# Problem set 7

S520

**Upload a HTML/PDF file with your answers to all questions and a .R/.Rmd file with your code through the Assignments tab on Canvas by 11:59 pm, Thursday 21st October.** Use R Markdown unless you have a good reason not to.

1. In the Powerball lottery, there are 59 white balls, numbered 1 to 59. Each week, five of the white balls are drawn, without replacement. In the past, the most frequently occurring white ball has been 23.

   (a) In the next lottery, will the probability of drawing the number 23 be greater than 5/59, less than 5/59, or equal to 5/59?

   Before one season, the Oakland A's were considered to be an average major league baseball team, predicted to win half (81) of their 162 games. They win the first six games of the season.

   (b) True or false: After the first six games, the best prediction of the total number of games the Oakland A's win that season is 162 out of 162.

   (c) True or false: After the first six games, the best prediction of the total number of games the Oakland A's win that season is still 81 out of 162.

   I survey a simple random sample of 1000 U.S. households and find out their income.

   (d) True or false: By the Central Limit Theorem, the incomes in the population will have an approximately normal distribution.

   (e) True or false: By the Central Limit Theorem, the incomes in the sample will have an approximately normal distribution.

2. Let $X$ be a discrete random variable with probability mass function

$$P(X = x) = \begin{cases} 0.3 & x = -2 \\ 0.6 & x = -1 \\ 0.1 & x = 12 \\ 0 & \text{otherwise.} \end{cases}$$

   Let $X_1, \ldots, X_n$ be an iid sequence of random variables with the same distribution as $X$. Let $\bar{X}$ be the sample mean (of $X_1, \ldots, X_n$.)

   (a) Find $E(X)$.

   (b) Find $\text{Var}(X)$.

   (c) What is the expected value of $\bar{X}$?

   (d) If $n = 100$, what is the variance of $\bar{X}$?

(e) Suppose $n = 100$. Use the R function `pnorm()` to find the approximate probability that $\bar{X}$ is greater than 0.5.

3. Trosset exercise 8.4.4. (Assume the sample size is large enough for the Central Limit Theorem to approximately hold.)

4. Trosset chapter 8.4 exercise 6.

5. I toss a fair coin 100 times.

   (a) Let $X$ be the number of heads. Using the binomial distribution, find the probability that $P(40 \leq X \leq 60)$.

   (b) Let $\bar{X}$ be the sample proportion of heads (i.e. the number of heads divided by 100.) Using the normal distribution as an approximation, find $P(0.395 \leq \bar{X} \leq 0.605)$. Is this close to your answer for (a)?

6. I want to find out the average number of people per household in the U.S. I survey a simple random sample of U.S. households and obtain the results displayed in the table below.

| Household size | Number of households |
|---|---|
| 1 | 27 |
| 2 | 34 |
| 3 | 16 |
| 4 | 13 |
| 5 | 6 |
| 6 | 3 |
| 7 | 1 |

   It may help to enter the data into R as follows:

```
household_sizes <- rep(1:7, c(27, 34, 16, 13, 6, 3, 1))
```

   (a) Lacking any other information, our best estimate for the population mean household size is the sample mean. What is the sample mean of our data?

   (b) What is our estimate for the standard deviation of household sizes?

   (c) What is the estimated standard error of the sample mean? (That is, based on our answer to (b), what is our estimate for the standard deviation of the distribution of the sample mean?)

   (d) Our error is the difference between the sample mean and the population mean. By the Central Limit Theorem, this error should have an approximately normal distribution with expected value 0 and standard deviation given by your answer to (c). Using the normal distribution, find the approximate probability that the absolute value of the error in a survey of this form and size is less than 0.5.

   (e) Can we be reasonably sure that the average household size for all U.S. households is between 2 and 3?