Week_1_Problem_Set

Dan Crownover

2024-12-02

Turn in one knitted PDF file. I want to be able to see your code chunks as well outputs and answers in paragraph form (not in the code chunks).

install.packages("tinytex") tinytex::install_tinytex()

library(tinytex)

#1. Flipping a coin To simulate flipping a coin, we can employ the binomial distribution which yields the number of successes in a sequence of yes/no experiments. We can implement this in R through the rbinom command. rbinom chooses random deviates from the binomial distribution. Let's flip a single coin:

```r
## n is the number of times we conduct the expiriement
## size is the number of times we flip the coin
rbinom(n = 20, size = 10, prob = 0.5)
```

```
##  [1] 4 5 5 5 3 3 6 2 3 5 4 3 2 5 5 3 3 6 5 6
```

In this case, a result of 1 means you successfully flipped a head, whereas a 0 means you flipped a tail. In this function size is the number of trials per "experiment" and n is the number of different "experiments". For example, let's say each one of our trials consists of flipping a fair coin 10 times, and we want to conduct the experiment 20 times. We could implement this as:

```r
## n = 20 we conducted the expiriment 20 times
##  flipping the coin 10 times
## assigning it to variable flip
flip <- rbinom(n = 20, size = 10, prob = 0.5)
```

flip tells us the number of heads (successes) out of 10 trials, and we can easily calculate the proportion of times we get heads versus tails:

```r
## calculting porporton of sucsess over number of times we flipped the coin = porportion of sucsess
flip/10
```

```
##  [1] 0.5 0.2 0.8 0.6 0.8 0.6 0.7 0.6 0.0 0.6 0.3 0.3 0.3 0.7 0.6 0.5 0.6 0.5 0.4
## [20] 0.6
```

you can take the average of that across all of the experiments to get the average proportion of heads.

```r
## .45
mean(flip)/10
```

```
## [1] 0.51
```

We can also flip a coin using the function "sample". Note, however, that we have no control over the probability of getting heads. Each element has an equal probability of being chosen.

```r
x <- sample(c("H","T"), 10, replace=T)
table(x)/10
```

```
## x
##   H   T
## 0.5 0.5
```

```
## the final thing is that this only has a .5 chance -- rbinom is better because we can alter the probab
```

#Question 1 (5 points) What is the average number of heads overall? What happens to the average if you flip a biased coin with a probability of 44% of getting a head?

```
## the average overall is 5 Heads if you flip the coin 10 times
```

```r
flipQ1 <- rbinom(n = 20, size = 10, prob = 0.44)
flipQ1
```

```
##  [1] 5 8 6 3 6 3 3 2 3 4 5 5 3 3 4 5 6 4 3 7
```

```r
mean(flipQ1)
```

```
## [1] 4.4
```

```
## if you flip it 10 times with a .44 probability of heads you land on heads 4.65 times
```

#Question 2 (10 points) Using the above coin-flipping function, flip a fair coin 50, 500, and 5000 times(the sample function or the binom). Then flip a coin 50, 500, and 5000 times with a 57% probability of getting a heads. What is the average proportion of heads? Show the code for each coin flip. Write a few sentances describing your results and the effect of flipping a coin more times.

```
## rbinom flipping the cone 50 times with a .57 probability of getting heads
## for the nature of this question its only asking if we conducted one trail and flipped it each 50 tim
```

```r
flipQ2_50 <- rbinom(n = 1, size = 50, prob = 0.57)

flipQ2_50
```

```
## [1] 33
```

```r
mean_flip50 <- mean(flipQ2_50)

p_flip_50  <- mean_flip50/50

p_flip_50
```

```
## [1] 0.66
```

```
## mean for 50 is 28.15 times
## probability is .567
```

```
## rbinom function of flipping the coin 500 times
## rbinom flipping the cone 50 times with a .57 probability of getting heads
flipQ2_500 <- rbinom(n = 20, size = 500, prob = 0.57)

flipQ2_500
```

```
##  [1] 281 284 281 299 265 284 284 293 279 272 274 283 299 275 287 274 298 312 272
## [20] 293
```

```
mean_flip500 <- mean(flipQ2_500)

p_flip_500 <- mean_flip500/500

p_flip_500
```

```
## [1] 0.5689
```

```
## average probablility number of heads flipping it 500 times is .5776 times
```

```
## rbinom function of flipping the coin 5000 times
```

```
flipQ2_5000 <- rbinom(n = 20, size = 5000, prob = 0.57)

flipQ2_5000
```

```
##  [1] 2845 2863 2827 2824 2792 2866 2813 2886 2789 2846 2838 2853 2848 2894 2853
## [16] 2860 2833 2828 2837 2815
```

```
mean_flip5000 <- mean(flipQ2_5000)

p_flip_5000 <- mean_flip5000/5000

p_flip_5000
```

```
## [1] 0.5681
```

```
## average  number of heads flipping it 5000 times is 0.5691 times
```

Q2 Write a few sentences describing your results and the effect of flipping a coin more times:

increasing the number of coin flips enhances the normality of the data distribution. With more sample means and porportions available (e.g., 50, 500, 5000), the proportion of heads approaches 57%(or the .57 porportion). As the number of coin flips increases, it consistently converges towards this value, in line with the Central Limit Theorem. The greater the number of coin flips, the more accurately the distribution resembles an equally distributed and normal bell-shaped curve, thanks to an increased number of sample means and bins.

#Question 3 (10 points) Last season, RJ Davis had a free throw percentage of 0.881, a three-point percentage of 0.362, and a field goal (2 points) percentage of 0.438. Would RJ have the best chances of making three

points by (a) shooting a single three-point shot, (b) shooting three free throws, or (c) shooting a two-point field goal and a free throw (ignoring the fact that he would have to get fouled first)? Show your work in R.

```
## Shooting 3 free throws

(.881)*(.881)*(.881)
```

```
## [1] 0.6837978
```

```
## probability is .68379 for shooting 3 free throws
```

```
##shooting a single three-point shot

point_shot <- (.362)

point_shot
```

```
## [1] 0.362
```

```
## probability is a single three point shot is .362
```

```
## shooting a two-point field goal + a free throw
##


RJ_fieldgoalPOINT_SHOT <- (.438)*(.881)

RJ_fieldgoalPOINT_SHOT
```

```
## [1] 0.385878
```

```
 ## prob of making a two point field goal and a free throw is .3858
```

Above displays the varying probabilities of making the different shots. The probability of shooting three free throws is .68379. The probability of shooting a single three point shot is .362 and the Probability of making a two point field goal and a free throw is .3858. From these probabilities, the best chance RJ has to score a 3 point shot is to go for the 3 free throws.

#Question 4 (15 points) Suppose we were interested in estimating the relative abundance of insect species in the Cape Fear River. We collect data on the abundances of 12 different species:

```
insects <- c(0.14, 15.49, 29.04, 6.36, 1.83, 5.40, 31.89, 3.92, 0.54, 2.01, 12.67, 48.75)
hist(insects, las = 1, ylab = "No. species", xlab = "Abundance", main="",
col="lightgrey")
```

```r
mean(insects)
```

```
## [1] 13.17
```
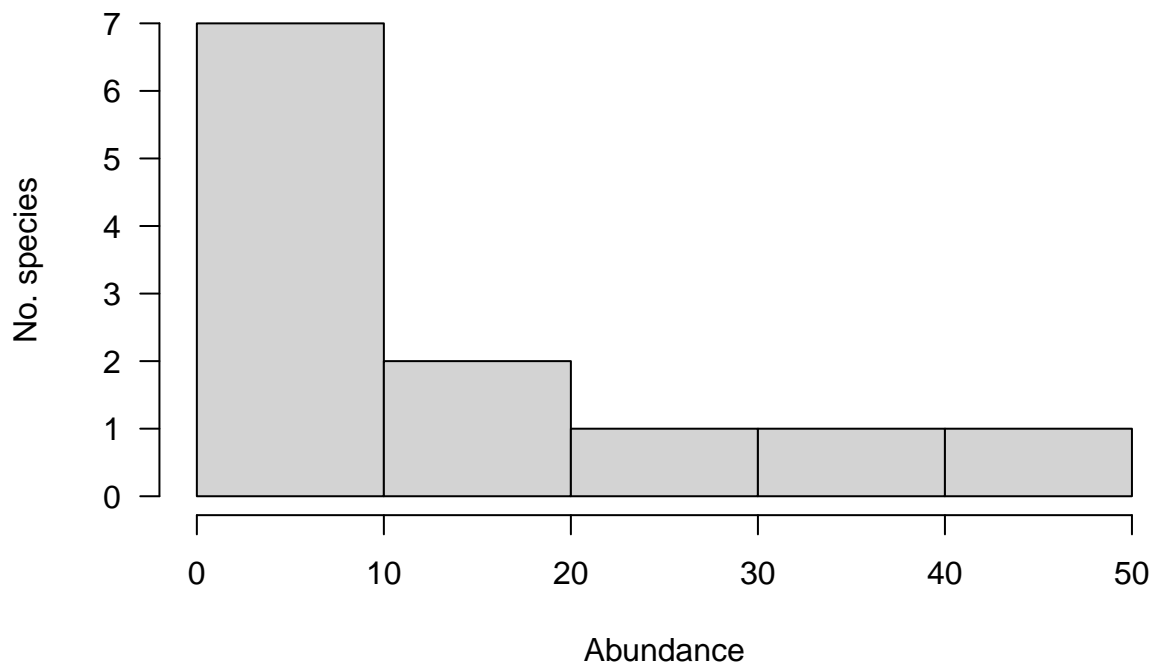
```r
##
set.seed(1001)
boot.mean <- c()
n <- 12
for(i in 1:1000){
  ## this is the bootstrapping part
boot.sample <- sample(insects, n, replace = T)
boot.mean[i] <- mean(boot.sample)
}

## mean of bootstrapped value(1. Bootstrap the data and calculate the mean of the bootstrapped means. H
mean(boot.mean)
```
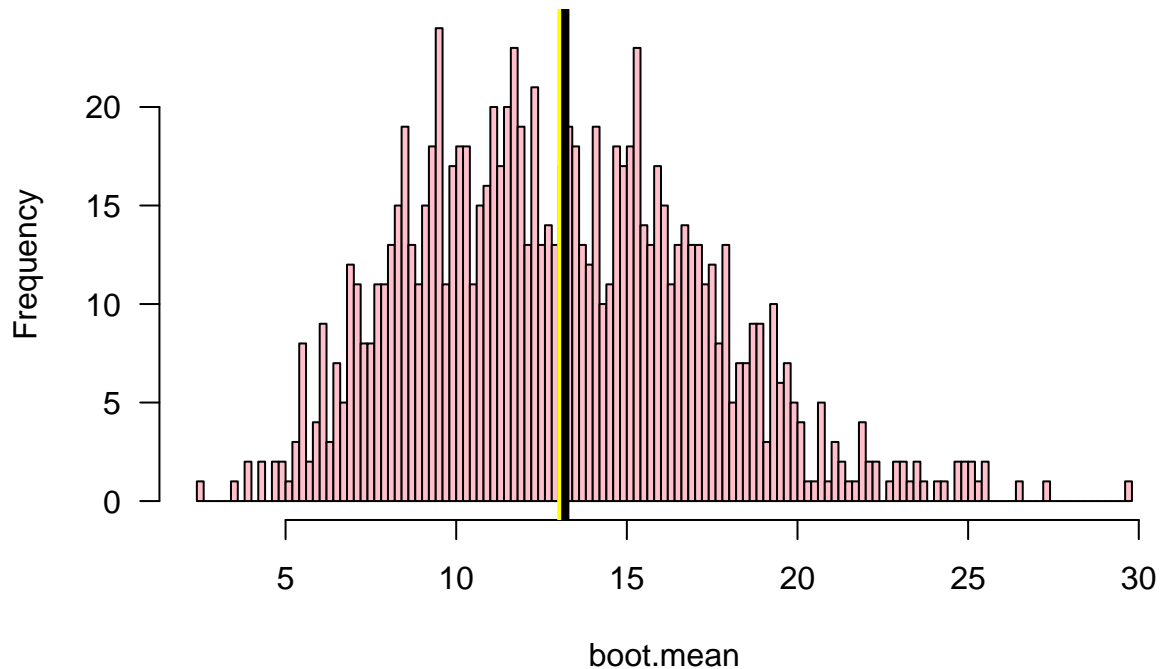
```
## [1] 13.01771
```

```r
mean(insects)
```

```
## [1] 13.17
```

```
## standard error function
se <- sd(boot.mean)
se
```

```
## [1] 4.261952
```

```
## creating the histogram(  2. Create a histogram of your bootstrapped means (5 points))
hist(boot.mean, breaks = 100,las = 1, col = "pink", main = "")
abline(v=mean(insects),lwd=5,col="black")
abline(v=mean(boot.mean),lwd=2,col="yellow")
```



```
## this round() function breaks the data into quantiles( Calculate the 90% confidence interval for the
## the confidence interval is .05 becasue half of .10 is .05
```

```
confinter <- round(c(quantile(boot.mean, 0.05), quantile(boot.mean, 0.95)), 2)
```

```
## Printing the confidence interval
confinter
```

```
##     5%    95%
##   6.64  19.96
```

Estimate the mean abundance of insect species using the bootstrap procedure(what we did in question 1 2 and 3), given that small sample size and lack of a normal distribution prohibit us from using parametric methods.

Please do the following: 1. Bootstrap the data and calculate the mean of the bootstrapped means. How does this compare to the sample mean? What is the standard error of the bootstrapped mean? (5 points)

the mean of sampling dist is 13.17, the mean of the bootstrapped value is 13.01771. Due to the central limit theome, the value of the bootstrapped mean is more accurate than the mean of the sample distrobution. which is why it is a more refined number.

2. Create a histogram of your bootstrapped means (5 points)

3. Calculate the 90% confidence interval for the bootstrapped mean of the relative abundance of insect species using the percentile CI method (i.e not the normal distribution).(5 points)

Knit your file to a PDF (or knit to HTML document and then save as PDF). Upload to Canvas Make sure the knitted version shows your r code and answers.