample size}} = \frac{100{,}000}{5} = 20,000$

1c)

i. Report the mean and standard deviation of the empirical sampling distribution.

```
## [1] "sampleMean is  1.00424768559808"
```

```
## [1] "sampleStdDev is  0.890695969843915"
```

ii. Recall from class that $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$. Verify this is the case (at least approximately). Show your work.

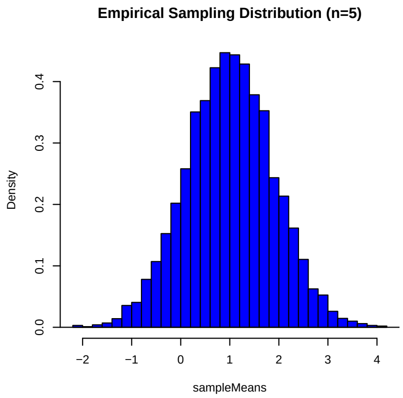
$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{5}}$

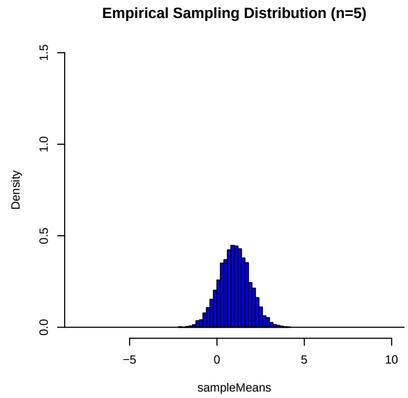
```
round(c(populationStDev / sqrt(sampleSize) ), 4)
```

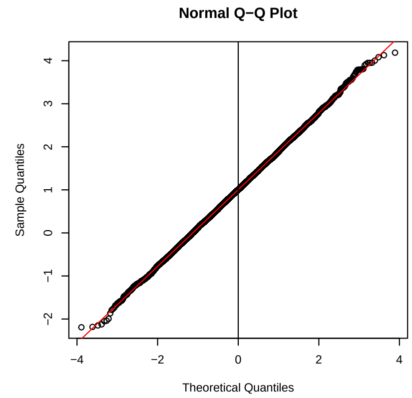
```
## [1] 0.8976
```

This is very close to the standard deviation of the empirical sampling distribution.

iii. Include the 3 graphs of the empirical sampling distribution in your lab sheet.







iv. What do you conclude about the shape of the empirical sampling distribution? Use complete sentences.

The empirical sampling distribution appears to be normally distributed, unimodal, with no skewedness. The center is around the sample Mean of 1.004, with a sample Standard Deviation of 0.8907.

2. Sampling distribution of the sample mean when the population is Normal and n = 10.

2a) Repeat the previous problem (all parts) when n = 10.

```
## [1] "sampleMean is  0.995981723894264"
```

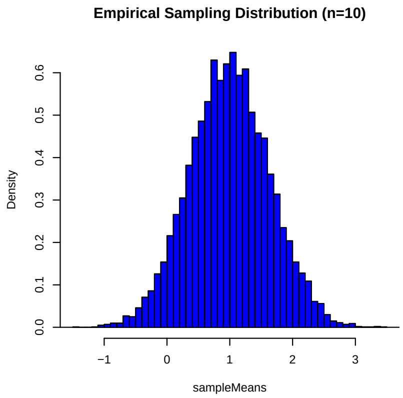
```
## [1] "sampleStdDev is  0.634065538299459"
```

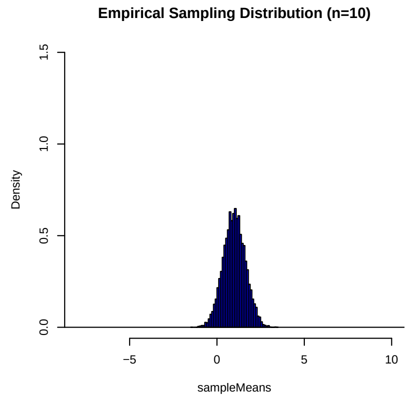
$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{10}}$

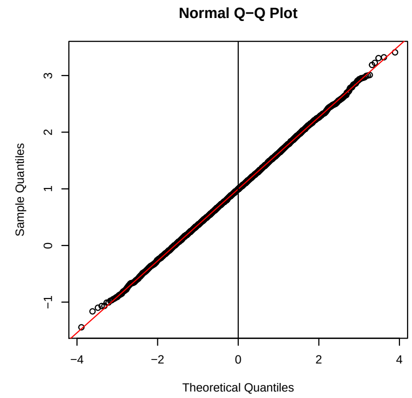
```
round(c(populationStdDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.6347
```

This is very close to the standard deviation of the empirical sampling distribution.







Because the population distribution is normal, we can conclude that the sampling distribution of the sample mean is also normal by the CLT. The center is around the sample Mean of 1.425, with a sample Standard Deviation of 0.981.

3. Sampling distribution of the sample mean when the population is Normal and n = 30.

3a) Repeat the previous problem when n = 30.

```
## [1] "sampleMean is  0.997825775977186"
```

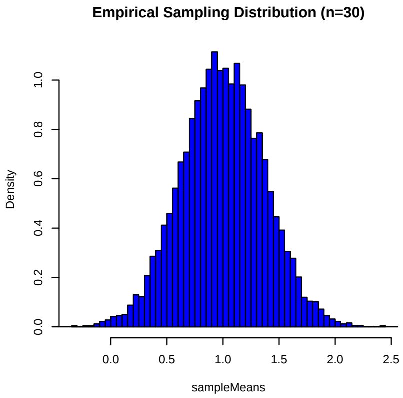
```
## [1] "sampleStdDev is  0.368589608322401"
```

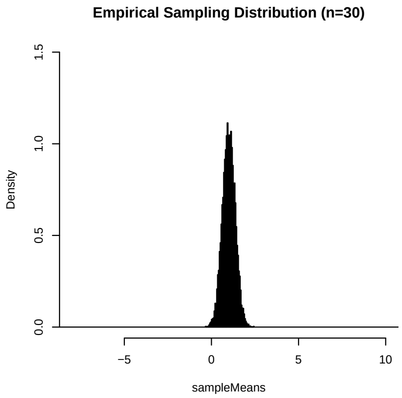
$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{30}}$

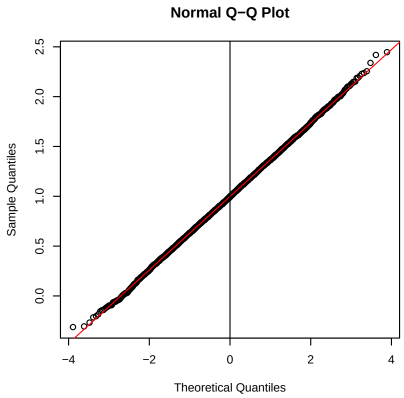
```r
round(c(populationStDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.3664
```

This is very close to the standard deviation of the empirical sampling distribution.







Because the population distribution is normal, we can conclude that the sampling distribution of the sample mean is also normal by the CLT. The center is around the sample Mean of 0.788, with a sample Standard Deviation of 0.951.

4. Sampling distribution of the sample mean when the population is Normal and n = 50.

4a) Repeat the previous problem when n = 50.

```
## [1] "sampleMean is  0.994952998631859"
```

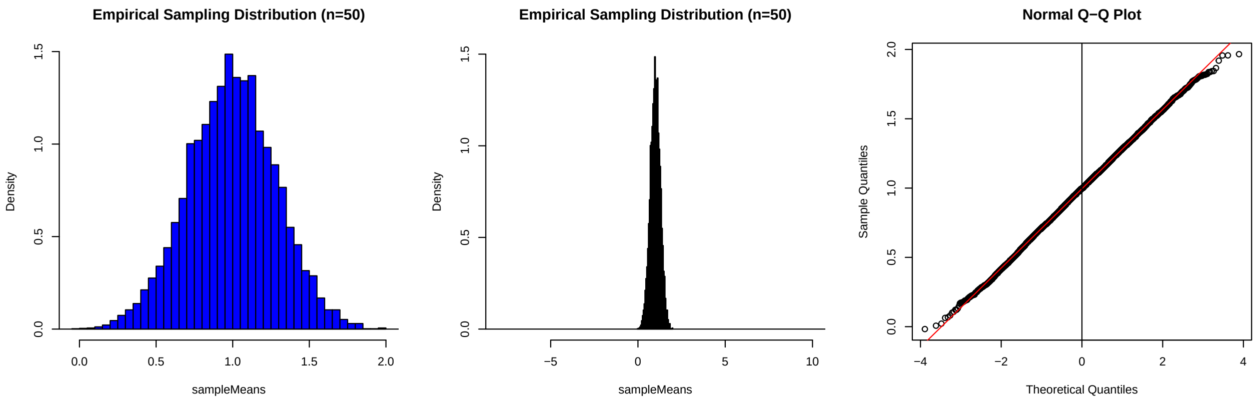
```
## [1] "sampleStdDev is  0.285862986951114"
```

$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{50}}$

```
round(c(populationStdDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.2838
```

This is very close to the standard deviation of the empirical sampling distribution.



Because the population distribution is normal, we can conclude that the sampling distribution of the sample mean is also normal by the CLT. The center is around the sample Mean of 0.921, with a sample Standard Deviation of 0.718.

5. The sampling distribution equals the population distribution when the sample size is 1.

5a) Describe what is happening to the histograms of the empirical sampling distribution n grows from 1 to 50. Compare the histograms that have the same x and y axis scales (this is why you made two histograms each time).

As n increases, the sampling distribution of the sampling mean reprents more of a normal distribution, and the range of values decreases (the range becomes smaller) while the height increases (becomes narrower and spikier). Furthermore, the sample mean and sample standard deviation approach the population mean and population standard deviation.

CLT Part 2: Sampling distribution of the sample mean when the population is non-Normal.

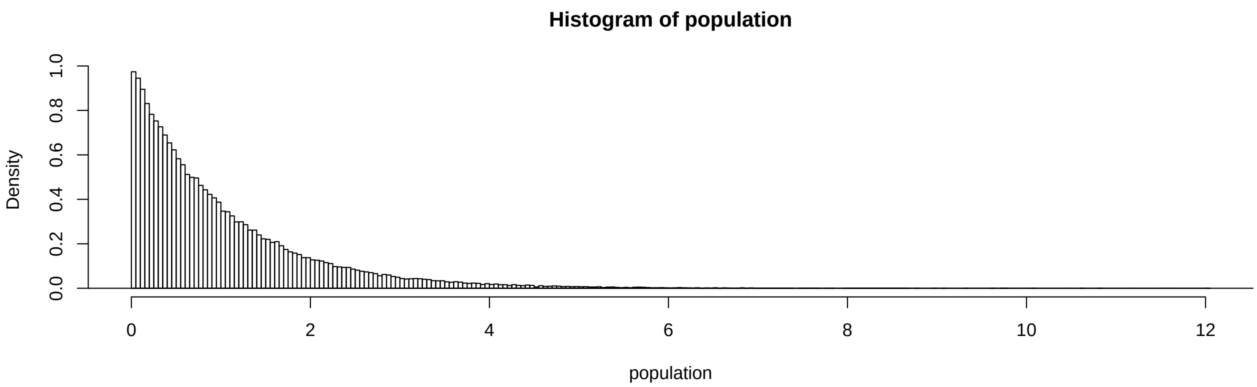
1.

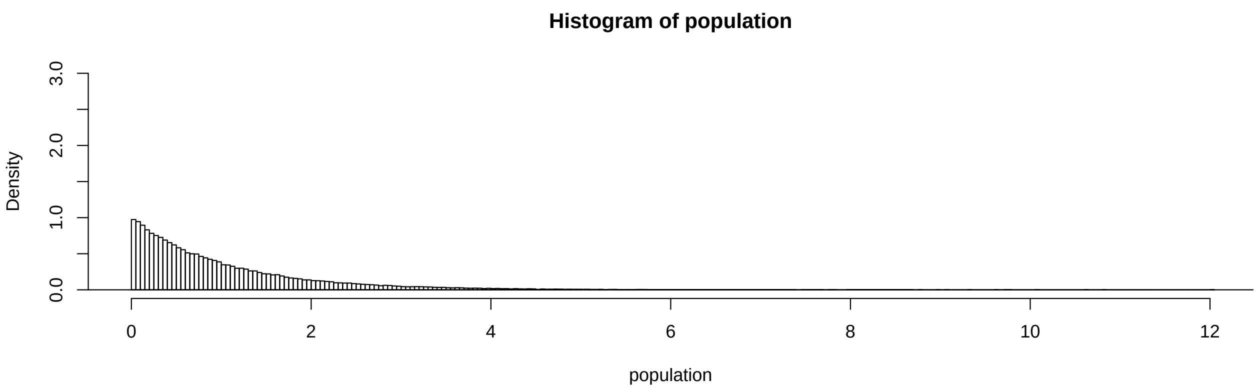
1a) Report the population mean, standard deviation, and the shape of the population distribution, and a graph.

```
## [1] "Population Mean is  1.00150229205488"
```

```
## [1] "Population StdDev is  1.00086419279743"
```

The shape of the population Distribution is unimodal and right skewed.





1b)

i. Report the mean and standard deviation of the empirical sampling distribution.

```
## [1] "Sample Mean is 0.997758992215806"
```

```
## [1] "Sample StdDev is 0.448046926974456"
```

ii. Recall from class that $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$. Verify this is the case (at least approximately). Show your work.

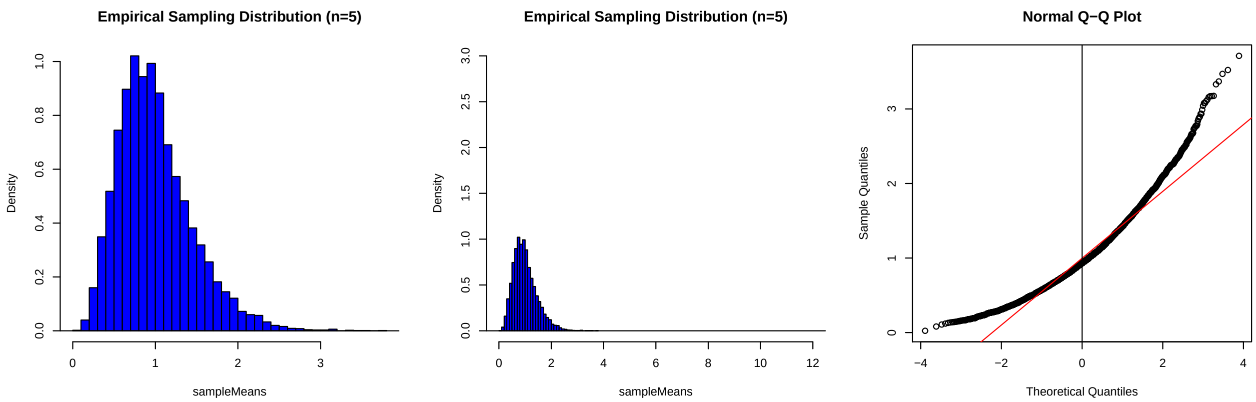
$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{50}}$

```r
round(c(populationStDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.4476
```

This is very close to the standard deviation of the empirical sampling distribution.

iii. Include the 3 graphs of the empirical sampling distribution in your lab sheet.



iv. What do you conclude about the shape of the enmpirical sampling distribution?

The population distribution is not normally distributed (it is right skewed). However, our sample size is 5, which is $\not\geq$ 30, so by the CLTFP we can say that the empirical sampling distribution is not approximately normally distributed. However by just looking at the plot, it looking fairly normally distributed, but is also right-skewed.

2. Sampling distribution of the sample mean when the population is Normal and n = 10.

2a) Repeat the previous problem when n = 10.

```
## [1] "Sample Mean is 0.996099697340699"
```

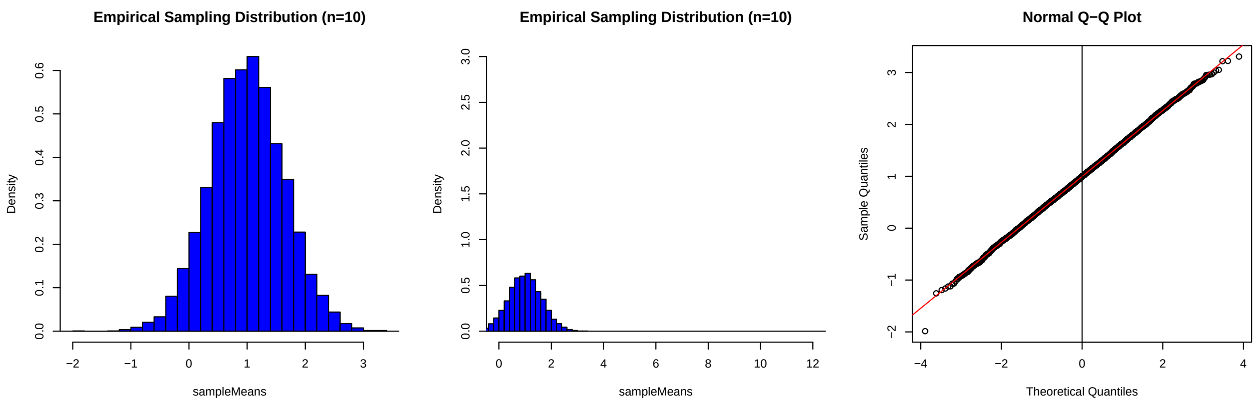
```
## [1] "Sample StdDev is 0.63507634115064"
```

$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{50}}$

```r
round(c(populationStDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.6347
```

This is very close to the standard deviation of the empirical sampling distribution.



The population distribution is not normally distributed (it is right skewed). However, our sample size is 10, which is $\ngeq$ 30, so by the CLTFP we can say that the empirical sampling distribution is not approximately normally distributed. However by just looking at the plot, it looking fairly normally distributed.

3. Sampling distribution of the sample mean when the population is Normal and n = 30.

3a) Repeat the previous problem when n = 30.

```
## [1] "Sample Mean is 0.997787922128622"
```

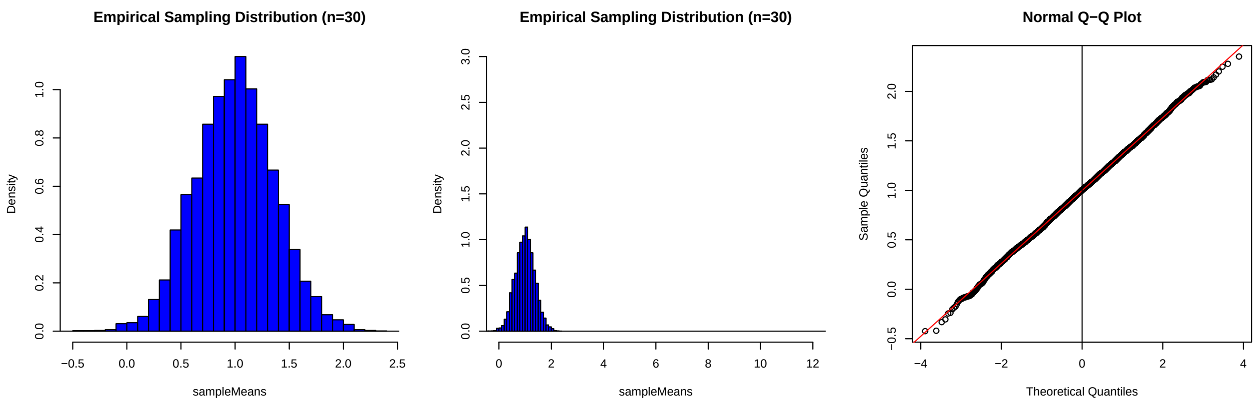
```
## [1] "Sample StdDev is 0.368317331522881"
```

$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{50}}$

```
round(c(populationStDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.3664
```

This is very close to the standard deviation of the empirical sampling distribution.



The population distribution is not normally distributed (it is right skewed). However, our sample size is 30, which is ≥ 30, so by the CLTFP we can say that the empirical sampling distribution is approximately normally distributed.

4. Sampling distribution of the sample mean when the population is Normal and n = 50.

4a) Repeat the previous problem when n = 50.

```
## [1] "Sample Mean is 0.994914323410688"
```

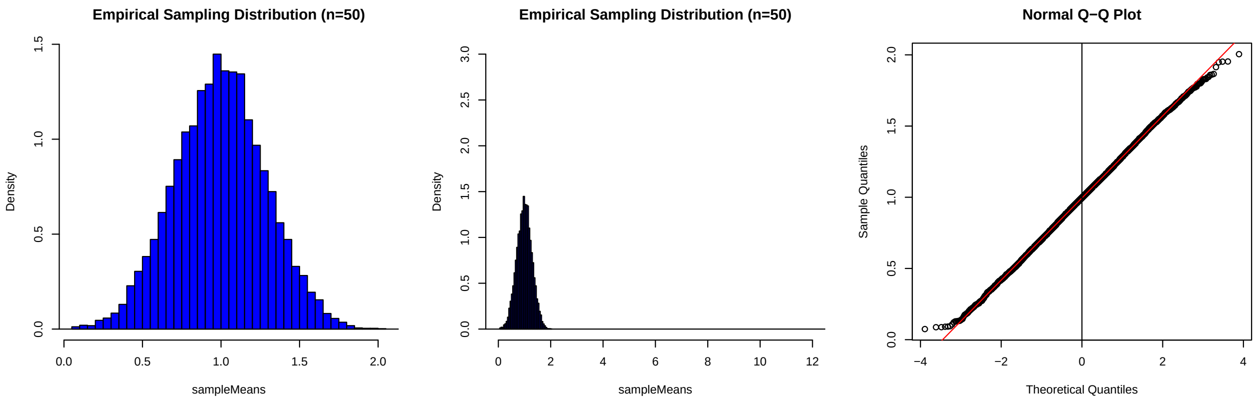
```
## [1] "Sample StdDev is 0.288388256192702"
```

$\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}} = \frac{2.00703643399632}{\sqrt{50}}$

```r
round(c(populationStDev / sqrt(sampleSize) ), 4)
```

```
## [1] 0.2838
```

This is very close to the standard deviation of the empirical sampling distribution.



The population distribution is not normally distributed (it is right skewed). However, our sample size is 50, which is ≥ 30, so by the CLTFP we can say that the empirical sampling distribution is approximately normally distributed.

5. The sampling distribution equals the population distribution when the sample size is 1. Describe what is happening to the histograms of the empirical sampling distribution n grows from 1 to 50. Compare the histograms that have the same x and y axis scales (this is why you made two histograms each time).

With a small sample size of 5, the empirical sampling distribution is slightly skewed and not approximately normally distributed. As the sample size increases (30), the empirical sampling distribution is fairly normally distributed. This gives us an experimental confirmation that the sampling distribution of the sample mean is normally distributed for any sample size, provided the original population is normally distributed. Also, observe that as the sample size increases, the mean of the sample means gets close to the population mean, and the standard deviations of the sample means gets close to the $\frac{\sigma}{\sqrt{n}}$.

CLT for the sample proportion.

1. Write down the expected value and standard deviation of the sample proportion in terms of the population proportion π and sample size n.

$\mu_p = \pi$

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

2. Write down the conditions that must be met in order for the sampling distribution of the sample proportion to be approximately normally distributed.

Use the CLTFP:

The sampling distribution of the sample proportion p is approximately Normally distributed provided than $n\pi \geq 5$ and $n(1-\pi) \geq 5$

3. Sampling distribution of the sample proportion when π = 0.5 and n = 100.

3a) Calculate nπ and $n(1-\pi)$.

$n\pi = 100 \cdot 0.5 = 50$

$n(1-\pi) = 100(1-0.5) = 100(0.5) = 50$

3b)

i. Report the mean and standard deviation of the empirical sampling distribution.

```
## [1] "Sample Mean is 0.500388"
```

```
## [1] "Sample StdDev is 0.0494060072395568"
```

ii. Is the mean what you expect? Explain.

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

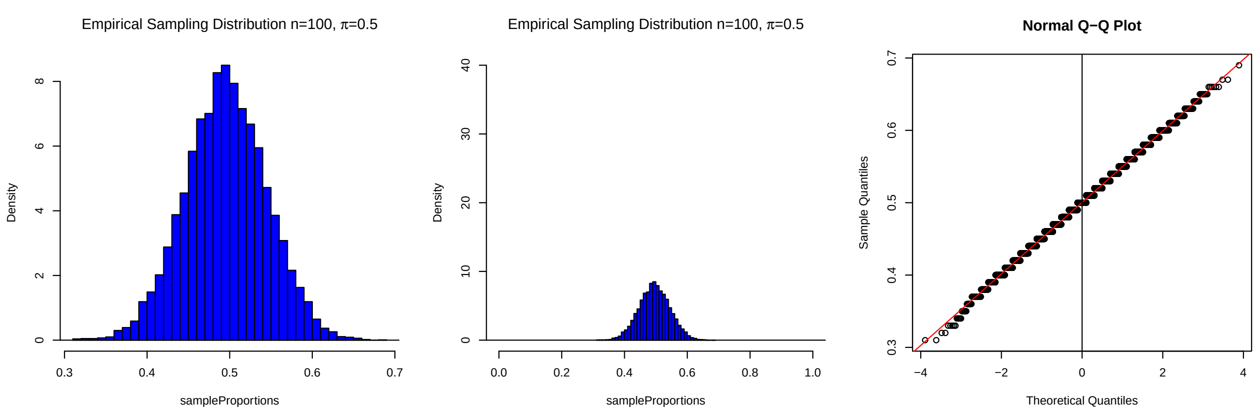
iii. Recall from class that $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

Verify this is the case (at least approximately). Show your work.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.50(1-0.50)}{100}} = \sqrt{\frac{0.50 \cdot 0.50}{100}} = 0.05$

Yes, this nearly exactly matches the sample std dev given from the example calculations.

iv. Include the 3 graphs of the empirical sampling distribution in your lab sheet.



v. What do you conclude about the shape of the empirical sampling distribution?

In part a), we found that both $n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

4. Sampling distribution of the sample proportion when π = 0.35 and n = 100.

4a) Repeat the previous question when π = 0.35 and n = 100.

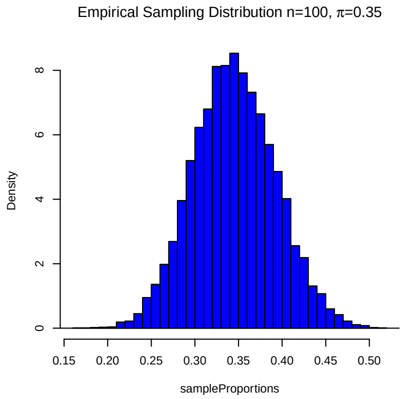
```
## [1] "Sample Mean is 0.350176"
```

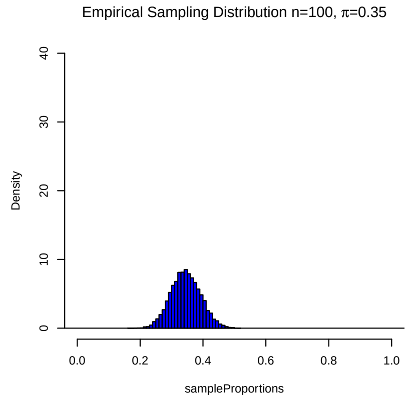
```
## [1] "Sample StdDev is 0.0472589924272251"
```

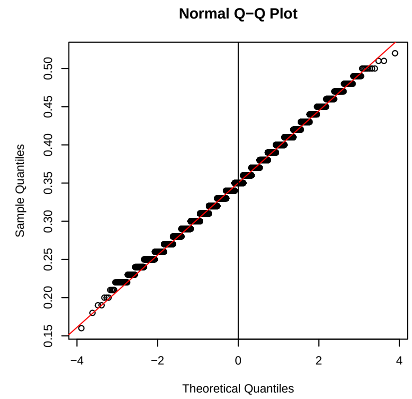
Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.35(1-0.35)}{100}} = \sqrt{\frac{0.35 \cdot 0.65}{100}} = 0.0477$

Yes, this nearly exactly matches the sample std dev given from the example calculations.







$n\pi = 100 \cdot 0.35 = 35 \geq 5$

$n(1-\pi) = 100(1-0.35) = 100(0.65) = 65 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

5. Sampling distribution of the sample proportion when $\pi = 0.65$ and n = 100.

5a) Repeat the previous question when $\pi = 0.65$ and n = 100.

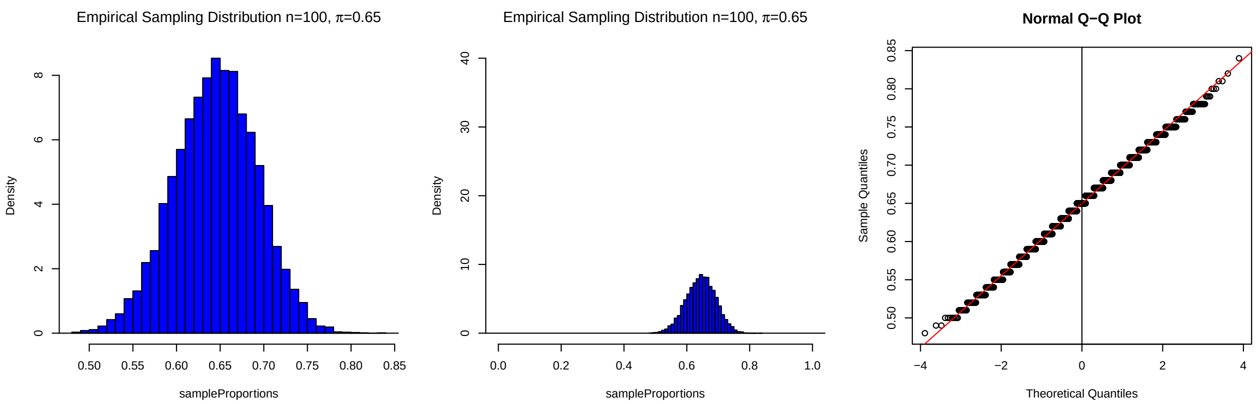
```
## [1] "Sample Mean is 0.649824"
```

```
## [1] "Sample StdDev is 0.0472589924272251"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.65(1-0.65)}{100}} = \sqrt{\frac{0.65 \cdot 0.35}{100}} = 0.0477$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.65 = 65 \geq 5$

$n(1 - \pi) = 100(1 - 0.65) = 100(0.35) = 35 \geq 5$

$n\pi$ and $n(1 - \pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

6. Sampling distribution of the sample proportion when π = 0.20 and n = 100.

6a) Repeat the previous question when π = 0.20 and n = 100.

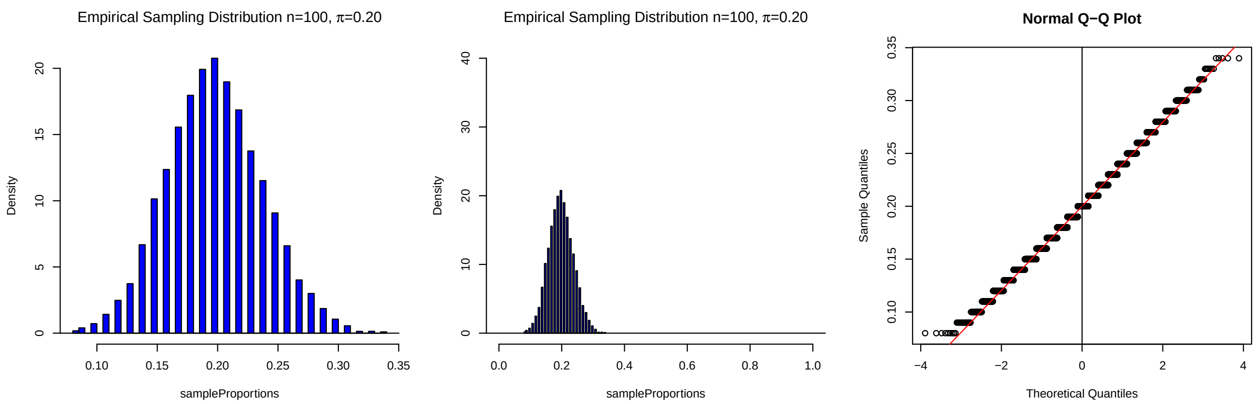
```
## [1] "Sample Mean is 0.200015"
```

```
## [1] "Sample StdDev is 0.0397691832175792"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.20(1-0.20)}{100}} = \sqrt{\frac{0.20 \cdot 0.80}{100}} = 0.04$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.20 = 20 \geq 5$

$n(1-\pi) = 100(1-0.20) = 100(0.80) = 80 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

7. Sampling distribution of the sample proportion when π = 0.80 and n = 100.

7a) Repeat the previous question when π = 0.80 and n = 100.

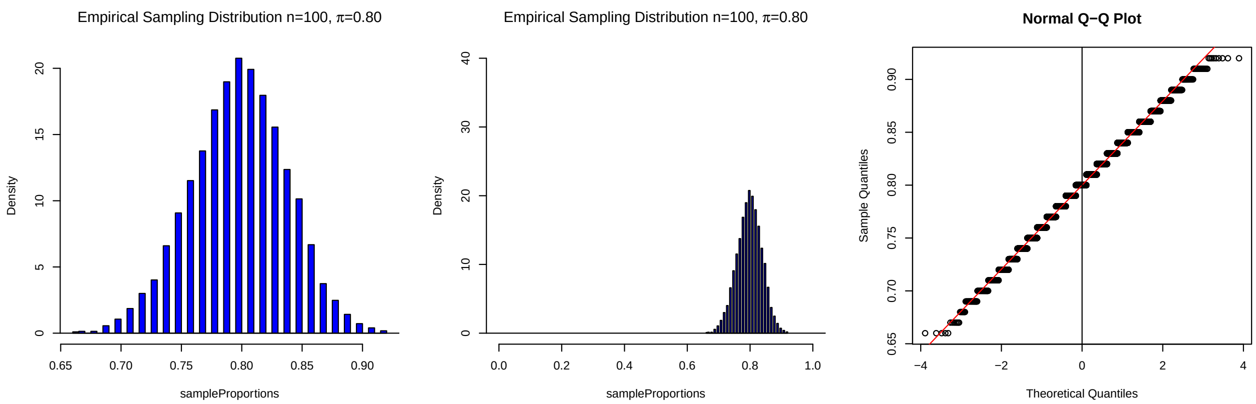
```
## [1] "Sample Mean is 0.799985"
```

```
## [1] "Sample StdDev is 0.0397691832175791"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.80(1-0.80)}{100}} = \sqrt{\frac{0.80 \cdot 0.20}{100}} = 0.04$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.80 = 80 \geq 5$

$n(1-\pi) = 100(1-0.80) = 100(0.20) = 20 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

8. Sampling distribution of the sample proportion when $\pi = 0.10$ and n = 100.

8a) Repeat the previous question when $\pi = 0.10$ and n = 100.

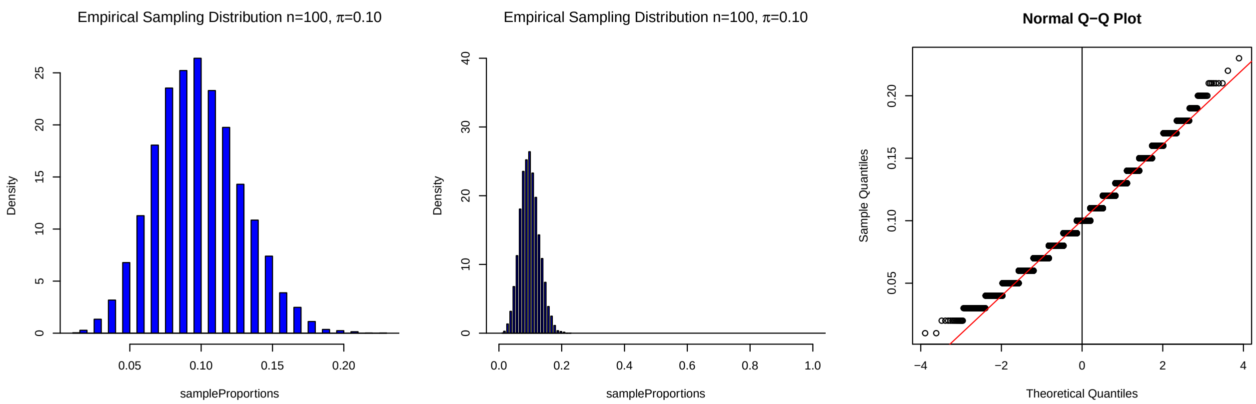
```
## [1] "Sample Mean is 0.100344"
```

```
## [1] "Sample StdDev is 0.0302670327803591"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.10(1-0.10)}{100}} = \sqrt{\frac{0.10 \cdot 0.90}{100}} = 0.03$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.10 = 10 \geq 5$

$n(1-\pi) = 100(1-0.10) = 100(0.90) = 90 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

9. Sampling distribution of the sample proportion when $\pi = 0.90$ and n = 100.

9a) Repeat the previous question when $\pi = 0.90$ and n = 100.

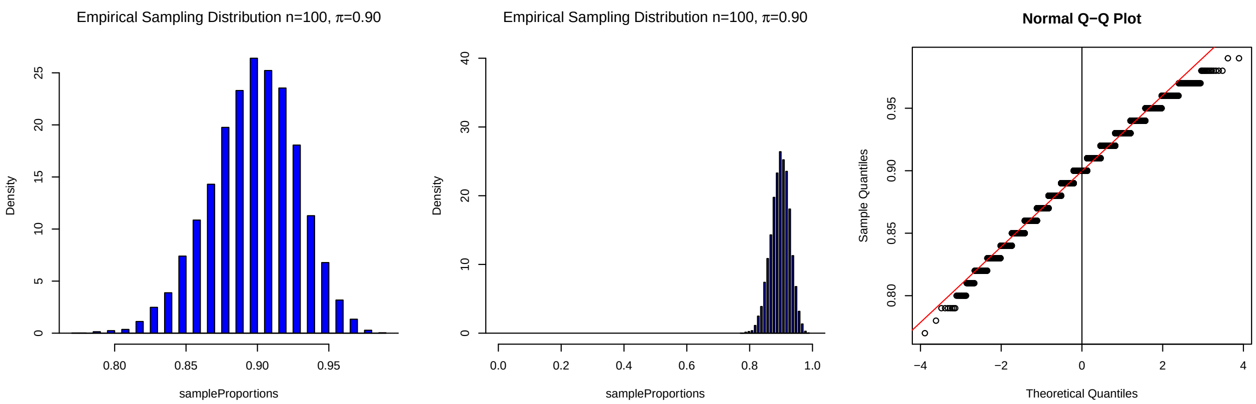
```
## [1] "Sample Mean is 0.899656"
```

```
## [1] "Sample StdDev is 0.0302670327803591"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.90(1-0.90)}{100}} = \sqrt{\frac{0.90 \cdot 0.10}{100}} = 0.03$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.90 = 90 \geq 5$

$n(1-\pi) = 100(1-0.90) = 100(0.10) = 10 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

10. Sampling distribution of the sample proportion when $\pi = 0.05$ and n = 100.

10a) Repeat the previous question when $\pi = 0.05$ and n = 100.

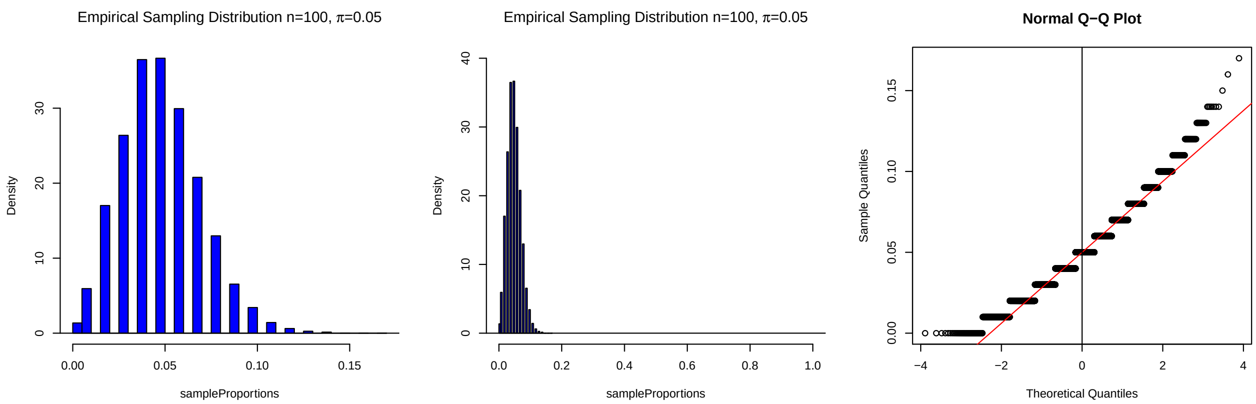
```
## [1] "Sample Mean is 0.049985"
```

```
## [1] "Sample StdDev is 0.021921629291252"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.05(1-0.95)}{100}} = \sqrt{\frac{0.05 \cdot 0.95}{100}} = 0.0218$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.05 = 5 \geq 5$

$n(1-\pi) = 100(1-0.05) = 100(0.95) = 95 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

11. Sampling distribution of the sample proportion when $\pi = 0.95$ and n = 100.

11a) Repeat the previous question when $\pi = 0.95$ and n = 100.

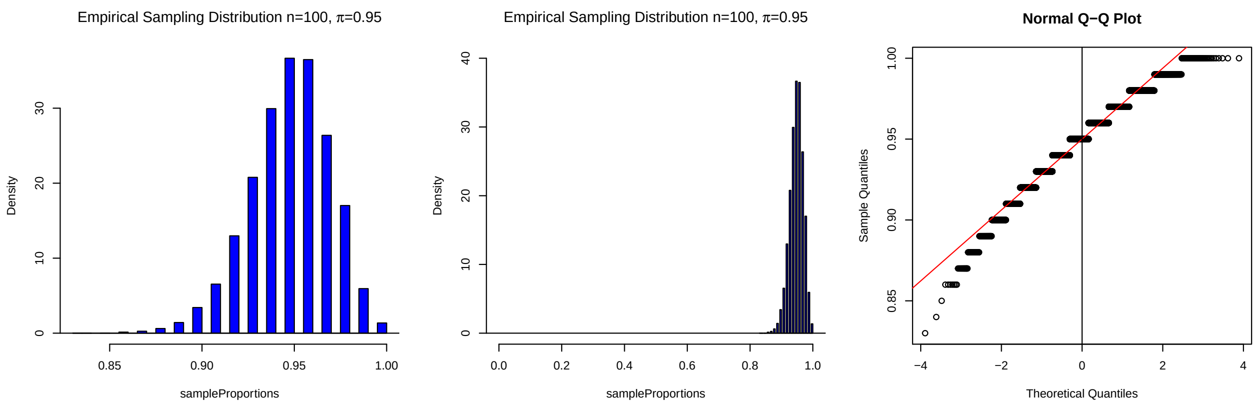
```
## [1] "Sample Mean is 0.950015"
```

```
## [1] "Sample StdDev is 0.021921629291252"
```

Yes, the sample mean should be close to the population mean, especially given that the sample population is 100. This is due to the law of large numbers, where the sample mean approaches the population mean under large sample sizes.

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.95(1-0.95)}{100}} = \sqrt{\frac{0.95 \cdot 0.05}{100}} = 0.0218$

Yes, this nearly exactly matches the sample std dev given from the example calculations.



$n\pi = 100 \cdot 0.95 = 95 \geq 5$

$n(1-\pi) = 100(1-0.95) = 100(0.05) = 5 \geq 5$

$n\pi$ and $n(1-\pi)$ are greater than or equal to 5. Therefore, by the CLTFP, we can conclude that the sampling distribution of the sample proportion is approximately normally distributed.

12. Compare the histograms of the empirical sampling distribution of the sample proportion for samples of size 100 as |0.5 - π| varies from 0.45 (the extremes) to 0 (when $\pi = 0.5$). This will be facilitated by the histograms you made that all have the same x-axis and y-axis axis scales.

The empirical sampling distribution for the sample proportion is very normally distributed when $\pi = 0.5$ (for any value of n). The empirical sampling distribution for the sample proportion remains fairly normally distributed when $\pi = 0.5$ for n = 100, provided π stays between 0.05 to 0.95. When π gets to either of these extremes, the sampling distribution exhibits more skew.

This is related to the cutoff values of $n\pi \geq 5$ and $n(1-\pi) \geq 5$

We can also see that for all of these simulations, the mean of the empirical sampling distribution of the sample proprotion is close to the population proportion π, and the standard deviation empirical sampling distribution of the sample proprotion is roughly $\pi(1-\pi)$.