L11. The Normal Approximation to the Binomial Distribution

Example 1: Gum Sales

Chewing gum sales have dropped by 15% over the past decade. One theory is that shoppers are too distracted by their smartphones to notice gum at the checkout. Let's suppose that the probability a shopper notices the gum is 0.4.

a. In a small observational study, 6 shoppers are monitored. What is the probability that 4 or 5 shoppers notice the gum?

b. In a different store, 10 shoppers were observed. What is the probability that at least 1 shopper notices the gum?

c. In a busy super market, 70 shoppers were observed. What is the probability that

   i. 30 or more shoppers notice the gum?

   ii. less than 15 notice the gum?

   iii. between 40 and 60 shoppers (inclusive) notice the gum?

Solution

Let the probability a shopper notices the gum be $p = 0.4$.

a. Let X = the number of customers who notice the gum.

   X is a binomial random variable with $n = 6$ and $p = 0.4$

$$
\begin{aligned}
P(X = 4 \cup X = 5) &= P(X = 4) + P(X = 5) \\
&= C_4^6 (0.4)^4 \cdot (0.6)^2 + C_5^6 (0.4)^5 \cdot (0.6) \\
&= 0.1382 + 0.0369 \\
&= 0.1751
\end{aligned}
$$

b. Want $P(X \geq 1)$ with $n = 10$

$$\begin{aligned} P(X \geq 1) &= P(X=1) + P(X=2) + P(X=3) + \cdots + P(X=10) \\ &= 1 - P(X=0) \\ &= 1 - C_0^{10}(0.4)^0 \cdot (0.6)^{10} \\ &= 1 - 0.0060 \\ &= 0.9940 \end{aligned}$$

c. i. Want $P(X \geq 30)$ with $n = 70$.

$$P(X \geq 30) = P(X=30) + P(X=31) + P(X=32) + \cdots + P(X=70)$$

Since this requires a lot of calculations, lets see if we can streamline the calculation using a normal approximation to the binomial.

Check: $n \cdot p = 70 \cdot (0.4) = 28 > 5$
$n \cdot (1-p) = 70 \cdot (0.6) = 42 > 5$
$\Rightarrow$ Yes we can.

The mean and standard deviation of a binomial random variable with $n = 70$ and $p = 0.4$ is

$$\mu = n \cdot p = 70 \cdot (0.4) = 28$$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)}\sqrt{70 \cdot (0.4) \cdot (0.6)} = \sqrt{16.8}$$

$$\begin{aligned} P(X \geq 30) &= P(X > 29.5) && \text{Add correction factor} \\ &= P\left(Z > \frac{X-\mu}{\sigma}\right) \\ &= P\left(Z > \frac{29.5-28}{\sqrt{16.8}}\right) \\ &= P(Z > 0.37) \\ &= 0.3557 \end{aligned}$$

ii. Want $P(Z < 15)$

$$\begin{aligned} P(X < 15) &= P(X < 14.5) && \text{Add correction factor} \\ &= P\left(Z < \frac{14.5-28}{\sqrt{16.8}}\right) \\ &= P(Z < -3.29) \\ &= 0.0005 \end{aligned}$$

iii. Want $P(40 \leq Z \leq 60) \approx P(39.5 < X < 60.5)$

$$\begin{aligned} P(40 \leq X \leq 60) &= P(39.5 < X < 60.5) \qquad \text{Add correction factors} \\ &= P\left(\frac{39.5 - 28}{\sqrt{16.8}} < Z < \frac{60.5 - 28}{\sqrt{16.8}}\right) \\ &= P(2.82 < Z < 7.96) \\ &= P(Z < 7.96) - P(Z < 2.82) \\ &= 1 - 0.9976 \\ &= 0.0024 \end{aligned}$$

Example 2: Amber Alert

In 2021, Texan authorities sent out an Amber alert asking citizens to look out for a 3ft1 28-year-old with red hair and blue eyes who is wearing denim overalls, a stripy shirt and carrying a large knife. The probability that a random Texan received the Amber alert is 0.25.

a. In a small town, 6 people were surveyed. What is the probability that 3 or 4 people received the alert?

b. A larger sample of 80 people was taken in a suburban area. What is the probability that at least 20 people received the alert?

c. In a state wide survey, 300 people were polled. Estimate the probability that between 65 and 90 people (exclusive) received the alert.

d. In a different sample of 400 people, estimate the probability that fewer than 270 people did not receive the alert.

Solution: Amber Alert

Let the probability that a person received the alert be $p = 0.25$.

a. Let X = the number of people who received the alert.

X is a binomial random variable with $n = 6$ and $p = 0.25$

$$\begin{aligned} P(X = 3 \cup X = 4) &= P(X = 3) + P(X = 4) \\ &= C_3^6(0.25)^3 \cdot (0.75)^3 + C_4^6(0.25)^4 \cdot (0.75)^2 \\ &= 0.1318 + 0.0329 \\ &= 0.1647 \end{aligned}$$

b. Let X be the number of people in a sample of $n = 80$ who received the alert.

Check: $n \cdot p = 80 \cdot 0.25 = 20 > 5$ and $n(1 - p) = 60 > 5 \Rightarrow$ Normal approximation is appropriate.

$$\mu = n \cdot p = 80 \cdot 0.25 = 20$$
$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{80 \cdot 0.25 \cdot 0.75} = \sqrt{15}$$

$$\begin{aligned} P(X \geq 20) &= P(X > 19.5) && \text{Add correction factor} \\ &= P\left(Z > \frac{19.5 - 20}{\sqrt{15}}\right) \\ &= P(Z > -0.13) \\ &= 0.5517 \end{aligned}$$

c. Let X be the number of people (out of $n = 300$) who received the alert.

Want $P(65 < X < 90)$.

Use normal approximation since $np = 75 > 5$ and $n(1-p) = 225 > 5$.

$$\mu = 300 \cdot 0.25 = 75$$
$$\sigma = \sqrt{300 \cdot 0.25 \cdot 0.75} = \sqrt{56.25}$$

$$\begin{aligned} P(65 < X < 90) &\approx P(65.5 < X < 89.5) && \text{Add correction factors} \\ &= P\left(\frac{65.5 - 75}{\sqrt{56.25}} < Z < \frac{89.5 - 75}{\sqrt{56.27}}\right) \\ &= P(-1.27 < Z < 1.93) \\ &= P(Z < 1.93) - P(Z < -1.27) \\ &= 0.9732 - 0.1020 \\ &= 0.8712 \end{aligned}$$

d. Let Y be the number of people (out of 400) who **did not** receive the alert. Want $P(X < 270)$.

Note: $P(\text{not receive}) = 1 - 0.25 = 0.75$, so Y is a binomial random variable with $n = 400$ and $p = 0.75$

$$\mu = 400 \cdot 0.75 = 300$$
$$\sigma = \sqrt{400 \cdot 0.75 \cdot 0.25} = \sqrt{75}$$

$$\begin{aligned} P(Y < 270) &= P(X < 269.5) && \text{Add correction factor} \\ &= P\left(Z < \frac{269.5 - 300}{\sqrt{75}}\right) \\ &= P(Z < -3.51) \\ &= 0.0002 \end{aligned}$$

Example 3: AI Detection Test

When the U.S. military tested their new AI human-movement recognition robot, a group of Marines were asked to approach it undetected. Despite the robot's state-of-the-art sensors, all of them succeeded—one disguised himself as a fir tree, another somersaulted for 300 metres, and two hid under a cardboard box while giggling the entire way.

The distance a Marine could crawl before being detected by the robot follows a normal distribution with a mean of 240 metres and a standard deviation of 35 metres.

a. What is the probability that a randomly selected Marine crawls more than 275 metres before being detected?

b. Given that a Marine is detected before reaching 260 metres, what is the probability that he crawled between 200 and 230 metres?

c. What crawling distance corresponds to the 99th percentile for Marines in this test?

d. In a training camp of 150 Marines, each attempts to approach the robot once. Suppose the probability that a Marine remains undetected is 0.84. What is the probability that more than 135 Marines go unnoticed?

e. In a smaller squad of 12 Marines, what is the probability that exactly 8 or 9 of them crawl more than 260 metres before being detected?

Solution

Let X = the distance that a Marine crawls before being detected

X is a normally distributed random variable with $\mu = 240$ and $\sigma = 35$.

a. Want $P(X \geq 275)$

$$\begin{aligned} P(X \geq 275) &= P\left(Z \geq \frac{275-240}{35}\right) \\ &= P(Z \geq 1) \\ &= 0.1587 \end{aligned}$$

b. Want $P(200 < X < 230 \mid X < 260)$.

$$\begin{aligned} P(200 < X < 230 \mid X < 260) = \frac{P(200 < X < 230)}{P(X < 260)} &= \frac{P\left(\frac{200-240}{35} < Z < \frac{230-240}{35}\right)}{P\left(Z < \frac{260-240}{35}\right)} \\ &= \frac{P(-1.14 < Z < -0.30)}{P(Z < 0.57)} \\ &= \frac{0.3821 - 0.1271}{0.7157} \\ &= 0.0.3563 \end{aligned}$$

c. Want X such that the area to the left of Z is 0.99.

$$
\begin{aligned}
Z_{0.99} &= 2.33 \\
X &= Z\sigma + \mu \\
&= 2.33 \cdot (240) + 35 \\
&= 594.2\, m
\end{aligned}
$$

d. Let Y = the number of marines that go undetected that go undetected by the robot.

Y is a binomial random variable with $p = 0.84$ and $n = 150$. Want $P(Y > 135)$

$\because \quad np = 150(0.84) = 126 > 5 \quad ; n(1-p) = 24 > 5 \quad \Rightarrow$ Normal Approximation to the Binomial with $\mu = 126$ and $\sigma = \sqrt{150(0.84)(0.16)} = \sqrt{20.16}$

$$
P(X > 135) = P(X > 135.5) = P\left(Z > \frac{135.5 - 126}{\sqrt{20.16}}\right) = P(Z > 2.11) = 0.0174
$$

e. First we need $P(X > 260)$

$P(X > 260) = P\left(Z > \frac{260-240}{35}\right) = P(Z > 0.57) = 0.2843$

Let Z = the number of marines who can crawl more than 260 meters before being detected

Z is a binomial random variable with $p = 0.2843$ and $n = 12$

$np = 12(0.2843) = 3.4166 < 5 \quad \Rightarrow$ Cannot use approximation

Want $P(X = 8 \cup X = 9)$

$$
\begin{aligned}
P(X = 8 \cup X = 9) &= P(X = 8) + P(X = 9) \\
&= C_8^{12}(0.2843)^8(1-0.2843)^4 + C_9^{12}(0.2843)^9(1-0.2843)^3 \\
&= 0.0055 + 0.0010 \\
&= 0.0065
\end{aligned}
$$

Example 4: Sunions

'Sunions' are a new type of genetically engineered onion that don't cause tears when chopped. Marketed for their mild flavor and tear-free properties, they are now being tested in grocery stores across the country.

The circumference of Sunions follows a normal distribution with a mean of 18 cm and a standard deviation of 2.5 cm.

a. What is the probability that a randomly selected Sunion has a circumference greater than 21 cm?

b. If a Sunion has a circumference that is greater than 19 cm, what is the probability that it is larger than 22 cm?

c. What circumference corresponds to the 95th percentile for Sunions?

d. In a test panel of 120 people, each person is asked to chop a Sunion and report if it made them tear up. Suppose the probability a person does not cry is 0.92. What is the probability that more than 110 people report no tears?

e. In a grocery shipment of 10 Sunions, what is the probability that 6 or 7 of them have a circumference greater than 20 cm?

Solution: Sunions

a. Let X = circumference of a Sunion in centimetres.

   X is normally distributed with $\mu = 18$ and $\sigma = 2.5$

   Want $P(X > 21)$.

$$\begin{aligned} P(X > 21) &= P\left(Z > \frac{21 - 18}{2.5}\right) \\ &= P(Z > 1.2) \\ &= 0.1151 \end{aligned}$$

b. Conditional probability: $P(X > 22 \mid X > 19)$

$$\begin{aligned} P(X > 22 \mid X > 19) = \frac{P[(X > 19) \cap (X > 22)]}{P(X > 19)} &= \frac{P(X > 22)}{P(X > 19)} \\ &= \frac{P(Z > \frac{22-18}{2.5})}{P(Z > \frac{19-18}{2.5})} \\ &= \frac{P(Z > 1.6)}{P(Z > 0.4)} \\ &= \frac{0.0548}{0.3446} \\ &= 0.1591 \end{aligned}$$

c. 95th percentile of the Sunion circumference: Want to find X such that the area to the left of Z is 0.95

$$\begin{aligned} Z_{0.95} &= 1.645 \quad \text{(from Z-table)} \\ x &= \mu + z\sigma = 18 + 1.645 \cdot 2.5 \\ &= 22.11\,\text{cm} \end{aligned}$$

d. Let K be the number of people (out of 120) who do not cry. $p = 0.92$

$np = 110.4, \quad n(1-p) = 9.6 \quad \Rightarrow$ Normal approximation is valid.

$$\begin{aligned} \mu &= 120 \cdot 0.92 = 110.4 \\ \sigma &= \sqrt{120 \cdot 0.92 \cdot 0.08} = \sqrt{8.832} \approx 2.9718 \end{aligned}$$

$$\begin{aligned} P(K > 110) &= P(X > 110.5) \\ &= P\left(Z > \frac{110.5 - 110.4}{2.9718}\right) \\ &= P(Z > 0.03) \\ &= 0.4880 \end{aligned}$$

e. Let G = number of sunions out of 10 that have a circumference greater than 20 cm.

First we need $P(X > 20)$

$$P(X > 20) = P\left(Z > \frac{20 - 18}{2.5}\right) = P(Z > 0.8) = 1 - 0.7881 = 0.2119$$

Now we want $P(6 \le G \le 7)$ and G is binomially distributed with $n = 10$ and $p = 0.2119$

$$\begin{aligned} P(6 \le X \le 7) &= P(X = 6) + P(X = 7) \\ &= C_6^{10}(0.2119)^6(0.7881)^4 + C_7^{10}(0.2119)^7(0.7881)^3 \\ &= 0.0080 + 0.0012 \\ &= 0.0092 \end{aligned}$$