Instructions: Submit one .pdf file that includes the answers to both R questions and conceptual questions. Also make sure you submit your .R file with the R code you used to get the answers.

Please feel free to discuss questions on the discussion board.

R Question

1. (2 points) Consider this game: You roll one die, and lose \$50 if you roll a 1, but win \$15 if you roll anything else. I've written a function for you, `play()` that plays this game. You can get the function by running:

```
source("https://gist.githubusercontent.com/cwickham/abe3b4c4ba5319e8e1dd5102541f2117/raw")
```

   Then you can play by calling the `play()` function

```
play()
```

   The function `play()` returns a numeric value of either `-50` or `15` depending on your roll. If you `play()` it will print a message with the outcome and return your payout. When you simulate many games, printing the outcome will be very time consuming, so use `play(silent = TRUE)` to play without printing results.

   Use R simulations to estimate your expected win/loss value for one roll. How many times did you play to find your estimate? How precise do you think your answer is? How much would you be willing to pay to play this game?

   Discussion Ideas (optional, and no extra credit beyond discussion points). You might want to take a look at the inner workings of `play()`. You can take a look at: `https://gist.githubusercontent.com/cwickham/abe3b4c4ba5319e8e1dd5102541f2117`. Can you come up with your own game? Can you challenge your discussion mates to decide how much to charge for playing your game? Or instead, can you explain to someone how to host their code in a github gist and share it with people?

2. (3 points) In lab you explored the Central Limit Theorem when the population distribution was a Gamma(5, 1). The amazing thing about the Central Limit Theorem is that it applies no matter the shape of the distribution (as long as the distribution has an expected value, and a finite variance). For this question, choose one of the following distributions, and replicate the exploration from the lab with sample sizes of 2, 5, 10, 30 and 100:

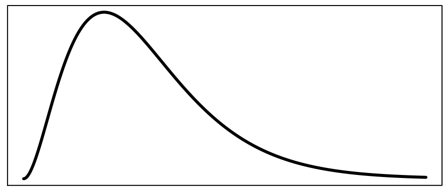
   - Continuous uniform on (0, 1), see `?runif`
   - Discrete uniform on $1, \dots, 10$, use `sample()`
   - A poisson distribution with your choice of parameter, see `?rpois`
   - Beta distribution with both parameters set to 0.5, see `?rbeta`

   Write up your exploration in a way that a reader can follow without having to understand your code.

Conceptual Questions

3. (2 points) Describe why we do not usually know the population mean. What statistic do we usually use to estimate the population mean and why?

Answer any three of the following four short answer questions.



4. (1 point) The above plot represents the distribution function for a random variable X:
   (a) Is this distribution skewed? Why or why not? If so, what is the direction of the skew?
   (b) Describe the relationship between the mean and the median for this distribution. If they are different from each other, why?
   (c) How many modes does this distribution have? Is it possible for a distribution to have more than one mode, median, or mean? Why?
5. (1 point) Consider the three independent random variables X_1, X_2, X_3 with standard deviations $\sigma, 2\sigma, \sigma$ Let $X = 4X_1 + \frac{1}{2}X_2 + X_3$.
   (a) Why do we *not* simply sum the deviations from the mean the mean to describe variation?
   (b) Why do we usually use variance instead of absolute deviations to describe variation?
   (c) What is an advantage of using variance over standard deviations?
   (d) Compare the variance of X_1 with the variance of X_2.
   (e) Compare the variance of X with the variance of X_1
6. (1 point) Participants in a cake frosting race must frost identical cakes to make them look like a given picture. Their cake-frosting times are recorded and their mean cake-frosting time is calculated. Ten cake makers are selected at random and their mean cake-frosting time is 8.73 minutes. One thousand cake makers are selected at random and their mean cake-frosting time is 10.19 minutes. One hundred thousand cake makers are selected at random and their mean cake-frosting time is 9.97 minutes. What would you guess is the mean cake-frosting time of the population of cake makers? Why?
7. (1 point) Pianos are rated on discrete a scale from 0 to 10 for having the correct pitch, with 10 being the best score. If larger and larger random samples of pianos are selected for a rating, and their mean pitch score is calculated, what will the distribution of the sample mean start to look like? What is the theorem that is responsible for this property? Does the existence of this phenomenon depend on the distribution of the population?