

## 812 Exercise 2

Hasaan Parker

### Binomial Distribution

Always use `set.seed(12345)`

1) Conduct a trial of flipping a fair coin 5 times. Let heads equal a success in the trial. What is the number of successes in the trial? Then flip the coin 10 times, 50 times, 100 times, 1000, and 10000 times. What happens as the number of coin flips increases? Hint use `rbinom()`

```
q1 <- rbinom(5,1, .5)
q1
```

```
## [1] 1 1 1 1 0
```

```
a <- c(10, 50, 100, 1000, 10000)
sapply((sapply(a, rbinom, size=1, prob = .5)), mean)
```

```
## [1] 0.4000 0.4600 0.6500 0.5190 0.5001
```

As we add more observations, we generally get closer to the true probability of 0.5. When I conduct a trial with only five observations, I got 2 heads.

2) What does the binomial distribution look like? Generate some data to visual the distribution. Use 1000 observations across 100 trials. Assume the coin is weighted to favor heads, so the probability of success equals 0.70. What is the mean number of successes across the trials?

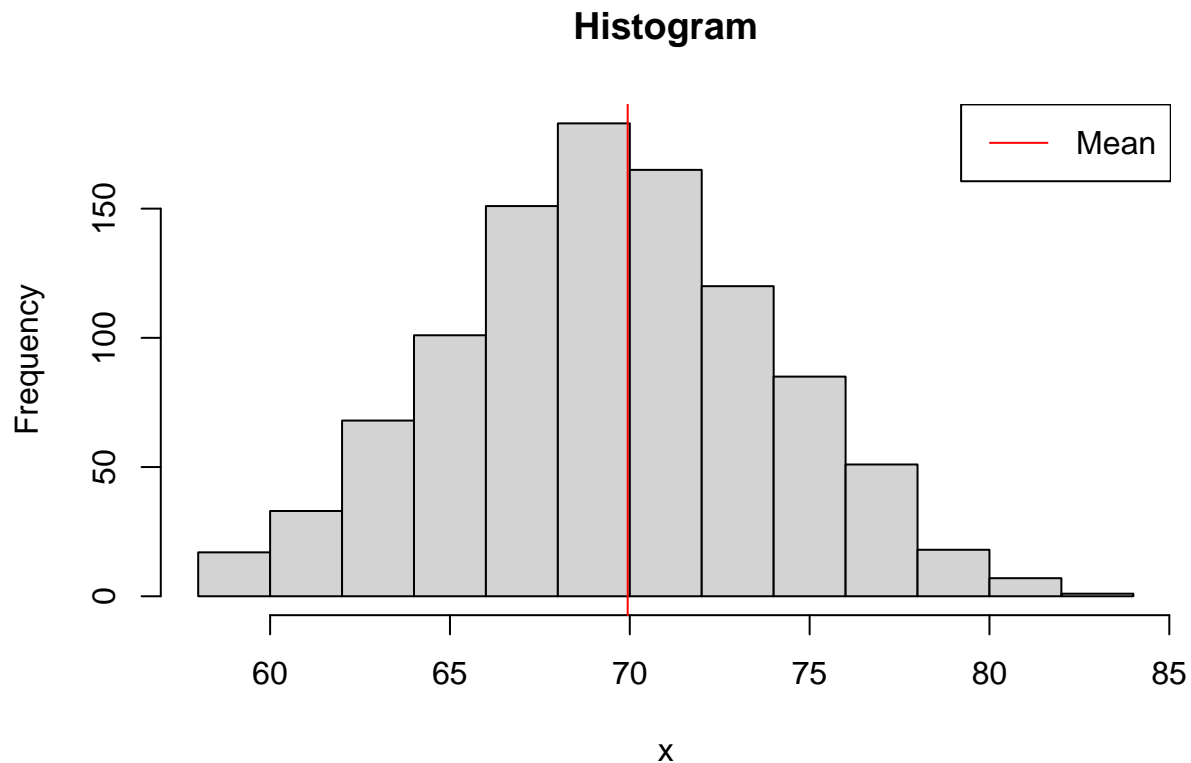
```
x <- rbinom(1000, 100, .7)
mean(x)
```

```
## [1] 69.941
```

The mean number of successes in each trial is 69.94.

3) Create a histogram of the trials. Generate a red dashed line at the mean. Make sure the labels/titles of the histogram are appropriately labeled. Hints: use `hist()` and `abline()`

```
hist(x, main = "Histogram")
abline(v = mean(x), col = "red")
legend("topright", col = "red", legend = "Mean", lty = 1)
```



4) Suppose that we flip the same coin 100 times. What is the probability that the coin lands on heads 70 times? How does this compare to the data you generated in question 2? Hint: use `dbinom()`

```
dbinom(70, 100, .7)
```

```
## [1] 0.08678386
```

```
mean(x == 70)
```

```
## [1] 0.103
```

I calculate that the probability of getting exactly 70 heads when you flip the coin 100 times is 8%. In our data from question 2, 10.3% of the trials ended up with 70 successes.

5) Suppose that we flip the same coin 100 times. What is the probability that the coin lands on heads 70 times or less? How does this compare to the data generated in question 2? Hint: use `pbinom()`

```
pbinom(70, 100, .7)
```

```
## [1] 0.5376603
```

```
mean(x <= 70)
```

```
## [1] 0.553
```

The probability of getting 70 or fewer heads is 53.77%, and in Q2 data 55.3% of the coin flips had 70 or fewer heads.

6) Every year, many people miss flights because they cancel at the last minute or are late to the airport. Research by Leder, et al. (2002) suggests that about 12% of airline passengers miss flights every year for these reasons. This tendency for a significant percentage of passengers to miss their flights is well known among airlines. To deal with this, airlines typically overbook seats, meaning they sell more tickets for a flight than there are seats available because they can count on a significant percentage of passengers to fail to show up. This is a somewhat risky strategy because there will be times when too many passengers show up and not all of them will be able to get seats. How often can we expect this to happen? We can use the binomial distribution to model an example. Suppose that there is a flight with 100 seats, and 110 tickets have been sold. What is the probability that more than 100 passengers will show up on time for this flight? I.e., how likely is it that at least one too many people will show up for this flight and be unable to board because there isn't any space?

```
pbinom(101, 110, .88, lower.tail = F)
```

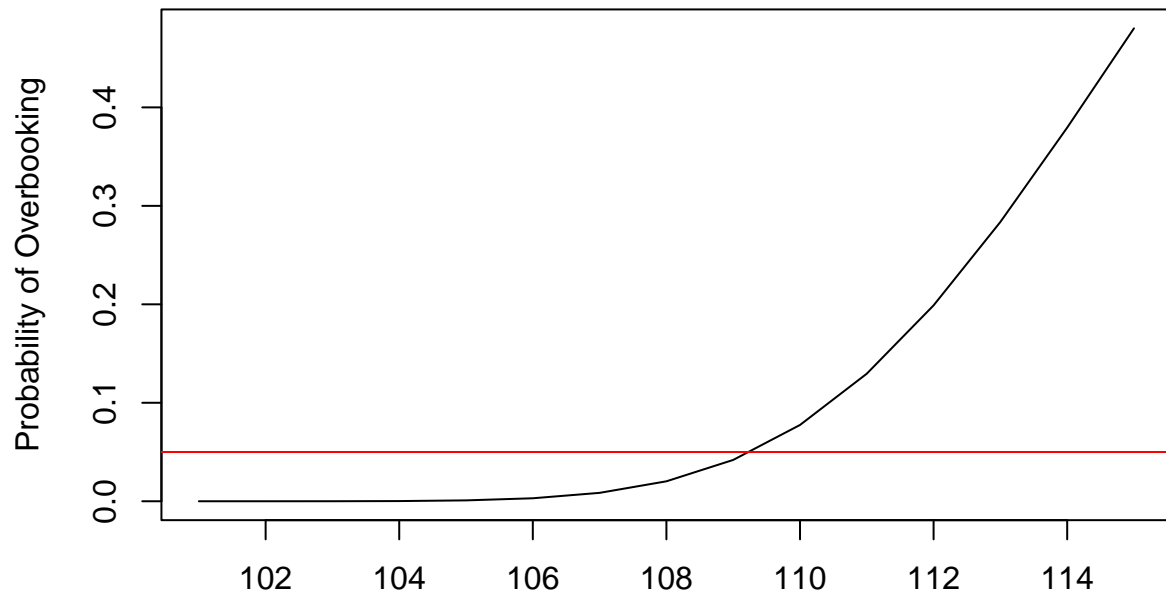
```
## [1] 0.07747508
```

There is a 7.74% chance that at least one too many passengers will show up to this flight.

7) Assume that airlines only want a 5% risk that at least one too many people show up for the flight and are unable to board. How many tickets should they sell? Plot the risk of overbooking from 101-115 tickets.

```
b <- 101:115
z <- pbinom(101, b, .88, lower.tail = F)
plot(b, z,
     type = "l",
     main = "Risk of Overbooking by Number of Tickets sold",
     xlab = "Number of Tickets sold",
     ylab = "Probability of Overbooking")
abline(h = .05, col = "red")
```

## Risk of Overbooking by Number of Tickets sold



Number of Tickets sold

They

should sell 109 tickets to get a 5% chance of overbooking.