7.9 Applying the Normal Approximation

MATHEMATICAL TECHNIQUES

♠ Find the z-score (the number of standard deviations from the mean) for the following measurements.

- EXERCISE **7.9.1**

  A value of 11.0 drawn from a normal distribution with mean 13.0 and standard deviation 1.2.

- EXERCISE **7.9.2**

  A value of 0.9 drawn from a normal distribution with mean 0.5 and standard deviation 0.3.

- EXERCISE **7.9.3**

  A value of 12.0 drawn from a normal distribution with mean 10.0 and variance 25.0.

- EXERCISE **7.9.4**

  A value of 7.0 drawn from a normal distribution with mean 10.0 and variance 4.0.

♠ Use the cumulative distribution function for the standard normal ($\Phi(z)$) to find the following probabilities. Shade the associated area on two graphs: the given normal distribution, and the standard normal distribution.

- EXERCISE **7.9.5**

  The probability of a value less than 0.7 drawn from a normal distribution with mean 0 and variance 1.

- EXERCISE **7.9.6**

  The probability of a value greater than -0.1 drawn from a normal distribution with mean 0 and variance 1.

- EXERCISE **7.9.7**

  The probability of a value greater than 11.0 drawn from a normal distribution with mean 13.0 and standard deviation 1.2 (as in exercise 7.9.1.

- EXERCISE **7.9.8**

  The probability of a value less than 0.9 drawn from a normal distribution with mean 0.5 and standard deviation 0.3 (as in exercise 7.9.2.

- EXERCISE **7.9.9**

  The probability of a value between 10.0 and 12.0 drawn from a normal distribution with mean 10.0 and variance 25.0 (as in exercise 7.9.3.

- EXERCISE **7.9.10**

  The probability of a value between 7.0 and 13.0 drawn from a normal distribution with mean 10.0 and variance 4.0 (as in exercise 7.9.4.

♠ Using a table or a program of the cumulative distribution function for the standard normal, find the following probabilities.

- EXERCISE **7.9.11**

  The masses of a type of insect are normally distributed with a mean of 0.38 gm and a standard deviation of 0.09 gm. What is the probability that a given insect has mass less than 0.40 gm?

- EXERCISE **7.9.12**

  Scores on a test are normally distributed with mean 70 and standard deviation 10. What is the probability that a student scores more than 85?

- EXERCISE **7.9.13**

  Measurement errors are normally distributed with a mean of 0 and a standard deviation of 0.01 mm. Find the probability that a given measurement is within 0.012 mm of the true value.

- EXERCISE **7.9.14**

  The number of insects captured in a trap on different nights is normally distributed with mean 2950 and standard deviation 550. What is the probability of capturing between 2500 and 3500 insects?

♠ Use the normal approximation to the binomial with and without the continuity correction to find the following probabilities. How much difference does the continuity correction make? Exact answers are given in parentheses.

- EXERCISE **7.9.15**

  43% of trees are infested by a certain insect. What is the chance of randomly choosing 40 trees fewer than 10 of which are infested? (0.0144).

- EXERCISE **7.9.16**
  30% of the cells in a small organism are not functioning. What is the probability that an organ consisting of 250 cells is functioning if it requires 170 cells to work? (0.777).
- EXERCISE **7.9.17**
  In a certain species of wasp 75% of individuals are female. Find the probability that between 70 and 80 (inclusive) are female in a sample of 100. (0.7967).
- EXERCISE **7.9.18**
  10% of people are known to carry a certain gene. In a sample of 200, what is the probability that between 5 and 15 carry the gene? (0.1431).

♠ Use the normal approximation to the Poisson with and without the continuity correction to find the following probabilities. Compare with the exact answers (given in parentheses).

- EXERCISE **7.9.19**
  Seeds have fallen into a region with an average density of 20 seeds per square meter. What is the probability that a particular square meter contains fewer than 15 seeds? (0.1565).
- EXERCISE **7.9.20**
  Mutations occur along a piece of DNA at a rate of 0.023 per thousand codons. What is the probability of 30 or more mutations in a piece of DNA one million codons long? (0.0915).
- EXERCISE **7.9.21**
  Insects are caught in a trap at a rate of 0.21 per minute. What is the probability of catching between 10 and 15 in an hour? (0.6040).
- EXERCISE **7.9.22**
  Molecules enter a cell at a rate of .045 per second. What is the probability that between 150 and 200 enter during an hour? (0.8352).

♠ The normal approximation cannot be used in the following cases. Why not? Find another way to compute the probabilities and compare with the normal approximation.

- EXERCISE **7.9.23**
  A fair coin is flipped 8 times. Find the probability of no more than 1 head.
- EXERCISE **7.9.24**
  5% of people are infected with a disease. Find the probability of choosing 50 people and finding none who are infected.
- EXERCISE **7.9.25**
  In exercise 7.9.21, find the probability of catching more than one insect in a minute.
- EXERCISE **7.9.26**
  In exercise 7.9.20, find the probability of no mutation in a piece of DNA ten thousand codons long.

♠ Suppose that if $X \sim N(\mu, \sigma^2)$ and $Z = \dfrac{X-\mu}{\sigma}$.

- EXERCISE **7.9.27**
  Show that $\mathrm{E}(Z) = 0$ without writing down any integrals.
- EXERCISE **7.9.28**
  Show that $\mathrm{Var}(Z) = 1$ without writing down any integrals.

♠ Recall the data describing the probabilities of the outcomes of 4 experiments counting the number of mutants in a bacterial culture (exercises 7.8.5–7.8.8). Each experiment is repeated (independently) and the total number of mutants is counted.

- EXERCISE **7.9.29**
  Experiment a is repeated 20 times (as in exercise 7.8.5). Find the probability that there is a total of between 30 and 40 mutants, and compare with the sketch in exercise 7.8.5.
- EXERCISE **7.9.30**
  Experiment b is repeated 80 times (as in exercise 7.8.6). Find the probability that there are between 30 and 40 mutants, and compare with the sketch in exercise 7.8.6.
- EXERCISE **7.9.31**
  Experiment c is repeated 100 times (as in exercise 7.8.7). Find the probability that there is an average of less than 1.8, and compare with the sketch in exercise 7.8.7.

• EXERCISE **7.9.32**

Experiment d is repeated 25 times (as in exercise 7.8.8). Find the probability that there is an average of less than 1.8, and compare with the sketch in exercise 7.8.8.

APPLICATIONS

♠ Consider again the populations in exercises 7.8.33-7.8.36. For each population, use the normal approximation to estimate the probability that the population is larger after the given periods of time.

• EXERCISE **7.9.33**

Immigrants arrive into population a for 20 years (as in exercise 7.8.33). Redo the calculation of the probability that the population grows if immigrants arrive for 40 years. Why does it become larger?

• EXERCISE **7.9.34**

Immigrants arrive into population b for 20 years (as in exercise 7.8.34). Redo the calculation of the probability that the population grows if immigrants arrive for 10 years. Why does it become smaller?

♠ Consider again the situation in exercises 7.8.29-7.8.32.

• EXERCISE **7.9.35**

What is the probability of an IQ above 100 with just the genes of large effect (as in exercise 7.8.29)?

• EXERCISE **7.9.36**

What is the probability of an IQ above 100 with just the genes of small effect (as in exercise 7.8.30)?

• EXERCISE **7.9.37**

What is the probability of an IQ above 100 with just environmental factors (as in exercise 7.8.31)?

• EXERCISE **7.9.38**

What is the probability of an IQ above 100 with both genetics and environmental factors (as in exercise 7.8.32)?

♠ Recall that the logarithm of population size can be approximated using the normal distribution and the central limit theorem (7.8.37-7.8.40). Consider populations growing for the given number of years with the given distribution for the random variable R giving per capita reproduction. Find the probability that the population size P_t lies in the given range assuming that $P_0 = 100$.

• EXERCISE **7.9.39**

$R_i = 4$ with probability 0.5, $R_i = 0.25$ with probability 0.5 (as in exercise 7.8.37). Find the probability that P_{50} lies between 1 and 10000.

• EXERCISE **7.9.40**

$R_i = 4$ with probability 0.25, $R_i = 0.25$ with probability 0.75 (as in exercise 7.8.38). Find the probability that P_{25} lies between 1 and 3.

• EXERCISE **7.9.41**

R has p.d.f. $g(x) = 5.0$ for $1.0 \leq x \leq 1.2$ (as in exercise 7.8.39). Find the probability that P_{50} lies between 50 and 150.

• EXERCISE **7.9.42**

R has p.d.f. $g(x) = 1.25$ for $0.7 \leq x \leq 1.5$ (as in exercise 7.8.40). Find the probability that P_{50} lies between 50 and 150.

♠ Suppose that the alleles **A** and **a** for height are additive, meaning that plants with genotype **AA** are tall, plants with genotype **Aa** are intermediate, and those with genotype **aa** are short. If an **Aa** plant is crossed with another **Aa** plant, 1/4 of the offspring should be tall, 1/2 should be intermediate, and 1/4 should be short. Assuming that the other conditions for the binomial distribution are met, find the probabilities of the following. Suppose such a cross produces 36 offspring.

• EXERCISE **7.9.43**

Find the binomial distribution and its normal approximation for the number of tall offspring. What is the probability of getting 10 or fewer?

• EXERCISE **7.9.44**

Find the binomial distribution and its normal approximation for the number of intermediate offspring. What is the probability of getting 8 or fewer?

♠ Starting with 50 molecules, each leaving with probability 0.8 per minute, find the normal probability distribution approximating the number remaining at the following times. Use it to estimate the probability that more than 40 remain.

- EXERCISE **7.9.45**

  1 minute.

- EXERCISE **7.9.46**

  2 minutes.

♠ Figure 6.3b, illustrating stochastic immigration, was generated by adding 2 individuals with probability 0.5, and 0 with probability 0.5 for 100 generations. The results in the figure show final populations of 106 and 96.

- EXERCISE **7.9.47**

  Use the normal approximation to find the probability that the results lie between 96 and 106 inclusive.

- EXERCISE **7.9.48**

  The population size can be thought of as twice a binomial random variable. Use this idea, and the continuity correction, to estimate the probability that the results lie between 96 and 106 inclusive. How do the results compare?

♠ Genes in different organisms have different rates of mutation. Use the normal approximation to compute the following probabilities.

- EXERCISE **7.9.49**

  A gene has a mutation rate of 0.02 mutations per generation. Estimate the probability of more than 50 mutations in a period of 2000 generations.

- EXERCISE **7.9.50**

  A gene has a mutation rate of 0.002 mutations per generation. Estimate the probability of exactly 1 mutation in a period of 2000 generations. How does this compare with the exact answer in exercise 7.7.39

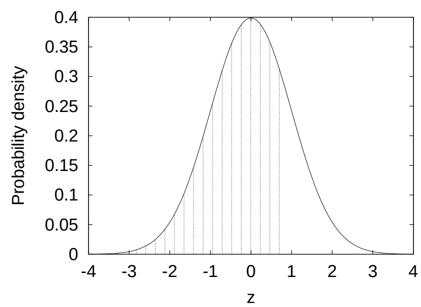
Chapter 8

Answers

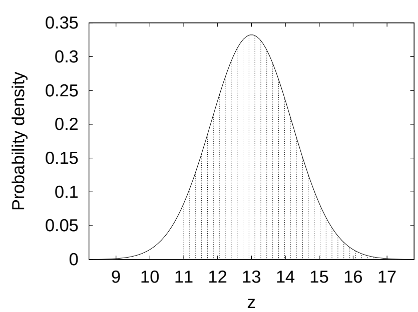
7.9.1. The z-score is $\frac{11.0-13.0}{1.2} = -1.67$.

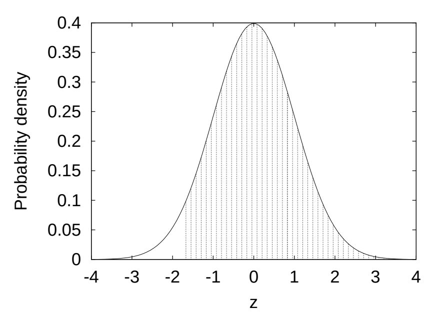
7.9.3. The standard deviation is $\sqrt{25.0} = 5.0$, so the z-score is $\frac{12.0-10.0}{5.0} = 0.4$.

7.9.5. Because the normal distribution with mean 0 and variance 1 is the standard normal, this probability is $\Phi(0.7)$ by definition, or 0.7508.

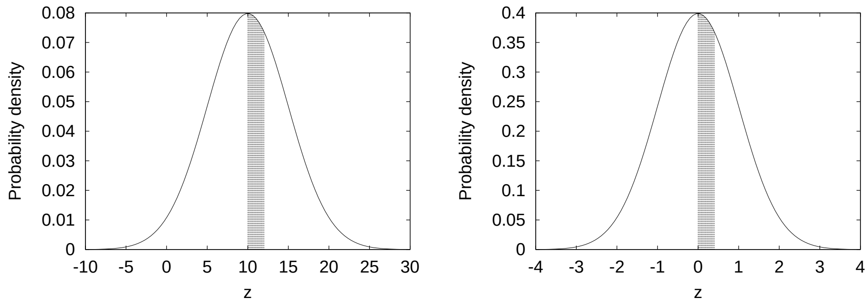


7.9.7. We found the z-score to be -1.67. Then $\Phi(-1.67)$ is the probability of a measurement less than 11.0 and $1-\Phi(-1.67) = 1-0.0475 = 0.9525$ is the probability of a value greater than 11.0.





7.9.9. We found the z-score associate with 12.0 to be 0.4. The z-score for the value 10.0 is 0 because 10.0 is the mean. The probability is then Then $\Phi(0.4) - \Phi(0.0) = 0.6554 - 0.5 = 0.1554$.



7.9.11. Let M represent the mass. Then

$$\begin{aligned}\Pr(M \le 0.4) &= \Pr\left(\frac{M-0.38}{0.09} \le \frac{0.4-0.38}{0.09}\right)\\ &= \Pr\left(Z \le \frac{0.4-0.38}{0.09}\right)\\ &= \Pr(Z \le 0.222) = \Phi(0.222) = 0.588.\end{aligned}$$

7.9.13. Let E represent the error. Then

$$\begin{aligned}\Pr(-0.012 \le E \le 0.012) &= \Pr\left(\frac{-0.012-0}{0.1} \le \frac{E-0}{0.1} \le \frac{0.012-0}{0.1}\right)\\ &= \Pr\left(-1.2 \le Z \le 1.2\right)\\ &= \Phi(1.2) - \Phi(-1.2) = 0.8849 - 0.1151 = 0.7698.\end{aligned}$$

7.9.15. $n = 40$ and $p = 0.43$. The mean is $np = 17.2$ and the variance is $np(1-p) = 9.804$. The standard deviation is $\sqrt{9.804} = 3.13$. Suppose $X \sim N(17.2, 9.804)$. With the continuity correction,

$$\Pr(T \le 10) \approx \Pr(X \le 10.5) = \Phi(\frac{10.5-17.2}{3.13}) = \Phi(-2.14) = 0.0162.$$

Without the continuity correction,

$$\Pr(T \le 10) \approx \Pr(X \le 10) = \Phi(\frac{10-17.2}{3.13}) = \Phi(-2.30) = 0.011.$$

Neither is extremely close in this case.

7.9.17. $n = 100$ and $p = 0.75$. The mean is $np = 75.0$ and the variance is $np(1-p) = 18.75$. The standard deviation is $\sqrt{18.75} = 4.33$. Suppose $X \sim N(75.0, 18.75)$. With the continuity correction,

$$\Pr(70 \le F \le 80) \approx \Pr(69.5 \le X \le 80.5) = \Phi(\frac{69.5-75.0}{4.33}) - \Phi(\frac{80.5-75.0}{4.33}) = \Phi(1.27) - \Phi(-1.27) = 0.7960.$$

Without the continuity correction,

$$\Pr(70 \le F \le 80) \approx \Pr(70 \le X \le 80) = \Phi(\frac{70-75.0}{4.33}) - \Phi(\frac{80-75.0}{4.33}) = \Phi(1.15) - \Phi(-1.15) = 0.7518.$$

The answer with the continuity correction is very close to the exact answer.

7.9.19. Let S represent the number of seeds that fall. $\Lambda = 20$. The mean is then 20 and the standard deviation is $\sqrt{20} = 4.47$. Let X be a normally distributed random variable with matching mean and standard deviation. Then with the continuity correction,

$$\begin{aligned}\Pr(S \le 15) &\approx \Pr(X \le 15.5)\\ &= \Phi(\frac{15.5-20.0}{4.47})\\ &= \Phi(-1.006) = 0.1572.\end{aligned}$$

Without the continuity correction,

$$\begin{aligned}\Pr(S \le 15) &\approx \Pr(X \le 15)\\ &= \Phi(\frac{15-20.0}{4.47})\\ &= \Phi(-1.118) = 0.1318.\end{aligned}$$

The answer with the continuity correction is much closer.

7.9.21. Let I represent the number of insects caught in an hour. $\Lambda = 0.21 \cdot 60 = 12.6$. The mean is then 12.6 and the standard deviation is $\sqrt{12.6} = 3.55$. Let X be a normally distributed random variable with matching mean and standard deviation. Then with the continuity correction,

$$\begin{aligned}\Pr(10 \le I \le 15) &\approx \Pr(9.5 \le X \le 15.5)\\ &= \Phi(\frac{9.5-12.6}{3.55}) - \Phi(\frac{15.5-12.6}{3.55})\\ &= \Phi(0.816) - \Phi(-0.873) = 0.6020.\end{aligned}$$

Without the continuity correction,

$$\begin{aligned}\Pr(10 \le I \le 15) &\approx \Pr(10 \le X \le 15)\\ &= \Phi(\frac{10-12.6}{3.55}) - \Phi(\frac{15-12.6}{3.55})\\ &= \Phi(0.676) - \Phi(-0.732) = 0.5186.\end{aligned}$$

The answer with the continuity correction is much closer.

7.9.23. The normal approximation cannot be used because $np = 4 < 5$. The probability of no more than one head is $b(0; 8, 0.5) + b(1; 8, 0.5) = 0.0352$. The mean is $np = 4$, the variance is $np(1-p) = 2.0$ and the standard deviation is $\sqrt{2.0} = 1.414$. Suppose $X \sim N(4.0, 2.0)$. With the continuity correction,

$$\Pr(H \le 1) \approx \Pr(X \le 1.5) = \Phi(\frac{1.5-4}{1.414}) = \Phi(-1.768) = 0.0385.$$

This turns out to be pretty close.

7.9.25. In one minute, the expected number is only 0.21, far less than 5. The exact probability of more than 1 is $1 - p(0; 0.21) - p(1; 0.21) = 0.192$. The normal approximation gives 0.0024.

7.9.27.

$$E(Z) = E(\frac{X-\mu}{\sigma}) = E(\frac{X}{\sigma}) - E(\frac{\mu}{\sigma}) = \frac{E(X)}{\sigma} - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0.$$

We used theorem 7.3.3 and the fact that the expectation of a constant is the constant itself.

7.9.29. We found that $S_{20} \sim N(42, 25.8)$. The standard deviation is $\sqrt{25.8} = 5.08$. Then

$$\Pr(30 \le S_{20} \le 40) \approx \Phi(\frac{40-42}{5.08}) - \Phi(\frac{30-42}{5.08}) = \Phi(-0.394) - \Phi(-2.37) = 0.338.$$

This is pretty close to the estimate of 0.4 found earlier.

7.9.31. We found that $A_{100} \sim N(1.6, 0.0184)$. The standard deviation is $\sqrt{0.0184} = 0.136$. Then

$$\Pr(A_{100} \le 1.8) \approx \Phi(\frac{1.8-1.6}{0.136}) = \Phi(1.47) = 0.929.$$

This is reasonably close to the estimate of 0.9 found earlier.

7.9.33. We found that $A_{20} \sim N(0.1, 0.0545)$, with standard deviation of $\sigma = 0.233$. Then

$$\Pr(A_{20} \ge 0) \approx 1 - \Phi(\frac{0-0.1}{0.233}) = 1 - \Phi(-0.429) = 0.666.$$

After 40 years, $A_{40} \sim N(0.2, 0.109)$, with standard deviation of $\sigma = \sqrt{0.109} = 0.330$. Then

$$\Pr(A_{40} \ge 0) \approx 1 - \Phi(\frac{0-0.2}{0.330}) = 1 - \Phi(-0.606) = 0.728.$$

This population grows on average, and is thus more likely to grow if given more time.

7.9.35.

$$\begin{aligned}\Pr(IQ \geq 100) &= 1 - \Phi(\frac{100.0 - 98.75)}{\sqrt{11.72}}) \\ &= 1 - \Phi(0.365) = 1 - 0.6425 = 0.3575.\end{aligned}$$

7.9.37.

$$\begin{aligned}\Pr(IQ \geq 100) &= 1 - \Phi(\frac{100.0 - 93.5)}{\sqrt{6.06}}) \\ &= 1 - \Phi(2.644) = 1 - 0.9959 = 0.0041.\end{aligned}$$

7.9.39. We found that $\ln(P_{50}) \sim N(4.6, 96)$ with standard deviation 9.8. We want to find the probability that $\ln(1) \leq \ln(P_{50}) \leq \ln(10000)$ or $0 \leq \ln(P_{50}) \leq 9.21$.

$$\Pr(0 \leq \ln(P_{50}) \leq 9.21) \approx \Phi(\frac{0 - 4.6}{9.8}) - \Phi(\frac{9.2 - 4.6}{9.8}) = \Phi(0.470) - \Phi(-0.469) = 0.362.$$

7.9.41. We found that $\ln(P_{50}) \sim N(4.695, 0.135)$ with standard deviation 0.37. We want to find the probability that $\ln(50) \leq \ln(P_{50}) \leq \ln(150)$ or $3.912 \leq \ln(P_{50}) \leq 5.011$.

$$\Pr(3.912 \leq \ln(P_{50}) \leq 5.011) \approx \Phi(\frac{3.912 - 4.695}{0.37}) - \Phi(\frac{5.011 - 4.695}{0.37}) = \Phi(0.853) - \Phi(-2.116) = 0.786.$$

7.9.43. The binomial has $n = 36$ and $p = 0.25$. The mean is then $np = 9$, the variance is $np(1 - p) = 6.75$ and the standard deviation is 2.598. The number $T \approx X \sim N(9.0, 6.75)$. With the continuity correction,

$$\begin{aligned}\Pr(T \leq 10) &\approx \Pr(X \leq 10.5) \\ &= \Pr((X - 9)/2.598 \leq (10.5 - 9)/2.598) \\ &= \Pr(Z \leq 0.577) = \Phi(0.577) = 0.718.\end{aligned}$$

7.9.45. The binomial has $n = 50$ and $p = 0.8$. The mean is then $np = 40$, the variance is $np(1 - p) = 8$ and the standard deviation is 2.82. The number $M \approx X \sim N(40.0, 8.0)$. With the continuity correction,

$$\begin{aligned}\Pr(M \geq 40) &\approx \Pr(X \geq 39.5) \\ &= \Pr((X - 40)/2.82 \geq (39.5 - 40)/2.82) \\ &= \Pr(Z \geq -0.177) = 1 - \Phi(-0.177) = 0.570.\end{aligned}$$

7.9.47. The mean number added per generation is 1.0, and the variance is also 1.0. After 100 generations, the mean is 100.0, the variance is 100.0, and the standard deviation is 10.0. The number $I \approx X \sim N(100.0, 100.0)$.

$$\begin{aligned}\Pr(96 \leq I \leq 106) &= \Pr\left(\frac{96 - 100}{10} \leq \frac{I - 100}{10} \leq \frac{106 - 100}{10}\right) \\ &= \Pr(-0.4 \leq Z \leq 0.6) \\ &= \Phi(0.6) - \Phi(-0.4) = 0.7257 - 0.3446 = 0.3811.\end{aligned}$$

7.9.49. The number of mutations will follow a Poisson distribution with parameter $\Lambda = 0.02 \cdot 2000 = 40.0$. The approximating normal is $X \sim N(40.0, 40.0)$ with standard deviation 6.32. Then

$$\begin{aligned}\Pr(M \geq 50) &\approx \Pr(X \geq 49.5) \\ &= \Pr((X - 40)/6.32 \geq (49.5 - 40)/6.32) \\ &= \Pr(Z \geq 1.502) = 1 - \Phi(1.502) = 0.0665.\end{aligned}$$