MTH 324, Lab #4, Fall 2022 Name ____

Instructions: Follow along with the tutorial portion of the lab. Replicate the code examples in R on your own, along with the demonstration. Then use those examples as a model to answer the questions/perform the tasks that follow. Copy and paste the results of your code to answer questions where directed. Submit your response file and the code used (both for the tutorial and part two). Your code file and your lab response file should each include your name inside.

Probability Distributions

Probability distributions built in to R come in four basic flavors: d, p, q, r. We'll use the binomial distribution as our base case.

The function dbinom() gives the probability distribution for a specific discrete value x. Here, we are calculating the probability of obtaining exactly 6 heads when we flip a fair coin 10 times.

```
dbinom(6,10,0.5)
```

We could use this function to create a graph of the distribution.

```
x <- seq(0,10,by=1)
y <- dbinom(x,10,0.5)
plot(x,y)
```

The function pbinom() works similarly, but finds the cumulative probability. We could find the probability of getting 6 or fewer heads if we flipped a coin 10 times. We could modify the code above to plot the cumulative distribution.

```
pbinom(6,10,0.5)

x <- seq(0,10,by=1)
y <- pbinom(x,10,0.5)
plot(x,y)
```

The function qbinom() is the inverse function. It finds the first discrete value whose cumulative probability that meets or exceeds a specified probability.

```
qbinom(0.95,10,0.5)
```

In this example, the cumulative probability of 8 heads is the first to exceed 95%.

Finally, the function rbinom() generates random samples from the distribution. Suppose we want to generate samples of 10 flips of a fair coin, and we want to see what could happen if we flipped those 10 coins repeatedly. We'll need to specify the number of samples. I've selected 100 samples.

```
rbinom(100,10,0.5)
```
.

We'll use the r version of distribution functions when we want to simulate data. This will come in particularly handy after the exam when we start looking at sampling distributions.

In the case of discrete probability distributions, like the binomial, we'll use the d version of the function fairly routinely, but for continuous distributions we will only use it if we want to plot a graph since it will give us the height of the distribution function rather than a probability.

It's also somewhat uncommon to use the q version of the function (the inverse) for discrete distributions. It comes up only rarely.

```
x <- seq(-10,20,by=.1)
y <- dnorm(x, mean=5,sd=4)
plot(x,y)
```

For the normal distribution, if we don't specify the mean and standard deviation, the default is the standard normal values.

A nice list of standard probabilities distributions available in R can be found here:
https://statisticsglobe.com/probability-distributions-in-r

Tasks

1. You are rolling an icosahedral die (20-sided). What is the probability of getting a number divisible by three? Use that probability of one success to find the probability of rolling that dice 10 times and getting a number divisible by 3 five times. State the result here.

2. Plot the binomial distribution for $n = 100, p = 0.03$. Plot the binomial distribution for $n = 100, p = 0.98$. Limit the range of x in your examples to the most common 10 observations (use the xlim option) so that we can see detail. Paste the graphs below and describe their shape. Do they meet the conditions for the normal approximation?

3. Plot the standard normal distribution, and the t-distribution with degrees of freedom 1, 4, and 50. On the same graph if you can or use the par() function to plot them on the same scale in a 2x2 grid. Use the range of x to be -4 to 4. Describe what you see.

4. ACT tests have a mean of about 21, and a standard deviation of about 5.2. What score represents the top 10%, top 5% and top 1% of scores on the ACT? Assume the test scores are normally distributed. Report the values you find below. Round to the nearest whole number.

5. A drive through window at a bank expects 10 customers in 30 minutes during a particular hour of the day. What it the probability that they will receive exactly 20 customers in the next hour? What is the probability that they will have more than 30 customers in that hour?

6. Choose one of the other distributions on the list of common distributions (linked above). Look up the documentation for it and experiment with modifying parameters. Make a graph of the

distribution with two different parameter settings. Paste the graphs below. Describe the effect of changing the parameters.

References:

1. Discovering Statistics Using R. Andy Field, Jeremy Miles, Zoe Field. (2012)
2. https://book.stat420.org/applied_statistics.pdf
3. https://scholarworks.montana.edu/xmlui/handle/1/2999
4. https://www.rstudio.com/resources/cheatsheets/
5. http://www.r-tutor.com/elementary-statistics/probability-distributions
6. https://www.cyclismo.org/tutorial/R/probability.html
7. https://statisticsglobe.com/probability-distributions-in-r