Chapter 6

The Binomial Distribution

In this chapter, we will discuss the following topics:

- We will look at the Binomial density with the R-function **dbinom**.
- We will look at the Cumulative Binomial distribution function with the R-function **pbinom**.
- We will consider the Normal approximation to the Binomial distribution which follows from the Central limit theorem.

We have three arguments to the functions **dbinom**(x, n, p) and **pbinom**(x, n, p), where x is the number of successes, n is the number of independent trials, and p is the success probability. Notice the following, where X follows a Binomial distribution with n trials and success probability p:

- The function **dbinom(x,n,p)** computes $P(X = x)$, which is the probability for exactly x successes.
- The function **pbinom(x,n,p,lower.tail=FALSE)** computes $P(X > x)$, which is the probability for more than x successes.
- The function **pbinom(x,n,p)** computes $P(X \leq x)$, which is the probability for at most x successes. (Here **lower.tail=TRUE** which is the default in R).

The Binomial Distribution

If X has the Binomial distribution with n independent trials and success probability p, we have that the probability for exactly k successes is given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, ..., n$.

Problem. Use R to find $\binom{10}{3}$

Solution. We obtain:

```
> choose(10,3)
[1] 120
```

so $\binom{10}{3} = 120$.

Problem. Suppose that X has the binomial distribution with $n = 20$ trials and success probability $p = 0.3$. Find $P(X = 4)$ which is the probability for exactly 4 successes in 20 trials.

Solution. We obtain:

```
> dbinom(4,20,0.3)
[1] 0.130421
```

so $P(X = 4) = 0.1304$.

Problem. During May 28th and September 13th the chances of precipitation in Tucson for any given day is on average 37% [2]. Pick 10 random days between May 28th and September 13th in Tucson.
(a) What is the probability that exactly 3 of the days have precipitation?
(b) What is the probability that at most 3 of the days have precipitation?
(c) What is the probability that at least 3 of the days have precipitation?

Solution to part (a). Let X be the number of days that have precipitation. X follows a binomial distribution with $n = 10$ trials and success probability $p = 0.37$. We want to find $P(X = 3)$:

```
> dbinom(3,10,0.37)
[1] 0.2394254
```

Hence, $P(X = 3) = 0.24$.

Solution to part (b). We want to find $P(X \leq 3)$:

```
> pbinom(3,10,0.37)
[1] 0.4599962
```

Hence, $P(X \leq 3) = 0.46$.

Solution to part (c). We want to find $P(X \geq 3)$. Notice that since X is a discrete random variable, this is the same as $P(X > 2)$:

```
> pbinom(2,10,0.37,lower.tail=FALSE)
[1] 0.7794292
```

Hence, $P(X \geq 3) = 0.78$. We could also have done the calculation this way in R using the fact that $P(X \geq 3) = 1 - P(X \leq 2)$:

```
> 1-pbinom(2,10,0.37)
[1] 0.7794292
```

The Normal Approximation to the Binomial Distribution

Suppose that X follows a binomial distribution with n independent observations and success probability p. The distribution of X is approximately Normal with mean np and standard deviation $\sqrt{np(1-p)}$ when n is sufficiently large. As a rough guide, we will use the Normal approximation when $np \geq 5$ and $n(1-p) \geq 5$.

Problem. The National Immunization Survey [1] estimated that the vaccine coverage for MMR (measles, mumps, rubella) vaccine among children aged 19-35 months in Arizona was 86.7% in 2011.
The national Healthy People 2020 target of MMR coverage is 90%. [1]
(a) In a sample of 1000 randomly selected children aged 19-35 months from Arizona, use the Normal approximation to the Binomial distribution to compute the probability that at least

900 of them are vaccinated against MMR?
(b) In a sample of 1000 randomly selected children aged 19-35 months from Arizona, compute the exact probability that at least 900 of them are vaccinated against MMR?
(c) The U.S. National vaccination coverage for MMR among children aged 19-35 months in 2011 was 91.6% [1]. In a sample of 1000 randomly selected U.S. children aged 19-35 months, use both the Normal approximation to the Binomial distribution and the exact Binomial distribution to compute the probability that at least 900 of them are vaccinated against MMR?

Solution to part (a). Let X be the number of children aged 19-35 months who are vaccinated for MMR. X follows a binomial distribution with n = 1000 and p = 0.867. We have $\mu = np = (1000) * (0.867) = 867$ and $\sigma = \sqrt{np(1-p)} = \sqrt{(1000)(0.8678)(1-0.867)} = 10.74325$. Since we have $np \geq 5$ and $n(1-p) \geq 5$, we can use the Normal approximation to the Binomial approximation. We want to find the probability that the number of children aged 19-35 months who are vaccinated for MMR is greater than or equal to 900, that is $P(X \geq 900)$. Using the continuity correction, we have that $P(X \geq 900) = P(X \geq 899.5)$, where by the Central limit theorem, X is approximately Normal with mean 867 and standard deviation 10.74325. We obtain:

```
> pnorm(899.5,867,10.74325,lower.tail=FALSE)
[1] 0.001242527
```

Hence, $P(X \geq 900) = 0.00124$ so the approximate probability that at least 900 of the children in the sample are vaccinated for MMR is 0.124%.

Solution to part (b). We obtain:

```
> pbinom(899,1000,0.867,lower.tail=FALSE)
[1] 0.0008684169
```

Hence the exact probability is 0.08684. The normal approximation to the Binomial distribution is 0.00037 units too high.

Solution to part (c). We have $\mu = np = (1000) * (0.916) = 916$ and $\sigma = \sqrt{np(1-p)} = \sqrt{(1000)(0.916)(1-0.916)} = 8.77177$. Since we have $np \geq 5$ and $n(1-p) \geq 5$, we can use the Normal approximation to the Binomial distribution:

```
> pnorm(899.5,916,8.77177,lower.tail=FALSE)
[1] 0.9700164
> pbinom(899,1000,0.916,lower.tail=FALSE)
[1] 0.9674523
```

The Normal approximation to the Binomial distribution gives 97% probability that there are 900 or more U.S. children aged 19-35 months in the sample who are vaccinated for MMR. The corresponding exact probability is 96.7%.

Explanation. The code can be explained as follows:

- The function **pnorm(x, μ, σ, lower.tail=FALSE)** computes $P(X \geq x) = P(X > x)$.

References

[1] Centers for Disease Control and Prevention. *National, State, and Local Area Vaccination Coverage Among Children Aged 1935 Months - United States, 2011.* MMWR, Weekly Vol. 61, No.35, 690-696, 2012.

[2] http://weatherspark.com/averages/31809/Tucson-Arizona-United-States

00