ISTA 116: Lab Assignment #7 (50 pts)

Due Wednesday, November 30th, 11:59 P.M.

Problem 1:

(9 pts)

This question does not require `R`.

Y is a continuous random variable representing the amount of time in minutes that it takes a grocery cashier to serve one customer. Suppose we know that 20% of people are checked out in less than 3.0 minutes, and the median checkout time is 5.0 minutes.

(a) (4 pts) Write down each of the two pieces of information given in terms of values of either the CDF or PDF of Y.

(b) (2 pts) Suppose there is no one in line when you walk up to the cashier. In terms of the random variable Y, what event represents the case where you take more than 3 and at most 5 minutes to check out?

(c) (3 pts) Using the information given, calculate the probability of the event described in (b).

Problem 2:

(18 pts)

Suppose a manufacturer makes a women's glove that fits people whose index finger is between 3.5 and 4.0 inches long. Suppose also that the length of female index fingers follows a Normal distribution with mean 3.20 inches and standard deviation 0.35 inches.

(a) (3 pts) Define a random variable that applies to the sample space of all women, and whose value represents the length of the women's index finger. Clearly state the name of the distribution, and all relevant parameters and their values.

(b) (2 pts) In terms of the random variable you defined in (a), what event represents a woman fitting comfortably into the glove?

(c) (3 pts) Assuming a woman is randomly selected from the population, calculate the probability of the event described in (b). (You'll need `R` for this).

(d) (3 pts) Suppose Z is a random variable representing the z-score of a woman's finger length. What distribution does it have? (Clearly state the name, and all relevant parameters and values)

(e) (4 pts) In terms of Z, what event represents a woman fitting comfortably into the glove?

(f) (3 pts) Assuming a woman is randomly selected from the population, calculate the probability of the event described in (e), using the distribution of Z.

Problem 3:

(23 pts)

You are the owner of an electronics store in a small town in a remote part of Cochise County, AZ. Based on data from a store in a similarly sized town in a neighboring county, you estimate that the number of plasma televisions sold in a given day can be described by the following probability distribution:

# of TVs sold per day	0	1	2	3	4	5	6	7	8	9
Probability	0.05	0.01	0.10	0.20	0.10	0.05	0.30	0.10	0.05	0.04

a. (2 pts) Compute the mean and standard deviation of the distribution of TVs sold per day.

b. (4 pts) Use the `sample()` function in `R` to simulate (i) a day, (ii) a week, (iii) a (30-day) month and (iv) a (non-leap) year of TV buying. State any assumptions you are making beyond what is stated in the problem description. Save your samples in variables.

c. (3 pts) Compute both the total number of TVs bought and the mean number of TVs bought per day over each of the time periods in (b).

d. (6 pts) Use `replicate()` to re-run each simulation 10,000 times. You will want to combine your code from (b) and (c) to produce a single line for each combination of time period (day, week, month, year) and statistic (sum, mean), so that you are creating a total of 8 variables, each a vector of 10,000 numbers.

The syntax looks like:

```
someVariableName <- replicate(n , mycode(...))
```

where `n` is the number of replicates, and `mycode(...)` is the code you used before, combined to do the sampling and then return a single value (in this case either a sample mean or a sample standard deviation) in a single line.

The result of each call to `replicate()` will be a vector called `someVariableName` whose first element is the outcome of the first replicate, whose second element is the outcome of the second replicate, etc.

Show your code, but in the name of all that is holy, don't include all the thousands of values!

e. (4 pts) Compute sample means and sample standard deviations for each of the 8 variables created in (d). Comment on their relationship to the mean and standard deviation you calculated in (a).

f. (4 pts) Produce two plots: one for the means and one for the sums. Each should contain density estimates from for the sampling distributions you simulated in (d) each of the four time periods (use `density(...)`). Use different colors for each time period, and label which curve is which, either with a legend on the plot itself, or in a caption beneath the image in your write-up.