# Lab 7 – distributions of sample means

In this lab, we are going to be investigating the distributions of **means** of samples of populations of various different distributions. For instance, suppose we take the means and standard deviations of **n** samples of size **m**, and then looked at the distribution of those **n** sample means. We will be adapting yourfunctions from Lab 4 and using some of the commands you learned in Lab 6 for this purpose.

**To submit: answers to all numbered questions. When the question asks you to write code or create graphs, submit the code and/or graphs in the Word document as part of your answer. Also submit a single .R file that contains all of your code.**

## Distribution of sample means in a uniformly distributed population

We are going to begin by looking at the distribution of sample means when the underlying population is uniform. A good example of such a population is die rolls: if you roll a fair die many times, the results 1, 2, 3, 4, 5, and 6 should occur with roughly equal frequency. (You saw this result in Lab 4.) We will now look at what happens when we take many samples of size **m** from a uniformly-distributed population and look at the means and standard deviations of those means. We can imagine this as rolling **m** dice many times, and looking at the means and standard deviations of the results of those dice.

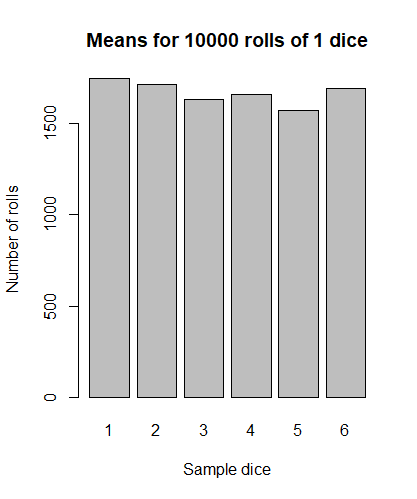
1. Get out the script you created in Lab 4. Save this script under a new name. Create a new function called **DiceMeans(n,m)** that will be based on your **RollSomeDice(n,m)** function. Your **DiceMeans(n,m)** function will simulate rolling **m** dice **n** times, and will compute (and keep track of) the sample mean for each of those **n** die rolls.Your function should return the following:

* The mean of the **n** sample means
* The standard deviation of the **n** sample means
* A bar plot of the **n** sample means

Note: The title of your bar plot should be “Means for **n** rolls of **m** dice”, where **n** and **m** are the actual values of **n** and **m** you entered. Choose your axis labels appropriately.

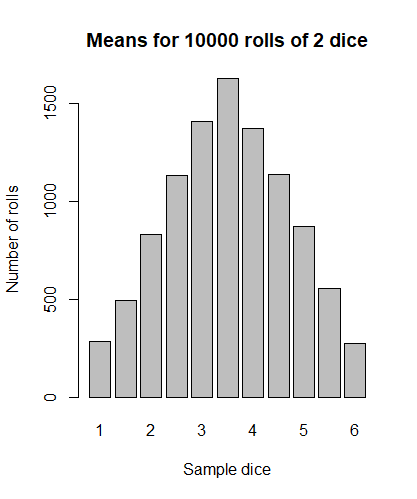
Run your function for **n=10000** and for the following values of **m**: 1 (ie, roll a single die 10000 times), 2 (roll 2 dice 10000 times and generate a list of the 10000 sums), 10, 50, 100.

Submit five outputs (bar plots, with the corresponding means and standard deviations pasted below each plot). How do the means and standard deviations compare as **m** increases? How do the shapes of the graphs compare?



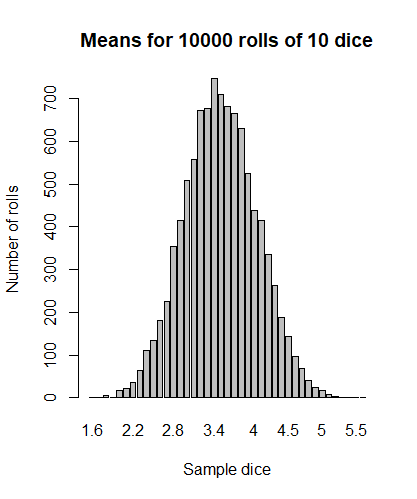
> DiceMeans(10000, 1)

[1] 3.466200 1.721846

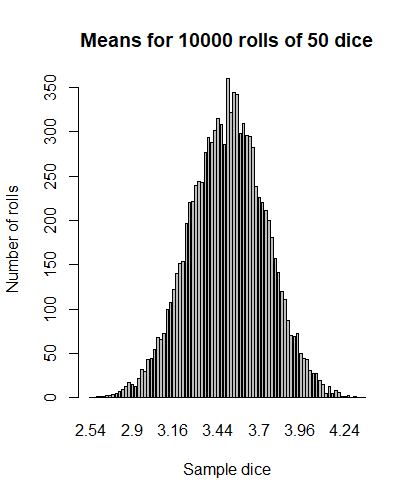


> DiceMeans(10000, 2)

[1] 3.514300 1.205463

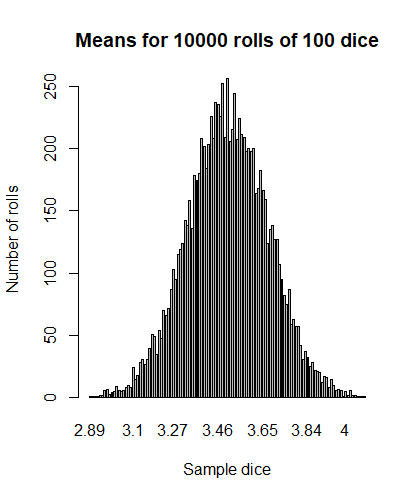


[1] 3.4974300 0.5426895



> DiceMeans(10000, 50)

[1] 3.5005860 0.2421192



> DiceMeans(10000, 100)

[1] 3.5010930 0.1723078

The more dice we have the more bell-shaped the graph is and more normalized distribution becomes.

**NOTE: WHILE THIS LAB IS DUE ONE WEEK FROM NOW, WE WILL BE DISCUSSING THE RESULTS OF THIS QUESTION DURING OUR NEXT LECTURE. HAVE YOUR RESULT FROM THIS QUESTION HANDY DURING LECTURE!**

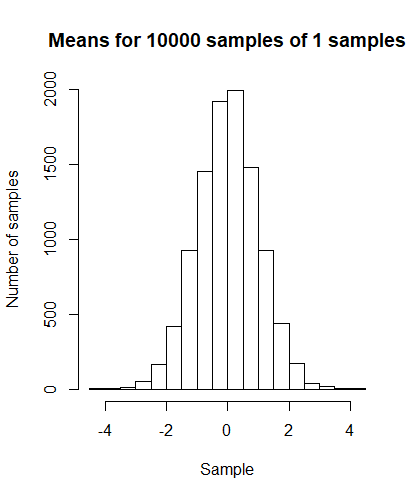
## Distribution of sample means in a normally distributed population

1. Create a function called **NormalMeans(n,m)** that simulates taking **n** samples of size **m** from normally-distributed populations with mean 0 and standard deviation 1. (You may have to revisit Lab 6 to simulate sampling from normally-distributed populations.) Like the function in Question 1, this new function will compute (and keep track of) the sample mean for each of those **n** samples.Your function should return the following:

* The mean of the **n** sample means
* The standard deviation of the **n** sample means
* An appropriately-named **histogram** of the **n** sample means (use the default bins)

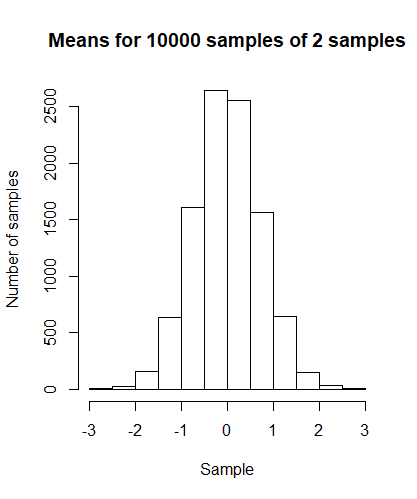
Run your function for **n=10000** and for the following values of **m**: 1, 2, 10, 50, 100.

Submit five outputs (five histograms, each with the means and standard deviations listed below). How do the means and standard deviations compare as **m** increases? How do the shapes of the graphs compare to one another, and to the graphs you obtained in Question 1?



> NormalMeans(10000, 1)

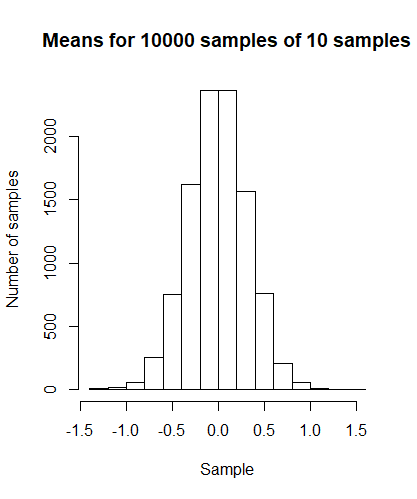
[1] 0.009724066 0.995789419



> NormalMeans(10000, 2)

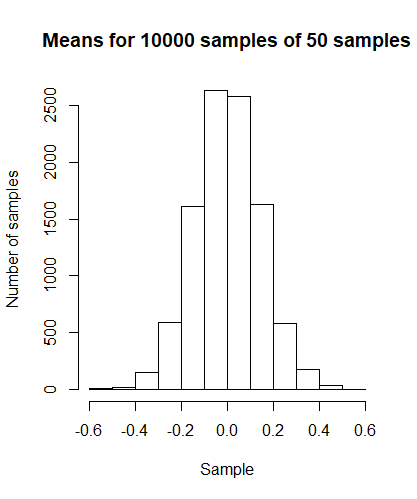
[1] -0.003381351 0.713395056

## Distribution of sample means in an exponentially-distributed population



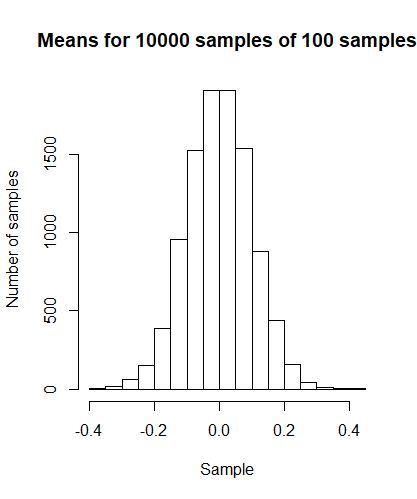
> NormalMeans(10000, 10)

[1] -0.004332744 0.317877168



> NormalMeans(10000, 50)

[1] 0.001213543 0.141272591



> NormalMeans(10000, 100)

[1] -0.0007143069 0.0994950088

When m increases, standard deviation becomes smaller, and mean approaches 0.

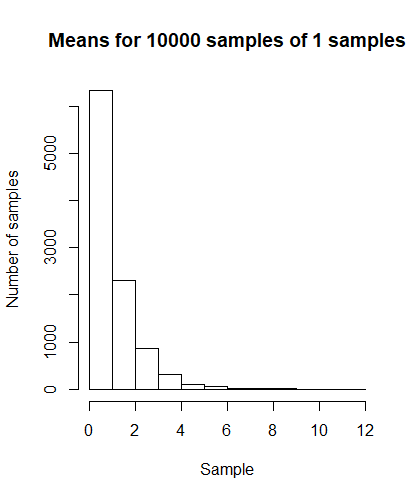
1. Create a function called **ExponentialMeans(n,m)** that simulates taking **n** samples of size **m** from exponentially-distributed populations with rate 1. As before, this new function will compute (and keep track of) the sample mean for each of those **n** samples.Your function should return the following:

* The mean of the **n** sample means
* The standard deviation of the **n** sample means
* An appropriately-named **histogram** of the **n** sample means (use the default bins)

Run your function for **n=10000** and for the following values of **m**: 1, 2, 10, 50, 100.

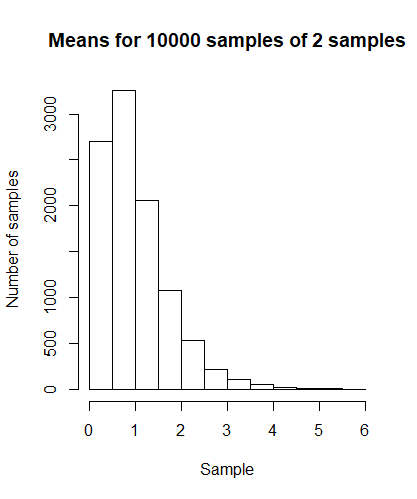
Submit five outputs (five histograms, each with the means and standard deviations listed below). How do the means and standard deviations compare as **m** increases? How do the shapes of the graphs compare to one another, and to the graphs you obtained earlier?

> ExponentialMeans(10000, 1)

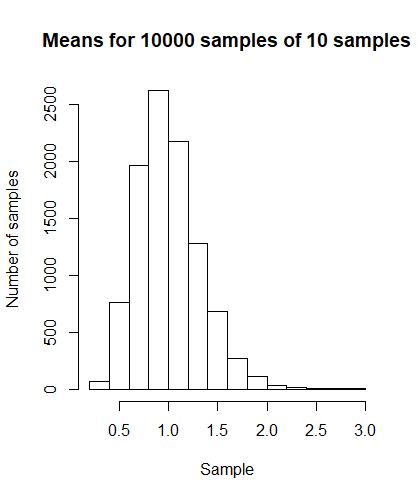
[1] 1.0007794 0.9953578  


> ExponentialMeans(10000, 2)

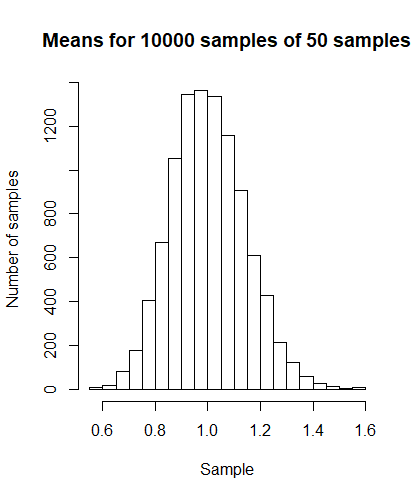
[1] 0.9938549 0.7003308



> ExponentialMeans(10000, 10)

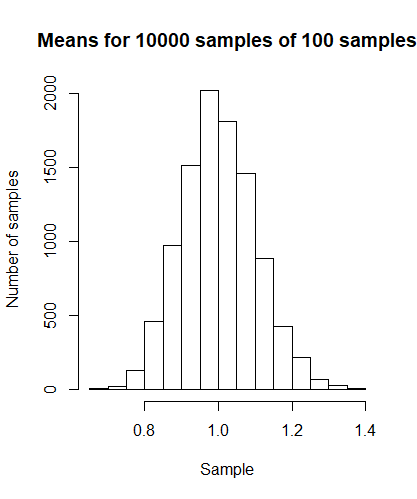
[1] 1.0018809 0.3191335  


> ExponentialMeans(10000, 50)

[1] 1.0020905 0.1413141  


> ExponentialMeans(10000, 100)

[1] 1.0013862 0.1002808



Standard deviation drops down significantly, and we can observe that the higher the sample the more bellshaped our graphs are.

## Distribution of sample means: a generalization

1. By now you’re probably starting to see a pattern. Based on your answers to Questions 1, 2, and 3, complete the following hypothesis: “If we take increasingly large samples from any set of data with mean µ, regardless of distribution, the mean of the sample means equals \_\_\_\_\_\_\_\_, the mean of the sample standard deviations gets [larger/smaller], and the distribution of sample means looks like \_\_\_\_\_\_\_\_.”

“If we take increasingly large samples from any set of data with mean µ, regardless of distribution, the mean of the sample means equals **1**, the mean of the sample standard deviations gets **smaller**, and the distribution of sample means looks like **bell-curve**.”