A Walk on the Wild Side

An infestation of whirling disease has had a dramatic effect on the trout populations in Montana's rivers. In this module, you discover how wildlife biologists investigate such problems.

Todd Fife • Pete Stabio

A Walk on the Wild Side

Introduction

In the spring of 1995, an 80-km stretch of Montana's Madison River, one of America's finest trout streams, was closed to fishing. Before the closure, studies by fisheries biologists showed a dramatic decline in the population of rainbow trout in that section of the river—from 2051 rainbows per kilometer in 1991 to 186 per kilometer in 1994. Wildlife managers ordered the closure to protect the trout population during the spawning season.

The rapid decline of the trout population in the Madison has been blamed on whirling disease. This incurable illness causes young fish to spin as if chasing their tails, leaving them vulnerable to predators and unable to feed.

Activity 1

Because it is not feasible to catch and examine every fish in a river, biologists rely on sampling methods when studying trout populations. When monitoring the spread of whirling disease, researchers use statistics to estimate the proportion of the population with the disease.

Exploration

In this exploration, you use a simulation to investigate sample proportions. **Note:** Save your work for use in Activity **2**.

a. Obtain a container of 100 beans from your teacher. In this simulation, the container represents a section of river, while the beans represent the trout population. The marked beans represent diseased fish. To estimate the proportion of diseased fish in your simulated population, complete Steps **1–3** below.

1. Draw one bean at random from the container and record whether or not it represents a diseased fish.
2. Return the bean to the container and mix it thoroughly with the others.
3. Repeat Steps **1** and **2** until you have obtained a sample of 4 beans. (This is an example of sampling with replacement.)

Mathematics Note:

The proportion of a population that displays a certain characteristic can be denoted as p. The value of p is a **population proportion** and can be found as follows:

$$p = \frac{\text{number in population with the characteristic}}{\text{total population}}$$

In the population of all citizens of the United States, for example, the proportion p of females is:

$$p = \frac{\text{number of females in U.S. population}}{\text{total U.S. population}}$$

When a sample is taken from a population, the proportion of the sample that displays the characteristic is denoted as $\hat{p}$. Sample proportions are often used to estimate population proportions. The value of $\hat{p}$ is the **sample proportion** and can be found as shown below:

$$\hat{p} = \frac{\text{number in sample with characteristic}}{\text{total number in sample}}$$

For example, consider a sample of the U.S. population that consists of all secondary school students. The proportion $\hat{p}$ of females in this sample can be used to estimate p, the proportion of females in the U.S population.

b. Determine the proportion $\hat{p}$ of diseased trout from the sample taken in Part **a**.

c. Estimate the proportion p of diseased trout in the population.

d.
1. Repeat Parts **a** and **b** 19 more times to obtain a total of 20 sample proportions.
2. Calculate the mean and standard deviation of the 20 sample proportions.

e. The **relative frequency** of an item is the ratio of its frequency to the total number of observations in the data set. A **relative frequency table** includes columns that describe a data item or interval, its frequency, and its relative frequency.

1. Complete a relative frequency table with headings like those in Table **1** for the 20 sample proportions from Part **d**.

Table 1: Relative frequency table for sample proportions

Sample Proportion	Frequency	Relative Frequency

2. Create a histogram of relative frequency versus sample proportion.

3. Use the graph to revise your estimate of the population proportion from Part **c**.

f. Use technology to repeat Parts **a–e** for a sample size of 30.

g.

1. Combine the class data for the samples of size 4. Create a relative frequency table and histogram of the results. **Note:** Save this data for use in Activity **2**.

2. Combine the class data for the samples of size 30. Create a relative frequency table and histogram of the results. **Note:** Save this data for use in Problem **1.3** and in Activity **2**.

3. Re-evaluate your estimate of p from Part **e**.

Discussion

a. Does the sampling method described in Part **a** of the exploration produce random samples? Explain your response.

b. Compare the results of the sampling process with samples of size 4 with the results for samples of size 30.

c. Imagine that biologists have only enough time and money to collect one sample when estimating the proportion of diseased fish in a population. Which of the sample sizes used in the exploration gave a better chance of obtaining a good estimate? Explain your response.

d.

1. Using the data obtained in the exploration for 20 samples of size 30, estimate the probability of obtaining a sample with exactly 15 diseased fish.

2. Using the combined class data for a sample size of 30, estimate the probability of obtaining a sample with exactly 15 diseased fish.

3. According to the law of large numbers, which set of data is more likely to provide a better estimate of the probability of obtaining a sample with exactly 15 diseased fish?

e. In practice, wildlife biologists often sample without replacement. To sample a fish population, for example, researchers might net fish from one section of a river. While the sample is being gathered, the fish are placed in holding tanks. After the desired data has been collected, the fish are returned to the water.

When is it reasonable to analyze such data as if the sampling had been done with replacement?

Assignment

1.1 In the exploration, you used sample sizes of 4 and 30 to obtain data about a population of fish. For a sample of size 4, the possible proportions of diseased fish are 0, 0.25, 0.5, 0.75, and 1.00. List the different proportions that could result for sample sizes of 30.

1.2 The following relative frequency table shows the data collected from 40 samples of size 4 from a population of fish. As you can see, some entries have been omitted from the table.

Proportion of Diseased Fish	Frequency	Relative Frequency
$0/4 = 0$	1	
$1/4 = 0.25$		0.175
$2/4 = 0.50$	14	
$3/4 = 0.75$		0.375
$4/4 = 1.00$		0.075

a. Determine the missing values in the table.

b. Describe how to use this data to estimate the proportion of diseased fish in the population.

1.3 Use the relative frequency table for the combined class data of samples of size 30 from Part **g** of the exploration to complete Parts **a–c**.

a. Estimate the probability of obtaining a sample with 8 diseased fish.

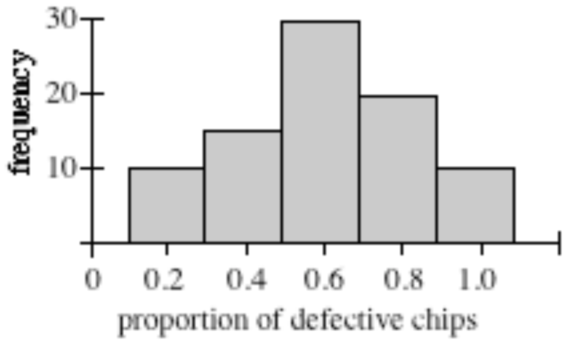
b. How is the relative frequency of a sample with 8 diseased fish related to the estimated probability?

c. Estimate the probability of obtaining each of the following:

1. a sample with 10 diseased fish
2. a sample with more than 10 diseased fish
3. a sample with 8 or fewer diseased fish.

* * * * *

1.4 The following frequency histogram shows the data collected from 85 samples of size 5 from a population of computer chips.



a. Using the histogram, estimate the probability of obtaining a sample with exactly 60% defective chips.

b. Using the histogram, estimate the probability of obtaining a sample with at least 60% defective chips.

c. Sketch a histogram that represents the relative frequency of each proportion in this data set.

d. Estimate the proportion of defective chips in the population.

1.5 The RoboSpace Company makes sprockets. As part of their quality-control process, the company take samples of 5 sprockets each hour and determines the proportion that are defective. The following table shows the data collected for 20 samples of size 5.

Proportion of Defective Sprockets	Frequency	Relative Frequency
$0/5 = 0.0$	5	
$1/5 = 0.20$	4	
$2/5 = 0.40$	5	
$3/5 = 0.60$	5	
$4/5 = 0.80$	1	
$5/5 = 1.00$	0	

a. Determine the relative frequency of each sample proportion.

b. Create a histogram of the relative frequencies.

c. Estimate the probability of obtaining a sample with exactly 2 defective sprockets.

d. Estimate the proportion of defective sprockets in the population.

* * * * * * * * * *

Activity 2

Even though the entire population may not be infected, wildlife researchers know that a random sample might contain all diseased fish. Though this is unlikely, it is possible. It is more likely, however, that the proportion of diseased fish in the sample will be close to the actual proportion of diseased fish in the population.

Discussion 1

a. As you may recall from previous modules, a **binomial experiment** has the following characteristics:

- The experiment consists of a fixed number of repetitions of the same action. Each repetition is a trial.
- The trials are independent of each other. In other words, the result of one trial does not influence the result of any other trial in the experiment.
- Each trial has only two possible outcomes: success or a failure.
- The probability of a success remains constant from trial to trial.
- The total number of successes is observed.

Explain why taking a sample of 4 fish from a lake and determining the number of diseased fish can be modeled by a binomial experiment.

b. In a sample of 4 fish, one possible outcome is HHDH, where H represents a healthy fish and D represents a diseased fish.

Describe how to use combinations to determine the number of different ways in which you could randomly draw 4 fish from a population, one at a time with replacement, and record exactly 1 diseased fish.

c. Recall that the **binomial probability formula** can be used to determine the probability of obtaining r successes in n trials in a binomial experiment. Symbolically, the binomial formula can be written as follows, where p is the probability of success in any one trial:

$$P(r \text{ successes in } n \text{ trials}) = C(n,r) \bullet p^r \bullet (1-p)^{n-r}$$

Given that the probability of catching a diseased fish is 0.6, describe how to determine the probability that a random sample of 4 fish contains exactly 1 diseased fish.

Mathematics Note

The **sampling distribution** of all possible sample proportions $\hat{p}$ for samples of size n is the set of probabilities associated with each possible value of $\hat{p}$.

For example, Table **2** below shows the sampling distribution for $\hat{p}$ given a sample size of 4 and a population proportion of 0.6.

Table 2: Sampling distribution for $n = 4$ and $p = 0.6$

Sample Proportion	Probability
$0/4 = 0$	0.0256
$1/4 = 0.25$	0.1536
$2/4 = 0.5$	0.3456
$3/4 = 0.75$	0.3456
$4/4 = 1$	0.1296

d. What is the sum of the probabilities in any sampling distribution of $\hat{p}$? Explain your response.

Exploration 1

In this exploration, you compare the data you collected in Activity **1** to the sampling distribution of $\hat{p}$ for samples of size 4 and 30.

a. Use the actual population proportion from the exploration in Activity **1** and the binomial probability formula to construct the sampling distribution for samples of size 4.

b. Calculate the mean of the sampling distribution from Part **a**. Compare this value with each of the following:

1. the population proportion
2. the mean of the sample proportions for samples of size 4 from Activity **1**.

Mathematics Note

For samples of size n, the sampling distribution of $\hat{p}$ has a mean μ where $\mu = p$, the population proportion.

The standard deviation σ of the sampling distribution of $\hat{p}$ can be determined using the following formula, where n is the sample size and p is the population proportion:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

For example, consider a population in which the proportion of females is 0.6. The mean of the sampling distribution of $\hat{p}$ for samples of size 30 also is 0.6. The standard deviation of the sampling distribution of $\hat{p}$ for samples of size 30 can be found as follows:

$$\sigma = \sqrt{\frac{0.6(1-0.6)}{30}} \approx 0.089$$

c. Calculate the standard deviation of the sampling distribution of $\hat{p}$ for samples of size 4.

Compare this value with the standard deviation for samples of size 4 determined in Activity **1**.

d. Repeat Parts **a–c** for samples of size 30.

Discussion 2

a. Describe the method you used in Part **b** of Exploration **1** to calculate the mean of the sampling distribution of $\hat{p}$.

b. As the number of samples grows very large, how would you expect a relative frequency table to compare to the corresponding sampling distribution for $\hat{p}$?

c. Consider a population of fish in which 60% are infected with whirling disease. How could you use the sampling distribution for samples of size 30 to show that a sample containing 10 or fewer diseased fish is very unlikely?

Exploration 2

In Exploration **1**, you used sampling distributions to examine the probabilities of obtaining specific sample proportions. In this exploration, you use technology to investigate the probability that a sample proportion will fall within 1, 2, or 3 standard deviations of the population proportion.

a. Consider a population in which the proportion with a certain characteristic is $p = 0.10$. Create connected scatterplots of the sampling distributions when $n = 4$, 10, and 30. Describe any trends you observe.

b. Repeat Part **a** for population proportions of 0.30, 0.50 and 0.80. Describe any trends you observe.

c. Consider a population of trout in which 60% are infected with whirling disease.

1. Determine a sample size n that results in a connected scatterplot which appears bell-shaped and symmetric about the population proportion.
2. Construct a table showing the sampling distribution of $\hat{p}$.
3. Record the mean μ and standard deviation σ of the sampling distribution.

d. Create a connected scatterplot of the sampling distribution in Part **c**. On the x-axis of your graph, mark the values that correspond to $\mu + \sigma$, $\mu + 2\sigma$, $\mu + 3\sigma$, μ, $\mu - \sigma$, $\mu - 2\sigma$, and $\mu - 3\sigma$.

Estimate the percentage of sample proportions that fall in each of the following intervals:

1. $[\mu + \sigma,\ \mu - \sigma]$
2. $[\mu + 2\sigma,\ \mu - 2\sigma]$
3. $[\mu + 3\sigma,\ \mu - 3\sigma]$

Discussion 3

a. Which sample sizes in Exploration **2** resulted in connected scatterplots that appear bell-shaped and symmetric about the population proportion?

b. In Part **d** of Exploration **2**, what was the probability that a sample proportion fell within each of the following intervals?

1. $[\mu + \sigma,\ \mu - \sigma]$
2. $[\mu + 2\sigma,\ \mu - 2\sigma]$
3. $[\mu + 3\sigma,\ \mu - 3\sigma]$

c. How do these probabilities compare to those predicted by the 68–95–99.7 rule for normal distributions?

Mathematics Note

The **central limit theorem** states that, regardless of the population, as the sample size increases, the sampling distribution of sample proportions approaches a normal distribution.

As a rule of thumb, the properties implied by the central limit theorem are approximately true when the sample size n satisfies the conditions $np > 5$ and $n(1 - p) > 5$. For example, suppose one is reasonably sure that $0.25 \le p \le 0.75$. If $p = 0.25$, then:

$$\begin{array}{rr} np > 5 & n(1-p) > 5 \\ n(0.25) > 5 & n(1-0.25) > 5 \\ n > 20 & n > 6 \end{array}$$

Similarly, if $p = 0.75$, then:

$$\begin{array}{rr} np > 5 & n(1-p) > 5 \\ n(0.75) > 5 & n(1-0.75) > 5 \\ n > 6 & n > 20 \end{array}$$

In this case, the central limit theorem may provide useful interpretations for sample sizes greater than 20.

d. If the population proportion is between 0.10 and 0.90, what sample sizes might you select to ensure that the sample distribution can be approximated by a normal distribution? Explain your reasoning.

Assignment

2.1 Assume that the proportion of females in a trout population is 50%.

a. Construct the sampling distribution of $\hat{p}$ for samples of size 10.

b. What is the probability of obtaining a random sample from this population with less than 3 female trout?

c. What is the probability of obtaining a random sample from this population in which the number of females is more than 3 but less than 8?

d. Determine the probability that the proportion of females in a random sample of 10 trout from this population will be within 1 standard deviation of the population proportion.

2.2 Some individuals in a certain species of bird carry a gene for an enzyme deficiency. The proportion of the bird population in your area that has this gene is 12%.

a. Construct the sampling distribution of $\hat{p}$ for samples of size 50.

b. Use the central limit theorem to determine an interval in which a sample proportion should fall 95% of the time.

2.3 In 1994, 35% of the rainbow trout in the Raynolds Pass area near the western border of Yellowstone National Park were infected with whirling disease.

a. What sample size is required to make the standard deviation of the sampling distribution of $\hat{p}$ less than or equal to 0.05?

b. Calculate each of the following intervals for a sample size of 400:

1. $[\mu + \sigma,\ \mu - \sigma]$

2. $[\mu + 2\sigma,\ \mu - 2\sigma]$

3. $[\mu + 3\sigma,\ \mu - 3\sigma]$

c. Describe the percentage of sample proportions you would expect to fall in each interval in Part **b**.

2.4 A wildlife biologist would like to estimate the proportion of white rabbits in a population with four different color phases: white, brown, black, and mixed. The biologist thinks that at least one-fifth are white and at least one-fifth are not white. What is the minimum sample size the researcher should use to be confident that the sampling distribution will be bell-shaped, symmetrical, and follow the 68–95–99.7 rule?

* * * * *

2.5 One day's production at the RoboSpace Company resulted in a population of sprockets with a 20% defective rate.

a. Determine the probability that a random sample of 4 sprockets will contain each of the following:

1. exactly 1 defective sprocket

2. exactly 0 defective sprockets

3. no more than 2 defective sprockets.

b. For a sample size of 200, in what interval would you expect 95% of the sample proportions to fall?

2.6 A bolt manufacturing company estimates that the probability of manufacturing a defective bolt is 30%.

a. If the company's quality-control process uses samples of 300 bolts, what would you expect the mean and the standard deviation to be for the sampling distribution of $\hat{p}$?

b. In what interval would you expect 95% of the sample proportions to fall for a sample size of 300?

c. What would you conclude if you drew a random sample of 300 bolts in which 40% were defective?

* * * * * * * * * *

Activity 3

Consider a population of 30,000 trout. Using the sampling method described in Activity **1** and samples of size 100, the number of possible samples is $30,000^{100}$. Approximately 95% of these sample proportions fall in an interval within 2 standard deviations of the population proportion.

If you knew that the proportion of diseased trout in the population was 56%, this interval would be $\left[0.56 - 2(0.05), 0.56 + 2(0.05)\right]$ or $\left[0.46, 0.66\right]$. Unfortunately, researchers seldom know the population proportion before studying a population. In this activity, you investigate what a single sample proportion can tell you about an unknown population proportion.

Exploration

In this exploration, you examine the number of sample proportions that fall within a certain interval of the population proportion.

a. **1.** On a sheet of graph paper, create a number line from 0 to 1 with increments of 0.1.

2. Mark the proportion p of diseased fish in the fish population from Activity **1** on the number line.

b. Take a random sample of 30 fish from this population and calculate $\hat{p}$, the proportion of diseased fish in the sample.

c. To estimate the standard deviation σ of the sampling distribution of $\hat{p}$ for samples of size 30, substitute $\hat{p}$ for p in the following expression, where n represents the sample size:

$$\sqrt{\frac{p(1-p)}{n}}$$

Call this estimate s.

d. Using your results from Part **c**, determine the interval $\left[\hat{p} - 2s, \hat{p} + 2s\right]$. Draw this interval, to the nearest 0.01, above your number line. This is a **confidence interval** for the population proportion based on your sample.

e. Determine whether or not the actual population proportion falls within the confidence interval.

f. **1.** Use technology to repeat Parts **b–e** 19 more times. Draw each confidence interval separately above your number line. For example, Figure **1** shows three confidence intervals for samples taken from a population where $p = 0.56$.

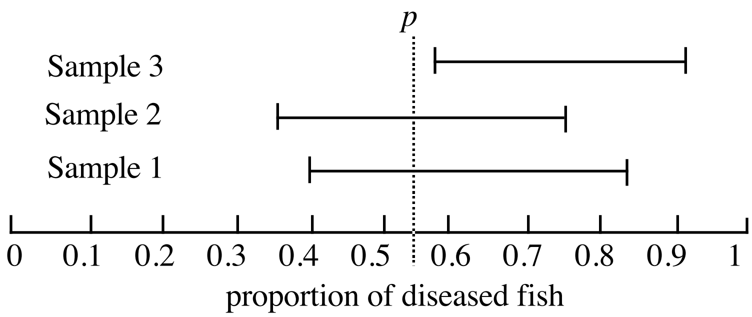


Figure 1: Number line with confidence intervals

2. Record the percentage of confidence intervals that contain the actual population proportion. In the example given in Figure **1**, approximately 67% of the confidence intervals contain p.

g. Repeat Parts **a–f** for samples of size 90.

Mathematics Note

A **confidence interval** for a population proportion is an interval in which the value of the population proportion is expected to be found. Every confidence interval has two aspects: an interval determined by the statistics collected from a random sample and a confidence level that gives the probability that the interval includes the parameter.

For example, a 95% confidence interval is generated by a process that results in an interval in which the probability that the parameter lies in that interval is 95%. In other words, you would expect 95% of the intervals produced by this process to contain the parameter, while 5% of the intervals would not.

Consider a population proportion p and a sample size n such that $np > 5$ and $n(1 - p) > 5$. The mean μ of all the sample proportions of size n equals the population proportion p. Since sample proportions $\hat{p}$ are normally distributed, the 68–95–99.7 rule can be applied.

For example, the proportion $\hat{p}$ of any one sample will fall within 2 standard deviations of p 95% of the time. This fact also indicates a 95% probability that p is within 2 standard deviations of $\hat{p}$. In other words, the 95% confidence interval for p is $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$, where

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

This assertion can be reworded as a **confidence statement** in the following form: "If $\hat{p}$ is the sample proportion for a random sample of size n, then you can be 95% confident that the actual population proportion falls in the interval $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$." When p is not known, the value of $\hat{p}$ is used to estimate σ and generate the confidence interval. This estimated standard deviation is denoted by s.

For example, consider a stream that contains several different species of fish. To determine the proportion of rainbow trout in the total fish population, biologists capture a random sample of 150 fish from the stream. Of this sample, 11 are rainbow trout. In this situation, $\hat{p} = 11/150 \approx 0.073$. Since p is not known, $\hat{p}$ is substituted for p to find s, the estimated standard deviation, as follows:

$$s \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.073\,(1-0.073)}{150}} = 0.021$$

Since $\hat{p} - 2s = 0.073 - 0.042 = 0.031$ and $\hat{p} + 2s = 0.073 + 0.042 = 0.115$, the 95% confidence interval for the population proportion is [0.031, 0.115].

Therefore, the researchers can conclude, with 95% confidence, that the proportion of rainbow trout in the total population is in the interval $[0.031, 0.115]$ or, in other words, that rainbow trout represent between 3.1% and 11.5% of the population in the stream.

Discussion

a.

1. Using a sample size of 30, what percentage of the confidence intervals in the exploration contained the population proportion?
2. Using a sample size of 90, what percentage of the confidence intervals in the exploration contained the population proportion?
3. Were these results consistent with the 68–95–99.7 rule?

b. Consider a random sample of 400 trout that contained 160 fish infected with whirling disease. Describe how to construct a confidence interval for the proportion of infected trout in the population using each of the following confidence levels:

1. 95%
2. 68%
3. 99.7%

c. Describe how sample size affects the width of the corresponding confidence intervals.

d. A newspaper reports that 40% of the rainbow trout in a certain section of the Madison River are diseased. To test that claim, a fisheries biologist takes a random sample of 300 rainbow trout from that section of the river.

1. What can the biologist conclude if 60 of the fish are diseased?
2. What can the biologist conclude if 100 of the fish are diseased?

Assignment

3.1 Fisheries managers are trying to determine the proportion of rainbow trout over 40 cm in length in a certain population. As part of their study, biologists collect random samples of 100 trout each day for 16 days. The following table shows the number of trout over 40 cm in each of these samples.

Day 1	6	Day 9	7
Day 2	8	Day 10	6
Day 3	6	Day 11	10
Day 4	10	Day 12	8
Day 5	9	Day 13	7
Day 6	7	Day 14	11
Day 7	8	Day 15	8
Day 8	14	Day 16	8

a. Use the data for one of the days to construct a 95% confidence interval for the proportion of rainbow trout over 40 cm in length and make a confidence statement about this interval.

b. Select a day with a different sample proportion from the one used in Part **a**. Construct a 95% confidence interval using this information and make a confidence statement about this interval.

c. Compare the two confidence intervals. How many different 95% confidence intervals would you expect if you determined one for each day in the table?

d. Would you expect all of the 95% confidence intervals obtained from the data to contain p, the actual proportion of rainbow trout in the river over 40 cm? Explain your response.

e. Describe how the data for all 16 days could be combined to determine a single 95% confidence interval.

f. Determine the 95% confidence interval that results from the technique you described in Part **e**.

3.2 To determine the proportion of rainbow trout in the total fish population, biologists capture a random sample of fish from a stream.

a. Determine a 95% confidence interval for the population proportion and make a confidence statement about this interval for each of the following situations:

1. a random sample of 50 fish contains 7 rainbow trout

2. a random sample of 100 fish contains 14 rainbow trout

3. a random sample of 400 fish contains 56 rainbow trout

b. Compare the sample proportions used to construct each confidence interval in Part **a.**

c. Write a paragraph explaining the significance of the results in Parts **a** and **b**.

3.3 In 1994, 35% of the rainbow trout in the Raynolds Pass area near Yellowstone Park were infected with whirling disease. It is feared that the proportion of diseased fish is increasing rapidly.

Imagine that you are a biologist assigned to determine if the disease rate in the Raynolds Pass area is still 35%. If a sample of 400 fish from the area contains 160 that are diseased, what can you conclude about the disease rate? Explain your response.

3.4 A wildlife biologist is studying the survival rates of calves in a population of elk. As part of her study, she fits a random sample of 100 calves with radio collars. At the end of 1 year, 73 of these calves are still alive.

a. Estimate the proportion of elk calves in the population that survive at least 1 year.

b. Estimate the standard deviation for the sampling distribution.

c. Determine a 95% confidence interval for the proportion of elk calves from this population that survive at least 1 year and make a confidence statement about this interval.

d. Describe how the confidence interval would be affected if the sample size was increased from 100 to 400 and the proportion of calves that survived remained the same.

e. Why is a small standard deviation advantageous to a researcher?

3.5 About 50 species of bats are found in North America. In the 1950s, bats became branded as carriers of rabies. Health departments around the country conducted rabies tests on dead bats found by local residents or their pets. Based on the thousands of bats tested each year, researchers estimated that 10% of the bats in North America carry rabies.

Suppose a group of scientists is studying bats in the southwestern United States. As part of the study, the group collects a random sample of 600 live bats from the region. Of this sample, 6 bats are found to carry rabies.

a. Using a confidence statement, estimate the proportion of rabid bats in the southwestern United States.

b. Based on your results, what conclusions would you draw regarding the claim that 10% of the bats in North America are rabid?

c. If the true proportion of rabid bats in the area you are studying is 10%, what is the probability of obtaining a random sample of 600 bats that contains 6 or fewer with rabies?

d. Describe some circumstances that may have affected the accuracy of the 10% estimate described above.

* * * * *

3.6 **a.** For samples of the same size from the same population, would a 99.7% confidence interval be wider or narrower than a 95% confidence interval? Explain your response.

b. Write an expression that will determine a 99.7% confidence interval given values for $\hat{p}$ and s.

c. Describe some factors that researchers should consider when deciding whether to use a 95% or a 99.7% confidence interval.

3.7 Before an election in Great Britain, a random sample of 100 voters indicated that 55% of them supported the party of the incumbent prime minister.

a. Determine a 95% confidence interval for the proportion of all voters who supported the incumbent's party.

b. Determine a 99.7% confidence interval for the proportion of all voters who supported the incumbent's party.

* * * * * * * * * *

Research Project

Select a population and a characteristic that you are interested in studying. For example, you may wish to estimate the percentage of students at your school who have part-time jobs. Design a sampling experiment that allows you to determine a reasonable estimate of the proportion of the population with this characteristic.

Activity 4

In 1991, biologists estimated that the rainbow trout population in one section of the Madison River was 2051 fish per km. In order to obtain this estimate, they used a sampling technique called **capture-recapture**. In this activity, you use this method to approximate the number of trout in a simulated river.

Exploration

Fishery biologists routinely "tag" or mark captured fish by clipping a small piece from one fin. This identifying mark allows researchers to recognize a recaptured fish. The fin will eventually grow back, but not before biologists have completed their study.

a. Obtain a container with an unknown number of beans. In this exploration, the container represents a river, while each bean represents a trout.

b. Select a random sample of 50 trout, without replacement, and "tag" each one with a marking pen.

c. Return the tagged fish to the river and mix them thoroughly with the rest of the population.

d. Select a random sample of 40 trout from the river and determine the proportion of tagged fish in the sample.

e. Using the estimated population proportion, estimate the number of trout in the river.

f. Determine a 68% confidence interval for the proportion of tagged fish.

g. Use the confidence interval from Part **f** to determine a 68% confidence interval for the number of rainbow trout in the river.

h.
1. Count the actual number of trout in the river.
2. Compare the estimate you made in Part **e** with the actual size of the population.

Discussion

a. What factors might affect the accuracy of a population estimate made using the capture-recapture method?

b. If no tagged fish are recaptured, what conclusions should researchers make about the population?

c. Was the actual size of the population within the confidence interval you determined in Part **g** of the exploration?

d. If you used the process described in the exploration to determine 100 estimates and their corresponding confidence intervals, how many of these intervals would you expect to contain the actual population size? Explain your response.

Assignment

4.1 In one capture-recapture experiment, biologists caught, tagged, and released 350 trout in a lake. Later, they caught 168 trout, 14 of which were tagged. Estimate the number of trout in the lake.

4.2 Imagine that you are a wildlife specialist monitoring the impact of oil-field development on wolves in Alaska.

a. Before any oil drilling occurred, you captured and tagged 50 wolves in one region of the state. Several days later, you captured 45 wolves, of which 15 were tagged. Estimate the wolf population in this region before any drilling took place.

b. One year after the last well was drilled, you captured and tagged 52 wolves with a different set of tags in the same region. One week later, you captured 40 wolves, of which 14 were wearing the new tags. Estimate the wolf population in this region after the oil field was developed.

c. Write a paragraph comparing your results from Parts **a** and **b.** Include statistical evidence to support whether you think that the wolf population is declining in this region of Alaska.

4.3 Seven fish are captured from a small pond. After being tagged, the fish are returned to the pond. A second sample of 6 fish from the pond contains 2 that were tagged earlier. Describe what conclusions, if any, you can make about the pond's fish population.

4.4 Imagine that you are a biologist monitoring the impact of a new sewage disposal site on the trout population downstream.

a. In the year before the site opened, you tagged a sample of 1000 trout in this section of the river. Several days later, you collected another sample of 1000 trout, of which 120 were tagged. Estimate the trout population in this section of the river before the site opened.

b. A year after the site opened, you tagged a sample of 750 trout in this section. Several days later, you captured a sample of 800 trout, of which 105 were tagged. Estimate the trout population in this section after the site opened.

c. Compare your results from Parts **a** and **b**. Do you think that the trout population is declining in this stretch of river? Explain your response.

d. How could you adjust your capture-recapture method to obtain more precise results and allow for more effective monitoring?

* * * * *

4.5 A team of researchers catches 100 bats in a large cave. After tagging the bats, they return them to the cave. A few days later, the scientists catch 500 bats in the same cave and find that 12 are tagged.

a. Determine a 68% confidence interval for the population of bats in the cave.

b. Determine a 95% confidence interval for the population of bats in the cave.

c. Describe any assumptions you made about this sampling method in order to determine your estimates in Parts **a** and **b**.

* * * * * * * * * *

Summary Assessment

1. Obtain a model population from your teacher. A certain proportion of this population has a given characteristic.

a. Use sampling to estimate the population proportion.

b. Describe the sampling distribution of $\hat{p}$ for your selected sample size, including the standard deviation of the distribution.

c. Write a report describing your best estimate of the population proportion and the procedure you used to obtain it. Include any data you collected, as well as the resulting statistics and confidence intervals.

2. Obtain a model population from your teacher.

a. Design a capture-recapture experiment to determine a confidence interval for the size of this population.

b. Write a report describing the process you used to determine your confidence interval.

Module Summary

- The proportion of a population that displays a certain characteristic can be denoted as p. The value of p is a **population proportion** and can be found as follows:

$$p = \frac{\text{number in population with the characteristic}}{\text{total population}}$$

- When a sample is taken from a population, the proportion of the sample that displays the characteristic is denoted as $\hat{p}$. The value of $\hat{p}$ is the **sample proportion** and can be found as shown below:

$$\hat{p} = \frac{\text{number in sample with characteristic}}{\text{total number in sample}}$$

- The **sampling distribution** of all possible sample proportions $\hat{p}$ for samples of size n is the set of probabilities associated with each possible value of $\hat{p}$.

- For samples of size n, the sampling distribution of $\hat{p}$ has a mean μ where $\mu = p$, the population proportion.

- The standard deviation σ of the sampling distribution of $\hat{p}$ can be determined using the following formula, where n is the sample size and p is the population proportion:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

- The **central limit theorem** states that, regardless of the population, as the sample size increases, the sampling distribution of sample proportions approaches a normal distribution. As a rule of thumb, the properties implied by the central limit theorem are approximately true when the sample size n satisfies the conditions $np > 5$ and $n(1-p) > 5$.

- A **confidence interval** for a population proportion is an interval in which the value of the population proportion is expected to be found. Every confidence interval has two aspects: an interval determined by the statistics collected from a random sample and a confidence level that gives the probability that the interval includes the parameter.

Selected References

Arya, J., and R. Lardner. *Mathematics for the Biological Sciences*. Englewood Cliffs, NJ: Prentice Hall, 1979.

Landwehr, J., J. Swift, and A. Watkins. *Quantitative Literacy Series: Exploring Surveys and Information from Samples*. Palo Alto, CA: Dale Seymour Publications, 1987.

Mendenhall, W., and T. Sincich. *Statistics for the Engineering and Computer Sciences*. Santa Clara, CA: Dellen Publishing Co., 1984.

Moore, D., and G. McCabe. *Introduction to the Practice of Statistics*. New York: W. H. Freeman and Co., 1993.

Piascik, C. *Applied Finite Mathematics for Business and the Social and Natural Sciences*. St. Paul, MN: West Publishing Co., 1992.

Pitman, J. *Probability*. New York: Springer-Verlag, 1993.

Triola, M. *Elementary Statistics*. Reading MA: Addison-Wesley, 1992.

Western Regional Environmental Education Council. *Project Wild: Secondary Activity Book*. Boulder, CO: Western Regional Environmental Education Council, 1985.