9.4 Normal Approximation to the Binomial Distribution

I. Characteristics of a Binomial Experiment

Experiments which have the following characteristics are called Binomial Experiments or Bernoulli Processes.

1. The same experiment is repeated many (n) times under the same conditions. Each repetition is called a trial or a Bernoulli trial.
2. Only the same two outcomes (success and failure) are possible each time the experiment is performed.
3. Each time the experiment is performed, the probabilities for the two outcomes remain the same. We use p for the probability of success and q = 1 – p for the probability of failure.
4. The trials are independent. The outcome of one trial has no influence on the outcome of another trial.
5. We are interested in the total number of successes, not the order in which they occur. There may be 0, 1, 2, 3, …, or n successes in n trials.

II. The General Formula for the Probability of a Binomial Experiment

Given a binomial experiment with n independent repeated trials, if p is the probability of success in a single trial, q = 1 – p is the probability of failure in a single trial, and x is the number of successes (with x ≤ n), then the probability of x successes in n trials is given by the formula:

\[ P(x \text{ successes in } n \text{ trials}) = C(n, x) \cdot p^x \cdot q^{n-x}. \]

III. The Binomial Distribution

The probability distribution for a binomial experiment is known as a binomial distribution. Such distributions are often represented graphically as histograms in which the area of each vertical rectangle is proportionate to its corresponding probability. For the following series of histograms, p = .3 and the number of trials (n) increases from 5 to 10 to 20 to 40. Notice that as n increases, the shape of the histogram more closely approximates the shape of the area under a normal curve.

IV. Normal Approximation to the Binomial Distribution

If we conduct a binomial experiment and the number of trials (n) is large and the probability of success (p) is not too close to 0 or 1, the binomial distribution may be approximated by a normal distribution with a mean of \( \mu = n \cdot p \) and a standard deviation of \( \sigma = \sqrt{n \cdot p \cdot q} \).

As a general rule, the normal curve approximation can be used as long as n·p and n·q are both at least 5.

Note: The normal distribution is continuous since the random variable can take on the value of any real number. The binomial distribution is discrete since the random variable can take on only integer values between 0 and n. Nevertheless, the normal distribution can be used to give a good approximation of binomial probability.
Example 1

Suppose a fair coin is flipped 15 times.

a. Find the mean and standard deviation for the number of heads.

b. Make a probability distribution for the number of heads and use it to construct a probability histogram.

c. Find the probability of getting exactly 9 heads using the binomial probability formula.

d. Find the probability of getting exactly 9 heads using the normal curve approximation.

a. \[ \mu = np = 15 \times 0.5 = 7.5 \]

\[ \sigma = \sqrt{npq} = \sqrt{15 \times 0.5 \times 0.5} = \sqrt{3.75} \approx 1.9365 \]

b. 

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00003</td>
</tr>
<tr>
<td>1</td>
<td>0.00046</td>
</tr>
<tr>
<td>2</td>
<td>0.00320</td>
</tr>
<tr>
<td>3</td>
<td>0.01389</td>
</tr>
<tr>
<td>4</td>
<td>0.04166</td>
</tr>
<tr>
<td>5</td>
<td>0.09164</td>
</tr>
<tr>
<td>6</td>
<td>0.15274</td>
</tr>
<tr>
<td>7</td>
<td>0.19638</td>
</tr>
<tr>
<td>8</td>
<td>0.19638</td>
</tr>
<tr>
<td>9</td>
<td>0.15274</td>
</tr>
<tr>
<td>10</td>
<td>0.09164</td>
</tr>
<tr>
<td>11</td>
<td>0.04166</td>
</tr>
<tr>
<td>12</td>
<td>0.01389</td>
</tr>
<tr>
<td>13</td>
<td>0.00320</td>
</tr>
<tr>
<td>14</td>
<td>0.00046</td>
</tr>
<tr>
<td>15</td>
<td>0.00003</td>
</tr>
</tbody>
</table>

To find \( P(9H) \) using the normal curve approximation, we must find the area under the curve from \( x = 8.5 \) to \( x = 9.5 \). We use 8.5 and 9.5 since these are the left and right boundary values for the histogram bar centered over 9.

To find this area we must first find the \( z \)-scores for 8.5 and 9.5. Then we use the table on A1 to find the area to the left of each of these scores. And finally we find the area between them by subtracting the smaller area from the larger area.

If \( x_1 = 8.5 \), then \( z_1 = \frac{8.5 - 7.5}{1.9365} \approx 0.52 \) and \( A_1 \approx 0.6985 \)

If \( x_2 = 9.5 \), then \( z_2 = \frac{9.5 - 7.5}{1.9365} \approx 1.03 \) and \( A_2 \approx 0.8485 \)

Thus \( P(9H) = P(8.5 \leq x \leq 9.5) = P(0.52 \leq z \leq 1.03) \approx 0.8485 - 0.6985 = 0.1500 \)

Our answer from part \( c \) is more accurate, but this approximation is still quite good.

Note: To find \( P(X) \) using the normal curve approximation, we must find \( P(x_1 \leq X \leq x_2) \) where \( x_1 = X - 0.5 \) and \( x_2 = X + 0.5 \)
Example 2
A flu vaccine has a probability of 80% of preventing a person who is inoculated from getting the flu. A county health office inoculates 134 people. Use the normal curve approximation to find the probabilities of the following.

a. Exactly 100 of the inoculated people get the flu
b. No more than 100 of the inoculated people get the flu
c. None of the inoculated people get the flu
d. Between 100 and 120 inoculated people (inclusive) get the flu

a. \( n = 134, x = 100, p = .8 \)

\[ \mu = 134(.8) = 107.2 \quad \sigma = \sqrt{134 \times .8 \times .2} = \sqrt{21.44} \approx 4.6303 \]

If \( x_1 = 99.5 \), then

\[ z_1 = \frac{99.5 - 107.2}{4.6303} = -1.66 \quad \text{and} \quad A_1 \approx .0485 \]

If \( x_2 = 100.5 \), then

\[ z_2 = \frac{100.5 - 107.2}{4.6303} \approx -1.45 \quad \text{and} \quad A_2 \approx .0735 \]

\[ P(100 \text{ get flu}) = P(99.5 \leq x \leq 100.5) = P(-1.66 \leq z \leq -1.45) = .0735 - .0485 = .025 \]

b. \( P(\text{no more than 100 get flu}) = P(x \leq 100.5) = P(z \leq -1.45) = .0735 \)

c. To find the probability that none of the inoculated people get the flu, we will use the Normal Cumulative Density Function, normalcdf, on our calculator. It is located under the DISTR menu.

If you are entering z-scores, the information is input as normalcdf\( (z_1, z_2) \).

If you are entering x-values, the information is input as normalcdf\( (x_1, x_2, \mu, \sigma) \).

\[ P(0 \text{ get the flu}) = \text{normalcdf}(-.5, .5, 107.2, 4.6303) = 0 \]

d. If \( x_1 = 99.5 \), then

\[ z_1 = \frac{99.5 - 107.2}{4.6303} = -1.66 \quad \text{and} \quad A_1 \approx .0485 \]

If \( x_2 = 120.5 \), then

\[ z_2 = \frac{120.5 - 107.2}{4.6303} \approx 2.87 \quad \text{and} \quad A_2 \approx .9979 \]

\[ P(100 - 120 \text{ get flu}) = P(99.5 \leq x \leq 120.5) = P(-1.66 \leq z \leq 2.87) = .9979 - .0485 = 94.94 \]

We can check our answer on our calculator by entering

\[ \text{normalcdf}(99.5, 120.5, 107.2, 4.6303) \] enter \( \rightarrow .9498 \)

Note: When we use the \text{normalcdf} function to find the area under a normal curve, the result is more accurate than when we use a z-table. This is due to rounding. The z-table rounds the areas to 4 decimal places while the calculator rounds to 10 decimal places. This can cause our answers from the calculator to be slightly different than our answers from the z-table.