2244 Triangle Assignment4 xjian59

Xinguang Jiang

2021/3/21

#2244 Triangle Assignment4

#Q1. The confidence interval can be understood as the "probability that a detected numerical parameter of something will fall between a set of values for a certain proportion of times", maybe this is a little bit complicated, I will explain the confidence interval to you. In statistics, confidence is another way to describe probability. You may want to ask why it is an interval but not a single value, the reason is when we analyze a bunch of survey data, again and again, we want to use this "Probability interval" (a range) to expect our estimate to fall between a certain percentage of the time. What we need is a range to illustrate the mean, the biggest, and smallest possible data for a specific statistic.

#Q2.a A binomial probability model would be appropriate to address the nursery's goal for this situation. There is a fixed number, n, for the observation of crop A's height and day. In this case, the nursery plan to take a simple random sample of 60 seeds from crop A, n = 60. The number n is independent. Each observation falls into one of two categories ("success" & "failure"), in this case, success = (plant height >= 10cm && day <= 5) while failure = (plant height < 10cm || day > 5). The probability of a plant which height >= 10cm and day<= 5 is the same for each sampled plant.

#Q2.b

```
#import t4data.csv file
data<- read.csv(file = "t4data.csv", header = TRUE, sep = ",")
Asubset <- subset(data, crop == "A",select = c(crop,height,day))

#success = plant which height >= 10 and day <= 5
#failure = plant which height < 10 or day >5
#simple random select 60A crop from all A crops
A60set <- Asubset[sample(1:nrow(Asubset), 60), ]

#Acount to represent the number of success in 60
Acount = 0
for(row in 1:nrow(A60set)) {
  H <- A60set[row, "height"]
  D <- A60set[row, "day"]
  if( 10 <= H & D <= 5){
    Acount = Acount +1
  }
}
#prob = the probability of success for every sample
prob <- Acount/60
#this is the probability of getting 24 or less plant which has 10+cm in height and 5-day in time
#, as the question asks us to show the probability that sample will contain 25 or more seeds
#that attain the threshold size by no more than 5 days, 1-pbinom(24, 60, prob) is the answer
#for this question
answer <- 1- pbinom(24, 60, prob)
```

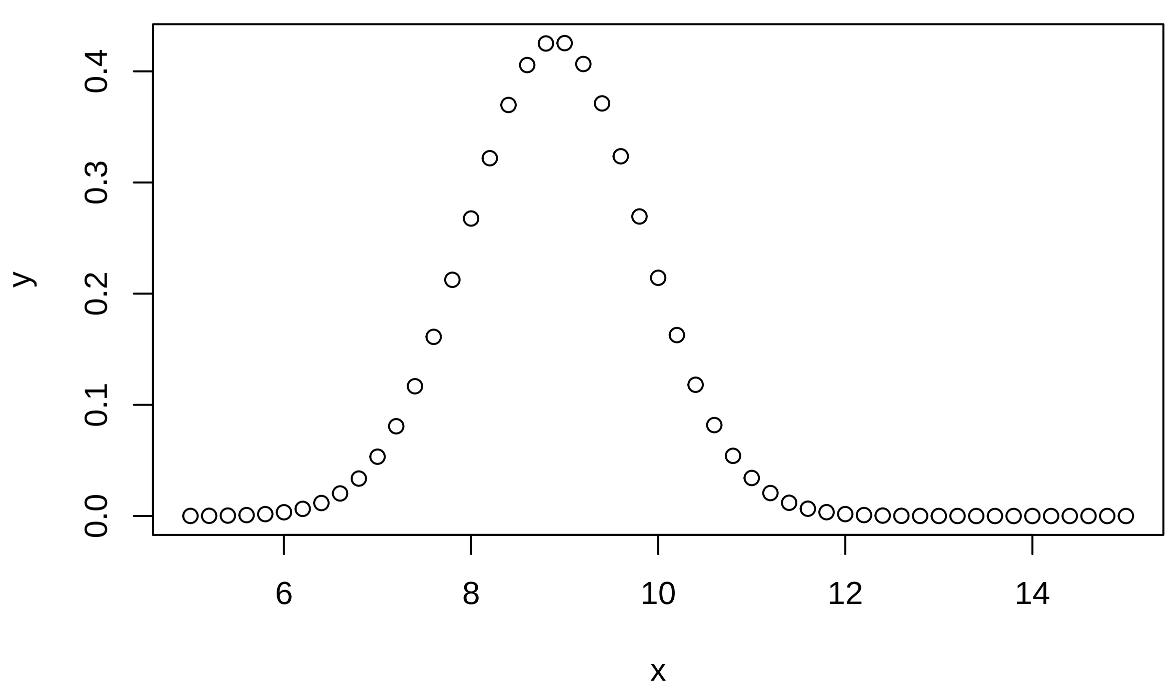
```r
print(paste( "The probability is:", answer))
```

```
## [1] "The probability is: 0.000390812533542229"
```

#Q3.a A normal probability model would be appropriate to address the nursery's goal for this situation. There is a fixed number, n, for the observation of crop B's weight on the 10th day of growth. In this case, the nursery plan to take a simple random sample of 60 seeds from crop B, n = 60. The number of B palnt n is independent. As the sampling distribution of B plant is a normal distribution. A normal probability model can be used to determine the percentage of a certain weight interval(8.7g-9.1g) in this case.

#Q3.b

```r
#import t4data.csv file
data<- read.csv(file = "t4data.csv", header = TRUE, sep = ",")
#select plant B which day=10, and store their weight in 10th day
Bsubset <- subset(data, crop == "B" & day == 10,select = c(crop,weight))
#simple random select 60 B plants from 10 day B plant, with replacement
B60set <- Bsubset[sample(1:nrow(Bsubset), 60, TRUE), ]
#try to calculate the mean and standard deviation of B
avg <- mean(B60set$weight)
sd <- sd(B60set$weight)
#this graph shows why we choose Normal but not binomial
x <- seq(5,15,by = .2)
y <- dnorm(x, avg, sd)
plot(x,y)
```




```r
#calculate 8.7&9.1 cumulative normal probability seperately.
percent87 <- pnorm(8.7, avg, sd)
```

```r
percent91 <- pnorm(9.1, avg, sd)

#percentage between 8.7-9.1
percent87to91 <- percent91 - percent87
#1- (percentage between 8.7-9.1) we'll get the percentage of samples <8.7g or >9.1g
answer <- 1 - percent87to91

print(paste( "The percentage is :", answer))
```

```
## [1] "The percentage is : 0.830233857761922"
```