Homework №8

Math 107, Spring 2016 Due: May 3 by 4:00 pm

Problem 1

A food safety inspector is called upon to investigate a restaurant with a few customer reports of poor sanitation practices. The food safety inspector uses a hypothesis testing framework to evaluate whether regulations are not being met. If they decide the restaurant is in gross violation, its license to serve food will be revoked.

(a) Write the hypotheses in words.

(b) What is a Type 1 Error in this context?

(c) What is a Type 2 Error in this context?

(d) Which error is more problematic for the restaurant owner? Why?

(e) Which error is more problematic for the diners? Why?

(f) As a diner, would you prefer that the food safety inspector requires strong evidence or very strong evidence of health concerns before revoking a restaurant's license? Explain your reasoning.

Problem 2

In this problem, we will explore properties of the sampling distribution and how it relates to the population distribtution. The population we will study consists of all Major Leauge Baseball (MLB) players in 2016. The file `mlb2016.csv` contains salary information for all 862 MLB players in 2016 (Source: http://www.usatoday.com/sports/mlb/salaries/). In this problem, we will consider player salaries.

(a) Calculate the population mean and standard deviation of the salaries for all 862 MLB players in 2016.

(b) Create a histogram of the salaries for all 862 MLB players in 2016. Include this plot in your homework submission along with a description of the distribution of the salaries.

(c) Use the following R code to obtain a single random sample of size $n = 25$ from the data set. Calculate the mean salary of the players included in the sample. Note that this would be one value in the sampling distribution.

```
# Load in the data first. I called the data set mlb2016.

# Randomly sample 100 players.
samp <- sample(mlb2016, size = 25)
```

(d) Create a histogram of the salaries for the 25 sampled players. Include this plot in your homework submission along with a description of the distribution of the salaries.

(e) Use the following R code to simulate the sampling distribution of the sample mean for samples of size $n = 25$. Calculate the mean and standard deviation of the salaries of the sampling distribution.

```
sampling_dsn25 <- do(10000) * sample(mlb2016, size = 25) %>%
  summarise(mean = mean(Salary))
```

(f) Create a histogram of the simulated sampling distribution of the sample mean for samples of size $n = 25$. Include this plot in your homework submission along with a description of the distribution.

(g) Use the following R code to simulate the sampling distribution of the sample mean for samples of size $n = 50$. Calculate the mean and standard deviation of the salaries of the sampling distribution.

```
sampling_dsn50 <- do(10000) * sample(mlb2016, size = 50) %>%
  summarise(mean = mean(Salary))
```

(h) Create a histogram of the simulated sampling distribution of the sample mean for samples of size $n = 50$. Include this plot in your homework submission along with a description of the distribution.

(i) Use the following R code to simulate the sampling distribution of the sample mean for samples of size $n = 100$. Calculate the mean and standard deviation of the salaries of the sampling distribution.

```
sampling_dsn100 <- do(10000) * sample(mlb2016, size = 100) %>%
  summarise(mean = mean(Salary))
```

(j) Create a histogram of the simulated sampling distribution of the sample mean for samples of size $n = 100$. Include this plot in your homework submission along with a description of the distribution.

(k) Compare the shapes of population distribution and the sampling distributions that you have simulated. Does the sample size impact the shape of the sampling distribution?

(l) Compare the means of the sampling distributions that you have simulated. How do they compare to the population mean? How does the sample size impact the mean of the sampling distribution?

(m) Compare the standard deviations (i.e. standard errors) of the sampling distributions that you have simulated. How do they compare to the population standard deviation? How does the sample size impact the standard error of the sampling distribution?