Discrete & Continuous Distributions

In this data analysis, we'll use R to better understand both discrete and continuous probability distributions. If you haven't already done so, work through the R tutorial provided on the Data Analysis 3 Canvas page. Once you've worked through the tutorial, write up your responses to the questions listed throughout the tutorial. The same questions are included below to help you format your submissions.

Submit a PDF copy of your responses to Gradescope by the deadline stated on Canvas.

Discrete Distributions

Question 1: (2 points) The table below displays the values the random variable X can take on, along with the number times each value appears on the wheel (labeled Count). Fill in the last row of the table with the correct probability mass function values. That is, for each value X can take on, x, determine the probability the wheel is spun and lands on the given value, $p(x)$. Round values to three decimal places.

x	$0	$300	$350	$400	$450	$500	$550	$600	$700	$800	$900	$5000
Count	2	5	1	2	1	3	1	3	1	2	2	1
$p(x)$	0.083	0.208	0.042	0.083	0.042	0.125	0.042	0.125	0.042	0.083	0.083	0.042

Question 2: (2 points) A contestant spins the wheel. What is the expected dollar amount they will win? Show your work. *Hint: use the PMF from the previous problem to calculate the expected value of X.*
E(x) =
0*0.083+300*0.208+350*0.042+400*0.083+450*0.042+500*0.125+550*0.042+600*0.125+700*0.042+800*0.083 +900*0.083+5000*0042 = 670.3

Question 3: (3 points) Suppose a contestant spins the wheel three times. How likely is it they will spin and land on $800 all three times? Show your work.
    According to the table above we know that the probability that spin will land on $800 is 0.083, we also know that land on $800 are independent events, therefore Probability that all 3 spins will land on $800 is: 0.083* 0.083*0.083 = 0.00057179

Question 4: (3 points) Suppose a contestant spins the wheel three times. How likely is it that they will spin and land on $800 at least once? Show your work.
  The probability does not land on $800:   1-0.083 = 0.917
  Probability 3 times do not land on $800:   0.917*0.917*0.917= 0.771
  Land on $800 at least once: 1- 0.771 = 0.229

Question 5: (1 point) What is the average dollar amount of the 1000 simulated spins. How does this compare to your answer in Question 2?

Code: spins_thousand <- sample (wheel_outcomes, size = 1000, replace = TRUE)
        mean(spins_thousand)
   Average dollar amount of 1000 simulated spins: 625, which is smaller than my answer(670.3) in Question 2

Question 6: (2 points) Include a table of your simulated probabilities. How different are the simulated probabilities to the theoretical probabilities in Question 1?

 Table of simulated probabilities:
 spins_thousand

0	300	350	400	450	500	550	600	700	800	900	5000
0.081	0.199	0.048	0.086	0.047	0.138	0.041	0.107	0.043	0.072	0.103	0.035

The simulated probability values are very close to the theoretical probabilities. However, they are not exactly same as the theoretical probability values form Question 1. I think this is expected since simulations are subject to random variation.

Question 7: (1 point) In general, what action will make the simulated values more like the theoretical ones?

Just simply increase the number of simulations, the more times we simulate an experiment, the more likely we are to get results that are closer to the theoretical probabilities. This is because the Law of Large Numbers states that as the number of trials increases, the average of the outcomes will converge to the expected value.

Continuous Distributions

Question 8: (4 points) For each of the two fast food restaurants, create a histogram to visualize the distributions of the number of calories from fat of the options from these two restaurants.

a. (2 points) Include the two histograms here. *Hint: You learned how to create a histogram in data analysis 1 using ggplot and geom_histogram.*

First:
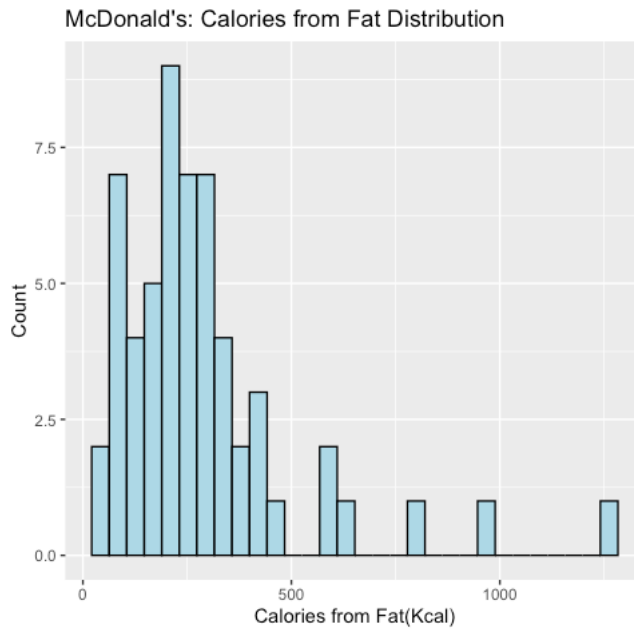
Second:
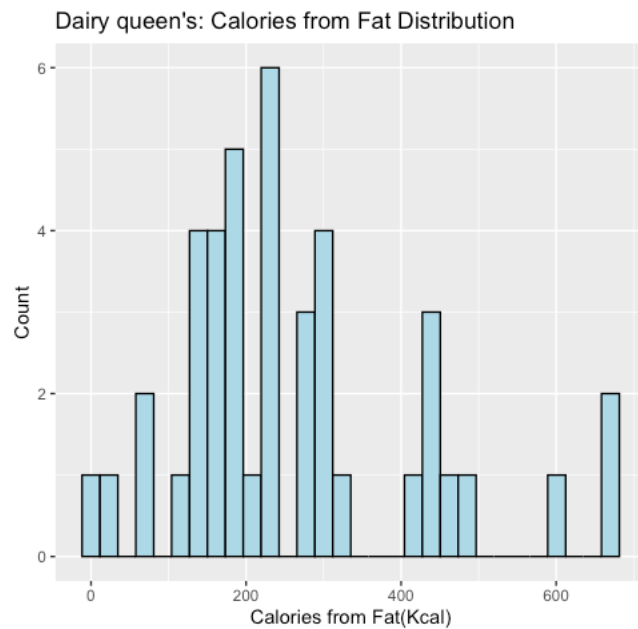
b. (2 points) How do the shapes, centers, and spreads of the two distributions compare?

Shapes: The two histograms are unimodal; the McDonalds' histogram is nearly right skewed, and the Dairy Queen histogram is nearly symmetric.

Centers: the center for McDonalds' histogram is about 240 Cal and the center for Dairy Queen's histogram is about 220 Cal. Therefore, the center of McDonald's (240 Cal) is greater than Dairy Queen(220cal)

Spreads: According to the two histograms, the standard deviation of Dairy Queen is bigger than the standard deviation of McDonald's, in other words, The histogram of Dairy Queen is more spread out than the histogram of McDonald'.

Question 9: (1 point) Based on this plot, does it appear that the calories from fat data from Dairy Queen follow a nearly normal distribution?
        Yes, based on the plot below, the calories from fat data from Dairy Queen follow a nearly normal distribution.
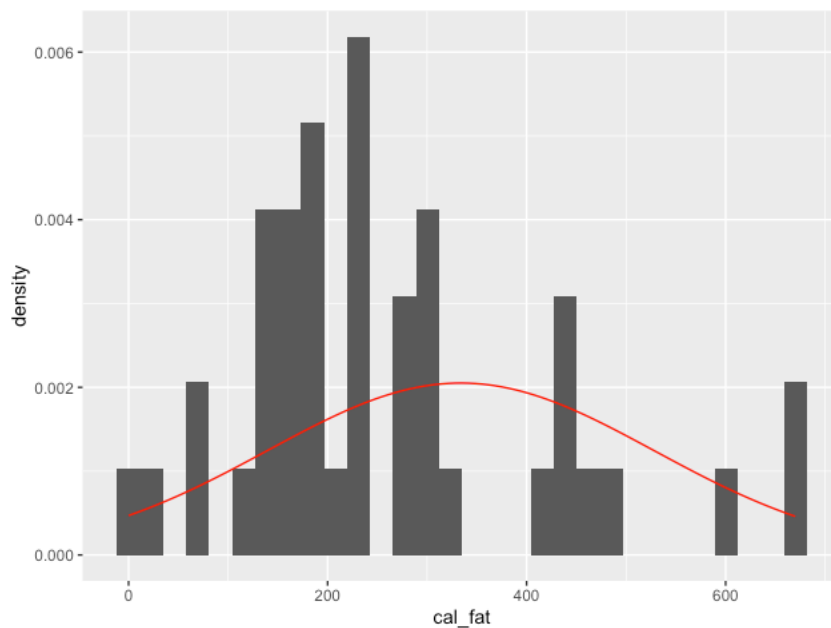
Question 10: (4 points)

a. (2 points) Write out a probability question that you would like to answer about the calories from fat content from any of the other restaurants in this dataset.

what is the probability that a randomly chosen Burger King item has more than 300 calories from fat?

b. (2 points) Calculate the probability using both the theoretical normal distribution as well as the empirical distribution. Compare the two values. *You should include the code you use to do this problem in your submission.*

The probability of using theoretical normal distribution: 0.568894

```
Code: dqmean <- mean(burger_king$cal_fat)
      dqsd <- sd(burger_king$cal_fat)
      1 - pnorm(q = 300, mean = dqmean, sd = dqsd)
```

The probability of using empirical distribution: 0.471

```
Code: burger_king %>%
      filter (cal_fat > 300) %>%
      summarise (percent = n () / nrow(burger_king))
```

According to the results I got, the probability using theoretical normal distribution is greater than the probability of using empirical distribution.

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".