Search this site...

STAT 414   Introduction to Probability Theory

28.1 - Normal Approximation to Binomial

As the title of this page suggests, we will now focus on using the normal distribution to approximate binomial probabilities. The Central Limit Theorem is the tool that allows us to do so. As usual, we'll use an example to motivate the material.

Example 28-1

Let X_i denote whether or not a randomly selected individual approves of the job the President is doing. More specifically:

- Let $X_i = 1$, if the person approves of the job the President is doing, with probability p

- Let $X_i = 0$, if the person does not approve of the job the President is doing with probability $1 - p$

Then, recall that X_i is a Bernoulli random variable with mean:

$$\mu = E(X) = (0)(1-p) + (1)(p) = p$$

and variance:

$$\sigma^2 = Var(X) = E[(X-p)^2] = (0-p)^2(1-p) + (1-p)^2(p) = p(1-p)[p+1-p] = p(1-p)$$

Now, take a random sample of n people, and let:

$$Y = X_1 + X_2 + \ldots + X_n$$

Then Y is a binomial(n, p) random variable, $y = 0, 1, 2, \ldots, n$, with mean:

$$\mu = np$$

and variance:

$$\sigma^2 = np(1-p)$$

Now, let $n = 10$ and $p = \frac{1}{2}$, so that Y is binomial$(10, \frac{1}{2})$. What is the probability that exactly five people approve of the job the President is doing?

Solution

There is really nothing new here. We can calculate the exact probability using the binomial table in the back of the book with $n = 10$ and $p = \frac{1}{2}$. Doing so, we get:

$$P(Y = 5) = P(Y \leq 5) - P(Y \leq 4)$$
$$= 0.6230 - 0.3770$$
$$= 0.2460$$

That is, there is a 24.6% chance that exactly five of the ten people selected approve of the job the President is doing.

Note, however, that Y in the above example is defined as a sum of independent, identically distributed random variables. Therefore, as long as n is sufficiently large, we can use the Central Limit Theorem to calculate probabilities for Y. Specifically, the Central Limit Theorem tells us that:

$$Z = \frac{Y - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

Let's use the normal distribution then to approximate some probabilities for Y. Again, what is the probability that exactly five people approve of the job the President is doing?
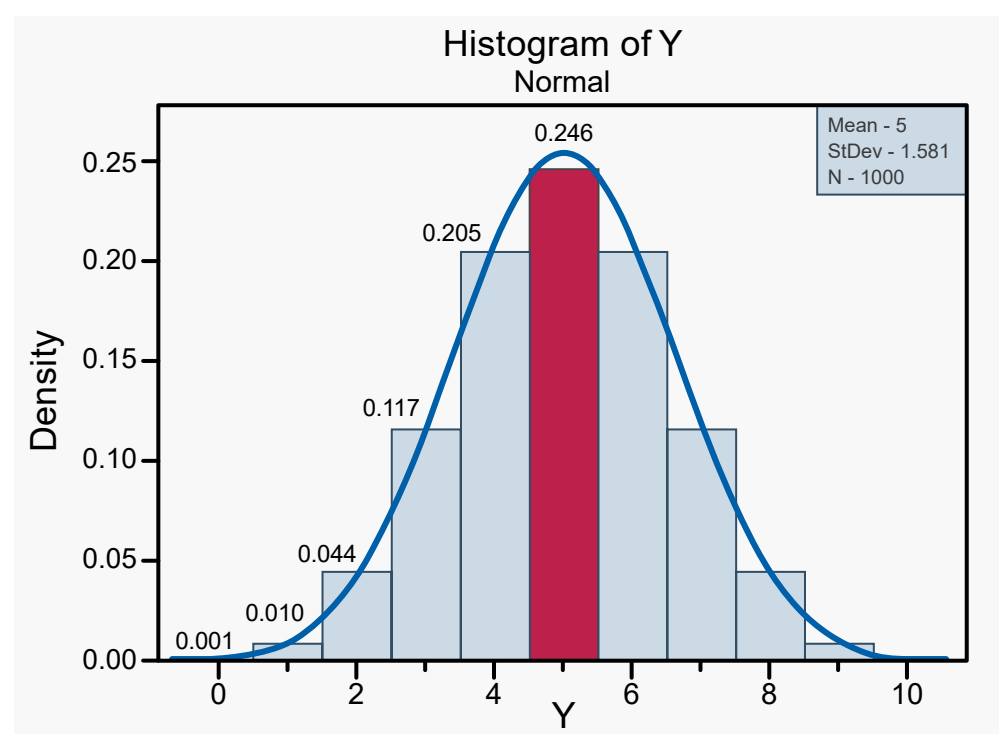
Solution

First, recognize in our case that the mean is:

$$\mu = np = 10\left(\frac{1}{2}\right) = 5$$

and the variance is:

$$\sigma^2 = np(1-p) = 10\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 2.5$$

Now, if we look at a graph of the binomial distribution with the rectangle corresponding to $Y = 5$ shaded in red:

we should see that we would benefit from making some kind of correction for the fact that we are using a continuous distribution to approximate a discrete distribution. Specifically, it seems that the rectangle $Y = 5$ really includes any Y greater than 4.5 but less than 5.5. That is:

$$P(Y = 5) = P(4.5 < Y < 5.5)$$

Such an adjustment is called a "**continuity correction**." Once we've made the continuity correction, the calculation reduces to a normal probability calculation:

Now, recall that we previous used the binomial distribution to determine that the probability that $Y = 5$ is exactly 0.246. Here, we used the normal distribution to determine that the probability that $Y = 5$ is approximately 0.251. That's not too shabby of an approximation, in light of the fact that we are dealing with a relative small sample size of $n = 10$!

Let's try a few more approximations. What is the probability that more than 7, but at most 9, of the ten people sampled approve of the job the President is doing?

Solution

If we look at a graph of the binomial distribution with the area corresponding to $7 < Y \leq 9$ shaded in red:
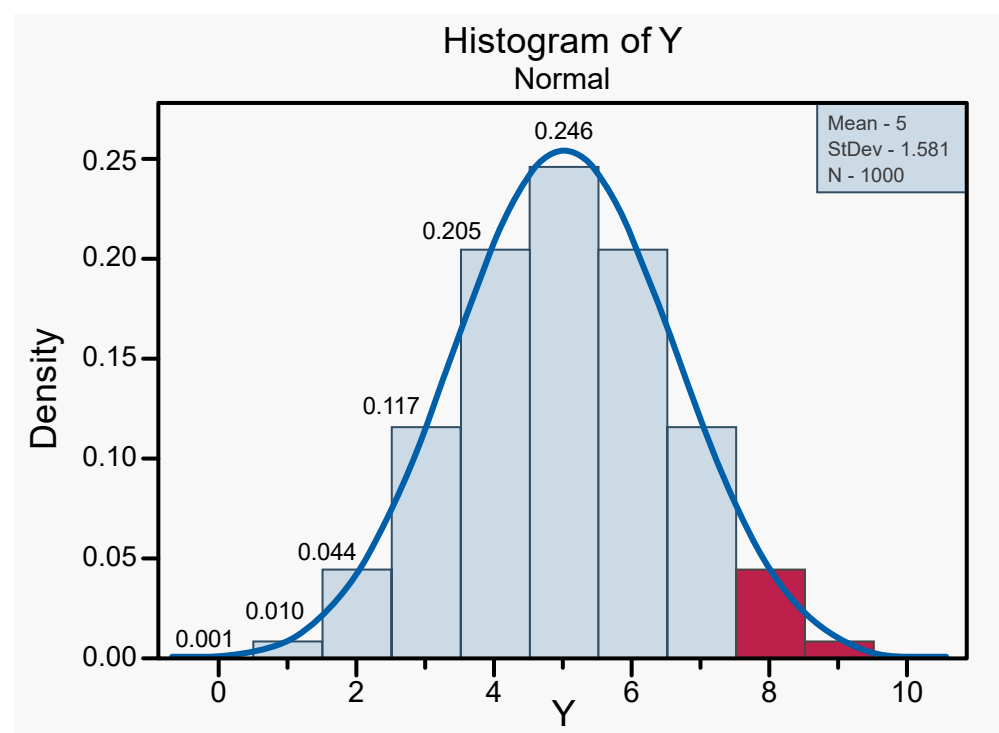
we should see that we'll want to make the following continuity correction:

$$P(7 < Y \leq 9) = P(7.5 < Y < 9.5)$$

Now again, once we've made the continuity correction, the calculation reduces to a normal probability calculation:


By the way, you might find it interesting to note that the approximate normal probability is quite close to the exact binomial probability. We showed that the approximate probability is 0.0549, whereas the following calculation shows that the exact probability (using the binomial table with $n = 10$ and $p = \frac{1}{2}$ is 0.0537:

$$P(7 < Y \leq 9) = P(Y \leq 9) - P(Y \leq 7) = 0.9990 - 0.9453 = 0.0537$$

Let's try one more approximation. What is the probability that at least 2, but less than 4, of the ten people sampled approve of the job the President is doing?

Solution

If we look at a graph of the binomial distribution with the area corresponding to $2 \leq Y < 4$ shaded in red:
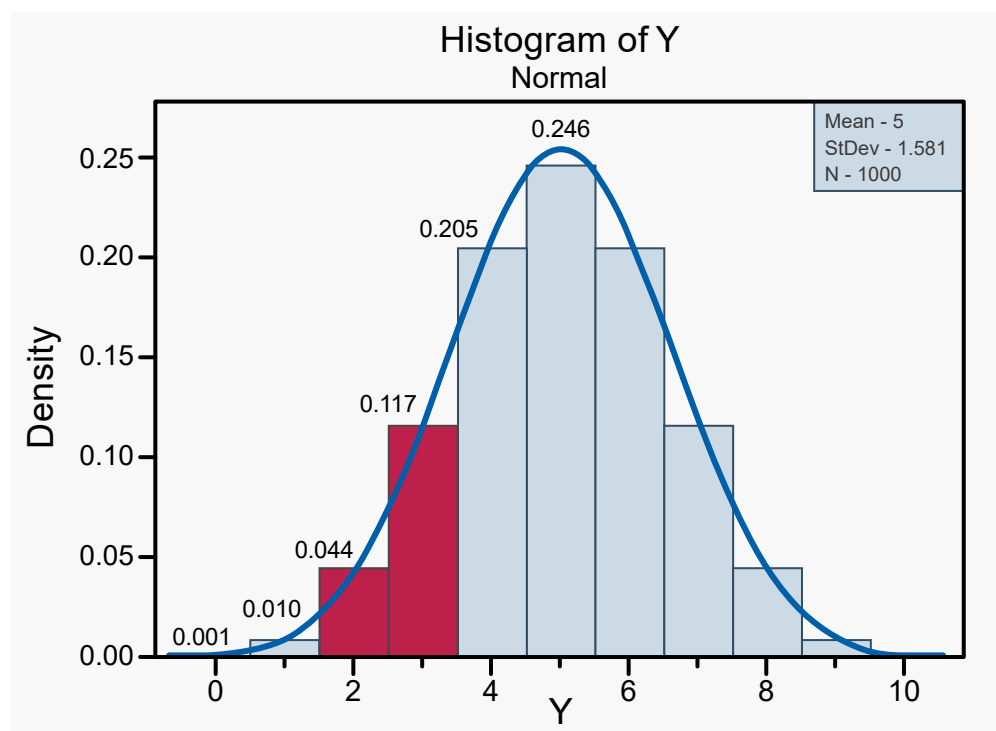
we should see that we'll want to make the following continuity correction:

$$P(2 \leq Y < 4) = P(1.5 < Y < 3.5)$$

Again, once we've made the continuity correction, the calculation reduces to a normal probability calculation:

$$
\begin{aligned}
P(2 \leq Y < 4) = P(1.5 < Y < 3.5) &= P(\frac{1.5 - 5}{\sqrt{2.5}} < Z < \frac{3.5 - 5}{\sqrt{2.5}}) \\
&= P(-2.21 < Z < -0.95) \\
&= P(Z > 0.95) - P(Z > 2.21) \\
&= 0.1711 - 0.0136 = 0.1575
\end{aligned}
$$

By the way, the exact binomial probability is 0.1612, as the following calculation illustrates:

$$P(2 \leq Y < 4) = P(Y \leq 3) - P(Y \leq 1) = 0.1719 - 0.0107 = 0.1612$$

Just a couple of comments before we close our discussion of the normal approximation to the binomial.

(1) First, we have not yet discussed what "sufficiently large" means in terms of when it is appropriate to use the normal approximation to the binomial. The general rule of thumb is that the sample size n is "sufficiently large" if:

$$np \geq 5 \text{ and } n(1-p) \geq 5$$

For example, in the above example, in which $p = 0.5$, the two conditions are met if:

$$np = n(0.5) \geq 5 \text{ and } n(1-p) = n(0.5) \geq 5$$

Now, both conditions are true if:

$$n \geq 5\left(\frac{10}{5}\right) = 10$$

Because our sample size was at least 10 (well, barely!), we now see why our approximations were quite close to the exact probabilities. In general, the farther p is away from 0.5, the larger the sample size n is needed. For example, suppose $p = 0.1$. Then, the two conditions are met if:

$$np = n(0.1) \geq 5 \text{ and } n(1-p) = n(0.9) \geq 5$$

Now, the first condition is met if:

$$n \geq 5(10) = 50$$

And, the second condition is met if:

$$n \geq 5 \left(\tfrac{10}{9} \right) = 5.5$$

That is, the only way both conditions are met is if $n \geq 50$. So, in summary, when $p = 0.5$, a sample size of $n = 10$ is sufficient. But, if $p = 0.1$, then we need a much larger sample size, namely $n = 50$.

(2) In truth, if you have the available tools, such as a binomial table or a statistical package, you'll probably want to calculate exact probabilities instead of approximate probabilities. Does that mean all of our discussion here is for naught? No, not at all! In reality, we'll most often use the Central Limit Theorem as applied to the sum of independent Bernoulli random variables to help us draw conclusions about a true population proportion p. If we take the Z random variable that we've been dealing with above, and divide the numerator by n and the denominator by n (and thereby not changing the overall quantity), we get the following result:

$$Z = \frac{\sum X_i - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \xrightarrow{d} N(0,1)$$

The quantity:

$$\hat{p} = \frac{\sum\limits_{i=1}^{n} X_i}{n}$$

that appears in the numerator is the "**sample proportion**," that is, the proportion in the sample meeting the condition of interest (approving of the President's job, for example). In Stat 415, we'll use the sample proportion in conjunction with the above result to draw conclusions about the unknown population proportion p. You'll definitely be seeing much more of this in Stat 415 !

« Previous   Next »