

R Report II

FORMAT

- * Use `set.seed(10)` at the beginning of your document.
- * Use complete sentences and proper grammar to answer all questions.
- * Use R Markdown to create an html document. Write in a report style format.
- * Code should not be in the body of the text, so be sure to add `echo = FALSE` in the preface to your R chunks. All code should be included at the end of the homework, as an appendix.

- I. Simulate a binomial random variable. Consider a class with 80 students, and the probability that a student does not turn in a homework is 0.15 (a “success”). Assume all students are independent of all other students, and the probability does not change.
- (a) Use `sample` to simulate drawing 80 students who either do, or do not, turn in their homework, and then find the total (out of 80) who did not turn in their homework. You should return one number, X = total # of students out of 80 who did not turn in their homework.
 - (b) Repeat (a) 1000000 times (you should have 1000000 values for “number of successes”, or X), plot a histogram of your result (**do not print out the 1000000 values!!**). For this particular binomial distribution, is the distribution symmetric? Explain.
 - (c) Find the average of the number of successes in 80 trials **and** the standard deviation based on your simulation from part (b).
 - (d) Estimate the probability that all students turned in their homework based on your simulation from (b).
 - (e) Estimate the probability that at least four students did not turn in their homework based on your simulation from part (b).
 - (f) What is the median number of students who will forget their homework based on your simulation from (b)?
- II. The goal of this problem is to simulate the distribution of the sample mean. We will use the built in dataset `lynx`. To load the dataset and avoid some problems, copy and paste the following command:

```
lynx = as.numeric(lynx)
```

Assume this vector represents the **population**. I.e., the mean of this vector is our “true mean”.

- (a) Draw a histogram of the population, find the “true” mean, and the “true” variance. Does this data look normally distributed?
- (b) Draw a random sample of size 10 from the population, and report back the mean of your sample.
- (c) Repeat part (b) 1000000 times. You should have a vector of 1000000 means. Find the mean of that vector, and the standard deviation of that vector. Plot a histogram.
- (d) Draw a random sample of size 50 from the population, and report back the mean of your sample.
- (e) Repeat part (d) 1000000 times. You should have a vector of 1000000 means. Did the mean and standard deviation increase, decrease, or stay approximately the same from (c)?
- (f) Plot a histogram of the vector in (e) and describe the shape. What did you notice as the sample size increased?