8.5 Repeated Sampling and Hypothesis Testing

Repeated Sampling

When you draw a sample from a population, you often use the sample mean, $\bar{x}$, as an estimate of the population mean, μ, and the sample standard deviation, s, as an estimate of the population standard deviation, σ. However, the statistics for a single sample may differ radically from those of the underlying population. Statisticians try to address this problem by repeated sampling. Do additional samples improve the accuracy of the estimate?

INVESTIGATE & INQUIRE: Simulating Repeated Sampling

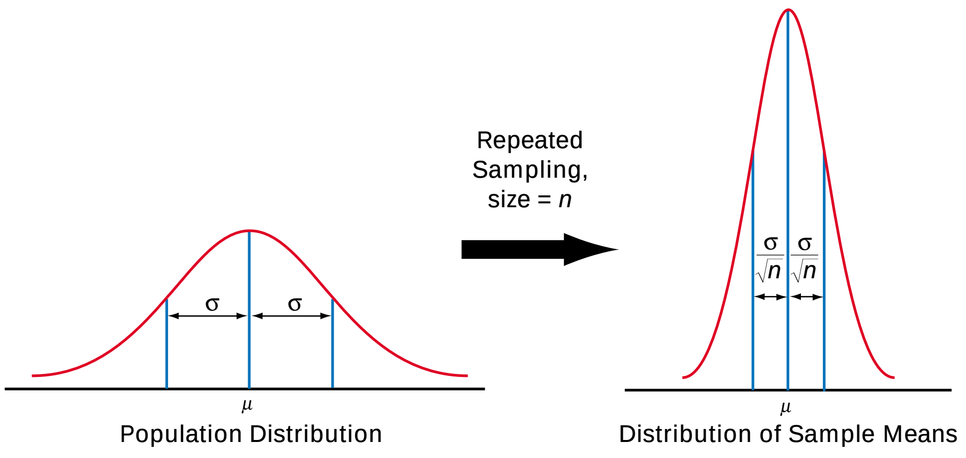
Simulate drawing samples of size 100 from a normally distributed population with mean $\mu = 10$ and standard deviation $\sigma = 5$. After 20 samples, examine the mean of the sample means.

The steps below outline a method using a graphing calculator. However, you can also simulate repeated sampling with a spreadsheet or statistical software such as Fathom™. See section 1.4 and the technology appendix descriptions of the software functions you could use.

1. Use the **mode settings** to set the number of decimal places to 2. Using the STAT EDIT menu, check that L1 and L2 are clear.
2. Place the cursor on L1. From the MATH PRB menu, select the **randNorm(function** and enter 10 as the mean, 5 as the standard deviation, and 100 as the number of trials. From the STAT CALC menu, select the **1-Var Stats command** to find the mean of the 100 random values in L1. Enter this mean in L2.
3. Repeat step 2 twenty times. You should then have 20 entries in L2. Each of these entries is the sample mean, $\bar{x}$, of a random sample of 100 drawn from a population with a mean $\mu = 10$ and a standard deviation $\sigma = 5$.
4. Use the **1-Var Stats command** to find the mean and standard deviation of L2.
5. Construct a histogram for L2. What does the shape of the histogram tell you about the distribution of the sample means?

When repeated samples of the same size are drawn from a normal population, the sample means will be normally distributed with a mean equal to the population mean μ.

The distribution of sample means will have a standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$, where n is the sample size. Notice that if $n = 1$, the samples are single data points and $\sigma_{\bar{x}} = \sigma$. As n increases, however, $\sigma_{\bar{x}}$ decreases, so the distribution of sample means becomes more tightly grouped around the true mean. Usually a sample size of at least 30 is sufficient to give a reasonably accurate estimate of μ.



Compare, from your investigation results, the mean from L2 with the true population mean, $\mu = 10$. Then, compare the standard deviation of L2 with $\dfrac{5}{\sqrt{100}} = 0.5$. How close are your results to these theoretical values?

Example 1 Tire Wear

The tires on a rental-car fleet have lifetimes that are normally distributed with a mean life of 64 000 km and a standard deviation of 4800 km. Every week a mechanic checks the tires on ten randomly selected cars.

a) What are the mean and standard deviation of these samples?

b) How likely is the mechanic to find a sample mean of 62 500 km or less?

Solution

a) The mean of the sample means will be

$$\begin{aligned}\mu_{\bar{x}} &= \mu \\ &= 64\ 000\end{aligned}$$

The sample size is 40, assuming all four tires on each of the ten cars are checked. Therefore, the standard deviation of the sample means will be

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
$$= \frac{4800}{\sqrt{40}}$$
$$\doteq 759$$

b) To find the z-score for a sample mean of 62 500 km, you need to use the mean and standard deviation of the sample means. Thus,

$$P(\bar{x} < 63\ 500) = P\left(Z < \frac{62\ 500 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$
$$\doteq P\left(Z < \frac{62\ 500 - 64\ 000}{759}\right)$$
$$= P(Z < -1.98)$$
$$= 0.0239$$

You can find this value easily using the table on pages 606 and 607 or the **normalcdf(function** on a graphing calculator. There is a 2.4% probability that the mechanic will find a sample mean of 62 500 km or less.

Common sense suggests that the mean of a good-sized random sample from a population gives a better estimate of the population mean than a single value taken randomly from the population. As you have just seen, this idea can be quantified: sample means from a normal population are distributed with much smaller standard deviations than single data points from the population. The table below compares the two distributions.

Statistic	Sample Size	Mean	Standard Deviation
Single data point, x	1	μ	σ
Sample mean, $\bar{x}$	n	μ	$\frac{\sigma}{\sqrt{n}}$

Hypothesis Testing

Statisticians use the standard deviation of the sample mean to quantify the reliability of statistical studies and conclusions. Often, it is not possible to determine with certainty whether a statement is true or false. However, it is possible to test the strength of the statement, based on a sample. This procedure is called a **hypothesis test**.

Consider this scenario. A large candy manufacturer produces chocolate bars with labels stating “45 g net weight.” A company spokesperson claims that the masses are normally distributed with a mean of 45 g and a standard deviation of 2 g. A mathematics class decides to check the truth of the 45 g label. They purchase and weigh 30 bars. The mean mass is 44.5 g. Is this evidence enough to challenge the company’s claim?

You can construct a hypothesis test to investigate the 45-g net-weight claim, using these steps.

- *Step 1* State the hypothesis being challenged, **null hypothesis** (null means no change in this context). The null hypothesis is usually denoted H_0. So, for the chocolate bars, H_0: $\mu = 45$.
- *Step 2* State the alternative hypothesis H_1 (sometimes called H_a). You suspect that the mean mass may be lower than 45 g; so, H_1: $\mu < 45$.
- *Step 3* Establish a decision rule. How strong must the evidence be to reject the null hypothesis? If, for a normal distribution with a population mean of 45 g, the probability of a sample with a mean of 44.5 g is *very small*, then getting such a sample would be strong evidence that the actual population mean is not 45 g. The **significance level**, α, is the probability threshold that you choose for deciding whether the observed results are rare enough to justify rejecting H_0. For example, if $\alpha = 0.05$, you are willing to be wrong 5% of the time.
- *Step 4* Conduct an experiment. For the chocolate bars, you weigh the sample of 30 bars.
- *Step 5* Assume H_0 is true. Calculate the probability of obtaining the results of the experiment given this assumption. If the standard deviation for chocolate bar weights is 2 g, then

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{30}} = 0.365$$

and

$$P(\bar{x} < 44.5) = P\left(Z < \frac{44.5 - 45}{0.365}\right) \doteq P(Z < -1.37) = 0.0853$$

The probability of a sample mean of 44.5 or less, given a sample of size 30 from an underlying normal distribution with a mean of 45, is 8.5%.

- *Step 6* Compare this probability to the significance level, α: 8.5% is greater than 5%.
- *Step* 7 Accept H_0 if the probability is greater than the significance level. Such probabilities show that the sample result is not sufficiently rare to support an alternative hypothesis. If the probability is less than the significance level reject H_0 and instead accept H_1. So, for the chocolate bars, you would accept H_0.
- *Step 8* Draw a conclusion. How reliable is the company's claim? In this case, the statistical evidence is slightly too weak to refute the company's claim.

Project Prep

You will be asked to conduct a hypothesis test when you complete your probability distributions project.

A significance level of 5% is the same as saying you want a **confidence level** of 95% that, if you should reject H_0, you are making the correct decision. Thus, with the chocolate bars, you can be only about 91% confident that H_1 is true. For a simple survey, a confidence level of 90% may be adequate, so you could choose to accept H_1 at that level. For more important decisions, such as the effect on humans of a new drug, a significance level of 1% or even less, corresponding to a confidence level of 99% or more, would be appropriate. The importance of hypothesis testing is that it allows a *quantifiable* check of such claims about value or effectiveness.

Example 2 Drug Effectiveness

A drug company tested a new drug on 250 pigs with swine flu. Historically, 20% of pigs contracting swine flu die from the disease. Of the 250 pigs treated with the new drug, 215 recovered. Make a hypothesis test of the drug's effectiveness, with a significance level of $\alpha = 1\%$.

a) Determine whether you can model this study with a normal distribution.

b) Set up a null hypothesis and an alternative hypothesis for the value of μ in the normal approximation.

c) Which hypothesis corresponds to the drug being effective? Explain.

d) Conduct the hypothesis test.

e) Can the company claim that its new drug is effective?

Solution

a) The data are discrete, there are only two possible outcomes, and successive trials are independent. So, the distribution is binomial with $n = 250$ and $p = 0.2$. Both np and nq are greater than 5. You can use the normal approximation to structure your test. So,

$$\begin{aligned} \mu &= np \\ &= 250 \times 0.2 \\ &= 50 \end{aligned} \qquad\qquad \begin{aligned} \sigma &= \sqrt{npq} \\ &= \sqrt{250(0.2)(0.8)} \\ &= \sqrt{40} \\ &\doteq 6.32 \end{aligned}$$

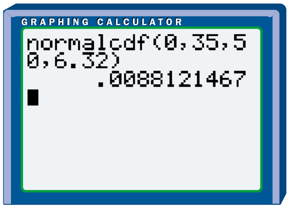
b) H_0: probability of death = 0.2 (the drug has no significant effect)

H_1: probability of death < 0.2 (the drug has some effect)

c) If the drug is effective, the probability of death will be reduced from the historical value of 20%, or 0.2. So, the alternate hypothesis, H_1, corresponds to the drug being effective.

d) Use software or a graphing calculator to find the probability that 35 or fewer pigs would die, using μ and σ from above.

$P(X < 35) \doteq 0.0088$



This probability is less than the significance level of $\alpha = 0.01$, indicating that the observed result would be a very rare event if the drug has no effect. Therefore, reject H_0.

e) The result of the hypothesis test is to accept H_1. The drug appears to have some effect on recovery rate, at the 1% significance level.

Key Concepts

- Sample means $\bar{x}$ from a normal population with mean μ and standard deviation σ are also normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$, where n is the sample size.
- Hypothesis tests assume the truth of the null hypothesis H_0 and investigate an alternative hypothesis H_1.
- The significance level α is the probability of error which the researcher is willing to accept.
- The confidence level $1 - \alpha$ is the probability that the decision is correct.
- If the probability of an experimental outcome is greater than the significance level, accept H_0.
- If the probability of an experimental outcome is less than the significance level, reject H_0.

Communicate Your Understanding

1. A researcher performed a hypothesis test, getting a result of $P(X < x) = 0.08$.
   a) Should the researcher accept or reject H_0 if $\alpha = 10\%$? Explain your answer.
   b) Should the researcher accept or reject H_0 if $\alpha = 5\%$? Explain your answer.

2. Outline the steps in a hypothesis test to determine whether getting 13 heads in 20 coin tosses is sufficient evidence to show that the coin is biased.

3. List at least four situations where hypothesis tests might be used. State H_0 and H_1 for each situation.

Practice

A

1. Copy and complete the table below.

Poplulation Mean	Population Standard Deviation	Sample Size	Mean of Sample Means	Standard Deviation of Sample Means
20	6	49		
12	4	25		
5	2	36		
40	8	100		
8.4	3.2	68		
17.6	10.4	87		
73.9	21.4	250		

B

2. For each situation, test the significance of the experimental results, given H_0 and H_1.

	Sample Size	Number of Successes	α	H_0	H_1
a)	50	23	10%	$p = 0.4$	$p > 0.4$
b)	200	55	5%	$p = 0.3$	$p < 0.3$
c)	250	175	1%	$p = 0.68$	$p > 0.68$
d)	40	8	10%	$p = 0.15$	$p > 0.15$
e)	400	80	1%	$p = 0.15$	$p > 0.15$

Apply , Solve, Communicate

3. **Application** A machine makes steel bearings with a mean diameter of 39 mm and a standard deviation of 3 mm. The bearing diameters are normally distributed. A quality-control technician found that in a sample of 50 bearings the mean diameter was 44 mm. Test the significance of this result with a significance level of 10%. Decide whether the machine needs to be adjusted.

4. **Application** A newspaper stated that 70% of the population supported a particular candidate's position on health care. In a random survey of 50 people, 31 agreed with the candidate's position. Test the significance of this result with a confidence level of 90%. Should the newspaper print a correction?

5. **Communication** A new drug will not be considered for acceptance by Health Canada unless it causes serious side effects in less than 0.01% of the population. In a trial with 80 000 people, 9 suffered serious side effects. Test the significance of this result with $\alpha = 0.01$. Do you recommend that this drug be accepted by Health Canada? Explain your answer.

6. A certain soft-drink manufacturer claims that its product holds 28% of the market. In a blind taste test, 13 out of 60 people chose this product. Does this test support or refute the soft drink manufacturer's claim? Choose a significance level you feel is appropriate for this situation.

7. **Inquiry/Problem Solving** An insurance company claims that 38% of automobile accidents occur within 5 km of home. The company examined 400 recent accidents and found that 120 occurred within 5 km of the driver's home. Does this result support or refute the company's claim? Choose a significance level and justify your choice.

8. **Inquiry/Problem Solving** A student-loan program claims that the average loan per student per year is $7500. Dana investigated this statement by asking 50 students about this year's student loan. The mean of the results was $5800. What additional information does Dana need to test the significance of this result? What significance level would be appropriate here? Why?

9. In finance, the *strong form* of the efficient-market hypothesis states that studying financial information about stocks is a waste of time, since all public and private information that might affect the stock's price is already reflected in the price of the stock. However, a study of 450 stocks found that only about 8% had price movements that could be accounted for in this way. At what significance level could you accept the strong form of the efficient-market hypothesis?

10. Does advertising influence behaviour? Before a recent advertising campaign, a children's breakfast cereal held 8% of the market. After the campaign, 18 families out of a sample of 200 families indicated they purchased the cereal. Was the advertising campaign a success? Select a confidence level you feel is appropriate for this situation.

11. Researchers often use repeated samples to test a hypothesis. Why do you think they use this method? Outline some of its advantages and disadvantages.

ACHIEVEMENT CHECK

Knowledge/Understanding	Thinking/Inquiry/Problem Solving	Communication	Application

12. A new medication is designed to lower cholesterol. The cholesterol level in a group of patients is normally distributed with a mean of 6.15 mmol/L and a standard deviation of 1.35 mmol/L. A sample of 40 people used the medication for 30 days, after which their mean cholesterol level was 5.87 mmol/L. The drug company wants to have a 95% confidence level that the drug is effective before releasing it.
   a) Should the company release the drug?
   b) The drug appears to have more effect if used for longer periods of time. What mean cholesterol level would the test group have to reach for the company to be confident about releasing the drug? Support your answer with mathematical calculations.

C

13. With a graphing calculator, you can use the **Z-Test instruction** to test a value for the mean of a normal distribution, based on a sample data set. Enter the first 20 years of the earthquake data on page 411. Use the **Z-Test instruction**. Switch the input to Data (as opposed to Stats). Choose and enter test values of μ_0 and σ, and also choose an alternative hypothesis. What information does this test give you?