### Part 1. (6 points) Normal Distribution Applications An industrial process yields a large number of steel cylinders. The length of the cylinder is a random variable with an average of 3.25 inches and a standard deviation of 0.003 inches. The distribution of cylinder lengths is symmetrical, where lengths are more likely to be close to the mean rather than further away from the mean. Based on the shape of the observed data, it is reasonable to assume the length of the cylinders are Normally distributed.

### (3 points) State the parameter values that describe the distribution.

Mean = 3.25 & Standard Deviation = 0.003

### (3 points) Suppose a steel cylinder is randomly selected from the population described above. Using the pnorm() function in R, determine the probability that the length of the selected cylinder is within one standard deviation of the mean. That is, find the probability that the length is between 3.247 inches and 3.253 inches. (Hint: for R help, see the Normal Probability Functions.R file in the Week 3 module or watch the lecture recording from April 14.)

A picture containing graphical user interface

Description automatically generated

### Part 2. (6 points) Normal Distributions The Normal distribution curve to the right displays the distribution of grades given to managers based on management performance at Ford. Of the large population of Ford managers, 10% were given A grades, 80% were given B grades, and 10% were given C grades. A’s were given to those who scored 356 or higher and C’s were given to those who scored 150 or lower.

### (2 point) What are the z scores associated with the 10th and 90th percentiles from the standard normal distribution? (Hint: the qnorm() function in R can be used to find these values.)

Text

Description automatically generated with medium confidence

### (2 point) What is the mean and standard deviation of the performance scores? Show work.

Mean = (356 + 150) / 2 = 253

Standard Deviation = (356 - 253) / 1.281552 = 80.37

### (2 point) Suppose the company adds grades D and F so there are 5 categories to grade performance. If they want to give A’s only to those in the top 3%, what performance score must a manager exceed to get an A?

Top 3% = qnorm(0.97) = 1.88

253 + (1.88 \* 80.37) = 404.095

If they want to give A’s only to those in the top 3%, a manager must exceed 404.095.

### Part 3. (13 points) Simulation of Gamma Random Variables

### Background: When we use the probability density function to find probabilities for a random variable, we are using the density function as a model. This is a smooth curve, based on the shape of observed outcomes for the random variable. The observed distribution will be rough and may not follow the model exactly. The probability density curve, or function, is still just a model for what is actually happening with the random variable. In other words, there can be some discrepancies between the actual proportion of values above x and the proportion of area under the curve above the same value x. Our expectation is as the number of observations increase, literally or theoretically, the observed distribution will align more with the density curve. Over the long run, the differences are negligible, the model is sufficient and more convenient to find desired information.

### Simulation: Use R to simulate 1000 observations from a gamma distribution. A random variable following a gamma distribution has the following probability density function.

for

represents the Gamma function and is defined as . If is a positive integer, the Gamma function simplifies to .

Here is some other useful information about the Gamma distribution:

* If *X* is random variable following a Gamma distribution with parameters *a* and *b*, the expectation of *X* is *.*
* If *X* is random variable following a Gamma distribution with parameters *a* and *b*, the variance of *X* is *.*

### If you haven’t already done so, download and open in R the Gamma\_Random\_Variable\_Simulation.R script provided in Canvas.

### To begin, set alpha = 2 and beta = 7. Highlight and run the parameters and observation values. Run the simulation code to plot the observations and fit the probability density function over the observations. You don’t need to change anything. You may run the section all at once by highlighting all of the section and running it by clicking the run button at the top of the script window.

### Given the values are from a gamma distribution with alpha= 2 and beta = 7,

### (1 points) What is the expression for the probability density function? (Hint: simplify as much as you can. Use the fact that )

### 

### f(x) =

### (1 point) What is the mean and standard deviation of the random variable? Show work in regards to how you derived these quantities.

E(x) = population mean = ab = 2 \* 7 = 14

Standard Deviation = = = 9.899

### (1 point) Review the definition of the probability density function presented in Week 3. What is the probability x is less than 4? You may solve this problem by hand, which will require using the integration by parts technique. You may also solve this problem using an integral calculator (e.g. Desmos, Wolfram). However you decide to solve the problem, you must state the integral that is computed to answer this question. It is not enough to just state the final answer.

= 0.113

### (2 point) Find the section of the R script titled Simulation (line 10). Run the simulation (lines 16-48) and paste your plot. Comment on the general shape of the distribution. How well does the density curve fit the observations?

Chart, histogram

Description automatically generated

The density curve fits the observations nearly well as it appears to be a Normal Distribution skewed right.

### (2 point) What is the exact proportion of values below 4? How does the actual proportion compare to the probability from the density curve in part 2-a-iii?

Text

Description automatically generated with low confidence

The actual proportion is actually pretty close to the probability of the density curve.

### (1 point) Increase the number of observations to 10000, rerun the simulation. Paste your plot. How does increasing the number of observations affect the fit of the density curve?

Chart, histogram

Description automatically generated

As we increase the number of observations, the density curve will become more and more accurate to fit the graph of the number of observations.

### (1 point) What is the exact proportion of values below 4 when the number of observations has been increased to 10,000? How does increasing the number of observations affect the accuracy of the model? Make a comparison between this proportion and 2-a-iii and 2c.

A picture containing company name

Description automatically generated

The increasing number of observations increases the accuracy of the values stated in 2-a-iii and 2c as they are all pretty close to each other.

### (1 point) Rerun the simulation with alpha = 1, beta = 7, and observations = 10000. Paste your plot. Comment on the general shape of the distribution.

Chart, histogram

Description automatically generated

The shape of this distribution appears to be a normal distribution cut in half with a longer tail. It’s decreasing exponentially and converging to 0.

### (1 point) The model in part (f) is a special case of the gamma distribution, called an Exponential Distribution. What is the expression for the probability density function? (Hint: )

### f(x) = f(x) =

### Optional: Change the parameter values and take note of the effect of increasing or decreasing parameter values.

### Gradescope Page Matching (2 points)

When you upload your PDF file to Gradescope, you will need to match each question on this assignment to the correct pages. Video instructions for doing this are available in the Start Here module on Canvas on the page “Submitting Assignments in Gradescope”. Failure to follow these instructions will result in a 2-point deduction on your assignment grade. Match this page to outline item “Gradescope Page Matching”.