Chapter 5: Normal Probability Distributions - Solutions

Note: All areas and z-scores are approximate. Your answers may vary *slightly.*

5.2 Normal Distributions: Finding Probabilities

If you are given that a random variable X has a normal distribution, finding **probabilities** corresponds to finding the **area** between the standard normal curve and the x-axis, using the table of z-scores. The mean (expected value) μ and standard deviation σ should be given in the problem.

- For the probability that $X < b$, convert b into a z-score using
$$z = \frac{b - \mu}{\sigma}$$
and use the table to find the area to the *left* of the z-value.
- For the probability that $X > a$, convert a into a z-score using
$$z = \frac{a - \mu}{\sigma}$$
and use the table to find the area to the *right* of the z-score.
- For the probability that $a < X < b$ (X is between two numbers, a and b), convert a and b into z-scores using
$$z = \frac{a - \mu}{\sigma} \text{ and } z = \frac{b - \mu}{\sigma}$$
and use the table to find the area *between* the two z-values.

1. The average speed of vehicles traveling on a stretch of highway is 67 miles per hour with a standard deviation of 3.5 miles per hour. A vehicle is selected at random.

   a. What is the probability that it is violating the 70 mile per hour speed limit? Assume that the speeds are normally distributed.
   Solution:
   The random variable X is ___speed___.

   We are told that X has a normal distribution.
   The mean $\mu =$ ___67___. The standard deviation $\sigma =$ ___3.5___.
   We are looking for the probability of the event that ___$X > 70$___.

Step 1: Convert 70 into a z-score:

$$z = \frac{70 - 67}{3.5} \approx 0.86$$

Step 2: Find the appropriate area between the normal curve and the axis using the table:

The table contains cumulative areas (to the *left* of the z-value). The area corresponding to a z-score of 0.86 in the table is 0.8051. Since we are interested in $X > 70$, we need the area to the *right* of the z-score, thus

$$P(X > 70) \approx 1 - 0.8051 \approx 0.1949$$

(a) What is the probability that a randomly selected vehicle is not violating the speed limit?

The z-score is the same: 0.86. We are interested in $P(X \leq 70)$, thus the area to the *left* of this z-score can be read directly off the table: 0.8051. OR, using complements and the answer to part a,

$$P(X \leq 70) = 1 - P(X > 70) \approx 1 - 0.1949 \approx 0.8051$$

(b) What is the probability that a randomly selected vehicle is traveling under 50 miles per hour?

We are interested in $P(X < 50)$. The z-score is

$$z = \frac{50 - 67}{3.5} = \frac{-17}{3.5} \approx -4.86$$

The area needed is to the *left* of this z-score. Notice that -4.86 is not even on the table, and the lowest z-score is -3.49, with a corresponding area of 0.0002. For z-scores less than -3.49, the area is even less than 0.0002 and very close to 0, and we may assume is approximately 0. (The more accurate answer is about 0.00000059: about 1 in 1.7 million probability.)

(c) What is the probability that a randomly selected vehicle is traveling between 50 and 70 miles per hour?

The z-score for 70 is 0.86 with a corresponding area of 0.8051. The z-score for 50 is -4.86 with a corresponding area of about 0. Thus, we subtract:

$$P(50 < X < 70) \approx 0.8051 - 0 \approx 0.8051$$

Practice Problem: A customer calling a call center spends an average of 45 minutes on hold during the peak season, with a standard deviation of 12 minutes. Suppose these times are normally distributed. Find the probability that the customer will be on hold for each interval of times:

a. More than 54 minutes.

Let X be the number of minutes the customer spends on hold. We want $P(X > 54)$. The mean is $\mu = 45$ and standard deviation is $\sigma = 12$. The z-score is

$$z = \frac{54 - 45}{12} = \frac{9}{12} = 0.75$$

The corresponding area is 0.7734. For $P(X > 54)$, the area to the *right* is needed. Thus,

$$P(X > 54) \approx 1 - 0.7734 = 0.2266$$

b. Less than 24 minutes.

We want $P(X < 24)$. The z-score is

$$z = \frac{24 - 45}{12} = \frac{-21}{12} = -1.75$$

The corresponding area is 0.0401. For $P(X < 24)$, the area to the *left* is needed. Thus, $P(X < 24) \approx 0.0401$

c. Between 24 and 54 minutes.

The z-score for 24 is -1.75 with a corresponding area of 0.0401, and the z-score for 54 is 0.75 with a corresponding area of 0.7734. Thus,

$$P(24 < X < 54) \approx 0.7734 - 0.0401 = 0.7333$$

d. More than 39 minutes.

We need $P(X > 39)$. The z-score for 39 is

$$z = \frac{39 - 45}{12} = -0.5$$

The corresponding area is 0.3085. We need the area to the *right* of the z-score. Thus,

$$P(X > 39) \approx 1 - 0.3085 = 0.6915$$

5.3 Normal Distributions: Finding Values

Now the process from 5.2 will be reversed. Starting with a probability, you will find a corresponding z-score. The same table will be used, but you will search the center of the table to find the probability first, and then determine the z-score that corresponds to that probability. To make this easier, first draw a picture.

2. Find the indicated z-scores. Draw a picture and include a short explanation

   a. The z-score that corresponds to a *cumulative area* of 0.3632 (the cumulative area is the area to the *left* of the z-score).

   Look for the given area in the table and find the corresponding z-score: -0.35

   b. The z-score that corresponds to 0.1075 of the distribution's area to its right.

   The table lists the cumulative area: to the *left* of the z-score. Thus, the z-score needed corresponds to a left area of $1 - 0.1075 = 0.8925$. This z-score is 1.24.

   c. The z-score that corresponds to 96.16% of the distribution's area to its right.

   First convert 96.16% into a probability (area): 0.9616. The z-score needed corresponds to a left area of $1 - 0.9616 = 0.0384$. This z-score is -1.77.

   d. The z-score that corresponds to the 90^{th} percentile (P_{90}) of the distribution's area.

   Convert 90% into a probability (area) first: 0.9000. Even though this exact area is not in the table, pick the closest areas. The desired z-score is between 1.28 and 1.29. (You may use either of these or average them).

<u>Practice Problem:</u>

a. The z-score that corresponds to a *cumulative area* of 0.8888.

From the table: $z = 1.22$.

b. The z-score that corresponds to 0.4090 of the distribution's area to its right.

Find the z-score corresponding to area $1 - 0.4090 = 0.5910$. This is 0.23.

c. The z-score that corresponds to 84.13% of the distribution's area to its right.

Convert 84.13% into a probability: 0.8413, and find the z-score corresponding to area $1 - 0.8413 = 0.1587$. This is -1.00.

d. The z-score that corresponds to the 30^{th} percentile (P_{30}) of the distribution's area.

Convert 30% into a probability: 0.3000, and find the z-score corresponding to this area: between -0.53 and -0.52.

Transforming a z-score into a data value

Given a z-score, it can be converted back into a data value by solving for x in the equation

$$z = \frac{x - \mu}{\sigma}$$

Given z, to find x, use the formula

$$x = \mu + z\sigma.$$

Procedure: Area $\rightarrow$ z-score $\rightarrow$ data value.

3. Scores for the California Police Officer Standards and Training test are normally distributed, with a mean of 50 and a standard deviation of 10.

   a. An agency will only hire applicants with scores in the top 10%. What is the lowest score you can earn and still be eligible to be hired by the agency?
   The mean is $\mu = 50$ and standard deviation is $\sigma = 10$. The top 10% corresponds to the 90^{th} percentile. The corresponding z-score was found earlier, which is about 1.285. Using the formula, this corresponds to a test score of

   $$x = \mu + z\sigma \approx 50 + 1.285(10) \approx 63$$

   b. Those officers scoring below the 20^{th} percentile are sent to undergo additional training. What is the minimum score needed to avoid this training?
   The 20^{th} percentile corresponds to a cumulative area of 0.2000. The closest z-scores are -0.85 and -0.84. We can use the average z-score -0.845. This corresponds to a test score of

   $$x = \mu + z\sigma \approx 50 + (-0.845)(10) \approx 42$$

Practice Problem: The length of time employees have worked at a particular company is normally distributed with mean 11.2 years and standard deviation 2.1 years.

a. If the lowest 10% of employees in seniority are to be layed-off in a cutback, what is the maximum length of time that an employee could have worked and still be laid off?
The 10^{th} percentile corresponds to a cumulative area of 0.1000. The closest z-scores are -1.29 and -1.28. We can use the average z-score -1.285. This corresponds to the length of time worked

$$x = \mu + z\sigma \approx 11.2 + (-1.285)(2.1) \approx 8.5 \text{ years}$$

b. If the highest 10% of employees in seniority are to be promoted, what is the minimum length of time that an employee could have worked and still be promoted?

The 90^{th} percentile corresponds to a cumulative area of 0.9000. The closest z-scores are 1.28 and 1.29. We can use the average z-score 1.285. This corresponds to the length of time worked

$$x = \mu + z\sigma \approx 11.2 + (1.285)(2.1) \approx 13.9 \text{ years}$$

5.4 Sampling Distributions and The Central Limit Theorem

Given:

i) a (large) population,

ii) a numerical characteristic associated with each member of the population,

iii) the population mean μ and population standard deviation σ for this characteristic,

You:

i) take a simple random sample of 100 members of the population and calculate the mean and standard deviation.

ii) repeat taking simple random samples of 100 members several times and calculate the mean and standard deviation each time.

The sample distribution is denoted by $\bar{x}$. In general, you cannot expect that the mean you obtain for each sample of 100 to be equal to μ, but

Theorem 0.1 (Central Limit Theorem)**.** *If samples of size n, ($n \geq 30$) are drawn from any population with mean μ and standard deviation σ, the sample mean will be approximately distributed according to a normal distribution with*

$$\begin{aligned} \textit{Mean } \mu_{\bar{x}} &= \mu \\ \textit{Variance } \sigma_{\bar{x}}^2 &= \frac{\sigma^2}{n} \\ \textit{Standard deviation } \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} \qquad \textit{“standard error of the mean”} \end{aligned}$$

If the population is already distributed normally, the restriction on the sample size, that $n \geq 30$ is not necessary. For any population, the larger the sample size, the better the normal approximation is.

Use the following analogue of the formula for the z-score to find the probability that a sample mean will fall into some given interval:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

4. Monthly cell phone bills for residents of a city have mean \$63 and standard deviation \$11. Simple random samples of 100 are drawn and the mean is determined for each sample.

   a. What is the sample size, n? 100

   b. Find the *mean* of the sampling distribution of sample means.

   $$\mu_{\bar{x}} = 63$$

   c. Find the *standard deviation* of the sampling distribution of sample means.

   $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11}{\sqrt{100}} = 1.1$$

   d. What is the probability that the mean of a sample is greater than \$74? (hint: first find the z-score)

   $$z = \frac{74 - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{74 - 63}{1.1} = 10$$

   Since this z-score is greater than 3.49, it corresponds to an area very close to (almost equal to) 1.

   The probability that the mean is *greater* than 74 corresponds to a *right* area, thus, the answer is approximately

   $$1 - 1 = 0$$

   e. What is the probability that the mean of a sample is less than \$63?

   $$z = \frac{63 - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{63 - 63}{1.1} = 0$$

   This z-score corresponds to an area of 0.5000.

   The probability that the mean is *less* than 74 corresponds to a *left* area, thus, the answer is 0.5000.

   f. What is the probability that the mean of a sample is between \$52 and \$74? For 52,

   $$z = \frac{52 - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{52 - 63}{1.1} = -10$$

   with a corresponding area of (a little bit above) 0.

   For 74, the z-score is 10, with a correponding area of (almost) 1.

   Thus, the probability that the sample mean is between 52 ad 74 is approximately

   $$1 - 0 = 1$$

Practice Problem: The mean room and board expense per year at four-year colleges is $7,540 and standard deviation is $1245. Assume the room and board yearly expense is normally distributed. You select a simple random sample of 9 colleges.

a. What is the probability that the mean room and board is less than $7800? The sample size is 9. Even though it is under 30, we can still use the results of section 5.4 because the original random variable, X, representing the yearly expense is (assumed to be) normally distributed.

$$z = \frac{7800 - 7540}{1245/\sqrt{9}} = \frac{260}{415} \approx 0.63$$

For the probability that the mean room and board is *less* than $7800, the area to the *left* of the z-score is needed, thus, the desired probability is 0.7357

b. What is the probability that the mean room and board is between $7295 and $8765?

$$z = \frac{7295 - 7540}{1245/\sqrt{9}} = \frac{-245}{415} \approx -0.59$$

with a corresponding area of 0.2776, and

$$z = \frac{8765 - 7540}{1245/\sqrt{9}} = \frac{1225}{415} \approx 2.95$$

with a corresponding area of 0.9984.

Thus, the probability that the mean room and board is between $7295 and $8765 is the difference of these areas:

$$0.9984 - 0.2776 = 0.7208$$

5.5 Normal Approximations to Binomial Distributions

Recall that a binomial random variable arises in a situation when there are n independent trials (repetitions) of the same experiment. Each trial has two outcomes: success or failure. p = probability of success of a single outcome and $q = 1 - p$ = probability of failure of a single outcome.

5. Suppose a doctor performs a surgical procedure on 150 patients. Each time, the procedure has an 85% success probability. So,

$$n = 150 \qquad p = 0.85 \qquad q = 0.15$$

Random variable X is the number of successes. Then the probability of *exactly* 100 successful surgeries is

$$P(X = 100) =_{150} C_{100}(0.85)^{100}(0.15)^{50} \approx 0.0000000112$$

But, what is the probability of 120 or fewer successful surgeries? This is impractical to calculate directly since we need $P(X = 120), P(X = 119), P(X = 118), P(X = 117)$, etc.

Theorem 0.2. *If $np \geq 5$ and $nq \geq 5$,*

the binomial distribution is well-approximated by the normal distribution with

$$\textit{mean } \mu = np \qquad \textit{and} \qquad \textit{standard deviation } \sigma = \sqrt{npq}.$$

<u>Note:</u> When converting from binomial to normal distributions, change the intervals in the following way:

- Add 0.5 to the maximum number of desired successes.
- Subtract 0.5 from the minimum number of desired successes.

Once this has been done, the number of successes ± 0.5 can be converted into a z-score.

a. Can the normal approximation to the binomial be used for the surgery example?

In this example, n = 150, p = 0.85, q = 0.15

We check: $np = 150 \cdot 0.85 = 127.5 \geq 5$ and $nq = 150 \cdot 0.15 = 22.5 \geq 5$. Since both np and nq are greater than 5, the normal approximation may be used.

b. What is the mean of the approximating normal distribution?

$$\mu = np = 150 \cdot 0.85 = 127.5$$

c. What is the standard deviation of the approximating normal distribution?

$$\sigma = \sqrt{npq} = \sqrt{150 \cdot 0.85 \cdot 0.15} \approx 4.37$$

d. To approximate $P(X \leq 120)$, convert 120.5 into a z-score.

$$z = \frac{120.5 - 127.5}{4.37} \approx -1.60$$

e. Use the table to look up the desired probability for part d.

The area to the *left* of the z-score is needed: $P(X \leq 120) \approx .0548$

f. Approximate $P(X \geq 130)$.

We first convert 129.5 into a z-score (since 130 is the minimum number of successes needed, subtract 0.5):

$$z = \frac{129.5 - 127.5}{4.37} \approx 0.46$$

The area to the *right* of the z-score is needed, thus

$$P(X \geq 130) \approx 1 - 0.6772 = .3228$$

g. Approximate $P(125 \leq X \leq 140)$.

First, the z-scores for 124.5 and 140.5:

$$z = \frac{124.5 - 127.5}{4.37} \approx -0.69$$

with a corresponding area of 0.2451, and

$$z = \frac{140.5 - 127.5}{4.37} \approx 2.97$$

with a corresponding area of 0.9985. Thus,

$$P(125 \leq X \leq 140) \approx 0.9985 - 0.2451 = 0.7534$$

<u>Practice Problem:</u> According to a survey, 70% of adults between 50 and 64 years old use the Internet. You randomly select 80 adults in that age range and ask them if they use the Internet.

a. Approximate the probability that 70 or more people say they use the Internet.

In this example, X is the number of people who use the internet, n = 80, p = 0.70, q = 0.30

We check: $np = 80 \cdot 0.70 = 56 \geq 5$ and $nq = 80 \cdot 0.30 = 24 \geq 5$. Since both np and nq are greater than 5, the normal approximation may be used.

The mean of the approximating normal distribution is

$$\mu = np = 80 \cdot 0.70 = 56$$

The standard deviation of the approximating normal distribution is

$$\sigma = \sqrt{npq} = \sqrt{80 \cdot 0.70 \cdot 0.30} \approx 4.10$$

To approximate $P(X \geq 70)$, convert 69.5 into a z-score.

$$z = \frac{69.5 - 56}{4.10} \approx 3.29$$

The area to the *right* of the z-score is needed: $P(X \geq 70) \approx 1 - 0.9995 = .0005$

b. Approximate the probability that between 50 and 70 people say they use the Internet.

The z-score for 70 (accounting for the continuity correction) is

$$z = \frac{70.5 - 56}{4.1} \approx 3.54$$

with corresponding area 0.9998.

The z-score for 50 is

$$z = \frac{49.5 - 56}{4.1} \approx -1.59$$

with corresponding area 0.0559. Thus,

$$P(50 \leq X \leq 70) \approx 0.9998 - 0.0559 = 0.9439$$