2025 CFA®
Exam Prep

SchweserNotes™
Quantitative Methods, Economics, and Corporate Issuers

Level I Book 1

KAPLAN SCHWESER

Kaplan Schweser's Path to Success

Level I CFA® Exam

CFA®

Welcome

As the head of Advanced Designations at Kaplan Schweser, I am pleased to have the opportunity to help you prepare for the CFA® exam. Kaplan Schweser has decades of experience in delivering the most effective CFA exam prep products in the market and I know you will find them to be invaluable in your studies.

Our products are designed to be an integrated study solution across print and digital media to provide you the best learning experience, whether you are studying with a physical book, online, or on your mobile device.

Our core product, the SchweserNotes™, addresses all of the Topics, Learning Modules, and LOS in the CFA curriculum. Each reading in the SchweserNotes has been broken into smaller, bite-sized modules with Module Quizzes interspersed throughout to help you continually assess your comprehension. Online Topic Quizzes appear at the end of each Topic to help you assess your knowledge of the material before you move on to the next section.

All purchasers of the SchweserNotes receive online access to the Kaplan Schweser online platform (our learning management system or LMS) at www.Schweser.com. In the LMS, you will see a dashboard that tracks your overall progress and performance and also includes an Activity Feed, which provides structure and organization to the tasks required to prepare for the CFA exam. You also have access to the SchweserNotes, Module Quizzes, and Topic Quizzes content as well as the Module Videos (if purchased), which contain a short video that complements each module in the SchweserNotes. Look for the icons indicating where video content, Module Quizzes, and Topic Quizzes are available online. I strongly encourage you to use the dashboard to track your progress and stay motivated.

Again, thank you for trusting Kaplan Schweser with your CFA exam preparation. We're here to help you throughout your journey to become a CFA charterholder.

Regards,

Derek Burkett

Derek Burkett, CFA, FRM, CAIA
Vice President (Advanced Designations)

Book 1: Quantitative Methods, Economics, and Corporate Issuers

SchweserNotes™ 2025

Level I CFA®

SCHWESERNOTES™ 2025 LEVEL I CFA® BOOK 1: QUANTITATIVE METHODS, ECONOMICS, AND CORPORATE ISSUERS

Published in 2024 by Kaplan, Inc.

ISBN: 978-1-0788-4560-1

CONTENTS

CORPORATE ISSUERS

Learning Outcome Statements (LOS)

1. Rates and Returns

The candidate should be able to:

a. interpret interest rates as required rates of return, discount rates, or opportunity costs and explain an interest rate as the sum of a real risk-free rate and premiums that compensate investors for bearing distinct types of risk.
b. calculate and interpret different approaches to return measurement over time and describe their appropriate uses.
c. compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures.
d. calculate and interpret annualized return measures and continuously compounded returns, and describe their appropriate uses.
e. calculate and interpret major return measures and describe their appropriate uses.

2. The Time Value of Money in Finance

The candidate should be able to:

a. calculate and interpret the present value (PV) of fixed-income and equity instruments based on expected future cash flows.
b. calculate and interpret the implied return of fixed-income instruments and required return and implied growth of equity instruments given the present value (PV) and cash flows.
c. explain the cash flow additivity principle, its importance for the no-arbitrage condition, and its use in calculating implied forward interest rates, forward exchange rates, and option values.

3. Statistical Measures of Asset Returns

The candidate should be able to:

a. calculate, interpret, and evaluate measures of central tendency and location to address an investment problem.
b. calculate, interpret, and evaluate measures of dispersion to address an investment problem.
c. interpret and evaluate measures of skewness and kurtosis to address an investment problem.
d. interpret correlation between two variables to address an investment problem.

4. Probability Trees and Conditional Expectations

The candidate should be able to:

a. calculate expected values, variances, and standard deviations and demonstrate their application to investment problems.
b. formulate an investment problem as a probability tree and explain the use of conditional expectations in investment application.
c. calculate and interpret an updated probability in an investment setting using Bayes' formula.

5. Portfolio Mathematics

The candidate should be able to:

a. calculate and interpret the expected value, variance, standard deviation, covariances, and correlations of portfolio returns.
b. calculate and interpret the covariance and correlation of portfolio returns using a joint probability function for returns.
c. define shortfall risk, calculate the safety-first ratio, and identify an optimal portfolio using Roy's safety-first criterion.

6. Simulation Methods

The candidate should be able to:

a. explain the relationship between normal and lognormal distributions and why the lognormal distribution is used to model asset prices when using continuously compounded asset returns.
b. describe Monte Carlo simulation and explain how it can be used in investment applications.

c. describe the use of bootstrap resampling in conducting a simulation based on observed data in investment applications.

7. Estimation and Inference

The candidate should be able to:

a. compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.
b. explain the central limit theorem and its importance for the distribution and standard error of the sample mean.
c. describe the use of resampling (bootstrap, jackknife) to estimate the sampling distribution of a statistic.

8. Hypothesis Testing

The candidate should be able to:

a. explain hypothesis testing and its components, including statistical significance, Type I and Type II errors, and the power of a test.
b. construct hypothesis tests and determine their statistical significance, the associated Type I and Type II errors, and power of the test given a significance level.
c. compare and contrast parametric and nonparametric tests, and describe situations where each is the more appropriate type of test.

9. Parametric and Non-Parametric Tests of Independence

The candidate should be able to:

a. explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.
b. explain tests of independence based on contingency table data.

10. Simple Linear Regression

The candidate should be able to:

a. describe a simple linear regression model, how the least squares criterion is used to estimate regression coefficients, and the interpretation of these coefficients.
b. explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated.
c. calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.
d. describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.
e. calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable.
f. describe different functional forms of simple linear regressions.

11. Introduction to Big Data Techniques

The candidate should be able to:

a. describe aspects of "fintech" that are directly relevant for the gathering and analyzing of financial data.
b. describe Big Data, artificial intelligence, and machine learning.
c. describe applications of Big Data and Data Science to investment management.

12. Firms and Market Structures

The candidate should be able to:

a. determine and interpret breakeven and shutdown points of production, as well as how economies and diseconomies of scale affect costs under perfect and imperfect competition.
b. describe characteristics of perfect competition, monopolistic competition, oligopoly, and pure monopoly.
c. explain supply and demand relationships under monopolistic competition, including the optimal price and output for firms as well as pricing strategy.
d. explain supply and demand relationships under oligopoly, including the optimal price and output for firms as well as pricing strategy.

e. identify the type of market structure within which a firm operates and describe the use and limitations of concentration measures.

13. Understanding Business Cycles

The candidate should be able to:

a. describe the business cycle and its phases.
b. describe credit cycles.
c. describe how resource use, consumer and business activity, housing sector activity, and external trade sector activity vary over the business cycle and describe their measurement using economic indicators.

14. Fiscal Policy

The candidate should be able to:

a. compare monetary and fiscal policy.
b. describe roles and objectives of fiscal policy as well as arguments as to whether the size of a national debt relative to GDP matters.
c. describe tools of fiscal policy, including their advantages and disadvantages.
d. explain the implementation of fiscal policy and difficulties of implementation as well as whether a fiscal policy is expansionary or contractionary.

15. Monetary Policy

The candidate should be able to:

a. describe the roles and objectives of central banks.
b. describe tools used to implement monetary policy tools and the monetary transmission mechanism, and explain the relationships between monetary policy and economic growth, inflation, interest, and exchange rates.
c. describe qualities of effective central banks; contrast their use of inflation, interest rate, and exchange rate targeting in expansionary or contractionary monetary policy; and describe the limitations of monetary policy.
d. explain the interaction of monetary and fiscal policy.

16. Introduction to Geopolitics

The candidate should be able to:

a. describe geopolitics from a cooperation versus competition perspective.
b. describe geopolitics and its relationship with globalization.
c. describe functions and objectives of the international organizations that facilitate trade, including the World Bank, the International Monetary Fund, and the World Trade Organization.
d. describe geopolitical risk.
e. describe tools of geopolitics and their impact on regions and economies.
f. describe the impact of geopolitical risk on investments.

17. International Trade

The candidate should be able to:

a. describe the benefits and costs of international trade.
b. compare types of trade restrictions, such as tariffs, quotas, and export subsidies, and their economic implications.
c. explain motivations for and advantages of trading blocs, common markets, and economic unions.

18. Capital Flows and the FX Market

The candidate should be able to:

a. describe the foreign exchange market, including its functions and participants, distinguish between nominal and real exchange rates, and calculate and interpret the percentage change in a currency relative to another currency.
b. describe exchange rate regimes and explain the effects of exchange rates on countries' international trade and capital flows.
c. describe common objectives of capital restrictions imposed by governments.

19. Exchange Rate Calculations

The candidate should be able to:

a. calculate and interpret currency cross-rates.
b. explain the arbitrage relationship between spot and forward exchange rates and interest rates, calculate a forward rate using points or in percentage terms, and interpret a forward discount or premium.

20. Organizational Forms, Corporate Issuer Features, and Ownership

The candidate should be able to:

a. compare the organizational forms of businesses.
b. describe key features of corporate issuers.
c. compare publicly and privately owned corporate issuers.

21. Investors and Other Stakeholders

The candidate should be able to:

a. compare the financial claims and motivations of lenders and shareholders.
b. describe a company's stakeholder groups and compare their interests.
c. describe environmental, social, and governance factors of corporate issuers considered by investors.

22. Corporate Governance: Conflicts, Mechanisms, Risks, and Benefits

The candidate should be able to:

a. describe the principal-agent relationship and conflicts that may arise between stakeholder groups.
b. describe corporate governance and mechanisms to manage stakeholder relationships and mitigate associated risks.
c. describe potential risks of poor corporate governance and stakeholder management and benefits of effective corporate governance and stakeholder management.

23. Working Capital and Liquidity

The candidate should be able to:

a. explain the cash conversion cycle and compare issuers' cash conversion cycles.
b. explain liquidity and compare issuers' liquidity levels.
c. describe issuers' objectives and compare methods for managing working capital and liquidity.

24. Capital Investments and Capital Allocation

The candidate should be able to:

a. describe types of capital investments.
b. describe the capital allocation process, calculate net present value (NPV), internal rate of return (IRR), and return on invested capital (ROIC), and contrast their use in capital allocation.
c. describe principles of capital allocation and common capital allocation pitfalls.
d. describe types of real options relevant to capital investments.

25. Capital Structure

The candidate should be able to:

a. calculate and interpret the weighted-average cost of capital for a company.
b. explain factors affecting capital structure and the weighted-average cost of capital.
c. explain the Modigliani–Miller propositions regarding capital structure.
d. describe optimal and target capital structures.

26. Business Models

The candidate should be able to:

a. describe key features of business models.
b. describe various types of business models.

READING 1

RATES AND RETURNS

MODULE 1.1: INTEREST RATES AND RETURN MEASUREMENT

Video covering this content is available online.

LOS 1.a: Interpret interest rates as required rates of return, discount rates, or opportunity costs and explain an interest rate as the sum of a real risk-free rate and premiums that compensate investors for bearing distinct types of risk.

Interest rates measure the time value of money, although risk differences in financial securities lead to differences in their equilibrium interest rates. Equilibrium interest rates are the **required rate of return** for a particular investment, in the sense that the market rate of return is the return that investors and savers require to get them to willingly lend their funds. Interest rates are also referred to as **discount rates** and, in fact, the terms are often used interchangeably. If an individual can borrow funds at an interest rate of 10%, then that individual should discount payments to be made in the future at that rate to get their equivalent value in current dollars or other currencies. Finally, we can also view interest rates as the **opportunity cost** of current consumption. If the market rate of interest on 1-year securities is 5%, earning an additional 5% is the opportunity forgone when current consumption is chosen rather than saving (postponing consumption).

The **real risk-free rate** of interest is a theoretical rate on a single-period loan that contains no expectation of inflation and zero probability of default. What the real risk-free rate represents in economic terms is **time preference**, the degree to which current consumption is preferred to equal future consumption.

When we speak of a real rate of return, we are referring to an investor's increase in purchasing power (after adjusting for inflation). Because expected inflation in future periods is not zero, the rates we observe on U.S. Treasury bills (T-bills), for example, are essentially risk-free rates, but not real rates of return. T-bill rates are nominal risk-free rates because they contain an **inflation premium**. This is the relation:

$$(1 + \text{nominal risk-free rate}) = (1 + \text{real risk-free rate})(1 + \text{expected inflation rate})$$

Often, including in many parts of the CFA curriculum, this relation is approximated as follows:

$$\text{nominal risk-free rate} \approx \text{real risk-free rate} + \text{expected inflation rate}$$

Securities may have one or more types of risk, and each added risk increases the required rate of return. These types of risks are as follows:

- **Default risk.** This is the risk that a borrower will not make the promised payments in a timely manner.
- **Liquidity risk.** This is the risk of receiving less than fair value for an investment if it must be sold quickly for cash.
- **Maturity risk.** As we will see in the Fixed Income topic area, the prices of longer-term bonds are more volatile than those of shorter-term bonds. Longer-maturity bonds have more maturity risk than shorter-term bonds and require a maturity risk premium.

Each of these risk factors is associated with a risk premium that we add to the nominal risk-free rate to adjust for greater default risk, less liquidity, and longer maturity relative to a liquid, short-term, default risk-free rate such as that on T-bills. We can write the following:

$$\begin{aligned}\text{nominal rate of interest} = {} & \text{real risk-free rate} \\ & + \text{inflation premium} \\ & + \text{default risk premium} \\ & + \text{liquidity premium} \\ & + \text{maturity premium}\end{aligned}$$

LOS 1.b: Calculate and interpret different approaches to return measurement over time and describe their appropriate uses.

Holding period return (HPR) is simply the percentage increase in the value of an investment over a given period:

$$\text{holding period return} = \frac{\text{end-of-period value}}{\text{beginning-of-period value}} - 1$$

For example, a stock that pays a dividend during a holding period has an HPR for that period equal to:

$$\frac{P_t + Div_t}{P_0} - 1, \text{ or } \frac{P_t - P_0 + Div_t}{P_0}$$

If a stock is valued at €20 at the beginning of the period, pays €1 in dividends over the period, and at the end of the period is valued at €22, the HPR is:

$$\text{HPR} = (22 + 1) / 20 - 1 = 0.15 = 15\%$$

Returns over multiple periods reflect compounding. For example, given HPRs for Years 1, 2, and 3, the HPR for the entire three-year period is:

$$\text{HPR} = (1 + \text{HPR}_{\text{Year 1}})(1 + \text{HPR}_{\text{Year 2}})(1 + \text{HPR}_{\text{Year 3}}) - 1$$

Later in this reading, we will see that a return over multiple years is typically stated as an *annualized return* rather than an HPR.

Average Returns

The **arithmetic mean return** is the simple average of a series of periodic returns. It has the statistical property of being an unbiased estimator of the true mean of the underlying distribution of returns:

$$\text{arithmetic mean return} = \frac{(R_1 + R_2 + R_3 + \ldots + R_n)}{n}$$

The **geometric mean return** is a compound rate. When periodic rates of return vary from period to period, the geometric mean return will have a value less than the arithmetic mean return:

$$\text{geometric mean return} = \sqrt[n]{(1 + R_1) \times (1 + R_2) \times (1 + R_3) \times \ldots \times (1 + R_n)} - 1$$

For example, for returns R_t over three annual periods, the geometric mean return is calculated as the following example shows.

EXAMPLE: Geometric mean return

For the last three years, the returns for Acme Corporation common stock have been −9.34%, 23.45%, and 8.92%. Calculate the compound annual rate of return over the three-year period.

Answer:

$$R_G = \sqrt[3]{(1 - 0.0934) \times (1 + 0.2345) \times (1 + 0.0892)} - 1$$

$$= \sqrt[3]{0.9066 \times 1.2345 \times 1.0892} - 1$$

$$= \sqrt[3]{1.21903} - 1$$

$$R_G = 1.06825 - 1 = 6.825\%$$

Solve this type of problem with your calculator as follows:

- On the TI, enter 1.21903 [y^x] 3 [1/x] [=]
- On the HP, enter 1.21903 [ENTER] 3 [1/x] [y^x]

In the previous example, the geometric mean results in an annual rate of return because the holding periods were years. If the holding periods are other than years, the geometric mean is not the same as the annual return. The root for the geometric mean is the number of *periods*, while the root for the annual return is the number of *years*.

EXAMPLE: Geometric mean vs. annual return

For the last four semiannual periods, the 6-month holding period returns on an investment were 2.0%, 0.5%, −1.0%, and 1.5%. Calculate the geometric mean and the annual rate of return.

Answer:

$$\text{Geometric mean} = \sqrt[4]{(1+0.02)(1+0.005)(1-0.10)(1+0.015)} - 1$$
$$= 0.007435 = 0.7435\%.$$

This is the geometric mean of the 6-month holding period returns.

$$\text{Annual return} = \sqrt[2]{(1+0.02)(1+0.005)(1-0.10)(1+0.015)} - 1$$
$$= 0.0149 = 1.49\%.$$

The four semiannual periods equal two years, so to get an annual return we use 2 as the root.

PROFESSOR'S NOTE

The geometric mean is always less than or equal to the arithmetic mean, and the difference increases as the dispersion of the observations increases. The only time the arithmetic and geometric means are equal is when there is no variability in the observations (i.e., all observations are equal).

A **harmonic mean** is used for certain computations, such as the average cost of shares purchased over time. The harmonic mean is calculated as $\frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}}$ where there are *N* values of X_i.

EXAMPLE: Calculating average cost with the harmonic mean

An investor purchases $1,000 of mutual fund shares each month, and over the last three months, the prices paid per share were $8, $9, and $10. What is the average cost per share?

Answer:

$$\overline{X}_H = \frac{3}{\frac{1}{8} + \frac{1}{9} + \frac{1}{10}} = \$8.926 \text{ per share}$$

To check this result, calculate the total shares purchased as follows:

$$\frac{1{,}000}{8} + \frac{1{,}000}{9} + \frac{1{,}000}{10} = 336.11 \text{ shares}$$

The average price is $\frac{\$3{,}000}{336.11} = \$8.926$ per share.

The previous example illustrates the interpretation of the harmonic mean in its most common application. Note that the average price paid per share ($8.93) is less than the arithmetic average of the share prices, which is $\frac{8+9+10}{3} = 9$.

We can only calculate a harmonic mean of positive numbers. For a set of returns that includes negative numbers, we can treat them the same way we did with geometric means, using (1 + return) for each period, then subtracting 1 from the result.

EXAMPLE: Harmonic mean with negative returns

For four periods, the returns on an investment were 2.0%, 0.5%, −1.0%, and 1.5%. Calculate the harmonic mean of these returns.

Answer:

$$\text{Harmonic mean} = \frac{4}{\frac{1}{(1+0.02)} + \frac{1}{(1+0.005)} + \frac{1}{(1-0.01)} + \frac{1}{(1+0.015)}} - 1$$

$$= 0.007369 = 0.7369\%$$

The relationship among arithmetic, geometric, and harmonic means can be stated as follows:

$$\text{arithmetic mean} \times \text{harmonic mean} = (\text{geometric mean})^2$$

PROFESSOR'S NOTE

The proof of this is beyond the scope of the Level I exam.

For values that are not all equal, harmonic mean < geometric mean < arithmetic mean. This mathematical fact is the basis for the claimed benefit of purchasing the same money amount of mutual fund shares each month or each week. Some refer to this practice as **cost averaging**.

Measures of average return can be affected by outliers, which are unusual observations in a dataset. Two of the methods for dealing with outliers are a *trimmed mean* and a *winsorized mean*. We will examine these in our reading on Statistical Measures of Asset Returns.

Appropriate uses for the various return measures are as follows:

- **Arithmetic mean.** Include all values, including outliers.
- **Geometric mean.** Compound the rate of returns over multiple periods.
- **Harmonic mean.** Calculate the average share cost from periodic purchases in a fixed money amount.
- **Trimmed or winsorized mean.** Decrease the effect of outliers.

MODULE QUIZ 1.1

1. An interest rate is *best* interpreted as a:
   A. discount rate or a measure of risk.
   B. measure of risk or a required rate of return.
   C. required rate of return or the opportunity cost of consumption.
2. An interest rate from which the inflation premium has been subtracted is known as a:
   A. real interest rate.
   B. risk-free interest rate.
   C. real risk-free interest rate.
3. The harmonic mean of 3, 4, and 5 is:
   A. 3.74.
   B. 3.83.

C. 4.12.

4. **XYZ Corp. Annual Stock Returns**

Year	20X1	20X2	20X3	20X4	20X5	20X6
Return	22%	5%	−7%	11%	2%	11%

The mean annual return on XYZ stock is *most appropriately* calculated using the:

A. harmonic mean.
B. arithmetic mean.
C. geometric mean.

MODULE 1.2: TIME-WEIGHTED AND MONEY-WEIGHTED RETURNS

Video covering this content is available online.

LOS 1.c: Compare the money-weighted and time-weighted rates of return and evaluate the performance of portfolios based on these measures.

The **money-weighted return** applies the concept of the **internal rate of return (IRR)** to investment portfolios. An IRR is the interest rate at which a series of cash inflows and outflows sum to zero when discounted to their present value. That is, they have a **net present value (NPV)** of zero. The IRR and NPV are built-in functions on financial calculators that CFA Institute permits candidates to use for the exam.

PROFESSOR'S NOTE
We have provided an online video in the Resource Library on how to use the TI calculator. You can view it by logging in to your account at www.schweser.com.

The **money-weighted rate of return** is defined as the IRR on a portfolio, taking into account all cash inflows and outflows. The beginning value of the account is an inflow, as are all deposits into the account. All withdrawals from the account are outflows, as is the ending value.

EXAMPLE: Money-weighted rate of return

Assume an investor buys a share of stock for $100 at t = 0, and at the end of the year (t = 1), she buys an additional share for $120. At the end of Year 2, the investor sells both shares for $130 each. At the end of each year in the holding period, the stock paid a $2 per share dividend. What is the money-weighted rate of return?

Step 1: Determine the timing of each cash flow and whether the cash flow is an inflow (+), into the account, or an outflow (−), available from the account.

$t = 0$:	purchase of first share	$= +\$100.00$	inflow to account
$t = 1$:	purchase of second share	$= +\$120.00$	
	dividends from first share	$= \underline{-\$2.00}$	
	subtotal, $t = 1$	$+\$118.00$	inflow to account
$t = 2$:	dividend from two shares	$= -\$4.00$	
	proceeds from selling shares	$= \underline{-\$260.00}$	
	subtotal, $t = 2$	$-\$264.00$	outflow from account

Step 2: Net the cash flows for each period and set the PV of cash inflows equal to the PV of cash outflows.

$$PV_{inflows} = PV_{outflows}$$

$$\$100 + \frac{\$118}{(1 + r)} = \frac{\$264}{(1 + r)^2}$$

Step 3: Solve for *r* to find the money-weighted rate of return. This can be done using trial and error or by using the IRR function on a financial calculator or spreadsheet.

The intuition here is that we deposited \$100 into the account at *t* = 0, then added \$118 to the account at *t* = 1 (which, with the \$2 dividend, funded the purchase of one more share at \$120), and ended with a total value of \$264.

To compute this value with a financial calculator, use these net cash flows and follow the procedure(s) described to calculate the IRR:

$$\text{net cash flows: } CF_0 = +100;\ CF_1 = +120 - 2 = +118;$$
$$CF_2 = -260 + -4 = -264$$

Calculating money-weighted return with the TI Business Analyst II Plus®

Note the values for F01, F02, and so on, are all equal to 1.

Keystrokes	Explanation	Display
[CF] [2nd][CLR WORK]	Clear cash flow registers	CF0 = 0.00000
100 [ENTER]	Initial cash outlay	CF0 = +100.00000
[↓] 118 [ENTER]	Period 1 cash flow	C01 = +118.00000
[↓] [↓] 264 [+/−] [ENTER]	Period 2 cash flow	C02 = −264.00000
[IRR] [CPT]	Calculate IRR	IRR = 13.86122

The money-weighted rate of return for this problem is 13.86%.

PROFESSOR'S NOTE

In the preceding example, we entered the flows into the account as a positive and the ending value as a negative (the investor could withdraw this amount from the account). Note that there is no difference in the solution if we enter the cash flows into the account as negative values (out of the investor's pocket) and the ending value as a positive value (into the investor's pocket). As long as payments into the account and payments out of the account (including the ending value) are entered with opposite signs, the computed IRR will be correct.

Time-weighted rate of return measures compound growth and is the rate at which \$1 compounds over a specified performance horizon. Time-weighting is the process of averaging a set of values over time. The annual time-weighted return for an investment may be computed by performing the following steps:

Step 1: Value the portfolio immediately preceding significant additions or withdrawals. Form subperiods over the evaluation period that correspond to the dates of deposits and withdrawals.

Step 2: Compute the holding period return (HPR) of the portfolio for each subperiod.

Step 3: Compute the product of (1 + HPR) for each subperiod to obtain a total return for the entire measurement period [i.e., $(1 + HPR_1) \times (1 + HPR_2) \ldots (1 + HPR_n)] - 1$.

If the total investment period is greater than one year, you must take the geometric mean of the measurement period return to find the annual time-weighted rate of return.

EXAMPLE: Time-weighted rate of return

An investor purchases a share of stock at $t = 0$ for \$100. At the end of the year, $t = 1$, the investor buys another share of the same stock for \$120. At the end of Year 2, the investor sells both shares for \$130 each. At the end of both Years 1 and 2, the stock paid a \$2 per share dividend. What is the annual time-weighted rate of return for this investment? (This is the same investment as the preceding example.)

Answer:

Step 1: Break the evaluation period into two subperiods based on timing of cash flows.

Holding period 1:	Beginning value	= \$100
	Dividends paid	= \$2
	Ending value	= \$120
Holding period 2:	Beginning value	= \$240 (2 shares)
	Dividends paid	= \$4 (\$2 per share)
	Ending value	= \$260 (2 shares)

Step 2: Calculate the HPR for each holding period.

$$HPR_1 = [(\$120 + 2)/\$100] - 1 = 22\%$$

$$HPR_2 = [(\$260 + 4)/\$240] - 1 = 10\%$$

Step 3: Find the compound annual rate that would have produced a total return equal to the return on the account over the two-year period.

$$(1 + \text{time-weighted rate of return})^2 = (1.22)(1.10)$$

$$\text{time-weighted rate of return} = [(1.22)(1.10)]^{0.5} - 1 = 15.84\%$$

The time-weighted rate of return is not affected by the timing of cash inflows and outflows. In the investment management industry, time-weighted return is the preferred method of performance measurement because portfolio managers typically do not control the timing of deposits to and withdrawals from the accounts they manage.

In the preceding examples, the time-weighted rate of return for the portfolio was 15.84%, while the money-weighted rate of return for the same portfolio was 13.86%. The results are different because the money-weighted rate of return gave a larger weight to the Year 2 HPR, which was 10%, versus the 22% HPR for Year 1. This is because there was more money in the account at the beginning of the second period.

If funds are contributed to an investment portfolio just before a period of relatively poor portfolio performance, the money-weighted rate of return will tend to be lower than the time-weighted rate of return. On the other hand, if funds are contributed to a portfolio at a favorable time (just before a period of relatively high returns), the money-weighted rate of return will be higher than the time-weighted rate of return. The use of the time-weighted return removes these distortions, and thus provides a better measure of a manager's ability to select investments over the period. If the manager has complete control over money flows into and out of an account, the money-weighted rate of return would be the more appropriate performance measure.

MODULE QUIZ 1.2

1. An investor buys a share of stock for $40 at time $t = 0$, buys another share of the same stock for $50 at $t = 1$, and sells both shares for $60 each at $t = 2$. The stock paid a dividend of $1 per share at $t = 1$ and at $t = 2$. The periodic money-weighted rate of return on the investment is *closest* to:
   A. 22.2%.
   B. 23.0%.
   C. 23.8%.
2. An investor buys a share of stock for $40 at time $t = 0$, buys another share of the same stock for $50 at $t = 1$, and sells both shares for $60 each at $t = 2$. The stock paid a dividend of $1 per share at $t = 1$ and at $t = 2$. The time-weighted rate of return on the investment for the period is *closest* to:
   A. 24.7%.
   B. 25.7%.
   C. 26.8%.

MODULE 1.3: COMMON MEASURES OF RETURN

Video covering this content is available online.

LOS 1.d: Calculate and interpret annualized return measures and continuously compounded returns, and describe their appropriate uses.

Interest rates and market returns are typically stated as **annualized returns**, regardless of the actual length of the time period over which they occur. To annualize an HPR that is realized over a specific number of days, use the following formula:

$$\text{annualized return} = (1 + \text{HPR})^{365/\text{days}} - 1$$

EXAMPLE: Annualized return, shorter than one year

A saver deposits $100 into a bank account. After 90 days, the account balance is $100.75. What is the saver's annualized rate of return?

Answer:

$$\text{HPR} = \frac{100.75}{100} - 1 = 0.0075 = 0.75\%$$
$$\text{annualized return} = (1 + 0.0075)^{365/90} - 1 = 0.0308 = 3.08\%$$

EXAMPLE: Annualized return, longer than one year

An investor buys a 500-day government bill for $970 and redeems it at maturity for $1,000. What is the investor's annualized return?

Answer:

$$\text{HPR} = \frac{1000}{970} - 1 = 0.0309 = 3.09\%$$
$$\text{annualized return} = (1 + 0.0309)^{365/500} - 1 = 0.0225 = 2.25\%$$

In time value of money calculations (which we will address in more detail in our reading on The Time Value of Money in Finance), more frequent compounding has an impact on future value and present value computations. Specifically, because an increase in the frequency of compounding increases the effective interest rate, it also *increases* the future value of a given cash flow and *decreases* the present value of a given cash flow.

This is the general formula for the present value of a future cash flow:

$$\text{PV} = \text{FV}_\text{N}\left(1 + \frac{\text{r}}{\text{m}}\right)^{-\text{mN}}$$

where:
r = quoted annual interest rate
N = number of years
m = compounding periods per year

EXAMPLE: The effect of compounding frequency on FV and PV

Compute the PV of $1,000 to be received one year from now using a stated annual interest rate of 6% with a range of compounding periods.

Answer:

With semiannual compounding, m = 2:

$$\text{PV} = 1{,}000\left(1 + \frac{0.06}{2}\right)^{-2} = 942.60$$

With quarterly compounding, m = 4:

$$PV = 1{,}000\left(1 + \frac{0.06}{4}\right)^{-4} = 942.18$$

With monthly compounding, $m = 12$:

$$PV = 1{,}000\left(1 + \frac{0.06}{12}\right)^{-12} = 941.91$$

With daily compounding, $m = 365$:

$$PV = 1{,}000\left(1 + \frac{0.06}{365}\right)^{-365} = 941.77$$

Compounding Frequency Effect

Compounding Frequency	Interest Rate per Period	Present Value
Annual ($m = 1$)	6.000%	$943.40
Semiannual ($m = 2$)	3.000	942.60
Quarterly ($m = 4$)	1.500	942.18
Monthly ($m = 12$)	0.500	941.91
Daily ($m = 365$)	0.016438	941.77

The mathematical limit of shortening the compounding period is known as continuous compounding. Given an HPR, we can use the natural logarithm (ln, or LN on your financial calculator) to state its associated **continuously compounded return**:

$$R_{CC} = \ln(1 + HPR) = \ln\left(\frac{\text{ending value}}{\text{beginning value}}\right)$$

Notice that because the calculation is based on 1 plus the HPR, we can also perform it directly from the **price relative**. The price relative is just the end-of-period value divided by the beginning-of-period value.

EXAMPLE: **Calculating continuously compounded returns**

A stock was purchased for $100 and sold one year later for $120. Calculate the investor's annual rate of return on a continuously compounded basis.

Answer:

$$\ln\left(\frac{120}{100}\right) = 18.232\%$$

If we had been given the return (20%) instead, the calculation is this:

$$\ln(1 + 0.20) = 18.232\%$$

A useful property of continuously compounded rates of return is that they are additive for multiple periods. That is, a continuously compounded return from $t = 0$ to $t = 2$ is the sum of the continuously compounded return from $t = 0$ to $t = 1$ and the continuously compounded return from $t = 1$ to $t = 2$.

LOS 1.e: Calculate and interpret major return measures and describe their appropriate uses.

Gross return refers to the total return on a security portfolio before deducting fees for the management and administration of the investment account. **Net return** refers to the return after these fees have been deducted. Commissions on trades and other costs that are necessary to generate the investment returns are deducted in both gross and net return measures.

Pretax nominal return refers to the return before paying taxes. Dividend income, interest income, short-term capital gains, and long-term capital gains may all be taxed at different rates. **After-tax nominal return** refers to the return after the tax liability is deducted.

Real return is nominal return adjusted for inflation. Consider an investor who earns a nominal return of 7% over a year when inflation is 2%. The investor's approximate real return is simply 7 − 2 = 5%. The investor's exact real return is slightly lower: 1.07 / 1.02 − 1 = 0.049 = 4.9%.

Using the components of an interest rate we defined earlier, we can state a real return as follows:

$$(1 + \text{real return}) = \frac{(1 + \text{nominal risk-free rate})(1 + \text{risk premium})}{(1 + \text{inflation premium})}$$

PROFESSOR'S NOTE

The Level I curriculum states this relationship as

$$(1 + \text{real return}) = \frac{(1 + \text{real risk-free rate})(1 + \text{risk premium})}{(1 + \text{inflation premium})}.$$

Stating it this way assumes the risk premium includes inflation risk.

Real return measures the increase in an investor's purchasing power—how much more goods she can purchase at the end of one year due to the increase in the value of her investments. If she invests $1,000 and earns a nominal return of 7%, she will have $1,070 at the end of the year. If the price of the goods she consumes has gone up 2%, from $1.00 to $1.02, she will be able to consume 1,070 / 1.02 = 1,049 units. She has given up consuming 1,000 units today, but instead, she is able to purchase 1,049 units at the end of one year. Her purchasing power has gone up 4.9%; this is her real return.

A **leveraged return** refers to a return to an investor that is a multiple of the return on the underlying asset. The leveraged return is calculated as the gain or loss on the investment as a percentage of an investor's cash investment. An investment in a derivative security, such as a futures contract, produces a leveraged return because the cash deposited is only a fraction of the value of the assets underlying the futures contract. Leveraged investments in real estate are common: investors pay only a portion of a property's cost in cash and borrow the rest.

To illustrate the effect of leverage on returns, consider a fund that can invest the amount V_0 without leverage, and earn the rate of return *r*. The fund's unleveraged return (as a money amount) is simply $r \times V_0$. Now let's say this fund can borrow the amount V_B at an interest rate of r_B, and earn *r* by investing the proceeds. The fund's leveraged return (again as a money amount), after subtracting the interest cost, then becomes $r \times (V_0 + V_B) - (r_B \times V_B)$.

Thus, stated as a rate of return on the initial value of V_0, the leveraged rate of return is as follows:

$$\text{leveraged return} = \frac{r(V_0 + V_B) - r_B V_B}{V_0}$$

MODULE QUIZ 1.3

1. If an investment loses 3% of its value over 120 days, its annualized return is *closest* to:
   A. −8.0%.
   B. −8.5%.
   C. −9.0%.
2. If a stock's initial price is \$20 and its price increases to \$23, its continuously compounded rate of return is *closest* to:
   A. 13.64%.
   B. 13.98%.
   C. 15.00%.
3. The value of an investment increases 5% before commissions and fees. This 5% increase represents:
   A. the investment's net return.
   B. the investment's gross return.
   C. neither the investment's gross return nor its net return.

KEY CONCEPTS

LOS 1.a

An interest rate can be interpreted as the rate of return required in equilibrium for a particular investment, the discount rate for calculating the present value of future cash flows, or as the opportunity cost of consuming now, rather than saving and investing.

The real risk-free rate reflects time preference for present goods versus future goods. Nominal risk-free rate ≈ real risk-free rate + expected inflation rate.

Securities may have several risks, and each increases the required rate of return. These include default risk, liquidity risk, and maturity risk.

We can view a nominal interest rate as the sum of a real risk-free rate, expected inflation, a default risk premium, a liquidity premium, and a maturity premium.

LOS 1.b

Holding period return is used to measure an investment's return over a specific period. Arithmetic mean return is the simple average of a series of periodic returns. Geometric mean return is a compound annual rate.

Arithmetic mean return includes all observations, including outliers. Geometric mean return is used for compound returns over multiple periods. Harmonic mean is used to calculate the average price paid with equal periodic investments. Trimmed mean or winsorized mean are used to reduce the effect of outliers.

LOS 1.c

The money-weighted rate of return is the IRR calculated using periodic cash flows into and out of an account and is the discount rate that makes the PV of cash inflows equal to the PV of cash outflows.

The time-weighted rate of return measures compound growth and is the rate at which money compounds over a specified performance horizon.

If funds are added to a portfolio just before a period of poor performance, the money-weighted return will be lower than the time-weighted return. If funds are added just before a period of high returns, the money-weighted return will be higher than the time-weighted return.

The time-weighted return is the preferred measure of a manager's ability to select investments. If the manager controls the money flows into and out of an account, the money-weighted return is the more appropriate performance measure.

LOS 1.d

Interest rates and market returns are typically stated on an annualized basis:

$$\text{annualized return} = (1 + \text{HPR})^{365/\text{days}} - 1$$

Given a holding period return, this is the associated continuously compounded return:

$$R_{cc} = \ln(1 + \text{HPR}) = \ln\left(\frac{\text{ending value}}{\text{beginning value}}\right)$$

LOS 1.e

Gross return is the total return after deducting commissions on trades and other costs necessary to generate the returns, but before deducting fees for the management and administration of the investment account. Net return is the return after management and administration fees have been deducted.

Pretax nominal return is the numerical percentage return of an investment, without considering the effects of taxes and inflation. After-tax nominal return is the numerical return after the tax liability is deducted, without adjusting for inflation.

Real return is the increase in an investor's purchasing power, roughly equal to nominal return minus inflation.

Leveraged return is the gain or loss on an investment as a percentage of an investor's cash investment.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 1.1

1. **C** Interest rates can be interpreted as required rates of return, discount rates, or opportunity costs of current consumption. A risk premium can be, but is not always, a component of an interest rate. (LOS 1.a)

2. **A** Real interest rates are those that have been adjusted for inflation. (LOS 1.a)

3. **B** $\overline{X}_H = \dfrac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = 3.83$

   (LOS 1.b)

4. **C** Because returns are compounded, the geometric mean is appropriate.

   $[(1.22)(1.05)(0.93)(1.11)(1.02)(1.11)]^{1/6} - 1 = 6.96\%$

   (LOS 1.b)

Module Quiz 1.2

1. **C** Using the cash flow functions on your financial calculator, enter CF_0 = −40; CF_1 = −50 + 1 = −49; CF_2 = 60 ´ 2 + 2 = 122; and CPT IRR = 23.82%. (LOS 1.c)

2. **A** $HPR_1 = \dfrac{50+1}{40} - 1 = 27.5\%$

   $HPR_2 = \dfrac{120+2}{100} - 1 = 22.0\%$

   $TWR = \sqrt{(1+0.275)(1+0.22)} - 1 = 24.72\%$

   (LOS 1.c)

Module Quiz 1.3

1. **C** $\text{Annualized return} = (1-0.03)^{365/120} - 1 = -0.0885 = -8.85\%$

   (LOS 1.d)

2. **B** ln (23/20) = 0.1398

   (LOS 1.d)

3. **C** Gross return is the total return after deducting commissions on trades and other costs necessary to generate the returns, but before deducting fees for the management and administration of the investment account. Net return is the return after management and administration fees have been deducted. (LOS 1.e)

READING 2

THE TIME VALUE OF MONEY IN FINANCE

PROFESSOR'S NOTE

The examples we use in this reading are meant to show how the time value of money appears throughout finance. Don't worry if you are not yet familiar with the securities we describe in this reading. We will see these examples again when we cover bonds and forward interest rates in Fixed Income, stocks in Equity Investments, foreign exchange in Economics, and options in Derivatives.

WARM-UP: USING A FINANCIAL CALCULATOR

For the exam, you must be able to use a financial calculator when working time value of money problems. You simply do not have the time to solve these problems any other way.

CFA Institute allows only two types of calculators to be used for the exam: (1) the Texas Instruments® TI BA II Plus™ (including the BA II Plus Professional™) and (2) the HP® 12C (including the HP 12C Platinum). This reading is written primarily with the TI BA II Plus in mind. If you do not already own a calculator, purchase a TI BA II Plus! However, if you already own the HP 12C and are comfortable with it, by all means, continue to use it.

Before we begin working with financial calculators, you should familiarize yourself with your TI BA II Plus by locating the keys noted below. These are the only keys you need to know to calculate virtually all of the time value of money problems:

- N = number of compounding periods
- I/Y = interest rate per compounding period
- PV = present value
- FV = future value
- PMT = annuity payments, or constant periodic cash flow
- CPT = compute

The TI BA II Plus comes preloaded from the factory with the periods per year function (P/Y) set to 12. This automatically converts the annual interest rate (I/Y) into monthly rates. While appropriate for many loan-type problems, this feature is not suitable for the vast majority of the time value of money applications we will be studying. So, before using our SchweserNotes™, please set your P/Y key to "1" using the following sequence of keystrokes:

[2nd] [P/Y] "1" [ENTER] [2nd] [QUIT]

As long as you do not change the P/Y setting, it will remain set at one period per year until the battery from your calculator is removed (it does not change when you turn the calculator on and off). If you want to check this setting at any time, press [2nd] [P/Y]. The display should read P/Y = 1.0. If it does, press [2nd] [QUIT] to get out of the "programming" mode. If it does not, repeat the procedure previously described to set the P/Y key. With P/Y set to equal 1, it is now possible to think of I/Y as the interest rate per compounding period and *N* as the number of compounding periods under analysis. Thinking of these keys in this way should help you keep things straight as we work through time value of money problems.

PROFESSOR'S NOTE
We have provided an online video in the Resource Library on how to use the TI calculator. You can view it by logging in to your account at www.schweser.com.

MODULE 2.1: DISCOUNTED CASH FLOW VALUATION

Video covering this content is available online.

LOS 2.a: Calculate and interpret the present value (PV) of fixed-income and equity instruments based on expected future cash flows.

In our Rates and Returns reading, we gave examples of the relationship between present values and future values. We can simplify that relationship as follows:

$$FV = PV\,(1 + r)^t$$

$$PV = \frac{FV}{(1 + r)^t} = FV(1 + r)^{-t}$$

where:
r = interest rate per compounding period
t = number of compounding periods

If we are using continuous compounding, this is the relationship:

$$FV = PV \times e^{rt}$$

$$PV = FV \times e^{-rt}$$

Fixed-Income Securities

One of the simplest examples of the time value of money concept is a **pure discount** debt instrument, such as a **zero-coupon bond**. With a pure discount instrument, the investor pays less than the face value to buy the instrument and receives the face value at maturity. The price the investor pays depends on the instrument's **yield to maturity** (the discount rate applied to the face value) and the time until maturity. The amount of interest the investor earns is the difference between the face value and the purchase price.

EXAMPLE: Zero-coupon bond

A zero-coupon bond with a face value of $1,000 will mature 15 years from today. The bond has a yield to maturity of 4%. Assuming annual compounding, what is the bond's price?

Answer:

$$\text{PV} = \frac{\$1{,}000}{(1 + 0.04)^{15}} = \$555.26$$

We can infer a bond's yield from its price using the same relationship. Rather than solving for *r* with algebra, we typically use our financial calculators. For this example, if we were given the price of $555.26, the face value of $1,000, and annual compounding over 15 years, we would enter the following:

$\text{PV} = -555.26$

$\text{FV} = 1{,}000$

$\text{PMT} = 0$

$\text{N} = 15$

Then, to get the yield, CPT I/Y = 4.00.

PROFESSOR'S NOTE

Remember to enter cash outflows as negative values and cash inflows as positive values. From the investor's point of view, the purchase price (PV) is an outflow, and the return of the face value at maturity (FV) is an inflow.

In some circumstances, interest rates can be negative. A zero-coupon bond with a negative yield would be priced at a **premium**, which means its price is greater than its face value.

EXAMPLE: Zero-coupon bond with a negative yield

If the bond in the previous example has a yield to maturity of −0.5%, what is its price, assuming annual compounding?

Answer:

$$\text{PV} = \frac{\$1{,}000}{(1 - 0.005)^{15}} = \$1{,}078.09$$

A **fixed-coupon bond** is only slightly more complex. With a coupon bond, the investor receives a cash interest payment each period in addition to the face value at maturity. The bond's **coupon rate** is a percentage of the face value and determines the amount of the interest payments. For example, a 3% annual coupon, $1,000 bond pays 3% of $1,000, or $30, each year.

The coupon rate and the yield to maturity are two different things. We only use the coupon rate to determine the coupon payment (PMT). The yield to maturity (I/Y) is the discount rate implied by the bond's price.

EXAMPLE: Price of an annual coupon bond

Consider a 10-year, $1,000 par value, 10% coupon, annual-pay bond. What is the value of this bond if its yield to maturity is 8%?

Answer:

The coupon payments will be 10% × $1,000 = $100 at the end of each year. The $1,000 par value will be paid at the end of Year 10, along with the last coupon payment.

The value of this bond with a discount rate (yield to maturity) of 8% is:

$$\frac{100}{1.08}+\frac{100}{1.08^2}+\frac{100}{1.08^3}+\ldots+\frac{100}{1.08^9}+\frac{1{,}100}{1.08^{10}}=1{,}134.20$$

The calculator solution is:

$$\text{N}=10;\ \text{PMT}=100;\ \text{FV}=1{,}000;\ \text{I/Y}=8;\ \text{CPT PV}=-1{,}134.20$$

The bond's value is $1,134.20.

PROFESSOR'S NOTE

For this reading where we want to illustrate time value of money concepts, we are only using annual coupon payments and compounding periods. In the Fixed Income topic area, we will also perform these calculations for semiannual-pay bonds.

Some bonds exist that have no maturity date. We refer to these as **perpetual bonds** or **perpetuities**. We cannot speak meaningfully of the future value of a perpetuity, but its present value simplifies mathematically to the following:

$$\text{PV of a perpetuity}=\frac{\text{payment}}{\text{r}}$$

An **amortizing bond** is one that pays a level amount each period, including its maturity period. The difference between an amortizing bond and a fixed-coupon bond is that for an amortizing bond, each payment includes some portion of the principal. With a fixed-coupon bond, the entire principal is paid to the investor on the maturity date.

Amortizing bonds are an example of an **annuity** instrument. For an annuity, the payment each period is calculated as follows:

$$\text{annuity payment} = \frac{r \times PV}{1 - (1+r)^{-t}}$$

where:

r = interest rate per period

t = number of periods

PV = present value (principal)

We can also determine an annuity payment using a financial calculator.

> **EXAMPLE: Computing a loan payment**
>
> Suppose you are considering applying for a $2,000 loan that will be repaid with equal end-of-year payments over the next 13 years. If the annual interest rate for the loan is 6%, how much are your payments?
>
> **Answer:**
>
> The size of the end-of-year loan payment can be determined by inputting values for the three known variables and computing PMT. Note that FV = 0 because the loan will be fully paid off after the last payment:
>
> $$N = 13;\ I/Y = 6;\ PV = -2{,}000;\ FV = 0;\ CPT \rightarrow PMT = \$225.92$$

Equity Securities

As with fixed-income securities, we value **equity securities** such as common and preferred stock as the present value of their future cash flows. The key differences are that equity securities do not mature, and their cash flows may change over time.

Preferred stock pays a fixed dividend that is stated as a percentage of its **par value** (similar to the face value of a bond). As with bonds, we must distinguish between the stated percentage that determines the cash flows and the discount rate we apply to the cash flows. We say that equity investors have a **required return** that will induce them to own an equity share. This required return is the discount rate we use to value equity securities.

Because we can consider a preferred stock's fixed stream of dividends to be infinite, we can use the perpetuity formula to determine its value:

$$\text{preferred stock value} = \frac{D_p}{k_p}$$

where:

D_p = dividend per period

k_p = the market's required return on the preferred stock

EXAMPLE: Preferred stock valuation

A company's $100 par preferred stock pays a $5.00 annual dividend and has a required return of 8%. Calculate the value of the preferred stock.

Answer:

Value of the preferred stock: D_p/k_P = $5.00/0.08 = $62.50

Common stock is a residual claim to a company's assets after it satisfies all other claims. Common stock typically does not promise a fixed dividend payment. Instead, the company's management decides whether and when to pay common dividends.

Because the future cash flows are uncertain, we must use models to estimate the value of common stock. Here, we will look at three approaches analysts use frequently, which we call **dividend discount models (DDMs)**. We will return to these examples in the Equity Investments topic area and explain when each model is appropriate.

1. *Assume a constant future dividend*. Under this assumption, we can value a common stock the same way we value a preferred stock, using the perpetuity formula.
2. *Assume a constant growth rate of dividends*. With this assumption, we can apply the **constant growth DDM**, also known as the **Gordon growth model**. In this model, we state the value of a common share as follows:

$$V_0 = \frac{D_1}{k_e - g_c}$$

where:

V_0 = value of a share *this* period

D_1 = dividend expected to be paid *next* period

k_e = required return on common equity

g_c = constant growth rate of dividends

In this model, V_0 represents the PV of all the dividends in future periods, beginning with D_1. Note that k_e must be greater than g_c or the math will not work.

EXAMPLE: Gordon growth model valuation

Calculate the value of a stock that is expected to pay a $1.62 dividend next year, if dividends are expected to grow at 8% forever and the required return on equity is 12%.

Answer:

$$\text{Calculate the stock's value} = D_1/(k_e - g_c)$$
$$= \$1.62/(0.12 - 0.08)$$
$$= \$40.50$$

3. *Assume a changing growth rate of dividends*. This can be done in many ways. The example we will use here (and the one that is required for the Level I CFA exam) is known as a **multistage DDM**. Essentially, we assume a pattern of dividends in the short term, such as a period of high growth, followed by a constant growth rate of dividends in the long term.

   To use a multistage DDM, we discount the expected dividends in the short term as individual cash flows, then apply the constant growth DDM to the long term. As we saw in the previous example, the constant growth DDM gives us a value for an equity share *one period before* the dividend we use in the numerator.

EXAMPLE: Multistage growth

Consider a stock with dividends that are expected to grow at 15% per year for two years, after which they are expected to grow at 5% per year, indefinitely. The last dividend paid was \$1.00, and k_e = 11%. Calculate the value of this stock using the multistage growth model.

Answer:

Calculate the dividends over the high growth period:

$$D_1 = D_0(1 + g^*) = 1.00(1.15) = \$1.15$$
$$D_2 = D_1(1 + g^*) = 1.15(1.15) = 1.15^2 = \$1.32$$

Calculate the first dividend of the constant-growth period:

$$D_3 = D_2(1 + g) = 1.32 \times 1.05 = \$1.386$$

Use the constant growth model to get P_2, a value for all the (infinite) dividends expected from time = 3 onward:

$$P_2 = \frac{D_3}{k_e - g_c} = \frac{1.386}{0.11 - 0.05} = \$23.10$$

Finally, we can sum the present values of dividends 1 and 2 and of P_2 to get the present value of all the expected future dividends during both the high-growth and constant-growth periods:

$$\frac{1.15}{1.11} + \frac{1.32 + 23.10}{(1.11)^2} = \$20.86$$

PROFESSOR'S NOTE

A key point to notice in this example is that when we applied the dividend in Period 3 to the constant growth model, it gave us a value for the stock in Period 2. To get a value for the stock today, we had to discount this value back by two periods, along with the dividend in Period 2 that was not included in the constant growth value.

MODULE QUIZ 2.1

1. Terry Corporation preferred stock is expected to pay a $9 annual dividend in perpetuity. If the required rate of return on an equivalent investment is 11%, one share of Terry preferred should be worth:
   A. $81.82.
   B. $99.00.
   C. $122.22.
2. Dover Company wants to issue a $10 million face value of 10-year bonds with an annual coupon rate of 5%. If the investors' required yield on Dover's bonds is 6%, the amount the company will receive when it issues these bonds (ignoring transactions costs) will be:
   A. less than $10 million.
   B. equal to $10 million.
   C. greater than $10 million.

MODULE 2.2: IMPLIED RETURNS AND CASH FLOW ADDITIVITY

Video covering this content is available online.

LOS 2.b: Calculate and interpret the implied return of fixed-income instruments and required return and implied growth of equity instruments given the present value (PV) and cash flows.

The examples we have seen so far illustrate the relationships among present value, future cash flows, and the required rate of return. We can easily rearrange these relationships and solve for the required rate of return, given a security's price and its future cash flows.

EXAMPLE: Rate of return for a pure discount bond

A zero-coupon bond with a face value of $1,000 will mature 15 years from today. The bond's price is $650. Assuming annual compounding, what is the investor's annualized return?

Answer:

$$\frac{\$1{,}000}{(1+r)^{15}} = \$650$$

$$(1+r)^{15} = \frac{\$1{,}000}{\$650} = 1.5385$$

$$r = 1.5385^{1/15} - 1 = 0.0291 = 2.91\%$$

EXAMPLE: Yield of an annual coupon bond

Consider the 10-year, $1,000 par value, 10% coupon, annual-pay bond we examined in an earlier example, when its price was $1,134.20 at a yield to maturity of 8%. What is its yield to maturity if its price decreases to $1,085.00?

Answer:

N = 10; PMT = 100; FV = 1,000; PV = −1,085; CPT I/Y = 8.6934

The bond's yield to maturity increased to 8.69%.

Notice that the relationship between prices and yields is inverse. *When the price decreases, the yield to maturity increases. When the price increases, the yield to maturity decreases.* Or, equivalently, *when the yield increases, the price decreases. When the yield decreases, the price increases*. We will use this concept again and again when we study bonds in the Fixed Income topic area.

In our examples for equity share values, we assumed the investor's required rate of return. In practice, the required rate of return on equity is not directly observable. Instead, we use share prices that we can observe in the market to derive implied required rates of return on equity, given our assumptions about their future cash flows.

For example, if we assume a constant rate of dividend growth, we can rearrange the constant growth DDM to solve for the required rate of return:

$$V_0 = \frac{D_1}{k_e - g_c}$$

$$k_e - g_c = \frac{D_1}{V_0}$$

$$k_e = \frac{D_1}{V_0} + g_c$$

That is, the required rate of return on equity is the ratio of the expected dividend to the current price (which we refer to as a share's **dividend yield**) plus the assumed constant growth rate.

We can also rearrange the model to solve for a stock's **implied growth rate**, given a required rate of return:

$$k_e = \frac{D_1}{V_0} + g_c$$

$$g_c = k_e - \frac{D_1}{V_0}$$

That is, the implied growth rate is the required rate of return minus the dividend yield.

LOS 2.c: Explain the cash flow additivity principle, its importance for the no-arbitrage condition, and its use in calculating implied forward interest rates,

forward exchange rates, and option values.

The **cash flow additivity principle** refers to the fact that the PV of any stream of cash flows equals the sum of the PVs of the cash flows. If we have two series of cash flows, the sum of the PVs of the two series is the same as the PVs of the two series taken together, adding cash flows that will be paid at the same point in time. We can also divide up a series of cash flows any way we like, and the PV of the "pieces" will equal the PV of the original series.

EXAMPLE: Cash flow additivity principle

A security will make the following payments at the end of the next four years: $100, $100, $400, and $100. Calculate the PV of these cash flows using the concept of the PV of an annuity when the appropriate discount rate is 10%.

Answer:

We can divide the cash flows so that we have:

$t = 1$	$t = 2$	$t = 3$	$t = 4$	
100	100	100	100	Cash flow series #1
0	0	300	0	Cash flow series #2
$100	$100	$400	$100	

The additivity principle tells us that to get the PV of the original series, we can just add the PVs of cash flow series #1 (a 4-period annuity) and cash flow series #2 (a single payment three periods from now).

For the annuity: N = 4; PMT = 100; FV = 0; I/Y = 10; CPT →
PV = −$316.99

For the single payment: N = 3; PMT = 0; FV = 300; I/Y = 10; CPT →
PV = −$225.39

The sum of these two values is 316.99 + 225.39 = $542.38.

The sum of these two (present) values is identical (except for rounding) to the sum of the present values of the payments of the original series:

$$\frac{100}{1.1} + \frac{100}{1.1^2} + \frac{400}{1.1^3} + \frac{100}{1.1^4} = \$542.38$$

This is a simple example of **replication**. In effect, we created the equivalent of the given series of uneven cash flows by combining a 4-year annuity of 100 with a 3-year zero-coupon bond of 300.

We rely on the cash flow additivity principle in many of the pricing models we see in the Level I CFA curriculum. It is the basis for the **no-arbitrage principle**, or "law of one price," which says that if two sets of future cash flows are identical under all conditions, they will have the same price today (or if they don't, investors will quickly buy the lower-priced one and sell the higher-priced one, which will drive their prices together).

Three examples of valuation based on the no-arbitrage condition are forward interest rates, forward exchange rates, and option pricing using a binomial model. We will explain each of these examples in greater detail when we address the related concepts in the Fixed Income, Economics, and Derivatives topic areas. For now, just focus on how they apply the principle that equivalent future cash flows must have the same present value.

Forward Interest Rates

A *forward interest rate* is the interest rate for a loan to be made at some future date. The notation used must identify both the length of the loan and when in the future the money will be borrowed. Thus, *1y1y* is the rate for a 1-year loan to be made one year from now; *2y1y* is the rate for a 1-year loan to be made two years from now; *3y2y* is the 2-year forward rate three years from now; and so on.

By contrast, a *spot interest rate* is an interest rate for a loan to be made today. We will use the notation S_1 for a 1-year rate today, S_2 for a 2-year rate today, and so on.

The way the cash flow additivity principle applies here is that, for example, borrowing for three years at the 3-year spot rate, or borrowing for one-year periods in three successive years, should have the same cost today. This relation is illustrated as follows: $(1 + S_3)^3 = (1 + S_1)(1 + 1y1y)(1 + 2y1y)$.

In fact, any combination of spot and forward interest rates that cover the same time period should have the same cost. Using this idea, we can derive **implied forward rates** from spot rates that are observable in the fixed-income markets.

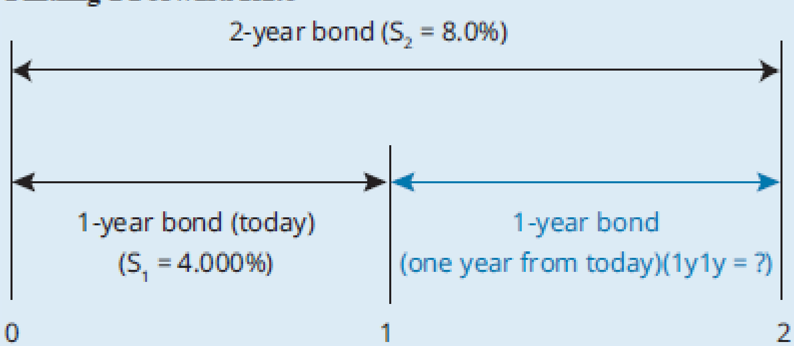
EXAMPLE: Computing a forward rate from spot rates

The 2-period spot rate, S_2, is 8%, and the 1-period spot rate, S_1, is 4%. Calculate the forward rate for one period, one period from now, *1y1y*.

Answer:

The following figure illustrates the problem.

Finding a Forward Rate



From our original equality, $(1 + S_2)^2 = (1 + S_1)(1 + 1y1y)$, we can get the following:

$$\frac{(1+S_2)^2}{(1+S_1)} = (1 + 1y1y)$$

Or, because we know that both choices have the same payoff in two years:

$$(1.08)^2 = (1.04)(1 + 1y1y)$$

$$(1 + 1y1y) = \frac{(1.08)^2}{(1.04)}$$

$$1y1y = \frac{(1.08)^2}{(1.04)} - 1 = \frac{1.1664}{1.04} - 1 = 12.154\%$$

In other words, investors are willing to accept 4.0% on the 1-year bond today (when they could get 8.0% on the 2-year bond today) only because they can get 12.154% on a 1-year bond one year from today. This future rate that can be locked in today is a forward rate.

Forward Currency Exchange Rates

An *exchange rate* is the price of one country's currency in terms of another country's currency. For example, an exchange rate of 1.416 USD/EUR means that one euro (EUR) is worth 1.416 U.S. dollars (USD). The Level I CFA curriculum refers to the currency in the numerator (USD, in this example) as the *price currency* and the one in the denominator (EUR in this example) as the *base currency*.

Like interest rates, exchange rates can be quoted as spot rates for currency exchanges to be made today, or as forward rates for currency exchanges to be made at a future date.

The percentage difference between forward and spot exchange rates is approximately the difference between the two countries' interest rates. This is because there is an arbitrage trade with a riskless profit to be made when this relation does not hold.

The possible arbitrage is as follows: borrow Currency A at Interest Rate A, convert it to Currency B at the spot rate and invest it to earn Interest Rate B, and sell the proceeds from this investment forward at the forward rate to turn it back into Currency A. If the forward rate does not correctly reflect the difference between interest rates, such an arbitrage could generate a profit to the extent that the return from investing Currency B and converting it back to Currency A with a forward contract is greater than the cost of borrowing Currency A for the period.

For spot and forward rates expressed as price currency/base currency, the no-arbitrage relation is as follows:

$$\frac{\text{forward}}{\text{spot}} = \frac{(1 + \text{interest rate}_{\text{price currency}})}{(1 + \text{interest rate}_{\text{base currency}})}$$

This formula can be rearranged as necessary to solve for specific values of the relevant terms.

EXAMPLE: Calculating the arbitrage-free forward exchange rate

Consider two currencies, the ABE and the DUB. The spot ABE/DUB exchange rate is 4.5671, the 1-year riskless ABE rate is 5%, and the 1-year riskless DUB rate is 3%. What is the 1-year forward exchange rate that will prevent arbitrage profits?

Answer:

Rearranging our formula, we have:

$$\text{forward} = \text{spot}\left(\frac{1 + \text{I}_{\text{ABE}}}{1 + \text{I}_{\text{DUB}}}\right)$$

and we can calculate the forward rate as:

$$\text{forward} = 4.5671\left(\frac{1.05}{1.03}\right) = 4.6558 \text{ ABE / DUB}$$

As you can see, the forward rate is greater than the spot rate by 4.6558 / 4.5671 − 1 = 1.94%. This is approximately equal to the interest rate differential of 5% − 3% = 2%.

Option Pricing Model

An *option* is the right, but not the obligation, to buy or sell an asset on a future date for a specified price. The right to buy an asset is a *call option*, and the right to sell an asset is a *put option*.

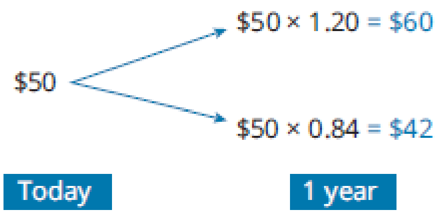
Valuing options is different from valuing other securities because the owner can let an option expire unexercised. A call option owner will let the option expire if the underlying asset can be bought in the market for less than the price specified in the option. A put option owner will let the option expire if the underlying asset can be sold in the market for more than the price specified in the option. In these cases, we say an option is *out of the money*. If an option is *in the money* on its expiration date, the owner has the right to buy the asset for less, or sell the asset for more, than its market price—and, therefore, will exercise the option.

An approach to valuing options that we will use in the Derivatives topic area is a **binomial model**. A binomial model is based on the idea that, over the next period, some value will change to one of two possible values. To construct a one-period binomial model for pricing an option, we need the following:

- A value for the underlying asset at the beginning of the period
- An exercise price for the option; the exercise price can be different from the value of the underlying, and we assume the option expires one period from now
- Returns that will result from an up-move and a down-move in the value of the underlying over one period
- The risk-free rate over the period

As an example, we can model a call option with an exercise price of $55 on a stock that is currently valued (S_0) at $50. Let us assume that in one period, the stock's value will either increase (S 1 u) to \$60 or decrease (S 1 d) to \$42. We state the return from an up-move (R^u) as \$60 / \$50 = 1.20, and the return from a down-move (R^d) as \$42 / \$50 = 0.84.

Figure 2.1: One-Period Binomial Tree



The call option will be in the money after an up-move or out of the money after a down-move. Its value at expiration after an up-move, c_1^u, is \$60 − \$55 = \$5. Its value after a down-move, c_1^d, is zero.

Now, we can use no-arbitrage pricing to determine the initial value of the call option (c_0). We do this by creating a portfolio of the option and the underlying stock, such that the portfolio will have the same value following either an up-move (V_1^u) or a down-move (V_1^d) in the stock. For our example, we would write the call option (that is, we grant someone else the option to buy the stock from us) and buy a number of shares of the stock that we will denote as *h*. We must solve for the *h* that results in $V_1^u = V_1^d$:

- The initial value of our portfolio, V_0, is $hS_0 - c_0$ (we subtract c_0 because we are short the call option).
- The portfolio value after an up-move, V_1^u, is $hS_1^u - c_1^u$.
- The portfolio value after a down-move, V_1^d, is $hS_1^d - c_1^d$.

In our example, $V_1^u = h(\$60) - \$5$, and $V_1^d = h(\$42) - 0$. Setting $V_1^u = V_1^d$ and solving for *h*, we get the following:

$$h(\$60) - \$5 = h(\$42)$$

$$h(\$60) - h(\$42) = \$5$$

$$h = \$5 \ / \ (\$60 - \$42) = 0.278$$

This result—the number of shares of the underlying we would buy for each call option we would write—is known as the hedge ratio for this option.

With $V_1^u = V_1^d$, the value of the portfolio after one period is known with certainty. This means we can say that either V_1^u or V_1^d must equal V_0 compounded at the risk-free rate for one period. In this example, V_1^d = 0.278(\$42) = \$11.68, or V_1^u = 0.278(\$60) − \$5 = \$11.68. Let us assume the risk-free rate over one period is 3%. Then, V_0 = \$11.68 / 1.03 = \$11.34.

Now, we can solve for the value of the call option, c_0. Recall that $V_0 = hS_0 - c_0$, so $c_0 = hS_0 - V_0$. Here, c_0 = 0.278(\$50) − \$11.34 = \$2.56.

MODULE QUIZ 2.2

1. For an equity share with a constant growth rate of dividends, we can estimate its:

A. value as the next dividend discounted at the required rate of return.
B. growth rate as the sum of its required rate of return and its dividend yield.
C. required return as the sum of its constant growth rate and its dividend yield.

2. An investment of €5 million today is expected to produce a one-time payoff of €7 million three years from today. The annual return on this investment, assuming annual compounding, is *closest* to:
   A. 12%.
   B. 13%.
   C. 14%.

KEY CONCEPTS

LOS 2.a

The value of a fixed-income instrument or an equity security is the present value of its future cash flows, discounted at the investor's required rate of return:

$$PV = \frac{FV}{(1+r)^t} = FV(1+r)^{-t}$$

where:
r = interest rate per compounding period
t = number of compounding periods

$$\text{annuity payment} = \frac{r \times PV}{1-(1+r)^{-t}}$$

where:
r = interest rate per period
t = number of periods
PV = present value (principal)

The PV of a perpetual bond or a preferred stock $= \frac{\text{payment}}{r}$, where r = required rate of return.

The PV of a common stock with a constant growth rate of dividends is:

$$V_0 = \frac{D_1}{k_e - g_c}$$

LOS 2.b

By rearranging the present value relationship, we can calculate a security's required rate of return based on its price and its future cash flows. The relationship between prices and required rates of return is inverse.

For an equity share with a constant rate of dividend growth, we can estimate the required rate of return as the dividend yield plus the assumed constant growth rate, or we can estimate the implied growth rate as the required rate of return minus the dividend yield.

LOS 2.c

Using the cash flow additivity principle, we can divide up a series of cash flows any way we like, and the present value of the pieces will equal the present value of the original series. This principle is the basis for the no-arbitrage condition, under which two sets of future cash flows that are identical must have the same present value.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 2.1

1. **A** 9 / 0.11 = $81.82 (LOS 2.a)

2. **A** Because the required yield is greater than the coupon rate, the present value of the bonds is less than their face value: N = 10; I/Y = 6; PMT = 0.05 × $10,000,00 = $500,000; FV = $10,000,000; and CPT PV = −$9,263,991. (LOS 2.a)

Module Quiz 2.2

1. **C** Using the constant growth dividend discount model, we can estimate the required rate of return as $k_e = \frac{D_1}{V_0} + g_c$. The estimated value of a share is *all* of its future dividends discounted at the required rate of return, which simplifies to $V_0 = \frac{D_1}{k_e - g_c}$ if we assume a constant growth rate. We can estimate the constant growth rate as the required rate of return *minus* the dividend yield. (LOS 2.b)

2. **A** $\left(\frac{7}{5}\right)^{\frac{1}{3}} - 1 = 0.1187$

   (LOS 2.b)

READING 3

STATISTICAL MEASURES OF ASSET RETURNS

MODULE 3.1: CENTRAL TENDENCY AND DISPERSION

Video covering this content is available online.

LOS 3.a: Calculate, interpret, and evaluate measures of central tendency and location to address an investment problem.

Measures of Central Tendency

Measures of central tendency identify the center, or average, of a dataset. This central point can then be used to represent the typical, or expected, value in the dataset.

The **arithmetic mean** is the sum of the observation values divided by the number of observations. It is the most widely used measure of central tendency. An example of an arithmetic mean is a **sample mean**, which is the sum of all the values in a sample of a population, ΣX, divided by the number of observations in the sample, *n*. It is used to make *inferences* about the population mean. The sample mean is expressed as follows:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

The **median** is the midpoint of a dataset, where the data are arranged in ascending or descending order. Half of the observations lie above the median, and half are below. To determine the median, arrange the data from the highest to lowest value, or lowest to highest value, and find the middle observation.

The median is important because the arithmetic mean can be affected by **outliers**, which are extremely large or small values. When this occurs, the median is a better measure of central tendency than the mean because it is not affected by extreme values that may actually be the result of errors in the data.

EXAMPLE: **The median using an odd number of observations**

What is the median return for five portfolio managers with a 10-year annualized total returns record of 30%, 15%, 25%, 21%, and 23%?

Answer:

First, arrange the returns in descending order:

30%, 25%, 23%, 21%, 15%

Then, select the observation that has an equal number of observations above and below it—the one in the middle. For the given dataset, the third observation, 23%, is the median value.

EXAMPLE: The median using an even number of observations

Suppose we add a sixth manager to the previous example with a return of 28%. What is the median return?

Answer:

Arranging the returns in descending order gives us this:

30%, 28%, 25%, 23%, 21%, 15%

With an even number of observations, there is no single middle value. The median value, in this case, is the arithmetic mean of the two middle observations, 25% and 23%. Thus, the median return for the six managers is 24% = 0.5(25 + 23).

The **mode** is the value that occurs most frequently in a dataset. A dataset may have more than one mode, or even no mode. When a distribution has one value that appears most frequently, it is said to be **unimodal**. When a dataset has two or three values that occur most frequently, it is said to be **bimodal** or **trimodal**, respectively.

EXAMPLE: The mode

What is the mode of the following dataset?

Dataset: [30%, 28%, 25%, 23%, 28%, 15%, 5%]

Answer:

The mode is 28% because it is the value appearing most frequently.

For continuous data, such as investment returns, we typically do not identify a single outcome as the mode. Instead, we divide the relevant range of outcomes into intervals, and we identify the **modal interval** as the one into which the largest number of observations fall.

Methods for Dealing With Outliers

In some cases, a researcher may decide that outliers should be excluded from a measure of central tendency. One technique for doing so is to use a **trimmed mean**. A trimmed mean excludes a stated percentage of the most extreme observations. A 1% trimmed mean, for example, would discard the lowest 0.5% and the highest 0.5% of the observations.

Another technique is to use a **winsorized mean**. Instead of discarding the highest and lowest observations, we substitute a value for them. To calculate a 90% winsorized mean, for example, we would determine the 5th and 95th percentile of the observations, substitute the 5th percentile for any values lower than that, substitute the 95th percentile for any values higher than that, and then calculate the mean of the revised dataset. Percentiles are measures of location, which we will address next.

Measures of Location

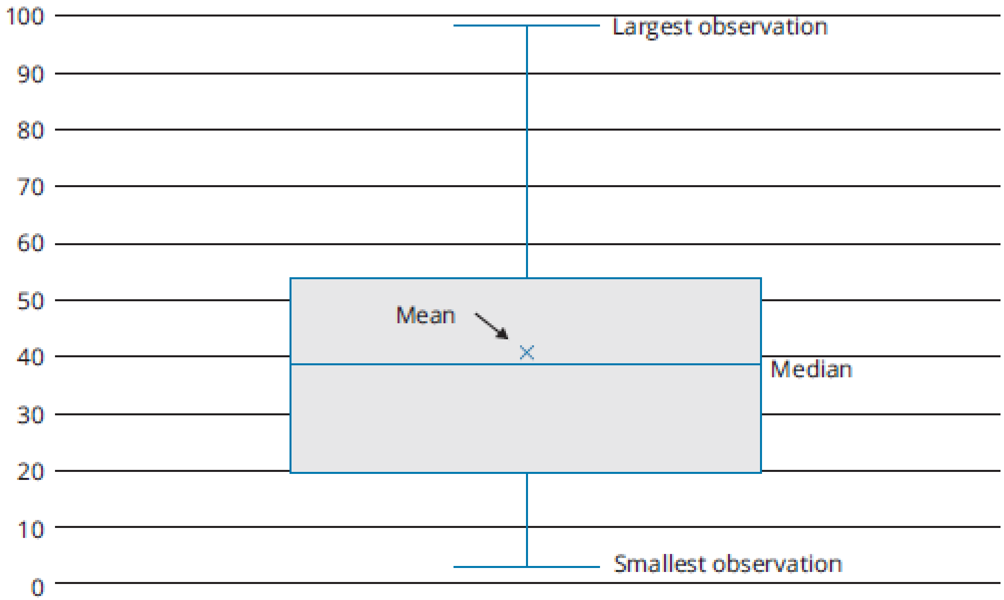
Quantile is the general term for a value at or below which a stated proportion of the data in a distribution lies. Examples of quantiles include the following:

- **Quartile.** The distribution is divided into quarters.
- **Quintile.** The distribution is divided into fifths.
- **Decile.** The distribution is divided into tenths.
- **Percentile.** The distribution is divided into hundredths (percentages).

Note that any quantile may be expressed as a percentile. For example, the third quartile partitions the distribution at a value such that three-fourths, or 75%, of the observations fall below that value. Thus, the third quartile is the 75th percentile. The difference between the third quartile and the first quartile (25th percentile) is known as the **interquartile range**.

To visualize a dataset based on quantiles, we can create a **box and whisker plot**, as shown in Figure 3.1. In a box and whisker plot, the box represents the central portion of the data, such as the interquartile range. The vertical line represents the entire range. In Figure 3.1, we can see that the largest observation is farther away from the center than is the smallest observation. This suggests that the data might include one or more outliers on the high side.

Figure 3.1: Box and Whisker Plot



LOS 3.b: Calculate, interpret, and evaluate measures of dispersion to address an investment problem.

Dispersion is defined as the *variability around the central tendency*. The common theme in finance and investments is the tradeoff between reward and variability, where the central tendency is the measure of the reward and dispersion is a measure of risk.

The **range** is a relatively simple measure of variability, but when used with other measures, it provides useful information. The range is the distance between the largest and the smallest value in the dataset:

range = maximum value − minimum value

> EXAMPLE: **The range**
>
> What is the range for the 5-year annualized total returns for five investment managers if the managers' individual returns were 30%, 12%, 25%, 20%, and 23%?
>
> **Answer:**
>
> range = 30 − 12 = 18%

The **mean absolute deviation (MAD)** is the average of the absolute values of the deviations of individual observations from the arithmetic mean:

$$\text{MAD} = \frac{\sum_{i=1}^{n} \left| X_i - \overline{X} \right|}{n}$$

The computation of the MAD uses the absolute values of each deviation from the mean because the sum of the actual deviations from the arithmetic mean is zero.

EXAMPLE: MAD

What is the MAD of the investment returns for the five managers discussed in the preceding example? How is it interpreted?

Answer:

Annualized returns: [30%, 12%, 25%, 20%, 23%]

$$\overline{X} = \frac{[30 + 12 + 25 + 20 + 23]}{5} = 22\%$$

$$MAD = \frac{[|30 - 22| + |12 - 22| + |25 - 22| + |20 - 22| + |23 - 22|]}{5}$$

$$MAD = \frac{[8 + 10 + 3 + 2 + 1]}{5} = 4.8\%$$

This result can be interpreted to mean that, on average, an individual return will deviate ±4.8% from the mean return of 22%.

The **sample variance**, s^2, is the measure of dispersion that applies when we are evaluating a sample of n observations from a population. The sample variance is calculated using the following formula:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1}$$

The denominator for s^2 is n − 1, one less than the sample size n. Based on the mathematical theory behind statistical procedures, the use of the entire number of sample observations, n, instead of n − 1 as the divisor in the computation of s^2, will systematically *underestimate* the population variance—particularly for small sample sizes. This systematic underestimation causes the sample variance to be a **biased estimator** of the population variance. Using n − 1 instead of n in the denominator, however, improves the statistical properties of s^2 as an estimator of the population variance.

EXAMPLE: Sample variance

Assume that the 5-year annualized total returns for the five investment managers used in the preceding examples represent only a sample of the managers at a large investment firm. What is the sample variance of these returns?

Answer:

$$\overline{X} = \frac{[30 + 12 + 25 + 20 + 23]}{5} = 22\%$$

$$s^2 = \frac{[(30 - 22)^2 + (12 - 22)^2 + (25 - 22)^2 + (20 - 22)^2 + (23 - 22)^2]}{5 - 1}$$

$$= 44.5(\%^2)$$

Thus, the sample variance of 44.5($\%^2$) can be interpreted to be an unbiased estimator of the population variance. Note that 44.5 "percent squared" is 0.00445, and you will get this value if you put the percentage returns in decimal form [e.g., $(0.30 - 0.22)^2$].

A major problem with using variance is the difficulty of interpreting it. The computed variance, unlike the mean, is in terms of squared units of measurement. How does one interpret squared percentages, squared dollars, or squared yen? This problem is mitigated through the use of the *standard deviation*. The units of standard deviation are the same as the units of the data (e.g., percentage return, dollars, euros). The **sample standard deviation** is the square root of the sample variance. The sample standard deviation, *s*, is calculated as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

EXAMPLE: Sample standard deviation

Compute the sample standard deviation based on the result of the preceding example.

Answer:

Because the sample variance for the preceding example was computed to be 44.5($\%^2$), this is the sample standard deviation:

$$s = [44.5(\%^2)]^{1/2} = 6.67\%, \ \sqrt{0.00445} = 0.0667$$

This means that on average, an individual return from the sample will deviate ±6.67% from the mean return of 22%. The sample standard deviation can be interpreted as an unbiased estimator of the population standard deviation.

A direct comparison between two or more measures of dispersion may be difficult. For instance, suppose you are comparing the annual returns distribution for retail stocks with a mean of 8% and an annual returns distribution for a real estate portfolio with a mean of 16%. A direct comparison between the dispersion of the two distributions is not meaningful because of the relatively large difference in their means. To make a meaningful comparison, a relative measure of dispersion must be used. **Relative dispersion** is the amount of variability in a distribution around a reference point or benchmark. Relative dispersion is commonly measured with the **coefficient of variation (CV)**, which is computed as follows:

$$CV = \frac{s_x}{\overline{X}} = \frac{\text{standard deviation of x}}{\text{average value of x}}$$

CV measures the amount of dispersion in a distribution relative to the distribution's mean. This is useful because it enables us to compare dispersion across different sets of data. In an investments setting, the CV is used to measure the risk (variability) per unit of expected return (mean). A lower CV is better.

EXAMPLE: Coefficient of variation

You have just been presented with a report that indicates that the mean monthly return on T-bills is 0.25% with a standard deviation of 0.36%, and the mean monthly return for the S&P 500 is 1.09% with a standard deviation of 7.30%. Your manager has asked you to compute the CV for these two investments and to interpret your results.

Answer:

$$CV_{\text{T-bills}} = \frac{0.36}{0.25} = 1.44$$

$$CV_{\text{S\&P 500}} = \frac{7.30}{1.09} = 6.70$$

These results indicate that there is less dispersion (risk) per unit of monthly return for T-bills than for the S&P 500 (1.44 vs. 6.70).

PROFESSOR'S NOTE

To remember the formula for CV, remember that the CV is a measure of variation, so standard deviation goes in the numerator. CV is variation per unit of return.

When we use variance or standard deviation as risk measures, we calculate risk based on outcomes both above and below the mean. In some situations, it may be more appropriate to consider only outcomes less than the mean (or some other specific value) in calculating a risk measure. In this case, we are measuring **downside risk**.

One measure of downside risk is **target downside deviation**, which is also known as **target semideviation**. Calculating target downside deviation is similar to calculating standard deviation, but in this case, we choose a target value against which to measure each outcome and only include deviations from the target value in our calculation if the outcomes are below that target.

The formula for target downside deviation is stated as follows:

$$s_{\text{target}} = \sqrt{\frac{\sum_{\text{all } X_i < B}^{n} (X_i - B)^2}{n - 1}}$$

where B = the target

Note that the denominator remains the sample size *n* minus one, even though we are not using all of the observations in the numerator.

EXAMPLE: Target downside deviation

Calculate the target downside deviation based on the data in the preceding examples, for a target return equal to the mean (22%), and for a target return of 24%.

Answer:

Return	Deviation From Mean	Deviation From Target Return
30%	30% − 22% = 8%	30% − 24% = 6%
12%	12% − 22% = −10%	12% − 24% = −12%
25%	25% − 22% = 3%	25% − 24% = 1%
20%	20% − 22% = −2%	20% − 24% = −4%
23%	23% − 22% = 1%	23% − 24% = −1%

$$s_{22\%} = \sqrt{\frac{(-10)^2 + (-2)^2}{5-1}} = 5.10\%$$

$$s_{24\%} = \sqrt{\frac{(-12)^2 + (-4)^2 + (-1)^2}{5-1}} = 6.34\%$$

MODULE QUIZ 3.1

1. A dataset has 100 observations. Which of the following measures of central tendency will be calculated using a denominator of 100?
   A. The winsorized mean, but not the trimmed mean.
   B. Both the trimmed mean and the winsorized mean.
   C. Neither the trimmed mean nor the winsorized mean.
2. **XYZ Corp. Annual Stock Returns**

20X1	20X2	20X3	20X4	20X5	20X6
22%	5%	−7%	11%	2%	11%

   What is the sample standard deviation?
   A. 9.8%.
   B. 72.4%.
   C. 96.3%.
3. **XYZ Corp. Annual Stock Returns**

20X1	20X2	20X3	20X4	20X5	20X6
22%	5%	−7%	11%	2%	11%

   Assume an investor has a target return of 11% for XYZ stock. What is the stock's target downside deviation?
   A. 9.39%.
   B. 12.10%.
   C. 14.80%.

MODULE 3.2: SKEWNESS, KURTOSIS, AND CORRELATION

Video covering this content is available online.

LOS 3.c: Interpret and evaluate measures of skewness and kurtosis to address an investment problem.

A distribution is symmetrical if it is shaped identically on both sides of its mean. Distributional symmetry implies that intervals of losses and gains will exhibit the same frequency. For example, a symmetrical distribution with a mean return of zero will have losses in the −6% to −4% interval as frequently as it will have gains in the +4% to +6% interval. The extent to which a returns distribution is symmetrical is important because the degree of symmetry tells analysts if deviations from the mean are more likely to be positive or negative.

Skewness, or skew, refers to the extent to which a distribution is not symmetrical. Nonsymmetrical distributions may be either positively or negatively skewed and result from the occurrence of outliers in the dataset. **Outliers** are observations extraordinarily far from the mean, either above or below:

- A *positively skewed* distribution is characterized by outliers greater than the mean (in the upper region, or right tail). A positively skewed distribution is said to be skewed right because of its relatively long upper (right) tail.
- A *negatively skewed* distribution has a disproportionately large amount of outliers less than the mean that fall within its lower (left) tail. A negatively skewed distribution is said to be skewed left because of its long lower tail.

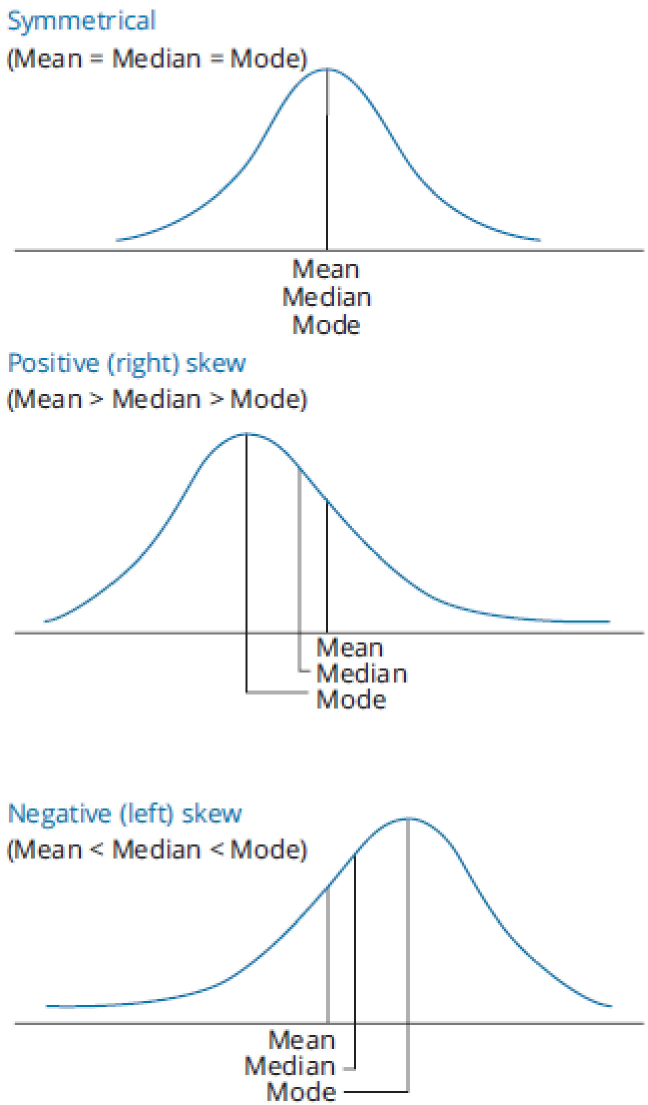
Skewness affects the location of the mean, median, and mode of a distribution:

- For a symmetrical distribution, the mean, median, and mode are equal.
- For a positively skewed, unimodal distribution, the mode is less than the median, which is less than the mean. The mean is affected by outliers; in a positively skewed distribution, there are large, positive outliers, which will tend to pull the mean upward, or more positive. An example of a positively skewed distribution is that of housing prices. Suppose you live in a neighborhood with 100 homes; 99 of them sell for $100,000, and one sells for $1,000,000. The median and the mode will be $100,000, but the mean will be $109,000. Hence, the mean has been pulled upward (to the right) by the existence of one home (outlier) in the neighborhood.
- For a negatively skewed, unimodal distribution, the mean is less than the median, which is less than the mode. In this case, there are large, negative outliers that tend to pull the mean downward (to the left).

PROFESSOR'S NOTE

The key to remembering how measures of central tendency are affected by skewed data is to recognize that skew affects the mean more than the median and mode, and the mean is pulled in the direction of the skew. The relative location of the mean, median, and mode for different distribution shapes is shown in Figure 3.2. Note that the median is between the other two measures for positively or negatively skewed distributions.

Figure 3.2: Effect of Skewness on Mean, Median, and Mode



Sample skewness is equal to the sum of the cubed deviations from the mean divided by the cubed standard deviation and by the number of observations. Sample skewness for large samples is approximated as follows:

$$\text{sample skewness} \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{s^3}$$

where:

s = sample standard deviation

PROFESSOR'S NOTE

The LOS requires us to "interpret and evaluate" measures of skewness and kurtosis, but not to calculate them.

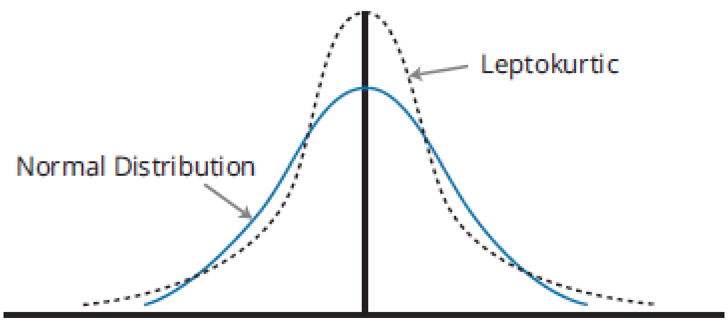
Note that the denominator is always positive, but that the numerator can be positive or negative depending on whether observations above the mean or observations below the mean tend to be farther from the mean, on average. When a distribution is right skewed, sample skewness is positive because the deviations above the mean are larger, on average. A left-skewed distribution has a negative sample skewness.

Dividing by standard deviation cubed standardizes the statistic and allows interpretation of the skewness measure. If relative skewness is equal to zero, the data are not skewed. Positive levels of relative skewness imply a positively skewed distribution, whereas negative values of relative skewness imply a negatively skewed distribution. Values of sample skewness in excess of 0.5 in absolute value are considered significant.

Kurtosis is a measure of the degree to which a distribution is more or less peaked than a normal distribution. **Leptokurtic** describes a distribution that is more peaked than a normal distribution, whereas **platykurtic** refers to a distribution that is less peaked, or flatter than a normal one. A distribution is **mesokurtic** if it has the same kurtosis as a normal distribution.

As indicated in Figure 3.3, a leptokurtic return distribution will have more returns clustered around the mean and more returns with large deviations from the mean (fatter tails). Relative to a normal distribution, a leptokurtic distribution will have a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean. This means that there is a relatively greater probability of an observed value being either close to the mean or far from the mean. Regarding an investment returns distribution, a greater likelihood of a large deviation from the mean return is often perceived as an increase in risk.

Figure 3.3: Kurtosis



A distribution is said to exhibit **excess kurtosis** if it has either more or less kurtosis than the normal distribution. The computed kurtosis for all normal distributions is three. Statisticians, however, sometimes report excess kurtosis, which is defined as kurtosis minus three. Thus, a normal distribution has excess kurtosis equal to zero, a leptokurtic distribution has excess kurtosis greater than zero, and platykurtic distributions will have excess kurtosis less than zero.

Kurtosis is critical in a risk management setting. Most research about the distribution of securities returns has shown that returns are not normally distributed. Actual securities returns tend to exhibit both skewness and kurtosis. Skewness and kurtosis are critical concepts for risk management because when securities returns are modeled using an assumed normal distribution, the predictions from the models will not take into account the potential for extremely large, negative outcomes. In fact, most risk managers put very little emphasis on the mean and standard deviation of a distribution and focus more on the distribution of returns in the tails of the distribution—that is where the risk is. In general, greater excess kurtosis and more negative skew in returns distributions indicate increased risk.

Sample kurtosis for large samples is approximated using deviations raised to the *fourth power*:

$$\text{sample kurtosis} \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4}$$

where:

s = sample standard deviation

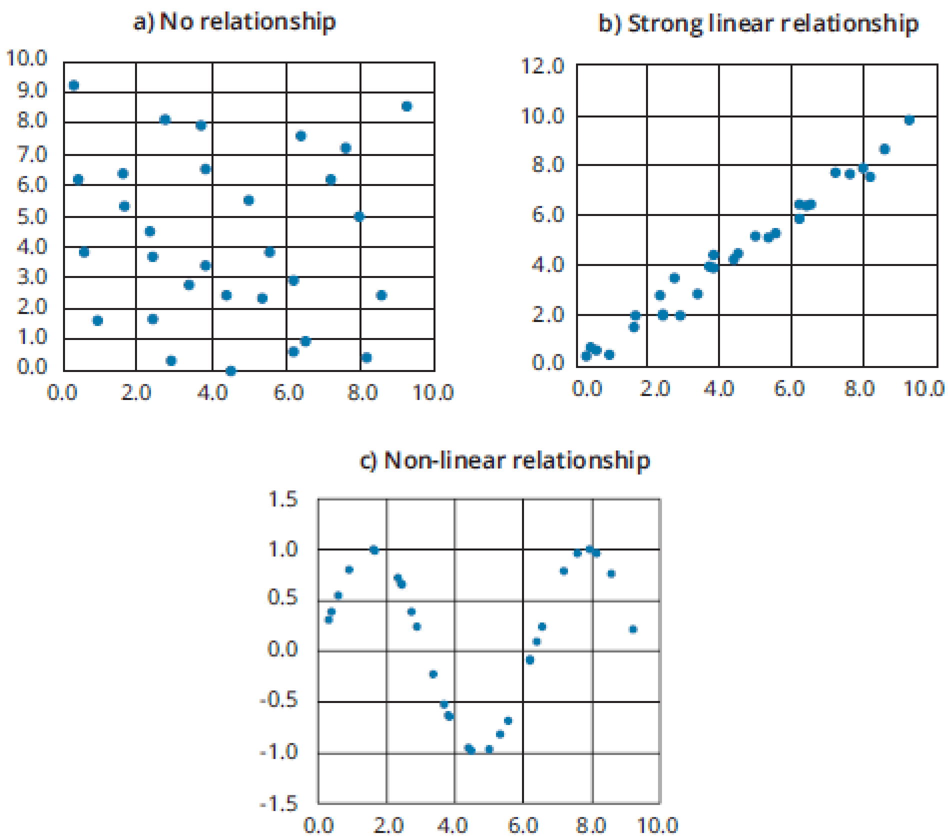
LOS 3.d: Interpret correlation between two variables to address an investment problem.

Scatter plots are a method for displaying the relationship between two variables. With one variable on the vertical axis and the other on the horizontal axis, their paired observations can each be plotted as a single point. For example, in Panel A of Figure 3.4, the point farthest to the upper right shows that when one of the variables (on the horizontal axis) equaled 9.2, the other variable (on the vertical axis) equaled 8.5.

The scatter plot in Panel A is typical of two variables that have no clear relationship. Panel B shows two variables that have a strong linear relationship—that is, a high correlation coefficient.

A key advantage of creating scatter plots is that they can reveal *nonlinear* relationships, which are not described by the correlation coefficient. Panel C illustrates such a relationship. Although the correlation coefficient for these two variables is close to zero, their scatter plot shows clearly that they are related in a predictable way.

Figure 3.4: Scatter Plots



Covariance is a measure of how two variables move together. The calculation of the **sample covariance** is based on the following formula:

$$s_{XY} = \frac{\sum_{i=1}^{n}\left\{\left[X_i - \overline{X}\right]\left[Y_i - \overline{Y}\right]\right\}}{n-1}$$

where:

X_i = an observation of variable X
Y_i = an observation of variable Y
$\overline{X}$ = mean of variable X
$\overline{Y}$ = mean of variable Y
n = number of periods

In practice, the covariance is difficult to interpret. The value of covariance depends on the units of the variables. The covariance of daily price changes of two securities priced in yen will be much greater than their covariance if the securities are priced in dollars. Like the variance, the units of covariance are the square of the units used for the data.

Additionally, we cannot interpret the relative strength of the relationship between two variables. Knowing that the covariance of *X* and *Y* is 0.8756 tells us only that they tend to move together because the covariance is positive. A standardized measure of the linear relationship between two variables is called the **correlation coefficient**, or simply *correlation*. The correlation between two variables, *X* and *Y*, is calculated as follows:

$$\rho_{XY} = \frac{s_{XY}}{s_X s_Y}, \text{ which implies:}$$

$$s_{XY} = \rho_{XY} s_X s_Y$$

The *properties of the correlation* of two random variables, *X* and *Y*, are summarized here:

- Correlation measures the strength of the linear relationship between two random variables.
- Correlation has no units.
- The correlation ranges from −1 to +1. That is, $-1 \leq \rho_{XY} \leq +1$.
- If $\rho_{XY} = 1.0$, the random variables have perfect positive correlation. This means that a movement in one random variable results in a proportional positive movement in the other relative to its mean.
- If $\rho_{XY} = -1.0$, the random variables have perfect negative correlation. This means that a movement in one random variable results in an exact opposite proportional movement in the other relative to its mean.
- If $\rho_{XY} = 0$, there is no linear relationship between the variables, indicating that prediction of *Y* cannot be made on the basis of *X* using linear methods.

EXAMPLE: Correlation

The variance of returns on Stock A is 0.0028, the variance of returns on Stock B is 0.0124, and their covariance of returns is 0.0058. Calculate and interpret the correlation of the returns for Stocks A and B.

Answer:

First, it is necessary to convert the variances to standard deviations:

$$s_A = (0.0028)^{½} = 0.0529$$

$$s_B = (0.0124)^{½} = 0.1114$$

Now, the correlation between the returns of Stock A and Stock B can be computed as follows:

$$\rho_{AB} = \frac{0.0058}{(0.0529)(0.1114)} = 0.9842$$

The fact that this value is close to +1 indicates that the linear relationship is not only positive, but also is very strong.

Care should be taken when drawing conclusions based on correlation. Causation is not implied just from significant correlation. Even if it were, which variable is causing change in the other is not revealed by correlation. It is more prudent to say that two variables exhibit positive (or negative) association, suggesting that the nature of any causal relationship is to be separately investigated or based on theory that can be subject to additional tests.

One question that can be investigated is the role of outliers (extreme values) in the correlation of two variables. If removing the outliers significantly reduces the calculated correlation, further inquiry is necessary into whether the outliers provide information or are caused by noise (randomness) in the data used.

Spurious correlation refers to correlation that is either the result of chance or present due to changes in both variables over time that is caused by their association with a third variable. For example, we can find instances where two variables that are both related to the inflation rate exhibit significant correlation, but for which causation in either direction is not present.

In his book *Spurious Correlation*,[1] Tyler Vigen presents the following examples. The correlation between the age of each year's Miss America and the number of films Nicolas Cage appeared in that year is 87%. This seems a bit random. The correlation between the U.S. spending on science, space, and technology and suicides by hanging, strangulation, and suffocation over the 1999–2009 period is 99.87%. Impressive correlation, but both variables increased in an approximately linear fashion over the period.

MODULE QUIZ 3.2

1. Which of the following is *most accurate* regarding a distribution of returns that has a mean greater than its median?
   A. It is positively skewed.
   B. It is a symmetric distribution.
   C. It has positive excess kurtosis.
2. A distribution of returns that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean compared with a normal distribution:
   A. is positively skewed.
   B. has positive excess kurtosis.
   C. has negative excess kurtosis.
3. The correlation between two variables is +0.25. The *most appropriate* way to interpret this value is to say:
   A. a scatter plot of the two variables is likely to show a strong linear relationship.
   B. when one variable is above its mean, the other variable tends to be above its mean as well.
   C. a change in one of the variables usually causes the other variable to change in the same direction.

KEY CONCEPTS

LOS 3.a

The arithmetic mean is the average of observations. The sample mean is the arithmetic mean of a sample:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

The median is the midpoint of a dataset when the data are arranged from largest to smallest.

The mode of a dataset is the value that occurs most frequently. The modal interval is a measure of mode for continuous data.

A trimmed mean omits outliers, and a winsorized mean replaces outliers with given values, reducing the effect of outliers on the mean in both cases.

Quantile is the general term for a value at or below which lies a stated proportion of the data in a distribution. Examples of quantiles include the following:

- *Quartile*. The distribution is divided into quarters.
- *Quintile*. The distribution is divided into fifths.
- *Decile*. The distribution is divided into tenths.
- *Percentile*. The distribution is divided into hundredths (percentages).

LOS 3.b

The range is the difference between the largest and smallest values in a dataset.

Mean absolute deviation (MAD) is the average of the absolute values of the deviations from the arithmetic mean:

$$\text{MAD} = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$$

Variance is defined as the mean of the squared deviations from the arithmetic mean:

$$\text{sample variance} = s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}$$

where:

$\overline{X}$ = sample mean

n = sample size

Standard deviation is the positive square root of the variance, and it is frequently used as a quantitative measure of risk.

The coefficient of variation (CV) for sample data, $CV = \frac{s}{\overline{X}}$, is the ratio of the standard deviation of the sample to its mean.

Target downside deviation or semideviation is a measure of downside risk:

$$s_{target} = \sqrt{\frac{\sum_{\text{all } X_i < B}^{n} (X_i - B)^2}{n - 1}}$$

where B = the target

LOS 3.c

Skewness describes the degree to which a distribution is not symmetric about its mean. A right-skewed distribution has positive skewness. A left-skewed distribution has negative skewness.

For a positively skewed, unimodal distribution, the mean is greater than the median, which is greater than the mode. For a negatively skewed, unimodal distribution, the mean is less than the median, which is less than the mode.

Kurtosis measures the peakedness of a distribution and the probability of extreme outcomes (thickness of tails):

- Excess kurtosis is measured relative to a normal distribution, which has a kurtosis of 3.
- Positive values of excess kurtosis indicate a distribution that is leptokurtic (fat tails, more peaked), so the probability of extreme outcomes is greater than for a normal distribution.
- Negative values of excess kurtosis indicate a platykurtic distribution (thin tails, less peaked).

LOS 3.d

Correlation is a standardized measure of association between two random variables. It ranges in value from −1 to +1 and is equal to $\frac{Cov_{A,B}}{\sigma_A \sigma_B}$.

Scatter plots are useful for revealing nonlinear relationships that are not measured by correlation.

Correlation does not imply that changes in one variable cause changes in the other. Spurious correlation may result by chance, or from the relationships of two variables to a third variable.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 3.1

1. **A** The winsorized mean substitutes a value for some of the largest and smallest observations. The trimmed mean removes some of the largest and smallest observations. (LOS 3.a)

2. **A** The sample mean is [22% + 5% + −7% + 11% + 2% + 11%] / 6 = 7.3%. The sample standard deviation is the square root of the sample variance:

$$s = \sqrt{\frac{(22-7.3)^2 + (5-7.3)^2 + (-7-7.3)^2 + (11-7.3)^2 + (2-7.3)^2 + (11-7.3)^2}{6-1}}$$

$$= \sqrt{96.3} = 9.8\%.$$

(LOS 3.b)

3. **A** Here are deviations from the target return:

$22\% - 11\% = 11\%$

$5\% - 11\% = -6\%$

$-7\% - 11\% = -18\%$

$11\% - 11\% = 0\%$

$2\% - 11\% = -9\%$

$11\% - 11\% = 0\%$

Target downside deviation =

$$\sqrt{\frac{(-6)^2 + (-18)^2 + (-9)^2}{6-1}} = \sqrt{88.2} = 9.39\%.$$

(LOS 3.b)

Module Quiz 3.2

1. **A** A distribution with a mean greater than its median is positively skewed, or skewed to the right. The skew pulls the mean. Kurtosis deals with the overall shape of a distribution, not its skewness. (LOS 3.c)
2. **B** A distribution that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean will be leptokurtic and will exhibit excess kurtosis (positive). The distribution will be more peaked and have fatter tails than a normal distribution. (LOS 3.c)
3. **B** A correlation of +0.25 indicates a positive linear relationship between the variables—one tends to be above its mean when the other is above its mean. The value 0.25 indicates that the linear relationship is not particularly strong. Correlation does not imply causation. (LOS 3.d)

---

[1] "Spurious Correlations," Tyler Vigen, www.tylervigen.com

READING 4

PROBABILITY TREES AND CONDITIONAL EXPECTATIONS

MODULE 4.1: PROBABILITY MODELS, EXPECTED VALUES, AND BAYES' FORMULA

Video covering this content is available online.

LOS 4.a: Calculate expected values, variances, and standard deviations and demonstrate their application to investment problems.

The **expected value** of a random variable is the weighted average of the possible outcomes for the variable. The mathematical representation for the expected value of random variable *X*, that can take on any of the values from x_1 to x_n, is:

$$E(X) = \Sigma P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$

where:

$P(x_i)$ = probability of outcome x_i

EXAMPLE: **Expected earnings per share**

The probability distribution of earnings per share (EPS) for Ron's Stores is given in the following figure. Calculate the expected EPS.

EPS Probability Distribution

Probability	EPS
10%	£1.80
20%	£1.60
40%	£1.20
30%	£1.00
100%	

Answer:

The expected EPS is simply a weighted average of each possible EPS, where the weights are the probabilities of each possible outcome:

$$E[EPS] = 0.10(1.80) + 0.20(1.60) + 0.40(1.20) + 0.30(1.00) = £1.28$$

Variance and *standard deviation* measure the dispersion of a random variable around its expected value, sometimes referred to as the **volatility** of a random variable. **Variance** (from a probability model) can be calculated as the probability-weighted sum of the squared deviations from the mean (or expected value). The standard deviation is the positive square root of the variance. The following example illustrates the calculations for a probability model of possible returns.

EXAMPLE: Expected value, variance, and standard deviation from a probability model

Using the probabilities given in the following table, calculate the expected return on Stock A, the variance of returns on Stock A, and the standard deviation of returns on Stock A.

Event	Prob	R_A	Prob × R_A	$R_A - E(R_A)$	$[R_A - E(R_A)]^2$	Prob × $[R_A - E(R_A)]^2$
Boom	30%	20%	0.06	0.07	0.0049	0.00147
Normal	50%	12%	0.06	−0.01	0.0001	0.00005
Slow	20%	5%	0.01	−0.08	0.0064	0.00128
			$E(R_A) = 0.13$			$Var(R_A) = 0.00280$

Answer:

$$E(R_A) = (0.30 \times 0.20) + (0.50 \times 0.12) + (0.20 \times 0.05) = 0.13 = 13\%$$

The expected return for Stock A is the probability-weighted sum of the returns under the three different economic scenarios.

In Column 5, we have calculated the differences between the returns under each economic scenario and the expected return of 13%.

In Column 6, we squared all the differences from Column 5. In the final column, we have multiplied the probabilities of each economic scenario times the squared deviation of returns from the expected returns, and their sum, 0.00280, is the variance of R_A.

The standard deviation of $R_A = \sqrt{0.0028} = 0.0529$.

Note that in a previous reading, we estimated the standard deviation of a distribution from sample data, rather than from a probability model of returns. For the sample standard deviation, we divided the sum of the squared deviations from the mean by n − 1, where *n* was the size of the sample. Here, we have no "n" because we have no observations; a probability model is forward-looking. We use the probability weights instead, as they describe the entire distribution of outcomes.

LOS 4.b: Formulate an investment problem as a probability tree and explain the use of conditional expectations in investment application.

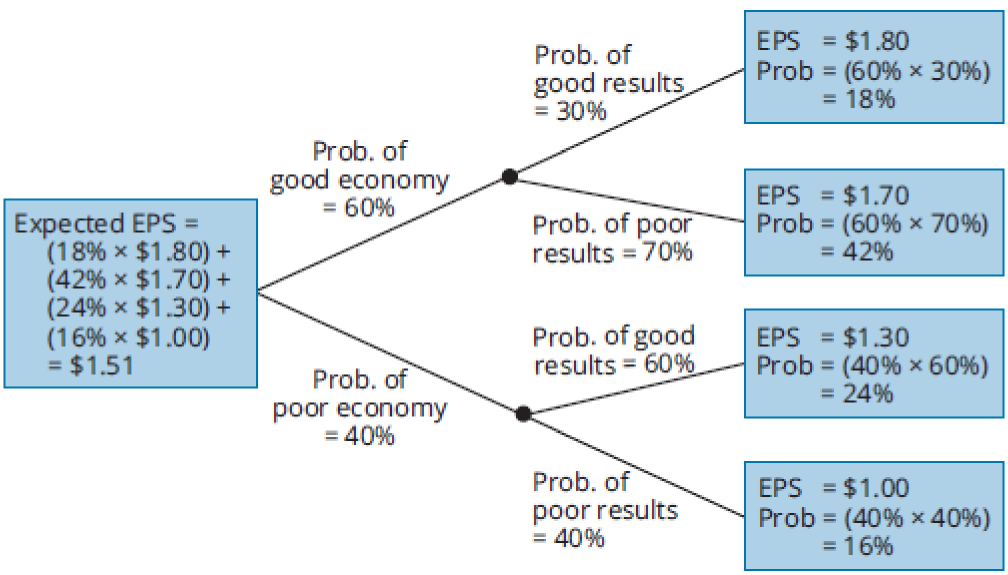
You may wonder where the returns and probabilities used in calculating expected values come from. A general framework, called a **probability tree**, is used to show the probabilities of various outcomes. In Figure 4.1, we have shown estimates of EPS for four different events: (1) a good economy and relatively good results at the company, (2) a good economy and relatively poor results at the company, (3) a poor economy and relatively good results at the company, and (4) a poor economy and relatively poor results at the company. Using the rules of probability, we can calculate the probabilities of each of the four EPS outcomes shown in the boxes on the right-hand side of the probability tree.

The expected EPS of \$1.51 is simply calculated as follows:

$$(0.18 \times 1.80) + (0.42 \times 1.70) + (0.24 \times 1.30) + (0.16 \times 1.00) = \$1.51$$

Note that the probabilities of the four possible outcomes sum to 1.

Figure 4.1: A Probability Tree



Expected values or expected returns can be calculated using conditional probabilities. As the name implies, **conditional expected values** are contingent on the outcome of some other event. An analyst would use a conditional expected value to revise his expectations when new information arrives.

Consider the effect a tariff on steel imports might have on the returns of a domestic steel producer's stock in the previous example. The stock's conditional expected return, given that the government imposes the tariff, will be higher than the conditional expected return if the tariff is not imposed.

LOS 4.c: Calculate and interpret an updated probability in an investment setting using Bayes' formula.

Bayes' formula is used to update a given set of prior probabilities for a given event in response to the arrival of new information. The rule for updating prior probability of an event is as follows:

$$\text{updated probability} = \frac{\text{probability of new information for a given event}}{\text{unconditional probability of new information}} \times \text{prior probability of event}$$

We can derive Bayes' formula using the multiplication rule and noting that P(AB) = P(BA):

$$P(B \mid A) \times P(A) = P(BA), \text{ and } P(A \mid B) \times P(B) = P(AB)$$

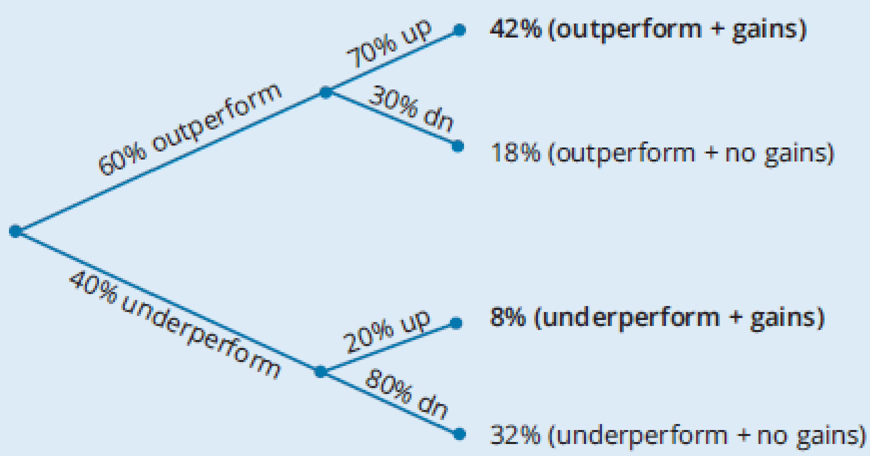
Because $P(BA) = P(AB)$, we can write $P(B \mid A)\,P(A) = P(A \mid B)\,P(B)$, and $\frac{P(B \mid A)P(A)}{P(B)}$ equals $\frac{P(BA)}{P(B)}$, the joint probability of A and B divided by the unconditional probability of B.

The following example illustrates the use of Bayes' formula. Note that A is *outperform* and A^C is *underperform*, P(BA) is (outperform + gains), $P(A^CB)$ is (underperform + gains), and the unconditional probability P(B) is P(AB) + $P(A^CB)$, by the total probability rule.

EXAMPLE: Bayes' formula

There is a 60% probability that the economy will outperform—and if it does, there is a 70% probability a stock will go up and a 30% probability the stock will go down. There is a 40% probability the economy will underperform, and if it does, there is a 20% probability the stock in question will increase in value (have gains) and an 80% probability it will not. Given that the stock increased in value, calculate the probability that the economy outperformed.

Answer:



In the figure just presented, we have multiplied the probabilities to calculate the probabilities of each of the four outcome pairs. Note that these sum to 1. Given that the stock has gains, what is our updated probability of an outperforming economy?

We sum the probability of stock gains in both states (outperform and underperform) to get 42% + 8% = 50%. Given that the stock has gains and using Bayes' formula, the probability that the economy has outperformed is $\frac{42\%}{50\%} = 84\%$.

MODULE QUIZ 4.1

1. Given the conditional probabilities in the following table and the unconditional probabilities P(Y = 1) = 0.3 and P(Y = 2) = 0.7, what is the expected value of *X*?

x_i	$P(x_i \mid Y = 1)$	$P(x_i \mid Y = 2)$
0	0.2	0.1
5	0.4	0.8
10	0.4	0.1

   A. 5.0.
   B. 5.3.
   C. 5.7.

2. An analyst believes that Davies Company has a 40% probability of earning more than $2 per share. She estimates that the probability that Davies Company's credit rating will be upgraded is 70% if its earnings per share (EPS) are greater than $2, and 20% if its EPS are $2 or less. Given the information that Davies Company's credit rating has been upgraded, what is the updated probability that its EPS are greater than $2?
   A. 50%.
   B. 60%.
   C. 70%.

KEY CONCEPTS

LOS 4.a

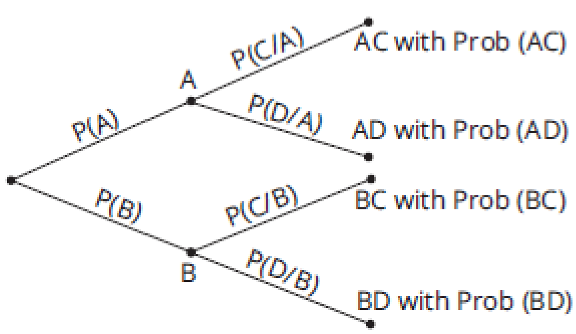
The expected value of a random variable is the weighted average of its possible outcomes:

$$E(X) = \Sigma P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$

Variance can be calculated as the probability-weighted sum of the squared deviations from the mean or expected value. The standard deviation is the positive square root of the variance.

LOS 4.b

A probability tree shows the probabilities of two events and the conditional probabilities of two subsequent events:



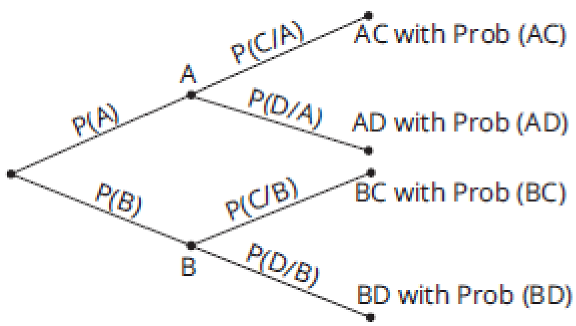
Conditional expected values depend on the outcome of some other event. Forecasts of expected values for a stock's return, earnings, and dividends can be refined, using conditional expected values, when new information arrives that affects the expected outcome.

LOS 4.c

Bayes' formula for updating probabilities based on the occurrence of an event *O* is as follows:

$$P(I|O) = \frac{P(O|I)}{P(O)} \times P(I)$$

Equivalently, based on the following tree diagram, $P(A|C) = \frac{P(AC)}{P(AC) + P(BC)}$:



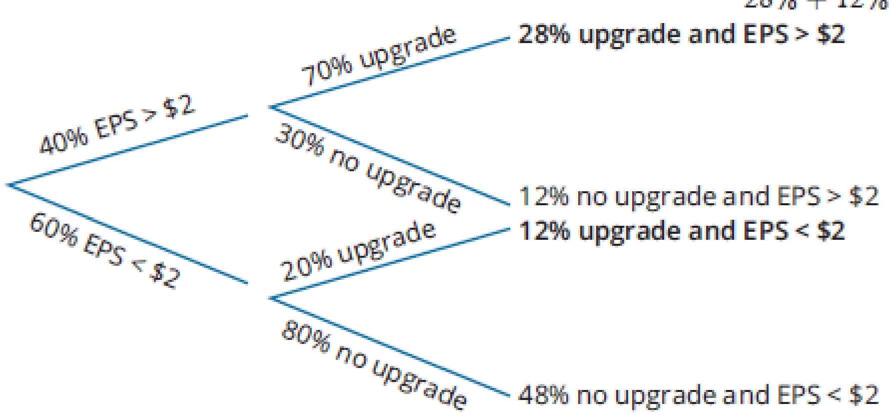
ANSWER KEY FOR MODULE QUIZZES

Module Quiz 4.1

1. **B** $E(X \mid Y = 1) = (0.2)(0) + (0.4)(5) + (0.4)(10) = 6$
$E(X \mid Y = 2) = (0.1)(0) + (0.8)(5) + (0.1)(10) = 5$
$E(X) = (0.3)(6) + (0.7)(5) = 5.30$

(LOS 4.a)

2. **C** This is an application of Bayes' formula. As the following tree diagram shows, the updated probability that EPS are greater than $2 is $\frac{28\%}{28\% + 12\%} = 70\%$:



(LOS 4.c)

READING 5

PORTFOLIO MATHEMATICS

MODULE 5.1: PROBABILITY MODELS FOR PORTFOLIO RETURN AND RISK

Video covering this content is available online.

LOS 5.a: Calculate and interpret the expected value, variance, standard deviation, covariances, and correlations of portfolio returns.

The **expected return of a portfolio** composed of *n* assets with weights, w_i, and expected returns, R_i, can be determined using the following formula:

$$E(R_p) = \sum_{i=1}^{n} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \ldots + w_n E(R_n)$$

The expected return and variance for a portfolio of assets can be determined using the properties of the individual assets in the portfolio. To do this, it is necessary to establish the portfolio weight for each asset. As indicated in the formula below, the weight, *w*, of Asset *i* is simply the market value currently invested in the asset divided by the current market value of the entire portfolio:

$$w_i = \frac{\text{market value of investment in Asset i}}{\text{market value of the portfolio}}$$

In many finance situations, we are interested in how two random variables move in relation to each other. For investment applications, one of the most frequently analyzed pairs of random variables is the returns of two assets. Investors and managers frequently ask questions such as, “What is the relationship between the return for Stock A and Stock B?” or, “What is the relationship between the performance of the S&P 500 and that of the automotive industry?”

Covariance is a measure of how two assets move together. It is the expected value of the product of the deviations of the two random variables from their respective expected values. A common symbol for the covariance between random variables *X* and *Y* is Cov(X,Y). Because we will be mostly concerned with the covariance of asset returns, the following formula has been written in terms of the covariance of the return of Asset *i*, R_i, and the return of Asset *j*, R_j:

$$\text{Cov}(R_i,R_j) = E\{[R_i - E(R_i)][R_j - E(R_j)]\}$$

The following are *properties of covariance*:

- The covariance of a random variable with itself is its variance; that is, $Cov(R_A,R_A) = Var(R_A)$.
- Covariance may range from negative infinity to positive infinity.
- A positive covariance indicates that when one random variable is above its mean, the other random variable also tends to be above its mean.
- A negative covariance indicates that when one random variable is above its mean, the other random variable tends to be below its mean.

The **sample covariance** for a sample of returns data can be calculated as follows:

$$s_{X,Y} = \frac{\sum_{i=1}^{n}\left\{\left[R_{1,i}-\overline{R_1}\right]\left[R_{2,i}-\overline{R_2}\right]\right\}}{n-1}$$

where:

$R_{1,i}$ = an observation of returns on Asset 1

$R_{2,i}$ = an observation of returns on Asset 2

$\overline{R_1}$ = mean return of Asset 1

$\overline{R_2}$ = mean return of Asset 2

n = number of observations in the sample

A **covariance matrix** shows the covariances between returns on a group of assets.

Figure 5.1: Covariance Matrix for Assets A, B, and C

Asset	A	B	C
A	$Cov(R_A,R_A)$	$Cov(R_A,R_B)$	$Cov(R_A,R_C)$
B	$Cov(R_B,R_A)$	$Cov(R_B,R_B)$	$Cov(R_B,R_C)$
C	$Cov(R_C,R_A)$	$Cov(R_C,R_B)$	$Cov(R_C,R_C)$

Note that the diagonal terms from top left are the variances of each asset's returns—in other words, $Cov(R_A,R_A) = Var(R_A)$.

The covariance between the returns on two assets does not depend on order—in other words, $Cov(R_A,R_B) = Cov(R_B,R_A)$—so in this covariance matrix, only three of the (off-diagonal) covariance terms are unique. In general for *n* assets, there are *n* variance terms (on the diagonal) and n(n − 1) / 2 unique covariance terms.

With **portfolio variance**, to calculate the variance of portfolio returns, we use the asset weights, returns variances, and returns covariances:

$$Var(R_P) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j Cov(R_i,R_j)$$

The variance of a portfolio composed of two risky assets, A and B, can be expressed as follows:

$$Var(R_p) = w_A w_A Cov(R_A,R_A) + w_A w_B Cov(R_A,R_B) + w_B w_A Cov(R_B,R_A) + w_B w_B Cov(R_B,R_B)$$

We can write this more simply as:

$$\text{Var}(R_p) = w_A^2 \text{Var}(R_A) + w_B^2 \text{Var}(R_B) + 2w_A w_B \text{Cov}(R_A, R_B)$$

or:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2w_A w_B \text{Cov}_{AB}$$

For a three-asset portfolio, the portfolio variance is:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + w_C^2 \sigma_C^2 + 2w_A w_B \text{Cov}_{AB} + 2w_A w_C \text{Cov}_{AC} + 2w_B w_C \text{Cov}_{BC}$$

Consider a portfolio with three assets: an index of domestic stocks (60%), an index of domestic bonds (30%), and an index of international equities (10%). A covariance matrix of the three assets is shown here.

Figure 5.2: Covariance Matrix for the Three Assets

Asset	Domestic Stocks	Domestic Bonds	International Equities
Domestic stocks	400	44	180
Domestic bonds	44	70	35
International equities	180	35	450

Portfolio returns variance =

$$(0.6^2)400 + (0.3^2)70 + (0.1^2)450 + 2(0.6)(0.3)44 + 2(0.6)(0.1)180 + 2(0.3)(0.1)35 = 194.34$$

Portfolio returns standard deviation = $\sqrt{194.34} = 13.94\%$

Note that the units of variance and covariance are $\%^2$ (i.e., 0.001). When we put these values in as whole numbers (in $\%^2$), the portfolio variance is in $\%^2$, and the standard deviation is in whole percentages. We could also put variance and covariance in as decimals and get both the portfolio returns variance and standard deviation as decimals.

From the formula for portfolio returns variance, we can see that the lower the covariance terms, the lower the portfolio variance (and standard deviation). This is true for positive values of covariance, as well as negative values.

Recall that the correlation coefficient for two variables is:

$$\rho_{AB} = \frac{\text{Cov}_{AB}}{\sigma_A \sigma_B}\sigma, \text{ so that } (\text{Cov}_{AB}) = \rho_{AB} \times \sigma_A \sigma_B$$

This can be substituted for Cov_{AB} in our formula for portfolio returns variance. With this substitution, we can use a correlation matrix to calculate portfolio returns variance, rather than using covariances.

Figure 5.3: Correlation Matrix for the Three Assets

Asset	Domestic Stocks	Domestic Bonds	International Equities
Domestic stocks	1.000	0.263	0.424
Domestic bonds	0.263	1.000	0.197
International equities	0.424	0.197	1.000

Note that the correlations of asset returns with themselves (the diagonal terms) are all 1.

LOS 5.b: Calculate and interpret the covariance and correlation of portfolio returns using a joint probability function for returns.

EXAMPLE: Covariance of returns from a joint probability function

Assume that the economy can be in three possible states (S) next year: boom, normal, or slow economic growth. An expert source has calculated that P(boom) = 0.30, P(normal) = 0.50, and P(slow) = 0.20. The returns for Asset A, R_A, and Asset B, R_B, under each of the economic states are provided in the probability model as follows. What is the covariance of the returns for Asset A and Asset B?

Joint Probability Function

	$R_B = 30\%$	$R_B = 10\%$	$R_B = 0\%$
$R_A = 20\%$	0.30	0	0
$R_A = 12\%$	0	0.50	0
$R_A = 5\%$	0	0	0.20

The table gives us the joint probability of returns on Assets A and B (e.g., there is a 30% probability that the return on Asset A is 20% and the return on Asset B is 30%, and there is a 50% probability that the return on Asset A is 12% and the return on Asset B is 10%).

Answer:

First, we must calculate the expected returns for each of the assets:

$$E(R_A) = (0.3)(0.20) + (0.5)(0.12) + (0.2)(0.05) = 0.13$$

$$E(R_B) = (0.3)(0.30) + (0.5)(0.10) + (0.2)(0.00) = 0.14$$

The covariance can now be computed using the procedure described in the following table.

Covariance Calculation

Probability	R_A	R_B	Probability × $[R_A - E(R_A)] \times [R_B - E(R_B)]$
0.3	0.20	0.30	$(0.3)(0.2 - 0.13)(0.3 - 0.14) = 0.00336$
0.5	0.12	0.10	$(0.5)(0.12 - 0.13)(0.1 - 0.14) = 0.00020$
0.2	0.05	0.00	$(0.2)(0.05 - 0.13)(0 - 0.14) = 0.00224$

The covariance of returns for Asset A and Asset B is 0.00336 + 0.00020 + 0.00224 = 0.0058.

LOS 5.c: Define shortfall risk, calculate the safety-first ratio, and identify an optimal portfolio using Roy's safety-first criterion.

Shortfall risk is the probability that a portfolio value or return will fall below a particular target value or return over a given period.

Roy's safety-first criterion states that the optimal portfolio minimizes the probability that the return of the portfolio falls below some minimum acceptable level. This minimum acceptable level is called the **threshold level**. Symbolically, Roy's safety-first criterion can be stated as follows:

$$\text{minimize } P(R_p < R_L)$$

where:

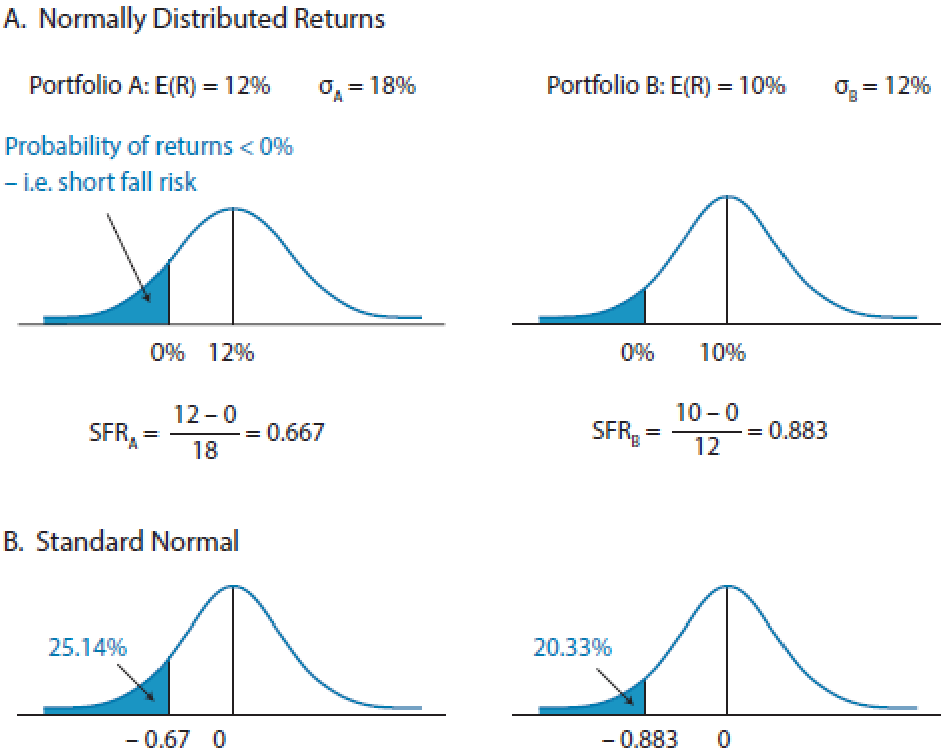
R_p = portfolio return
R_L = threshold level return

If portfolio returns are normally distributed, then Roy's safety-first criterion can be stated as follows:

$$\text{maximize safety-first ratio, which equals } \frac{E(R_P) - R_L}{\sigma_P}$$

The reasoning behind the safety-first criterion is illustrated in Figure 5.4. Assume an investor is choosing between two portfolios: Portfolio A, with an expected return of 12% and standard deviation of returns of 18%, and Portfolio B, with an expected return of 10% and standard deviation of returns of 12%. The investor has stated that he wants to minimize the probability of losing money (negative returns). Assuming that returns are normally distributed, the portfolio with the larger safety-first ratio using 0% as the threshold return (R_L) will be the one with the lower probability of negative returns.

Figure 5.4: The Safety-First Criterion and Shortfall Risk



Panel B of Figure 5.4 relates the safety-first ratio to the standard normal distribution. Note that the safety-first ratio is the number of standard deviations *below* the mean. Thus, the portfolio with the larger safety-first ratio has the lower probability of returns below the threshold return, which is a return of 0% in our example. Using a *z*-table for negative values, we can find the probabilities in the left-hand tails as indicated. These probabilities (25% for Portfolio A and 20% for Portfolio B) are also the shortfall risk for a target return of 0%—that is, the probability of negative returns. Portfolio B has the higher safety-first ratio, which means it has the lower probability of negative returns.

In summary, when choosing among portfolios with normally distributed returns using Roy's safety-first criterion, there are two steps:

Step 1: Calculate the safety-first ratio $= \frac{E(R_P) - R_L}{\sigma_P}$.

Step 2: Choose the portfolio that has the *largest* safety-first ratio.

EXAMPLE: Roy's safety-first criterion

For the next year, the managers of a $120 million college endowment plan have set a minimum acceptable end-of-year portfolio value of $123.6 million. Three portfolios are being considered that have the expected returns and standard deviation shown in the first two rows of the following table. Determine which of these portfolios is the most desirable using Roy's safety-first criterion and the probability that the portfolio value will fall short of the target amount.

Answer:

The threshold return is R_L = (123.6 − 120) / 120 = 0.030 = 3%. The safety-first ratios are shown in the following table. As indicated, the best choice is Portfolio A because it has the largest safety-first ratio.

Roy's Safety-First Ratios

Portfolio	Portfolio A	Portfolio B	Portfolio C
$E(R_p)$	9%	11%	6.6%
σ_p	12%	20%	8.2%
SFRatio	$0.5 = \frac{(9-3)}{12}$	$0.4 = \frac{(11-3)}{20}$	$0.44 = \frac{(6.6-3)}{8.2}$

The probability of an ending value for Portfolio A less than $123.6 million (a return less than 3%) is simply F(−0.5), which we can find on the *z*-table for negative values. The probability is 0.3085 = 30.85%.

MODULE QUIZ 5.1

1. The correlation of returns between Stocks A and B is 0.50. The covariance between these two securities is 0.0043, and the standard deviation of the return of Stock B is 26%. The variance of returns for Stock A is:
   A. 0.0011.
   B. 0.0331.
   C. 0.2656.

2. Given the following joint probability table for the returns on Assets P and Q:

	Q = 7%	Q = 4%	Q = 0%
P = 15%	0.2	0	0
P = 12%	0	0.2	0
P = 0%	0	0	0.6

   The covariance between P and Q is *closest* to:

   A. 18.0.
   B. 18.7.
   C. 19.3.

3. Expected returns and standard deviations of returns for three portfolios are as follows:

	Portfolio A	Portfolio B	Portfolio C
$E(R_p)$	5%	11%	18%
σ_p	8%	21%	40%

   Given a threshold level of return of 4%, the optimal portfolio using Roy's safety-first criterion is:

   A. Portfolio A.
   B. Portfolio B.

C. Portfolio C.

KEY CONCEPTS

LOS 5.a

The expected return of a portfolio composed of n assets with weights, w_i, and expected returns, R_i, is:

$$E(R_P) = \sum_{i=1}^{n} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \ldots + w_n E(R_n)$$

The variance of a portfolio composed of two risky assets, A and B, can be expressed as follows:

$$\sigma_P^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \text{Cov}_{AB}$$

where Cov_{AB} is the expected value of the product of the deviations of the two assets' returns from their respective expected values.

The variance of a two-asset portfolio can also be expressed as follows:

$$\text{Var}(R_P) = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_A \sigma_B \rho_{A,B}$$

where $\rho_{A,B}$ is the correlation of the two assets' returns.

LOS 5.b

Given the joint probabilities for A_i and B_i, the covariance is calculated as follows:

$$\sum_{i=1}^{n} P(A_i B_i)[A_i - E(A)][B_i - E(B)]$$

LOS 5.c

Shortfall risk is the probability that a portfolio's value (or return) will fall below a specific value over a given period.

The safety-first ratio for Portfolio P, based on a target return R_T, is:

$$\text{SFRatio} = \frac{E(R_P) - R_T}{\sigma_P}$$

Greater safety-first ratios are preferred and indicate a smaller shortfall probability. Roy's safety-first criterion states that the optimal portfolio minimizes shortfall risk.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 5.1

1. **A** $\text{Corr}(R_A, R_B) = \dfrac{\text{Cov}(R_A, R_B)}{[\sigma(R_A)][\sigma(R_B)]}$

$$\sigma^2(R_A) = \left[\frac{\text{Cov}(R_A, R_B)}{\sigma(R_B)\text{Corr}(R_A, R_B)}\right]^2$$

$$= \left[\frac{0.0043}{(0.26)(0.5)}\right]^2 = 0.0331^2 = 0.0011$$

(LOS 5.a)

2. **B** Expected value of P $= 0.2 \times 15 + 0.2 \times 12 + 0.6 \times 0 = 5.4\%$

Expected value of Q $= 0.2 \times 7 + 0.2 \times 4 + 0.6 \times 0 = 2.2\%$

Covariance $= 0.2 \times (15 - 5.4) \times (7 - 2.2) + 0.2 \times (12 - 5.4) \times (4 - 2.2) + 0.6\,(0 - 5.4)\,(0 - 2.2) = 18.72$

(LOS 5.b)

3. **C** Safety-first ratio for Portfolio A = (5 − 4) / 8 = 0.125

Safety-first ratio for Portfolio B = (11 − 4) / 21 = 0.300

Safety-first ratio for Portfolio C = (18 − 4) / 40 = 0.350

The largest value is 0.35, so Portfolio C has the smallest probability of a return below the threshold. (LOS 5.c)

READING 6

SIMULATION METHODS

MODULE 6.1: LOGNORMAL DISTRIBUTIONS AND SIMULATION TECHNIQUES

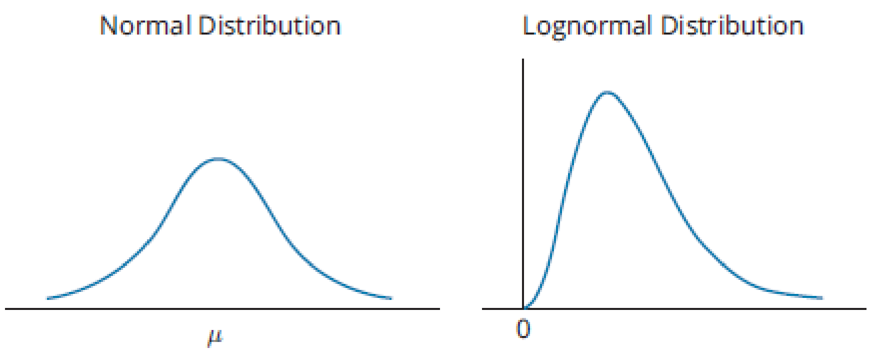
Video covering this content is available online.

LOS 6.a: Explain the relationship between normal and lognormal distributions and why the lognormal distribution is used to model asset prices when using continuously compounded asset returns.

The **lognormal distribution** is generated by the function e^x, where x is normally distributed. Because the natural logarithm, *ln*, of e^x is x, the logarithms of lognormally distributed random variables are normally distributed, thus the name.

Figure 6.1 illustrates the differences between a normal distribution and a lognormal distribution.

Figure 6.1: Normal vs. Lognormal Distributions



Recall from our reading about rates and returns that we use the natural logarithm to calculate continuously compounded returns. The lognormal distribution is useful for modeling asset prices if we think of an asset's future price as the result of a continuously compounded return on its current price. That is:

$$P_T = P_0 e^{r_{0,T}}$$

where:

P_T = asset price at time T

P_0 = asset price at time 0 (today)

$r_{0,T}$ = continuously compounded return on the asset from time 0 to time T

Because continuously compounded returns are additive, we can divide the time from 0 to T into shorter periods and state that $r_{0,T}$ is the sum of the continuously compounded returns over each of these shorter periods. Then, if we assume each of these returns is normally distributed, we can state that $r_{0,T}$ is normally distributed. Even if they are not normally distributed, the central limit theorem implies that their sum ($r_{0,T}$) is approximately normally distributed. This allows us to say P_T is lognormally distributed, because it is proportional to the logarithm of a normally distributed variable.

In many of the pricing models that we will see in the CFA curriculum, we assume returns are **independently and identically distributed**. If returns are *independently distributed*, past returns are not useful for predicting future returns. If returns are *identically distributed*, their mean and variance do not change over time (a property known as *stationarity* that is important in time series modeling, a Level II topic).

LOS 6.b: Describe Monte Carlo simulation and explain how it can be used in investment applications.

Monte Carlo simulation is a technique based on the repeated generation of one or more risk factors that affect security values to generate a distribution of security values. For each of the risk factors, the analyst must specify the parameters of the probability distribution that the risk factor is assumed to follow. A computer is then used to generate random values for each risk factor based on its assumed probability distributions. Each set of randomly generated risk factors is used with a pricing model to value the security. This procedure is repeated many times (100s, 1,000s, or 10,000s), and the distribution of simulated asset values is used to draw inferences about the expected (mean) value of the security—and possibly the variance of security values about the mean as well.

As an example, consider the valuation of stock options that can only be exercised on a particular date. The main risk factor is the value of the stock itself, but interest rates could affect the valuation as well. The simulation procedure would be to do the following:

1. Specify the probability distributions of stock prices and of the relevant interest rate, as well as the parameters (e.g., mean, variance, skewness) of the distributions.
2. Randomly generate values for both stock prices and interest rates.
3. Value the options for each pair of risk factor values.
4. After many iterations, calculate the mean option value and use that as your estimate of the option's value.

Monte Carlo simulation is used to do the following:

- Value complex securities.
- Simulate the profits/losses from a trading strategy.
- Calculate estimates of value at risk (VaR) to determine the riskiness of a portfolio of assets and liabilities.

- Simulate pension fund assets and liabilities over time to examine the variability of the difference between the two.
- Value portfolios of assets that have nonnormal return distributions.

An advantage of Monte Carlo simulation is that its inputs are not limited to the range of historical data. This allows an analyst to test scenarios that have not occurred in the past. The limitations of Monte Carlo simulation are that it is fairly complex and will provide answers that are no better than the assumptions about the distributions of the risk factors and the pricing/valuation model that is used. Also, simulation is not an analytic method, but a statistical one, and cannot offer the insights provided by an analytic method.

LOS 6.c: Describe the use of bootstrap resampling in conducting a simulation based on observed data in investment applications.

Resampling is another method for generating data inputs to use in a simulation. Often, we do not (or cannot) have data for a population, and can only approximate the population by sampling from it. (For example, we may think of the observed historical returns on an investment as a sample from the population of possible return outcomes.) To conduct resampling, we start with the observed sample and repeatedly draw subsamples from it, each with the same number of observations. From these samples, we can infer parameters for the population, such as its mean and variance.

In our reading on Estimation and Inference, we will describe some of the available resampling techniques. One of these is known as **bootstrap resampling**. In bootstrap resampling, we draw repeated samples of size *n* from the full dataset, replacing the sampled observations each time so that they might be redrawn in another sample. We can then directly calculate the standard deviation of these sample means as our estimate of the standard error of the sample mean.

Simulation using data from bootstrap resampling follows the same procedure as Monte Carlo simulation. The difference is the source and scope of the data. For example, if a simulation uses bootstrap resampling of historical returns data, its inputs are limited by the distribution of actual outcomes.

MODULE QUIZ 6.1

1. For a lognormal distribution the:
   A. mean equals the median.
   B. probability of a negative outcome is zero.
   C. probability of a positive outcome is 50%.
2. Which of the following is *least likely* to be a limitation of Monte Carlo analysis?
   A. Monte Carlo simulation is a statistical rather than an analytic method.
   B. Results of the analysis are no better than the assumptions used to generate it.
   C. Monte Carlo simulation is unable to provide answers to “what if” questions.
3. Which of the following is *most likely* a strength of bootstrapping?
   A. Offers a representation of the statistical features of a population.
   B. Provides only statistical estimates, not exact results.

C. Inputs may be limited by the distribution of actual outcomes.

KEY CONCEPTS

LOS 6.a

If x is normally distributed, e^x follows a lognormal distribution. The lognormal distribution is useful for modeling an asset's future price as the result of a continuously compounded return on its current price.

If investment returns are independently distributed, past returns are not useful for predicting future returns.

If investment returns are identically distributed, their mean and variance do not change over time.

LOS 6.b

Monte Carlo simulation uses randomly generated values for risk factors, based on their assumed distributions, to produce a distribution of possible security values. Its limitations are that it is fairly complex and will provide answers that are no better than the assumptions used.

LOS 6.c

Bootstrap resampling involves drawing repeated samples from a sample that represents the population, replacing the sampled observations each time so that they might be redrawn in another sample. The standard deviation of these sample means is an estimate of the standard error of the sample mean. As with all simulation techniques, its answers are no better than the assumptions used.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 6.1

1. **B** A lognormally distributed variable is never negative. (LOS 6.a)
2. **C** The ability to address "what if" questions (i.e., using input data outside the range that has been observed historically) is an advantage of Monte Carlo simulation. (LOS 6.b)
3. **A** One of the strengths of bootstrapping is that it offers a good representation of the statistical features of a population. However, this method does not provide exact results, and the inputs can be limited by the distribution of actual outcomes. (LOS 6.c)

READING 7

ESTIMATION AND INFERENCE

MODULE 7.1: SAMPLING TECHNIQUES AND THE CENTRAL LIMIT THEOREM

Video covering this content is available online.

LOS 7.a: Compare and contrast simple random, stratified random, cluster, convenience, and judgmental sampling and their implications for sampling error in an investment problem.

Probability sampling refers to selecting a sample when we know the probability of each sample member in the overall population. With **random sampling**, each item is assumed to have the same probability of being selected. If we have a population of data and select our sample by using a computer to randomly select a number of observations from the population, each data point has an equal probability of being selected—we call this **simple random sampling**. If we want to estimate the mean profitability for a population of firms, this may be an appropriate method.

Nonprobability sampling is based on either low cost and easy access to some data items, or on using the judgment of the researcher in selecting specific data items. Less randomness in selection may lead to greater sampling error.

Probability Sampling Methods

Simple random sampling is a method of selecting a sample in such a way that each item or person in the population being studied has the same likelihood of being included in the sample. As an example of simple random sampling, assume that you want to draw a sample of 5 items out of a group of 50 items. This can be accomplished by numbering each of the 50 items, placing them in a hat, and shaking the hat. Next, one number can be drawn randomly from the hat. Repeating this process (experiment) four more times results in a set of five numbers. The five drawn numbers (items) comprise a simple random sample from the population. In applications like this one, a random-number table or a computer random-number generator is often used to create the sample. Another way to form an approximately random sample is **systematic sampling**—selecting every *n*th member from a population.

Stratified random sampling uses a classification system to separate the population into smaller groups based on one or more distinguishing characteristics. From each subgroup, or stratum, a random sample is taken and the results are pooled. The size of the samples from each stratum is based on the size of the stratum relative to the population.

Stratified sampling is often used in bond indexing because of the difficulty and cost of completely replicating the entire population of bonds. In this case, bonds in a population are categorized (stratified) according to major bond risk factors including, but not limited to, duration, maturity, and coupon rate. Then, samples are drawn from each separate category and combined to form a final sample.

To see how this works, suppose you want to construct a portfolio of 100 bonds that is indexed to a major municipal bond index of 1,000 bonds, using a stratified random sampling approach. First, the entire population of 1,000 municipal bonds in the index can be classified on the basis of maturity and coupon rate. Then, cells (stratum) can be created for different maturity/coupon combinations, and random samples can be drawn from each of the maturity/coupon cells. To sample from a cell containing 50 bonds with 2-to-4-year maturities and coupon rates less than 5%, we would select five bonds. The number of bonds drawn from a given cell corresponds to the cell's weight relative to the population (index), or (50/1,000) × 100 = 5 bonds. This process is repeated for all the maturity/coupon cells, and the individual samples are combined to form the portfolio.

By using stratified sampling, we guarantee that we sample five bonds from this cell. If we had used simple random sampling, there would be no guarantee that we would sample any of the bonds in the cell. Or, we may have selected more than five bonds from this cell.

Cluster sampling is also based on subsets of a population, but in this case, we are assuming that each subset (cluster) is representative of the overall population with respect to the item we are sampling. For example, we may have data on personal incomes for a state's residents by county. The data for each county is a cluster.

In **one-stage cluster sampling**, a random sample of clusters is selected, and all the data in those clusters comprise the sample. In **two-stage cluster sampling**, random samples from each of the selected clusters comprise the sample. Contrast this with stratified random sampling, in which random samples are selected from every subgroup.

To the extent that the subgroups do not have the same distribution as the entire population of the characteristic of interest, cluster sampling will have greater sampling error than simple random sampling. Two-stage cluster sampling can be expected to have greater sampling error than one-stage cluster sampling. Lower cost and less time required to assemble the sample are the primary advantages of cluster sampling, and it may be most appropriate for a smaller pilot study.

Nonprobability Sampling Methods

Convenience sampling refers to selecting sample data based on ease of access, using data that are readily available. Because such a sample is typically not random, sampling error will be greater. An analyst should initially look at the data before adopting a sampling method with less sampling error.

Judgmental sampling refers to samples for which each observation is selected from a larger dataset by the researcher, based on one's experience and judgment. As an example, a researcher interested in assessing company compliance with accounting standards may have experience suggesting that evidence of noncompliance is typically found in certain ratios derived from the financial statements. The researcher may select only data on these items. Researcher bias (or simply poor judgment) may lead to samples that have excessive sampling error. In the absence of bias or poor judgment, judgmental sampling may produce a more representative sample or allow the researcher to focus on a sample that offers good data on the characteristic or statistic of interest.

An important consideration when sampling is ensuring that the distribution of data of interest is constant for the whole population being sampled. For example, judging a characteristic of U.S. banks using data from 2005 to 2015 may not be appropriate. Regulatory reform of the banking industry after the financial crisis of 2007–2008 may have resulted in significant changes in banking practices, so that the mean of a statistic precrisis and its mean value across the population of banks postcrisis are quite different. Pooling the data over the entire period from 2005 to 2015 would not be appropriate if this is the case, and the sample mean calculated from these data would not be a good estimate of either precrisis or postcrisis mean values.

LOS 7.b: Explain the central limit theorem and its importance for the distribution and standard error of the sample mean.

The **central limit theorem** states that for simple random samples of size *n* from a population with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean $\bar{x}$ approaches a normal probability distribution with mean μ and a variance equal to $\frac{\sigma^2}{n}$ as the sample size becomes large.

The central limit theorem is extremely useful because the normal distribution is relatively easy to apply to hypothesis testing and to the construction of confidence intervals. Specific inferences about the population mean can be made from the sample mean, *regardless of the population's distribution*, as long as the sample size is sufficiently large, which usually means $n \geq 30$.

Important properties of the central limit theorem include the following:

- If the sample size *n* is sufficiently large ($n \geq 30$), the sampling distribution of the sample means will be approximately normal. Remember what's going on here: random samples of size *n* are repeatedly being taken from an overall larger population. Each of these random samples has its own mean, which is itself a random variable, and this set of sample means has a distribution that is approximately normal.
- The mean of the population, μ, and the mean of the distribution of all possible sample means are equal.
- The variance of the distribution of sample means is $\frac{\sigma^2}{n}$, the population variance divided by the sample size.

The **standard error of the sample mean** is the standard deviation of the distribution of the sample means.

When the standard deviation of the population, σ, is *known*, the standard error of the sample mean is calculated as:

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

where:
$\sigma_{\bar{x}}$ = standard error of the sample mean
σ = standard deviation of the population
n = size of the sample

However, practically speaking, the *population's standard deviation is almost never known*. Instead, the standard error of the sample mean must be estimated by dividing the standard deviation of *the sample* by $\sqrt{n}$:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

EXAMPLE: **Standard error of sample mean (unknown population variance)**

Suppose a sample contains the past 30 monthly returns for McCreary, Inc. The mean return is 2%, and the *sample* standard deviation is 20%. Calculate and interpret the standard error of the sample mean.

Answer:

Because σ is unknown, this is the standard error of the sample mean:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{30}} = 3.6\%$$

This implies that if we took all possible samples of the size of 30 from McCreary's monthly returns and prepared a sampling distribution of the sample means, the mean would be 2% with a standard error of 3.6%.

EXAMPLE: **Standard error of sample mean (unknown population variance)**

Continuing with our example, suppose that instead of a sample size of 30, we take a sample of the past 200 monthly returns for McCreary, Inc. To highlight the effect of sample size on the sample standard error, let's assume that the mean return and standard deviation of this larger sample remain at 2% and 20%, respectively. Now, calculate the standard error of the sample mean for the 200-return sample.

Answer:

The standard error of the sample mean is computed as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20\%}{\sqrt{200}} = 1.4\%$$

The result of the preceding two examples illustrates an important property of sampling distributions. Notice that the value of the standard error of the sample mean decreased from 3.6% to 1.4% as the sample size increased from 30 to 200. This is because as the sample size increases, the sample mean gets closer, on average, to the true mean of the population. In other words, the distribution of the sample means about the population mean gets smaller and smaller, so the standard error of the sample mean decreases.

LOS 7.c: Describe the use of resampling (bootstrap, jackknife) to estimate the sampling distribution of a statistic.

Previously, we used the sample variance to calculate the standard error of our estimate of the mean. The standard error provides better estimates of the distribution of sample means when the sample is unbiased and the distribution of sample means is approximately normal.

Two alternative methods of estimating the standard error of the sample mean involve resampling of the data. The first of these, termed the **jackknife**, calculates multiple sample means, each with one of the observations removed from the sample. The standard deviation of these sample means can then be used as an estimate of the standard error of sample means. The jackknife is a computationally simple tool and can be used when the number of observations available is relatively small. This method can remove bias from statistical estimates.

The jackknife (so named because it is a handy and readily available tool) was developed when computational power was not as readily available and as low cost as today. A **bootstrap** method is more computationally demanding, but it has some advantages. To estimate the standard error of the sample mean, we draw repeated samples of size n from the full dataset (replacing the sampled observations each time). We can then directly calculate the standard deviation of these sample means as our estimate of the standard error of the sample mean.

The bootstrap method can improve accuracy compared to using only the data in a single sample, and it can be used to construct confidence intervals for various statistics in addition to the mean, such as the median. This method can also be used to estimate the distributions of complex statistics, including those that do not have an analytic form.

MODULE QUIZ 7.1

1. A simple random sample is a sample drawn in such a way that each member of the population has:
   A. some chance of being selected in the sample.
   B. an equal chance of being included in the sample.
   C. a 1% chance of being included in the sample.
2. To apply the central limit theorem to the sampling distribution of the sample mean, the sample is usually considered to be large if n is at least:
   A. 20.
   B. 25.

C. 30.

3. Which of the following techniques to improve the accuracy of confidence intervals on a statistic is *most* computationally demanding?
   A. Jackknife resampling.
   B. Systematic resampling.
   C. Bootstrap resampling.

KEY CONCEPTS

LOS 7.a

Simple random sampling is a method of selecting a sample in such a way that each item or person in the population being studied has the same probability of being included in the sample.

Stratified random sampling involves randomly selecting samples proportionally from subgroups that are formed based on one or more distinguishing characteristics of the data, so that random samples from the subgroups will have the same distribution of these characteristics as the overall population.

Cluster sampling is also based on subgroups (not necessarily based on data characteristics) of a larger dataset. In one-stage cluster sampling, the sample is formed from randomly chosen clusters (subsets) of the overall dataset. In two-stage cluster sampling, random samples are taken from each of the randomly chosen clusters (subgroups).

Convenience sampling refers to selecting sample data based on ease of access, using data that are readily available. Judgmental sampling refers to samples for which each observation is selected from a larger dataset by the researcher, based on the researcher's experience and judgment. Both are examples of nonprobability sampling and are nonrandom.

LOS 7.b

The central limit theorem states that for a population with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean of all possible samples of size *n* (for $n \geq 30$) will be approximately normally distributed with a mean equal to μ and a variance equal to σ^2/n.

The standard error of the sample mean is the standard deviation of the distribution of the sample means and is calculated as $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ (where σ, the population standard deviation, is known) and as $s_{\bar{X}} = \frac{s}{\sqrt{n}}$ (where *s*, the sample standard deviation, is used because the population standard deviation is unknown).

LOS 7.c

Two resampling techniques to improve our estimates of the distribution of sample statistics are the jackknife and bootstrap. With the jackknife, we calculate *n* sample means, one with each observation in a sample of size *n* removed, and base our estimate on the standard error of sample means of size *n*. This can remove bias from our estimates based on the sample standard deviation without resampling.

With bootstrap resampling, we use the distribution of sample means (or other statistics) from a large number of samples of size *n*, drawn from a large dataset. Bootstrap resampling can improve our estimates of the distribution of various sample statistics and provide such estimates when analytical methods will not.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 7.1

1. **B** In a simple random sample, each element of the population has an equal probability of being selected. The 1% chance answer option allows for an equal chance, but only if there are 100 elements in the population from which the random sample is drawn. (LOS 7.a)

2. **C** Sample sizes of 30 or greater are typically considered large. (LOS 7.b)

3. **C** Bootstrap resampling, repeatedly drawing samples of equal size from a large dataset, is more computationally demanding than the jackknife. We have not defined *systematic resampling* as a specific technique. (LOS 7.c)

READING 8

HYPOTHESIS TESTING

MODULE 8.1: HYPOTHESIS TESTING BASICS

Video covering this content is available online.

LOS 8.a: Explain hypothesis testing and its components, including statistical significance, Type I and Type II errors, and the power of a test.

A hypothesis is a statement about the value of a population parameter developed for the purpose of testing a theory or belief. Hypotheses are stated in terms of the population parameter to be tested, like the population mean, μ. For example, a researcher may be interested in the mean daily return on stock options. Hence, the hypothesis may be that the mean daily return on a portfolio of stock options is positive.

Hypothesis testing procedures, based on sample statistics and probability theory, are used to determine whether a hypothesis is a reasonable statement and should not be rejected, or if it is an unreasonable statement and should be rejected. The process of hypothesis testing consists of a series of steps shown in Figure 8.1.

Figure 8.1: Hypothesis Testing Procedure

State the hypothesis

↓

Select the appropriate test statistic

↓

Specify the level of significance

↓

State the decision rule regarding the hypothesis

↓

Collect the sample and calculate the sample statistics

↓

Make a decision regarding the hypothesis

↓

Make a decision based on the results of the test

Source: Wayne W. Daniel and James C. Terrell, *Business Statistics, Basic Concepts and Methodology*, Houghton Mifflin, Boston, 1997.

The **null hypothesis**, designated H_0, is the hypothesis that the researcher wants to reject. It is the hypothesis that is actually tested and is the basis for the selection of the test statistics. The null is generally stated as a simple statement about a population parameter. A typical statement of the null hypothesis for the population mean is H_0: $\mu = \mu_0$, where μ is the population mean and μ_0 is the hypothesized value of the population mean.

PROFESSOR'S NOTE
The null hypothesis always includes the "equal to" condition.

The **alternative hypothesis**, designated H_a, is what is concluded if there is sufficient evidence to reject the null hypothesis and is usually what you are really trying to assess. Why? You can never really prove anything with statistics—when the null hypothesis is discredited, the implication is that the alternative hypothesis is valid.

For a null hypothesis H_0: $\mu = \mu_0$, the alternative hypothesis is H_a: $\mu \neq \mu_0$. Notice that the null and alternative hypotheses are *mutually exclusive and exhaustive*. They include all possible outcomes of the test (they are exhaustive), and no possible outcome of the test satisfies both hypotheses (they are mutually exclusive).

Let's look at the development of the decision rule for a two-tailed test using a z-distributed test statistic (a z-test) at a 5% level of significance, $\alpha = 0.05$:

- At $\alpha = 0.05$, the computed test statistic is compared with the critical z-values of ±1.96. The values of ±1.96 correspond to $\pm z_{\alpha/2} = \pm z_{0.025}$, which is the range of z-values within which 95% of the probability lies. These values are obtained from the cumulative probability table for the standard normal distribution (z-table), which is included in the Appendix section of this book.
- If the computed test statistic falls outside the range of critical z-values (i.e., test statistic > 1.96, or test statistic < −1.96), we reject the null and conclude that the sample statistic is sufficiently different from the hypothesized value.
- If the computed test statistic falls within the range ±1.96, we conclude that the sample statistic is not sufficiently different from the hypothesized value ($\mu = \mu_0$, in this case), and we fail to reject the null hypothesis.

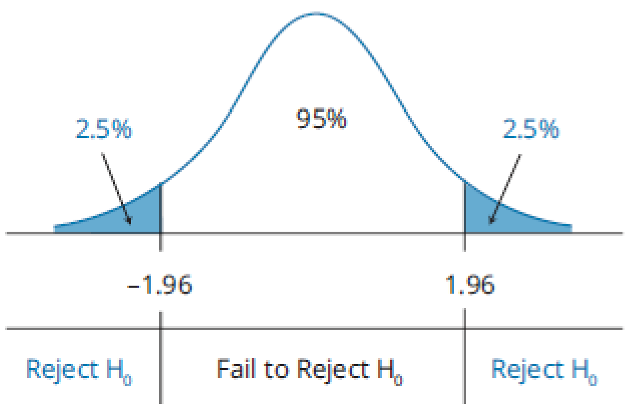
The **decision rule** (rejection rule) for a two-tailed z-test at $\alpha = 0.05$ can be stated as follows:

$$\text{Reject } H_0 \text{ if test statistic} < -1.96 \text{ or test statistic} > 1.96$$

Figure 8.2 shows the standard normal distribution for a two-tailed hypothesis test using the z-distribution. Notice that the significance level of 0.05 means that there is 0.05 / 2 = 0.025 probability (area) under each tail of the distribution beyond ±1.96.

Figure 8.2: Two-Tailed Hypothesis Test Using the Standard Normal (z)

Distribution



Hypothesis testing involves two statistics: (1) the test statistic calculated from the sample data and (2) the **critical value** of the test statistic. The value of the computed test statistic relative to the critical value is a key step in assessing the validity of a hypothesis.

PROFESSOR'S NOTE

Recall from the prerequisite readings that candidates are expected to know the most frequently used critical values for hypothesis tests with *z*-statistics:

Critical Value	Use for:
1.65	2-tailed test with 10% significance level, *or* 1-tailed test with 5% significance level
1.96	2 tailed test with 5% significance level
2.33	1-tailed test with 1% significance level
2.58	2-tailed test with 1% significance level

A test statistic is calculated by comparing the point estimate of the population parameter with the hypothesized value of the parameter (i.e., the value specified in the null hypothesis). As indicated in the following expression, the **test statistic** is the difference between the sample statistic and the hypothesized value, scaled by the standard error of the sample statistic:

$$\text{test statistic} = \frac{\text{sample statistic} - \text{hypothesized value}}{\text{standard error of the sample statistic}}$$

The standard error of the sample statistic is the adjusted standard deviation of the sample. When the sample statistic is the sample mean, $\bar{x}$, the standard error of the sample statistic for sample size *n*, is calculated as:

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

when the population standard deviation, σ, *is known*, or:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

when the population standard deviation, σ, *is not known*. In this case, it is estimated using the standard deviation of the sample, *s*.

PROFESSOR'S NOTE

Don't be confused by the notation here. A lot of the literature you will encounter in your studies simply uses the term $\sigma_{\bar{x}}$ for the standard error of the test statistic, regardless of whether the population standard deviation or sample standard deviation was used in its computation.

As you will soon see, a test statistic is a random variable that may follow one of several distributions, depending on the characteristics of the sample and the population. We will look at four distributions for test statistics: the *t*-distribution, the *z*-distribution (standard normal distribution), the chi-square distribution, and the *F*-distribution. The critical value for the appropriate test statistic—the value against which the computed test statistic is compared—depends on its distribution.

Type I and Type II Errors

Keep in mind that hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population. We must be aware that there is some probability that the sample, in some way, does not represent the population, and any conclusion based on the sample about the population may be made in error.

When drawing inferences from a hypothesis test, there are two types of errors:

- **Type I error.** This is the rejection of the null hypothesis when it is actually true.
- **Type II error.** This is the failure to reject the null hypothesis when it is actually false.

The **significance level** is the probability of making a Type I error (rejecting the null when it is true) and is designated by the Greek letter alpha (α). For instance, a significance level of 5% ($\alpha = 0.05$) means there is a 5% chance of rejecting a true null hypothesis. When conducting hypothesis tests, a significance level must be specified to identify the critical values needed to evaluate the test statistic.

While the significance level of a test is the probability of rejecting the null hypothesis when it is true, the **power of a test** is the probability of correctly rejecting the null hypothesis when it is false. The power of a test is actually one minus the probability of making a Type II error, or 1 − P(Type II error). In other words, the probability of rejecting the null when it is false (power of the test) equals one minus the probability of *not* rejecting the null when it is false (Type II error). When more than one test statistic may be used, the power of the test for the competing test statistics may be useful in deciding which test statistic to use. Ordinarily, we wish to use the test statistic that provides the most powerful test among all possible tests.

Figure 8.3 shows the relationship between the level of significance, the power of a test, and the two types of errors.

Figure 8.3: Type I and Type II Errors in Hypothesis Testing

Decision	True Condition	
	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Incorrect decision **Type II error**
Reject H_0	Incorrect decision **Type I error** Significance level, α, = P(Type I error)	Correct decision Power of the test = 1 – P(Type II error)

Sample size and the choice of significance level (Type I error probability) will together determine the probability of a Type II error. The relation is not simple, however, and calculating the probability of a Type II error in practice is quite difficult. Decreasing the significance level (probability of a Type I error) from 5% to 1%, for example, will increase the probability of failing to reject a false null (Type II error)—and, therefore, reduce the power of the test. Conversely, for a given sample size, we can increase the power of a test only with the cost that the probability of rejecting a true null (Type I error) increases. For a given significance level, we can decrease the probability of a Type II error and increase the power of a test only by increasing the sample size.

The decision for a hypothesis test is to either reject the null hypothesis or fail to reject the null hypothesis. Note that it is statistically incorrect to say "accept" the null hypothesis; it can only be supported or rejected. The **decision rule** for rejecting or failing to reject the null hypothesis is based on the distribution of the test statistic. For example, if the test statistic follows a normal distribution, the decision rule is based on critical values determined from the standard normal distribution (z-distribution). Regardless of the appropriate distribution, it must be determined if a one-tailed or two-tailed hypothesis test is appropriate before a decision rule (rejection rule) can be determined.

A decision rule is specific and quantitative. Once we have determined whether a one- or two-tailed test is appropriate, the significance level we require, and the distribution of the test statistic, we can calculate the exact critical value for the test statistic. Then, we have a decision rule of the following form: if the test statistic is (greater, less than) the value X, reject the null.

The ***p-value*** is the probability of obtaining a test statistic that would lead to a rejection of the null hypothesis, assuming the null hypothesis is true. It is the smallest level of significance for which the null hypothesis can be rejected.

MODULE QUIZ 8.1

1. For a hypothesis test with a probability of a Type II error of 60% and a probability of a Type I error of 5%, which of the following statements is *most accurate*?
   A. The power of the test is 40%, and there is a 5% probability that the test statistic will exceed the critical value(s).
   B. There is a 95% probability that the test statistic will be between the critical values, if this is a two-tailed test.

C. There is a 5% probability that the null hypothesis will be rejected when actually true, and the probability of rejecting the null when it is false is 40%.

2. If the significance level of a test is 0.05 and the probability of a Type II error is 0.15, what is the power of the test?
   A. 0.850.
   B. 0.950.
   C. 0.975.

MODULE 8.2: TYPES OF HYPOTHESIS TESTS

Video covering this content is available online.

LOS 8.b: Construct hypothesis tests and determine their statistical significance, the associated Type I and Type II errors, and power of the test given a significance level.

Here, we will show examples of hypothesis tests for population means and population variances:

- For a hypothesis concerning the *value of a population mean*, we use a *t*-test (or a *z*-test if the sample size is large enough).
- To test a hypothesis concerning the *equality of two population means*, we use a *t*-test. The nature of that test depends on whether the samples are independent (a difference in means test) or dependent (a paired comparisons test).
- For a hypothesis concerning the *value of a population variance*, we use a chi-square test.
- To test a hypothesis concerning the *equality of two population variances*, we use an *F*-test.

Value of a Population Mean

EXAMPLE: Hypothesis test of a population mean

A researcher has gathered data on the daily returns on a portfolio of call options over a recent 250-day period. The mean daily return has been 0.1%, and the sample standard deviation of daily portfolio returns is 0.25%. The researcher believes that the mean daily portfolio return is not equal to zero. Construct a hypothesis test of the researcher's belief.

Answer:

First, we need to specify the null and alternative hypotheses:

$$H_0: \mu_0 = 0 \text{ versus } H_a: \mu_0 \neq 0$$

The *t*-distribution is used to test a hypothesis about the value of a population mean. With 250 observations, however, this sample is considered to be large, so the *z*-distribution is acceptable. Because our sample is so large, the critical values for the *t* and z are almost identical. Hence, there is almost no difference in the likelihood of rejecting a true null.

At a 5% level of significance, the critical z-values for a two-tailed test are ±1.96, so the decision rule can be stated as:

Reject H_0 if test statistic < −1.96 or test statistic > +1.96

Given a sample size of 250 with a standard deviation of 0.25%, the standard error can be computed as:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.25\%}{\sqrt{250}} = 0.0158\%$$

Using the standard error of the sample mean, our test statistic is:

$$\frac{0.001}{\left(\frac{0.0025}{\sqrt{250}}\right)} = \frac{0.001}{0.000158} = 6.33$$

Because 6.33 > 1.96, we can reject the null hypothesis that the mean daily option return is equal to zero.

Difference Between Means (Independent Samples)

We frequently want to know if there is a difference between the means of two populations. The t-test for **differences between means** requires that we are reasonably certain that our samples are independent and that they are taken from two populations that are normally distributed.

PROFESSOR'S NOTE
Please note the language of the LOS here. Candidates must "[c]onstruct hypothesis tests and determine their statistical significance …" Certainly, you should know that this is a t-test, and that we reject the hypothesis of equality when the test statistic is outside the critical t-values. Don't worry about memorizing the following formulas.

A pooled variance is used with the t-test for testing the hypothesis that the means of two normally distributed populations are equal, when the variances of the populations are unknown but assumed to be equal.

Assuming independent samples, the t-statistic is computed as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

where:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

s_1^2 = variance of the first sample

s_2^2 = variance of the second sample

n_1 = number of observations in the first sample

n_2 = number of observations in the second sample

Note: degrees of freedom, *df*, is $(n_1 + n_2 - 2)$

Because we assume that the variances are equal, we just add the variances of the two sample means to calculate the standard error in the denominator.

The intuition here is straightforward. If the sample means are close together, the numerator of the *t*-statistic (and the *t*-statistic itself) are small, and we do not reject equality. If the sample means are far apart, the numerator of the *t*-statistic (and the *t*-statistic itself) are large, and we reject equality. Perhaps not as easy to remember is the fact that this test is only valid for two populations that are independent and normally distributed.

EXAMPLE: Difference between means—equal variances

Sue Smith is investigating whether the abnormal returns for acquiring firms during merger announcement periods differ for horizontal and vertical mergers. She estimates the abnormal returns for a sample of acquiring firms associated with horizontal mergers and a sample of acquiring firms involved in vertical mergers. Smith finds that abnormal returns from horizontal mergers have a mean of 1.0% and a standard deviation of 1.0%, while abnormal returns from vertical mergers have a mean of 2.5% and a standard deviation of 2.0%.

Smith assumes the samples are independent, the population means are normally distributed, and the population variances are equal.

Smith calculates the *t*-statistic as −5.474 and the degrees of freedom as 120. Using a 5% significance level, should Smith reject or fail to reject the null hypothesis that the abnormal returns to acquiring firms during the announcement period are the same for horizontal and vertical mergers?

Answer:

Because this is a two-tailed test, the structure of the hypotheses takes the following form:

$$H_0: \mu_1 - \mu_2 = 0 \text{ versus } H_a: \mu_1 - \mu_2 \neq 0$$

where:

μ_1 = the mean of the abnormal returns for the horizontal mergers

μ_2 = the mean of the abnormal returns for the vertical mergers

From the following *t*-table segment, the critical *t*-value for a 5% level of significance at α / 2 = p = 0.025 with df = 120, is 1.980.

Partial *t*-Table

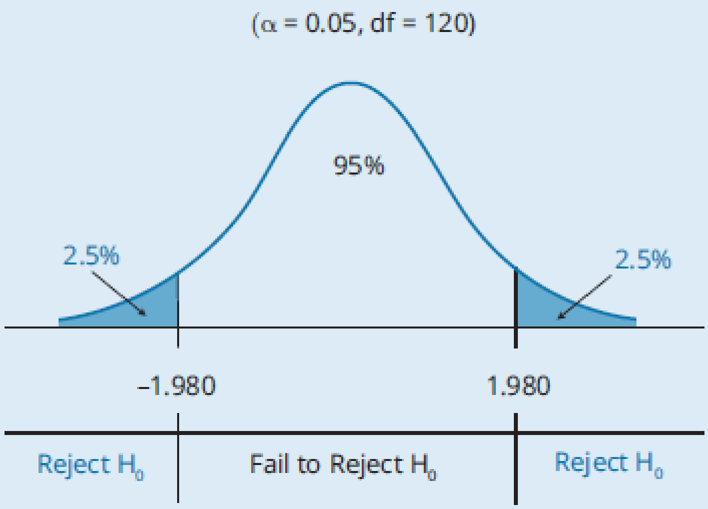
df	One-Tailed Probabilities (p)		
	p = 0.10	p = 0.05	p = 0.025
110	1.289	1.659	1.982
120	1.289	1.658	1.980
200	1.286	1.653	1.972

Thus, the decision rule can be stated as follows:

Reject H_0 if t-statistic < -1.980 or t-statistic > 1.980

The rejection region for this test is illustrated in the following figure.

Decision Rule for Two-Tailed *t*-Test



Because the test statistic, −5.474, falls to the left of the lower critical *t*-value, Smith can reject the null hypothesis and conclude that mean abnormal returns are different for horizontal and vertical mergers.

Paired Comparisons (Means of Dependent Samples)

While the test in the previous section was of the difference between the means of two independent samples, sometimes, our samples may be dependent. If the observations in the two samples both depend on some other factor, we can construct a **paired comparisons test** of whether the means of the differences between observations for the two samples are different. Dependence may result from an event that affects both sets of observations for a number of companies, or because observations for two firms over time are both influenced by market returns or economic conditions.

For an example of a paired comparisons test, consider a test of whether the returns on two steel firms were equal over a five-year period. We can't use the difference in means test because we have reason to believe that the samples are not independent. To some extent, both will depend on the returns on the overall market (market risk) and the conditions in the steel industry (industry-specific risk). In this case, our pairs will be the returns on each firm over the same time periods, so we use the differences in monthly returns for the two companies. The paired comparisons test is just a test of whether the average difference between monthly returns is significantly different from some value, typically zero, based on the standard error of the differences in monthly returns.

Remember, the paired comparisons test also requires that the sample data be normally distributed. Although we frequently just want to test the hypothesis that the mean of the differences in the pairs is zero ($\mu_{dz} = 0$), the general form of the test for any hypothesized mean difference, μ_{dz}, is as follows:

H_0: $\mu_d = \mu_{dz}$ versus H_a: $\mu_d \neq \mu_{dz}$

where:

μ_d = mean of the population of paired differences

μ_{dz} = hypothesized mean of paired differences, which is commonly zero

For one-tail tests, the hypotheses are structured as either of these:

H_0: $\mu_d \leq \mu_{dz}$ versus H_a: $\mu_d > \mu_{dz}$, or H_0: $\mu_d \geq \mu_{dz}$ versus H_a: $\mu_d < \mu_{dz}$

For the paired comparisons test, the *t*-statistic with n − 1 degrees of freedom is computed as follows:

$$t = \frac{\bar{d} - \mu_{dz}}{s_{\bar{d}}}$$

where:

$\bar{d}$ = sample mean difference $= \frac{1}{n}\sum_{i=1}^{n} d_i$

d_i = difference between the *i*th pair of observations

$s_{\bar{d}}$ = standard error of the mean difference $= \frac{s_d}{\sqrt{n}}$

s_d = sample standard deviation $= \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$

n = number of paired observations

EXAMPLE: Paired comparisons test

Joe Andrews is examining changes in estimated betas for the common stock of companies in the telecommunications industry before and after deregulation. Andrews believes the betas may decline because of deregulation, because companies are no longer subject to the uncertainties of rate regulation—or that they may increase because there is more uncertainty regarding competition in the industry. Andrews calculates a t-statistic of 10.26 for this hypothesis test, based on a sample size of 39. Using a 5% significance level, determine whether there is a change in betas.

Answer:

Because the mean difference may be positive or negative, a two-tailed test is in order here. Thus, the hypotheses are structured as:

$$H_0: \mu_d = 0 \text{ versus } H_a: \mu_d \neq 0$$

There are 39 − 1 = 38 degrees of freedom. Using the t-distribution, the two-tailed critical t-values for a 5% level of significance with df = 38 is ±2.024. As indicated in the following table, the critical t-value of 2.024 is located at the intersection of the p = 0.025 column and the df = 38 row. The one-tailed probability of 0.025 is used because we need 2.5% in each tail for 5% significance with a two-tailed test.

Partial *t*-Table

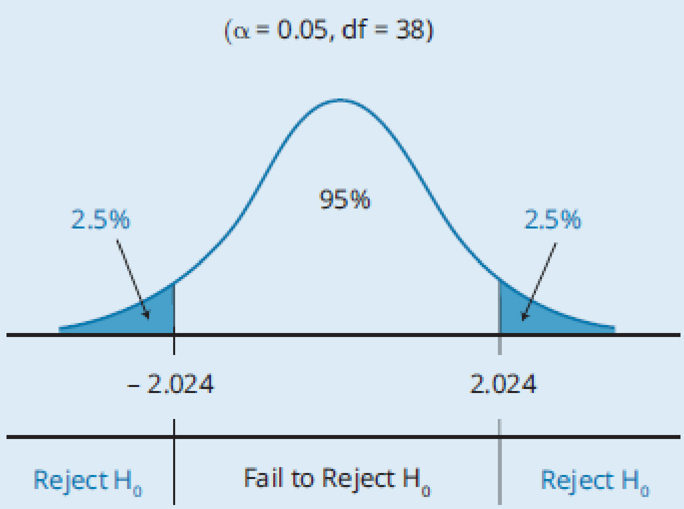
df	One-Tailed Probabilities (p)		
	p = 0.10	p = 0.05	p = 0.025
38	1.304	1.686	2.024
39	1.304	1.685	2.023
40	1.303	1.684	2.021

Thus, this is the decision rule:

$$\text{Reject } H_0 \text{ if } t\text{-statistic} < -2.024, \text{ or } t\text{-statistic} > 2.024$$

This decision rule is illustrated in the following figure.

Decision Rule for a Two-Tailed Paired Comparisons Test



The test statistic, 10.26, is greater than the critical t-value, 2.024—it falls in the rejection region to the right of 2.024 in the previous figure. Thus, we reject the null hypothesis of no difference, concluding that there *is* a statistically significant difference between mean firm betas before and after deregulation.

Keep in mind that we have been describing two distinct hypothesis tests: one about the significance of the difference between the means of two populations and one about the significance of the mean of the differences between pairs of observations. Here are rules for when these tests may be applied:

- The test of the differences in means is used when there are two *independent samples*.
- A test of the significance of the mean of the differences between paired observations is used when the samples are *not independent*.

PROFESSOR'S NOTE

Again, the LOS here says "[c]onstruct hypothesis tests and determine their statistical significance …" We can't believe candidates are expected to memorize these formulas (or that you would be a better analyst if you did). You should instead focus on the fact that both of these tests involve *t*-statistics and depend on the degrees of freedom. Also note that when samples are independent, you can use the difference in means test, and when they are dependent, we must use the paired comparison (mean differences) test. In that case, with a null hypothesis that there is no difference in means, the test statistic is simply the mean of the differences between each pair of observations, divided by the standard error of those differences. This is just a straightforward *t*-test of whether the mean of a sample is zero, which might be considered fair game for the exam.

Value of a Population Variance

The *chi-square test* is used for hypothesis tests concerning the variance of a normally distributed population. Letting σ^2 represent the true population variance and $\sigma_0{}^2$ represent the hypothesized variance, the hypotheses for a two-tailed test of a single population variance are structured as follows:

$$H_0: \sigma^2 = \sigma_0^2 \text{ versus } H_a: \sigma^2 \neq \sigma_0^2$$

The hypotheses for one-tailed tests are structured as follows:

$$H_0: \sigma^2 \leq \sigma_0^2 \text{ versus } H_a: \sigma^2 > \sigma_0^2 \text{ or } H_0: \sigma^2 \geq \sigma_0^2 \text{ versus } H_a: \sigma^2 < \sigma_0^2$$

Hypothesis testing of the population variance requires the use of a chi-square distributed test statistic, denoted χ^2. The chi-square distribution is asymmetrical and approaches the normal distribution in shape as the degrees of freedom increase.

To illustrate the chi-square distribution, consider a two-tailed test with a 5% level of significance and 30 degrees of freedom. As displayed in Figure 8.4, the critical chi-square values are 16.791 and 46.979 for the lower and upper bounds, respectively. These values are obtained from a chi-square table, which is used in the same manner as a *t*-table. A portion of a chi-square table is presented in Figure 8.5.

Note that the chi-square values in Figure 8.5 correspond to the probabilities in the right tail of the distribution. As such, the 16.791 in Figure 8.4 is from the column headed *0.975* because 95% + 2.5% of the probability is to the right of it. The 46.979 is from the column headed *0.025* because only 2.5% probability is to the right of it. Similarly, at a 5% level of significance with 10 degrees of freedom, Figure 8.5 shows that the critical chi-square values for a two-tailed test are 3.247 and 20.483.

Figure 8.4: Decision Rule for a Two-Tailed Chi-Square Test

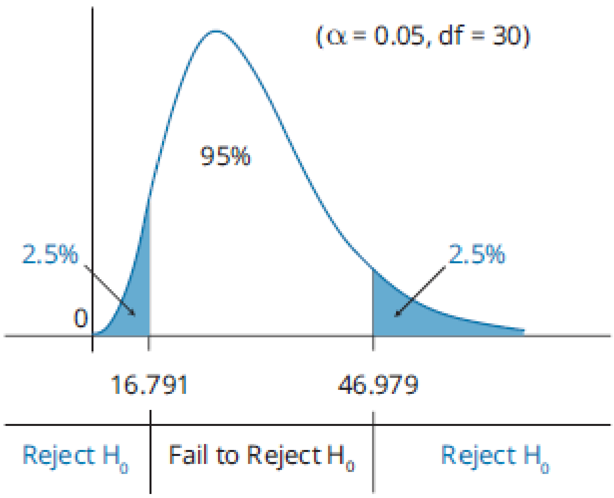


Figure 8.5: Chi-Square Table

	Probability in Right Tail					
Degrees of Freedom	0.975	0.95	0.90	0.1	0.05	0.025
9	2.700	3.325	4.168	14.684	16.919	19.023
10	3.247	3.940	4.865	15.987	18.307	20.483
11	3.816	4.575	5.578	17.275	19.675	21.920
30	16.791	18.493	20.599	40.256	43.773	46.979

The chi-square test statistic, χ^2, with n − 1 degrees of freedom, is computed as follows:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$$

where:

n = sample size

s^2 = sample variance

σ_0^2 = hypothesized value for the population variance

Similar to other hypothesis tests, the chi-square test compares the test statistic, χ^2_{n-1}, to a critical chi-square value at a given level of significance and n − 1 degrees of freedom. Because the chi-square distribution is bounded below by zero, chi-square values cannot be negative.

EXAMPLE: Chi-square test for a single population variance

Historically, the High-Return Equity Fund has advertised that its monthly returns have a standard deviation equal to 4%. This was based on estimates from the 2005–2013 period. High-Return wants to verify whether this claim still adequately describes the standard deviation of the fund's returns. High-Return collected monthly returns for the 24-month period between 2013 and 2015 and measured a standard deviation of monthly returns of 3.8%. High-Return calculates a test statistic of 20.76. Using a 5% significance level, determine if the more recent standard deviation is different from the advertised standard deviation.

Answer:

The null hypothesis is that the standard deviation is equal to 4% and, therefore, the variance of monthly returns for the population is $(0.04)^2 = 0.0016$. Because High-Return simply wants to test whether the standard deviation has changed, up or down, a two-sided test should be used. The hypothesis test structure takes this form:

$$H_0: \sigma_0^2 = 0.0016 \text{ versus } H_a: \sigma^2 \neq 0.0016$$

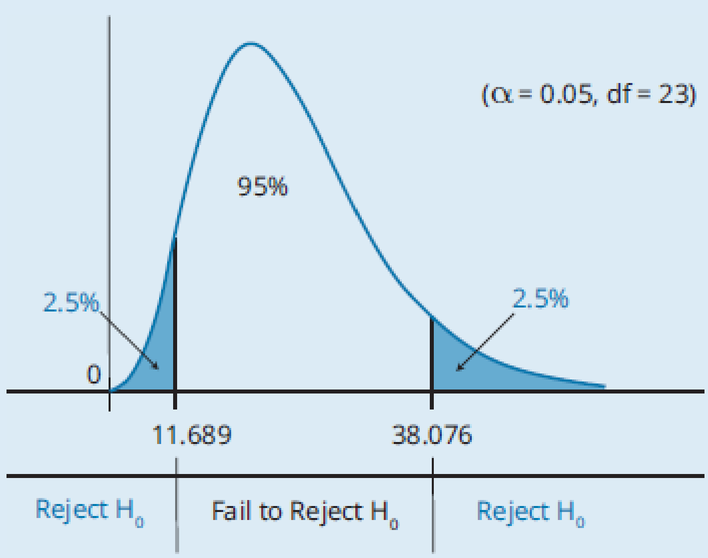
The appropriate test statistic for tests of variance is a chi-square statistic.

With a 24-month sample, there are 23 degrees of freedom. Using the table of chi-square values in Appendix E of this book, for 23 degrees of freedom and probabilities of 0.975 and 0.025, we find two critical values, 11.689 and 38.076. Thus, this is the decision rule:

$$\text{Reject } H_0 \text{ if } \chi^2 < 11.689, \text{ or } \chi^2 > 38.076$$

This decision rule is illustrated in the following figure.

Decision Rule for a Two-Tailed Chi-Square Test of a Single Population Variance



Because the computed test statistic, χ^2, falls between the two critical values, we cannot reject the null hypothesis that the variance is equal to 0.0016. The recently measured standard deviation is close enough to the advertised standard deviation that we cannot say that it is different from 4%, at a 5% level of significance.

Comparing Two Population Variances

The hypotheses concerned with the equality of the variances of two populations are tested with an *F*-distributed test statistic. Hypothesis testing using a test statistic that follows an *F*-distribution is referred to as the *F*-test. The *F*-test is used under the assumption that the populations from which samples are drawn are normally distributed, and that the samples are independent.

If we let σ_1^2 and σ_2^2 represent the variances of normal Population 1 and Population 2, respectively, the hypotheses for the two-tailed *F*-test of differences in the variances can be structured as follows:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ versus } H_a: \sigma_1^2 \neq \sigma_2^2$$

The one-sided test structures can be specified as follows:

$$H_0: \sigma_1^2 \leq \sigma_2^2 \text{ versus } H_a: \sigma_1^2 > \sigma_2^2, \text{ or } H_0: \sigma_1^2 \geq \sigma_2^2 \text{ versus } H_a: \sigma_1^2 < \sigma_2^2$$

The test statistic for the *F*-test is the ratio of the sample variances. The *F*-statistic is computed as follows:

$$F = \frac{s_1^2}{s_2^2}$$

where:

S_1^2 = variance of the sample of n_1 observations drawn from Population 1

S_2^2 = variance of the sample of n_2 observations drawn from Population 2

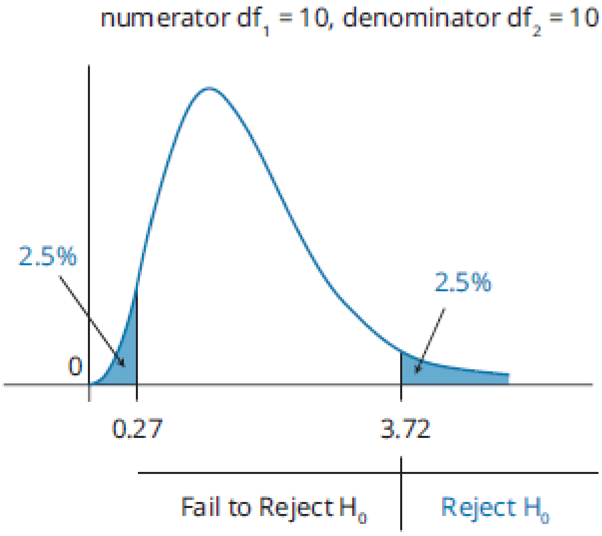
Note that $n_1 - 1$ and $n_2 - 1$ are the degrees of freedom used to identify the appropriate critical value from the *F*-table (provided in this book's Appendix).

PROFESSOR'S NOTE

Always put the larger variance in the numerator (s_1^2). Following this convention means that we only have to consider the critical value for the right-hand tail.

An ***F*-distribution** is presented in Figure 8.6. As indicated, the *F*-distribution is right skewed and is bounded by zero on the left-hand side. The shape of the *F*-distribution is determined by *two separate degrees of freedom*: the numerator degrees of freedom, df_1, and the denominator degrees of freedom, df_2.

Figure 8.6: *F*-Distribution



Note that when the sample variances are equal, the value of the test statistic is 1. The upper critical value is always greater than one (the numerator is significantly greater than the denominator), and the lower critical value is always less than one (the numerator is significantly smaller than the denominator). In fact, the lower critical value is the reciprocal of the upper critical value. For this reason, in practice, we put the larger sample variance in the numerator and consider only the upper critical value.

EXAMPLE: *F*-test for equal variances

Annie Cower is examining the earnings for two different industries. Cower suspects that the variance of earnings in the textile industry is different from the variance of earnings in the paper industry. To confirm this suspicion, Cower has looked at a sample of 31 textile manufacturers and a sample of 41 paper companies. She measured the sample standard deviation of earnings across the textile industry to be $4.30, and that of the paper industry companies to be $3.80. Cower calculates a test statistic of 1.2805. Using a 5% significance level, determine if the earnings of the textile industry have a different standard deviation than those of the paper industry.

Answer:

In this example, we are concerned with whether the variance of earnings for companies in the textile industry is equal to the variance of earnings for companies in the paper industry. As such, the test hypotheses can be appropriately structured as follows:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ versus } H_a: \sigma_1^2 \neq \sigma_2^2$$

For tests of difference between variances, the appropriate test statistic is:

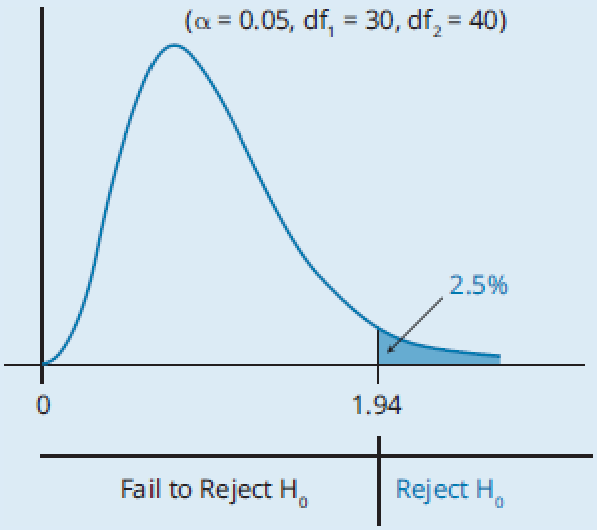
$$F = \frac{s_1^2}{s_2^2}$$

where s_1^2 is the larger sample variance.

Using the sample sizes for the two industries, the critical *F*-value for our test is found to be 1.94. This value is obtained from the table of the *F*-distribution for 2.5% in the upper tail, with $df_1 = 30$ and $df_2 = 40$. Thus, if the computed *F*-statistic is greater than the critical value of 1.94, the null hypothesis is rejected. The decision rule, illustrated in the following figure, can be stated as follows:

Reject H_0 if $F > 1.94$

Decision Rule for *F*-Test



Because the calculated *F*-statistic of 1.2805 is less than the critical *F*-statistic of 1.94, Cower cannot reject the null hypothesis. Cower should conclude that the earnings variances of the industries are not significantly different from one another at a 5% level of significance.

LOS 8.c: Compare and contrast parametric and nonparametric tests, and describe situations where each is the more appropriate type of test.

Parametric tests rely on assumptions regarding the distribution of the population and are specific to population parameters. For example, the *z*-test relies upon a mean and a standard deviation to define the normal distribution. The *z*-test also requires that either the sample is large, relying on the central limit theorem to assure a normal sampling distribution, or that the population is normally distributed.

Nonparametric tests either do not consider a particular population parameter or have few assumptions about the population that is sampled. Nonparametric tests are used when there is concern about quantities other than the parameters of a distribution, or when the assumptions of parametric tests can't be supported. They are also used when the data are not suitable for parametric tests (e.g., ranked observations).

Some situations where a nonparametric test is called for may include the following:

1. The assumptions about the distribution of the random variable that support a parametric test are not met. An example would be a hypothesis test of the mean

value for a variable that comes from a distribution that is not normal and is of small size, so that neither the *t*-test nor the *z*-test is appropriate.

2. A nonparametric test is called for when data are ranks (an ordinal measurement scale) rather than values.
3. The hypothesis does not involve the parameters of the distribution, such as testing whether a variable is normally distributed. We can use a nonparametric test, called a runs test, to determine whether data are random. A runs test provides an estimate of the probability that a series of changes (e.g., +, +, −, −, +, −,....) are random.

MODULE QUIZ 8.2

1. Which of the following assumptions is *least likely* required for the difference in means test based on two samples?
   A. The two samples are independent.
   B. The two populations are normally distributed.
   C. The two populations have known variances.
2. The appropriate test statistic for a test of the equality of variances for two normally distributed random variables, based on two independent random samples, is the:
   A. *t*-test.
   B. *F*-test.
   C. χ^2 test.
3. The appropriate test statistic to test the hypothesis that the variance of a normally distributed population is equal to 13 is the:
   A. *t*-test.
   B. *F*-test.
   C. χ^2 test.

KEY CONCEPTS

LOS 8.a

The hypothesis testing process requires a statement of a null and an alternative hypothesis, the selection of the appropriate test statistic, specification of the significance level, a decision rule, the calculation of a sample statistic, a decision regarding the hypotheses based on the test, and a decision based on the test results.

The null hypothesis is what the researcher wants to reject. The alternative hypothesis is what the researcher wants to support, and it is accepted when the null hypothesis is rejected.

A Type I error is the rejection of the null hypothesis when it is actually true, while a Type II error is the failure to reject the null hypothesis when it is actually false.

The significance level can be interpreted as the probability that a test statistic will reject the null hypothesis by chance when it is actually true (i.e., the probability of a Type I error). A significance level must be specified to select the critical values for the test.

The power of a test is the probability of rejecting the null when it is false. The power of a test = 1 − P(Type II error).

The *p*-value for a hypothesis test is the smallest significance level for which the hypothesis would be rejected.

LOS 8.b

Hypothesis tests of:	Use a:	With degrees of freedom:
One population mean	*t*-statistic	n − 1
Two population means	*t*-statistic	n − 1
One population variance	Chi-square statistic	n − 1
Two population variances	*F*-statistic	$n_1 - 1, n_2 - 1$

LOS 8.c

Parametric tests, like the *t*-test, *F*-test, and chi-square test, make assumptions regarding the distribution of the population from which samples are drawn. Nonparametric tests either do not consider a particular population parameter or have few assumptions about the sampled population. Nonparametric tests are used when the assumptions of parametric tests can't be supported, or when the data are not suitable for parametric tests.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 8.1

1. **C** A Type I error is rejecting the null hypothesis when it is true. The probability of rejecting a false null is [1 − Prob Type II] = [1 − 0.60] = 40%, which is called the power of the test. The other answer choices are not necessarily true, because the null may be false and the probability of rejection unknown. (LOS 8.a)
2. **A** The power of a test is 1 − P(Type II error) = 1 − 0.15 = 0.85. (LOS 8.a)

Module Quiz 8.2

1. **C** The difference in means test does not require the two population variances to be known. (LOS 8.b)
2. **B** The *F*-test is the appropriate test. (LOS 8.b)
3. **C** A test of the population variance is a chi-square test. (LOS 8.b)

READING 9

PARAMETRIC AND NON-PARAMETRIC TESTS OF INDEPENDENCE

MODULE 9.1: TESTS FOR INDEPENDENCE

Video covering this content is available online.

LOS 9.a: Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Correlation measures the strength of the relationship between two variables. If the correlation between two variables is zero, there is no linear relationship between them. When the sample correlation coefficient for two variables is different from zero, we must address the question of whether the true population correlation coefficient (ρ) is equal to zero. The appropriate test statistic for the hypothesis that the population correlation equals zero, when the two variables are normally distributed, is as follows:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where:
r = sample correlation
n = sample size

This test statistic follows a *t*-distribution with n − 2 degrees of freedom. Note that the test statistic increases, not only with the sample correlation coefficient, but also with sample size.

EXAMPLE: **Test of the hypothesis that the population correlation coefficient equals zero**

A researcher computes the sample correlation coefficient for two normally distributed random variables as 0.35, based on a sample size of 42. Determine whether to reject the hypothesis that the population correlation coefficient is equal to zero at a 5% significance level.

Answer:

Our test statistic is $\frac{0.35\sqrt{42-2}}{\sqrt{1-0.35^2}} = 2.363.$

Using the *t*-table with 42 − 2 = 40 degrees of freedom for a two-tailed test and a significance level of 5%, we can find the critical value of 2.021. Because our computed test statistic of 2.363 is greater than 2.021, we reject the hypothesis that the population mean is zero and conclude that it is not equal to zero. That is, the two populations are correlated—in this case, positively.

PROFESSOR'S NOTE

The correlation coefficient we refer to here is the Pearson correlation coefficient, which is a measure of the linear relationship between two variables. There are other correlation coefficients that better measure the strength of any nonlinear relationship between two variables.

The **Spearman rank correlation test**, a nonparametric test, can be used to test whether two sets of ranks are correlated. Ranks are simply ordered values. If there is a tie (equal values), the ranks are shared—so if second and third rank is the same, the ranks are shared, and each gets a rank of (2 + 3) / 2 = 2.5.

The Spearman rank correlation, r_s (when all ranks are integer values), is calculated as follows:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where:

r_S = rank correlation

n = sample size

d_i = difference between two ranks

We can test the significance of the Spearman rank correlation calculated with the formula just listed using the same test statistic we used for estimating the significance of a parametric correlation coefficient:

$$\frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

When the sample size is greater than 30, the test statistic follows a *t*-distribution with n − 2 degrees of freedom.

LOS 9.b: Explain tests of independence based on contingency table data.

A contingency or two-way table shows the number of observations from a sample that have a combination of two characteristics. Figure 9.1 is a contingency table where the characteristics are earnings growth (low, medium, or high) and dividend yield (low, medium, or high). We can use the data in the table to test the hypothesis that the two characteristics, earnings growth and dividend yield, are independent of each other.

Figure 9.1: Contingency Table for Categorical Data

Earnings Growth	Dividend Yield			
	Low	Medium	High	Total
Low	28	53	42	123
Medium	42	32	39	113
High	49	25	14	88
Total	119	110	95	324

We index our three categories of earnings growth from low to high with $i = 1, 2$, or 3, and our three categories of dividend yield from low to high with $j = 1, 2$, or 3. From the table, we see in Cell 1,1 that 28 firms have both low earnings growth and low dividend yield. We see in Cell 3,2 that 25 firms have high earnings growth and medium dividend yields.

For our test, we are going to compare the actual table values to what the values would be if the two characteristics were independent. The test statistic is a chi-square test statistic calculated as follows:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where:

O_{ij} = the number of observations in Cell i,j: Row i and Column j (i.e., observed frequency)

E_{ij} = the expected number of observations for Cell i,j

r = the number of row categories

c = the number of column categories

The degrees of freedom are $(r - 1) \times (c - 1)$, which is 4 in our example for dividend yield and earnings growth.

E_{ij}, the expected number of observations in Cell i,j, is:

$$\frac{\text{total for Row i} \times \text{total for Column j}}{\text{total for all columns and rows}}$$

The expected number of observations for Cell 2,2 is:

$$\frac{110 \times 113}{324} = 38.4$$

In calculating our test statistic, the term for Cell 2,2 is:

$$\frac{(32-38.4)^2}{38.4} = 1.0667$$

Figure 9.2 shows the expected frequencies for each pair of categories in our earnings growth and dividend yield contingency table.

Figure 9.2: Contingency Table for Expected Frequencies

Earnings Growth	Dividend Yield		
	Low	Medium	High
Low	45.2	41.8	36.1
Medium	41.5	38.4	33.1
High	32.3	29.9	25.8

For our test statistic, we sum, for all nine cells, the squared difference between the expected frequency and observed frequency, divided by the expected frequency. The resulting sum is 27.47. Figure 9.3 shows the results for each cell in calculating the test statistic.

Figure 9.3: Squared Differences from Contingency Table

Earnings Growth	Dividend Yield		
	Low	Medium	High
Low	6.5451	3.0010	0.9643
Medium	0.0060	1.0667	1.0517
High	8.6344	0.8030	5.3969
	Sum = 27.4691		

Our degrees of freedom are (3 − 1) × (3 − 1) = 4. The critical value for a significance level of 5% (from the chi-square table in the Appendix) with 4 degrees of freedom is 9.488. Based on our sample data, we can reject the hypothesis that the earnings growth and dividend yield categories are independent.

MODULE QUIZ 9.1

1. The test statistic for a Spearman rank correlation test for a sample size greater than 30 follows a:
   A. *t*-distribution.
   B. normal distribution.
   C. chi-square distribution.
2. A contingency table can be used to test:
   A. a null hypothesis that rank correlations are equal to zero.
   B. whether multiple characteristics of a population are independent.
   C. the number of *p*-values from multiple tests that are less than adjusted critical values.
3. For a parametric test of whether a correlation coefficient is equal to zero, it is *least likely* that:
   A. degrees of freedom are n − 1.
   B. the test statistic follows a *t*-distribution.
   C. the test statistic increases with a greater sample size.

KEY CONCEPTS

LOS 9.a

To test a hypothesis that a population correlation coefficient equals zero, the appropriate test statistic is a *t*-statistic with n − 2 degrees of freedom, calculated as $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, where *r* is the sample correlation coefficient.

A nonparametric test of correlation can be performed when we have only ranks (e.g., deciles of investment performance). The Spearman rank correlation test examines whether the ranks for multiple periods are correlated. The rank correlation is $r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$, where d_i^2 is the sum of the squared differences in pairs of ranks and *n* is the number of sample periods. The test statistic follows a *t*-distribution for samples sizes greater than 30.

LOS 9.b

A contingency table can be used to test the hypothesis that two characteristics (categories) of a sample of items are independent. A contingency table shows the number of the sample items (e.g., firms that have both of two characteristics). The test statistic follows a chi-square distribution and is calculated as follows:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where:

O_{ij} = the number of observations in Cell i,j: Row *i* and Column *j* (i.e., observed frequency)

E_{ij} = the expected number of observations for Cell i,j of the contingency table with independence

r = the number of row categories

c = the number of column categories

The degrees of freedom are (r − 1) × (c − 1). If the test statistic is greater than the critical chi-square value for a given level of significance, we reject the hypothesis that the two characteristics are independent.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 9.1

1. **A** The test statistic for a Spearman rank correlation test follows a *t*-distribution. (LOS 9.a)

2. **B** A contingency table is used to determine whether two characteristics of a group are independent. (LOS 9.b)

3. **A** Degrees of freedom are n − 2 for a test of the hypothesis that correlation is equal to zero. The test statistic increases with sample size (degrees of freedom increase) and follows a *t*-distribution. (LOS 9.a)

READING 10

SIMPLE LINEAR REGRESSION

MODULE 10.1: LINEAR REGRESSION BASICS

Video covering this content is available online.

LOS 10.a: Describe a simple linear regression model, how the least squares criterion is used to estimate regression coefficients, and the interpretation of these coefficients.

The purpose of **simple linear regression** is to explain the variation in a dependent variable in terms of the variation in a single independent variable. Here, the term *variation* is interpreted as the degree to which a variable differs from its mean value. Don't confuse *variation* with *variance*—they are related, but they are not the same.

$$\text{variation in Y} = \sum_{1}^{n}(Y_i - \overline{Y})^2$$

- The **dependent variable** is the variable whose variation is explained by the independent variable. We are interested in answering the question, "What explains fluctuations in the dependent variable?" The dependent variable is also referred to as the terms *explained variable, endogenous variable,* or *predicted variable*.
- The **independent variable** is the variable used to explain the variation of the dependent variable. The independent variable is also referred to as the terms *explanatory variable, exogenous variable,* or *predicting variable*.

EXAMPLE: **Dependent vs. independent variables**

Suppose you want to predict stock returns with GDP growth. Which variable is the independent variable?

Answer:

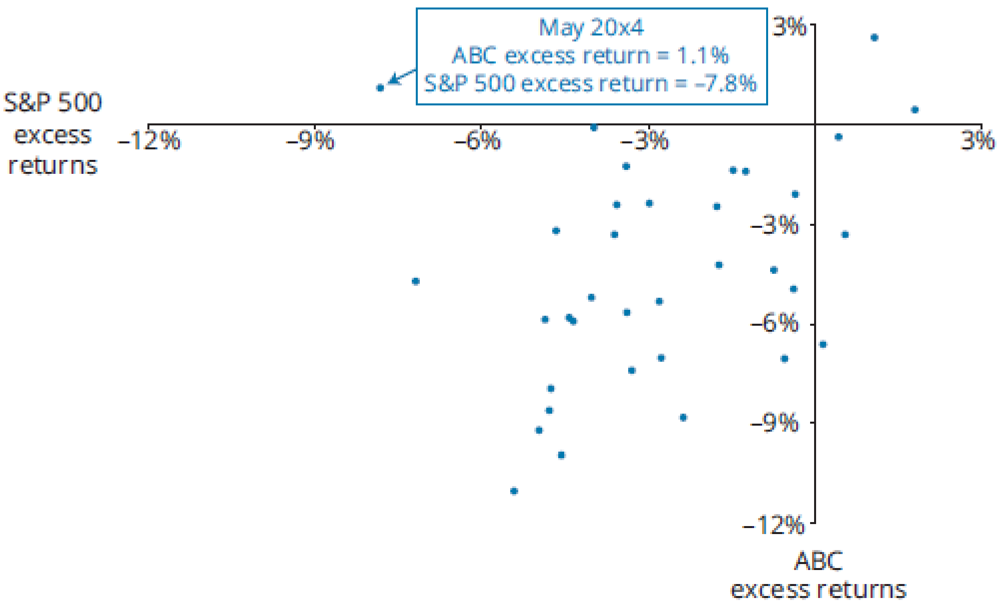
Because GDP is going to be used as a predictor of stock returns, stock returns are being *explained* by GDP. Hence, stock returns are the dependent (explained) variable, and GDP is the independent (explanatory) variable.

Suppose we want to use excess returns on the S&P 500 (the independent variable) to explain the variation in excess returns on ABC common stock (the dependent variable).

For this model, we define excess return as the difference between the actual return and the return on 1-month Treasury bills.

We would start by creating a scatter plot with ABC excess returns on the vertical axis and S&P 500 excess returns on the horizontal axis. Monthly excess returns for both variables from June 20X2 to May 20X5 are plotted in Figure 10.1. For example, look at the point labeled May 20X4. In that month, the excess return on the S&P 500 was −7.8%, and the excess return on ABC was 1.1%.

Figure 10.1: Scatter Plot of ABC Excess Returns vs. S&P 500 Index Excess Returns



The two variables in Figure 10.1 appear to be positively correlated: excess ABC returns tended to be positive (negative) in the same month that S&P 500 excess returns were positive (negative). This is not the case for all the observations, however (for example, May 20X4). In fact, the correlation between these variables is approximately 0.40.

Simple Linear Regression Model

The following linear regression model is used to describe the relationship between two variables, *X* and *Y*:

$$Y_i = b_0 + b_1X_i + \varepsilon_i, \ldots i = 1, \ldots, n$$

where:

Y_i = *i*th observation of the dependent variable, *Y*
X_i = *i*th observation of the independent variable, *X*
b_0 = regression intercept term
b_1 = regression slope coefficient
ε_i = **residual** for the *i*th observation (also referred to as the disturbance term or error term)

Based on this regression model, the regression process estimates an equation for a line through a scatter plot of the data that "best" explains the observed values for *Y* in terms of the observed values for *X*.

The linear equation, often called the line of best fit or **regression line**, takes the following form:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i, \ i = 1,2,3\ldots,n$$

where:

$\hat{Y}_i$ = estimated value of Y_i given X_i

$\hat{b}_0$ = estimated intercept term

$\hat{b}_1$ = estimated slope coefficient

PROFESSOR'S NOTE

The hat "^" above a variable or parameter indicates a predicted value.

The regression line is just one of the many possible lines that can be drawn through the scatter plot of *X* and *Y*. The criteria used to estimate this line is the essence of linear regression. The regression line is the line that minimizes the sum of the squared differences (vertical distances) between the *Y*-values predicted by the regression equation $(\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i)$ and the actual *Y*-values, Y_i. The sum of the squared vertical distances between the estimated and actual *Y*-values is referred to as the **sum of squared errors (SSE)**.

Thus, the regression line is the line that minimizes the SSE. This explains why simple linear regression is frequently referred to as **ordinary least squares (OLS)** regression, and the values determined by the estimated regression equation, $\hat{Y}_i$, are called least squares estimates.

The estimated **slope coefficient** $(\hat{b}_1)$ for the regression line describes the change in *Y* for a one-unit change in *X*. It can be positive, negative, or zero, depending on the relationship between the regression variables. The slope term is calculated as follows:

$$\hat{b}_1 = \frac{Cov_{XY}}{\sigma_X^2}$$

The intercept term $(\hat{b}_0)$ is the line's intersection with the *Y*-axis at X = 0. It can be positive, negative, or zero. A property of the least squares method is that the intercept term may be expressed as follows:

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}$$

where:

$\overline{Y}$ = mean of Y

$\overline{X}$ = mean of X

The intercept equation highlights the fact that the regression line passes through a point with coordinates equal to the mean of the independent and dependent variables (i.e., the point X, Y).

EXAMPLE: Computing the slope coefficient and intercept term

Compute the slope coefficient and intercept term using the following information:

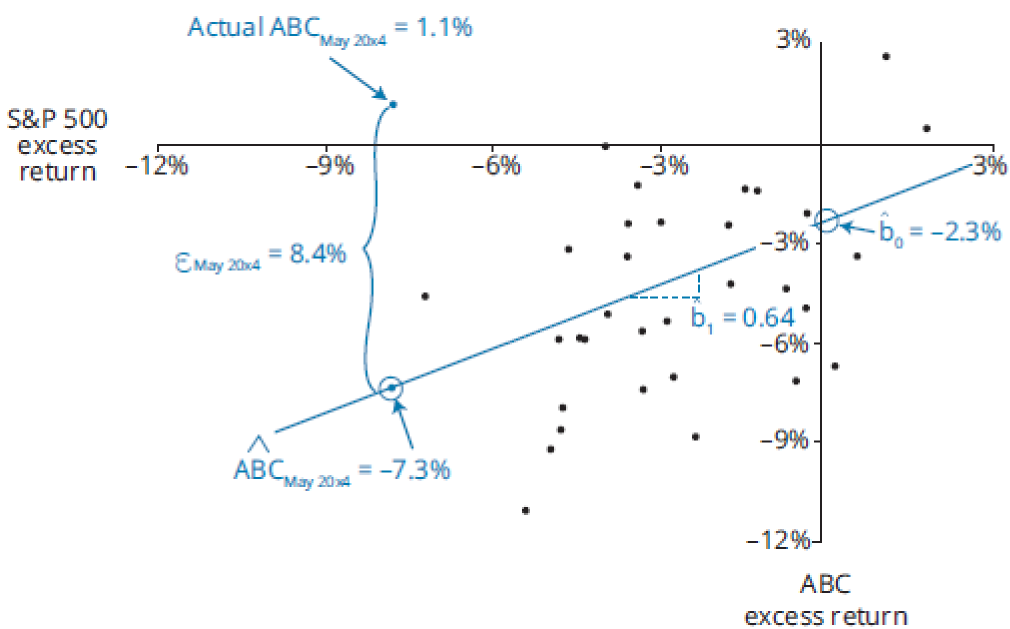
Cov(S&P 500, ABC)	= 0.000336	Mean return, S&P 500	= −2.70%
Var(S&P 500)	= 0.000522	Mean return, ABC	= −4.05%

> **Answer:**
>
> The slope coefficient is calculated as $\hat{b}_1 = 0.000336 / 0.000522 = 0.64$.
>
> The intercept term is calculated as follows:
>
> $$\bar{b}_0 = \overline{ABC} - \hat{b}_1 \overline{S\&P\ 500} = -4.05\% - 0.64(-2.70\%) = -2.3\%$$

The estimated regression line that minimizes the SSE in our ABC stock return example is shown in Figure 10.2.

This regression line has an intercept of −2.3% and a slope of 0.64. The model predicts that if the S&P 500 excess return is −7.8% (May 20X4 value), then the ABC excess return would be −2.3% + (0.64)(−7.8%) = −7.3%. The residual (error) for the May 20X4 ABC prediction is 8.4%—the difference between the actual ABC excess return of 1.1% and the predicted return of −7.3%.

Figure 10.2: Estimated Regression Equation for ABC vs. S&P 500 Excess Returns



Interpreting a Regression Coefficient

The estimated intercept represents the value of the dependent variable at the point of intersection of the regression line and the axis of the dependent variable (usually, the vertical axis). In other words, the intercept is an estimate of the dependent variable when the independent variable is zero.

We also mentioned earlier that the estimated slope coefficient is interpreted as the expected change in the dependent variable for a one-unit change in the independent variable. For example, an estimated slope coefficient of 2 would indicate that the dependent variable is expected to change by two units for every one-unit change in the independent variable.

EXAMPLE: Interpreting regression coefficients

In the previous example, the estimated slope coefficient was 0.64 and the estimated intercept term was −2.3%. Interpret each coefficient estimate.

Answer:

The slope coefficient of 0.64 can be interpreted to mean that when excess S&P 500 returns increase (decrease) by 1%, ABC excess returns is expected to increase (decrease) by 0.64%.

The intercept term of −2.3% can be interpreted to mean that when the excess return on the S&P 500 is zero, the expected return on ABC stock is −2.3%.

PROFESSOR'S NOTE

The slope coefficient in a regression of the excess returns of an individual security (the *y*-variable) on the return on the market (the *x*-variable) is called the stock's beta, which is an estimate of systematic risk of ABC stock. Notice that ABC is less risky than the average stock, because its returns tend to increase or decrease by less than the overall change in the market returns. A stock with a beta (regression slope coefficient) of 1 has an average level of systematic risk, and a stock with a beta greater than 1 has more-than-average systematic risk. We will apply this concept in the Portfolio Management topic area.

Keep in mind, however, that any conclusions regarding the importance of an independent variable in explaining a dependent variable are based on the statistical significance of the slope coefficient. The magnitude of the slope coefficient tells us nothing about the strength of the linear relationship between the dependent and independent variables. A hypothesis test must be conducted, or a confidence interval must be formed, to assess the explanatory power of the independent variable. Later in this reading we will perform these hypothesis tests.

LOS 10.b: Explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated.

Linear regression is based on numerous assumptions. Most of the major assumptions pertain to the regression model's residual term (ε). Linear regression assumes the following:

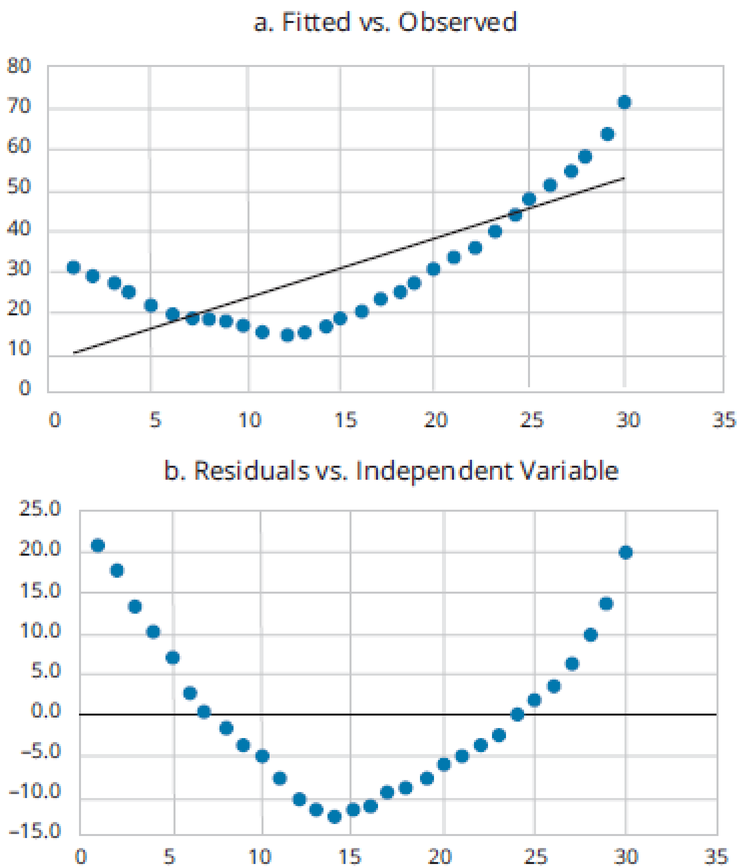
1. A linear relationship exists between the dependent and the independent variables.
2. The variance of the residual term is constant for all observations (homoskedasticity).
3. The residual term is independently distributed; that is, the residual for one observation is not correlated with that of another observation (or, the paired *x* and *y* observations are independent of each other).

4. The residual term is normally distributed.

Linear Relationship

A linear regression model is not appropriate when the underlying relationship between X and Y is nonlinear. In Panel A of Figure 10.3, we illustrate a regression line fitted to a nonlinear relationship. Note that the prediction errors (vertical distances from the dots to the line) are positive for low values of X, then increasingly negative for higher values of X, and then turning positive for still-greater values of X. One way of checking for linearity is to examine the model residuals (prediction errors) in relation to the independent regression variable. In Panel B, we show the pattern of residuals over the range of the independent variable: positive, negative, then positive.

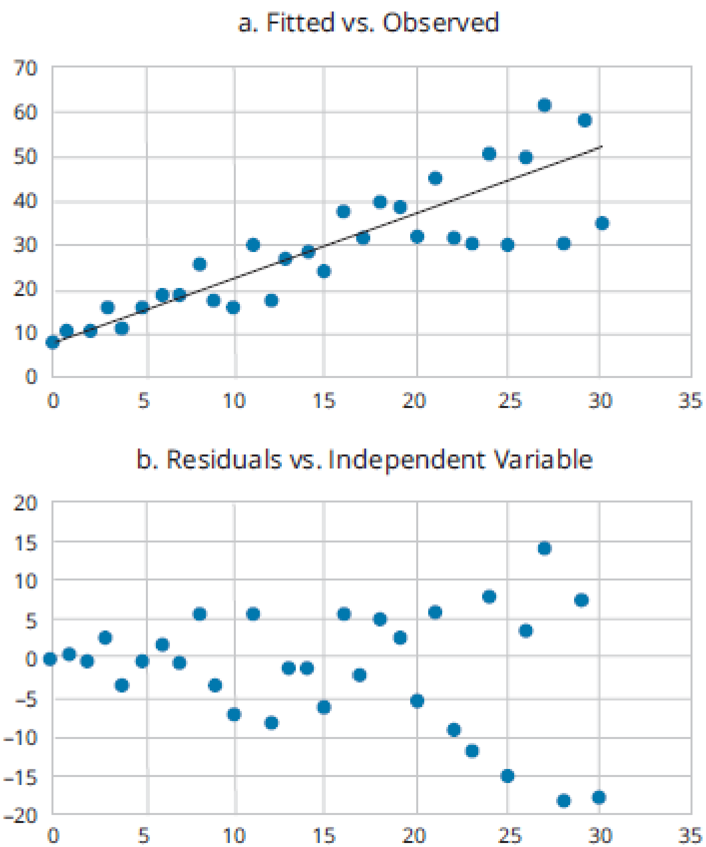
Figure 10.3: Nonlinear Relationship



Homoskedasticity

Homoskedasticity refers to the case where prediction errors all have the same variance. **Heteroskedasticity** refers to the situation when the assumption of homoskedasticity is violated. Figure 10.4, Panel A shows a scatter plot of observations around a fitted regression line where the residuals (prediction errors) increase in magnitude with larger values of the independent variable X. Panel B shows the residuals plotted versus the value of the independent variable, and it also illustrates that the variance of the error terms is not likely constant for all observations.

Figure 10.4: Heteroskedasticity

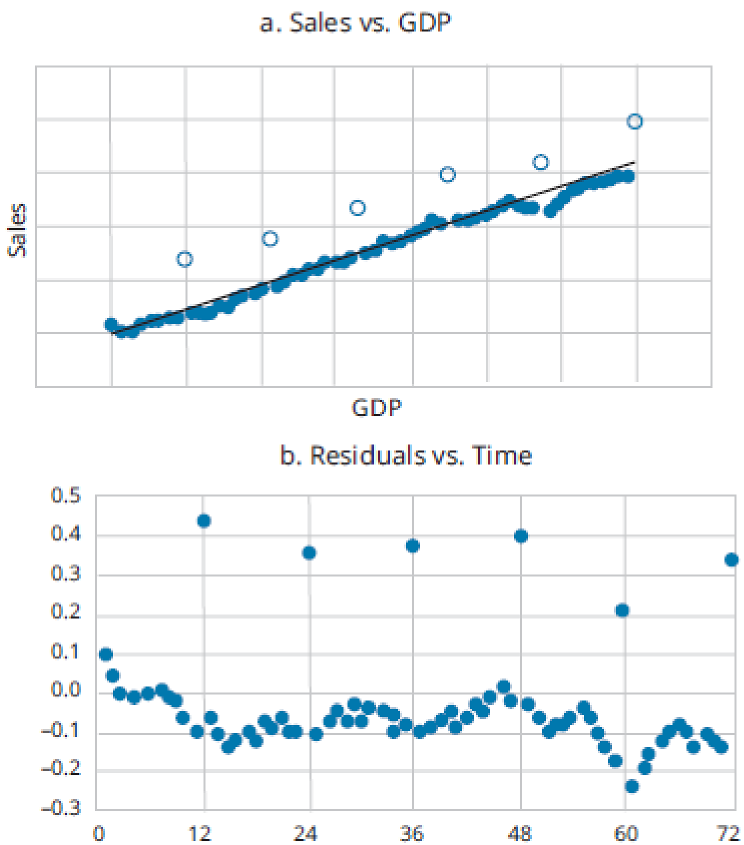


Another type of heteroskedasticity results if the variance of the error term changes over time (rather than with the magnitude of the independent variable). We could observe this by plotting the residuals from a linear regression model versus the dates of each observation and finding that the magnitude of the errors exhibits a pattern of changing over time. To illustrate this, we could plot the residuals versus a time index (as the *x*-variable). Residuals would exhibit a pattern of increasing over time.

Independence

Suppose we collect a company's monthly sales and plot them against monthly GDP as in Figure 10.5, Panel A, and observe that some prediction errors (the unfilled dots) are noticeably larger than others. To investigate this, we plot the residuals versus time, as in Panel B. The residuals plot illustrates that there are large negative prediction errors every 12 months (in December). This suggests that there is seasonality in sales such that December sales (the unfilled dots in Figure 10.5) are noticeably farther from their predicted values than sales for the other months. If the relationship between X and Y is not independent, the residuals are not independent, and our estimates of the model parameters' variances will not be correct.

Figure 10.5: Independence



Normality

When the residuals (prediction errors) are normally distributed, we can conduct hypothesis testing for evaluating the goodness of fit of the model (discussed later). With a large sample size, based on the central limit theorem, our parameter estimates may be valid, even when the residuals are not normally distributed.

Outliers are observations (one or a few) that are far from our regression line (have large prediction errors or *X* values that are far from the others). Outliers will influence our parameter estimates so that the OLS model will not fit the other observations well.

MODULE QUIZ 10.1

1. What is the *most appropriate* interpretation of a slope coefficient estimate equal to 10.0?
   A. The predicted value of the dependent variable when the independent variable is 0 is 10.0.
   B. For every 1-unit change in the independent variable, the model predicts that the dependent variable will change by 10 units.
   C. For every 1-unit change in the independent variable, the model predicts that the dependent variable will change by 0.1 units.
2. Which of the following is *least likely* a necessary assumption of simple linear regression analysis?
   A. The residuals are normally distributed.
   B. There is a constant variance of the error term.

C. The dependent variable is uncorrelated with the residuals.

MODULE 10.2: ANALYSIS OF VARIANCE (ANOVA) AND GOODNESS OF FIT

Video covering this content is available online.

LOS 10.c: Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

LOS 10.d: Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Analysis of variance (ANOVA) is a statistical procedure for analyzing the total variability of the dependent variable. Let's define some terms before we move on to ANOVA tables:

- The **total sum of squares (SST)** measures the total variation in the dependent variable. SST is equal to the sum of the squared differences between the actual *Y*-values and the mean of *Y*:

  $$SST = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- The **sum of squares regression (SSR)** measures the variation in the dependent variable that is explained by the independent variable. SSR is the sum of the squared distances between the predicted *Y*-values and the mean of *Y*:

  $$SSR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$$

- The **mean square regression (MSR)** is the SSR divided by the number of independent variables. A simple linear regression has only one independent variable, so in this case, MSR = SSR.

PROFESSOR'S NOTE
Multiple regression (i.e., with more than one independent variable) is addressed in the Level II CFA curriculum.

- The **sum of squared errors (SSE)** measures the unexplained variation in the dependent variable. It's also known as the sum of squared residuals or the residual sum of squares. SSE is the sum of the squared vertical distances between the actual *Y*-values and the predicted *Y*-values on the regression line:

  $$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y})^2$$

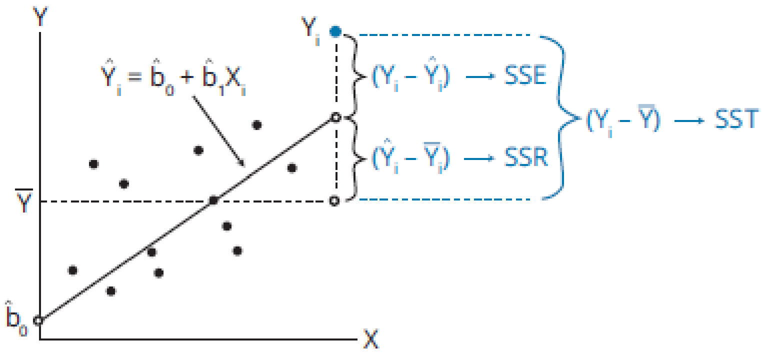
- The **mean squared error (MSE)** is the SSE divided by the degrees of freedom, which is $n - 1$ minus the number of independent variables. A simple linear regression has only one independent variable, so in this case, degrees of freedom are $n - 2$.

You probably will not be surprised to learn the following:

total variation = explained variation + unexplained variation
or:
SST = SSR + SSE

Figure 10.6 illustrates how the total variation in the dependent variable (SST) is composed of SSR and SSE.

Figure 10.6: Components of Total Variation



The output of the ANOVA procedure is an ANOVA table, which is a summary of the variation in the dependent variable. ANOVA tables are included in the regression output of many statistical software packages. You can think of the ANOVA table as the source of the data for the computation of many of the regression concepts discussed in this reading. A generic ANOVA table for a simple linear regression (one independent variable) is presented in Figure 10.7.

Figure 10.7: ANOVA Table for a Simple Linear Regression

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares
Regression (explained)	1	SSR	$\text{MSR} = \frac{\text{SSR}}{k} = \frac{\text{SSR}}{1} = \text{SSR}$
Error (unexplained)	$n - 2$	SSE	$\text{MSE} = \frac{\text{SSE}}{n-2}$
Total	$n - 1$	SST	

Standard Error of Estimate (SEE)

The SEE for a regression is the standard deviation of its residuals. The lower the SEE, the better the model fit:

$$\text{SEE} = \sqrt{\text{MSE}}$$

Coefficient of Determination (R^2)

The **coefficient of determination** (R^2) is defined as the percentage of the total variation in the dependent variable explained by the independent variable. For example, an R^2 of 0.63 indicates that the variation of the independent variable explains 63% of the variation in the dependent variable:

$$R^2 = \text{SSR} / \text{SST}$$

PROFESSOR'S NOTE

For simple linear regression (i.e., with one independent variable), the coefficient of determination, R^2, may be computed by simply squaring the correlation coefficient, *r*. In other words, $R^2 = r^2$ for a regression with one independent variable.

EXAMPLE: Using the ANOVA table

Given the following ANOVA table based on 36 observations, calculate the R^2 and the standard error of estimate (SEE).

Completed ANOVA table for ABC regression

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares
Regression (explained)	1	0.0076	0.0076
Error (unexplained)	34	0.0406	0.0012
Total	35	0.0482	

Answer:

$$R^2 = \frac{\text{explained variation (SSR)}}{\text{total variation (SST)}} = \frac{0.0076}{0.0482} = 0.158 \text{ or } 15.8\%$$

$$\text{SEE} = \sqrt{\text{MSE}} = \sqrt{0.0012} = 0.035$$

The *F*-Statistic

An *F*-test assesses how well a set of independent variables, as a group, explains the variation in the dependent variable.

The *F*-statistic is calculated as follows:

$$F = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/n-k-1}$$

where:

MSR = mean regression sum of squares
MSE = mean squared error
Important: This is always a one-tailed test!

For simple linear regression, there is only one independent variable, so the *F*-test is equivalent to a *t*-test of the statistical significance of the slope coefficient:

$$H_0\text{: } b_1 = 0 \text{ versus } H_a\text{: } b_1 \neq 0$$

To determine whether b_1 is statistically significant using the *F*-test, the calculated *F*-statistic is compared with the critical *F*-value, F_c, at the appropriate level of significance. The degrees of freedom for the numerator and denominator with one independent variable are as follows:

$$df_{numerator} = k = 1$$
$$df_{denominator} = n - k - 1 = n - 2$$

where:

n = number of observations

The decision rule for the *F*-test is to reject H_0 if $F > F_c$.

Rejecting the null hypothesis that the value of the slope coefficient equals zero at a stated level of significance indicates that the independent variable and the dependent variable have a significant linear relationship.

EXAMPLE: Calculating and interpreting the *F*-statistic

Use the ANOVA table from the previous example to calculate and interpret the *F*-statistic. Test the null hypothesis at the 5% significance level that the slope coefficient is equal to 0.

Answer:

$$F = \frac{MSR}{MSE} = \frac{0.0076}{0.0012} = 6.33$$
$$df_{numerator} = k = 1$$
$$df_{denominator} = n - k - 1 = 36 - 1 - 1 = 34$$

The null and alternative hypotheses are h_0: $b_1 = 0$ versus h_a: $b_1 \neq 0$. The critical *F*-value for 1 and 34 degrees of freedom at a 5% significance level is approximately 4.1. (Remember, it's a one-tailed test, so we use the 5% *F*-table.) Therefore, we can reject the null hypothesis and conclude that the slope coefficient is significantly different than zero.

Hypothesis Test of a Regression Coefficient

A *t*-test may also be used to test the hypothesis that the true slope coefficient, b_1, is equal to a hypothesized value. Letting $\hat{b}_1$ be the point estimate for b_1, the appropriate test statistic with n − 2 degrees of freedom is:

$$t_{b_1} = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}}$$

The decision rule for tests of significance for regression coefficients is:

$$\text{Reject } H_0 \text{ if } t > +t_{critical} \text{ or } t < -t_{critical}$$

Rejection of the null supports the alternative hypothesis that the slope coefficient is *different* from the hypothesized value of b_1. To test whether an independent variable explains the variation in the dependent variable (i.e., it is statistically significant), the null hypothesis is that the true slope is zero ($b_1 = 0$). The appropriate test structure for the null and alternative hypotheses is:

$$H_0: b_1 = 0 \text{ versus } H_a: b_1 \neq 0$$

EXAMPLE: Hypothesis test for significance of regression coefficients

The estimated slope coefficient from the ABC example is 0.64 with a standard error equal to 0.26. Assuming that the sample has 36 observations, determine if the estimated slope coefficient is significantly different than zero at a 5% level of significance.

Answer:

The calculated test statistic is:

$$t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{0.64 - 0}{0.26} = 2.46$$

The critical two-tailed *t*-values are ±2.03 (from the *t*-table with df = 36 − 2 = 34). Because $t > t_{critical}$ (i.e., 2.46 > 2.03), we reject the null hypothesis and conclude that the slope is different from zero.

Note that the *t*-test for a simple linear regression is equivalent to a *t*-test for the correlation coefficient between *x* and *y*:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

MODULE QUIZ 10.2

1. Consider the following statement: "In a simple linear regression, the appropriate degrees of freedom for the critical *t*-value used to calculate a confidence interval around both a parameter estimate and a predicted *Y*-value is the same as the number of observations minus two." This statement is:
   A. justified.
   B. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a parameter estimate is the number of observations.
   C. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a predicted *Y*-value is the number of observations.
2. What is the appropriate alternative hypothesis to test the statistical significance of the intercept term in the following regression?

   $Y = a_1 + a_2(X) + \varepsilon$

   A. H_A: $a_1 \neq 0$.
   B. H_A: $a_1 > 0$.
   C. H_A: $a_2 \neq 0$.
3. The variation in the dependent variable explained by the independent variable is measured by the:
   A. mean squared error.
   B. sum of squared errors.

C. regression sum of squares.

MODULE 10.3: PREDICTED VALUES AND FUNCTIONAL FORMS OF REGRESSION

Video covering this content is available online.

LOS 10.e: Calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable.

Predicted values are values of the dependent variable based on the estimated regression coefficients and a prediction about the value of the independent variable. They are the values that are *predicted* by the regression equation, given an estimate of the independent variable.

For a simple regression, this is the predicted (or forecast) value of *Y*:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_p$$

where:

$\hat{Y}$ = predicted value of the dependent variable

X_p = forecasted value of the independent variable

EXAMPLE: Predicting the dependent variable

Given the ABC regression equation as follows:

$$\widehat{ABC} = -2.3\% + (0.64)(\widehat{S\&P\ 500})$$

Calculate the predicted value of ABC excess returns if forecast S&P 500 excess returns are 10%.

Answer:

The predicted value for ABC excess returns is determined as follows:

$$\widehat{ABC} = -2.3\% + (0.64)(10\%) = 4.1\%$$

Confidence Intervals for Predicted Values

This is the equation for the confidence interval for a predicted value of *Y*:

$$\hat{Y} \pm (t_c \times s_f) \Rightarrow [\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)]$$

where:

t_c = two-tailed critical *t*-value at the desired level of significance with df = n − 2

s_f = standard error of the forecast

The challenge with computing a confidence interval for a predicted value is calculating s_f. On the Level I exam, it's highly unlikely that you will have to calculate the standard error of the forecast (it will probably be provided if you need to compute a confidence interval for the dependent variable). However, if you do need to calculate s_f, it can be done with the following formula for the variance of the forecast:

$$s_f^2 = SEE^2\left[1 + \frac{1}{n} + \frac{(X - \overline{X})^2}{(n-1)s_x^2}\right]$$

where:

SEE^2 = variance of the residuals = the square of the standard error of estimate

s_x^2 = variance of the independent variable

X = value of the independent variable for which the forecast was made

EXAMPLE: Confidence interval for a predicted value

Calculate a 95% prediction interval on the predicted value of ABC excess returns from the previous example. Assume the standard error of the forecast is 3.67, and the forecast value of S&P 500 excess returns is 10%.

Answer:

This is the predicted value for ABC excess returns:

$$\widehat{ABC} = -2.3\% + (0.64)(10\%) = 4.1\%$$

The 5% two-tailed critical *t*-value with 34 degrees of freedom is 2.03. This is the prediction interval at the 95% confidence level:

$$\widehat{ABC} \pm (t_c \times s_f) \Rightarrow [4.1\% \pm (2.03 \times 3.67\%)] = 4.1\% \pm 7.5\%$$

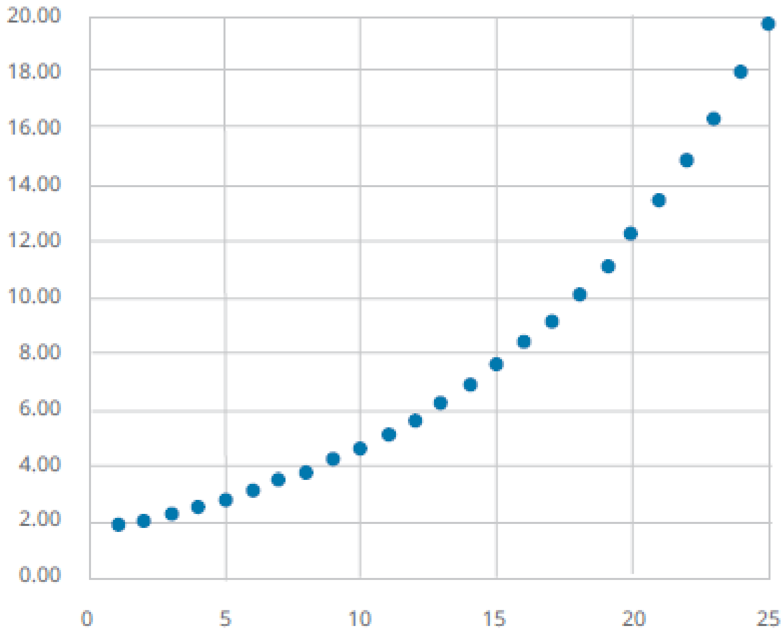
Or, −3.4% to 11.6%.

We can interpret this range to mean that, given a forecast value for S&P 500 excess returns of 10%, we can be 95% confident that the ABC excess returns will be between −3.4% and 11.6%.

LOS 10.f: Describe different functional forms of simple linear regressions.

One of the assumptions of linear regression is that the relationship between *X* and *Y* is linear. What if that assumption is violated? Consider *Y* = EPS for a company and *X* = time index. Suppose that EPS is growing at approximately 10% annually. Figure 10.8 shows the plot of actual EPS versus time.

Figure 10.8: Nonlinear Relationship



In such a situation, transforming one or both of the variables can produce a linear relationship. The appropriate transformation depends on the relationship between the two variables. One often-used transformation is to take the natural log of one or both of the variables. Here are some examples:

- **Log-lin model.** This is if the dependent variable is logarithmic, while the independent variable is linear.
- **Lin-log model.** This is if the dependent variable is linear, while the independent variable is logarithmic.
- **Log-log model.** Both the dependent variable and the independent variable are logarithmic.

Selecting the correct functional form involves determining the nature of the variables and evaluating the goodness-of-fit measures (e.g., R^2, SEE, *F*-stat).

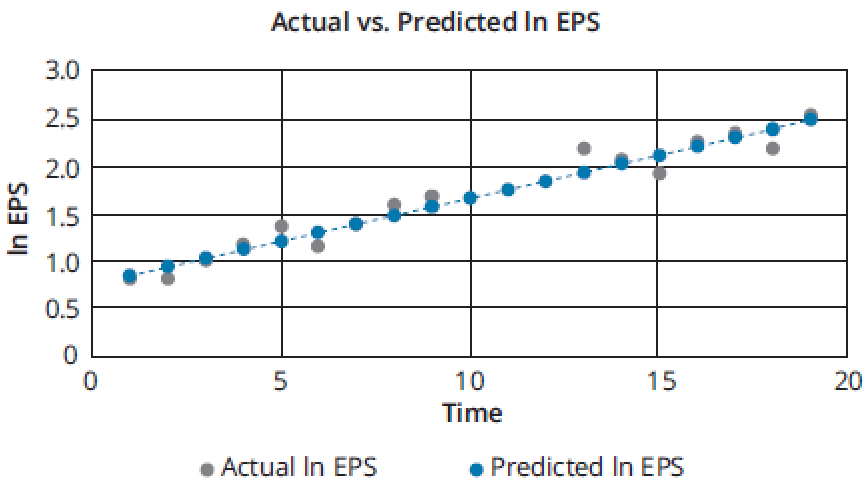
Log-Lin Model

Taking the natural logarithm of the dependent variable, our model now becomes this:

$$\ln Y_i = b_0 + b_1 X_i + \varepsilon_i$$

In this model, the slope coefficient is interpreted as the *relative* change in dependent variable for an absolute change in the independent variable. Figure 10.9 shows the results after taking the natural log of EPS, and fitting that data using a log-lin model.

Figure 10.9: Log-Lin Model, EPS Data



Lin-Log Model

Taking the natural logarithm of the independent variable, our model now becomes this:

$$Y_i = b_0 + b_1 \ln(X)_i + \varepsilon_i$$

In this model, the slope coefficient is interpreted as the *absolute* change in dependent variable for a *relative* change in the independent variable.

Log-Log Model

Taking the natural logarithm of both variables, our model now becomes this:

$$\ln Y_i = b_0 + b_1 \ln(X)_i + \varepsilon_i$$

In this model, the slope coefficient is interpreted as the relative change in dependent variable for a relative change in the independent variable.

MODULE QUIZ 10.3

1. For a regression model of Y = 5 + 3.5X, the analysis (based on a large data sample) provides the standard error of the forecast as 2.5 and the standard error of the slope coefficient as 0.8. A 90% confidence interval for the estimate of *Y* when the value of the independent variable is 10 is *closest* to:

   A. 35.1 to 44.9.
   B. 35.6 to 44.4.
   C. 35.9 to 44.1.

2. The appropriate regression model for a linear relationship between the relative change in an independent variable and the absolute change in the dependent variable is a:

   A. log-lin model.
   B. lin-log model.
   C. lin-lin model.

KEY CONCEPTS

LOS 10.a

Linear regression provides an estimate of the linear relationship between an independent variable (the explanatory variable) and a dependent variable (the predicted variable).

The general form of a simple linear regression model is as follows:

$$Y_i = b_0 + b_1X_i + \varepsilon_i$$

The least squares model minimizes the sum of squared errors:

- $\hat{b}_0$ = fitted intercept = $\overline{Y} - \hat{b}_1\overline{X}$
- $\hat{b}_1$ = fitted slope coefficient = Cov (X,Y) / variance of X

The estimated intercept, $\hat{b}_0$, represents the value of the dependent variable at the point of intersection of the regression line and the axis of the dependent variable (usually, the vertical axis). The estimated slope coefficient, $\hat{b}_1$, is interpreted as the change in the dependent variable for a one-unit change in the independent variable.

LOS 10.b

Assumptions made regarding simple linear regression include the following:

1. A linear relationship exists between the dependent and the independent variable.
2. The variance of the residual term is constant (homoskedasticity).
3. The residual term is independently distributed (residuals are uncorrelated).
4. The residual term is normally distributed.

LOS 10.c

The total sum of squares (SST) measures the total variation in the dependent variable and equals the sum of the squared differences between its actual values and its mean.

The sum of squares regression (SSR) measures the variation in the dependent variable that is explained by the independent variable, and equals the sum of the squared distances between the predicted values and the mean of the dependent variable.

The mean square regression (MSR) is the SSR divided by the number of independent variables. For a simple linear regression, MSR = SSR.

The sum of squared errors (SSE) measures the unexplained variation in the dependent variable and is the sum of the squared vertical distances between the actual and the predicted values of the dependent variable.

The mean squared error (MSE) is the SSE divided by the degrees of freedom (n − 2 for a simple linear regression).

The coefficient of determination, R^2, is the proportion of the total variation of the dependent variable explained by the regression:

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

In simple linear regression, because there is only one independent variable (k = 1), the *F*-test tests the same null hypothesis as testing the statistical significance of b_1 using the t-test: H_0: $b_1 = 0$ versus H_a: $b_1 \neq 0$. With only one independent variable, F is calculated as follows:

$$\text{F-stat} = \frac{\text{MSR}}{\text{MSE}} \text{ with 1 and n} - 2 \text{ degrees of freedom}$$

LOS 10.d

ANOVA Table for Simple Linear Regression (k = 1)

Source of Variation	Degrees of Freedom (df)	Sum of Squares	Mean Sum of Squares
Regression (explained)	1	SSR	$\text{MSR} = \frac{\text{SSR}}{k} = \frac{\text{SSR}}{1} = \text{SSR}$
Error (unexplained)	$n - 2$	SSE	$\text{MSE} = \frac{\text{SSE}}{n-2}$
Total	$n - 1$	SST	

The standard error of the estimate in a simple linear regression is calculated as follows:

$$\text{SEE} = \sqrt{\frac{\text{SSE}}{n-2}}$$

LOS 10.e

A predicted value of the dependent variable, $\hat{Y}$, is determined by inserting the predicted value of the independent variable, X_p, in the regression equation and calculating $\hat{Y}_p = \hat{b}_0 + \hat{b}_1 X_p$.

The confidence interval for a predicted Y-value is $\left[\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)\right]$ where s_f is the standard error of the forecast.

LOS 10.f

Dependent Variable	Independent Variable	Model	Slope Interpretation
Logarithmic	Linear	Log-lin	*Relative* change in dependent variable for an *absolute* change in the independent variable
Linear	Logarithmic	Lin-log	*Absolute* change in dependent variable for a *relative* change in the independent variable
Logarithmic	Logarithmic	Log-log	*Relative* change in dependent variable for a *relative* change in the independent variable

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 10.1

1. **B** The slope coefficient is best interpreted as the predicted change in the dependent variable for a 1-unit change in the independent variable; if the slope coefficient estimate is 10.0 and the independent variable changes by 1 unit, the dependent variable is expected to change by 10 units. The intercept term is best interpreted

as the value of the dependent variable when the independent variable is equal to zero. (LOS 10.a)

2. **C** The model does not assume that the dependent variable is uncorrelated with the residuals. It does assume that the independent variable is uncorrelated with the residuals. (LOS 10.b)

Module Quiz 10.2

1. **A** In simple linear regression, the appropriate degrees of freedom for both confidence intervals is the number of observations in the sample (*n*) minus two. (LOS 10.c)

2. **A** In this regression, a_1 is the intercept term. To test the statistical significance means to test the null hypothesis that a_1 is equal to zero, versus the alternative that a_1 is not equal to zero. (LOS 10.c)

3. **C** The regression sum of squares measures the amount of variation in the dependent variable explained by the independent variable (i.e., the explained variation). The sum of squared errors measures the variation in the dependent variable not explained by the independent variable. The mean squared error is equal to the sum of squared errors divided by its degrees of freedom. (LOS 10.d)

Module Quiz 10.3

1. **C** The estimate of Y, given $X = 10$, is $Y = 5 + 3.5(10) = 40$. The critical value for a 90% confidence interval with a large sample size (*z*-statistic) is approximately 1.65. Given the standard error of the forecast of 2.5, the confidence interval for the estimated value of Y is $40 +/- 1.65(2.5) = 35.875$ to 44.125. (LOS 10.e)

2. **B** The appropriate model would be a lin-log model, in which the values of the dependent variable (Y) are regressed on the natural logarithms of the independent variable (X): $Y = b_0 + b_1 \ln(X)$. (LOS 10.f)

READING 11

INTRODUCTION TO BIG DATA TECHNIQUES

MODULE 11.1: INTRODUCTION TO FINTECH

Video covering this content is available online.

LOS 11.a: Describe aspects of "fintech" that are directly relevant for the gathering and analyzing of financial data.

The term **fintech** refers to developments in technology that can be applied to the financial services industry. Companies that are in the business of developing technologies for the finance industry are often referred to as fintech companies.

Some of the primary areas where fintech is developing include the following:

- Increasing functionality to handle large sets of data that may come from many sources and exist in various forms
- Tools and techniques such as artificial intelligence for analyzing very large datasets

LOS 11.b: Describe Big Data, artificial intelligence, and machine learning.

Big Data is a widely used expression that refers to all the potentially useful information that is generated in the economy. This includes not only data from traditional sources, such as financial markets, company financial reports, and government economic statistics, but also **alternative data** from nontraditional sources. Some of these nontraditional sources are as follows:

- Individuals generate usable data such as social media posts, online reviews, email, and website visits.
- Businesses generate potentially useful information such as bank records and retail scanner data. These kinds of data are referred to as **corporate exhaust**.
- Sensors, such as radio frequency identification chips, are embedded in numerous devices such as smartphones and smart buildings. The broad network of such devices is referred to as the **Internet of Things**.

Characteristics of Big Data include its volume, velocity, and variety.

The *volume* of data continues to grow by orders of magnitude. The units in which data can be measured have increased from megabytes and gigabytes to terabytes (1,000 gigabytes) and even petabytes (1,000 terabytes).

Velocity refers to how quickly data are communicated. Real-time data such as stock market price feeds are said to have low **latency**. Data that are only communicated periodically or with a lag are said to have high latency.

The *variety* of data refers to the varying degrees of structure in which data may exist. These range from structured forms (e.g., spreadsheets and databases), to semistructured forms (e.g., photos and web page code), to unstructured forms (e.g., video).

The field of **data science** concerns how we extract information from Big Data. Data science describes methods for processing and visualizing data. Processing methods include the following:

- *Capture*. This is collecting data and transforming it into usable forms.
- *Curation*. This is assuring data quality by adjusting for bad or missing data.
- *Storage*. This is archiving and accessing data.
- *Search*. This is examining stored data to find needed information.
- *Transfer*. This is moving data from their source or a storage medium to where they are needed.

Visualization techniques include the familiar charts and graphs that display structured data. To visualize less structured data requires other methods. Some examples of these are word clouds that illustrate the frequency with which words appear in a sample of text, or mind maps that display logical relations among concepts.

Taking advantage of Big Data presents numerous challenges. Analysts must ensure that the data they use are of high quality, accounting for the possibilities of outliers, bad or missing data, or sampling biases. The volume of data collected must be sufficient and appropriate for its intended use.

The need to process and organize data before using it can be especially problematic with qualitative and unstructured data. This is a process to which **artificial intelligence**, or computer systems that can be programmed to simulate human cognition, may be applied usefully. **Neural networks** are an example of artificial intelligence in that they are programmed to process information in a way similar to the human brain.

An important development in the field of artificial intelligence is **machine learning**. In machine learning, a computer algorithm is given inputs of source data, with no assumptions about their probability distributions, and may be given outputs of target data. The algorithm is designed to learn, without human assistance, how to model the output data based on the input data, or to learn how to detect and recognize patterns in the input data.

Machine learning typically requires vast amounts of data. A typical process begins with a *training* dataset in which the algorithm looks for relationships. A *validation* dataset is then used to refine these relationship models, which can then be applied to a *test* dataset to analyze their predictive ability.

In **supervised learning**, the input and output data are labeled, the machine learns to model the outputs from the inputs, and then the machine is given new data on which to use the model. In **unsupervised learning**, the input data are not labeled, and the machine learns to describe the structure of the data. **Deep learning** is a technique that uses layers of neural networks to identify patterns, beginning with simple patterns and advancing to more complex ones. Deep learning may employ supervised or unsupervised learning. Some of the applications of deep learning include image and speech recognition.

Machine learning can produce models that overfit or underfit the data. **Overfitting** occurs when the machine learns the input and output data too exactly, treats noise as true parameters, and identifies spurious patterns and relationships. In effect, the machine creates a model that is too complex. **Underfitting** occurs when the machine fails to identify actual patterns and relationships, treating true parameters as noise. This means that the model is not complex enough to describe the data. A further challenge with machine learning is that its results can be a “black box,” producing outcomes based on relationships that are not readily explainable.

LOS 11.c: Describe applications of Big Data and Data Science to investment management.

Applications of fintech that are relevant to investment management include text analytics, natural language processing, risk governance, and algorithmic trading.

Text analytics refers to the analysis of unstructured data in text or voice forms. An example of text analytics is analyzing the frequency of words and phrases. In the finance industry, text analytics have the potential to partially automate specific tasks such as evaluating company regulatory filings.

Natural language processing refers to the use of computers and artificial intelligence to interpret human language. Speech recognition and language translation are among the uses of natural language processing. Possible applications in finance could be to check for regulatory compliance in an examination of employee communications, or to evaluate large volumes of research reports to detect more subtle changes in sentiment than can be discerned from analysts’ recommendations alone.

Risk governance requires an understanding of a firm’s exposure to a wide variety of risks. Financial regulators require firms to perform risk assessments and stress testing. The simulations, scenario analysis, and other techniques used for risk analysis require large amounts of quantitative data along with a great deal of qualitative information. Machine learning and other techniques related to Big Data can be useful in modeling and testing risk, particularly if firms use real-time data to monitor risk exposures.

Algorithmic trading refers to computerized securities trading based on a predetermined set of rules. For example, algorithms may be designed to enter the optimal execution instructions for any given trade based on real-time price and volume data. Algorithmic trading can also be useful for executing large orders by determining the best way to divide the orders across exchanges. Another application of algorithmic trading is **high-frequency trading** that identifies and takes advantage of intraday securities mispricings.

MODULE QUIZ 11.1

1. Fintech is *most accurately* described as the:
   A. application of technology to the financial services industry.
   B. replacement of government-issued money with electronic currencies.
   C. clearing and settling of securities trades through distributed ledger technology.
2. Which of the following technological developments is *most likely* to be useful for analyzing Big Data?
   A. Machine learning.
   B. High-latency capture.
   C. The Internet of Things.

KEY CONCEPTS

LOS 11.a

Fintech refers to developments in technology that can be applied to the financial services industry. Companies that develop technologies for the finance industry are referred to as fintech companies.

LOS 11.b

Big Data refers to the potentially useful information that is generated in the economy, including data from traditional and nontraditional sources. Characteristics of Big Data include its volume, velocity, and variety.

Artificial intelligence refers to computer systems that can be programmed to simulate human cognition. Neural networks are an example of artificial intelligence.

Machine learning is programming that gives a computer system the ability to improve its performance of a task over time and is often used to detect patterns in large sets of data.

LOS 11.c

Applications of fintech to investment management include text analytics, natural language processing, risk governance, and algorithmic trading.

Text analytics refers to analyzing unstructured data in text or voice forms. Natural language processing is the use of computers and artificial intelligence to interpret human language. Algorithmic trading refers to computerized securities trading based on predetermined rules.

ANSWER KEY FOR MODULE QUIZZES

Module Quiz 11.1

1. **A** Fintech is the application of technology to the financial services industry and to companies that are involved in developing and applying technology for financial services. Cryptocurrencies and distributed ledger technology are examples of fintech-related developments. (LOS 11.a)

2. **A** Machine learning is a computer programming technique useful for identifying and modeling patterns in large volumes of data. The Internet of Things is the network of devices that is one of the sources of Big Data. Capture is one aspect of processing data. Latency is the lag between when data is generated and when it is needed. (LOS 11.b)

TOPIC QUIZ: QUANTITATIVE METHODS

You have now finished the Quantitative Methods topic section. Please log into your Schweser online dashboard and take the Topic Quiz on this section. The Topic Quiz provides immediate feedback on how effective your study has been for this material. Questions are more exam-like than typical Module Quiz or QBank questions; a score of less than 70% indicates that your study likely needs improvement. These tests are best taken timed; allow 1.5 minutes per question.

READING 12

FIRMS AND MARKET STRUCTURES

MODULE 12.1: BREAKEVEN, SHUTDOWN, AND SCALE

Video covering this content is available online.

LOS 12.a: Determine and interpret breakeven and shutdown points of production, as well as how economies and diseconomies of scale affect costs under perfect and imperfect competition.

In economics, we define the **short run** for a firm as the time period over which some factors of production are fixed. Typically, we assume that capital is fixed in the short run so that a firm cannot change its scale of operations (plant and equipment) over the short run. All factors of production (costs) are variable in the **long run**. The firm can let its leases expire and sell its equipment, thereby avoiding costs that are fixed in the short run.

Shutdown and Breakeven Under Perfect Competition

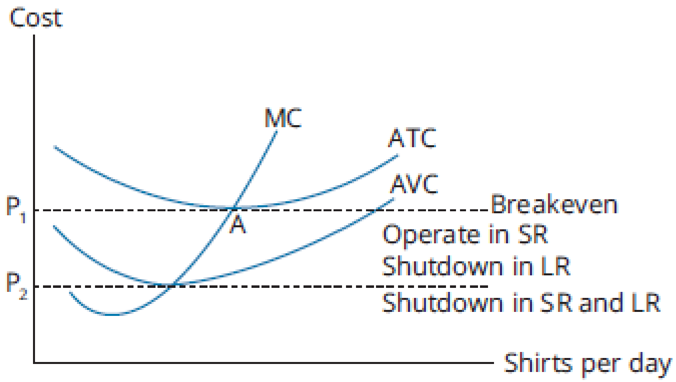
As a simple example of shutdown and breakeven analysis, consider a retail store with a one-year lease (fixed cost) and one employee (quasi-fixed cost), so that variable costs are simply the store's cost of merchandise. If the total sales (total revenue) just cover both fixed and variable costs, price equals both average revenue and average total cost —so we are at the breakeven output quantity, and economic profit equals zero.

During the period of the lease (the short run), as long as items are being sold for more than their variable cost, the store should continue to operate to minimize losses. If items are being sold for less than their average variable cost, losses would be reduced by shutting down the business in the short run.

In the long run, a firm should shut down if the price is less than average total cost, regardless of the relation between price and average variable cost.

For a firm under perfect competition (a price-taker), we can use a graph of cost functions to examine the profitability of the firm at different output prices. In Figure 12.1, at price P_1, price and average revenue equal average total cost. At the output level of Point A, the firm is making an economic profit of zero. At a price above P_1, economic profit is positive, and at prices less than P_1, economic profit is negative (the firm has economic losses).

Figure 12.1: Shutdown and Breakeven



Because some costs are fixed in the short run, it will be better for the firm to continue production in the short run as long as average revenue is greater than average variable costs. At prices between P_1 and P_2 in Figure 12.1, the firm has losses, but the losses are smaller than would occur if all production were stopped. As long as total revenue is greater than total variable cost, at least some of the firm's fixed costs are covered by continuing to produce and sell its product. If the firm were to shut down, losses would be equal to the fixed costs that still must be paid. As long as price is greater than average variable costs, the firm will minimize its losses in the short run by continuing in business.

If average revenue is less than average variable cost, the firm's losses are greater than its fixed costs, and it will minimize its losses by shutting down production in the short run. In this case (a price less than P_2 in Figure 12.1), the loss from continuing to operate is greater than the loss (total fixed costs) if the firm is shut down.

In the long run, all costs are variable, so a firm can avoid its (short-run) fixed costs by shutting down. For this reason, if price is expected to remain below minimum average total cost (Point A in Figure 12.1) in the long run, the firm will shut down rather than continue to generate losses.

To sum up, if average revenue is less than average variable cost in the short run, the firm should shut down. This is its **short-run shutdown point**. If average revenue is greater than average variable cost in the short run, the firm should continue to operate, even if it has losses. In the long run, the firm should shut down if average revenue is less than average total cost. This is the **long-run shutdown point**. If average revenue is just equal to average total cost, total revenue is just equal to total (economic) cost, and this is the firm's **breakeven point**.

- If AR ≥ ATC, the firm should stay in the market in both the short and long run.
- If AR ≥ AVC, but AR < ATC, the firm should stay in the market in the short run but will exit the market in the long run.
- If AR < AVC, the firm should shut down in the short run and exit the market in the long run.

Shutdown and Breakeven Under Imperfect Competition

For price-searcher firms (those that face downward-sloping demand curves), we could compare average revenue to ATC and AVC, just as we did for price-taker firms, to identify shutdown and breakeven points. However, marginal revenue is no longer equal to price.

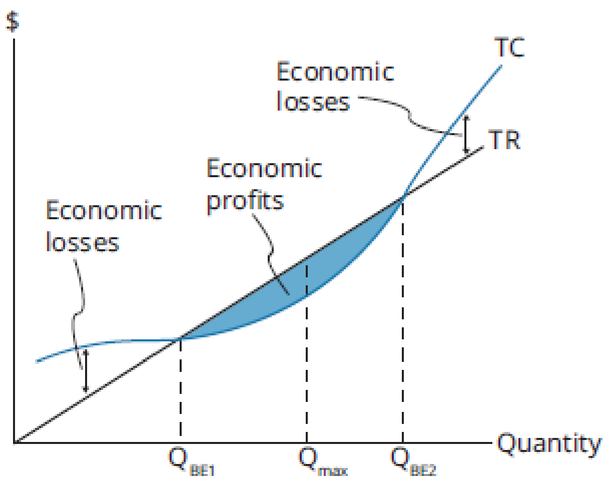
We can, however, still identify the conditions under which a firm is breaking even, should shut down in the short run, and should shut down in the long run in terms of total costs and total revenue. These conditions are as follows:

- TR = TC: break even
- TC > TR > TVC: firm should continue to operate in the short run but shut down in the long run
- TR < TVC: firm should shut down in the short run and the long run

Because price does not equal marginal revenue for a firm in imperfect competition, analysis based on total costs and revenues is better suited for examining breakeven and shutdown points.

The previously described relations hold for both price-taker and price-searcher firms. We illustrate these relations in Figure 12.2 for a price-taker firm (TR increases at a constant rate with quantity). Total cost equals total revenue at the breakeven quantities Q_{BE1} and Q_{BE2}. The quantity for which economic profit is maximized is shown as Q_{max}.

Figure 12.2: Breakeven Point Using the Total Revenue/Total Cost Approach



If the entire TC curve exceeds TR (i.e., no breakeven point), the firm will want to minimize the economic loss in the short run by operating at the quantity corresponding to the smallest (negative) value of TR − TC.

EXAMPLE: Short-run shutdown decision

For the last fiscal year, Legion Gaming reported total revenue of $700,000, total variable costs of $800,000, and total fixed costs of $400,000. Should the firm continue to operate in the short run?

Answer:

The firm should shut down. Total revenue of $700,000 is less than total costs of $1,200,000, and it is also less than total variable costs of $800,000. By shutting down, the firm will lose an amount equal to fixed costs of $400,000. This is less than the loss of operating, which is TR − TC = $500,000.

EXAMPLE: **Long-run shutdown decision**

Suppose, instead, that Legion Gaming reported total revenue of $850,000. Should the firm continue to operate in the short run? Should it continue to operate in the long run?

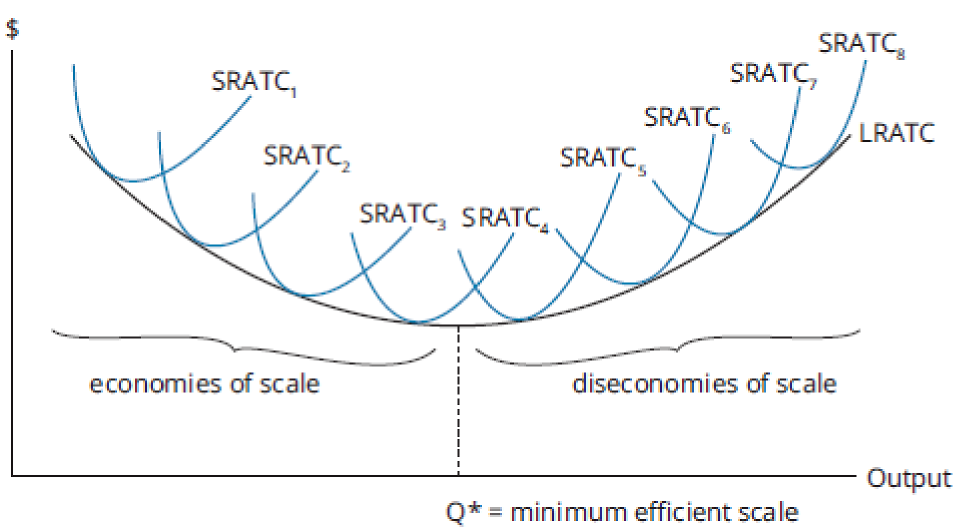
Answer:

In the short run, TR > TVC, and the firm should continue operating. The firm should consider exiting the market in the long run, as TR is not sufficient to cover all of the fixed costs and variable costs.

Economies and Diseconomies of Scale

While plant size is fixed in the short run, in the long run, firms can choose their most profitable scale of operations. Because the long-run average total cost (LRATC) curve is drawn for many different plant sizes or scales of operation, each point along the curve represents the minimum ATC for a given plant size or scale of operations. In Figure 12.3, we show a firm's LRATC curve along with short-run average total cost (SRATC) curves for many different plant sizes, with $SRATC_{n+1}$ representing a larger scale of operations than $SRATC_n$.

Figure 12.3: Economies and Diseconomies of Scale



We draw the LRATC curve as U-shaped. Average total costs first decrease with larger scale, but eventually begin to increase with larger scale. The lowest point on the LRATC corresponds to the scale or plant size at which the average total cost of production is at a minimum. This scale is sometimes called the **minimum efficient scale**. Under perfect competition, firms must operate at minimum efficient scale in long-run equilibrium, and LRATC will equal the market price. Recall that under perfect competition, firms earn zero economic profit in long-run equilibrium. Firms that have chosen a different scale of operations with higher average total costs will have economic losses and must either leave the industry or change to the minimum efficient scale.

The downward-sloping segment of the LRATC curve presented in Figure 12.3 indicates that **economies of scale** (or *increasing returns to scale*) are present. Economies of scale result from factors such as labor specialization, mass production, and investment in more efficient equipment and technology. In addition, the firm may be able to negotiate lower input prices with suppliers as it increases in size and purchases more resources. A firm operating with economies of scale can increase its competitiveness by expanding production and reducing costs.

The upward-sloping segment of the LRATC curve indicates that **diseconomies of scale** are present. Diseconomies of scale may result as the increasing bureaucracy of larger firms leads to inefficiency, problems with motivating a larger workforce, and greater barriers to innovation and entrepreneurial activity. A firm operating under diseconomies of scale will want to decrease output and move back toward the minimum efficient scale. The U.S. auto industry is an example of an industry that has exhibited diseconomies of scale.

There may be a relatively flat portion at the bottom of the LRATC curve that exhibits *constant returns to scale*, or relatively constant costs across a range of plant sizes.

MODULE QUIZ 12.1

1. In a purely competitive market, economic losses indicate that:
   A. price is below average total costs.
   B. collusion is occurring in the marketplace.
   C. firms need to expand output to reduce costs.
2. A firm is likely to operate in the short run as long as price is at least as great as:
   A. marginal cost.
   B. average total cost.
   C. average variable cost.
3. A firm's average revenue is greater than its average variable cost and less than its average total cost. If this situation is expected to persist, the firm should:
   A. shut down in the short run and in the long run.
   B. shut down in the short run, but operate in the long run.
   C. operate in the short run, but shut down in the long run.
4. If a firm increases its plant size by 10% and its minimum average total cost increases by 10%, the firm is experiencing:
   A. constant returns to scale.
   B. diseconomies of scale.

C. economies of scale.

MODULE 12.2: CHARACTERISTICS OF MARKET STRUCTURES

Video covering this content is available online.

LOS 12.b: Describe characteristics of perfect competition, monopolistic competition, oligopoly, and pure monopoly.

Recall from the prerequisite readings that perfect competition results in firm demand that is horizontal (perfectly elastic) at the market price. The firm demand curves for the three other market structures we discuss are all downward sloping. When a firm's demand curve slopes downward, marginal revenue (MR) is less than price. For both horizontal and downward-sloping demand curves, a firm will maximize profits by producing the quantity for which MR is just equal to marginal cost.

While it may not be true in every case, it may be useful to think of firms under pure monopoly as having the steepest demand curves. Firms under monopolistic competition typically have relatively elastic downward-sloping demand curves, while firms in an oligopoly market will face downward-sloping demand curves somewhere between these two extremes.

We can analyze where a market falls along the spectrum from perfect competition to pure monopoly by examining five factors:

1. Number of firms and their relative sizes
2. Degree to which firms differentiate their products
3. Bargaining power of firms with respect to pricing
4. Barriers to entry into or exit from the industry
5. Degree to which firms compete on factors other than price

Perfect competition refers to a market in which many firms produce identical products, barriers to entry into the market are very low, and firms compete for sales only on the basis of price. Firms face perfectly elastic (horizontal) demand curves at the price determined in the market because no firm has a large enough portion of the overall market to affect the market price of the good. The market for wheat in a region is a good approximation of such a market. Overall market supply and demand determine the price of wheat, and each producer can sell all that they choose to at that price.

Monopolistic competition differs from perfect competition in that products are not identical. Each firm differentiates its product(s) from those of other firms through some combination of differences in product quality, product features, and marketing. The demand curve faced by each firm is downward sloping (i.e., neither perfectly elastic nor perfectly inelastic). Prices that producers charge are not identical because of perceived differences among their products, and typically, barriers to entry are low. The market for toothpaste is a good example of monopolistic competition. Firms differentiate their products through features and marketing with claims of more attractiveness, whiter teeth, fresher breath, and even actually cleaning your teeth and preventing decay. If the price of your personal favorite increases, you are not likely to immediately switch to another brand as we assume under perfect competition. Some customers may switch brands in response to a 10% increase in price, and some may not. This is why firm demand is downward sloping rather than perfectly elastic.

The most important characteristic of an **oligopoly** market is that only a few firms are in the industry. In such a market, each firm must consider the actions and responses of other firms in setting price and business strategy. We say that such firms are *interdependent*. While the firms' products are typically good substitutes for each other, they may be either quite similar or differentiated through features, branding, marketing, and quality. Barriers to entry are typically high, often because of economies of scale in production or marketing, which accounts for the existence of a small number of firms with relatively large market shares. Demand can be more or less elastic than for firms in monopolistic competition. The automobile market is dominated by a small number of large firms and can be characterized as an oligopoly. The product and pricing decisions of Toyota certainly affect those of Ford, and vice versa. Automobile makers compete based on price, but they also compete through marketing, product features, and quality, which are often signaled strongly through the brand name. The oil industry also has a few firms with large market shares, but their products are, in most cases, good substitutes for each other.

A **monopoly** market is characterized by a single seller of a product with no good substitutes. This fact alone means that the firm faces a downward-sloping demand curve (the market demand curve) and has the power to choose the price at which it sells its product. High barriers to entry protect a monopoly producer from competition. One source of monopoly power is the protection offered by copyrights and patents. Another possible source of monopoly power is control over a resource specifically needed to produce the product. Most frequently, monopoly power is supported by specific laws or government regulation (e.g., a local electric utility).

Figure 12.4 shows the key features of each market structure.

Figure 12.4: Characteristics of Market Structures

	Perfect Competition	Monopolistic Competition	Oligopoly	Monopoly
Number of sellers	Many firms	Many firms	Few firms	Single firm
Barriers to entry	Very low	Low	High	Very high
Nature of substitute products	Very good substitutes	Good substitutes, but differentiated	Good substitutes or differentiated	No good substitutes
Nature of competition	Price only	Price, marketing, features	Price, marketing, features	Advertising
Pricing power	None	Some	Some to significant	Significant

LOS 12.c: Explain supply and demand relationships under monopolistic competition, including the optimal price and output for firms as well as pricing strategy.

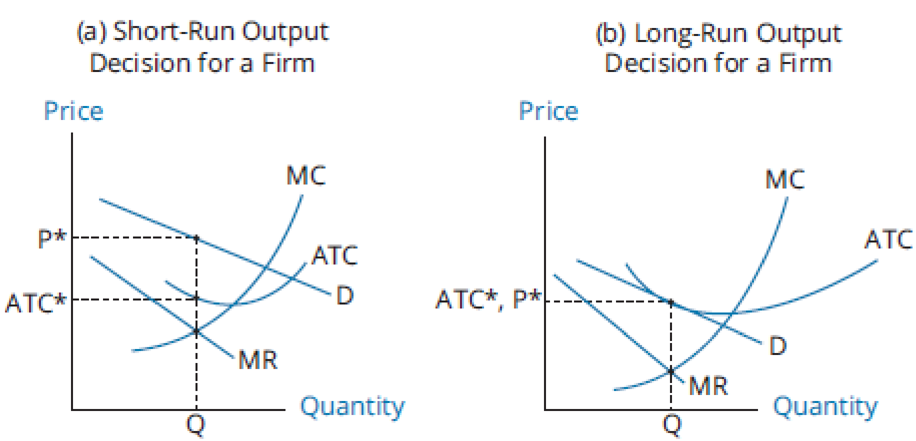
Monopolistic competition has the following market characteristics:

- *A large number of independent sellers*. (1) Each firm has a relatively small market share, so no individual firm has any significant power over price. (2) Firms only need to pay attention to average market price, not the prices of individual competitors. (3) There are too many firms in the industry for collusion (price-fixing) to be possible.
- *Differentiated products*. Each producer has a product that is, in some way, different from those of its competitors (in the minds of consumers). The competing products are considered close substitutes for one another.
- *Firms compete less on price and more on marketing, perceived quality, and differences in features*. Firms must make price and output decisions because they face downward-sloping demand curves.
- *Low barriers to entry*. The cost of entering the market and exiting the market are relatively low.

Think about the market for toothpaste. Brands of toothpaste are quite similar, and it is reasonable to assume that toothpaste is not too difficult or costly to produce. But brands are differentiated based on specific features, on influential advertising and marketing, and on the reputations of the producers.

The price/output decision for monopolistic competition is illustrated in Figure 12.5. Panel A of Figure 12.5 illustrates the short-run price/output characteristics of monopolistic competition for a single firm. As indicated, firms in monopolistic competition maximize economic profits by producing where marginal revenue (MR) equals marginal cost (MC), and by charging the price for that quantity from the demand curve, *D*. Here, the firm earns positive economic profits because price, *P**, exceeds average total cost, *ATC**. Due to low barriers to entry, competitors can enter the market in pursuit of these economic profits.

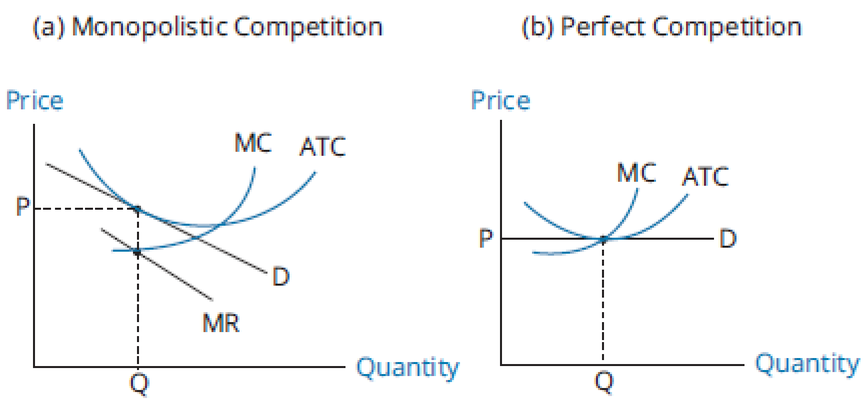
Figure 12.5: Short-Run and Long-Run Output Under Monopolistic Competition



Panel B of Figure 12.5 illustrates long-run equilibrium for a *representative* firm after new firms have entered the market. As indicated, the entry of new firms shifts the demand curve faced by each individual firm down to the point where price equals average total cost (P* = ATC*), such that economic profit is zero. At this point, there is no longer an incentive for new firms to enter the market, and long-run market equilibrium is established. A firm in monopolistic competition continues to produce at the quantity where MR = MC, but it no longer earns positive economic profits. We can get to a similar long-run equilibrium even without entry of new firms if each firm increases its marketing spending (a component of ATC) to either increase or defend its market share until ATC has increased to that shown in Panel B. If all firms compete in this way, each firm will produce Q* and sell at P* but earn no economic profit because marketing costs have increased ATC to ATC*, which is equal to price. Advertising expenses are often relatively high for firms in monopolistic competition.

Figure 12.6 illustrates the differences between long-run equilibrium in markets with monopolistic competition and markets with perfect competition. Note that with monopolistic competition, price is greater than MC (i.e., producers can realize an economic profit), average total cost is not at a minimum for the quantity produced (suggesting excess capacity, or an inefficient scale of production), and the price is slightly higher than under perfect competition. The point to consider here, however, is that perfect competition is characterized by no product differentiation. The question of the efficiency of monopolistic competition becomes, "Is there an economically efficient amount of product differentiation?"

Figure 12.6: Firm Output Under Monopolistic and Perfect Competition



In a world with only one brand of toothpaste, clearly, average production costs would be lower. That fact alone probably does not mean that a world with only one brand or type of toothpaste would be a better world. While product differentiation has costs, it also has benefits to consumers. Consider the market for a pharmaceutical that reduces blood pressure to prolong life. There may be several competing drugs that are more or less effective for, or well or poorly tolerated by, different groups of patients. In this case, we may find that firm demand curves are relatively steep compared to those for brands of toothpaste, indicating that the competing drugs are not considered good substitutes for many patients.

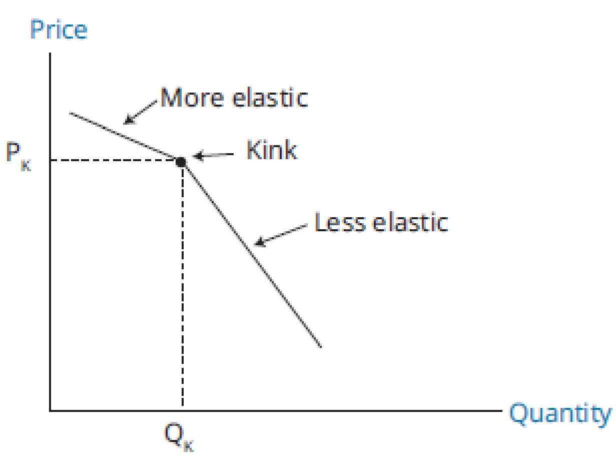
LOS 12.d: Explain supply and demand relationships under oligopoly, including the optimal price and output for firms as well as pricing strategy.

Compared to monopolistic competition, an oligopoly market has higher barriers to entry and fewer firms. The other key difference is that the firms are interdependent; a price change by one firm can be expected to be met by a price change by its competitors in response. This means that the actions of another firm will directly affect a given firm's demand curve for the product. Given this complicating fact, models of oligopoly pricing and profits must make numerous important assumptions. In the following, we describe four of these models and their implications for price and quantity:

1. Kinked demand curve model
2. Cournot duopoly model
3. Nash equilibrium model
4. Stackelberg dominant firm model

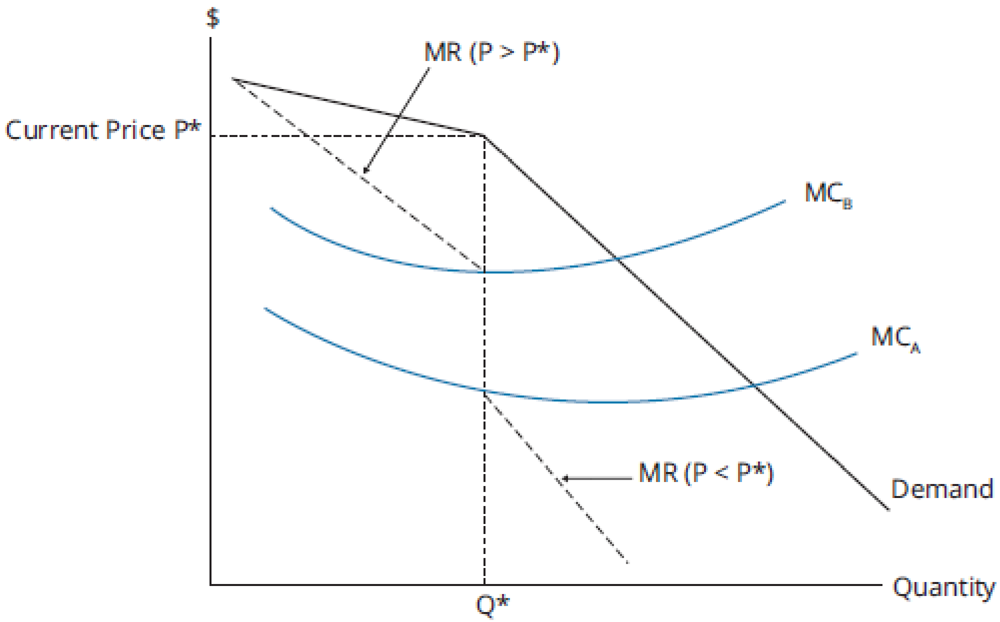
One traditional model of oligopoly, the **kinked demand curve model**, is based on the assumption that competitors are unlikely to match a price increase by a competitor, but very likely to match a price decrease by a competitor. This results in a kink in the demand curves faced by each producer, at the current market price. Each firm believes that it faces a demand curve that is more elastic (flatter) above the current price (the kink in the demand curve) than it is below the given price. The kinked demand curve model is illustrated in Figure 12.7. The kink price is at price P_K, where a firm produces Q_K. A firm believes that if it raises its price above P_K, its competitors will remain at P_K, and it will lose market share because it has the highest price. Above P_K, the demand curve is considered to be relatively elastic (i.e., a small price increase will result in a large decrease in demand). On the other hand, if a firm decreases its price below P_K, other firms will match the price cut, and all firms will experience a relatively small increase in sales relative to any price reduction. Therefore, Q_K is the profit-maximizing level of output.

Figure 12.7: Kinked Demand Curve Model



With a kink in the demand curve, we also get a gap in the associated MR curve, as shown in Figure 12.8. For any firm with a MC curve passing through this gap, the price where the kink is located is the firm's profit-maximizing price.

Figure 12.8: Marginal Revenue With Kinked Demand Curve



We say that the decisions of firms in an oligopoly are interdependent; that is, the pricing decision of one firm depends on the pricing decisions of other firms. Some models of market price equilibrium have a set of rules for the actions of oligopolists. These rules assume they choose prices based on the choices of the other firms. By specifying the decision rules that each firm follows, we can design a model that allows us to determine the equilibrium prices and quantities for firms operating in an oligopoly market.

An early model of oligopoly pricing decisions is the **Cournot model**. In Cournot's duopoly model, two firms with identical MC curves each choose their preferred selling price based on the price the other firm chose in the previous period. Firms assume that the competitor's price will not change. The long-run equilibrium for an oligopoly with two firms (duopoly), in the Cournot model, is for both firms to sell the same quantity, dividing the market equally at the equilibrium price. The equilibrium price is less than the price that a single monopolist would charge, but greater than the equilibrium price that would result under perfect competition. With a greater number of producers, the long-run market equilibrium price moves toward the competitive price.

Another model, the **Stackelberg model**, uses a different set of rules and produces a different result. While the Cournot model assumes the competitors choose price simultaneously each period, the Stackelberg model assumes pricing decisions are made sequentially. One firm, the "leader," chooses its price first, and the other firm chooses a price based on the leader's price. In long-run equilibrium, under these rules, the leader charges a higher price and receives a greater proportion of the firms' total profits.

These models are early versions of *rules-based models*, which fall under the heading of what are now generally termed *strategic games*. Strategic games comprise decision