Qiuying Li

Stat 424 Project 2

Oct. 22th, 2015

1. Select three sets of lognormal parameters to generate populations with

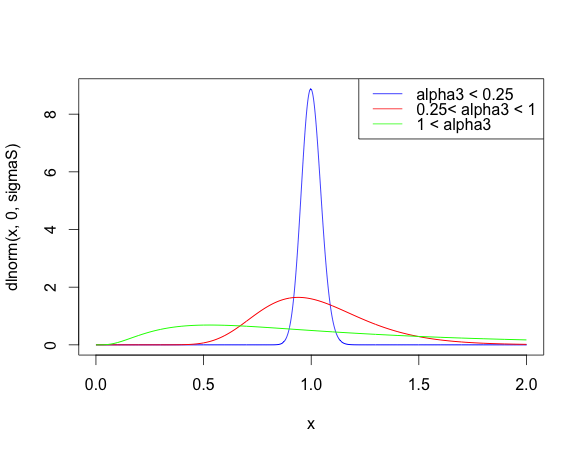
• Small skew coefficient , 𝛼3 < .25.

• Moderate skew coefficient , 0. 25 < 𝛼3 < 1.0.

• Large skew coefficient , 1. 0 < 𝛼3.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Small (<0.25) | Moderate (0.25<<1) | Large (1<) |
|  | 0.051 | 0.21 | 0.6 |
|  | 0.153 | 0.65 | 2.26 |

2. Create a plot of the population model. That is, a graph of the lognormal distributions above.



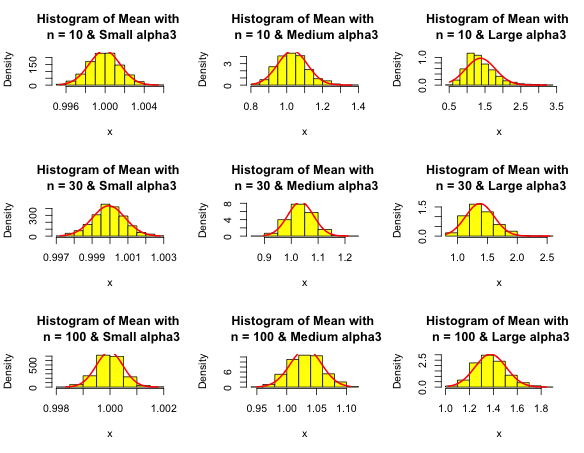
**Plot description:**

As we can see from the plot of population model above, when alpha is increasing, the respect population distribution is getting wider, but its tail is getting worse. For example, When 𝛼3 less than 0.25, its populaiton distribution is better than other two population distribution with larger 𝛼3. So the skew coefficient 𝛼3, is expected to be small to assume a normal population.

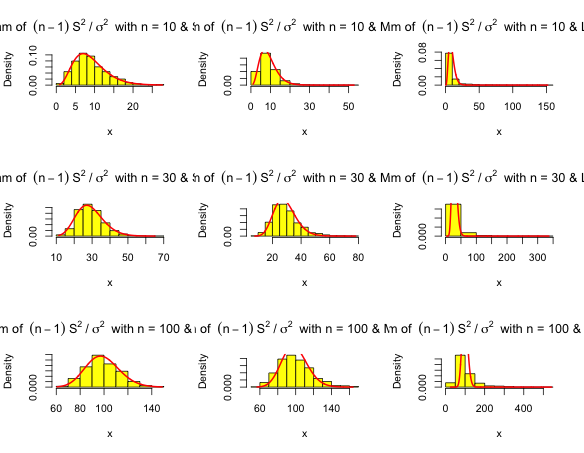
3.Calculate the simulated standard error by finding the standard deviation of your 1,000 means, and compare to the theoretical value of 𝜎 /√𝑛 .

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Small | | Moderate | | Large | |
| Simulated | Theoretical | Simulated | Theoretical | Simulated | Theoretical |
| n=10 | 0.016159 | 0.01617752 | 0.069056 | 0.068643 | 0.23561 | 0.24921 |
| n=30 | 0.009290 | 0.009329 | 0.03846 | 0.03963 | 0.1467 | 0.1438 |
| n=100 | 0.005206 | 0.005109 | 0.02063 | 0.021707 | 0.0806 | 0.07881 |
| From all the comparison above, we can see that theoretical values are very close to the simulated values. The result support the central limit theorem since in all the cases, the simulated standard error is similar to the theoretical value. And with larger sample size, the standard error, whatever the simulated or theoretical, is getting smaller. In a word, with large sample size, central limit theorem is true. | | | | | | |

• Create histograms and normal probability plots for the mean.



The 9 histograms above are all bell-shaped, which match the shape of the normal distribution. The 9 graphs above are the histograms and normal probability plots for the means of different sample size with the three levels of sigma. Compare the all the nine histograms; the means are following with the theoretical distributions, so it proves the model of Central Limit Theorem is true.

 Create histograms and Chi-square probability plots for the variance.

These 9 plot above are the histograms and Chi-square probability plots for the variance.

Although 9 histograms are a little bit different from each other, however, all the 9 histograms follow the shape of the chi-square distribution. In addition, when we compare the 9 histograms, we can find out that the variances are following with the theoretical distributions. There may have some slightly imperfect matching, which may because the sample is generated from a random sample, so it may not perfectly fit with chi-square distribution. So it proves the model of Central Limit Theorem is true

Generate 95% confidence intervals for the population mean, and count the number that cover the mean

|  |  |  |  |
| --- | --- | --- | --- |
|  | Small | Moderate | Large |
| n=10 | 947 | 950 | 945 |
| n=30 | 945 | 959 | 959 |
| n=100 | 966 | 952 | 946 |

The meaning of the 95% confidence interval means that we have 95% confidence to conclude that the true mean is within the range. The 9 results above show that we were correct with the model since numbers in the 9 cases that cover the mean are close to 950. Since in each case, we have 1000 intervals for the mean, and we are 95% confidence that those intervals would contain the true mean. So we should report the number be close to 950. In a word, the 95% confidence interval for the population mean is correct in the cases above, and we can show the validation of the central limit theorem.

Generate 95% confidence intervals for the population standard deviation, and count the number that cover the standard deviation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Small | Moderate | Large |
| n=10 | 893 | 850 | 581 |
| n=30 | 920 | 866 | 512 |
| n=100 | 948 | 899 | 491 |

For the confidence interval about standard deviation, we can see that none of the cases are above the count 950, it’s not “perfect” like it for the population mean that the numbers are not close to 950. However, for small , we can see that the confidence interval is from 90%-95%, and the result of is better than the moderate and large .So with the small and larger sample size, the result of 95% confidence interval is more accurate.

Conclusion:

In conclusion, when the skew coefficient is small, we could assume the distribution of population is normal. And with a large sample, (at least n=100), we could have more evidence to suggest that the Central Limit Theorem is valid.