2018 SchweserNotes™ Part I

FRM®
Exam Prep

Quantitative Analysis

eBook 2

KAPLAN

Getting Started

FRM®

FRM® Exam Part I

Welcome

As the VP of Advanced Designations at Kaplan Schweser, I am pleased to have the opportunity to help you prepare for the 2018 FRM® Exam. Getting an early start on your study program is important for you to sufficiently **prepare**, **practice**, and **perform** on exam day. Proper planning will allow you to set aside enough time to master the learning objectives in the Part I curriculum.

Now that you've received your SchweserNotes™, here's how to get started:

Step 1: Access Your Online Tools

Visit www.schweser.com/frm and log in to your online account using the button located in the top navigation bar. After logging in, select the appropriate part and proceed to the dashboard where you can access your online products.

Step 2: Create a Study Plan

Create a study plan with the Schweser Study Calendar (located on the Schweser dashboard). Then view the Candidate Resource Library on-demand videos for an introduction to core concepts.

Step 3: Prepare and Practice

Read your SchweserNotes™

Our clear, concise study notes will help you **prepare** for the exam. At the end of each reading, you can answer the Concept Checker questions for better understanding of the curriculum.

Attend a Weekly Class

Attend our Live Online Weekly Class or review the on-demand archives as often as you like. Our expert faculty will guide you through the FRM curriculum with a structured approach to help you **prepare** for the exam. (See our instruction packages to the right. Visit www.schweser.com/frm to order.)

Practice with SchweserPro™ QBank

Maximize your retention of important concepts and **practice** answering exam-style questions in the SchweserPro™ QBank and taking several Practice Exams. Use Schweser's QuickSheet for continuous review on the go. (Visit www.schweser.com/frm to order.)

Step 4: Final Review

A few weeks before the exam, make use of our Online Review Workshop Package. Review key curriculum concepts in every topic, **perform** by working through demonstration problems, and **practice** your exam techniques with our 8-hour live Online Review Workshop. Use Schweser's Secret Sauce® for convenient study on the go.

Step 5: Perform

As part of our Online Review Workshop Package, take a Schweser Mock Exam to ensure you are ready to **perform** on the actual FRM Exam. Put your skills and knowledge to the test and gain confidence before the exam.

Again, thank you for trusting Kaplan Schweser with your FRM Exam preparation!

Sincerely,

Derek Burkett

Derek Burkett, CFA, FRM, CAIA
VP, Advanced Designations, Kaplan Schweser

The Kaplan Way

Prepare — Acquire new knowledge through demonstration and examples.

Practice — Apply new knowledge through simulation and practice.

Perform — Evaluate mastery of new knowledge and identify achieved outcomes.

FRM® Instruction Packages:

- PremiumPlus™ Package
- Premium Instruction Package

Live Instruction*:

Remember to join our Live Online Weekly Class. Register online today at www.schweser.com/frm.

*Dates, times, and instructors subject to change

MKT-005735

FRM Part I Book 2: Quantitative Analysis

FRM 2018 PART I BOOK 2: QUANTITATIVE ANALYSIS

Published in 2018 by Kaplan, Inc.

Printed in the United States of America.

ISBN: 978-1-4754-6998-1

Reading Assignments and Learning Objectives

The following material is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by the Global Association of Risk Professionals.

Reading Assignments

Learning Objectives

15. Probabilities

After completing this reading, you should be able to:

1. Describe and distinguish between continuous and discrete random variables. (page 13)
2. Define and distinguish between the probability density function, the cumulative distribution function, and the inverse cumulative distribution function. (page 15)
3. Calculate the probability of an event given a discrete probability function. (page 16)
4. Distinguish between independent and mutually exclusive events. (page 19)
5. Define joint probability, describe a probability matrix, and calculate joint probabilities using probability matrices. (page 21)
6. Define and calculate a conditional probability, and distinguish between conditional and unconditional probabilities. (page 18)

16. Basic Statistics

After completing this reading, you should be able to:

1. Interpret and apply the mean, standard deviation, and variance of a random variable. (page 29)
2. Calculate the mean, standard deviation, and variance of a discrete random variable. (page 29)
3. Interpret and calculate the expected value of a discrete random variable. (page 34)
4. Calculate and interpret the covariance and correlation between two random variables. (page 38)
5. Calculate the mean and variance of sums of variables. (page 34)
6. Describe the four central moments of a statistical variable or distribution: mean, variance, skewness and kurtosis. (page 42)
7. Interpret the skewness and kurtosis of a statistical distribution, and interpret the concepts of coskewness and cokurtosis. (page 44)
8. Describe and interpret the best linear unbiased estimator. (page 48)

17. Distributions

After completing this reading, you should be able to:

1. Distinguish the key properties among the following distributions: uniform distribution, Bernoulli distribution, Binomial distribution, Poisson distribution, normal distribution, lognormal distribution, Chi-squared distribution, Student's t, and F-distributions, and identify common occurrences of each distribution. (page 53)
2. Describe the central limit theorem and the implications it has when combining independent and identically distributed (i.i.d.) random variables. (page 66)
3. Describe i.i.d. random variables and the implications of the i.i.d. assumption when combining random variables. (page 66)
4. Describe a mixture distribution and explain the creation and characteristics of mixture distributions. (page 70)

18. Bayesian Analysis

After completing this reading, you should be able to:

1. Describe Bayes' theorem and apply this theorem in the calculation of conditional probabilities. (page 75)

2. Compare the Bayesian approach to the frequentist approach. (page 80)
3. Apply Bayes' theorem to scenarios with more than two possible outcomes and calculate posterior probabilities. (page 81)

19. Hypothesis Testing and Confidence Intervals

After completing this reading, you should be able to:
1. Calculate and interpret the sample mean and sample variance. (page 90)
2. Construct and interpret a confidence interval. (page 96)
3. Construct an appropriate null and alternative hypothesis, and calculate an appropriate test statistic. (page 100)
4. Differentiate between a one-tailed and a two-tailed test and identify when to use each test. (page 102)
5. Interpret the results of hypothesis tests with a specific level of confidence. (page 113)
6. Demonstrate the process of backtesting VaR by calculating the number of exceedances. (page 121)

20. Linear Regression with One Regressor

After completing this reading, you should be able to:
1. Explain how regression analysis in econometrics measures the relationship between dependent and independent variables. (page 128)
2. Interpret a population regression function, regression coefficients, parameters, slope, intercept, and the error term. (page 129)
3. Interpret a sample regression function, regression coefficients, parameters, slope, intercept, and the error term. (page 130)
4. Describe the key properties of a linear regression. (page 131)
5. Define an ordinary least squares (OLS) regression and calculate the intercept and slope of the regression. (page 132)
6. Describe the method and three key assumptions of OLS for estimation of parameters. (page 133)
7. Summarize the benefits of using OLS estimators. (page 133)
8. Describe the properties of OLS estimators and their sampling distributions, and explain the properties of consistent estimators in general. (page 133)
9. Interpret the explained sum of squares, the total sum of squares, the residual sum of squares, the standard error of the regression, and the regression R^2. (page 134)
10. Interpret the results of an OLS regression. (page 134)

21. Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

After completing this reading, you should be able to:
1. Calculate and interpret confidence intervals for regression coefficients. (page 142)
2. Interpret the p-value. (page 144)
3. Interpret hypothesis tests about regression coefficients. (page 143)
4. Evaluate the implications of homoskedasticity and heteroskedasticity. (page 147)
5. Determine the conditions under which the OLS is the best linear conditionally unbiased estimator. (page 149)
6. Explain the Gauss-Markov Theorem and its limitations, and alternatives to the OLS. (page 149)
7. Apply and interpret the t-statistic when the sample size is small. (page 150)

22. Linear Regression with Multiple Regressors

After completing this reading, you should be able to:

1. Define and interpret omitted variable bias, and describe the methods for addressing this bias. (page 156)
2. Distinguish between single and multiple regression. (page 157)
3. Interpret the slope coefficient in a multiple regression. (page 158)
4. Describe homoskedasticity and heteroskedasticity in a multiple regression. (page 159)
5. Describe the OLS estimator in a multiple regression. (page 157)
6. Calculate and interpret measures of fit in multiple regression. (page 159)
7. Explain the assumptions of the multiple linear regression model. (page 162)
8. Explain the concepts of imperfect and perfect multicollinearity and their implications. (page 162)

23. Hypothesis Tests and Confidence Intervals in Multiple Regression

After completing this reading, you should be able to:

1. Construct, apply, and interpret hypothesis tests and confidence intervals for a single coefficient in a multiple regression. (page 170)
2. Construct, apply, and interpret joint hypothesis tests and confidence intervals for multiple coefficients in a multiple regression. (page 176)
3. Interpret the F-statistic. (page 176)
4. Interpret tests of a single restriction involving multiple coefficients. (page 182)
5. Interpret confidence sets for multiple coefficients. (page 176)
6. Identify examples of omitted variable bias in multiple regressions. (page 183)
7. Interpret the R^2 and adjusted R^2 in a multiple regression. (page 181)

24. Modeling and Forecasting Trend

After completing this reading, you should be able to:

1. Describe linear and nonlinear trends. (page 189)
2. Describe trend models to estimate and forecast trends. (page 192)
3. Compare and evaluate model selection criteria, including mean squared error (MSE), s^2, the Akaike information criterion (AIC), and the Schwarz information criterion (SIC). (page 197)
4. Explain the necessary conditions for a model selection criterion to demonstrate consistency. (page 200)

25. Modeling and Forecasting Seasonality

After completing this reading, you should be able to:

1. Describe the sources of seasonality and how to deal with it in time series analysis. (page 206)
2. Explain how to use regression analysis to model seasonality. (page 208)
3. Explain how to construct an h-step-ahead point forecast. (page 210)

26. Characterizing Cycles

After completing this reading, you should be able to:

1. Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression. (page 214)
2. Describe the requirements for a series to be covariance stationary. (page 214)
3. Explain the implications of working with models that are not covariance stationary. (page 217)

4. Define white noise, and describe independent white noise and normal (Gaussian) white noise. (page 218)
5. Explain the characteristics of the dynamic structure of white noise. (page 218)
6. Explain how a lag operator works. (page 220)
7. Describe Wold's theorem. (page 221)
8. Define a general linear process. (page 221)
9. Relate rational distributed lags to Wold's theorem. (page 221)
10. Calculate the sample mean and sample autocorrelation, and describe the Box-Pierce Q-statistic and the Ljung-Box Q-statistic. (page 221)
11. Describe sample partial autocorrelation. (page 221)

27. Modeling Cycles: MA, AR, and ARMA Models

After completing this reading, you should be able to:

1. Describe the properties of the first-order moving average (MA(1)) process, and distinguish between autoregressive representation and moving average representation. (page 229)
2. Describe the properties of a general finite-order process of order q (MA(q)) process. (page 231)
3. Describe the properties of the first-order autoregressive (AR(1)) process, and define and explain the Yule-Walker equation. (page 231)
4. Describe the properties of a general p^{th} order autoregressive (AR(p)) process. (page 233)
5. Define and describe the properties of the autoregressive moving average (ARMA) process. (page 233)
6. Describe the application of AR and ARMA processes. (page 234)

28. Volatility

After completing this reading, you should be able to:

1. Define and distinguish between volatility, variance rate, and implied volatility. (page 239)
2. Describe the power law. (page 240)
3. Explain how various weighting schemes can be used in estimating volatility. (page 242)
4. Apply the exponentially weighted moving average (EWMA) model to estimate volatility. (page 243)
5. Describe the generalized autoregressive conditional heteroskedasticity (GARCH (p,q)) model for estimating volatility and its properties. (page 244)
6. Calculate volatility using the GARCH(1,1) model. (page 244)
7. Explain mean reversion and how it is captured in the GARCH(1,1) model. (page 245)
8. Explain the weights in the EWMA and GARCH(1,1) models. (page 243)
9. Explain how GARCH models perform in volatility forecasting. (page 246)
10. Describe the volatility term structure and the impact of volatility changes. (page 246)

29. Correlations and Copulas

After completing this reading, you should be able to:

1. Define correlation and covariance and differentiate between correlation and dependence. (page 251)
2. Calculate covariance using the EWMA and GARCH(1,1) models. (page 253)

3. Apply the consistency condition to covariance. (page 256)
4. Describe the procedure of generating samples from a bivariate normal distribution. (page 257)
5. Describe properties of correlations between normally distributed variables when using a one-factor model. (page 258)
6. Define copula and describe the key properties of copulas and copula correlation. (page 258)
8. Describe the Gaussian copula, Student's t-copula, multivariate copula, and one-factor copula. (page 260)
7. Explain tail dependence. (page 262)

30. Simulation Methods

After completing this reading, you should be able to:

1. Describe the basic steps to conduct a Monte Carlo simulation. (page 269)
2. Describe ways to reduce Monte Carlo sampling error. (page 270)
3. Explain how to use antithetic variate technique to reduce Monte Carlo sampling error. (page 271)
4. Explain how to use control variates to reduce Monte Carlo sampling error and when it is effective. (page 272)
5. Describe the benefits of reusing sets of random number draws across Monte Carlo experiments and how to reuse them. (page 273)
6. Describe the bootstrapping method and its advantage over Monte Carlo simulation. (page 274)
8. Describe situations where the bootstrapping method is ineffective. (page 275)
7. Describe the pseudo-random number generation method and how a good simulation design alleviates the effects the choice of the seed has on the properties of the generated series. (page 275)
9. Describe disadvantages of the simulation approach to financial problem solving. (page 276)

The Time Value of Money

Exam Focus

This is an optional reading that provides a tutorial for time value of money (TVM) calculations. Understanding how to use your financial calculator to make these calculations will be very beneficial as you proceed through the curriculum. In particular, for the fixed income material in Book 4, FRM candidates should be able to perform present value calculations using TVM functions. We have included Concept Checkers at the end of this reading for additional practice with these concepts.

Time Value of Money Concepts and Applications

The concept of **compound interest** or **interest on interest** is deeply embedded in time value of money (TVM) procedures. When an investment is subjected to compound interest, the growth in the value of the investment from period to period reflects not only the interest earned on the original principal amount but also on the interest earned on the previous period's interest earnings—the interest on interest.

TVM applications frequently call for determining the **future value** (FV) of an investment's cash flows as a result of the effects of compound interest. Computing FV involves projecting the cash flows forward, on the basis of an appropriate compound interest rate, to the end of the investment's life. The computation of the **present value** (PV) works in the opposite direction—it brings the cash flows from an investment back to the beginning of the investment's life based on an appropriate compound rate of return.

Being able to measure the PV and/or FV of an investment's cash flows becomes useful when comparing investment alternatives because the value of the investment's cash flows must be measured at some common point in time, typically at the end of the investment horizon (FV) or at the beginning of the investment horizon (PV).

Using a Financial Calculator

It is very important that you be able to use a financial calculator when working TVM problems because the FRM exam is constructed under the assumption that candidates have the ability to do so. There is simply no other way that you will have time to solve TVM problems. *GARP allows only four types of calculators to be used for the exam—the TI BAII Plus® (including the BAII Plus Professional), the HP 12C® (including the HP 12C Platinum), the HP 10bII®, and the HP 20b®. This reading is written primarily with the TI BAII Plus in mind.* If you don't already own a calculator, go out and buy a *TI BAII Plus!* However, if you already own one of the HP models listed and are comfortable with it, by all means continue to use it.

The TI BAII 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 TVM applications we will be studying. So prior to using our Study Notes, 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 doesn't, 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 TVM problems.

Before we begin working with financial calculators, you should familiarize yourself with your TI by locating the TVM keys noted here. These are the only keys you need to know to work virtually all TVM problems.

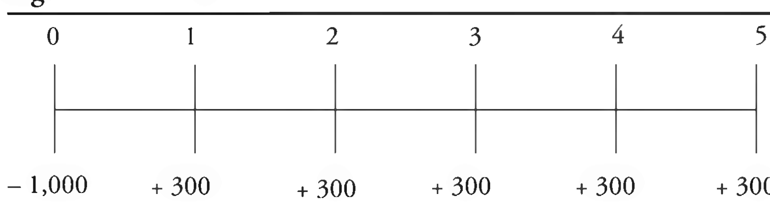
- N = Number of compounding periods
- I/Y = Interest rate per compounding period
- PV = Present value
- FV = Future value
- PMT = Annuity payments, or constant periodic cash flow
- CPT = Compute

Time Lines

It is often a good idea to draw a time line before you start to solve a TVM problem. A **time line** is simply a diagram of the cash flows associated with a TVM problem. A cash flow that occurs in the present (today) is put at time zero. Cash outflows (payments) are given a negative sign, and cash inflows (receipts) are given a positive sign. Once the cash flows are assigned to a time line, they may be moved to the beginning of the investment period to calculate the PV through a process called **discounting** or to the end of the period to calculate the FV using a process called **compounding**.

Figure 1 illustrates a time line for an investment that costs $1,000 today (outflow) and will return a stream of cash payments (inflows) of $300 per year at the end of each of the next five years.

Figure 1: Time Line



Please recognize that the cash flows occur at the end of the period depicted on the time line. Furthermore, note that the end of one period is the same as the beginning of the next period. For example, the end of the second year (t = 2) is the same as the beginning of the third year, so a cash flow at the beginning of year 3 appears at time t = 2 on the time line. Keeping this convention in mind will help you keep things straight when you are setting up TVM problems.

Professor's Note: Throughout the problems in this reading, rounding differences may occur between the use of different calculators or techniques presented in this document. So don't panic if you are a few cents off in your calculations.

Interest rates are our measure of 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 in order to get their equivalent value in current dollars or other currency. Finally, we can also view interest rates as the **opportunity cost** of current consumption. If the market rate of interest on one-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 has no expectation of inflation in it. When we speak of a real rate of return, we are referring to an investor's increase in purchasing power (after adjusting for inflation). Since expected inflation in future periods is not zero, the rates we observe on U.S. Treasury bills (T-bills), for example, are risk-free rates but not *real* rates of return. T-bill rates are *nominal risk-free rates* because they contain an *inflation premium*. The approximate relation here is:

nominal risk-free rate = real risk-free rate + expected inflation rate

Securities may have one or more **types of risk**, and each added risk increases the required rate of return on the security. These types of risk are:

- **Default risk.** The risk that a borrower will not make the promised payments in a timely manner.
- **Liquidity risk.** The risk of receiving less than fair value for an investment if it must be sold for cash quickly.
- **Maturity risk.** As we will cover in detail in the readings on debt securities in Book 4, 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 very liquid, short-term, default risk-free rate such as that on T-bills. We can write:

required interest rate on a security	=	nominal risk-free rate
	+	default risk premium
	+	liquidity premium
	+	maturity risk premium

Present Value (PV) of a Single Sum

The PV of a single sum is today's value of a cash flow that is to be received at some point in the future. In other words, it is the amount of money that must be invested today, at a given rate of return over a given period of time, in order to end up with a specified FV. As previously mentioned, the process for finding the PV of a cash flow is known as *discounting* (i.e., future cash flows are "discounted" back to the present). The interest rate used in the discounting process is commonly referred to as the **discount rate** but may also be referred to as the **opportunity cost**, **required rate of return**, and the **cost of capital**. Whatever you want to call it, it represents the annual compound rate of return that can be earned on an investment.

The relationship between PV and FV is as follows:

$$PV = FV \times \left[\frac{1}{(1+I/Y)^N}\right] = \frac{FV}{(1+I/Y)^N}$$

Note that for a single future cash flow, PV is always less than the FV whenever the discount rate is positive.

The quantity $1/(1 + I/Y)^N$ in the PV equation is frequently referred to as the **present value factor, present value interest factor**, or **discount factor** for a single cash flow at I/Y over N compounding periods.

Example: PV of a single sum

Given a discount rate of 9%, **calculate** the PV of a $1,000 cash flow that will be received in five years.

Answer:

To solve this problem, input the relevant data and compute PV.

N = 5; I/Y = 9; FV = 1,000; CPT → PV = –$649.93 (ignore the sign)

Professor's Note: With single sum PV problems, you can either enter FV as a positive number and ignore the negative sign on PV or enter FV as a negative number.

This relatively simple problem could also be solved using the following PV equation.

$$PV = \frac{1{,}000}{(1+0.09)^5} = \$649.93$$

On the TI, enter 1.09 [y^x] 5 [=] [1/x] [×] 1,000 [=].

The PV computed here implies that at a rate of 9%, an investor will be indifferent between $1,000 in five years and $649.93 today. Put another way, $649.93 is the amount that must be invested today at a 9% rate of return in order to generate a cash flow of $1,000 at the end of five years.

Annuities

An **annuity** is a stream of *equal cash flows* that occurs at *equal intervals* over a given period. Receiving $1,000 per year at the end of each of the next eight years is an example of an annuity. The *ordinary annuity* is the most common type of annuity. It is characterized by cash flows that occur at the *end* of each compounding period. This is a typical cash flow pattern for many investment and business finance applications.

Computing the FV or PV of an annuity with your calculator is no more difficult than it is for a single cash flow. You will know four of the five relevant variables and solve for the fifth (either PV or FV). The difference between single sum and annuity TVM problems is that instead of solving for the PV or FV of a single cash flow, we solve for the PV or FV of a stream of equal periodic cash flows, where the size of the periodic cash flow is defined by the payment (PMT) variable on your calculator.

Example: FV of an ordinary annuity

What is the future value of an ordinary annuity that pays $150 per year at the end of each of the next 15 years, given the investment is expected to earn a 7% rate of return?

Answer:

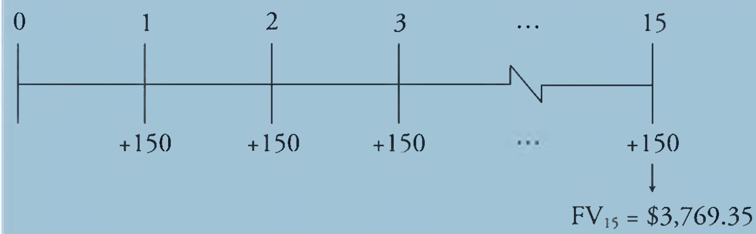
This problem can be solved by entering the relevant data and computing FV.

N = 15; I/Y = 7; PMT = –150; CPT → FV = $3,769.35

Implicit here is that PV = 0.

The time line for the cash flows in this problem is depicted in Figure 2.

Figure 2: FV of an Ordinary Annuity



As indicated here, the sum of the compounded values of the individual cash flows in this 15-year ordinary annuity is \$3,769.35. Note that the annuity payments themselves amounted to \$2,250 = 15 × \$150, and the balance is the interest earned at the rate of 7% per year.

To find the PV of an ordinary annuity, we use the future cash flow stream, PMT, that we used with FV annuity problems, but we discount the cash flows back to the present (time = 0) rather than compounding them forward to the terminal date of the annuity.

Here again, the PMT variable is a *single* periodic payment, *not* the total of all the payments (or deposits) in the annuity. The PVA_O measures the collective PV of a stream of equal cash flows received at the end of each compounding period over a stated number of periods, N, given a specified rate of return, I/Y. The following example illustrates how to determine the PV of an ordinary annuity using a financial calculator.

Example: PV of an ordinary annuity

What is the PV of an annuity that pays \$200 per year at the end of each of the next 13 years given a 6% discount rate?

Answer:

To solve this problem, enter the relevant information and compute PV.

N = 13; I/Y = 6; PMT = –200; CPT → PV = \$1,770.54

The \$1,770.54 computed here represents the amount of money that an investor would need to invest *today* at a 6% rate of return to generate 13 end-of-year cash flows of \$200 each.

Present Value of a Perpetuity

A **perpetuity** is a financial instrument that pays a fixed amount of money at set intervals over an *infinite* period of time. In essence, a perpetuity is a perpetual annuity. British consul bonds and most preferred stocks are examples of perpetuities since they promise fixed interest or dividend payments forever. Without going into all the mathematical details, the discount factor for a perpetuity is just one divided by the appropriate rate of return (i.e., 1/r). Given this, we can compute the PV of a perpetuity.

$$PV_{perpetuity} = \frac{PMT}{I/Y}$$

The PV of a perpetuity is the fixed periodic cash flow divided by the appropriate periodic rate of return.

As with other TVM applications, it is possible to solve for unknown variables in the $PV_{perpetuity}$ equation. In fact, you can solve for any one of the three relevant variables, given the values for the other two.

Example: PV of a perpetuity

Assume the preferred stock of Kodon Corporation pays \$4.50 per year in annual dividends and plans to follow this dividend policy forever. Given an 8% rate of return, what is the value of Kodon's preferred stock?

Answer:

Given that the value of the stock is the PV of all future dividends, we have:

$$PV_{perpetuity} = \frac{4.50}{0.08} = \$56.25$$

Thus, if an investor requires an 8% rate of return, the investor should be willing to pay \$56.25 for each share of Kodon's preferred stock.

Example: Rate of return for a perpetuity

Using the Kodon preferred stock described in the preceding example, **determine** the rate of return that an investor would realize if she paid \$75.00 per share for the stock.

Answer:

Rearranging the equation for $PV_{perpetuity}$, we get:

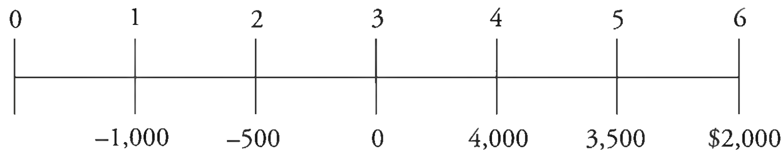
$$I/Y = \frac{PMT}{PV_{perpetuity}} = \frac{4.50}{75.00} = 0.06 = 6.0\%$$

This implies that the return (yield) on a \$75 preferred stock that pays a \$4.50 annual dividend is 6.0%.

PV and FV of Uneven Cash Flow Series

It is not uncommon to have applications in investments and corporate finance where it is necessary to evaluate a cash flow stream that is not equal from period to period. The time line in Figure 3 depicts such a cash flow stream.

Figure 3: Time Line for Uneven Cash Flows



This 6-year cash flow series is not an annuity since the cash flows are different every year. In fact, there is one year with zero cash flow and two others with negative cash flows. In essence, this series of uneven cash flows is nothing more than a stream of annual single sum cash flows. Thus, to find the PV or FV of this cash flow stream, all we need to do is sum the PVs or FVs of the individual cash flows.

Example: Computing the FV of an uneven cash flow series

Using a rate of return of 10%, **compute** the future value of the 6-year uneven cash flow stream described in Figure 3 at the end of the sixth year.

Answer:

The FV for the cash flow stream is determined by first computing the FV of each individual cash flow, then summing the FVs of the individual cash flows. Note that we need to preserve the signs of the cash flows.

FV_1: PV = –1,000; I/Y = 10; N = 5; CPT → FV = FV_1 = –1,610.51

FV_2: PV = –500; I/Y = 10; N = 4; CPT → FV = FV_2 = –732.05

FV_3: PV = 0; I/Y = 10; N = 3; CPT → FV = FV_3 = 0.00

FV_4: PV = 4,000; I/Y = 10; N = 2; CPT → FV = FV_4 = 4,840.00

FV_5: PV = 3,500; I/Y = 10; N = 1; CPT → FV = FV_5 = 3,850.00

FV_6: PV = 2,000; I/Y = 10; N = 0; CPT → FV = FV_6 = $\underline{2{,}000.00}$

FV of cash flow stream = $\sum FV_{\text{individual}}$ = 8,347.44

Example: Computing PV of an uneven cash flow series

Compute the present value of this 6-year uneven cash flow stream described in Figure 3 using a 10% rate of return.

Answer:

This problem is solved by first computing the PV of each individual cash flow, then summing the PVs of the individual cash flows, which yields the PV of the cash flow stream. Again the signs of the cash flows are preserved.

PV_1: FV = –1,000; I/Y = 10; N = 1; CPT → PV = PV_1 = –909.09

PV_2: FV = –500; I/Y = 10; N = 2; CPT → PV = PV_2 = –413.22

PV_3: FV = 0; I/Y = 10; N = 3; CPT → PV = PV_3 = 0

PV_4: FV = 4,000; I/Y = 10; N = 4; CPT → PV = PV_4 = 2,732.05

PV_5: FV = 3,500; I/Y = 10; N = 5; CPT → PV = PV_5 = 2,173.22

PV_6: FV = 2,000; I/Y = 10; N = 6; CPT → PV = PV_6 = 1,128.95

PV of cash flow stream = $\sum PV_{individual}$ = $4,711.91

Solving TVM Problems When Compounding Periods are Other Than Annual

While the conceptual foundations of TVM calculations are not affected by the compounding period, more frequent compounding does have an impact on FV and PV computations. Specifically, since an increase in the frequency of compounding increases the effective rate of interest, it also *increases* the FV of a given cash flow and *decreases* the PV of a given cash flow.

Example: The effect of compounding frequency on FV and PV

Compute the FV and PV of a $1,000 single sum for an investment horizon of one year using a stated annual interest rate of 6.0% with a range of compounding periods.

Answer:

Figure 4: Compounding Frequency Effect

Compounding Frequency	*Interest Rate per Period*	*Effective Rate of Interest*	*Future Value*	*Present Value*
Annual (m = 1)	6.000%	6.000%	$1,060.00	$943.396
Semiannual (m = 2)	3.000	6.090	1,060.90	942.596
Quarterly (m = 4)	1.500	6.136	1,061.36	942.184
Monthly (m = 12)	0.500	6.168	1,061.68	941.905
Daily (m = 365)	0.016438	6.183	1,061.83	941.769

There are two ways to use your financial calculator to compute PVs and FVs under different compounding frequencies:

1. Adjust the number of periods per year (P/Y) mode on your calculator to correspond to the compounding frequency (e.g., for quarterly, P/Y = 4). **We do not recommend this approach!**
2. Keep the calculator in the annual compounding mode (P/Y = 1) and enter I/Y as the interest rate per compounding period, and *N* as the number of compounding periods in the investment horizon. Letting *m* equal the number of compounding periods per year, the basic formulas for the calculator input data are determined as follows:

    I/Y = the annual interest rate / m

    N = the number of years × m

The computations for the FV and PV amounts in the previous example are:

PV_A: FV = –1,000; I/Y = 6/1 = 6; N = 1 × 1 = 1:
CPT → PV = PV_A = 943.396

PV_S: FV = –1,000; I/Y = 6/2 = 3; N = 1 × 2 = 2:
CPT → PV = PV_S = 942.596

PV_Q: FV = –1,000; I/Y = 6/4 = 1.5; N = 1 × 4 = 4:
CPT → PV = PV_Q = 942.184

PV_M: FV = –1,000; I/Y = 6/12 = 0.5; N = 1 × 12 = 12:
CPT → PV = PV_M = 941.905

PV_D: FV = –1,000; I/Y = 6/365 = 0.016438; N = 1 × 365 = 365:
CPT → PV = PV_D = 941.769

FV_A: PV = –1,000; I/Y = 6/1 = 6; N = 1 × 1 = 1:
CPT → FV = FV_A = 1,060.00

FV_S: PV = –1,000; I/Y = 6/2 = 3; N = 1 × 2 = 2:
CPT → FV = FV_S = 1,060.90

FV_Q: PV = –1,000; I/Y = 6/4 = 1.5; N = 1 × 4 = 4:
CPT → FV = FV_Q =1,061.36

FV_M: PV = –1,000; I/Y = 6/12 = 0.5; N = 1 × 12 = 12:
CPT → FV = FV_M = 1,061.68

FV_D: PV = –1,000; I/Y = 6/365 = 0.016438; N = 1 × 365 = 365:
CPT → FV = FV_D = 1,061.83

Example: FV of a single sum using quarterly compounding

Compute the FV of $2,000 today, five years from today using an interest rate of 12%, compounded quarterly.

Answer:

To solve this problem, enter the relevant data and compute FV:

N = 5 × 4 = 20; I/Y = 12 / 4 = 3; PV = –$2,000; CPT → FV = $3,612.22

Concept Checkers

1. The amount an investor will have in 15 years if $1,000 is invested today at an annual interest rate of 9% will be closest to:
   A. $1,350.
   B. $3,518.
   C. $3,642.
   D. $9,000.

2. How much must be invested today, at 8% interest, to accumulate enough to retire a $10,000 debt due seven years from today? The amount that must be invested today is closest to:
   A. $3,265.
   B. $5,835.
   C. $6,123.
   D. $8,794.

3. An analyst estimates that XYZ's earnings will grow from $3.00 a share to $4.50 per share over the next eight years. The rate of growth in XYZ's earnings is closest to:
   A. 4.9%.
   B. 5.2%.
   C. 6.7%.
   D. 7.0%.

4. If $5,000 is invested in a fund offering a rate of return of 12% per year, approximately how many years will it take for the investment to reach $10,000?
   A. 4 years.
   B. 5 years.
   C. 6 years.
   D. 7 years.

5. An investor is looking at a $150,000 home. If 20% must be put down and the balance is financed at 9% over the next 30 years, what is the monthly mortgage payment?
   A. $652.25.
   B. $799.33.
   C. $895.21.
   D. $965.55.

Concept Checker Answers

1. **C** $N = 15$; $I/Y = 9$; $PV = -1{,}000$; $PMT = 0$; $CPT \rightarrow FV = \$3{,}642.48$

2. **B** $N = 7$; $I/Y = 8$; $FV = -10{,}000$; $PMT = 0$; $CPT \rightarrow PV = \$5{,}834.90$

3. **B** $N = 8$; $PV = -3$; $FV = 4.50$; $PMT = 0$; $CPT \rightarrow I/Y = 5.1989$

4. **C** $PV = -5{,}000$; $I/Y = 12$; $FV = 10{,}000$; $PMT = 0$; $CPT \rightarrow N = 6.12$. Rule of 72 $\rightarrow$ 72/12 = six years.

   Note to HP12C users: One known problem with the HP12C is that it does not have the capability to round. In this particular question, you will come up with 7, although the correct answer is 6.1163.

5. **D** $N = 30 \times 12 = 360$; $I/Y = 9 / 12 = 0.75$; $PV = -150{,}000(1 - 0.2) = -120{,}000$; $FV = 0$; $CPT \rightarrow PMT = \$965.55$

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Probabilities

Topic 15

Exam Focus

This topic covers important terms and concepts associated with probability theory. Random variables, events, outcomes, conditional probability, and joint probability are described. Specifically, we will examine the difference between discrete and continuous probability distributions, the difference between independent and mutually exclusive events, and the difference between unconditional and conditional probabilities. For the exam, be able to calculate probabilities based on the probability functions discussed.

Random Variables

LO 15.1: Describe and distinguish between continuous and discrete random variables.

- A **random variable** is an uncertain quantity/number.
- An **outcome** is an observed value of a random variable.
- An **event** is a single outcome or a set of outcomes.
- **Mutually exclusive events** are events that cannot happen at the same time.
- **Exhaustive events** are those that include all possible outcomes.

Consider rolling a 6-sided die. The number that comes up is a *random variable*. If you roll a 4, that is an *outcome*. Rolling a 4 is an event, and rolling an even number is an *event*. Rolling a 4 and rolling a 6 are *mutually exclusive events*. Rolling an even number and rolling an odd number is a set of mutually exclusive and *exhaustive events*.

A **probability distribution** describes the probabilities of all the possible outcomes for a random variable. The probabilities of all possible outcomes must sum to 1. A simple probability distribution is that for the roll of one fair die there are six possible outcomes and each one has a probability of 1/6, so they sum to 1. The probability distribution of all the possible returns on the S&P 500 Index for the next year is a more complex version of the same idea.

A **discrete random variable** is one for which the number of possible outcomes can be counted, and for each possible outcome, there is a measurable and positive probability. An example of a discrete random variable is the number of days it rains in a given month because there is a finite number of possible outcomes—the number of days it can rain in a month is defined by the number of days in the month.

A **probability function**, denoted p(x), specifies the probability that a random variable is equal to a specific value. More formally, p(x) is the probability that random variable X takes on the value x, or p(x) = P(X = x).

The two key properties of a probability function are:

- $0 \leq p(x) \leq 1$.
- $\Sigma p(x) = 1$, the sum of the probabilities for *all* possible outcomes, *x*, for a random variable, *X*, equals 1.

Example: Evaluating a probability function

Consider the following function: $X = \{1, 2, 3, 4\}$, $p(x) = \frac{x}{10}$, else $p(x) = 0$

Determine whether this function satisfies the conditions for a probability function.

Answer:

Note that all of the probabilities are between 0 and 1, and the sum of all probabilities equals 1:

$$\sum p(x) = \frac{1}{10} + \frac{2}{10} + \frac{3}{10} + \frac{4}{10} = 0.1 + 0.2 + 0.3 + 0.4 = 1$$

Both conditions for a probability function are satisfied.

A **continuous random variable** is one for which the number of possible outcomes is infinite, even if lower and upper bounds exist. The actual amount of daily rainfall between zero and 100 inches is an example of a continuous random variable because the actual amount of rainfall can take on an infinite number of values. Daily rainfall can be measured in inches, half inches, quarter inches, thousandths of inches, or even smaller increments. Thus, the number of possible daily rainfall amounts between zero and 100 inches is essentially infinite.

The assignment of probabilities to the possible outcomes for discrete and continuous random variables provides us with discrete probability distributions and continuous probability distributions. The difference between these types of distributions is most apparent for the following properties:

- For a *discrete distribution*, $p(x) = 0$ when *x* cannot occur, or $p(x) > 0$ if it can. Recall that $p(x)$ is read: "the probability that random variable X = x." For example, the probability of it raining 33 days in June is zero because this cannot occur, but the probability of it raining 25 days in June has some positive value.
- For a *continuous distribution*, $p(x) = 0$ even though *x* can occur. We can only consider $P(x_1 \leq X \leq x_2)$ where x_1 and x_2 are actual numbers. For example, the probability of receiving two inches of rain in June is zero because two inches is a single point in an infinite range of possible values. On the other hand, the probability of the amount of rain being between 1.99999999 and 2.00000001 inches has some positive value. In the case of continuous distributions, $P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2)$ because $p(x_1) = p(x_2) = 0$.

In finance, some discrete distributions are treated as though they are continuous because the number of possible outcomes is very large. For example, the increase or decrease in the price of a stock traded on an American exchange is recorded in dollars and cents. Yet, the probability of a change of exactly \$1.33 or \$1.34 or any other specific change is almost zero. It is customary, therefore, to speak in terms of the probability of a range of possible price change, say between \$1.00 and \$2.00. In other words p(price change = 1.33) is essentially zero, but p(\$1 < price change < \$2) is greater than zero.

Distribution Functions

LO 15.2: Define and distinguish between the probability density function, the cumulative distribution function, and the inverse cumulative distribution function.

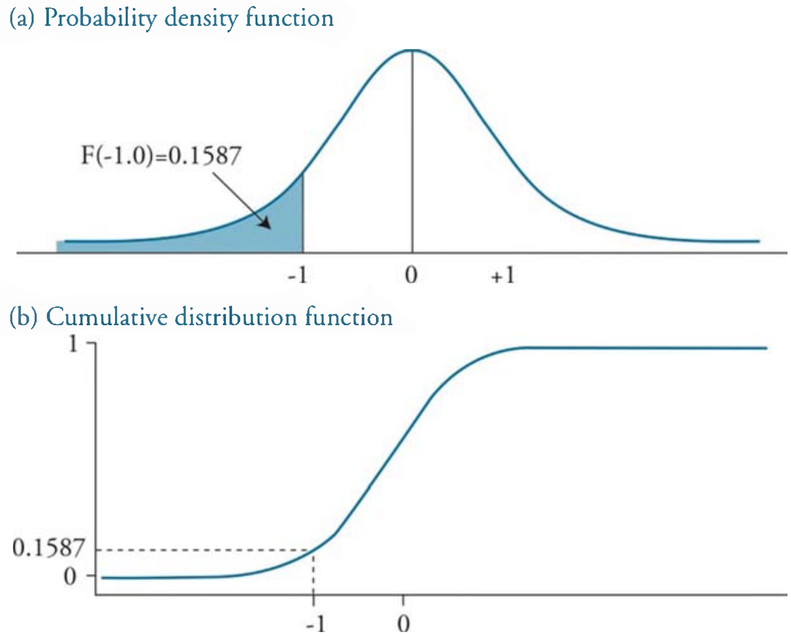
A **probability density function** (pdf) is a function, denoted f(x), that can be used to generate the probability that outcomes of a continuous distribution lie within a particular range of outcomes. For a continuous distribution, it is the equivalent of a *probability function* for a discrete distribution. Know that for a continuous distribution, the probability of any one particular outcome (of the infinite possible outcomes) is zero (e.g., the probability of receiving exactly two inches of rain in June is zero because two inches is a single point in an infinite range of possible values). A pdf is used to calculate the probability of an outcome between two values (i.e., the probability of the outcome falling within a specified range).

A **cumulative distribution function** (cdf), or simply *distribution function*, defines the probability that a random variable, *X*, takes on a value equal to or less than a specific value, *x*. It represents the sum, or *cumulative value*, of the probabilities for the outcomes up to and including a specified outcome. The cumulative distribution function for a random variable, *X*, may be expressed as $F(x) = P(X \leq x)$.

Consider the probability function defined earlier for X = {1, 2, 3, 4}, p(x) = x / 10. For this distribution, F(3) = 0.6 = 0.1 + 0.2 + 0.3, and F(4) = 1 = 0.1 + 0.2 + 0.3 + 0.4. This means that F(3) is the cumulative probability that outcomes 1, 2, or 3 occur, and F(4) is the cumulative probability that one of the possible outcomes occurs.

Figure 1 shows an example of a cumulative distribution function (for a standard normal distribution, described in Topic 17). There is a 15.87% probability of a value less than –1. This is the total area to the left of –1 in the pdf in Panel (a), and the y-axis value of the cdf for a value of –1 in Panel (b).

Figure 1: Standard Normal Probability Density and Cumulative Distribution Functions



Instead of finding the probability less than or equal to a specific value, x, the **inverse cumulative distribution function** can be used to find the value that corresponds to a specific probability. For example, it may be useful to know the value, x, where 15.87% of the distribution is less than or equal to x. From Figure 1, this value would be –1.

Consider a cumulative distribution function, $F(x) = p = x^2 / 25$, where $0 \leq x \leq 5$. F(3) finds the probability less than or equal to 3. In this case, $F(3) = 3^2 / 25 = 36\%$. The inverse function rearranges this cumulative function to instead input a probability and solve for x. Thus, the inverse cumulative distribution function in this example is: $F^{-1}(p) = x = 5\sqrt{p}$.

We can check the accuracy of this inverse function by testing the limits of the distribution ($0 \leq x \leq 5$). At p = 0, the minimum value is equal to 0, and at p = 1, the maximum value is equal to 5. By inputting a probability of 36% into the inverse function, we again see that 36% of the distribution is less than or equal to 3: $F^{-1}(0.36) = x = 5\sqrt{0.36} = 3$.

Discrete Probability Function

LO 15.3: Calculate the probability of an event given a discrete probability function.

A **discrete uniform random variable** is one for which the probabilities for all possible outcomes for a discrete random variable are equal. For example, consider the *discrete uniform probability distribution* defined as X = {1, 2, 3, 4, 5}, p(x) = 0.2. Here, the probability for each outcome is equal to 0.2 [i.e., p(1) = p(2) = p(3) = p(4) = p(5) = 0.2]. Also, the cumulative distribution function for the *n*th outcome, $F(x_n) = np(x)$, and the probability for a range of outcomes is p(x)k, where *k* is the number of possible outcomes in the range.

Example: Discrete uniform distribution

Determine p(6), F(6), and P(2 ≤ X ≤ 8) for the discrete uniform distribution function defined as:

$$X = \{2, 4, 6, 8, 10\},\ p(x) = 0.2$$

Answer:

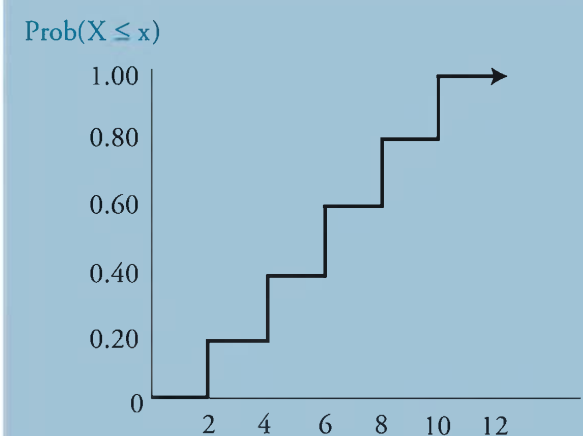
p(6) = 0.2, since p(x) = 0.2 for all *x*. F(6) = P(X ≤ 6) = np(x) = 3(0.2) = 0.6. Note that n = 3 since 6 is the third outcome in the range of possible outcomes.

P(2 ≤ X ≤ 8) = 4(0.2) = 0.8. Note that k = 4, since there are four outcomes in the range 2 ≤ X ≤ 8. The following figures illustrate the concepts of a probability function and cumulative distribution function for this distribution.

Probability and Cumulative Distribution Functions

X = x	*Probability of x* *Prob (X = x)*	*Cumulative Distribution Function* *Prob (X < x)*
2	0.20	0.20
4	0.20	0.40
6	0.20	0.60
8	0.20	0.80

Cumulative Distribution Function for X ~ Uniform {2, 4, 6, 8, 10}



Conditional Probabilities

LO 15.6: Define and calculate a conditional probability, and distinguish between conditional and unconditional probabilities.

As noted earlier, there are two defining properties of probability:

- The probability of occurrence of any event (E_i) is between 0 and 1 (i.e., $0 \leq P(E_i) \leq 1$).
- If a set of events, E_1, E_2, ... E_n, is mutually exclusive and exhaustive, the probabilities of those events sum to 1 (i.e., $\Sigma P(E_i) = 1$).

The first of the defining properties introduces the term $P(E_i)$, which is shorthand for the "probability of event *i*." If $P(E_i) = 0$, the event will never happen. If $P(E_i) = 1$, the event is certain to occur, and the outcome is not random.

The probability of rolling any one of the numbers 1–6 with a fair die is 1/6 = 0.1667 = 16.7%. The set of events—rolling a number equal to 1, 2, 3, 4, 5, or 6—is exhaustive, and the individual events are mutually exclusive, so the probability of this set of events is equal to 1. We are certain that one of the values in this set of events will occur.

Unconditional probability (i.e., *marginal probability*) refers to the probability of an event regardless of the past or future occurrence of other events. If we are concerned with the probability of an economic recession, regardless of the occurrence of changes in interest rates or inflation, we are concerned with the unconditional probability of a recession.

A **conditional probability** is one where the occurrence of one event affects the probability of the occurrence of another event. For example, we might be concerned with the probability of a recession *given* that the monetary authority increases interest rates. This is a conditional probability. The key word to watch for here is "given." Using probability notation, "the probability of A *given* the occurrence of B" is expressed as $P(A \mid B)$, where the vertical bar (|) indicates "given," or "conditional upon." For example, the probability of a recession *given* an increase in interest rates is expressed as P(*recession* | *increase in interest rates*). A conditional probability of an occurrence is also called its likelihood.

The **joint probability** of two events is the probability that they will both occur. We can calculate this from the conditional probability that A will occur given B occurs (a conditional probability) and the probability that B will occur (the unconditional probability of B). This calculation is sometimes referred to as the *multiplication rule of probability*. Using the notation for conditional and unconditional probabilities, we can express this rule as:

$$P(AB) = P(A \mid B) \times P(B)$$

This expression is read as follows: "The joint probability of A and B, P(AB), is equal to the conditional probability of A *given* B, $P(A \mid B)$, times the unconditional probability of B, P(B)."

This relationship can be rearranged to define the conditional probability of A given B as follows:

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

Example: Multiplication rule of probability

Consider the following information:

- P(I) = 0.4, the probability of the monetary authority increasing interest rates (I) is 40%.
- P(R | I) = 0.7, the probability of a recession (R) given an increase in interest rates is 70%.

What is P(RI), the joint probability of a recession *and* an increase in interest rates?

Answer:

Applying the multiplication rule, we get the following result:

$$P(RI) = P(R \mid I) \times P(I)$$

$$P(RI) = 0.7 \times 0.4$$

$$P(RI) = 0.28$$

Don't let the cumbersome notation obscure the simple logic of this result. If an interest rate increase will occur 40% of the time and lead to a recession 70% of the time when it occurs, the joint probability of an interest rate increase and a resulting recession is (0.4)(0.7) = (0.28) = 28%.

Independent and Mutually Exclusive Events

LO 15.4: Distinguish between independent and mutually exclusive events.

Independent events refer to events for which the occurrence of one has no influence on the occurrence of the others. The definition of independent events can be expressed in terms of conditional probabilities. Events A and B are independent if and only if:

$$P(A \mid B) = P(A), \text{ or equivalently, } P(B \mid A) = P(B)$$

If this condition is not satisfied, the events are **dependent events** (i.e., the occurrence of one is dependent on the occurrence of the other).

In our interest rate and recession example, recall that events I and R are not independent; the occurrence of I affects the probability of the occurrence of R. In this example, the independence conditions for I and R are violated because:

P(R) = 0.34, but P(R | I) = 0.7; the probability of a recession is greater when there is an increase in interest rates.

The best examples of independent events are found with the probabilities of dice tosses or coin flips. A die has "no memory." Therefore, the event of rolling a 4 on the second toss is independent of rolling a 4 on the first toss. This idea may be expressed as:

P(4 on second toss | 4 on first toss) = P(4 on second toss) = 1/6 or 0.167

The idea of independent events also applies to flips of a coin:

P(heads on first coin | heads on second coin) = P(heads on first coin) = 1/2 or 0.50

Calculating the Probability That at Least One of Two Events Will Occur

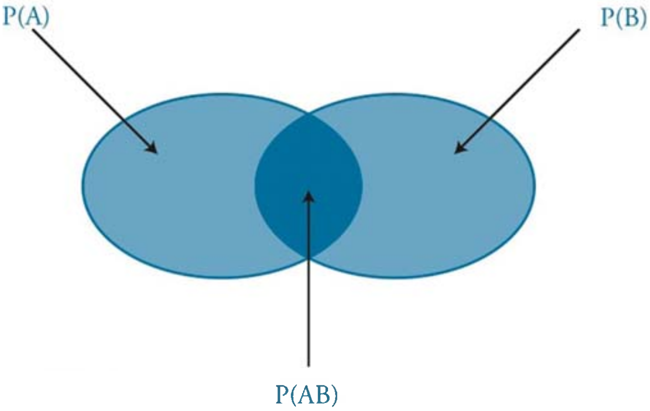
The *addition rule for probabilities* is used to determine the probability that at least one of two events will occur. For example, given two events, A and B, the addition rule can be used to determine the probability that either A or B will occur. If the events are *not* mutually exclusive, double counting must be avoided by subtracting the joint probability that *both* A and B will occur from the sum of the unconditional probabilities. This is reflected in the following general expression for the addition rule:

P(A or B) = P(A) + P(B) – P(AB)

For **mutually exclusive events** where the joint probability, P(AB), is zero, the probability that either A or B will occur is simply the sum of the unconditional probabilities for each event, P(A or B) = P(A) + P(B).

Figure 2 illustrates the addition rule with a Venn diagram and highlights why the joint probability must be subtracted from the sum of the unconditional probabilities. Note that if the events are mutually exclusive the sets do not intersect, P(AB) = 0, and the probability that one of the two events will occur is simply P(A) + P(B).

Figure 2: Venn Diagram for Events That Are Not Mutually Exclusive



Example: Addition rule of probability

Using the information in our previous interest rate and recession example and the fact that the unconditional probability of a recession, P(R), is 34%, **determine** the probability that either interest rates will increase *or* a recession will occur.

Answer:

Given that P(R) = 0.34, P(I) = 0.40, and P(RI) = 0.28, we can compute P(R or I) as follows:

$$P(R \text{ or } I) = P(R) + P(I) - P(RI)$$

$$P(R \text{ or } I) = 0.34 + 0.40 - 0.28$$

$$P(R \text{ or } I) = 0.46$$

Calculating a Joint Probability of Any Number of Independent Events

LO 15.5: Define joint probability, describe a probability matrix, and calculate joint probabilities using probability matrices.

On the roll of two dice, the joint probability of getting two 4s is calculated as:

$$P(4 \text{ on first die and } 4 \text{ on second die}) = P(4 \text{ on first die}) \times P(4 \text{ on second die}) = 1/6 \times 1/6 = 1/36 = 0.0278$$

On the flip of two coins, the probability of getting two heads is:

$$P(\text{heads on first coin and heads on second coin}) = 1/2 \times 1/2 = 1/4 = 0.25$$

Hint: When dealing with *independent events*, the word *and* indicates multiplication, and the word *or* indicates addition. In probability notation:

$$P(A \text{ or } B) = P(A) + P(B), \text{ and } P(A \text{ and } B) = P(A) \times P(B)$$

Professor's Note: On the exam, you may see A and B represented as $A \cap B$. This notation means "the intersection of A and B" and refers to the event "both A and B." Similarly, you may see A or B represented as $A \cup B$, which is "the union of A and B" and refers to the event "either A or B or both."

The multiplication rule we used to calculate the joint probability of two independent events may be applied to any number of independent events, as the following examples illustrate.

Example: Joint probability for more than two independent events (1)

What is the probability of rolling three 4s in one simultaneous toss of three dice?

Answer:

Since the probability of rolling a 4 for each die is 1/6, the probability of rolling three 4s is:

$$P(\text{three 4s on the roll of three dice}) = 1/6 \times 1/6 \times 1/6 = 1/216 = 0.00463$$

Similarly:

$$P(\text{four heads on the flip of four coins}) = 1/2 \times 1/2 \times 1/2 \times 1/2 = 1/16 = 0.0625$$

Example: Joint probability for more than two independent events (2)

Using empirical probabilities, suppose we observe that the DJIA has closed higher on two-thirds of all days in the past few decades. Furthermore, it has been determined that up and down days are independent. Based on this information, **compute** the probability of the DJIA closing higher for five consecutive days.

Answer:

$$P(\text{DJIA up five days in a row}) = 2/3 \times 2/3 \times 2/3 \times 2/3 \times 2/3 = (2/3)^5 = 0.132$$

Similarly:

$$P(\text{DJIA down five days in a row}) = 1/3 \times 1/3 \times 1/3 \times 1/3 \times 1/3 = (1/3)^5 = 0.004$$

Probability Matrix

Joint probabilities of independent events can be conveniently summarized using a probability matrix (sometimes known as a probability table). Suppose, for example, that we wanted to view how the state of the economy relates to the direction of interest rates. The probability matrix in Figure 3 shows the joint and unconditional probabilities of these two variables.

Figure 3: Joint and Unconditional Probabilities

		Interest Rates		
		Increase	No Increase	
Economy	Good	14%	6%	**20%**
	Normal	20%	30%	**50%**
	Poor	6%	24%	**30%**
		40%	**60%**	**100%**

From this probability matrix, we see that the joint probability of a poor economy and an increase in interest rates is 6%. Similarly, the joint probability of a normal economy and no increase in interest rates is 30%. Unconditional probabilities are shown as the sum of each column and each row. For example, the unconditional probability of a rate increase, irrespective of the state of the economy, is the sum of the joint probabilities, 14% + 20% + 6% = 40%. Also, the sum of all joint probabilities is equal to 100%, since one of these events must happen.

Example: Calculating joint probabilities using a probability matrix

Given the following incomplete probability matrix, **calculate** the joint probability of a normal economy and an increase in rates, and the unconditional probability of a good economy.

		Interest Rates		
		Increase	No Increase	
Economy	Good	15%	X2	**X3**
	Normal	X1	25%	**X4**
	Poor	10%	20%	**30%**
		50%	**50%**	**100%**

Answer:

Since the unconditional probability of an increase in rates, irrespective of the state of the economy, is 50%, we know the sum of each joint probability in the first column must equal 50%. By solving for X1, we find the joint probability of a normal economy and an increase in rates:

$$15\% + X1 + 10\% = 50\%$$

$$X1 = 50\% - 15\% - 10\% = 25\%$$

The unconditional probability of a good economy, X3, can be computed by first solving for X2 (the joint probability of a good economy and no increase in interest rates) and then summing both joint probabilities in the first row.

$$X2 + 25\% + 20\% = 50\%$$

$$X2 = 50\% - 25\% - 20\% = 5\%$$

$$X3 = 15\% + X2 = 15\% + 5\% = 20\%$$

Key Concepts

LO 15.1

A discrete random variable has positive probabilities associated with a finite number of outcomes.

A continuous random variable has positive probabilities associated with a range of outcome values—the probability of any single value is zero.

LO 15.2

A probability function specifies the probability that a random variable is equal to a specific value; $P(X = x) = p(x)$.

A probability density function (pdf) is the expression for a probability function for a continuous random variable.

A cumulative distribution function (cdf) gives the probability of the random variable being equal to or less than each specific value. It is the area under the probability distribution to the left of a specified value.

LO 15.3

A discrete uniform distribution is one where there are *n* discrete, equally likely outcomes, so that for each outcome $p(x) = 1/n$.

LO 15.4

The probability of an independent event is unaffected by the occurrence of other events, but the probability of a dependent event is changed by the occurrence of another event.

Events A and B are independent if and only if:

$$P(A \mid B) = P(A), \text{ or equivalently, } P(B \mid A) = P(B)$$

The probability that at least one of two events will occur is $P(A \text{ or } B) = P(A) + P(B) - P(AB)$. For mutually exclusive events, $P(A \text{ or } B) = P(A) + P(B)$, since $P(AB) = 0$.

LO 15.5

The joint probability of two events, $P(AB)$, is the probability that they will both occur. $P(AB) = P(A \mid B) \times P(B)$. For independent events, $P(A \mid B) = P(A)$, so that $P(AB) = P(A) \times P(B)$.

LO 15.6

Unconditional probability (marginal probability) is the probability of an event occurring.

Conditional probability, $P(A \mid B)$, is the probability of an event A occurring given that event B has occurred.

Concept Checkers

1. If events A and B are mutually exclusive, then:
   A. P(A | B) = P(A).
   B. P(A | B) = P(B).
   C. P(AB) = P(A) × P(B).
   D. P(A or B) = P(A) + P(B).

2. At a charity ball, 800 names were put into a hat. Four of the names are identical. On a random draw, what is the probability that one of these four names will be drawn?
   A. 0.004.
   B. 0.005.
   C. 0.010.
   D. 0.025.

3. Two events are said to be independent if the occurrence of one event:
   A. means the second event cannot occur.
   B. means the second event is certain to occur.
   C. affects the probability of the occurrence of the other event.
   D. does not affect the probability of the occurrence of the other event.

4. For a continuous random variable *X*, the probability of any single value of *X* is:
   A. one.
   B. zero.
   C. determined by the cdf.
   D. determined by the pdf.

5. Given the following incomplete probability matrix, what is the joint probability of a good economy and no increase in interest rates?

		Interest Rates		
		Increase	No Increase	
Economy	Good	20%	A	**B**
	Normal	C	20%	**D**
	Poor	10%	E	**20%**
		60%	**40%**	**100%**

   A. 0%.
   B. 10%.
   C. 20%.
   D. 30%.

Concept Checker Answers

1. **D** There is no intersection of events when events are mutually exclusive. P(AB) = P(A) × P(B) is only true for independent events. Note that since A and B are mutually exclusive (cannot both happen), P(A | B) and P(AB) must both be equal to zero, making answers A, B, and C incorrect.

2. **B** P(name 1 or name 2 or name 3 or name 4) = 1/800 + 1/800 + 1/800 + 1/800 = 4/800 = 0.005

3. **D** Two events are said to be independent if the occurrence of one event does not affect the probability of the occurrence of the other event.

4. **B** For a continuous distribution p(x) = 0 for all *X*; only ranges of value of *X* have positive probabilities.

5. **B** Because the unconditional probability of a poor economy, irrespective of interest rates, is 20%, we know that the sum of each joint probability in the poor economy row must equal 20%. By solving for *E*, we find the joint probability of a poor economy and no increase in rates:

   10% + E = 20%

   E = 20% – 10% = 10%

   The joint probability of a good economy and no increase in interest rates, *A*, can be computed by subtracting the joint probability of a normal economy and no increase in rates and the joint probability of a poor economy and no increase in rates from the unconditional probability of no increase in interest rates.

   A = 40% – 20% – E

   A = 40% – 20% – 10% = 10%

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Basic Statistics

Topic 16

Exam Focus

This topic addresses the concepts of expected value, variance, standard deviation, covariance, correlation, skewness, and kurtosis. The characteristics and calculations of these measures will be discussed. For the exam, be able to calculate the mean and variance of a random variable, and the covariance and correlation between two random variables. Also, be able to identify and interpret the first four moments of a statistical distribution.

The word *statistics* is used to refer to data (e.g., the average return on XYZ stock was 8% over the last ten years) and the methods we use to analyze data. Statistical methods fall into one of two categories, descriptive statistics or inferential statistics.

Descriptive statistics are used to summarize the important characteristics of large data sets. The focus of this topic is on the use of descriptive statistics to consolidate a mass of numerical data into useful information.

Inferential statistics, which will be discussed in subsequent topics, pertain to the procedures used to make forecasts, estimates, or judgments about a large set of data on the basis of the statistical characteristics of a smaller set (a sample).

A *population* is defined as the set of all possible members of a stated group. A cross-section of the returns of all of the stocks traded on the New York Stock Exchange (NYSE) is an example of a population.

It is frequently too costly or time consuming to obtain measurements for every member of a population, if it is even possible. In this case, a sample may be used. A sample is defined as a subset of the population of interest. Once a population has been defined, a sample can be drawn from the population, and the sample's characteristics can be used to describe the population as a whole. For example, a sample of 30 stocks may be selected from all of the stocks listed on the NYSE to represent the population of all NYSE-traded stocks.

Measures of Central Tendency

LO 16.1: Interpret and apply the mean, standard deviation, and variance of a random variable.

LO 16.2: Calculate the mean, standard deviation, and variance of a discrete random variable.

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

To compute the **population mean**, all the observed values in the population are summed (ΣX) and divided by the number of observations in the population, *N*. Note that the population mean is unique in that a given population only has one mean. The population mean is expressed as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

The **sample mean** 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:

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

Note the use of *n*, the sample size, versus *N*, the population size.

Example: Population mean and sample mean

Assume you and your research assistant are evaluating the stock of AXZ Corporation. You have calculated the stock returns for AXZ over the last 12 years to develop the following data set. Your research assistant has decided to conduct his analysis using only the returns for the five most recent years, which are displayed as the bold numbers in the data set. Given this information, **calculate** the population mean and the sample mean.

Data set: 12%, **25%**, **34%**, 15%, **19%**, 44%, **54%**, 33%, 22%, 28%, **17%**, 24%

Answer:

$$\mu = \text{population mean} = \frac{12+25+34+15+19+44+54+33+22+28+17+24}{12}$$
$$= 27.25\%$$

$$\overline{X} = \text{sample mean} = \frac{25+34+19+54+17}{5} = 29.8\%$$

The population mean and sample mean are both examples of arithmetic means. 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 and has the following properties:

- All interval and ratio data sets have an arithmetic mean.
- All data values are considered and included in the arithmetic mean computation.
- A data set has only one arithmetic mean (i.e., the arithmetic mean is unique).
- The sum of the deviations of each observation in the data set from the mean is always zero.

The arithmetic mean is the only measure of central tendency for which the sum of the deviations from the mean is zero. Mathematically, this property can be expressed as follows:

$$\text{sum of mean deviations} = \sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

Example: Arithmetic mean and deviations from the mean

Compute the arithmetic mean for a data set described as:

Data set: [5, 9, 4, 10]

Answer:

The arithmetic mean of these numbers is:

$$\overline{X} = \frac{5+9+4+10}{4} = 7$$

The sum of the deviations from the mean (of 7) is:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = (5-7)+(9-7)+(4-7)+(10-7) = -2 \; +2-3+3=0$$

Unusually large or small values can have a disproportionate effect on the computed value for the arithmetic mean. The mean of 1, 2, 3, and 50 is 14 and is not a good indication of what the individual data values really are. On the positive side, the arithmetic mean uses all the information available about the observations. The arithmetic mean of a sample from a population is the best estimate of both the true mean of the sample and the value of the next observation.

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

The median is important because the arithmetic mean can be affected by extremely large or small values (outliers). 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 10-year annualized total returns 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 data set, 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:

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% = 0.5(25 + 23).

Consider that while we calculated the mean of 1, 2, 3, and 50 as 14, the median is 2.5. If the data were 1, 2, 3, and 4 instead, the arithmetic mean and median would both be 2.5.

The **mode** is the value that occurs most frequently in a data set. A data set 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 set of data 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 data set?

Data set: [30%, 28%, 25%, 23%, 28%, 15%, 5%]

Answer:

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

The **geometric mean** is often used when calculating investment returns over multiple periods or when measuring compound growth rates. The general formula for the geometric mean, *G*, is as follows:

$$G = \sqrt[n]{X_1 \times X_2 \times \ldots \times X_n} = (X_1 \times X_2 \times \ldots \times X_n)^{1/n}$$

Note that this equation has a solution only if the product under the radical sign is non-negative.

When calculating the geometric mean for a returns data set, it is necessary to add 1 to each value under the radical and then subtract 1 from the result. The geometric mean return (R_G) can be computed using the following equation:

$$1 + R_G = \sqrt[n]{(1+R_1) \times (1+R_2) \times \ldots \times (1+R_n)}$$

where:
R_t = the return for period t

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%. **Compute** the compound annual rate of return over the 3-year period.

Answer:

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

$$1 + R_G = \sqrt[3]{0.9066 \times 1.2345 \times 1.0892} = \sqrt[3]{1.21903} = (1.21903)^{1/3} = 1.06825$$

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

Solve this type of problem with your calculator as follows:

- On the TI, enter 1.21903 [y^x] 0.33333 [=], or 1.21903 [y^x] 3 [1/x] [=]
- On the HP, enter 1.21903 [ENTER] 0.33333 [y^x], or 1.21903 [ENTER] 3 [1/x] [y^x]

Note that the 0.33333 represents the one-third power.

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).

Expectations

LO 16.3: Interpret and calculate the expected value of a discrete random variable.

LO 16.5: Calculate the mean and variance of sums of variables.

The **expected value** is the weighted average of the possible outcomes of a random variable, where the weights are the probabilities that the outcomes will occur. The mathematical representation for the expected value of random variable *X* is:

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

Here, *E* is referred to as the expectations operator and is used to indicate the computation of a probability-weighted average. The symbol x_1 represents the first observed value (observation) for random variable *X*; x_2 is the second observation, and so on through the *n*th observation. The concept of expected value may be demonstrated using probabilities associated with a coin toss. On the flip of one coin, the occurrence of the event "heads" may be used to assign the value of one to a random variable. Alternatively, the event "tails" means the random variable equals zero. Statistically, we would formally write:

if heads, then X = 1

if tails, then X = 0

For a fair coin, P(heads) = P(X = 1) = 0.5, and P(tails) = P(X = 0) = 0.5. The expected value can be computed as follows:

$$E(X) = \Sigma P(x_i)x_i = P(X = 0)(0) + P(X = 1)(1) = (0.5)(0) + (0.5)(1) = 0.5$$

In any individual flip of a coin, *X* cannot assume a value of 0.5. Over the long term, however, the average of all the outcomes is expected to be 0.5. Similarly, the expected value of the roll of a fair die, where X = number that faces up on the die, is determined to be:

$$E(X) = \Sigma P(x_i)x_i = (1/6)(1) + (1/6)(2) + (1/6)(3) + (1/6)(4) + (1/6)(5) + (1/6)(6)$$

$$E(X) = 3.5$$

We can never roll a 3.5 on a die, but over the long term, 3.5 should be the average value of all outcomes.

The expected value is, statistically speaking, our "best guess" of the outcome of a random variable. While a 3.5 will never appear when a die is rolled, the average amount by which our guess differs from the actual outcomes is minimized when we use the expected value calculated this way.

Professor's Note: When we had historical data earlier, we calculated the mean or simple arithmetic average. The calculations given here for the expected value (or weighted mean) are based on probability models, whereas our earlier calculations were based on samples or populations of outcomes. Note that when the probabilities are equal, the simple mean is the expected value. For the roll of a die, all six outcomes are equally likely, so $\frac{1+2+3+4+5+6}{6} = 3.5$ *gives us the same expected value as the probability model. However, with a probability model, the probabilities of the possible outcomes need not be equal, and the simple mean is not necessarily the expected outcome, as the following example illustrates.*

Example: Expected earnings per share

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

EPS Probability Distribution

Probability	*Earnings Per Share*
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.

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

Properties of expectation include:

1. If c is any constant, then:

   $$E(cX) = cE(X)$$

2. If X and Y are any random variables, then:

   $$E(X + Y) = E(X) + E(Y)$$

Professor's Note: This property displays the mean of the sum of random variables. It is simply the sum of the individual random variable means.

3. If c and a are constants, then:

   $$E(cX + a) = cE(X) + a$$

4. If X and Y are independent random variables, then:

   $$E(XY) = E(X) \times E(Y)$$

5. If X and Y are NOT independent, then:

   $$E(XY) \neq E(X) \times E(Y)$$

6. If X is a random variable, then:

   $$E(X^2) \neq [E(X)]^2$$

Variance and Standard Deviation

The mean and variance of a distribution are defined as the first and second moments of the distribution, respectively. **Variance** is defined as:

$$\text{Var}(X) = E[(X - \mu)^2]$$

The square root of the variance is called the **standard deviation**. The variance and standard deviation provide a measure of the extent of the dispersion in the values of the random variable around the mean.

Properties of variance include:

1. $\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$

   where $\mu = E(X)$

2. If *c* is any constant, then:

   $\text{Var}(c) = 0$

3. If *c* is any constant, then:

   $\text{Var}(cX) = c^2 \times \text{Var}(X)$

4. If *c* is any constant, then:

   $\text{Var}(X + c) = \text{Var}(X)$

5. If *a* and *c* are constants, then:

   $\text{Var}(aX + c) = a^2 \times \text{Var}(X)$

6. If *X* and *Y* are independent random variables, then:

   $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

   $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$

7. If *X* and *Y* are independent and *a* and *c* are constants, then:

   $\text{Var}(aX + cY) = a^2 \times \text{Var}(X) + c^2 \times \text{Var}(Y)$

Example: Computing variance and standard deviation

What is the variance and standard deviation of the sum of points in tossing a single coin if heads = 2 points and tails = 10 points?

Answer:

$$\mu = (2 + 10) / 2 = 6$$

$$\text{Var}(X) = (2 - 6)^2 \times 0.5 + (10 - 6)^2 \times 0.5$$

$$\text{Var}(X) = 8 + 8 = 16$$

$$\text{standard deviation}(X) = \sqrt{16} = 4$$

Covariance and Correlation

LO 16.4: Calculate and interpret the covariance and correlation between two random variables.

The variance and standard deviation measure the dispersion, or volatility, of only one variable. In many finance situations, however, 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?" As you will soon see, the covariance provides useful information about how two random variables, such as asset returns, are related.

Covariance 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). Since 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:

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

This equation simplifies to:

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

Properties of covariance include:

1. If X and Y are independent random variables, then:

   $$\text{Cov}(X,Y) = 0$$

2. The covariance of random variable X with itself is the variance of X.

   $$\text{Cov}(X,X) = \text{Var}(X)$$

3. If a, b, c, and d are constants, then:

   $$\text{Cov}(a + bX, c + dY) = b \times d \times \text{Cov}(X,Y)$$

4. If X and Y are NOT independent, then:

   $$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \times \text{Cov}(X,Y)$$

   $$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2 \times \text{Cov}(X,Y)$$

Professor's Note: When discussing the properties of variance, we showed the variance of the sum of independent random variable variances. The covariance term was not present in this earlier expression because the variables did not influence each other. However, when random variables are not independent, two times the covariance of the random variables must be included as demonstrated in the fourth property listed.

To aid in the interpretation of covariance, consider the returns of a stock and of a put option on the stock. These two returns will have a negative covariance because they move in opposite directions. The returns of two automotive stocks would likely have a positive covariance, and the returns of a stock and a riskless asset would have a zero covariance because the riskless asset's returns never move, regardless of movements in the stock's return.

Example: Covariance

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 Stock A, R_A, and Stock B, R_B, under each of the economic states are provided in the following table. What is the covariance of the returns for Stock A and Stock B?

Answer:

First, the expected returns for each of the stocks must be determined.

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

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

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

Covariance Computation

Event	*P(S)*	R_A	R_B	$P(S) \times [R_A - E(R_A)] \times [R_B - E(R_B)]$
Boom	0.3	0.20	0.30	$(0.3)(0.2 - 0.13)(0.3 - 0.14) = 0.00336$
Normal	0.5	0.12	0.10	$(0.5)(0.12 - 0.13)(0.1 - 0.14) = 0.00020$
Slow	0.2	0.05	0.00	$(0.2)(0.05 - 0.13)(0 - 0.14) = 0.00224$
		$\text{Cov}(R_A, R_B) = \Sigma P(S) \times [R_A - E(R_A)] \times [R_B - E(R_B)]$		$= 0.00580$

In practice, the covariance is difficult to interpret. This is mostly because it can take on extremely large values, ranging from negative to positive infinity, and, like the variance, these values are expressed in terms of squared units.

To make the covariance of two random variables easier to interpret, it may be divided by the product of the random variables' standard deviations. The resulting value is called the correlation coefficient, or simply, **correlation**. The relationship between covariances, standard deviations, and correlations can be seen in the following expression for the correlation of the returns for asset *i* and *j*:

$$\text{Corr}(R_i, R_j) = \frac{\text{Cov}(R_i, R_j)}{\sigma(R_i)\sigma(R_j)}, \text{ which implies } \text{Cov}(R_i, R_j) = \text{Corr}(R_i, R_j)\sigma(R_i)\sigma(R_j)$$

The correlation between two random return variables may also be expressed as $\rho(R_i, R_j)$, or $\rho_{i,j}$.

Properties of correlation of two random variables R_i and R_j 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, $-1 \leq \text{Corr}(R_i, R_j) \leq +1$.
- If $\text{Corr}(R_i, R_j) = 1.0$, 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 $Corr(R_i, R_j) = -1.0$, 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 $Corr(R_i, R_j) = 0$, there is no linear relationship between the variables, indicating that prediction of R_i cannot be made on the basis of R_j using linear methods.

Example: Correlation

Using our previous example, **compute** and **interpret** the correlation of the returns for stocks A and B, given that $\sigma^2(R_A) = 0.0028$ and $\sigma^2(R_B) = 0.0124$ and recalling that $Cov(R_A,R_B) = 0.0058$.

Answer:

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

$$\sigma(R_A) = (0.0028)^{1/2} = 0.0529$$

$$\sigma(R_B) = (0.0124)^{1/2} = 0.1114$$

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

$$Corr(R_A, R_B) = \frac{0.0058}{(0.0529)(0.1114)} = 0.9842$$

The interpretation of the possible correlation values is summarized in Figure 1.

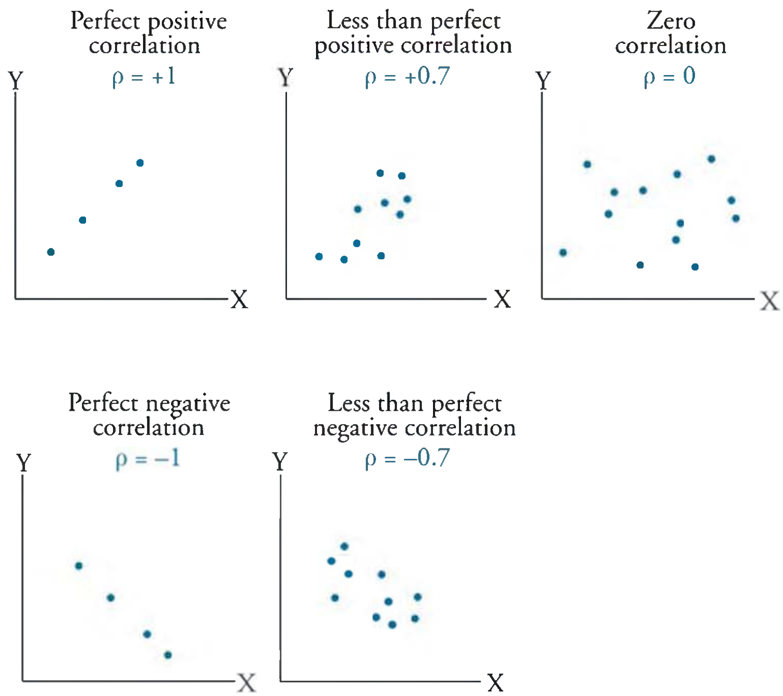
Figure 1: Interpretation of Correlation Coefficients

Correlation Coefficient (ρ)	*Interpretation*
$\rho = +1$	perfect positive correlation
$0 < \rho < +1$	a positive linear relationship
$\rho = 0$	no linear relationship
$-1 < \rho < 0$	a negative linear relationship
$\rho = -1$	perfect negative correlation

Interpreting a Scatter Plot

A scatter plot is a collection of points on a graph where each point represents the values of two variables (i.e., an X/Y pair). Figure 2 shows several scatter plots for the two random variables *X* and *Y* and the corresponding interpretation of correlation. As shown, an upward-sweeping scatter plot indicates a positive correlation between the two variables, while a downward-sweeping plot implies a negative correlation. Also illustrated in Figure 2 is that as we move from left to right in the rows of scatter plots, the extent of the linear relationship between the two variables deteriorates, and the correlation gets closer to zero. Note that for $\rho = 1$ and $\rho = -1$, the data points lie exactly on a line, but the slope of that line is not necessarily +1 or –1.

Figure 2: Interpretations of Correlation



Moments and Central Moments

LO 16.6: Describe the four central moments of a statistical variable or distribution: mean, variance, skewness and kurtosis.

The shape of a probability distribution can be described by the "moments" of the distribution. Raw moments are measured relative to an expected value raised to the appropriate power. The first raw moment is the **mean** of the distribution, which is the expected value of returns:

$$E(R) = \mu = \sum_{i=1}^{n} p_i R_i^1$$

where:
p_i = probability of event i
R_i = return associated with event i

Generalizing, the *k*th raw moment is the expected value of R^k:

$$E(R^k) = \sum_{i=1}^{n} p_i R_i^k$$

Raw moments for k > 1 are not very useful for our purposes, however, central moments for k > 1 are important.

Central moments are measured relative to the mean (i.e., central around the mean). The *k*th central moment is defined as:

$$E(R-\mu)^k = \sum_{i=1}^{n} p_i (R_i - \mu)^k$$

Professor's Note: Since central moments are measured relative to the mean, the first central moment equals zero and is, therefore, not typically used.

The second central moment is the **variance** of the distribution, which measures the dispersion of data.

$$\text{variance} = \sigma^2 = E\left[(R-\mu)^2\right]$$

Professor's Note: Since moments higher than the second central moment can be difficult to interpret, they are typically standardized by dividing the central moment by σ^k.

The third central moment measures the departure from symmetry in the distribution. This moment will equal zero for a symmetric distribution (such as the normal distribution).

$$\text{third central moment} = E\left[(R-\mu)^3\right]$$

The **skewness** statistic is the standardized third central moment. Skewness (sometimes called *relative skewness*) refers to the extent to which the distribution of data is not symmetric around its mean. It is calculated as:

$$\text{skewness} = \frac{E\left[(R-\mu)^3\right]}{\sigma^3}$$

The fourth central moment measures the degree of clustering in the distribution.

$$\text{fourth central moment} = E\left[(R-\mu)^4\right]$$

The **kurtosis** statistic is the standardized fourth central moment of the distribution. Kurtosis refers to the degree of peakedness or clustering in the data distribution and is calculated as:

$$\text{kurtosis} = \frac{E\left[(R-\mu)^4\right]}{\sigma^4}$$

Kurtosis for the normal distribution equals 3. Therefore, the **excess kurtosis** for any distribution equals:

$$\text{excess kurtosis} = \text{kurtosis} - 3$$

Although additional central moments can be calculated, risk management is not often concerned with anything beyond the fourth central moment.

Skewness and Kurtosis

LO 16.7: Interpret the skewness and kurtosis of a statistical distribution, and interpret the concepts of coskewness and cokurtosis.

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 data set. Outliers are observations with extraordinarily large values, either positive or negative.

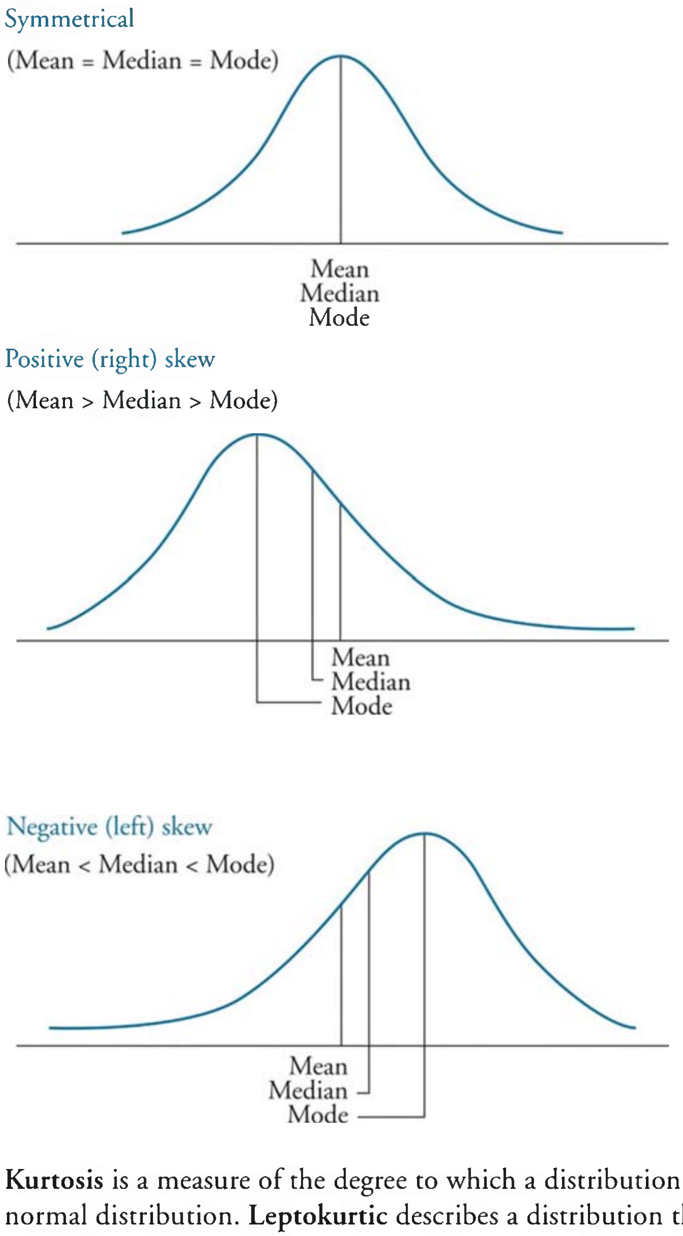
- A *positively skewed* distribution is characterized by many outliers in the upper region, or right tail. A positively skewed distribution is said to be skewed right because of its relatively long upper (right) tail.
- A *negatively skewed* distribution has a disproportionately large amount of outliers 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. Note the median is between the other two measures for positively or negatively skewed distributions.

Figure 3: Effect of Skewness on Mean, Median, and Mode

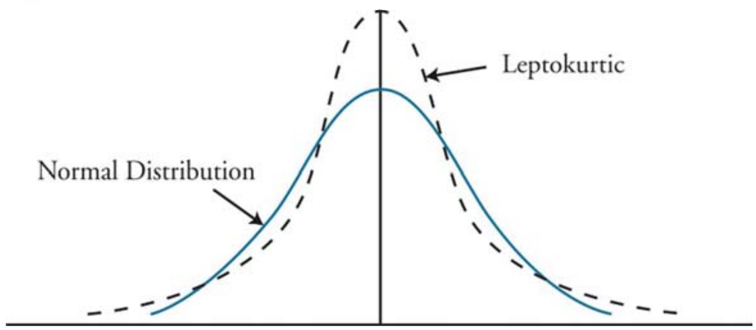


Kurtosis is a measure of the degree to which a distribution is more or less "peaked" than a normal distribution. **Leptokurtic** describes a distribution that is more peaked than a normal distribution, whereas **platykurtic** refers to a distribution that is less peaked (or flatter) than a normal distribution.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

As indicated in Figure 4, 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 there is a relatively greater probability of an observed value being either close to the mean or far from the mean. With regard to an investment returns distribution, a greater likelihood of a large deviation from the mean return is often perceived as an increase in risk.

Figure 4: 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 positive kurtosis and more negative skew in returns distributions indicates increased risk.

Coskewness and Cokurtosis

Previously, we identified moments and central moments for mean and variance. In a similar fashion, we can identify cross central moments for the concept of covariance. The third cross central moment is known as **coskewness** and the fourth cross central moment is known as **cokurtosis**.

To illustrate the importance of these concepts in risk management, suppose we are analyzing the returns data from four different stocks over a 7-year time period (shown in Figure 5). Although returns vary over time, the mean, variance, skewness, and kurtosis of all stock returns are the same under this scenario. In addition, the covariance between returns for Stock 1 and Stock 2 is equal to the covariance between returns for Stock 3 and Stock 4.

Figure 5: Stock Returns

	Stocks			
Time	1	2	3	4
1	0.0%	–2.4%	–12.6%	–12.6%
2	–2.4%	–12.6%	–5.3%	–5.3%
3	–12.6%	2.4%	0.0%	–2.4%
4	–5.3%	–5.3%	–2.4%	12.6%
5	2.4%	0.0%	2.4%	0.0%
6	5.3%	5.3%	5.3%	5.3%
7	12.6%	12.6%	12.6%	2.4%

By combining Stock 1 and Stock 2 into Portfolio A, and Stock 3 and Stock 4 into Portfolio B (shown in Figure 6), we find that the returns for Portfolio A and Portfolio B have the same mean and variance. However, these combined return sets do not have the same skewness (i.e., the coskewness between stocks in the portfolios is different). The reason for this difference is that the ranking of returns over time (e.g., from best to worst) is different for each stock, and when combined in a portfolio, these differences skew the portfolio returns distribution. For example, the worst return for Stock 1 occurred during time period 3, but in Portfolio A, the worst return occurred during time period 2. Similarly, the best return for Stock 4 occurred during time period 4, but in Portfolio B, the best return occurred during time period 7.

Figure 6: Portfolio Returns

	Portfolio	
Time	A	B
1	–1.2%	–12.6%
2	–7.5%	–5.3%
3	–5.1%	–1.2%
4	–5.3%	5.1%
5	1.2%	1.2%
6	5.3%	5.3%
7	12.6%	7.5%

From a risk management standpoint, it is helpful to know that the worst outcome in Portfolio B is 1.7 times greater than the worst outcome in Portfolio A. So, although the mean and variance of these portfolios are equal, shortfall risk expectations can differ depending on time period. This is important information to know, however, most risk models choose to ignore the effects of coskewness and cokurtosis. The reason being is that as the number of variables increase, the number of coskewness and cokurtosis terms will increase rapidly, making the data much more difficult to analyze. Practitioners instead opt to use more tractable risk models, such as GARCH (see Topic 28), which capture the essence of coskewness and cokurtosis by incorporating time-varying volatility and/or time-varying correlation.

The Best Linear Unbiased Estimator

LO 16.8: Describe and interpret the best linear unbiased estimator.

In upcoming topics, we will continue to discuss statistics and explore how sample parameters can be used to draw conclusions about population parameters. **Point estimates** are single (sample) values used to estimate population parameters, and the formula used to compute a point estimate is known as an **estimator**.

There are certain statistical properties that make some estimates more desirable than others. These desirable properties of an estimator are unbiasedness, efficiency, consistency, and linearity.

- An *unbiased* estimator is one for which the expected value of the estimator is equal to the parameter you are trying to estimate. For example, because the expected value of the sample mean is equal to the population mean $[E(\bar{x}) = \mu]$, the sample mean is an unbiased estimator of the population mean.
- An unbiased estimator is also *efficient* if the variance of its sampling distribution is smaller than all the other unbiased estimators of the parameter you are trying to estimate. The sample mean, for example, is an unbiased and efficient estimator of the population mean.
- A *consistent* estimator is one for which the accuracy of the parameter estimate increases as the sample size increases. As the sample size increases, the sampling distribution bunches more closely around the population mean.
- A point estimate is a *linear* estimator when it can be used as a linear function of sample data.

If the estimator is the best available (i.e., has the minimum variance), exhibits linearity, and is unbiased, it is said to be the **best linear unbiased estimator** (BLUE).

Key Concepts

LO 16.1

To compute the population mean, all the observed values in the population are summed and divided by the number of observations in the population.

Variance and standard deviation provide a measure of the extent of the dispersion in the values of the random variable around the mean.

LO 16.2

The mean of a population is expressed as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

Variance of a random variable is defined as:

$$\text{Var}(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$$
where $\mu = E(X)$

The square root of the variance is called the standard deviation.

LO 16.3

Expected value is the weighted average of the possible outcomes of a random variable, where the weights are the probabilities that the outcomes will occur. The expectation of a random variable *X* having possible values $x_1, \ldots, x_n$ is defined as:

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

LO 16.4

Covariance measures the extent to which two random variables tend to be above and below their respective means for each joint realization. It can be calculated as:

$$\text{Cov}(A,B) = \sum_{i=1}^{N} P_i\left(A_i - \bar{A}\right)\left(B_i - \bar{B}\right)$$

Correlation is a standardized measure of association between two random variables; it ranges in value from –1 to +1 and is equal to:

$$\frac{\text{Cov}(A,B)}{\sigma_A \sigma_B}$$

LO 16.5

If X and Y are any random variables, then:

$$E(X + Y) = E(X) + E(Y)$$

If X and Y are independent random variables, then:

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

If X and Y are NOT independent, then:

$$Var(X + Y) = Var(X) + Var(Y) + 2 \times Cov(X,Y)$$

$$Var(X - Y) = Var(X) + Var(Y) - 2 \times Cov(X,Y)$$

LO 16.6

The shape of a probability distribution is characterized by its raw moments and central moments. The first raw moment is the mean of the distribution. The second central moment is the variance. The third central moment divided by the cube of the standard deviation measures the skewness of the distribution, and the fourth central moment divided by the fourth power of the standard deviation measures the kurtosis of the distribution.

LO 16.7

Skewness describes the degree to which a distribution is nonsymmetric about its mean.

- A right-skewed distribution has positive skewness and a mean that is higher than the median that is higher than the mode.
- A left-skewed distribution has negative skewness and a mean that is lower than the median that is lower than the mode.

Kurtosis measures the peakedness of a distribution and the probability of extreme outcomes.

- Excess kurtosis is measured relative to a normal distribution, which has a kurtosis of three.
- Positive values of excess kurtosis indicate a distribution that is leptokurtic (fat tails, more peaked).
- Negative values of excess kurtosis indicate a platykurtic distribution (thin tails, less peaked).

Like mean and variance, we can generalize covariance to cross central moments. The third cross central moment is coskewness and the fourth cross central moment is cokurtosis.

LO 16.8

Desirable statistical properties of an estimator include unbiasedness (sign of estimation error is random), efficiency (lower sampling error than any other unbiased estimator), consistency (variance of sampling error decreases with sample size), and linearity (used as a linear function of sample data).

Concept Checkers

1. 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:
   A. is positively skewed.
   B. is a symmetric distribution.
   C. has positive excess kurtosis.
   D. has negative excess kurtosis.

2. 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. 0.0331.
   B. 0.0011.
   C. 0.2656.
   D. 0.0112.

Use the following data to answer Questions 3 and 4.

Probability Matrix

Returns	R_B = 50%	R_B = 20%	R_B = –30%
R_A = –10%	40%	0%	0%
R_A = 10%	0%	30%	0%
R_A = 30%	0%	0%	30%

3. Given this probability matrix, the standard deviation of Stock B is closest to:
   A. 0.11.
   B. 0.22.
   C. 0.33.
   D. 0.15.

4. Given this probability matrix, the covariance between Stock A and B is closest to:
   A. –0.160.
   B. –0.055.
   C. 0.004.
   D. 0.020.

5. A discrete uniform distribution (each event has an equal probability of occurrence) has the following possible outcomes for X: [1, 2, 3, 4]. The variance of this distribution is closest to:
   A. 1.00.
   B. 1.25.
   C. 1.50.
   D. 2.00.

Concept Checker Answers

1. C 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 taller and have fatter tails than a normal distribution.

2. B $\text{Corr}(R_A,R_B)=\dfrac{\text{Cov}(R_A,R_B)}{[\sigma(R_A)][\sigma(R_B)]}$

$$\sigma^2(R_A)=\left[\frac{\text{Cov}(R_A,R_B)}{\sigma(R_B)\text{Corr}(R_A,R_B)}\right]^2=\left[\frac{0.0043}{(0.26)(0.5)}\right]^2=0.0331^2=0.0011$$

3. C Expected return of Stock B = (0.4)(0.5) + (0.3)(0.2) + (0.3)(–0.3) = 0.17

   $\text{Var}(R_B) = 0.4(0.5 - 0.17)^2 + 0.3(0.2 - 0.17)^2 + 0.3(-0.3 - 0.17)^2 = 0.1101$

   Standard deviation = $\sqrt{0.1101}$ = 0.3318

4. B $\text{Cov}(R_A,R_B) = 0.4(-0.1 - 0.08)(0.5 - 0.17) + 0.3(0.1 - 0.08)(0.2 - 0.17) + 0.3(0.3 - 0.08)(-0.3 - 0.17) = -0.0546$

5. B Expected value = (1/4)(1 + 2 + 3 + 4) = 2.5

   Variance = $(1/4)[(1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2] = 1.25$

   Note that since each observation is equally likely, each has 25% (1/4) chance of occurrence.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Distributions

Topic 17

Exam Focus

This topic explores common probability distributions: uniform, Bernoulli, binomial, Poisson, normal, lognormal, chi-squared, Student's t, and F. You will learn the properties, parameters, and common occurrences of these distributions. Also discussed is the central limit theorem, which allows us to use sampling statistics to construct confidence intervals for point estimates of population means. For the exam, focus most of your attention on the binomial, normal, and Student's t distributions. Also, know how to standardize a normally distributed random variable, how to use a z-table, and how to construct confidence intervals.

Parametric and Nonparametric Distributions

Probability distributions are classified into two categories: parametric and nonparametric. **Parametric distributions**, such as a normal distribution, can be described by using a mathematical function. These types of distributions make it easier to draw conclusions about the data; however, they also make restrictive assumptions, which are not necessarily supported by real-world patterns. **Nonparametric distributions**, such as a historical distribution, cannot be described by using a mathematical function. Instead of making restrictive assumptions, these types of distributions fit the data perfectly; however, without generalizing the data, it can be difficult for a researcher to draw any conclusions.

LO 17.1: Distinguish the key properties among the following distributions: uniform distribution, Bernoulli distribution, Binomial distribution, Poisson distribution, normal distribution, lognormal distribution, Chi-squared distribution, Student's t, and F-distributions, and identify common occurrences of each distribution.

The Uniform Distribution

The **continuous uniform distribution** is defined over a range that spans between some lower limit, *a*, and some upper limit, *b*, which serve as the parameters of the distribution. Outcomes can only occur between *a* and *b*, and since we are dealing with a continuous distribution, even if $a < x < b$, $P(X = x) = 0$. Formally, the properties of a continuous uniform distribution may be described as follows:

- For all $a \leq x_1 < x_2 \leq b$ (i.e., for all x_1 and x_2 between the boundaries *a* and *b*).
- $P(X < a \text{ or } X > b) = 0$ (i.e., the probability of X outside the boundaries is zero).
- $P(x_1 \leq X \leq x_2) = (x_2 - x_1)/(b - a)$. This defines the probability of outcomes between x_1 and x_2.

Don't miss how simple this is just because the notation is so mathematical. For a continuous uniform distribution, the probability of outcomes in a range that is one-half the whole range is 50%. The probability of outcomes in a range that is one-quarter as large as the whole possible range is 25%.

Example: Continuous uniform distribution

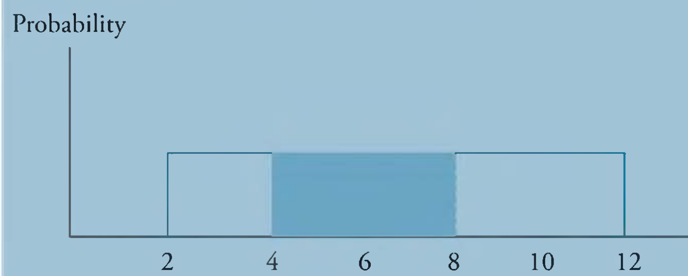
X is uniformly distributed between 2 and 12. **Calculate** the probability that *X* will be between 4 and 8.

Answer:

$$\frac{8-4}{12-2}=\frac{4}{10}=40\%$$

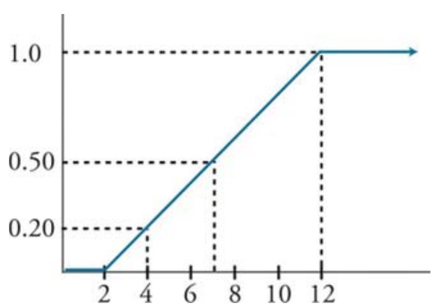
The following figure illustrates this continuous uniform distribution. Note that the area bounded by 4 and 8 is 40% of the total probability between 2 and 12 (which is 100%).

Continuous Uniform Distribution



Since outcomes are equal over equal-size possible intervals, the cumulative distribution function (cdf) is linear over the variable's range. The cdf for the distribution in the previous example, Prob (X < x), is shown in Figure 1.

Figure 1: CDF for a Continuous Uniform Variable



The probability function for a continuous random variable is called the probability density function (pdf) and is denoted f(x). Symbolically, the probability density function for a continuous uniform distribution is expressed as:

$$f(x)=\frac{1}{b-a} \text{ for } a \leq x \leq b, \text{ else } f(x)=0$$

The mean and variance, respectively, of a uniform distribution are:

$$E(x) = \frac{a+b}{2}$$
$$Var(x) = \frac{(b-a)^2}{12}$$

The Bernoulli Distribution

A Bernoulli distributed random variable only has two possible outcomes. The outcomes can be defined as either a "success" or a "failure." The probability of success, *p*, may be denoted with the value "1" and the probability of failure, *1 – p*, may be denoted with the value "0." Bernoulli distributed random variables are commonly used for assessing whether or not a company defaults during a specified time period. In the default example, the random variable equals "1" in the event of default and "0" in the event of survival.

The Binomial Distribution

A **binomial random variable** may be defined as the number of "successes" in a given number of trials, whereby the outcome can be either "success" or "failure." The probability of success, *p*, is constant for each trial and the trials are independent. A binomial random variable for which the number of trials is 1 is called a Bernoulli random variable. Think of a trial as a mini-experiment (or Bernoulli trial). The final outcome is the number of successes in a series of *n* trials. Under these conditions, the binomial probability function defines the probability of *x* successes in *n* trials. It can be expressed using the following formula:

$$p(x) = P(X = x) = (\text{number of ways to choose } x \text{ from } n)p^x(1 - p)^{n-x}$$

where:

$$(\text{number of ways to choose x from n}) = \frac{n!}{(n-x)!x!}$$

p = the probability of "success" on each trial [don't confuse it with p(x)]

So the probability of exactly *x* successes in *n* trials is:

$$p(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

Example: Binomial probability

Assuming a binomial distribution, **compute** the probability of drawing three black beans from a bowl of black and white beans if the probability of selecting a black bean in any given attempt is 0.6. You will draw five beans from the bowl.

Answer:

$$P(X=3)=p(3)=\frac{5!}{2!3!}(0.6)^3(0.4)^2=(120\ /\ 12)(0.216)(0.160)=0.3456$$

Some intuition about these results may help you remember the calculations. Consider that a (very large) bowl of black and white beans has 60% black beans and that each time you select a bean, you replace it in the bowl before drawing again. We want to know the probability of selecting exactly three black beans in five draws, as in the previous example.

One way this might happen is BBBWW. Since the draws are independent, the probability of this is easy to calculate. The probability of drawing a black bean is 60%, and the probability of drawing a white bean is 1 – 60% = 40%. Therefore, the probability of selecting BBBWW, in order, is 0.6 × 0.6 × 0.6 × 0.4 × 0.4 = 3.456%. This is the $p^3(1 - p)^2$ from the formula and p is 60%, the probability of selecting a black bean on any single draw from the bowl. BBBWW is not, however, the only way to choose exactly three black beans in five trials. Another possibility is BBWWB, and a third is BWWBB. Each of these will have exactly the same probability of occurring as our initial outcome, BBBWW. That's why we need to answer the question of how many ways (different orders) there are for us to choose three black beans in five draws. Using the formula, there are $\frac{5!}{(5-3)!3!}=10$ ways; 10 × 3.456% = 34.56%, the answer we computed previously.

Expected Value and Variance of a Binomial Random Variable

For a given series of n trials, the expected number of successes, or E(X), is given by the following formula:

expected value of X = E(X) = np

The intuition is straightforward; if we perform n trials and the probability of success on each trial is p, we expect np successes.

The variance of a binomial random variable is given by:

variance of X = np(1 – p) = npq

Professor's Note: $q = 1 - p$ is the probability that the event will fail to occur in a single trial (i.e., the probability of failure).

Example: Expected value of a binomial random variable

Based on empirical data, the probability that the Dow Jones Industrial Average (DJIA) will increase on any given day has been determined to equal 0.67. Assuming the only other outcome is that it decreases, we can state p(UP) = 0.67 and p(DOWN) = 0.33. Further, assume that movements in the DJIA are independent (i.e., an increase in one day is independent of what happened on another day).

Using the information provided, **compute** the expected value of the number of up days in a 5-day period.

Answer:

Using binomial terminology, we define success as UP, so p = 0.67. Note that the definition of success is critical to any binomial problem.

$$E(X \mid n = 5, p = 0.67) = (5)(0.67) = 3.35$$

Recall that the "|" symbol means *given*. Hence, the preceding statement is read as: the expected value of X given that n = 5, and the probability of success = 67% is 3.35.

Using the equation for the variance of a binomial distribution, we find the variance of X to be:

$$\text{Var}(X) = np(1 - p) = 5(0.67)(0.33) = 1.106$$

We should note that since the binomial distribution is a discrete distribution, the result X = 3.35 is not possible. However, if we were to record the results of many 5-day periods, the average number of up days (successes) would converge to 3.35.

Binomial distributions are used extensively in the investment world where outcomes are typically seen as successes or failures. In general, if the price of a security goes up, it is viewed as a success. If the price of a security goes down, it is a failure. In this context, binomial distributions are often used to create models to aid in the process of asset valuation.

Professor's Note: We will examine binomial trees for stock option valuation in Book 4.

The Poisson Distribution

The Poisson distribution is another discrete probability distribution with a number of real-world applications. For example, the number of defects per batch in a production process or the number of calls per hour arriving at the 911 emergency switchboard are discrete random variables that follow a Poisson distribution.

While the Poisson random variable X refers to the *number of successes per unit*, the parameter lambda (λ) refers to the average or *expected number of successes per unit*. The mathematical expression for the Poisson distribution for obtaining X successes, given that λ successes are expected, is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

An interesting feature of the Poisson distribution is that both its mean and variance are equal to the parameter, λ.

Example: Using the Poisson distribution (1)

On average, the 911 emergency switchboards receive 0.1 incoming calls per second. What is the probability that in a given minute exactly 5.0 phone calls will be received, assuming the arrival of calls follows a Poisson distribution?

Answer:

We first need to convert the seconds into minutes. Note that λ, the expected number of calls per minute, is (0.1)(60) = 6.0. Hence:

$$P(X = 5) = \frac{6^5 e^{-6}}{5!} = 0.1606 = 16.06\%$$

This means that, given the average of 0.1 incoming calls per second, there is a 16.06% chance there will be five incoming phone calls in a minute.

Example: Using the Poisson distribution (2)

Assume there is a 0.01 probability of a patient experiencing severe weight loss as a side effect from taking a recently approved drug used to treat heart disease. What is the probability that out of 200 such procedures conducted on different patients, five patients will develop this complication? Assume that the number of patients developing the complication from the procedure is Poisson-distributed.

Answer:

Let X = expected number of patients developing the complication from the procedure
= np = (200)(0.01) = 2

$$P(X = 5) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^5 e^{-2}}{5!} = 0.036 = 3.6\%$$

This means that given a complication rate of 0.01, there is a 3.6% probability that 5 out of every 200 patients will experience severe weight loss from taking the drug.

The Normal Distribution

The normal distribution is important for many reasons. Many of the random variables that are relevant to finance and other professional disciplines follow a normal distribution. In the area of investment and portfolio management, the normal distribution plays a central role in portfolio theory.

The probability density function for the normal distribution is:

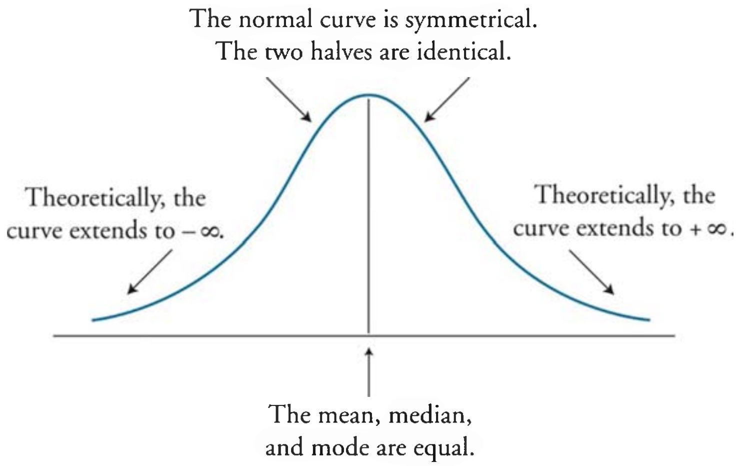
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The normal distribution has the following key properties:

- It is completely described by its mean, μ, and variance, σ^2, stated as $X \sim N(\mu, \sigma^2)$. In words, this says that "*X* is normally distributed with mean μ and variance σ^2."
- Skewness = 0, meaning the normal distribution is symmetric about its mean, so that $P(X \leq \mu) = P(\mu \leq X) = 0.5$, and mean = median = mode.
- Kurtosis = 3; this is a measure of how flat the distribution is. Recall that excess kurtosis is measured relative to 3, the kurtosis of the normal distribution.
- A linear combination of normally distributed independent random variables is also normally distributed.
- The probabilities of outcomes further above and below the mean get smaller and smaller but do not go to zero (the tails get very thin but extend infinitely).

Many of these properties are evident from examining the graph of a normal distribution's probability density function as illustrated in Figure 2.

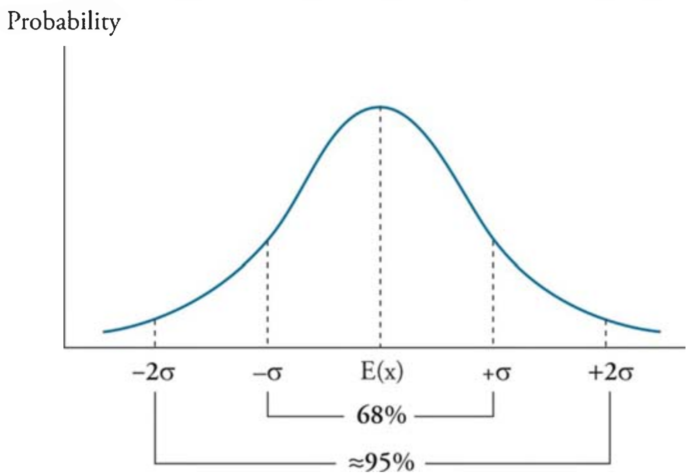
Figure 2: Normal Distribution Probability Density Function



A **confidence interval** is a range of values around the expected outcome within which we expect the actual outcome to be some specified percentage of the time. A 95% confidence interval is a range that we expect the random variable to be in 95% of the time. For a normal distribution, this interval is based on the expected value (sometimes called a point estimate) of the random variable and on its variability, which we measure with standard deviation.

Confidence intervals for a normal distribution are illustrated in Figure 3. For any normally distributed random variable, 68% of the outcomes are within one standard deviation of the expected value (mean), and approximately 95% of the outcomes are within two standard deviations of the expected value.

Figure 3: Confidence Intervals for a Normal Distribution



In practice, we will not know the actual values for the mean and standard deviation of the distribution, but will have estimated them as $\bar{X}$ and s. The three confidence intervals of most interest are given by:

- The 90% confidence interval for X is $\bar{X} - 1.65s$ to $\bar{X} + 1.65s$.
- The 95% confidence interval for X is $\bar{X} - 1.96s$ to $\bar{X} + 1.96s$.
- The 99% confidence interval for X is $\bar{X} - 2.58s$ to $\bar{X} + 2.58s$.

Example: Confidence intervals

The average return of a mutual fund is 10.5% per year and the standard deviation of annual returns is 18%. If returns are approximately normal, what is the 95% confidence interval for the mutual fund return next year?

Answer:

Here μ and σ are 10.5% and 18%, respectively. Thus, the 95% confidence interval for the return, R, is:

$$10.5 \pm 1.96(18) = -24.78\% \text{ to } 45.78\%$$

Symbolically, this result can be expressed as:

$$P(-24.78 < R < 45.78) = 0.95 \text{ or } 95\%$$

The interpretation is that the annual return is expected to be within this interval 95% of the time, or 95 out of 100 years.

The Standard Normal Distribution

A standard normal distribution (i.e., *z*-distribution) is a normal distribution that has been standardized so it has a mean of zero and a standard deviation of 1 [i.e., N~(0,1)]. To standardize an observation from a given normal distribution, the *z-value* of the observation must be calculated. The *z*-value represents the number of standard deviations a given observation is from the population mean. *Standardization* is the process of converting an observed value for a random variable to its *z*-value. The following formula is used to standardize a random variable:

$$z = \frac{\text{observation} - \text{population mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

Professor's Note: The term z-value will be used for a standardized observation in this topic. The terms z-score and z-statistic are also commonly used.

Example: Standardizing a random variable (calculating *z*-values)

Assume the annual earnings per share (EPS) for a population of firms are normally distributed with a mean of \$6 and a standard deviation of \$2.

What are the *z*-values for EPS of \$2 and \$8?

Answer:

If EPS = x = \$8, then $z = (x - \mu) / \sigma = (\$8 - \$6) / \$2 = +1$

If EPS = x = \$2, then $z = (x - \mu) / \sigma = (\$2 - \$6) / \$2 = -2$

Here, $z = +1$ indicates that an EPS of \$8 is one standard deviation above the mean, and $z = -2$ means that an EPS of \$2 is two standard deviations below the mean.

Calculating Probabilities Using *z*-Values

Now we will show how to use standardized values (*z*-values) and a table of probabilities for *Z* to determine probabilities. A portion of a table of the cumulative distribution function for a standard normal distribution is shown in Figure 4. We will refer to this table as the *z*-table, as it contains values generated using the cumulative density function for a standard normal distribution, denoted by F(Z). Thus, the values in the *z*-table are the probabilities of observing a *z*-value that is less than a given value, *z* [i.e., $P(Z < z)$]. The numbers in the first column are *z*-values that have only one decimal place. The columns to the right supply probabilities for *z*-values with two decimal places.

Note that the *z*-table in Figure 4 only provides probabilities for positive *z*-values. This is not a problem because we know from the symmetry of the standard normal distribution that $F(-Z) = 1 - F(Z)$. The tables in the back of many texts actually provide probabilities for negative *z*-values, but we will work with only the positive portion of the table because this may be all you get on the exam. In Figure 4, we can find the probability that a standard normal random variable will be less than 1.66, for example. The table value is 95.15%. The probability that the random variable will be less than –1.66 is simply $1 - 0.9515 = 0.0485 =$ 4.85%, which is also the probability that the variable will be greater than +1.66.

Figure 4: Cumulative Probabilities for a Standard Normal Distribution

CDF Values for the Standard Normal Distribution: The z-Table										
z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.5	.6915	Please note that several of the rows have been deleted to save space.*								
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

*A complete cumulative standard normal table is included in the Appendix.

Professor's Note: When you use the standard normal probabilities, you have formulated the problem in terms of standard deviations from the mean. Consider a security with returns that are approximately normal, an expected return of 10%, and standard deviation of returns of 12%. The probability of returns greater than 30% is calculated based on the number of standard deviations that 30% is above the expected return of 10%. 30% is 20% above the expected return of 10%, which is 20 / 12 = 1.67 standard deviations above the mean. We look up the probability of returns less than 1.67 standard deviations above the mean (0.9525 or 95.25% from Figure 4) and calculate the probability of returns more than 1.67 standard deviations above the mean as 1 – 0.9525 = 4.75%.

Example: Using the z-table (1)

Considering again EPS distributed with μ = \$6 and σ = \$2, what is the probability that EPS will be \$9.70 or more?

Answer:

Here we want to know P(EPS > \$9.70), which is the area under the curve to the right of the *z*-value corresponding to EPS = \$9.70 (see the distribution that follows).

The *z*-value for EPS = \$9.70 is:

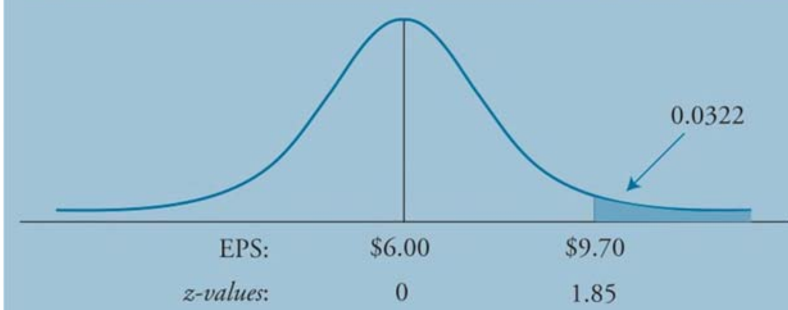
$$z = \frac{(x - \mu)}{\sigma} = \frac{(9.70 - 6)}{2} = 1.85$$

That is, \$9.70 is 1.85 standard deviations above the mean EPS value of \$6.

From the z-table we have F(1.85) = 0.9678, but this is P(EPS ≤ 9.70). We want P(EPS > 9.70), which is 1 – P(EPS ≤ 9.70).

P(EPS > 9.70) = 1 – 0.9678 = 0.0322, or 3.2%

P(EPS>$9.70)



Example: Using the z-table (2)

Using the distribution of EPS with μ = $6 and σ = $2 again, what percent of the observed EPS values are likely to be less than $4.10?

Answer:

As shown graphically in the distribution that follows, we want to know P(EPS < $4.10). This requires a 2-step approach like the one taken in the preceding example.

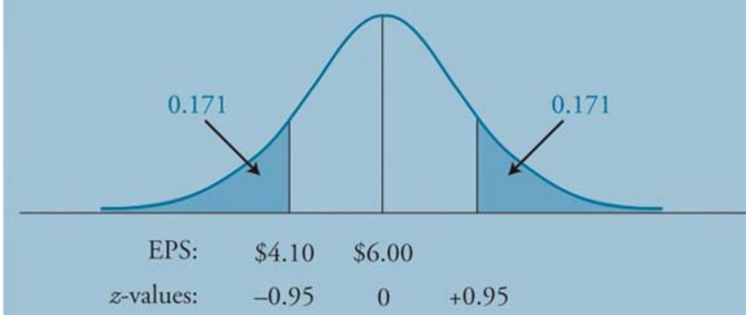
First, the corresponding z-value must be determined as follows:

$$z = \frac{(\$4.10 - \$6)}{2} = -0.95,$$

So $4.10 is 0.95 standard deviations below the mean of $6.00.

Now, from the z-table for negative values in the back of this book, we find that F(–0.95) = 0.1711, or 17.11%.

Finding a Left-Tail Probability



The *z*-table gives us the probability that the outcome will be more than 0.95 standard deviations below the mean.

The Lognormal Distribution

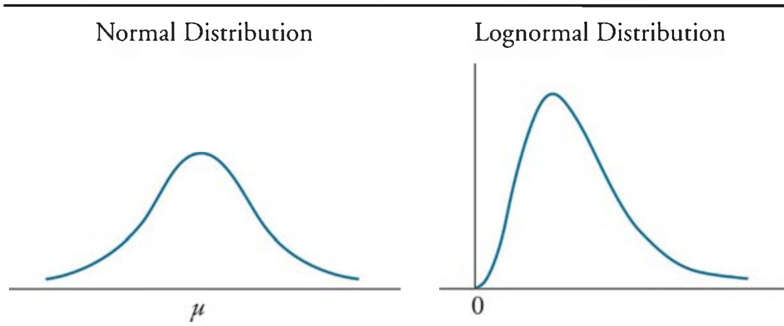
The lognormal distribution is generated by the function e^x, where x is normally distributed. Since the natural logarithm, ln, of e^x is x, the logarithms of lognormally distributed random variables are normally distributed, thus the name.

The probability density function for the lognormal distribution is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}$$

Figure 5 illustrates the differences between a normal distribution and a lognormal distribution.

Figure 5: Normal vs. Lognormal Distributions



In Figure 5, we can see that:

- The lognormal distribution is skewed to the right.
- The lognormal distribution is bounded from below by zero so that it is useful for modeling asset prices which never take negative values.

If we used a normal distribution of returns to model asset prices over time, we would admit the possibility of returns less than –100%, which would admit the possibility of asset prices less than zero. Using a lognormal distribution to model *price relatives* avoids this problem. A price relative is just the end-of-period price of the asset divided by the beginning price (S_1/S_0) and is equal to (1 + the holding period return). To get the end-of-period asset price, we can simply multiply the price relative times the beginning-of-period asset price. Since a lognormal distribution takes a minimum value of zero, end-of-period asset prices cannot be less than zero. A price relative of zero corresponds to a holding period return of –100% (i.e., the asset price has gone to zero).

The Central Limit Theorem

LO 17.2: Describe the central limit theorem and the implications it has when combining independent and identically distributed (i.i.d.) random variables.

LO 17.3: Describe i.i.d. random variables and the implications of the i.i.d. assumption when combining random variables.

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 variance equal to $\frac{\sigma^2}{n}$ as the sample size becomes large. This is possible because, when the sample size is large, the sums of independent and identically distributed (i.i.d.) random variables (the individual items drawn for the sample) will be normally distributed.

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.

Student's *t*-Distribution

Student's *t*-distribution, or simply the *t*-distribution, is a bell-shaped probability distribution that is symmetrical about its mean. It is the appropriate distribution to use when constructing confidence intervals based on *small samples* ($n < 30$) from populations with *unknown variance* and a normal, or approximately normal, distribution. It may also be appropriate to use the *t*-distribution when the population variance is unknown and the sample size is large enough that the central limit theorem will assure that the sampling distribution is approximately normal.

Student's t-distribution has the following properties:

- It is symmetrical.
- It is defined by a single parameter, the degrees of freedom (df), where the degrees of freedom are equal to the number of sample observations minus 1, n – 1, for sample means.
- It has more probability in the tails (fatter tails) than the normal distribution.
- As the degrees of freedom (the sample size) gets larger, the shape of the t-distribution more closely approaches a standard normal distribution.

When *compared to the normal distribution*, the t-distribution is flatter with more area under the tails (i.e., it has fatter tails). As the degrees of freedom for the t-distribution increase, however, its shape approaches that of the normal distribution.

The degrees of freedom for tests based on sample means are n – 1 because, given the mean, only n – 1 observations can be unique.

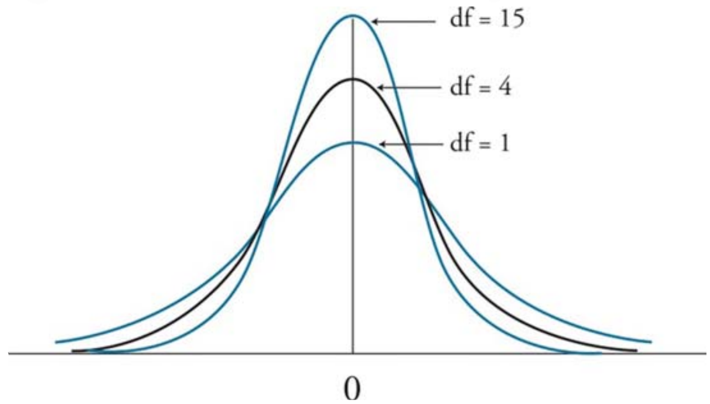
The table in Figure 6 contains one-tailed critical values for the t-distribution at the 0.05 and 0.025 levels of significance with various degrees of freedom (df). Note that, unlike the z-table, the t-values are contained within the table and the probabilities are located at the column headings. Also note that the level of significance of a t-test corresponds to the *one-tailed probabilities*, p, that head the columns in the t-table.

Figure 6: Table of Critical t-Values

	One-Tailed Probabilities, p	
df	$p = 0.05$	$p = 0.025$
5	2.015	2.571
10	1.812	2.228
15	1.753	2.131
20	1.725	2.086
25	1.708	2.060
30	1.697	2.042
40	1.684	2.021
50	1.676	2.009
60	1.671	2.000
70	1.667	1.994
80	1.664	1.990
90	1.662	1.987
100	1.660	1.984
120	1.658	1.980
∞	1.645	1.960

Figure 7 illustrates the different shapes of the t-distribution associated with different degrees of freedom. The tendency is for the t-distribution to look more and more like the normal distribution as the degrees of freedom increase. Practically speaking, the greater the degrees of freedom, the greater the percentage of observations near the center of the distribution and the lower the percentage of observations in the tails, which are thinner as degrees of freedom increase. This means that confidence intervals for a random variable that follows a *t*-distribution must be wider (narrower) when the degrees of freedom are less (more) for a given significance level.

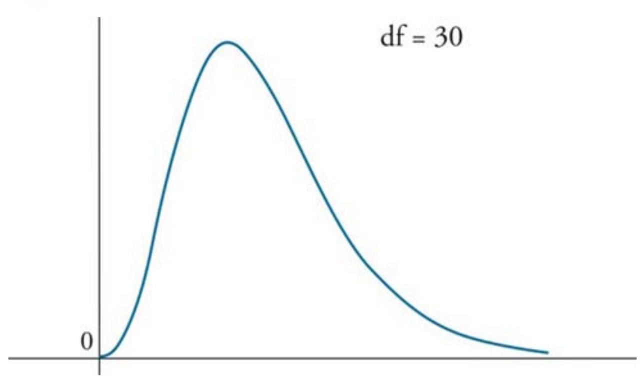
Figure 7: *t*-Distributions for Different Degrees of Freedom (df)



The Chi-Squared Distribution

As you will see in Topic 19, hypothesis testing of the population variance requires the use of a chi-squared distributed test statistic, denoted χ^2. The chi-squared distribution is asymmetrical, bounded below by zero, and approaches the normal distribution in shape as the degrees of freedom increase.

Figure 8: Chi-Squared Distribution



The chi-squared test statistic, χ^2, with n – 1 degrees of freedom, is computed as:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$$

where:

n = sample size
s^2 = sample variance
σ_0^2 = hypothesized value for the population variance

The chi-squared test compares the test statistic to a critical chi-squared value at a given level of significance to determine whether to reject or fail to reject a null hypothesis. Note that since the chi-squared distribution is bounded below by zero, chi-squared values cannot be negative.

The *F*-Distribution

As you will also see in Topic 19, 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.

The test statistic for the *F*-test is the ratio of the sample variances. The *F*-statistic is computed as:

$$F = \frac{s_1^2}{s_2^2}$$

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

An *F*-distribution is presented in Figure 9. As indicated, the *F*-distribution is right-skewed and is truncated at 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.

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 the Appendix).

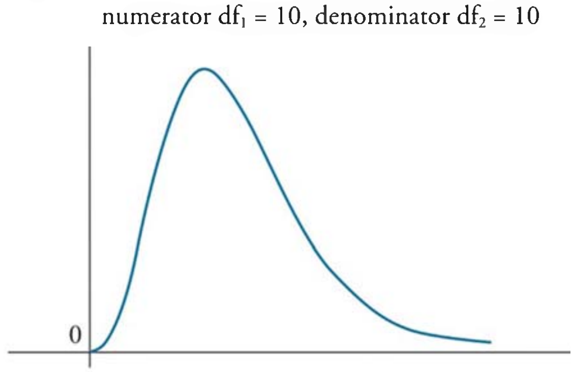
Some additional properties of the *F*-distribution include the following:

- The *F*-distribution approaches the normal distribution as the number of observations increases (just as with the *t*-distribution and chi-squared distribution).
- A random variable's *t*-value squared (t^2) with n – 1 degrees of freedom is *F*-distributed with 1 degree of freedom in the numerator and n – 1 degrees of freedom in the denominator.
- There exists a relationship between the *F*- and chi-squared distributions such that:

$$F = \frac{\chi^2}{\text{\# of observations in numerator}}$$

as the # of observations in denominator $\rightarrow \infty$

Figure 9: *F*-Distribution



Mixture Distributions

LO 17.4: Describe a mixture distribution and explain the creation and characteristics of mixture distributions.

The distributions discussed in this topic, as well as others, can be combined to create unique probability density functions. It may be helpful to create a new distribution if the underlying data you are working with does not currently fit a predetermined distribution. In this case, a newly created distribution may assist with explaining the relevant data.

To illustrate a mixture distribution, suppose that the returns of a stock follow a normal distribution with low volatility 75% of the time and high volatility 25% of the time. Here we have two normal distributions with the same mean, but different risk levels. To create a mixture distribution from these scenarios, we randomly choose either the low or high volatility distribution, placing a 75% probability on selecting the low volatility distribution. We then generate a random return from the selected distribution. By repeating this process several times, we will create a probability distribution that reflects both levels of volatility.

Mixture distributions contain elements of both parametric and nonparametric distributions. The distributions used as inputs (i.e., the component distributions) are parametric, while the weights of each distribution within the mixture are nonparametric. The more component distributions used as inputs, the more closely the mixture distribution will follow the actual data. However, more component distributions will make it difficult to draw conclusions given that the newly created distribution will be very specific to the data.

By mixing distributions, it is easy to see how we can alter skewness and kurtosis of the component distributions. Skewness can be changed by combining distributions with different means, and kurtosis can be changed by combining distributions with different variances. Also, by combining distributions that have significantly different means, we can create a mixture distribution with multiple modes (e.g., a bimodal distribution).

Creating a more robust distribution is clearly beneficial to risk managers. Different levels of skew and/or kurtosis can reveal extreme events that were previously difficult to identify. By creating these mixture distributions, we can improve risk models by incorporating the potential for low-frequency, high-severity events.

Key Concepts

LO 17.1

A continuous uniform distribution is one where the probability of *X* occurring in a possible range is the length of the range relative to the total of all possible values. Letting *a* and *b* be the lower and upper limit of the uniform distribution, respectively, then for: $a \leq x_1 < x_2 \leq b$,

$$P(x_1 \leq X \leq x_2) = \frac{(x_2 - x_1)}{(b - a)}$$

The binomial distribution is a discrete probability distribution for a random variable, *X*, that has one of two possible outcomes, success or failure. The probability of a specific number of successes in *n* independent binomial trials is:

$$p(x) = P(X = x) = \frac{n!}{(n - x)!x!} p^x (1 - p)^{n-x}$$

where p = the probability of success in a given trial

The Poisson random variable *X* refers to a specific number of successes per unit. The probability for obtaining *X* successes, given a Poisson distribution with parameter λ is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The normal probability distribution has the following characteristics:

- The normal curve is symmetrical and bell-shaped with a single peak at the exact center of the distribution.
- Mean = median = mode, and all are in the exact center of the distribution.
- The normal distribution can be completely defined by its mean and standard deviation because the skew is always zero and kurtosis is always three.

A lognormal distribution exists for random variable *Y*, when $Y = e^X$, and *X* is normally distributed.

The *t*-distribution is similar, but not identical, to the normal distribution in shape—it is defined by the degrees of freedom, has a lower peak, and has fatter tails. The *t*-distribution is used to construct confidence intervals for the population mean when the population variance is not known.

Degrees of freedom for the *t*-distribution is equal to $n - 1$; Student's *t*-distribution is closer to the normal distribution when *df* is greater, and confidence intervals are narrower when df is greater.

The chi-squared distribution is asymmetrical, bounded below by zero, and approaches the normal distribution in shape as the degrees of freedom increase.

The *F*-distribution is right-skewed and is truncated at zero on the left-hand side. The shape of the *F*-distribution is determined by two separate degrees of freedom.

LO 17.2

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* will be approximately normally distributed with a mean equal to μ and a variance equal to σ^2/n.

LO 17.3

When a sample size is large, the sums of independent and identically distributed (i.i.d.) random variables will be normally distributed.

LO 17.4

Mixture distributions combine the concepts of parametric and nonparametric distributions. The component distributions used as inputs are parametric while the weights of each distribution within the mixture are based on historical data, which is nonparametric.

Concept Checkers

1. Which of the following statements about the *F*-distribution and chi-squared distribution is least accurate? Both distributions:
   A. are asymmetrical.
   B. are bound by zero on the left.
   C. are defined by degrees of freedom.
   D. have means that are less than their standard deviations.

2. The probability that a standard normally distributed random variable will be more than two standard deviations above its mean is:
   A. 0.0217.
   B. 0.0228.
   C. 0.4772.
   D. 0.9772.

3. If 5% of the cars coming off the assembly line have some defect in them, what is the probability that out of three cars chosen at random, exactly one car will be defective? Assume that the number of defective cars has a Poisson distribution.
   A. 0.129.
   B. 0.135.
   C. 0.151.
   D. 0.174.

4. A recent study indicated that 60% of all businesses have a fax machine. Assuming a binomial probability distribution, what is the probability that exactly four businesses will have a fax machine in a random selection of six businesses?
   A. 0.138.
   B. 0.276.
   C. 0.311.
   D. 0.324.

5. What is the probability of an outcome being between 15 and 25 for a random variable that follows a continuous uniform distribution over the range of 12 to 28?
   A. 0.509.
   B. 0.625.
   C. 1.000.
   D. 1.600.

Concept Checker Answers

1. D There is no consistent relationship between the mean and standard deviation of the chi-squared distribution or *F*-distribution.

2. B 1 – F(2) = 1 – 0.9772 = 0.0228

3. A The probability of a defective car (*p*) is 0.05; hence, the probability of a non-defective car (q) = 1 – 0.05 = 0.95. Assuming a Poisson distribution:

   $\lambda = np = (3)(0.05) = 0.15$

   Then,

   $$P(X=1)=\frac{\lambda^x e^{-\lambda}}{x!}=\frac{(0.15)^1 e^{-0.15}}{1!}=0.129106$$

4. C Success = having a fax machine:

   $[6! / 4!(6-4)!](0.6)^4(0.4)^{6-4} = 15(0.1296)(0.16) = 0.311$

5. B Since a = 12 and b = 28:

   $$P(15 \le X \le 25)=\frac{(25-15)}{(28-12)}=\frac{10}{16}=0.625$$

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

BAYESIAN ANALYSIS

Topic 18

EXAM FOCUS

Bayes' theorem is used to update a given set of prior probabilities for a given event in response to the arrival of new information. Updating a prior probability of an event requires knowledge of both conditional and unconditional probabilities. For the exam, be prepared to calculate updated probabilities when applying Bayesian analysis based on the probability of conditional and unconditional events occurring. Also, be prepared to contrast the Bayesian approach with the frequentist approach.

BAYES' THEOREM

LO 18.1: Describe Bayes' theorem and apply this theorem in the calculation of conditional probabilities.

Bayesian analysis is applied in numerous disciplines and is growing in interest in finance and risk management. The foundation of Bayesian analysis is **Bayes' theorem**. Bayes' theorem for two random variables *A* and *B* is defined as follows:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

For this topic, it is helpful to recall the notation and definitions of conditional, unconditional, and joint probabilities. The notation for a **conditional probability** is shown on the left-hand side of the equation, P(A | B). The conditional probability is read as the probability of event *A* occurring, given that event *B* has already occurred. The **unconditional probability** of event *A* occurring is noted as P(A). This is an overall probability of event *A* occurring regardless of the outcome of other events.

The numerator of the equation, [P(B | A) × P(A)], is the joint probability of events *A* and *B*. The joint probability of two events occurring at the same time can also be stated as P(AB). Therefore, another way of expressing Bayes' theorem based on the joint probability of both events occurring is shown as follows:

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

The joint probability of both events *A* and *B* occurring can be determined by the following two equations. Notice that it does not matter which event occurred first. The first equation is used if event *B* occurred first and the second equation is used if event *A* occurred first.

$$P(AB) = P(A \mid B) \times P(B)$$
$$P(AB) = P(B \mid A) \times P(A)$$

Regardless of which unconditional event occurred first, the joint probability of both occurring is the same. Thus, these two equations can be combined. Notice that if we divide each side of this equation by P(B), we have the first derivation of Bayes' theorem introduced in this topic.

$$P(A \mid B) \times P(B) = P(B \mid A) \times P(A)$$

Bayes' theorem provides a framework for determining the probability of one random event occurring given that another random event has already occurred. This is known as a conditional probability. The following example illustrates how to determine the probability of one bond defaulting given that another bond has already defaulted.

Suppose a bond manager is interested in knowing the probability of Bond A defaulting given that Bond B is already in default. Figure 1 provides a probability matrix defining two events for both bonds, default and no default. Bonds A and B each have a 12% probability of default and an 88% probability of not defaulting. The bottom row of Figure 1 sums the total probabilities for Bond A for no default and default as 88% and 12%, respectively. Likewise, the last column of Figure 1 sums the total of no default and default for Bond B as 88% and 12%, respectively. The joint probability of both bonds defaulting is 4% in this example. Similarly, the joint probability of no defaults for either bond is 80%.

Figure 1: Probability Matrix for Bond A and Bond B

		Bond A		
		No Default	Default	
Bond B	No Default	80%	8%	88%
	Default	8%	4%	12%
		88%	12%	100%

Professor's Note: The two events for each bond must sum to 100% (88% + 12% = 100%). Each bond will either be in a state of default or no default.

The recent financial crisis beginning in 2007 illustrated that bond defaults are highly correlated. If the probabilities of bond defaults were independent, then the probability of both bonds defaulting would be calculated as 1.44% (i.e., 12% × 12%). However, the actual joint probability of both bonds defaulting is much higher at 4%. In addition, the joint probability that both bonds do not default is 80%. This probability is higher than the probability for two independent events each with an 88% probability of occurring (i.e., 88% × 88% = 77.44%).

As mentioned, an unconditional probability is a random event that is not contingent on any additional information or events occurring. The unconditional probability of Bond A defaulting is the overall probability of Bond A default given in the example as 12%. In other words, there is a 12% probability of Bond A defaulting regardless of the state of Bond B.

The conditional probability of Bond A defaulting given that Bond B is already in default is defined by: $P(A \mid B) = P(AB) / P(B)$. The numerator is the joint probability of both defaulting, $P(AB) = 4\%$. The denominator is the unconditional probability of Bond B defaulting, $P(B)$. Thus, the conditional probability can be computed as:

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{4\%}{12\%} = \frac{1}{3} \text{ or } 33.3333\%$$

Professor's Note: If two events are highly correlated, the conditional probability of the event occurring (e.g., Bond A defaults given that Bond B is in default) is always higher than the unconditional probability of the event occurring.

Now we will look at another example that does not have everything neatly presented in a probability matrix.

Example: Bayes' theorem (1)

Suppose you are an equity analyst for ABC Insurance Company. You manage an equity fund of funds and use historical data to categorize the managers as excellent or average. Excellent managers are expected to outperform the market 70% of the time. Average managers are expected to outperform the market only 50% of the time. Assume that the probabilities of managers outperforming the market for any given year is independent of their performance in prior years. ABC Insurance Company has found that only 20% of all fund managers are excellent managers and the remaining 80% are average managers.

A new fund manager to the portfolio started three years ago and outperformed the market all three years. What is the probability that the new manager was an excellent manager when she first started managing portfolios three years ago?

Answer:

The last probabilities stated in the problem are the probabilities that a random fund manager is either an excellent manager [P(E) = 20%] or an average manager [P(A) = 80%].

The unconditional probability will answer the question related to the new manager (a random event occurring given no other information). There was a 20% probability that the new manager was an excellent manager when she first joined three years ago.

Bayesian analysis requires updating prior beliefs based on new information. In the prior example, we have new information that the manager outperformed the market three years in a row. Therefore, this information will change our prior beliefs regarding the probabilities that the manager is either excellent or average. This next example illustrates how Bayesian analysis updates prior beliefs based on new information.

Example: Bayes' theorem (2)

Using the same information given in the previous example, what are the probabilities that the new manager is an excellent or average manager today?

Answer:

To solve this problem, we first summarize the conditional probabilities related to the probability of outperforming the market given that the fund manager is either excellent or average.

- The probability of an excellent manager outperforming the market is 70% [$P(O \mid E) = 70\%$]. The notation is read as the probability that a manager outperforms the market given she is an excellent manager equals 70%.
- The probability of an average manager outperforming the market is 50% [$P(O \mid A) = 50\%$].

Next, we need to use Bayes' theorem to determine the probability that the new manager is excellent given that the manager outperformed the market three years in a row.

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)}$$

The numerator of Bayes' theorem is the probability that an excellent manager outperforms the market three years in a row [$P(O \mid E) \times P(E)$]. In other words, it is a joint probability of a manager being excellent and outperforming the market three years in a row. The manager's performance each year is independent of the performance in prior years. The probability of an excellent manager outperforming the market in any given year was given as 70%. Thus, the probability of an excellent manager outperforming the market three years in a row is 70% to the third power or 34.3% [$P(O \mid E) = 0.7^3 = 0.343$].

The denominator of Bayes' theorem is the unconditional probability of outperforming the market for three years in a row [$P(O)$]. This is calculated by finding the weighted average probability of both manager types outperforming the market three years in a row. If there is a 20% probability that a manager is excellent, then there is an 80% probability that a manager is average. The probabilities of the manager being excellent or average are used as the weights of 20% and 80%, respectively.

We are given that excellent managers are expected to outperform the market 70% of the time and we just determined that the probability of an excellent manager outperforming three years in a row is 34.3%. Similarly, the probability of an average manager outperforming the market three years in a row is determined by taking the 50% probability to the third power: ($0.5^3 = 0.125$).

With this information, we can solve for the unconditional probability of a random manager outperforming the market for three years in a row. This is computed as a weighted average of the probabilities of outperforming three years in a row for each type of manager:

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.7^3 \times 0.2) + (0.5^3 \times 0.8) \\ &= 0.0686 + 0.1 \\ &= 0.1686 \end{aligned}$$

We can now answer the question, "What is the probability that the new manager is excellent or average after outperforming the market three years in a row?" by incorporating the information required for Bayes' theorem.

Probability for excellent manager:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)} = \frac{0.343 \times 0.2}{0.1686} = 0.4069 = 40.7\%$$

Probability for average manager:

$$P(A \mid O) = \frac{P(O \mid A) \times P(A)}{P(O)} = \frac{0.125 \times 0.8}{0.1686} = 0.5931 = 59.3\%$$

The fact that the new manager outperformed the market three years in a row increases the probability that the new manager is an excellent manager from 20% to 40.7%. The probability that the new manager is an average manager goes down from 80% to 59.3%.

Professor's Note: The denominator is the same for both calculations as it is the unconditional probability of a random manager outperforming the market for three years in a row. In addition, the sum of the updated probabilities must still equal 100% (i.e., 40.7% + 59.3%), because the manager must be excellent or average.

Example: Bayes' theorem (3)

Using the same information given in the previous two examples, what is the probability that the new manager will beat the market next year, given that the new manager outperformed the market the last three years?

Answer:

This question is determined by finding the unconditional probability of the new manager outperforming the market. However, now we will use 40.7% as the weight for the probability that the manager is excellent and 59.3% as the weight for the probability that the manager is average:

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.7 \times 0.407) + (0.5 \times 0.593) \\ &= 0.2849 + 0.2965 \\ &= 0.5814 \end{aligned}$$

Thus, the probability that the new manager will outperform the market next year is 58.14%.

Bayesian Approach vs. Frequentist Approach

LO 18.2: Compare the Bayesian approach to the frequentist approach.

The **frequentist approach** involves drawing conclusions from sample data based on the frequency of that data. For example, the approach suggests that the probability of a positive event will be 100% if the sample data consists of only observations that are positive events. The primary difference between the Bayesian approach and the frequentist approach is that the Bayesian approach is instead based on a prior belief regarding the probability of an event occurring.

In the previous examples, we began under the assumptions that excellent managers outperform the market 70% of the time, average managers outperform the market only 50% of the time, and only 20% of all managers are excellent. The Bayesian approach was used to update the probabilities that the new manager is either an excellent manager (updated from 20% to 40.7%) or an average manager (updated from 80% to 59.3%). These updated probabilities were based on the new information that the manager outperformed the market three years in a row. Next, under the Bayesian approach, the updated probabilities were used to determine the probability that the new manager outperforms the market next year. The Bayesian approach determined that there is a 58.14% probability that the new manager will outperform the market next year.

Conversely, under the frequentist approach there is a 100% probability that the new manager outperforms the market next year. There was a sample of three years with the manager outperforming the market each year (i.e., 3 out of 3 = 100%). The frequentist approach is simply based on the observed frequency of positive events occurring.

Obviously, the frequentist approach is questionable with a small sample size. It is difficult to believe that there is no way the new manager can underperform the market next year. However, individuals who apply the frequentist approach point out the weakness in relying on prior beliefs in the Bayesian approach. The Bayesian approach requires a beginning assumption regarding probabilities. In the prior examples, we assumed specific probabilities for a manager being excellent or average and specific probabilities related to the probability of outperforming the market for each type of manager. These prior assumptions are often based on a frequentist approach (i.e., number of events occurring during a sample period) or some other subjective analysis.

With small sample sizes, such as three years of historical performance, the Bayesian approach is often used in practice. With larger sample sizes, most analysts tend to use the frequentist approach. The frequentist approach is also often used because it is easier to implement and understand than the Bayesian approach.

Bayes' Theorem with Multiple States

LO 18.3: Apply Bayes' theorem to scenarios with more than two possible outcomes and calculate posterior probabilities.

In prior examples, we assumed there were only two possible outcomes where either a manager was excellent or average. Suppose now that we add another possible outcome where a manager is below average. The prior belief regarding the probabilities of a manager outperforming the market are 80% for an excellent manager, 50% for an average manager, and 20% for a below average manager. Furthermore, there is a 15% probability that a manager is excellent, a 55% probability that a manager is average, and a 30% probability that a manager is below average. These probabilities of manager performance are noted as follows:

$$P(p = 0.8) = 15\%$$
$$P(p = 0.5) = 55\%$$
$$P(p = 0.2) = 30\%$$

Example: Bayes' theorem with three outcomes

Suppose a new fund manager outperforms the market two years in a row. Given the manager performance probabilities listed previously, how is Bayesian analysis applied to update prior expectations regarding the new manager's ability?

Answer:

The first step is to calculate the probability of each type of manager outperforming the market two years in a row, assuming the probability of outperforming the market is independent for each year. The probability that an excellent manager outperforms the market two years in a row is calculated by multiplying 80% by 80%. Thus, the probability that an excellent manager outperforms the market two years in a row is 64%.

$$P(O \mid p = 0.8) = 0.8^2 = 0.64$$

The probability that an average manager outperforms the market two years in a row is 25%.

$$P(O \mid p = 0.5) = 0.5^2 = 0.25$$

The probability that a below average manager outperforms the market two years in a row is 4%.

$$P(O \mid p = 0.2) = 0.2^2 = 0.04$$

Next, we calculate the unconditional probability of a random manager outperforming the market two years in a row. Previously, with two possible outcomes, we used a weighted average of probabilities to calculate unconditional probabilities. This weighted average is now updated to include a third possible outcome for below average managers. The weights are based on prior beliefs regarding the probabilities that a manager is excellent (15%), average (55%), or below average (30%). The following calculation determines the unconditional probability that a manager outperforms the market two years in a row.

$$P(O) = (15\% \times 64\%) + (55\% \times 25\%) + (30\% \times 4\%) = 0.096 + 0.1375 + 0.012 = 0.2455$$

We now use Bayes' theorem to update our beliefs that the manager is excellent, average, or below average by calculating the following **posterior probabilities:**

$$P(p = 0.8 \mid O) = \frac{P(O \mid p = 0.8) \times P(p = 0.8)}{P(O)} = \frac{0.64 \times 0.15}{0.2455} = 39.1\%$$

$$P(p = 0.5 \mid O) = \frac{P(O \mid p = 0.5) \times P(p = 0.5)}{P(O)} = \frac{0.25 \times 0.55}{0.2455} = 56.01\%$$

$$P(p = 0.2 \mid O) = \frac{P(O \mid p = 0.2) \times P(p = 0.2)}{P(O)} = \frac{0.04 \times 0.3}{0.2455} = 4.89\%$$

Notice that after the new manager outperforms the market for two consecutive years, the probability that the manager is an excellent manager more than doubles from 15% to 39.1%. In this example, the 15% is known as a prior belief, which is set *before* seeing the manager outperform the market two years in a row. The 39.1% is known as a posterior belief, which is set *after* seeing the manager outperform the market two years in a row. The updated probability that the manager is average goes up slightly from 55% to 56.01%, and the updated probability that the manager is below average goes down significantly from 30% to 4.89%. Notice that the updated probabilities still sum to 100% (= 39.1% + 56.01% + 4.89%).

Key Concepts

LO 18.1

Bayes' theorem is defined for two random variables *A* and *B* as follows:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

LO 18.2

The primary difference between the Bayesian and frequentist approaches is that the Bayesian approach is based on a prior belief regarding the probability of an event occurring, while the frequentist approach is based on a number or frequency of events occurring during the most recent sample.

LO 18.3

Bayes' theorem can be extended to include more than two possible outcomes. Given the numerous calculations involved when incorporating multiple states, it is helpful to solve these types of problems using spreadsheet software.

Concept Checkers

Use the following information to answer Questions 1 through 3

Suppose a manager for a fund of funds uses historical data to categorize managers as excellent or average. Based on historical performance, the probabilities of excellent and average managers outperforming the market are 80% and 50%, respectively. Assume that the probabilities for managers outperforming the market is independent of their performance in prior years. In addition, the fund of funds manager believes that only 15% of total fund managers are excellent managers. Assume that a new manager started three years ago and beat the market in each of the past three years.

1. Using the Bayesian approach, what is the approximate probability that the new manager is an excellent manager today?
   A. 18.3%.
   B. 27.5%.
   C. 32.1%.
   D. 42.0%.

2. What is the approximate probability that the new manager will outperform the market next year using the Bayesian approach?
   A. 31.9%.
   B. 51.2%.
   C. 62.6%.
   D. 80.0%.

3. What is the probability that the new manager will outperform the market next year using the frequentist approach?
   A. 41.9%.
   B. 51.2%.
   C. 80.0%.
   D. 100.0%.

Use the following information to answer Questions 4 and 5

Suppose a pension fund gathers information on portfolio managers to rank their abilities as excellent, average, or below average. The analyst for the pension fund forms prior beliefs regarding the probabilities of a manager outperforming the market based on historical performances of all managers. There is a 10% probability that a manager is excellent, a 60% probability that a manager is average, and a 30% probability that a manager is below average. In addition, the probabilities of a manager outperforming the market are 75% for an excellent manager, 50% for an average manager, and 25% for a below average manager. Assume the probability of the manager outperforming the market is independent of the prior year performance.

4. What is the probability of a new manager outperforming the market two years in a row?
   A. 18.50%.
   B. 22.50%.
   C. 37.25%.
   D. 56.25%.

5. Suppose a new manager just outperformed the market two years in a row. Using Bayesian analysis, what is the updated belief or probability that the new manager is excellent?
   A. 20.0%.
   B. 22.5%.
   C. 25.0%.
   D. 27.5%.

Concept Checker Answers

1. **D** Excellent managers are expected to outperform the market 80% of the time. The probability of an excellent manager outperforming three years in a row is 0.8^3 or 51.2%. Similarly, the probability of an average manager outperforming the market three years in a row is determined by taking the 50% probability to the third power: $0.5^3 = 0.125$.

The probability that the new manager is excellent after beating the market three years in a row is determined by the following Bayesian approach:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)}$$

The denominator is the unconditional probability of outperforming the market for three years in a row. This is computed as a weighted average of the probabilities of outperforming three years in a row for each type of manager.

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.512 \times 0.15) + (0.125 \times 0.85) \\ &= 0.0768 + 0.10625 \\ &= 0.18305 \end{aligned}$$

With this information, we can now apply the Bayesian approach as follows:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)} = \frac{0.512 \times 0.15}{0.18305} = 41.956\%$$

2. **C** The probability of the new manager outperforming the market next year is the unconditional probability of outperforming the market based on the new probability that the new manager is an excellent manager after outperforming the market three years in a row. From Question 1, we determined the probability that the new manager is excellent after beating the market three years in a row as:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)} = \frac{0.512 \times 0.15}{0.18305} = 41.956\%$$

The probability that the new manager is average after beating the market three years in a row is determined as:

$$P(A \mid O) = \frac{P(O \mid A) \times P(A)}{P(O)} = \frac{0.125 \times 0.85}{0.18305} = 58.044\%$$

Next, these new probabilities are now used to determine the unconditional probability of outperforming the market next year.

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.8 \times 0.41956) + (0.5 \times 0.58044) \\ &= 0.3356 + 0.2902 \\ &= 0.6258 \text{ or } 62.58\% \end{aligned}$$

3. **D** The frequentist approach determines the probability based on the outperformance for the most recent sample size. In this example, there are only three years of data and the new manager outperformed the market every year. Thus, there is a 100% probability under this approach (3 out of 3) that the new manager will outperform the market next year.

4. **B** To answer this question, you need to determine the unconditional probability of outperforming the market two years in a row. The first step is to calculate the probability of each type of manager outperforming the market two years in a row.

   The probability that an excellent manager outperforms the market two years in a row is:

   $P(O \mid p = 0.75) = 0.75^2 = 0.5625$

   The probability that an average manager outperforms the market two years in a row is:

   $P(O \mid p = 0.5) = 0.5^2 = 0.25$

   The probability that a below average manager outperforms the market two years in a row is:

   $P(O \mid p = 0.25) = 0.25^2 = 0.0625$

   Next, calculate the unconditional probability that a new manager outperforms the market two years in a row based on prior expectations or beliefs:

   $P(O) = (10\% \times 56.25\%) + (60\% \times 25\%) + (30\% \times 6.25\%) = 0.05625 + 0.15 + 0.01875 = 0.225$ or 22.5%

5. **C** From Question 4, we know the unconditional probability that a new manager outperforms the market two years in a row based on prior expectations or beliefs is:

   $P(O) = (10\% \times 56.25\%) + (60\% \times 25\%) + (30\% \times 6.25\%) = 0.05625 + 0.15 + 0.01875 = 0.225$ or 22.5%

   With this information, we can apply Bayes' theorem to update our beliefs that the manager is excellent:

$$P(p = 0.75 \mid O) = \frac{P(O \mid p = 0.75) \times P(p = 0.75)}{P(O)} = \frac{0.5625 \times 0.1}{0.225} = 25\%$$

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Hypothesis Testing and Confidence Intervals

Topic 19

Exam Focus

This topic provides insight into how risk managers make portfolio decisions on the basis of statistical analysis of samples of investment returns or other random economic and financial variables. We first focus on the estimation of sample statistics and the construction of confidence intervals for population parameters based on sample statistics. We then discuss hypothesis testing procedures used to conduct tests concerned with population means and population variances. Specific tests reviewed include the z-test and the t-test. For the exam, you should be able to construct and interpret a confidence interval and know when and how to apply each of the test statistics discussed when conducting hypothesis testing.

Applied Statistics

In many real-world statistics applications, it is impractical (or impossible) to study an entire population. When this is the case, a subgroup of the population, called a sample, can be evaluated. Based upon this sample, the parameters of the underlying population can be estimated.

For example, rather than attempting to measure the performance of the U.S. stock market by observing the performance of all 10,000 or so stocks trading in the United States at any one time, the performance of the subgroup of 500 stocks in the S&P 500 can be measured. The results of the statistical analysis of this sample can then be used to draw conclusions about the entire population of U.S. stocks.

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 you want to draw a sample of five 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 nth member from a population.

Sampling error is the difference between a sample statistic (the mean, variance, or standard deviation of the sample) and its corresponding population parameter (the true mean, variance, or standard deviation of the population). For example, the sampling error for the mean is as follows:

$$\text{sampling error of the mean} = \text{sample mean} - \text{population mean} = \bar{x} - \mu$$

Mean and Variance of the Sample Average

It is important to recognize that the sample statistic itself is a random variable and, therefore, has a probability distribution. The **sampling distribution** of the sample statistic is a probability distribution of all possible sample statistics computed from a set of equal-size samples that were randomly drawn from the same population. Think of it as the probability distribution of a statistic from many samples.

For example, suppose a random sample of 100 bonds is selected from a population of a major municipal bond index consisting of 1,000 bonds, and then the mean return of the 100-bond sample is calculated. Repeating this process many times will result in many different estimates of the population mean return (i.e., one for each sample). The distribution of these estimates of the mean is the *sampling distribution of the mean*. It is important to note that this sampling distribution is distinct from the distribution of the actual prices of the 1,000 bonds in the underlying population and will have different parameters.

To illustrate the mean of the sample average, suppose we have selected two independent and identically distributed (i.i.d.) variables at random, X_1 and X_2, from a population. Since these two variables are i.i.d., the mean and variance for both observations will be the same, respectively.

Recall from Topic 16, the mean of the sum of two random variables is equal to:

$$E(X_1 + X_2) = \mu_X + \mu_X = 2\mu_X$$

Thus, the mean of the sample average, $E(\overline{X})$, will be equal to:

$$E\left(\frac{X_1 + X_2}{2}\right) = \frac{2\mu_X}{2} = \mu_X$$

More generally, we can say that for *n* observations:

$$E(\overline{X}) = \mu_X$$

By applying the properties of variance for the sums of independent random variables, we can also calculate the variance of the sample average. Recall, that for independent variables, the covariance term in the variance equation will equal zero. For two observations, the variance of the sum of two random variables will equal:

$$Var(X_1 + X_2) = 2\sigma_x^2$$

Thus, when applying the following variance property:

$$\text{Var}(aX_1 + cX_2) = a^2 \times \text{Var}(X_1) + c^2 \times \text{Var}(X_2)$$

and assuming *a* and *c* are equal to 0.5, the variance of the sample average, $\text{Var}(\overline{X})$, will be equal to $\frac{\sigma_X^2}{2}$. In more general terms, $\text{Var}(\overline{X}) = \frac{\sigma_X^2}{n}$ for *n* observations, and the standard deviation of the sample average is equal to $\frac{\sigma_X}{\sqrt{n}}$. This standard deviation measure is known as the **standard error**.

These properties help define the distributional characteristics of the sample distribution of the mean and allow us to make assumptions about the distribution when the sample size is large.

LO 19.1: Calculate and interpret the sample mean and sample variance.

Population and Sample Mean

Recall from Topic 16, that in order to compute the **population mean**, all the observed values in the population are summed (ΣX) and divided by the number of observations in the population, *N*.

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

The **sample mean** 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.

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

Population and Sample Variance

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 **population variance** is defined as the average of the squared deviations from the mean. The population variance (σ^2) uses the values for all members of a population and is calculated using the following formula:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Example: Population variance, σ^2

Assume the following 5-year annualized total returns represent all of the managers at a small investment firm (30%, 12%, 25%, 20%, 23%). What is the population variance of these returns?

Answer:

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

$$\sigma^2 = \frac{\left[(30-22)^2 + (12-22)^2 + (25-22)^2 + (20-22)^2 + (23-22)^2\right]}{5} = 35.60\left(\%^2\right)$$

Interpreting this result, we can say that the average variation from the mean return is 35.60% squared. Had we done the calculation using decimals instead of whole percents, the variance would be 0.00356.

A major problem with using the 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 percents, squared dollars, or squared yen? This problem is mitigated through the use of the *standard deviation*. The population standard deviation, σ, is the square root of the population variance and is calculated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X - \mu)^2}{N}}$$

Example: Population standard deviation, σ

Using the data from the preceding example, **compute** the population standard deviation.

Answer:

$$\sigma = \sqrt{\frac{(30-22)^2 + (12-22)^2 + (25-22)^2 + (20-22)^2 + (23-22)^2}{5}}$$

$$= \sqrt{35.60} = 5.97\%$$

Calculated with decimals instead of whole percents, we would get:

$$\sigma^2 = 0.00356 \text{ and } \sigma = \sqrt{0.00356} = 0.05966 = 5.97\%$$

Since the population standard deviation and population mean are both expressed in the same units (percent), these values are easy to relate. The outcome of this example indicates that the mean return is 22% and the standard deviation about the mean is 5.97%.

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 most noteworthy difference from the formula for population variance is that the denominator for s^2 is $n - 1$, one less than the sample size *n*, where σ^2 uses the entire population size *N*. Another difference is the use of the sample mean, $\overline{X}$, instead of the population mean, μ. 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 parameter, σ^2, particularly for small sample sizes. This systematic underestimation causes the sample variance to be what is referred to as 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 σ^2. Thus, s^2, as expressed in the equation, is considered to be an unbiased estimator of σ^2.

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:

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

$$s^2 = \frac{\left[(30-22)^2+(12-22)^2+(25-22)^2+(20-22)^2+(23-22)^2\right]}{5-1} = 44.5\left(\%^2\right)$$

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 percent returns in decimal form [e.g., $(0.30 - 0.22)^2$, and so forth.].

As with the population standard deviation, the sample standard deviation can be calculated by taking the square root of the sample variance. The sample standard deviation, *s*, is defined as:

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

Example: Sample standard deviation

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

Answer:

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

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

The results shown here mean that the sample standard deviation, $s = 6.67\%$, can be interpreted as an unbiased estimator of the population standard deviation, σ.

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

Example: Standard error of sample mean (known population variance)

The mean hourly wage for Iowa farm workers is \$13.50 with a *population standard deviation* of \$2.90. **Calculate** and **interpret** the standard error of the sample mean for a sample size of 30.

Answer:

Because the population standard deviation, σ, *is known*, the standard error of the sample mean is expressed as:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$2.90}{\sqrt{30}} = \$0.53$$

Professor's Note: On the TI BAII Plus, the use of the square root key is obvious. On the HP 12C, the square root of 30 is computed as: [30] [g] [$\sqrt{x}$].

This means that if we were to take all possible samples of size 30 from the Iowa farm worker population and prepare a sampling distribution of the sample means, we would get a distribution with a mean of \$13.50 and standard error of \$0.53.

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* mean 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:

Since σ is unknown, the standard error of the sample mean is:

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

This implies that if we took all possible samples of size 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. In order 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:

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

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.

Population and Sample Covariance

The covariance between two random variables is a statistical measure of the degree to which the two variables move together. The covariance captures the linear relationship between one variable and another. A positive covariance indicates that the variables tend to move together; a negative covariance indicates that the variables tend to move in opposite directions.

The population and sample covariances are calculated as:

$$\text{population cov}_{XY} = \frac{\sum_{i=1}^{N}(X_i - \mu_x)(Y_i - \mu_Y)}{N}$$

$$\text{sample cov}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

The actual value of the covariance is not very meaningful because its measurement is extremely sensitive to the scale of the two variables. Also, the covariance may range from negative to positive infinity and it is presented in terms of squared units (e.g., percent squared). For these reasons, we take the additional step of calculating the correlation coefficient (see Topic 16), which converts the covariance into a measure that is easier to interpret.

Confidence Intervals

LO 19.2: Construct and interpret a confidence interval.

Confidence interval estimates result in a range of values within which the actual value of a parameter will lie, given the probability of $1 - \alpha$. Here, alpha, α, is called the *level of significance* for the confidence interval, and the probability $1 - \alpha$ is referred to as the *degree of confidence*. For example, we might estimate that the population mean of random variables will range from 15 to 25 with a 95% degree of confidence, or at the 5% level of significance.

Confidence intervals are usually constructed by adding or subtracting an appropriate value from the point estimate. In general, confidence intervals take on the following form:

point estimate ± (reliability factor × standard error)

where:
point estimate = value of a sample statistic of the population parameter
reliability factor = number that depends on the sampling distribution of the point estimate and the probability that the point estimate falls in the confidence interval, $(1 - \alpha)$
standard error = standard error of the point estimate

If the population has a *normal distribution with a known variance*, a confidence interval for the population mean can be calculated as:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where:

$\bar{x}$ = point estimate of the population mean (sample mean)

$z_{\alpha/2}$ = reliability factor, a standard normal random variable for which the probability in the right-hand tail of the distribution is $\alpha/2$. In other words, this is the *z*-score that leaves $\alpha/2$ of probability in the upper tail.

$\frac{\sigma}{\sqrt{n}}$ = the standard error of the sample mean where σ is the known standard deviation of the population, and *n* is the sample size

The most commonly used standard normal distribution reliability factors are:

$z_{\alpha/2}$ = 1.65 for 90% confidence intervals (the significance level is 10%, 5% in each tail).

$z_{\alpha/2}$ = 1.96 for 95% confidence intervals (the significance level is 5%, 2.5% in each tail).

$z_{\alpha/2}$ = 2.58 for 99% confidence intervals (the significance level is 1%, 0.5% in each tail).

Do these numbers look familiar? They should! In Topic 17, we found the probability under the standard normal curve between z = –1.96 and z = +1.96 to be 0.95, or 95%. Owing to symmetry, this leaves a probability of 0.025 under each tail of the curve beyond z = –1.96 or z = +1.96, for a total of 0.05, or 5%—just what we need for a significance level of 0.05, or 5%.

Example: Confidence interval

Consider a practice exam that was administered to 36 FRM Part I candidates. The mean score on this practice exam was 80. Assuming a population standard deviation equal to 15, **construct** and **interpret** a 99% confidence interval for the mean score on the practice exam for 36 candidates. *Note that in this example the population standard deviation is known, so we don't have to estimate it.*

Answer:

At a confidence level of 99%, $z_{\alpha/2} = z_{0.005} = 2.58$. So, the 99% confidence interval is calculated as follows:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 80 \pm 2.58 \frac{15}{\sqrt{36}} = 80 \pm 6.45$$

Thus, the 99% confidence interval ranges from 73.55 to 86.45.

Confidence intervals can be interpreted from a probabilistic perspective or a practical perspective. With regard to the outcome of the practice exam example, these two perspectives can be described as follows:

- *Probabilistic interpretation*. After repeatedly taking samples of exam candidates, administering the practice exam, and constructing confidence intervals for each sample's mean, 99% of the resulting confidence intervals will, in the long run, include the population mean.
- *Practical interpretation*. We are 99% confident that the population mean score is between 73.55 and 86.45 for candidates from this population.

Confidence Intervals for a Population Mean: Normal With Unknown Variance

If the distribution of the *population is normal with unknown variance*, we can use the *t*-distribution to construct a confidence interval:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where:

$\bar{x}$ = the point estimate of the population mean

$t_{\alpha/2}$ = the *t*-reliability factor (i.e., *t*-statistic or critical *t*-value) corresponding to a *t*-distributed random variable with n – 1 degrees of freedom, where *n* is the sample size. The area under the tail of the *t*-distribution to the right of $t_{\alpha/2}$ is $\alpha/2$.

$\frac{s}{\sqrt{n}}$ = standard error of the sample mean

s = sample standard deviation

Unlike the standard normal distribution, the reliability factors for the *t*-distribution depend on the sample size, so we can't rely on a commonly used set of reliability factors. Instead, reliability factors for the *t*-distribution have to be looked up in a table of Student's *t*-distribution, like the one at the back of this book.

Owing to the relatively fatter tails of the *t*-distribution, confidence intervals constructed using *t*-reliability factors ($t_{\alpha/2}$) will be more conservative (wider) than those constructed using *z*-reliability factors ($z_{\alpha/2}$).

Example: Confidence intervals

Let's return to the McCreary, Inc. example. Recall that we took a sample of the past 30 monthly stock returns for McCreary, Inc. and determined that the mean return was 2% and the sample standard deviation was 20%. Since the population variance is unknown, the standard error of the sample was estimated to be:

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

Now, let's construct a 95% confidence interval for the mean monthly return.

Answer:

Here, we will use the *t*-reliability factor because the population variance is unknown. Since there are 30 observations, the degrees of freedom are 29 = 30 – 1. Remember, because this is a two-tailed test at the 95% confidence level, the probability under each tail must be $\alpha/2 = 2.5\%$, for a total of 5%. So, referencing the one-tailed probabilities for Student's *t*-distribution at the back of this book, we find the critical *t*-value (reliability factor) for $\alpha/2 = 0.025$ and df = 29 to be $t_{29,\,2.5} = 2.045$. Thus, the 95% confidence interval for the population mean is:

$$2\% \pm 2.045\left(\frac{20\%}{\sqrt{30}}\right) = 2\% \pm 2.045(3.6\%) = 2\% \pm 7.4\%$$

Thus, the 95% confidence has a lower limit of –5.4% and an upper limit of +9.4%.

We can interpret this confidence interval by saying we are 95% confident that the population mean monthly return for McCreary stock is between –5.4% and +9.4%.

Professor's Note: You should practice looking up reliability factors (i.e., critical t-values or t-statistics) in a t-table. The first step is always to compute the degrees of freedom, which is n – 1. The second step is to find the appropriate level of alpha or significance. This depends on whether the test you're concerned with is one-tailed (use α) or two-tailed (use $\alpha/2$). To look up $t_{29,\,2.5}$, find the 29 df row and match it with the 0.025 column; t = 2.045 is the result. We'll do more of this in our study of hypothesis testing.

Confidence Interval for a Population Mean: Nonnormal With Unknown Variance

We now know that the *z*-statistic should be used to construct confidence intervals when the population distribution is normal and the variance is known, and the *t*-statistic should be used when the distribution is normal but the variance is unknown. But what do we do when the distribution is *nonnormal*?

As it turns out, the size of the sample influences whether or not we can construct the appropriate confidence interval for the sample mean.

- If the *distribution is nonnormal* but the *population variance is known*, the *z*-statistic can be used as long as the sample size is large ($n \geq 30$). We can do this because the central limit theorem assures us that the distribution of the sample mean is approximately normal when the sample is large.
- If the *distribution is nonnormal* and the *population variance is unknown*, the *t*-statistic can be used as long as the sample size is large ($n \geq 30$). It is also acceptable to use the *z*-statistic, although use of the *t*-statistic is more conservative.

This means that if we are sampling from a nonnormal distribution (which is sometimes the case in finance), *we cannot create a confidence interval if the sample size is less than 30*. So, all else equal, make sure you have a sample of at least 30, and the larger, the better.

Figure 1: Criteria for Selecting the Appropriate Test Statistic

When sampling from a:	*Test Statistic*	
	Small Sample $(n < 30)$	*Large Sample* $(n \geq 30)$
Normal distribution with *known* variance	*z*-statistic	*z*-statistic
Normal distribution with *unknown* variance	*t*-statistic	*t*-statistic*
Nonnormal distribution with *known* variance	not available	*z*-statistic
Nonnormal distribution with *unknown* variance	not available	*t*-statistic*

* The *z*-statistic is theoretically acceptable here, but use of the *t*-statistic is more conservative.

All of the preceding analysis depends on the sample we draw from the population being random. If the sample isn't random, the central limit theorem doesn't apply, our estimates won't have the desirable properties, and we can't form unbiased confidence intervals. Surprisingly, creating a *random sample* is not as easy as one might believe. There are a number of potential mistakes in sampling methods that can bias the results. These biases are particularly problematic in financial research, where available historical data are plentiful, but the creation of new sample data by experimentation is restricted.

Hypothesis Testing

LO 19.3: Construct an appropriate null and alternative hypothesis, and calculate an appropriate test statistic.

Hypothesis testing is the statistical assessment of a statement or idea regarding a population. For instance, a statement could be, "The mean return for the U.S. equity market is greater than zero." Given the relevant returns data, hypothesis testing procedures can be employed to test the validity of this statement at a given significance level.

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 2.

Figure 2: 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 and Alternative Hypothesis

The **null hypothesis**, designated H_0, is the hypothesis 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 a simple statement about a population parameter. Typical statements of the null hypothesis for the population mean include H_0: $\mu = \mu_0$, H_0: $\mu \leq \mu_0$, and H_0: $\mu \geq \mu_0$, 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. It is usually the alternative hypothesis the researcher is really trying to assess. Why? Since you can never really prove anything with statistics, when the null hypothesis is discredited, the implication is that the alternative hypothesis is valid.

The Choice of the Null and Alternative Hypotheses

The most common null hypothesis will be an "equal to" hypothesis. The alternative is often the hoped-for hypothesis. When the null is that a coefficient is equal to zero, we hope to reject it and show the significance of the relationship.

When the null is less than or equal to, the (mutually exclusive) alternative is framed as greater than. If we are trying to demonstrate that a return is greater than the risk-free rate, this would be the correct formulation. We will have set up the null and alternative hypothesis so rejection of the null will lead to acceptance of the alternative, our goal in performing the test.

Hypothesis testing involves two statistics: the *test statistic* calculated from the sample data and 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.

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). With reference to our option return example, this means we are concerned with the difference between the mean return of the sample and the hypothesized mean return. 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:

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

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-squared 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.

One-Tailed and Two-Tailed Tests of Hypotheses

LO 19.4: Differentiate between a one-tailed and a two-tailed test and identify when to use each test.

The alternative hypothesis can be one-sided or two-sided. A one-sided test is referred to as a **one-tailed test**, and a two-sided test is referred to as a **two-tailed test**. Whether the test is one- or two-sided depends on the proposition being tested. If a researcher wants to test whether the return on stock options is greater than zero, a one-tailed test should be used. However, a two-tailed test should be used if the research question is whether the return on options is simply different from zero. Two-sided tests allow for deviation on both sides of the hypothesized value (zero). In practice, most hypothesis tests are constructed as two-tailed tests.

A two-tailed test for the population mean may be structured as:

H_0: $\mu = \mu_0$ versus H_A: $\mu \neq \mu_0$

Since the alternative hypothesis allows for values above and below the hypothesized parameter, a two-tailed test uses two critical values (or rejection points).

The ***general decision rule for a two-tailed test*** is:

Reject H_0 if: test statistic > upper critical value or
test statistic < lower critical value

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, $\alpha = 0.05$.

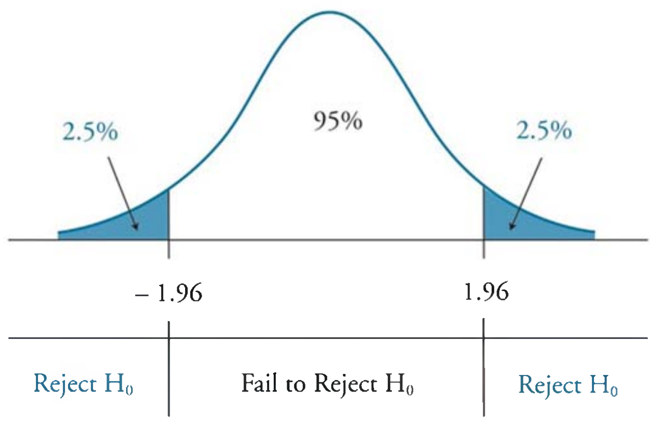
- At $\alpha = 0.05$, the computed test statistic is compared with the critical z-values of ±1.96. The values of ±1.96 correspond to $\pm z_{\alpha/2} = \pm z_{0.025}$, which is the range of *z*-values within which 95% of the probability lies. These values are obtained from the cumulative probability table for the standard normal distribution (*z*-table), which is included at the back 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:

Reject H_0 if: test statistic < –1.96 or
test statistic > 1.96

Figure 3 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 3: Two-Tailed Hypothesis Test Using the Standard Normal (*z*) Distribution



Example: Two-tailed test

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 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. The null hypothesis is the one the researcher expects to reject.

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

Since the null hypothesis is an equality, this is a two-tailed test. 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

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

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

Since our sample statistic here is a sample mean, the standard error of the sample mean for a sample size of 250 is $\frac{0.0025}{\sqrt{250}}$ and our test statistic is:

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

Since 6.33 > 1.96, we reject the null hypothesis that the mean daily option return is equal to zero. Note that when we reject the null, we conclude that the sample value is significantly different from the hypothesized value. We are saying that the two values are different from one another *after considering the variation in the sample*. That is, the mean daily return of 0.001 is statistically different from zero given the sample's standard deviation and size.

For a one-tailed hypothesis test of the population mean, the null and alternative hypotheses are either:

Upper tail: H_0: $\mu \leq \mu_0$ versus H_A: $\mu > \mu_0$, or
Lower tail: H_0: $\mu \geq \mu_0$ versus H_A: $\mu < \mu_0$

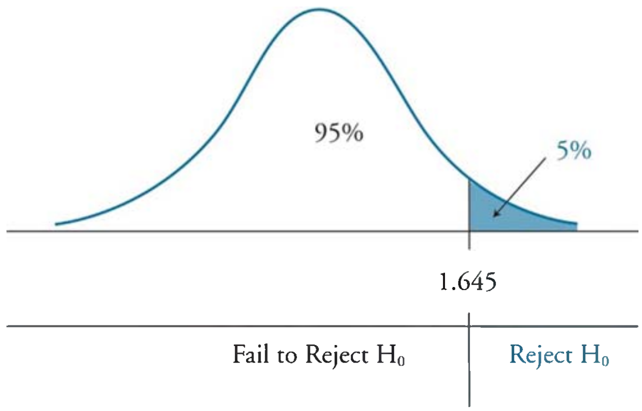
The appropriate set of hypotheses depends on whether we believe the population mean, μ, to be greater than (upper tail) or less than (lower tail) the hypothesized value, μ_0. Using a *z*-test at the 5% level of significance, the computed test statistic is compared with the critical values of 1.645 for the upper tail tests (i.e., H_A: $\mu > \mu_0$) or –1.645 for lower tail tests (i.e., H_A: $\mu < \mu_0$). These critical values are obtained from a *z*-table, where $-z_{0.05} = -1.645$ corresponds to a cumulative probability equal to 5%, and the $z_{0.05} = 1.645$ corresponds to a cumulative probability of 95% (1 – 0.05).

Let's use the upper tail test structure where H_0: $\mu \leq \mu_0$ and H_A: $\mu > \mu_0$.

- If the calculated test statistic is greater than 1.645, we conclude that the sample statistic is sufficiently greater than the hypothesized value. In other words, we reject the null hypothesis.
- If the calculated test statistic is less than 1.645, we conclude that the sample statistic is not sufficiently different from the hypothesized value, and we fail to reject the null hypothesis.

Figure 4 shows the standard normal distribution and the rejection region for a one-tailed test (upper tail) at the 5% level of significance.

Figure 4: One-Tailed Hypothesis Test Using the Standard Normal (z) Distribution



Example: One-tailed test

Perform a z-test using the option portfolio data from the previous example to test the belief that option returns are positive.

Answer:

In this case, we use a one-tailed test with the following structure:

H_0: $\mu \leq 0$ versus H_A: $\mu > 0$

The appropriate decision rule for this one-tailed z-test at a significance level of 5% is:

Reject H_0 if: test statistic > 1.645

The test statistic is computed the same way, regardless of whether we are using a one-tailed or two-tailed test. From the previous example, we know the test statistic for the option return sample is 6.33. Since 6.33 > 1.645, we reject the null hypothesis and conclude that mean returns are statistically greater than zero at a 5% level of significance.

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: the rejection of the null hypothesis when it is actually true.
- Type II error: 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 in order to identify the critical values needed to evaluate the test statistic.

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 Power of a Test

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

Figure 5: Type I and Type II Errors in Hypothesis Testing

	True Condition	
Decision	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 Relation Between Confidence Intervals and Hypothesis Tests

A confidence interval is a range of values within which the researcher believes the true population parameter may lie.

A confidence interval is determined as:

$$\left\{\left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix} - \begin{pmatrix}\text{critical}\\\text{value}\end{pmatrix}\begin{pmatrix}\text{standard}\\\text{error}\end{pmatrix}\right] \leq \begin{matrix}\text{population}\\\text{parameter}\end{matrix} \leq \left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix} + \begin{pmatrix}\text{critical}\\\text{value}\end{pmatrix}\begin{pmatrix}\text{standard}\\\text{error}\end{pmatrix}\right]\right\}$$

The interpretation of a confidence interval is that for a level of confidence of 95%, for example, there is a 95% probability that the true population parameter is contained in the interval.

From the previous expression, we see that a confidence interval and a hypothesis test are linked by the critical value. For example, a 95% confidence interval uses a critical value associated with a given distribution at the 5% level of significance. Similarly, a hypothesis test would compare a test statistic to a critical value at the 5% level of significance. To see this relationship more clearly, the expression for the confidence interval can be manipulated and restated as:

$$-\text{critical value} \leq \text{test statistic} \leq +\text{critical value}$$

This is the range within which we fail to reject the null for a two-tailed hypothesis test at a given level of significance.

Example: Confidence interval

Using option portfolio data from the previous examples, **construct** a 95% confidence interval for the population mean daily return over the 250-day sample period. Use a *z*-distribution. **Decide** if the hypothesis μ = 0 should be rejected.

Answer:

Given a sample size of 250 with a standard deviation of 0.25%, the standard error can be computed as $s_{\bar{x}} = \frac{s}{\sqrt{n}} = 0.25/\sqrt{250} = 0.0158\%$.

At the 5% level of significance, the critical *z*-values for the confidence interval are $z_{0.025}$ = 1.96 and $-z_{0.025}$ = −1.96. Thus, given a sample mean equal to 0.1%, the 95% confidence interval for the population mean is:

$$0.1 - 1.96(0.0158) \leq \mu \leq 0.1 + 1.96(0.0158), \text{ or}$$
$$0.069\% \leq \mu \leq 0.1310\%$$

Since there is a 95% probability that the true mean is within this confidence interval, we can reject the hypothesis μ = 0 because 0 is not within the confidence interval.

Notice the similarity of this analysis with our test of whether μ = 0. We rejected the hypothesis μ = 0 because the sample mean of 0.1% is more than 1.96 standard errors from zero. Based on the 95% confidence interval, we reject μ = 0 because zero is more than 1.96 standard errors from the sample mean of 0.1%.

Statistical Significance vs. Economic Significance

Statistical significance does not necessarily imply economic significance. For example, we may have tested a null hypothesis that a strategy of going long all the stocks that satisfy some criteria and shorting all the stocks that do not satisfy the criteria resulted in returns that were less than or equal to zero over a 20-year period. Assume we have rejected the null in favor of the alternative hypothesis that the returns to the strategy are greater than zero (positive). This does not necessarily mean that investing in that strategy will result in economically meaningful positive returns. Several factors must be considered.

One important consideration is transactions costs. Once we consider the costs of buying and selling the securities, we may find that the mean positive returns to the strategy are not enough to generate positive returns. Taxes are another factor that may make a seemingly attractive strategy a poor one in practice. A third reason that statistically significant results may not be economically significant is risk. In the strategy just discussed, we have additional risk from short sales (they may have to be closed out earlier than in the test strategy). Since the statistically significant results were for a period of 20 years, it may be the case that there is significant variation from year to year in the returns from the strategy, even though the mean strategy return is greater than zero. This variation in returns from period to period is an additional risk to the strategy that is not accounted for in our test of statistical significance.

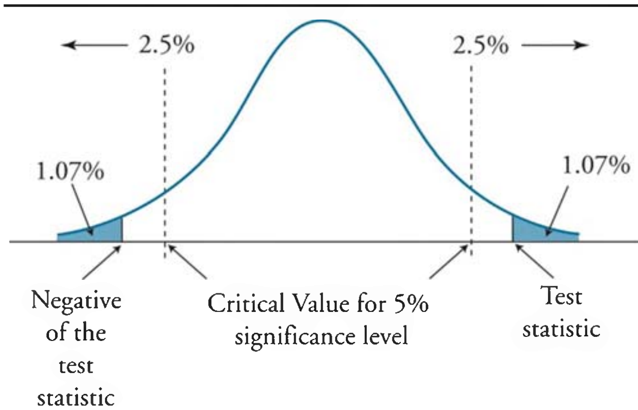
Any of these factors could make committing funds to a strategy unattractive, even though the statistical evidence of positive returns is highly significant. By the nature of statistical tests, a very large sample size can result in highly (statistically) significant results that are quite small in absolute terms.

THE *p*-VALUE

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. For one-tailed tests, the *p*-value is the probability that lies above the computed test statistic for upper tail tests or below the computed test statistic for lower tail tests. For two-tailed tests, the *p*-value is the probability that lies above the positive value of the computed test statistic *plus* the probability that lies below the negative value of the computed test statistic.

Consider a two-tailed hypothesis test about the mean value of a random variable at the 95% significance level where the test statistic is 2.3, greater than the upper critical value of 1.96. If we consult the *z*-table, we find the probability of getting a value greater than 2.3 is $(1 - 0.9893) = 1.07\%$. Since it's a two-tailed test, our *p*-value is $2 \times 1.07 = 2.14\%$, as illustrated in Figure 6. At a 3%, 4%, or 5% significance level, we would reject the null hypothesis, but at a 2% or 1% significance level, we would not. Many researchers report *p*-values without selecting a significance level and allow the reader to judge how strong the evidence for rejection is.

Figure 6: Two-Tailed Hypothesis Test with *p*-Value = 2.14%



THE *t*-TEST

When hypothesis testing, the choice between using a critical value based on the *t*-distribution or the *z*-distribution depends on sample size, the distribution of the population, and whether the variance of the population is known.

The *t*-test is a widely used hypothesis test that employs a test statistic that is distributed according to a *t*-distribution. Following are the rules for when it is appropriate to use the *t*-test for hypothesis tests of the population mean.

Use the t*-test if the population variance is unknown* and either of the following conditions exist:

- The sample is large ($n \geq 30$).
- The sample is small ($n < 30$), but the distribution of the population is normal or approximately normal.

If the sample is small and the distribution is non-normal, we have no reliable statistical test.

The computed value for the test statistic based on the *t*-distribution is referred to as the *t*-statistic. For hypothesis tests of a population mean, a *t*-statistic with n – 1 degrees of freedom is computed as:

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

where:
$\bar{x}$ = sample mean
μ_0 = hypothesized population mean (i.e., the null)
s = standard deviation of the sample
n = sample size

Professor's Note: This computation is not new. It is the same test statistic computation that we have been performing all along. Note the use of the sample standard deviation, s, *in the standard error term in the denominator.*

To conduct a *t*-test, the *t*-statistic is compared to a critical *t*-value at the desired level of significance with the appropriate degrees of freedom.

In the real world, the underlying variance of the population is rarely known, so the *t*-test enjoys widespread application.

The *z*-Test

The *z*-test is the appropriate hypothesis test of the population mean when the *population is normally distributed with known variance.* The computed test statistic used with the *z*-test is referred to as the *z*-statistic. The *z*-statistic for a hypothesis test for a population mean is computed as follows:

$$\text{z-statistic} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

where:
$\bar{x}$ = sample mean
μ_0 = hypothesized population mean
σ = standard deviation of the population
n = sample size

To test a hypothesis, the *z*-statistic is compared to the critical *z*-value corresponding to the significance of the test. Critical *z*-values for the most common levels of significance are displayed in Figure 7. You should memorize these critical values for the exam.

Figure 7: Critical *z*-Values

Level of Significance	*Two-Tailed Test*	*One-Tailed Test*
0.10 = 10%	±1.65	+1.28 or –1.28
0.05 = 5%	±1.96	+1.65 or –1.65
0.01 = 1%	±2.58	+2.33 or –2.33

When the *sample size is large* and the *population variance is unknown*, the *z*-statistic is:

$$\text{z-statistic} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

where:
$\bar{x}$ = sample mean
μ_0 = hypothesized population mean
s = standard deviation of the sample
n = sample size

Note the use of the sample standard deviation, *s*, versus the population standard deviation, σ. Remember, this is acceptable if the sample size is large, although the *t*-statistic is the more conservative measure when the population variance is unknown.

Example: *z*-test or *t*-test?

Referring to our previous option portfolio mean return problem once more, **determine** which test statistic (*z* or *t*) should be used and the difference in the likelihood of rejecting a true null with each distribution.

Answer:

The population variance for our sample of returns is unknown. Hence, the *t*-distribution is appropriate. With 250 observations, however, the sample is considered to be large, so the *z*-distribution would also be acceptable. This is a trick question—either distribution, *t* or *z*, is appropriate. With regard to the difference in the likelihood of rejecting a true null, since 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.

LO 19.5: Interpret the results of hypothesis tests with a specific level of confidence.

Example: The *z*-test

When your company's gizmo machine is working properly, the mean length of gizmos is 2.5 inches. However, from time to time the machine gets out of alignment and produces gizmos that are either too long or too short. When this happens, production is stopped and the machine is adjusted. To check the machine, the quality control department takes a gizmo sample each day. Today, a random sample of 49 gizmos showed a mean length of 2.49 inches. The population standard deviation is known to be 0.021 inches. Using a 5% significance level, **determine** if the machine should be shut down and adjusted.

Answer:

Let μ be the mean length of all gizmos made by this machine, and let $\bar{x}$ be the corresponding mean for the sample.

Let's follow the hypothesis testing procedure presented earlier in Figure 2. Again, you should know this process.

Statement of hypothesis. For the information provided, the null and alternative hypotheses are appropriately structured as:

H_0: μ = 2.5 (The machine does not need an adjustment.)
H_A: μ ≠ 2.5 (The machine needs an adjustment.)

Note that since this is a two-tailed test, H_A allows for values above and below 2.5.

Select the appropriate test statistic. Since the population variance is known and the sample size is > 30, the *z*-statistic is the appropriate test statistic. The *z*-statistic is computed as:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Specify the level of significance. The level of significance is given at 5%, implying that we are willing to accept a 5% probability of rejecting a true null hypothesis.

State the decision rule regarding the hypothesis. The ≠ sign in the alternative hypothesis indicates that the test is two-tailed with two rejection regions, one in each tail of the standard normal distribution curve. Because the total area of both rejection regions combined is 0.05 (the significance level), the area of the rejection region in each tail is 0.025. You should know that the critical *z*-values for $\pm z_{0.025}$ are ±1.96. This means that the null hypothesis should not be rejected if the computed z-statistic lies between –1.96 and +1.96 and should be rejected if it lies outside of these critical values. The decision rule can be stated as:

Reject H_0 if: z-statistic < $-z_{0.025}$ or z-statistic > $z_{0.025}$, or equivalently,
Reject H_0 if: z-statistic < –1.96 or z-statistic > + 1.96

Collect the sample and calculate the test statistic. The value of $\bar{x}$ from the sample is 2.49. Since σ is given as 0.021, we calculate the *z*-statistic using σ as follows:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{2.49 - 2.5}{0.021 / \sqrt{49}} = \frac{-0.01}{0.003} = -3.33$$

Make a decision regarding the hypothesis. The calculated value of the *z*-statistic is –3.33. Since this value is less than the critical value, $-z_{0.025} = -1.96$, it falls in the rejection region in the left tail of the *z*-distribution. Hence, there is sufficient evidence to reject H_0.

Make a decision based on the results of the test. Based on the sample information and the results of the test, it is concluded that the machine is out of adjustment and should be shut down for repair.

The Chi-Squared Test

The *chi-squared test* is used for hypothesis tests concerning the variance of a normally distributed population. Letting σ^2 represent the true population variance and σ_0^2 represent the hypothesized variance, the hypotheses for a two-tailed test of a single population variance are structured as:

$H_0: \sigma^2 = \sigma_0^2$ versus $H_A: \sigma^2 \neq \sigma_0^2$

The hypotheses for one-tailed tests are structured as:

$H_0: \sigma^2 \leq \sigma_0^2$ versus $H_A: \sigma^2 > \sigma_0^2$, or
$H_0: \sigma^2 \geq \sigma_0^2$ versus $H_A: \sigma^2 < \sigma_0^2$

Hypothesis testing of the population variance requires the use of a chi-squared distributed test statistic, denoted χ^2. The chi-squared distribution is asymmetrical and approaches the normal distribution in shape as the degrees of freedom increase.

To illustrate the chi-squared distribution, consider a two-tailed test with a 5% level of significance and 30 degrees of freedom. As displayed in Figure 8, the critical chi-squared values are 16.791 and 46.979 for the lower and upper bounds, respectively. These values are obtained from a chi-squared table, which is used in the same manner as a *t*-table. A portion of a chi-squared table is presented in Figure 9.

Note that the chi-squared values in Figure 9 correspond to the probabilities in the right tail of the distribution. As such, the 16.791 in Figure 8 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 9 shows that the critical chi-squared values for a two-tailed test are 3.247 and 20.483.

Figure 8: Decision Rule for a Two-Tailed Chi-Squared Test

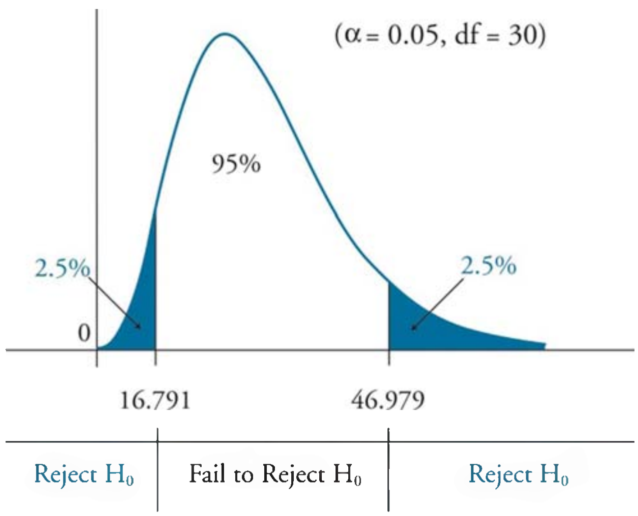


Figure 9: Chi-Squared Table

Degrees of Freedom	*Probability in Right Tail*					
	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	8.307	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-squared test statistic, χ^2, with n – 1 degrees of freedom, is computed as:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$$

where:
n = sample size
s^2 = sample variance
σ_0^2 = hypothesized value for the population variance

Similar to other hypothesis tests, the chi-squared test compares the test statistic, χ^2_{n-1}, to a critical chi-squared value at a given level of significance and n – 1 degrees of freedom.

Example: Chi-squared test for a single population variance

Historically, High-Return Equity Fund has advertised that its monthly returns have a standard deviation equal to 4%. This was based on estimates from the 1990–1998 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 1998 and 2000 and measured a standard deviation of monthly returns of 3.8%. **Determine** if the more recent standard deviation is different from the advertised standard deviation.

Answer:

State the hypothesis. The null hypothesis is that the standard deviation is equal to 4% and, therefore, the variance of monthly returns for the population is $(0.04)^2 = 0.0016$. Since 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 the form:

$$H_0: \sigma^2 = 0.0016 \text{ versus } H_A: \sigma^2 \neq 0.0016$$

Select the appropriate test statistic. The appropriate test statistic for tests of variance using the chi-squared distribution is computed as follows:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

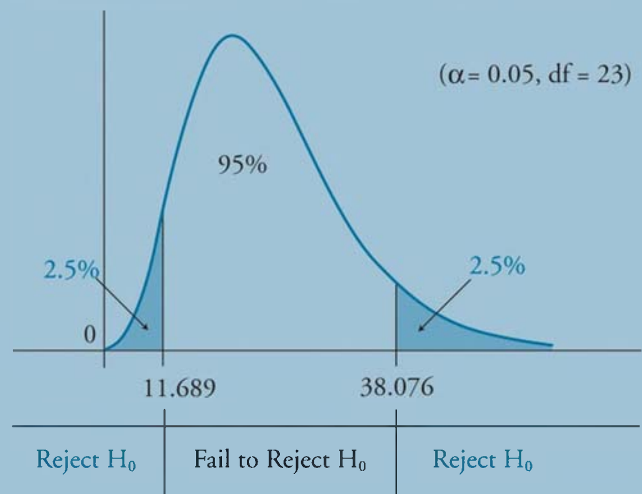
Specify the level of significance. Let's use a 5% level of significance, meaning there will be 2.5% probability in each tail of the chi-squared distribution.

State the decision rule regarding the hypothesis. With a 24-month sample, there are 23 degrees of freedom. Using the table of chi-squared values at the back 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, the decision rule is:

$$\text{Reject } H_0 \text{ if: } \chi^2 < 11.689, \text{ or } \chi^2 > 38.076$$

This decision rule is illustrated in the following distribution.

Decision Rule for a Two-Tailed Chi-Squared Test of a Single Population Variance



Collect the sample and calculate the sample statistics. Using the information provided, the test statistic is computed as:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(23)(0.001444)}{0.0016} = \frac{0.033212}{0.0016} = 20.7575$$

Make a decision regarding the hypothesis. Since the computed test statistic, χ^2, falls between the two critical values, we fail to reject the null hypothesis that the variance is equal to 0.0016.

Make a decision based on the results of the test. It can be concluded that the recently measured standard deviation is close enough to the advertised standard deviation that we cannot say it is different from 4%, at a 5% level of significance.

The *F*-Test

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:

$$H_0\text{: } \sigma_1^2 = \sigma_2^2 \text{ versus } H_A\text{: } \sigma_1^2 \neq \sigma_2^2$$

and the one-sided test structures can be specified as:

$$H_0: \sigma_1^2 \leq \sigma_2^2 \text{ versus } H_A: \sigma_1^2 > \sigma_2^2, \text{ or } H_0: \sigma_1^2 \geq \sigma_2^2 \text{ versus } H_A: \sigma_1^2 < \sigma_2^2$$

The test statistic for the *F*-test is the ratio of the sample variances. The *F*-statistic is computed as:

$$F = \frac{s_1^2}{s_2^2}$$

where:

s_1^2 = variance of the sample of n_1 observations drawn from Population 1

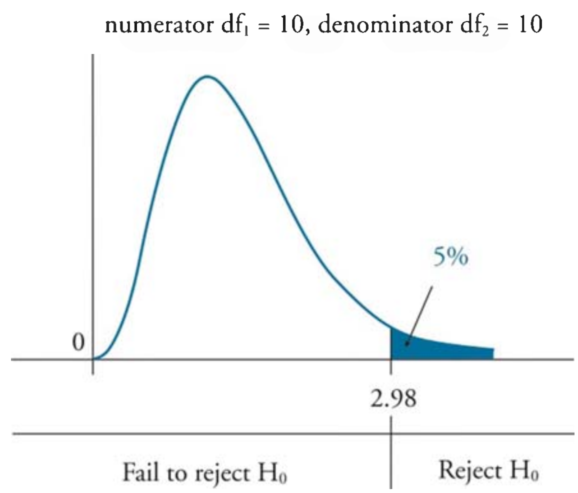
s_2^2 = variance of the sample of n_2 observations drawn from Population 2

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 the Appendix).

Professor's Note: Always put the larger variance in the numerator (s_1^2). Following this convention means we only have to consider the critical value for the right-hand tail.

An *F*-distribution is presented in Figure 10. As indicated, the *F*-distribution is right-skewed and is truncated at 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. Also shown in Figure 10 is that the *rejection region is in the right-side tail* of the distribution. This will always be the case as long as the *F*-statistic is computed with the largest sample variance in the numerator. The labeling of 1 and 2 is arbitrary anyway.

Figure 10: *F*-Distribution



Example: *F*-test for equal variances

Annie Cower is examining the earnings for two different industries. Cower suspects that the earnings of the textile industry are more divergent than those of 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. **Determine** if the earnings of the textile industry have greater standard deviation than those of the paper industry.

Answer:

State the hypothesis. In this example, we are concerned with whether the variance of the earnings of the textile industry is greater (more divergent) than the variance of the earnings of the paper industry. As such, the test hypotheses can be appropriately structured as:

$$H_0: \sigma_1^2 \leq \sigma_2^2 \text{ versus } H_A: \sigma_1^2 > \sigma_2^2$$

where:

σ_1^2 = variance of earnings for the textile industry

σ_2^2 = variance of earnings for the paper industry

Note: $\sigma_1^2 > \sigma_2^2$

Select the appropriate test statistic. For tests of difference between variances, the appropriate test statistic is:

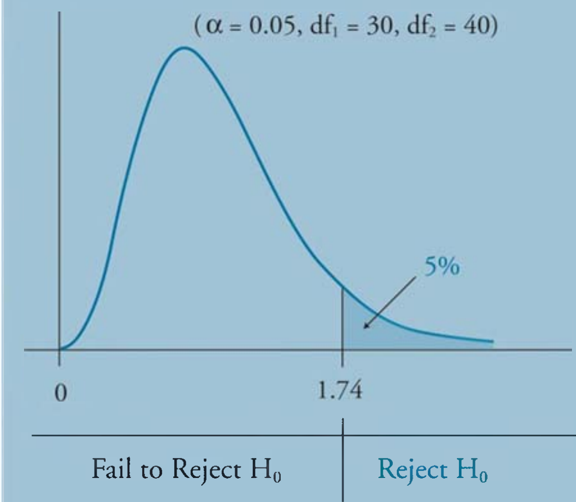
$$F = \frac{s_1^2}{s_2^2}$$

Specify the level of significance. Let's conduct our hypothesis test at the 5% level of significance.

State the decision rule regarding the hypothesis. Using the sample sizes for the two industries, the critical *F*-value for our test is found to be 1.74. This value is obtained from the table of the *F*-distribution at the 5% level of significance with $df_1 = 30$ and $df_2 = 40$. Thus, if the computed *F*-statistic is greater than the critical value of 1.74, the null hypothesis is rejected. The decision rule, illustrated in the distribution that follows, can be stated as:

Reject H_0 if: $F > 1.74$

Decision Rule for *F*-Test



Collect the sample and calculate the sample statistics. Using the information provided, the *F*-statistic can be computed as:

$$F = \frac{s_1^2}{s_2^2} = \frac{\$4.30^2}{\$3.80^2} = \frac{\$18.49}{\$14.44} = 1.2805$$

Professor's Note: Remember to square the standard deviations to get the variances.

Make a decision regarding the hypothesis. Since the calculated *F*-statistic of 1.2805 is less than the critical *F*-statistic of 1.74, we fail to reject the null hypothesis.

Make a decision based on the results of the test. Based on the results of the hypothesis test, Cower should conclude that the earnings variances of the industries are not statistically significantly different from one another at a 5% level of significance. More pointedly, the earnings of the textile industry are not more divergent than those of the paper industry.

Chebyshev's Inequality

Chebyshev's inequality states that for any set of observations, whether sample or population data and regardless of the shape of the distribution, the percentage of the observations that lie within k standard deviations of the mean is *at least* $1 - 1/k^2$ for all $k > 1$.

Example: Chebyshev's inequality

What is the minimum percentage of any distribution that will lie within ±2 standard deviations of the mean?

Answer:

Applying Chebyshev's inequality, we have:

$$1 - 1/k^2 = 1 - 1/2^2 = 1 - 1/4 = 0.75 \text{ or } 75\%$$

According to Chebyshev's inequality, the following relationships hold for any distribution. At least:

- 36% of observations lie within ±1.25 standard deviations of the mean.
- 56% of observations lie within ±1.50 standard deviations of the mean.
- 75% of observations lie within ±2 standard deviations of the mean.
- 89% of observations lie within ±3 standard deviations of the mean.
- 94% of observations lie within ±4 standard deviations of the mean.

The importance of Chebyshev's inequality is that it applies to any distribution. If we know the underlying distribution is actually normal, we can be even more precise about the percentage of observations that will fall within a given number of standard deviations of the mean.

Note that with a normal distribution, extreme events beyond ±3 standard deviations are very rare (occurring only 0.26% of the time). However, as Chebyshev's inequality points out, events that are ±3 standard deviations may not be so rare for nonnormal distributions (potentially occurring 11% of the time). Therefore, simply assuming normality, without knowing the parameters of the underlying distribution, could lead to a severe underestimation of risk.

Backtesting

LO 19.6: Demonstrate the process of backtesting VaR by calculating the number of exceedances.

The process of **backtesting** involves comparing expected outcomes against actual data. For example, if we apply a 95% confidence interval, we are expecting an event to exceed the confidence interval with a 5% probability. Recall that the 5% in this example is known as the level of significance.

It is common for risk managers to backtest their value at risk (VaR) models to ensure that the model is forecasting losses with the same frequency predicted by the confidence interval (VaR models typically use a 95% confidence interval). When the VaR measure is exceeded during a given testing period, it is known as an **exception** or an **exceedance**. After backtesting the VaR model, if the number of exceptions if greater than expected, the risk manager may be underestimating actual risk. Conversely, if the number of exceptions is less than expected, the risk manager may be overestimating actual risk.

Example: Calculating the number of exceedances

Assume that the value at risk (VaR) of a portfolio, at a 95% confidence interval, is \$100 million. Also assume that given a 100-day trading period, the actual number of daily losses exceeding \$100 million occurred eight times. Is this VaR model underestimating or overestimating the actual level of risk?

Answer:

With a 95% confidence interval, we expect to have exceptions (i.e., losses exceeding \$100 million) 5% of the time. If the losses exceeding \$100 million occurred eight times during the 100-day period, exceptions occurred 8% of the time. Therefore, this VaR model is underestimating risk because the number of exceptions is greater than expected according to the 95% confidence interval.

One of the main issues with backtesting VaR models is that exceptions are often serially correlated. In other words, there is a high probability that an exception will occur after the previous period had an exception. Another issue is that the occurrence of exceptions tends to be correlated with overall market volatility. In other words, VaR exceptions tend to be higher (lower) when market volatility is high (low). This may be the result of a VaR model failing to quickly react to changes in risk levels.

Professor's Note: We will discuss VaR methodologies and backtesting VaR in more detail in Book 4.

Key Concepts

LO 19.1

Population variance $= \sigma^2 = \dfrac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$, where μ = population mean and N = size

Sample variance $= s^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$, where $\overline{X}$ = sample mean and n = sample size

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

LO 19.2

For a normally distributed population, a confidence interval for its mean can be constructed using a *z*-statistic when variance is known, and a *t*-statistic when the variance is unknown. The *z*-statistic is acceptable in the case of a normal population with an unknown variance if the sample size is large (30+).

In general, we have:

- $\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$ when the variance is known
- $\bar{x} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$ when the variance is unknown

LO 19.3

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 test statistic is the value that a decision about a hypothesis will be based on. For a test about the value of the mean of a distribution:

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

With unknown population variance, the *t*-statistic is used for tests about the mean of a normally distributed population: $t_{n-1} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$. If the population variance is known, the appropriate test statistic is $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ for tests about the mean of a population.

LO 19.4

A two-tailed test results from a two-sided alternative hypothesis (e.g., H_A: $\mu \neq \mu_0$). A one-tailed test results from a one-sided alternative hypothesis (e.g., H_A: $\mu > \mu_0$, or H_A: $\mu < \mu_0$).

LO 19.5

Hypothesis testing compares a computed test statistic to a critical value at a stated level of significance, which is the decision rule for the test.

A hypothesis about a population parameter is rejected when the sample statistic lies outside a confidence interval around the hypothesized value for the chosen level of significance.

LO 19.6

Backtesting is the process of comparing losses predicted by the value at risk (VaR) model to those actually experienced over the sample testing period. If a model were completely accurate, we would expect VaR to be exceeded with the same frequency predicted by the confidence level used in the VaR model. In other words, the probability of observing a loss amount greater than VaR should be equal to the level of significance.

Concept Checkers

1. An analyst observes that the variance of daily stock returns for Stock X during a certain period is 0.003. He assumes daily stock returns are normally distributed and wants to conduct a hypothesis test to determine whether the variance of daily returns on Stock X is different from 0.005. The analyst looks up the critical values for his test, which are 9.59 and 34.17. He calculates a test statistic of 11.40 for his set of data. What kind of test statistic did the analyst calculate, and should he conclude that the variance is different from 0.005?

	Test statistic	Variance $\neq$ 0.005
A.	t-statistic	Yes
B.	Chi-squared statistic	Yes
C.	t-statistic	No
D.	Chi-squared statistic	No

Use the following data to answer Questions 2 and 3.

Austin Roberts believes the mean price of houses in the area is greater than \$145,000. A random sample of 36 houses in the area has a mean price of \$149,750. The population standard deviation is \$24,000, and Roberts wants to conduct a hypothesis test at a 1% level of significance.

2. The appropriate alternative hypothesis is:
   A. H_A: μ < \$145,000.
   B. H_A: μ ± \$145,000.
   C. H_A: $\mu \geq$ \$145,000.
   D. H_A: μ > \$145,000.

3. The value of the calculated test statistic is closest to:
   A. $z = 0.67$.
   B. $z = 1.19$.
   C. $z = 4.00$.
   D. $z = 8.13$.

4. The 95% confidence interval of the sample mean of employee age for a major corporation is 19 years to 44 years based on a z-statistic. The population of employees is more than 5,000 and the sample size of this test is 100. Assuming the population is normally distributed, the standard error of mean employee age is closest to:
   A. 1.96.
   B. 2.58.
   C. 6.38.
   D. 12.50.

Use the following data to answer Question 5.

XYZ Corp. Annual Stock Prices					
1995	1996	1997	1998	1999	2000
22%	5%	–7%	11%	2%	11%

5. Assuming the distribution of XYZ stock returns is a sample, what is the sample standard deviation?
   A. 7.4%.
   B. 9.8%.
   C. 72.4%.
   D. 96.3%.

Concept Checker Answers

1. **D** Hypothesis tests concerning the variance of a normally distributed population use the chi-squared statistic. The null hypothesis is that the variance is equal to 0.005. Since the test statistic falls within the range of the critical values, the test fails to reject the null hypothesis. The analyst cannot conclude that the variance of daily returns on Stock X is different from 0.005.

2. **D** H_A: $\mu > \$145{,}000$.

3. **B** $$z = \frac{149{,}750 - 145{,}000}{24{,}000 / \sqrt{36}} = 1.1875.$$

4. **C** At the 95% confidence level, with sample size n = 100 and mean 31.5 years, the appropriate test statistic is $z_{\alpha/2} = 1.96$. *Note: The mean of 31.5 is calculated as the midpoint of the interval, or (19 + 44) / 2.* Thus, the confidence interval is $31.5 \pm 1.96s_x$, where s_x is the standard error of the sample mean. If we take the upper bound, we know that $31.5 + 1.96s_x = 44$, or $1.96s_x = 12.5$, or $s_x = 6.38$ years.

5. **B** The sample standard deviation is the square root of the sample variance:

$$s = \left\{ \frac{\left[(22-7.3)^2 + (5-7.3)^2 + (-7-7.3)^2 + (11-7.3)^2 + (2-7.3)^2 + (11-7.3)^2\right]}{6-1} \right\}^{1/2}$$

$$= \left(96.3\%^2\right)^{1/2} = 9.8\%$$

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Linear Regression with One Regressor

Topic 20

Exam Focus

Linear regression refers to the process of representing relationships with linear equations where there is one dependent variable being explained by one or more independent variables. There will be deviations from the expected value of the dependent variable called error terms, which represent the effect of independent variables not included in the population regression function. Typically we do not know the population regression function; instead, we estimate it with a method such as ordinary least squares (OLS). For the exam, be able to apply the concepts of simple linear regression and understand how sample data can be used to estimate population regression parameters (i.e., the intercept and slope of the linear regression).

Regression Analysis

LO 20.1: Explain how regression analysis in econometrics measures the relationship between dependent and independent variables.

A regression analysis has the goal of measuring how changes in one variable, called a **dependent** or **explained** variable can be explained by changes in one or more other variables called the **independent** or **explanatory** variables. The regression analysis measures the relationship by estimating an equation (e.g., linear regression model). The **parameters** of the equation indicate the relationship.

A **scatter plot** is a visual representation of the relationship between the dependent variable and a given independent variable. It uses a standard two-dimensional graph where the values of the dependent, or Y variable, are on the vertical axis, and those of the independent, or X variable, are on the horizontal axis.

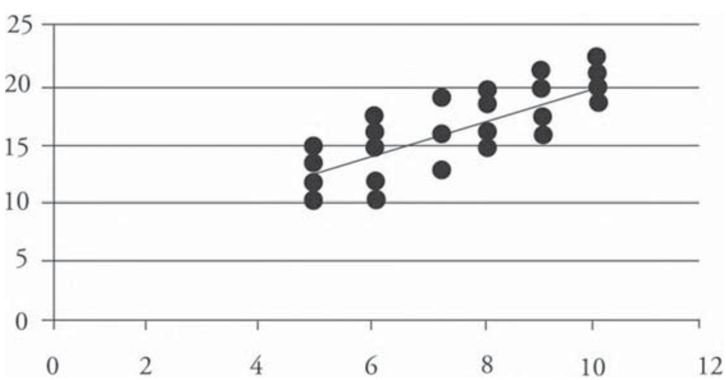
A scatter plot can indicate the nature of the relationship between the dependent and independent variable. The most basic property indicated by a scatter plot is whether there is a positive or negative relationship between the dependent variable and the independent variable. A closer inspection can indicate if the relationship is linear or nonlinear.

As an example, let us assume that we have access to all the returns data for a certain class of hedge funds over a given year. The population consists of 30 hedge funds that follow the same strategy, but they differ by the length of the lockup period. The lockup period is the minimum number of years an investor must keep funds invested. For this given strategy of hedge funds, the lockup periods range from five to ten years. Figure 1 contains the hedge fund data, and Figure 2 is a scatter plot that illustrates the relationship.

Figure 1: Hedge Fund Data

Lockup (yrs)	*Returns (%) per year*					*Average Return*
5	10	14	14	15	12	13
6	17	12	15	16	10	14
7	16	19	19	13	13	16
8	15	20	19	15	16	17
9	21	20	16	20	18	19
10	20	17	21	23	19	20

Figure 2: Return Over Lockup Period



The scatter plot indicates that there is a positive relationship between the hedge fund returns and the lockup period. We should keep in mind that the data represents returns over the same period (i.e., one year). The factor that varies is the amount of time a manager knows that he will control the funds. One interpretation of the graph could be that managers who know that they can control the funds over a longer period can engage in strategies that reap a higher return in any given year. As a final note, the scatter plot in this example indicates a fairly linear relationship. With each 1-year increase in the lockup period, according to the graph, the corresponding returns seem to increase by a similar amount.

Population Regression Function

LO 20.2: Interpret a population regression function, regression coefficients, parameters, slope, intercept, and the error term.

Assuming that the 30 observations represent the population of hedge funds that are in the same class (i.e., have the same basic investment strategy) then their relationship can provide a **population regression function**. Such a function would consist of parameters called **regression coefficients**. The regression equation (or function) will include an intercept term and one slope coefficient for each independent variable. For this simple two-variable case, the function is:

$$E(\text{return} \mid \text{lockup period}) = B_0 + B_1 \times (\text{lockup period})$$

Or more generally:

$$E(Y_i \mid X_i) = B_0 + B_1 \times (X_i)$$

In the equation, B_0 is the **intercept coefficient**, which is the expected value of the return if $X = 0$. B_1 is the **slope coefficient**, which is the expected change in *Y* for a unit change in *X*. In this example, for every additional year of lockup, a hedge fund is expected to earn an additional B_1 per year in return.

The Error Term

There is a dispersion of *Y*-values around each conditional expected value. The difference between each *Y* and its corresponding conditional expectation (i.e., the line that fits the data) is the **error term** or **noise component** denoted ε_i.

$$\varepsilon_i = Y_i - E(Y_i \mid X_i)$$

The deviation from the expected value is the result of factors other than the included *X*-variable. One way to break down the equation is to say that $E(Y_i \mid X_i) = B_0 + B_1 \times X_i$ is the deterministic or systematic component, and ε_i is the nonsystematic or random component. The error term provides another way of expressing the population regression function:

$$Y_i = B_0 + B_1 \times X_i + \varepsilon_i$$

The error term represents effects from independent variables not included in the model. In the case of the hedge fund example, ε_i is probably a function of the individual manager's unique trading tactics and management activities within the style classification. Variables that might explain this error term are the number of positions and trades a manager makes over time. Another variable might be the years of experience of the manager. An analyst may need to include several of these variables (e.g., trading style and experience) into the population regression function to reduce the error term by a noticeable amount. Often, it is found that limiting an equation to the one or two independent variables with the most explanatory power is the best choice.

Sample Regression Function

LO 20.3: Interpret a sample regression function, regression coefficients, parameters, slope, intercept, and the error term.

The **sample regression function** is an equation that represents a relationship between the *Y* and *X* variable(s) that is based only on the information in a sample of the population. In almost all cases the slope and intercept coefficients of a sample regression function will be different from that of the population regression function. If the sample of *X* and *Y* variables is truly a random sample, then the difference between the sample coefficients and the population coefficients will be random too. There are various ways to use notation to distinguish the components of the sample regression function from the population regression function. Here we have denoted the population parameters with capital letters (i.e., B_0 and B_1) and the sample coefficients with small letters as indicated in the following sample regression function:

$$Y_i = b_0 + b_1 \times X_i + e_i$$

The sample regression coefficients are b_0 and b_1, which are the intercept and slope. There is also an extra term on the end called the **residual**: $e_i = Y_i - (b_0 + b_1 \times X_i)$. Since the population and sample coefficients are almost always different, the residual will very rarely equal the corresponding population error term (i.e., generally $e_i \neq \varepsilon_i$).

Properties of Regression

LO 20.4: Describe the key properties of a linear regression.

Under certain, basic assumptions, we can use a linear regression to estimate the population regression function. The term "linear" has implications for both the independent variable and the coefficients. One interpretation of the term *linear* relates to the independent variable(s) and specifies that the independent variable(s) enters into the equation without a transformation such as a square root or logarithm. If it is the case that the relationship between the dependent variable and an independent variable is non-linear, then an analyst would do that transformation first and then enter the transformed value into the linear equation as *X*. For example, in estimating a utility function as a function of consumption, we might allow for the property of diminishing marginal utility by transforming consumption into a logarithm of consumption. In other words, the actual relationship is:

$$E(\text{utility} \mid \text{amount consumed}) = B_0 + B_1 \times \ln(\text{amount consumed})$$

Here we let Y = utility and X = ln (amount consumed) and estimate: $E(Y_i \mid X_i) = B_0 + B_1 \times (X_i)$ using linear techniques.

A second interpretation for the term linear applies to the parameters. It specifies that the dependent variable is a linear function of the parameters, but does not require that there is linearity in the variables. Two examples of non-linear relationships are as follows:

$$E(Y_i \mid X_i) = B_0 + (B_1)^2 \times (X_i)$$

$$E(Y_i \mid X_i) = B_0 + (1/B_1) \times (X_i)$$

It would not be appropriate to apply linear regression to estimate the parameters of these functions. The primary concern for linear models is that they display linearity in the parameters. Therefore, when we refer to a linear regression model we generally assume that the equation is linear in the parameters; it may or may not be linear in the variables.

Ordinary Least Squares Regression

LO 20.5: Define an ordinary least squares (OLS) regression and calculate the intercept and slope of the regression.

Ordinary least squares (OLS) estimation is a process that estimates the population parameters B_i with corresponding values for b_i that minimize the squared residuals (i.e., error terms). Recall the expression $e_i = Y_i - (b_0 + b_1 \times X_i)$; the OLS sample coefficients are those that:

$$\text{minimize } \sum e_i^2 = \Sigma[Y_i - (b_0 + b_1 \times X_i)]^2$$

The estimated **slope coefficient** (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:

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

The **intercept** term (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:

$$b_0 = \overline{Y} - b_1\overline{X}$$

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, $\overline{X}, \overline{Y}$).

Assumptions Underlying Linear Regression

LO 20.6: Describe the method and three key assumptions of OLS for estimation of parameters.

OLS regression requires a number of assumptions. Most of the major assumptions pertain to the regression model's residual term (i.e., error term). Three key assumptions are as follows:

- The expected value of the error term, conditional on the independent variable, is zero ($E(\varepsilon_i|X_i) = 0$).
- All (X, Y) observations are independent and identically distributed (i.i.d.).
- It is unlikely that large outliers will be observed in the data. Large outliers have the potential to create misleading regression results.

Additional assumptions include:

- A linear relationship exists between the dependent and independent variable.
- The model is correctly specified in that it includes the appropriate independent variable and does not omit variables.
- The independent variable is uncorrelated with the error terms.
- The variance of ε_i is constant for all X_i: $\text{Var}(\varepsilon_i|X_i) = \sigma^2$.
- No serial correlation of the error terms exists [i.e., $\text{Corr}(\varepsilon_i, \varepsilon_{i+j}) = 0$ for j=1, 2, 3…]. The point being that knowing the value of an error for one observation does not reveal information concerning the value of an error for another observation.
- The error term is normally distributed.

Properties of OLS Estimators

LO 20.7: Summarize the benefits of using OLS estimators.

OLS estimators and terminology are used widely in practice when applying regression analysis techniques. In fields such as economics, finance, and statistics, the presentation of OLS regression results is the same. This means that the calculation of b_0 and b_1 and the interpretation and analysis of regression output is easily understood across multiple fields of study. As a result, statistical software packages make it easy for users to apply OLS estimators. In addition to practical benefits, OLS estimators also have theoretical benefits. OLS estimated coefficients are unbiased, consistent, and (under special conditions) efficient. Recall from Topic 16, that these characteristics are desirable properties of an estimator.

LO 20.8: Describe the properties of OLS estimators and their sampling distributions, and explain the properties of consistent estimators in general.

Since OLS estimators are derived from random samples, these estimators are also random variables because they vary from one sample to the next. Therefore, OLS estimators will have their own probability distributions (i.e., sampling distributions). These sampling distributions allow us to estimate population parameters, such as the population mean, the population regression intercept term, and the population regression slope coefficient.

Drawing multiple samples from a population will produce multiple sample means. The distribution of these sample means is referred to as the *sampling distribution of the sample mean*. The mean of this sampling distribution is used as an estimator of the population mean and is said to be an **unbiased estimator** of the population mean. Recall that an unbiased estimator is one for which the expected value of the estimator is equal to the parameter you are trying to estimate.

Given the **central limit theorem**, for large sample sizes, it is reasonable to assume that the sampling distribution will approach the normal distribution. This means that the estimator is also a **consistent estimator**. Recall that a consistent estimator is one for which the accuracy of the parameter estimate increases as the sample size increases. Note that a general guideline for a large sample size in regression analysis is a sample greater than 100.

Like the sampling distribution of the sample mean, OLS estimators for the population intercept term and slope coefficient also have sampling distributions. The sampling distributions of OLS estimators, b_0 and b_1, are unbiased and consistent estimators of population parameters, B_0 and B_1. Being able to assume that b_0 and b_1 are normally distributed is a key property in allowing us to make statistical inferences about population coefficients.

OLS Regression Results

LO 20.9: Interpret the explained sum of squares, the total sum of squares, the residual sum of squares, the standard error of the regression, and the regression R^2.

LO 20.10: Interpret the results of an OLS regression.

The **sum of squared residuals** (SSR), sometimes denoted SSE, for sum of squared errors, is the sum of squares that results from placing a given intercept and slope coefficient into the equation and computing the residuals, squaring the residuals and summing them. It is represented by $\sum e_i^2$. The sum is an indicator of how well the sample regression function explains the data.

Assuming certain conditions exist, an analyst can use the results of an ordinary least squares regression in place of the unknown population regression function to describe the relationship between the dependent and independent variable(s). In our earlier example concerning hedge fund returns and lockup periods, we might assume that an analyst only has access to a sample of returns data (e.g., six observations). This may be the result of the fact that hedge funds are not regulated and the reporting of returns is voluntary. In any case, we will assume that the data in Figure 3 is the sample of six observations and includes the corresponding computations for computing OLS estimates.

Figure 3: Sample of Returns and Corresponding Lockup Periods

	Lockup	*Returns*	$(X-\overline{X})$	$(Y-\overline{Y})$	*Cov(X,Y)*	*Var(X)*
	5	10	–2.5	–6	15	6.25
	6	12	–1.5	–4	6	2.25
	7	19	–0.5	3	–1.5	0.25
	8	16	0.5	0	0	0.25
	9	18	1.5	2	3	2.25
	10	21	2.5	5	12.5	6.25
Sum	45	96	0	0	35	17.50
Average	7.5	16				

From Figure 3, we can compute the sample coefficients:

$$b_1 = \frac{35}{17.5} = 2$$

$$b_0 = 16 - 2 \times 7.5 = 1$$

Thus, the sample regression function is: $Y_i = 1 + 2 \times X_i + e_i$. This means that, according to the data, on average a hedge fund with a lockup period of six years will have a 2% higher return than a hedge fund with a 5-year lockup period.

The Coefficient of Determination

The **coefficient of determination**, represented by R^2, is a measure of the "goodness of fit" of the regression. It is interpreted as a percentage of variation in the dependent variable explained by the independent variable. The underlying concept is that for the dependent variable, there is a total sum of squares (TSS) around the sample mean. The regression equation explains some portion of that TSS. Since the explained portion is determined by the independent variables, which are assumed independent of the errors, the total sum of squares can be broken down as follows:

$$\text{Total sum of squares} = \text{explained sum of squares} + \text{sum of squared residuals}$$

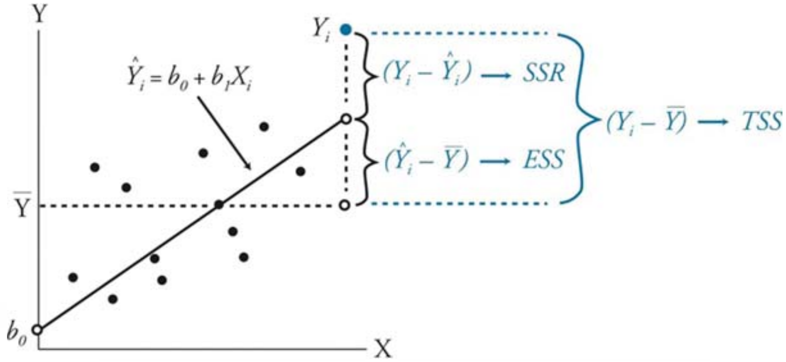
$$\sum(Y_i - \overline{Y})^2 = \sum(\hat{Y} - \overline{Y})^2 + \sum(Y_i - \hat{Y})^2$$

$$\text{TSS} = \text{ESS} + \text{SSR}$$

Professor's Note: As mentioned previously, sum of squared residuals (SSR) is also known as the sum of squared errors (SSE). In the same regard, total sum of squares (TSS) is also known as sum of squares total (SST), and explained sum of squares (ESS) is also known as regression sum of squares (RSS).

Figure 4 illustrates how the total variation in the dependent variable (TSS) is composed of SSR and ESS.

Figure 4: Components of the Total Variation



The coefficient of determination can be calculated as follows:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2}$$

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \overline{Y})^2}$$

Example: Computing R^2

Figure 5 contains the relevant information from our hedge fund example where the average of the hedge fund returns was 16% (i.e., $\overline{Y} = 16$). **Compute** the coefficient of determination for the hedge fund regression line.

Figure 5: Computing the Coefficient of Determination

	Lockup	Returns, Y_i	e_i	e_i^2	$\sum(Y_i - \overline{Y})^2$	$\hat{Y}_i$	$\sum(Y_i - \hat{Y}_i)^2$
	5	10	–1	1	36	11	1
	6	12	–1	1	16	13	1
	7	19	4	16	9	15	16
	8	16	–1	1	0	17	1
	9	18	–1	1	4	19	1
	10	21	0	0	25	21	0
Sum	45	96	0	20	90	96	20

Answer:

The coefficient of determination is 77.8%, which is calculated as follows:

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \overline{Y})^2} = 1 - \frac{20}{90} = 0.778$$

In a simple two-variable regression, the square root of R^2 is the **correlation coefficient** (r) between X_i and Y_i. If the relationship is positive, then:

$$r = \sqrt{R^2}$$

For the hedge fund data, the correlation coefficient is: $r = \sqrt{0.778} = 0.882$

The correlation coefficient is a standard measure of the strength of the linear relationship between two variables. Initially it may seem similar to the coefficient of determination, but it is not for two reasons. First, the correlation coefficient indicates the sign of the relationship, whereas the coefficient of determination does not. Second, the coefficient of determination can apply to an equation with several independent variables, and it implies a causation or explanatory power, while the correlation coefficient only applies to two variables and does not imply causation between the variables.

The Standard Error of the Regression

The standard error of the regression (SER) measures the degree of variability of the actual Y-values relative to the estimated Y-values from a regression equation. The SER gauges the "fit" of the regression line. The smaller the standard error, the better the fit.

The SER is the standard deviation of the error terms in the regression. As such, SER is also referred to as the standard error of the residual, or the standard error of estimate (SEE).

In some regressions, the relationship between the independent and dependent variables is very strong (e.g., the relationship between 10-year Treasury bond yields and mortgage rates). In other cases, the relationship is much weaker (e.g., the relationship between stock returns and inflation). SER will be low (relative to total variability) if the relationship is very strong and high if the relationship is weak.

Key Concepts

LO 20.1

Regression analysis attempts to measure the relationship between a dependent variable and one or more independent variables.

A scatter plot (a.k.a. scattergram) is a collection of points on a graph where each point represents the values of two variables (i.e., an X/Y pair).

LO 20.2

A population regression line indicates the expected value of a dependent variable conditional on one or more independent variables: $E(Y_i \mid X_i) = B_0 + B_1 \times (X_i)$.

The difference between an actual dependent variable and a given expected value is the error term or noise component denoted $\varepsilon_i = Y_i - E(Y_i \mid X_i)$.

LO 20.3

The sample regression function is an equation that represents a relationship between the *Y* and *X* variable(s) using only a sample of the total data. It uses symbols that are similar but still distinct from that of the population $Y_i = b_0 + b_1 \times X_i + e_i$.

LO 20.4

In a linear regression model, we generally assume that the equation is linear in the parameters, and that it may or may not be linear in the variables.

LO 20.5

Ordinary least squares estimation is a process that estimates the population parameters B_i with corresponding values for b_i that minimize $\sum e_i^2 = \sum[Y_i - (b_0 + b_1 \times X_i)]^2$. The formulas for the coefficients are:

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

$$b_0 = \overline{Y} - b_1\overline{X}$$

LO 20.6

Three key assumptions made with simple linear regression include:

- The expected value of the error term, conditional on the independent variable, is zero.
- All (X, Y) observations are independent and identically distributed (i.i.d.).
- It is unlikely that large outliers will be observed in the data.

LO 20.7

OLS estimators are used widely in practice. In addition to practical benefits, OLS estimators exhibit desirable properties of an estimator.

LO 20.8

Since OLS estimators are random variables, they have their own sampling distributions. These sampling distributions are used to estimate population parameters. Given that the expected value of the estimator is equal to the parameter being estimated and the accuracy of the parameter estimate increases as the sample size increases, we can say that OLS estimators are both unbiased and consistent.

LO 20.9

Explained sum of squares (ESS) measures the variation in the dependent variable that is explained by the independent variable.

Total sum of squares (TSS) measures the total variation in the dependent variable. TSS is equal to the sum of the squared differences between the actual Y-values and the mean of Y.

Sum of squared residuals (SSR) measures the unexplained variation in the dependent variable.

The standard error of the regression (SER) measures the degree of variability of the actual Y-values relative to the estimated Y-values from a regression equation.

The coefficient of determination, represented by R^2, is a measure of the "goodness of fit" of the regression.

LO 20.10

Assuming certain conditions exist, an analyst can use the results of an ordinary least squares regression in place of an unknown population regression function to describe the relationship between the dependent and independent variable.

Concept Checkers

1. If the value of the independent variable is zero, then the expected value of the dependent variable would be equal to the:
   A. slope coefficient.
   B. intercept coefficient.
   C. error term.
   D. residual.

2. The error term represents the portion of the:
   A. dependent variable that is not explained by the independent variable(s) but could possibly be explained by adding additional independent variables.
   B. dependent variable that is explained by the independent variable(s).
   C. independent variables that are explained by the dependent variable.
   D. dependent variable that is explained by the error in the independent variable(s).

3. 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 zero is 10.0.
   B. The predicted value of the independent variable when the dependent variable is zero is 0.1.
   C. For every one unit change in the independent variable the model predicts that the dependent variable will change by 10 units.
   D. For every one unit change in the independent variable the model predicts that the dependent variable will change by 0.1 units.

4. A linear regression function assumes that the equation must be linear in:
   A. both the variables and the coefficients.
   B. the coefficients but not necessarily the variables.
   C. the variables but not necessarily the coefficients.
   D. neither the variables nor the coefficients.

5. Ordinary least squares refers to the process that:
   A. maximizes the number of independent variables.
   B. minimizes the number of independent variables.
   C. produces sample regression coefficients.
   D. minimizes the sum of the squared error terms.

Concept Checker Answers

1. **B** The equation is $E(Y \mid X) = b_0 + b_1 \times X$. If $X = 0$, then $Y = b_0$ (i.e., the intercept coefficient).

2. **A** The error term represents effects from independent variables not included in the model. It could be explained by additional independent variables.

3. **C** 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 one unit, the dependent variable will 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.

4. **B** Linear regression refers to a regression that is linear in the coefficients/parameters; it may or may not be linear in the variables.

5. **D** OLS is a process that minimizes the sum of squared residuals to produce estimates of the population parameters known as sample regression coefficients.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

Topic 21

Exam Focus

As shown in the previous topic, the classical linear regression model requires several assumptions. One of those assumptions is homoskedasticity, which means a constant variance of the errors over the sample. If the assumptions are true, the estimated coefficients have the desirable properties of being unbiased and having a minimum variance when compared to other estimators. It is usually assumed that the errors are normally distributed, which allows for standard methods of hypothesis testing of the estimated coefficients. For the exam, be able to construct confidence intervals and perform hypothesis tests on regression coefficients, and understand how to detect heteroskedasticity.

Regression Coefficient Confidence Intervals

LO 21.1: Calculate and interpret confidence intervals for regression coefficients.

Hypothesis testing for a regression coefficient may use the confidence interval for the coefficient being tested. For instance, a frequently asked question is whether an estimated slope coefficient is statistically different from zero. In other words, the null hypothesis is H_0: $B_1 = 0$ and the alternative hypothesis is H_A: $B_1 \neq 0$. If the confidence interval at the desired level of significance does not include zero, the null is rejected, and the coefficient is said to be statistically different from zero.

The confidence interval for the regression coefficient, B_1, is calculated as:

$$b_1 \pm \left(t_c \times s_{b_1}\right), \text{or} \left[b_1 - \left(t_c \times s_{b_1}\right) < B_1 < b_1 + \left(t_c \times s_{b_1}\right)\right]$$

In this expression, t_c is the critical two-tailed *t*-value for the selected confidence level with the appropriate number of degrees of freedom, which is equal to the number of sample observations minus 2 (i.e., n – 2).

The standard error of the regression coefficient is denoted as s_{b_1}. It is a function of the SER: as SER rises, s_{b_1} also increases, and the confidence interval widens. This makes sense because SER measures the variability of the data about the regression line, and the more variable the data, the less confidence there is in the regression model to estimate a coefficient.

Professor's Note: It is highly unlikely you will have to calculate s_{b_1} on the exam. It is included in the output of all statistical software packages and should be given to you if you need it.

Example: Calculating the confidence interval for a regression coefficient

The estimated slope coefficient, B_1, from a regression run on WPO stock is 0.64 with a standard error equal to 0.26. Assuming that the sample had 36 observations, **calculate** the 95% confidence interval for B_1.

Answer:

The confidence interval for b_1 is:

$$b_1 \pm \left(t_c \times s_{b_1}\right), \text{or} \left[b_1 - \left(t_c \times s_{b_1}\right) < B_1 < b_1 + \left(t_c \times s_{b_1}\right)\right]$$

The critical two-tail *t*-values are $\pm$ 2.03 (from the *t*-table with n – 2 = 34 degrees of freedom). We can compute the 95% confidence interval as:

$$0.64 \pm (2.03)(0.26) = 0.64 \pm 0.53 = 0.11 \text{ to } 1.17$$

Because this confidence interval does not include zero, we can conclude that the slope coefficient is significantly different from zero.

Regression Coefficient Hypothesis Testing

LO 21.3: Interpret hypothesis tests about regression coefficients.

A *t*-test may also be used to test the hypothesis that the true slope coefficient, B_1, is equal to some hypothesized value. Letting b_1 be the point estimate for B_1, the appropriate test statistic with n – 2 degrees of freedom is:

$$t = \frac{b_1 - B_1}{s_{b_1}}$$

The decision rule for tests of significance for regression coefficients is:

Reject H_0 if $t > +t_{critical}$ or $t < -t_{critical}$

Rejection of the null means 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 hypothesis that is tested is whether the true slope is zero ($B_1 = 0$). The appropriate test structure for the null and alternative hypotheses is:

H_0: $B_1 = 0$ versus H_A: $B_1 \neq 0$

Example: Hypothesis test for significance of regression coefficients

Again, suppose that the estimated slope coefficient for the WPO regression 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 $t = \frac{b_1 - B_1}{s_{b_1}} = \frac{0.64 - 0}{0.26} = 2.46$.

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 and the confidence interval lead to the same conclusion to reject the null hypothesis and conclude that the slope coefficient is statistically significant.

LO 21.2: Interpret the p-value.

Comparing a test statistic to critical values is the preferred method for testing statistical significance. Another method involves the computation and interpretation of a ***p*-value**. Recall from Topic 19, the *p*-value is the smallest level of significance for which the null hypothesis can be rejected.

For two-tailed tests, the *p*-value is the probability that lies above the positive value of the computed test statistic *plus* the probability that lies below the negative value of the computed test statistic. For example, by consulting the *z*-table, the probability that lies above a test statistic of 2.46 is: (1 – 0.9931) = 0.0069 = 0.69%. With a two-tailed test, this *p*-value is: 2 × 0.69% = 1.38%. Therefore, the null hypothesis can be rejected at any level of significance greater than 1.38%. However, with a level of significance of, say, 1%, we would fail to reject the null.

A very small *p*-value provides support for rejecting the null hypothesis. This would indicate a large test statistic that is likely greater than critical values for a common level of significance (e.g., 5%). Many statistical software packages for regression analysis report *p*-values for regression coefficients. This output gives researchers a general idea of statistical significance without selecting a significance level.

Predicted Values

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, the predicted (or forecast) value of *Y* is:

$$\hat{Y} = b_0 + b_1 X_p$$

where:
$\hat{Y}$ = predicted value of the dependent variable
X_p = forecasted value of the independent variable

Example: Predicting the dependent variable

Given the regression equation:

$$\widehat{WPO} = -2.3\% + (0.64)\ (\widehat{S\&P\ 500})$$

Calculate the predicted value of WPO excess returns if forecasted S&P 500 excess returns are 10%.

Answer:

The predicted value for WPO excess returns is determined as follows:

$$\widehat{WPO} = -2.3\% + (0.64)(10\%) = 4.1\%$$

Confidence Intervals for Predicted Values

Confidence intervals for the predicted value of a dependent variable are calculated in a manner similar to the confidence interval for the regression coefficients. The equation for the confidence interval for a predicted value of *Y* is:

$$\hat{Y} \pm (t_c \times s_f) \Rightarrow \left[\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)\right]$$

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. 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:

$$s_f^2 = SER^2\left[1+\frac{1}{n}+\frac{(X-\bar{X})^2}{(n-1)s_x^2}\right]$$

where:
SER^2 = variance of the residuals = the square of the standard error of the regression
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 WPO from the previous example. Assume the standard error of the forecast is 3.67, and the forecasted value of S&P 500 excess returns is 10%.

Answer:

The predicted value for WPO is:

$$\widehat{WPO} = -2.3\% + (0.64)(10\%) = 4.1\%$$

The 5% two-tailed critical *t*-value with 34 degrees of freedom is 2.03. The prediction interval at the 95% confidence level is:

$$\widehat{WPO} \pm (t_c \times s_f) \Rightarrow [4.1\% \pm (2.03 \times 3.67\%)] = 4.1\% \pm 7.5\%$$

or

–3.4% to 11.6%

This range can be interpreted as, given a forecasted value for S&P 500 excess returns of 10%, we can be 95% confident that the WPO excess returns will be between –3.4% and 11.6%.