Approximating the Binomial using the Normal Distribution

Updated: March 2025

Example

Statistics held by the Road Safety Division of the Police show that 78% of drivers being tested for their licence pass at the first attempt.

If a group of 120 drivers are tested in one centre in a year, find the probability that more than 99 pass at the first attempt.

This is a binomial question with: $n = 120$, $p = 0.78$ and $P(X > 99)$ required.

Why not use the binomial distribution here?

Using the binomial formula would require calculations:

$$P(X = 100) + P(X = 101) + P(X = 102) + \ldots + P(X = 120).$$

That is

$$^{120}C_{100} \times 0.78^{100} \times 0.22^{20} + ^{120}C_{101} \times 0.78^{101} \times 0.22^{19} + ^{120}C_{102} \times 0.78^{102} \times 0.22^{18} + \ldots$$

This is difficult and time consuming to calculate.

In situations like this, that is, when

- n is large, and
- p is not too close to 0 or 1,

the normal distribution is a close approximation to the binomial distribution.

The conditions that ensure this are: $n \times p \geq 5$ **AND** $n(1 - p) \geq 5$

Check

In our example, $np = 120 \times 0.78 = 93.6$ and $n(1 - p) = 120 \times 0.22 = 26.4$.

Both ≥ 5, so we can use the normal approximation.

Convert

To use the normal distribution, we need a *mean* and a *standard deviation.*

We use the ***binomial*** mean and standard deviation:

In our example, $\bar{x} = np = 120 \times 0.78 = 93.6$

$$s = \sqrt{np(1 - p)} = \sqrt{(120 \times 0.78 \times 0.22)} = 4.538$$

We are approximating a discrete distribution with a continuous one, and so we must use a continuity correction.

Since there are no values between discrete (counting) numbers, half the “space” between the digits on the number line is “given” to the lower value and half to the upper value.

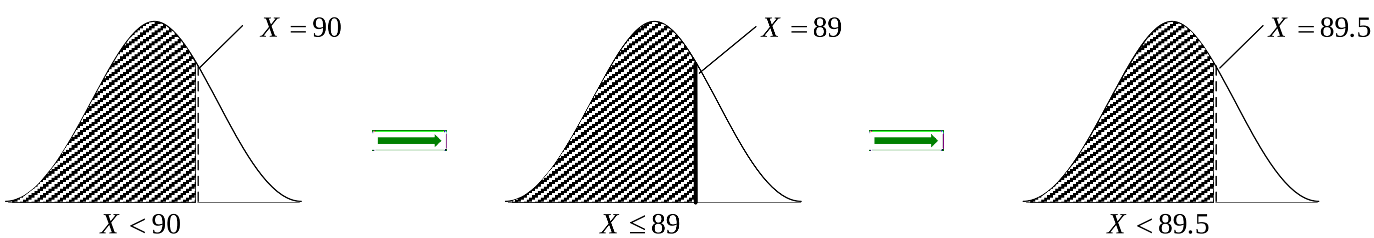
To do this,

- Draw the diagram and shade the required area.
- Use a broken line for the “boundary” if the probability is an inequality
  - >, greater than or <, less than

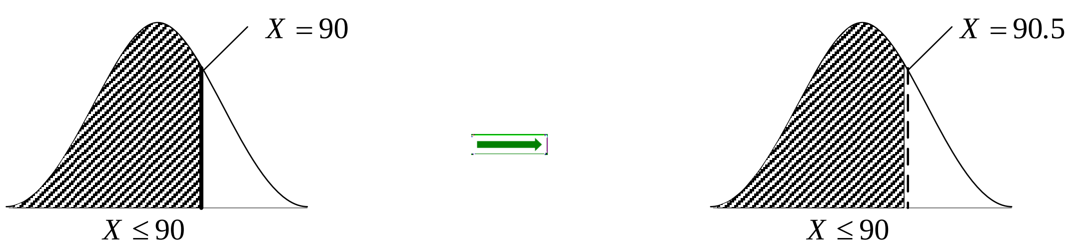
  and a solid line otherwise
  - ≥, greater than or equal to or ≤, less than or equal to
- Adjust any inequality by extending to the next “included” integer
  - For example, >2 becomes ≥ 3, < 5 becomes ≤ 4
- The shaded area is now ***extended*** by half a unit

The following four examples demonstrate this process. (Assume a mean of 80)

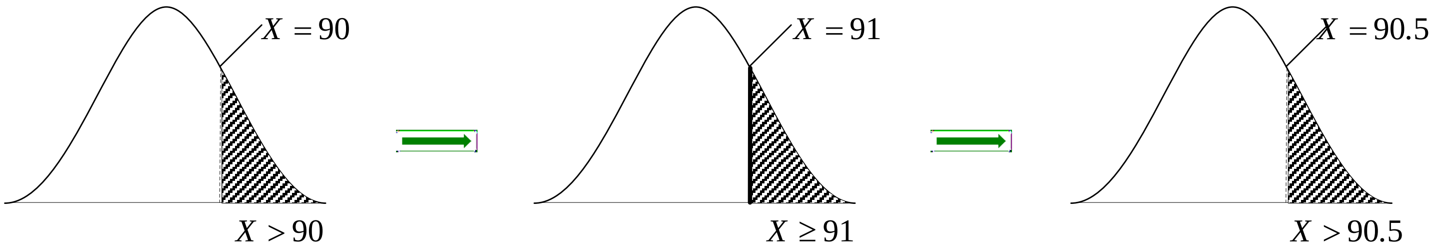
Example 1



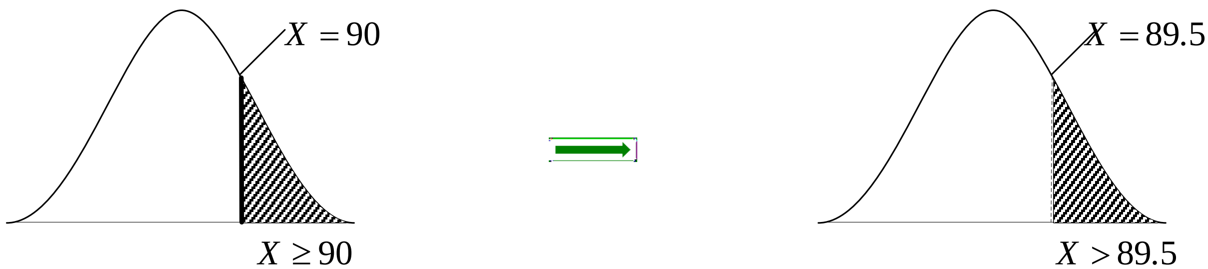
Example 2



Example 3



Example 4



Note: Each situation needs careful analysis, but in general, if shading is to the left, extend to the right and if shading is to the right, extend to the left.

So, continuing, the process is still:

1. Draw and label a diagram
2. Shade required area
3. Change raw score to standard (Z_{score}):
4. Look up $P(Z < z)$ in tables
5. Adjust probability from table to required area

Practise:

1. For a binomial distribution, X , with $n = 1000$ and $p = 0.5$, use the normal distribution to calculate:

   (a) $P(X \leq 510)$ (b) $P(X > 495)$

2. On the average about 90% of the prospective guests at a hotel arrive on a given night arrive. The hotel has 200 rooms and 215 reservations have been accepted for one night. The problem is to find the probability that all guests can be accommodated this night.

   a) The number, *X*, of guests arriving at the hotel is assumed to be a binomial distribution with parameters, *n* = 215 and π = 0.9. Find the mean and variance of *X* and state the assumption needed for the arrivals to be binomial.

   (b) Using the normal approximation to the binomial distribution firn the probability that all guests can be accommodated.

3. Data recorded over recent years show that one in every fifty male adults develops a serious respiratory illness.

   (a) In a random sample of *n* = 400 male adults, find, using results for a binomial distribution, the mean and standard deviation of the number who will suffer from this respiratory illness.

   (b) A random sample of 400 workers in a certain occupation shows 15 with the respiratory illness. Find, using the normal approximation to the binomial distribution, the probability that 15 or more in a sample of 400 will suffer from the respiratory illness. What conclusion do you draw from this probability?

4. 2% of all births in New Zealand are twins. If there are 500 births in one week, calculate the following:

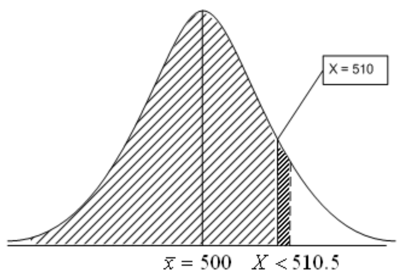
   (a) The probability that more than 10 births in one week would result in twins.

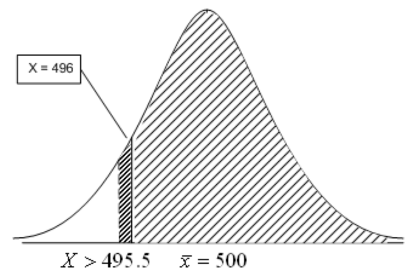
   (b) The probability that at least 5 births result in twins.

5. The probability that a new restaurant in Christchurch fails in the first year of operation is 0.1. If 15 new restaurants are sampled, use Normal approximation to calculate the probability that more than 5 out of the 15 fail.

Answers:

1. $$n = 1000,\ \ p = 0.5 \ \Rightarrow \bar{x} = np = 500,\ \ \ s = \sqrt{np(1-p)} = 15.81$$



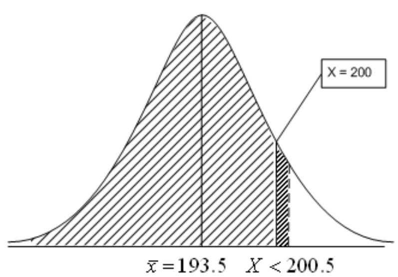


a)
$$P(X \le 510) = P(X < 510.5)$$
$$ie\ \ P\left(Z < \frac{510.5 - 500}{15.81}\right) = P(Z < 0.66)$$
$$= 0.7454$$

b)
$$P(X > 495) = P(X \ge 496) = P(X > 495.5)$$
$$ie\ \ P\left(Z > \frac{495.5 - 500}{15.81}\right) = 1 - P(Z < -0.29)$$
$$= 1 - 0.3859 = 0.6151$$

2. a) $$n = 215,\ \ p = 0.9 \ \Rightarrow \bar{x} = np = 193.5,\ \ \ s = \sqrt{np(1-p)} = 4.399$$

b)

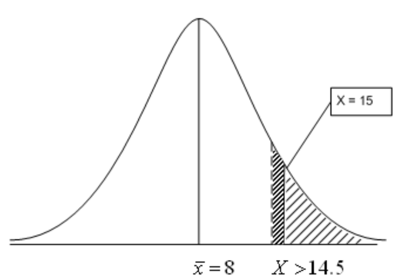


$$P(X \le 200) = P(X < 200.5)$$
$$ie\ \ P\left(Z < \frac{200.5 - 193.5}{4.399}\right) = P(Z < 1.59) = 0.9441$$

ie it is 94% certain that all guests can be accommodated in any one night.

3. a) $$n = 400,\ \ p = 0.02 \ \Rightarrow \bar{x} = np = 8,$$
$$s = \sqrt{np(1-p)} = 2.8$$

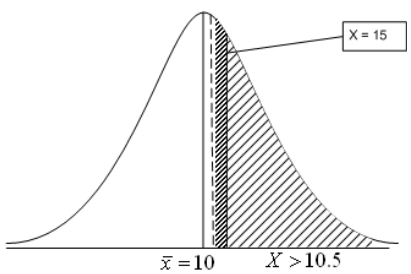
b)



$$P(X \ge 15) = P(X > 14.5)$$
$$ie\ \ P\left(Z > \frac{14.5 - 8}{2.8}\right) = 1 - P(Z < 2.32)$$
$$= 1 - 0.9898 = 0.0102$$

4. $n = 500, \; p = 0.02 \Rightarrow \bar{x} = np = 10, \quad s = \sqrt{np(1-p)} = 3.13$

a)



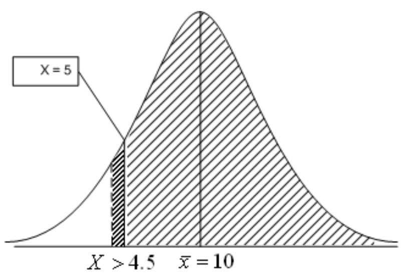
$P(X > 10) = P(X > 10.5)$

$ie \;\; P\left(Z > \frac{10.5 - 10}{3.13}\right) = 1 - P(Z < -0.16)$

$= 1 - 0.5636 = 0.4364$

Note: $P(X > 10) = P(X \geq 11) = P(X > 10.5)$

b)



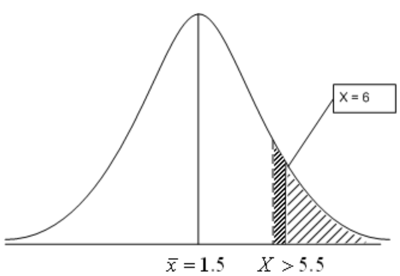
b) $P(X \geq 5) = P(X > 4.5)$

$ie \;\; P\left(Z > \frac{4.5 - 10}{3.13}\right) = 1 - P(Z < -1.76)$

$= 1 - 0.0392 = 0.9608$

5.

$n = 15, \; p = 0.1 \Rightarrow \bar{x} = np = 1.5, \quad s = \sqrt{np(1-p)} = 1.16$



$P(X > 5) = P(X \geq 6) = P(X > 5.5)$

$ie \;\; P\left(Z > \frac{5.5 - 1.5}{1.16}\right) = 1 - P(Z < 3.45) = 1 - 0.99971 = 0.00029$