2016 SchweserNotes™ Part I

FRM® Exam Prep

Quantitative Analysis

eBook 2

KAPLAN UNIVERSITY SCHOOL OF PROFESSIONAL AND CONTINUING EDUCATION | SCHWESER

FRM官方交流群 392417118

FRM官方交流群 392417118

Getting Started

Part I FRM® Exam

Welcome

As the VP of Advanced Designations at Kaplan Schweser, I am pleased to have the opportunity to help you prepare for the 2016 FRM® Exam. Getting an early start on your study program is important for you to sufficiently **Prepare ▸ Practice ▸ 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,

Timothy Smaby

Timothy Smaby, PhD, CFA, FRM
Vice President, Advanced Designations, Kaplan Schweser

The Kaplan Way for Learning

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.**

May Exam Instructor
Dr. John Broussard
CFA, FRM

November Exam Instructor
Dr. Greg Filbeck
CFA, FRM, CAIA

*Dates, times, and instructors subject to change

Contact us for questions about your study package, upgrading your package, purchasing additional study materials, or for additional information:

www.schweser.com/frm | Toll-Free: 888.325.5072 | International: +1 608.779.8397

FRM官方交流群 392417118

FRM官方交流群 392417118

FRM Part I Book 2: Quantitative Analysis

Reading Assignments and Learning Objectives v

Quantitative Analysis

The Time Value of Money 1

15: Probabilities 13

16: Basic Statistics 29

17: Distributions 53

18: Bayesian Analysis 75

19: Hypothesis Testing and Confidence Intervals 88

20: Correlations and Copulas 128

21: Linear Regression with One Regressor 147

22: Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals 161

23: Linear Regression with Multiple Regressors 176

24: Hypothesis Tests and Confidence Intervals in Multiple Regression 190

25: Modeling and Forecasting Trend 209

26: Characterizing Cycles 218

27: Modeling Cycles: MA, AR, and ARMA Models 227

28: Estimating Volatilities and Correlations for Risk Management 237

29: Simulation Methods 247

Self-Test: Quantitative Analysis 260

Formulas 267

Appendix 274

Index 281

最新视频课程 QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

FRM 2016 PART I BOOK 2: QUANTITATIVE ANALYSIS
©2015 Kaplan, Inc., d.b.a. Kaplan Schweser. All rights reserved.
Printed in the United States of America.
ISBN: 978-1-4754-3804-8
PPN: 3200-7212

Required Disclaimer: GARP® does not endorse, promote, review, or warrant the accuracy of the products or services offered by Kaplan Schweser of FRM® related information, nor does it endorse any pass rates claimed by the provider. Further, GARP® is not responsible for any fees or costs paid by the user to Kaplan Schweser, nor is GARP® responsible for any fees or costs of any person or entity providing any services to Kaplan Schweser. FRM®, GARP®, and Global Association of Risk Professionals™ are trademarks owned by the Global Association of Risk Professionals, Inc.

These materials may not be copied without written permission from the author. The unauthorized duplication of these notes is a violation of global copyright laws. Your assistance in pursuing potential violators of this law is greatly appreciated.

Disclaimer: The SchweserNotes should be used in conjunction with the original readings as set forth by GARP®. The information contained in these books is based on the original readings and is believed to be accurate. However, their accuracy cannot be guaranteed nor is any warranty conveyed as to your ultimate exam success.

FRM官方交流群 392417118

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

Michael Miller, *Mathematics and Statistics for Financial Risk Management, 2nd Edition* (Hoboken, NJ: John Wiley & Sons, 2013).

15. "Probabilities," Chapter 2 (page 13)
16. "Basic Statistics," Chapter 3 (page 29)
17. "Distributions," Chapter 4 (page 53)
18. "Bayesian Analysis," Chapter 6 (page 75)
19. "Hypothesis Testing and Confidence Intervals," Chapter 7 (page 88)

John Hull, *Risk Management and Financial Institutions, 4th Edition* (Hoboken, NJ: John Wiley & Sons, 2015).

20. "Correlations and Copulas," Chapter 11 (page 128)

James Stock and Mark Watson, *Introduction to Econometrics, Brief Edition* (Boston: Pearson Education, 2008).

21. "Linear Regression with One Regressor," Chapter 4 (page 147)
22. "Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals," Chapter 5 (page 161)
23. "Linear Regression with Multiple Regressors," Chapter 6 (page 176)
24. "Hypothesis Tests and Confidence Intervals in Multiple Regression," Chapter 7 (page 190)

Francis X. Diebold, *Elements of Forecasting, 4th Edition* (Mason, Ohio: Cengage Learning, 2006).

25. "Modeling and Forecasting Trend," Chapter 5 (page 209)
26. "Characterizing Cycles," Chapter 7 (page 218)
27. "Modeling Cycles: MA, AR, and ARMA Models," Chapter 8 (page 227)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

John Hull, *Options, Futures, and Other Derivatives, 9th Edition* (New York: Pearson Prentice Hall, 2014).

28. "Estimating Volatilities and Correlations for Risk Management," Chapter 23 (page 237)

    Chris Brooks, *Introductory Econometrics for Finance, 3rd Edition* (Cambridge, UK: Cambridge University Press, 2014).

29. "Simulation Methods," Chapter 13 (page 247)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 i.i.d. random variables. (page 66)
3. Describe independent and identically distributed (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)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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. Correlation and Copulas

After completing this reading, you should be able to:

1. Define correlation and covariance, differentiate between correlation and dependence. (page 128)
2. Calculate covariance using the EWMA and GARCH(1,1) models. (page 130)
3. Apply the consistency condition to covariance. (page 133)
4. Describe the procedure of generating samples from a bivariate normal distribution. (page 134)
5. Describe properties of correlations between normally distributed variables when using a one-factor model. (page 135)
6. Define copula and describe the key properties of copulas and copula correlation. (page 136)
7. Explain tail dependence. (page 140)
8. Describe the Gaussian copula, Student's t-copula, multivariate copula, and one factor copula. (page 138)

21. 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 147)
2. Interpret a population regression function, regression coefficients, parameters, slope, intercept, and the error term. (page 148)
3. Interpret a sample regression function, regression coefficients, parameters, slope, intercept, and the error term. (page 149)
4. Describe the key properties of a linear regression. (page 150)
5. Define an ordinary least squares (OLS) regression and calculate the intercept and slope of the regression. (page 151)
6. Describe the method and three key assumptions of OLS for estimation of parameters. (page 152)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

7. Summarize the benefits of using OLS estimators. (page 152)
8. Describe the properties of OLS estimators and their sampling distributions, and explain the properties of consistent estimators in general. (page 152)
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 153)
10. Interpret the results of an OLS regression. (page 153)

22. 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 161)
2. Interpret the p-value. (page 163)
3. Interpret hypothesis tests about regression coefficients. (page 162)
4. Evaluate the implications of homoskedasticity and heteroskedasticity. (page 166)
5. Determine the conditions under which the OLS is the best linear conditionally unbiased estimator. (page 169)
6. Explain the Gauss-Markov Theorem and its limitations, and alternatives to the OLS. (page 169)
7. Apply and interpret the t-statistic when the sample size is small. (page 169)

23. 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 176)
2. Distinguish between single and multiple regression. (page 177)
3. Interpret the slope coefficient in a multiple regression. (page 178)
4. Describe homoskedasticity and heteroskedasticity in a multiple regression. (page 179)
5. Describe the OLS estimator in a multiple regression. (page 177)
6. Calculate and interpret measures of fit in multiple regression. (page 179)
7. Explain the assumptions of the multiple linear regression model. (page 182)
8. Explain the concept of imperfect and perfect multicollinearity and their implications. (page 182)

24. 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 190)
2. Construct, apply, and interpret joint hypothesis tests and confidence intervals for multiple coefficients in a multiple regression. (page 196)
3. Interpret the F-statistic. (page 196)
4. Interpret tests of single restrictions involving multiple coefficients. (page 201)
5. Interpret confidence sets for multiple coefficients. (page 196)
6. Identify examples of omitted variable bias in multiple regressions. (page 202)
7. Interpret the R^2 and adjusted-R^2 in a multiple regression. (page 201)

25. Modeling and Forecasting Trend

After completing this reading, you should be able to:
1. Define mean squared error (MSE) and explain the implications of MSE in model selection. (page 209)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

2. Explain how to reduce the bias associated with MSE and similar measures. (page 210)
3. Compare and evaluate model selection criteria, including s^2, the Akaike information criterion (AIC), and the Schwarz information criterion (SIC). (page 211)
4. Explain the necessary conditions for a model selection criterion to demonstrate consistency. (page 212)

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 218)
2. Describe the requirements for a series to be covariance stationary. (page 219)
3. Explain the implications of working with models that are not covariance stationary. (page 219)
4. Define white noise, describe independent white noise and normal (Gaussian) white noise. (page 219)
5. Explain the characteristics of the dynamic structure of white noise. (page 219)
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 227)
2. Describe the properties of a general finite-order process of order q (MA(q)) process. (page 229)
3. Describe the properties of the first-order autoregressive (AR(1)) process, and define and explain the Yule-Walker equation. (page 229)
4. Describe the properties of a general pth order autoregressive (AR(p)) process. (page 231)
5. Define and describe the properties of the autoregressive moving average (ARMA) process. (page 231)
6. Describe the application of AR and ARMA processes. (page 232)

28. Estimating Volatilities and Correlations for Risk Management

After completing this reading, you should be able to:

1. Explain how various weighting schemes can be used in estimating volatility. (page 237)
2. Apply the exponentially weighted moving average (EWMA) model to estimate volatility. (page 238)
3. Describe the generalized autoregressive conditional heteroskedasticity (GARCH(p,q)) model for estimating volatility and its properties:
   - Calculate volatility using the GARCH(1,1) model
   - Explain mean reversion and how it is captured in the GARCH(1,1) model (page 239)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

4. Explain the weights in the EWMA and GARCH(1,1) models. (page 238)
5. Explain how GARCH models perform in volatility forecasting. (page 241)
6. Describe the volatility term structure and the impact of volatility changes. (page 241)
7. Describe how correlations and covariances are calculated, and explain the consistency condition for covariances. (page 242)

29. Simulation Methods

After completing this reading, you should be able to:

1. Describe the basic steps to conduct a Monte Carlo simulation. (page 247)
2. Describe ways to reduce Monte Carlo sampling error. (page 248)
3. Explain how to use antithetic variate technique to reduce Monte Carlo sampling error. (page 249)
4. Explain how to use control variates to reduce Monte Carlo sampling error when it is effective. (page 250)
5. Describe the benefits of reusing sets of random number draws across Monte Carlo experiments and how to reuse them. (page 251)
6. Describe the bootstrapping method and its advantage over Monte Carlo simulation. (page 251)
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 253)
8. Describe situations where the bootstrapping method is ineffective. (page 253)
9. Describe the disadvantages of the simulations approach to financial problem solving. (page 254)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

FRM官方交流群 392417118

The Time Value of Money

Exam Focus

This optional reading 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 below. These are the only keys you need to know to work virtually all TVM problems.

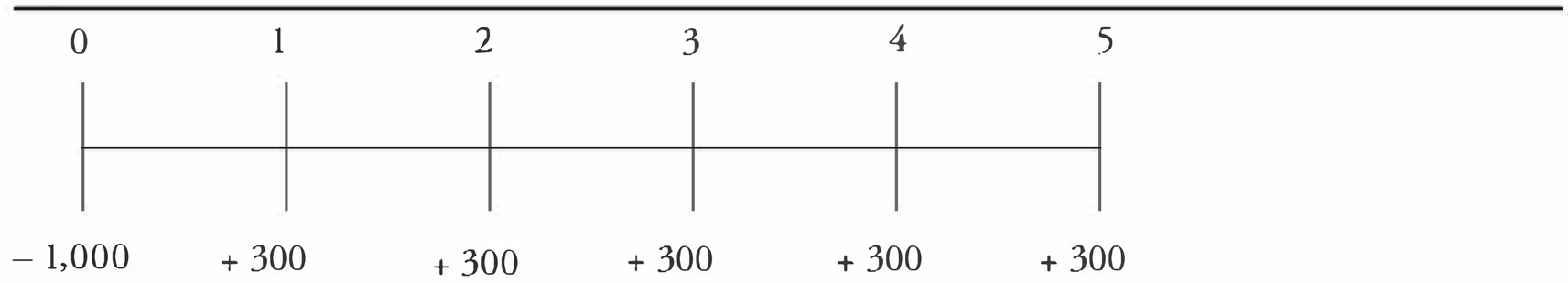
- N = Number of compounding periods.
- I/Y = Interest rate per compounding period.
- PV = Present value.
- FV = Future value.
- PMT = Annuity payments, or constant periodic cash flow.
- CPT = Compute.

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



©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 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)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

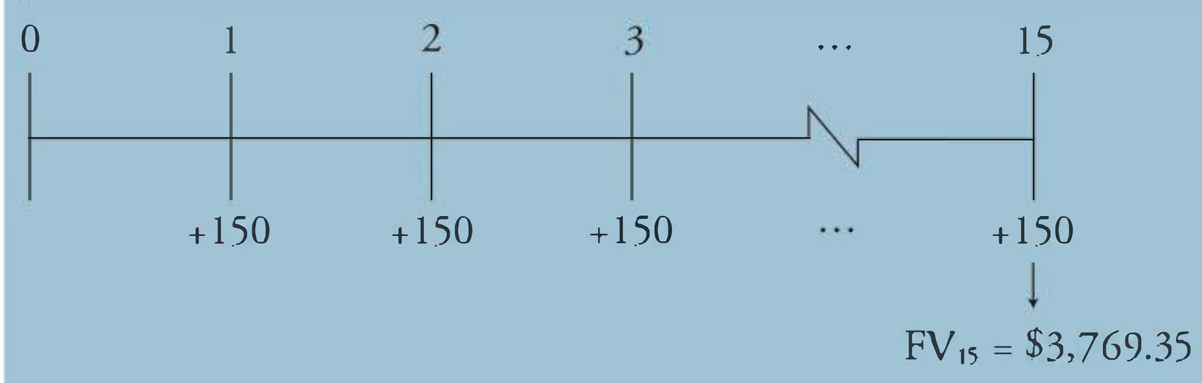
Implicit here is that PV = 0.

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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%.

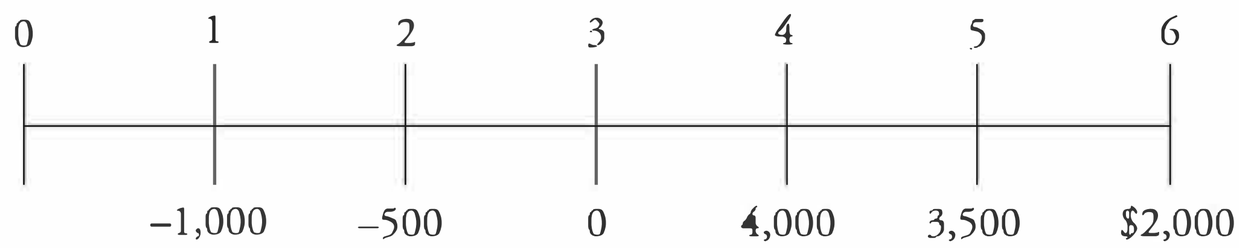
©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 = 2,000.00

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

最新视频课程 QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checker Answers

1. **C** N = 15; I/Y = 9; PV = –1,000; PMT = 0; CPT → FV = $3,642.48

2. **B** N = 7; I/Y = 8; FV = –10,000; PMT = 0; CPT → PV = $5,834.90

3. **B** N = 8; PV = –3; FV = 4.50; PMT = 0; CPT → I/Y = 5.1989

4. **C** PV = –5,000; I/Y = 12; FV = 10,000; PMT = 0; CPT → N = 6.12. Rule of 72 → 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 × 12 = 360; I/Y = 9 / 12 = 0.75; PV = –150,000(1 – 0.2) = –120,000; FV = 0; CPT → PMT = $965.55

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 $p(x) = P(X = x)$.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

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 $p(x_1) = p(x_2) = 0$.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

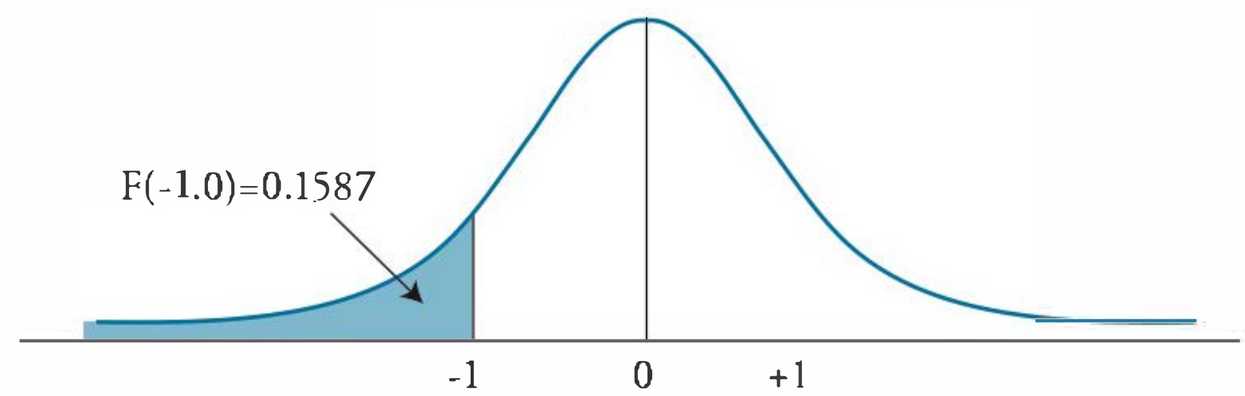
Figure 1 shows an example of a cumulative distribution function (for a standard normal distribution, described in Topic 17). There is a 15.87% probability of a value less than –1. This is the total area to the left of –1 in the pdf in Panel (a), and the y-axis value of the cdf for a value of –1 in Panel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

©2015 Kaplan, Inc.

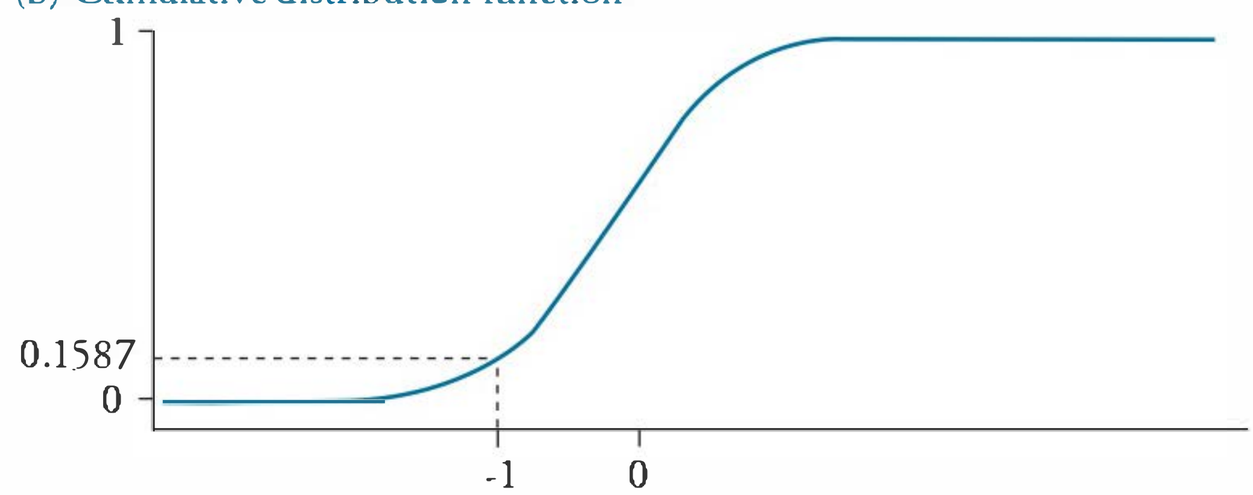
FRM官方交流群 392417118

Figure 1: Standard Normal Probability Density and Cumulative Distribution Functions

(a) Probability density function



(b) Cumulative distribution function



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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Example: Discrete uniform distribution

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

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

Answer:

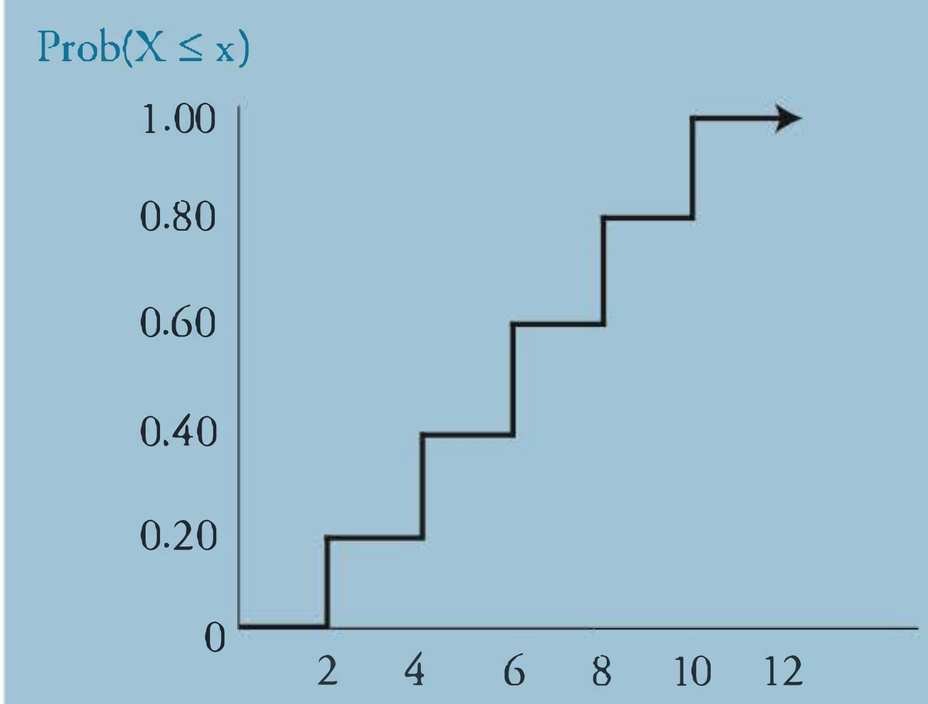
p(6) = 0.2, since p(x) = 0.2 for all *x*. $F(6) = P(X \leq 6) = np(x) = 3(0.2) = 0.6$. Note that n = 3 since 6 is the third outcome in the range of possible outcomes.

$P(2 \leq X \leq 8) = 4(0.2) = 0.8$. Note that k = 4, since there are four outcomes in the range $2 \leq X \leq 8$. 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}



FRM官方交流群 392417118

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, $E_1, E_2, \ldots E_n$, 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 | 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 | B), times the unconditional probability of B, P(B)."

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 | I) × P(I)

P(RI) = 0.7 × 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 | B) = P(A), or equivalently, P(B | 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).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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)

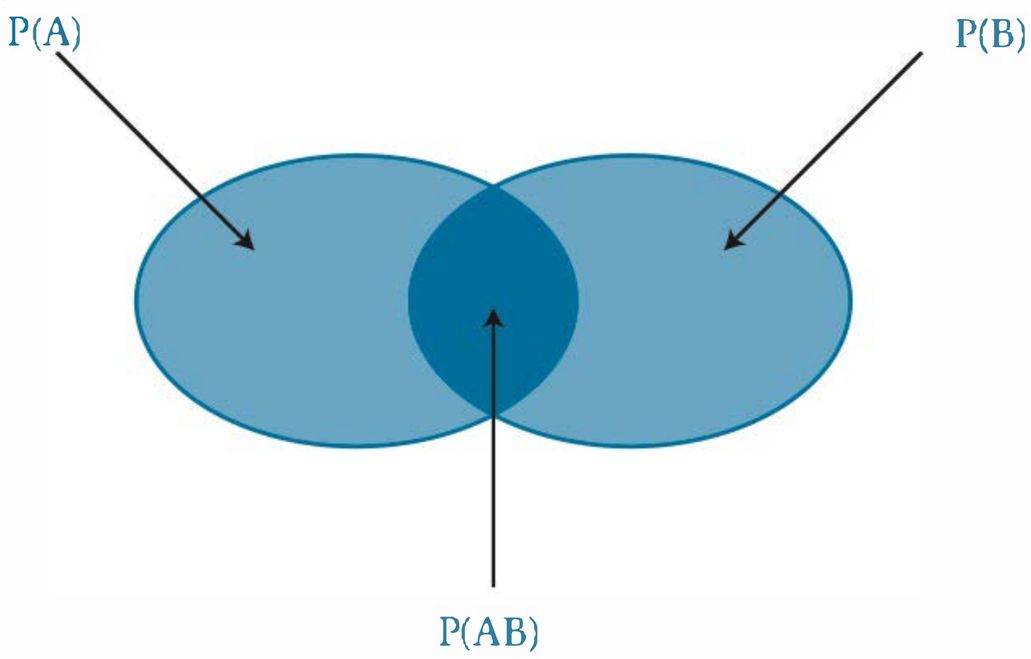
For **mutually exclusive events** where the joint probability, P(AB), is zero, the probability that either A or B will occur is simply the sum of the unconditional probabilities for each event, P(A or B) = P(A) + P(B).

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).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 or I) = P(R) + P(I) – P(RI)

P(R or I) = 0.34 + 0.40 – 0.28

P(R 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 on first die and 4 on second die) = P(4 on first die) × P(4 on second die) = 1/6 × 1/6 = 1/36 = 0.0278

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

P(heads on first coin and heads on second coin) = 1/2 × 1/2 = 1/4 = 0.25

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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%

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

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

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

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 below 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%.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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%

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

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

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

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

FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 figure below. **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%	

FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

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

   where $\mu = E(X)$

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

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

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

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

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

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

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

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

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

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

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

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

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

最新视频课程 QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

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

$Var(X) = 8 + 8 = 16$

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:

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

This equation simplifies to:

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Properties of covariance include:

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

   $Cov(X,Y) = 0$

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

   $Cov(X,X) = Var(X)$

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

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

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

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

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

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 above property.

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 table below. What is the covariance of the returns for Stock A and Stock B?

FRM官方交流群 392417118

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
				$Cov(R_A, R_B) = \Sigma P(S) \times [R_A - E(R_A)] \times [R_B - E(R_B)]$ = 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, $-1 \leq Corr(R_i, R_j) \leq +1$.
- If $Corr(R_i, R_j) = 1.0$, 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

- 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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

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

FRM官方交流群 392417118

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

excess kurtosis = 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.

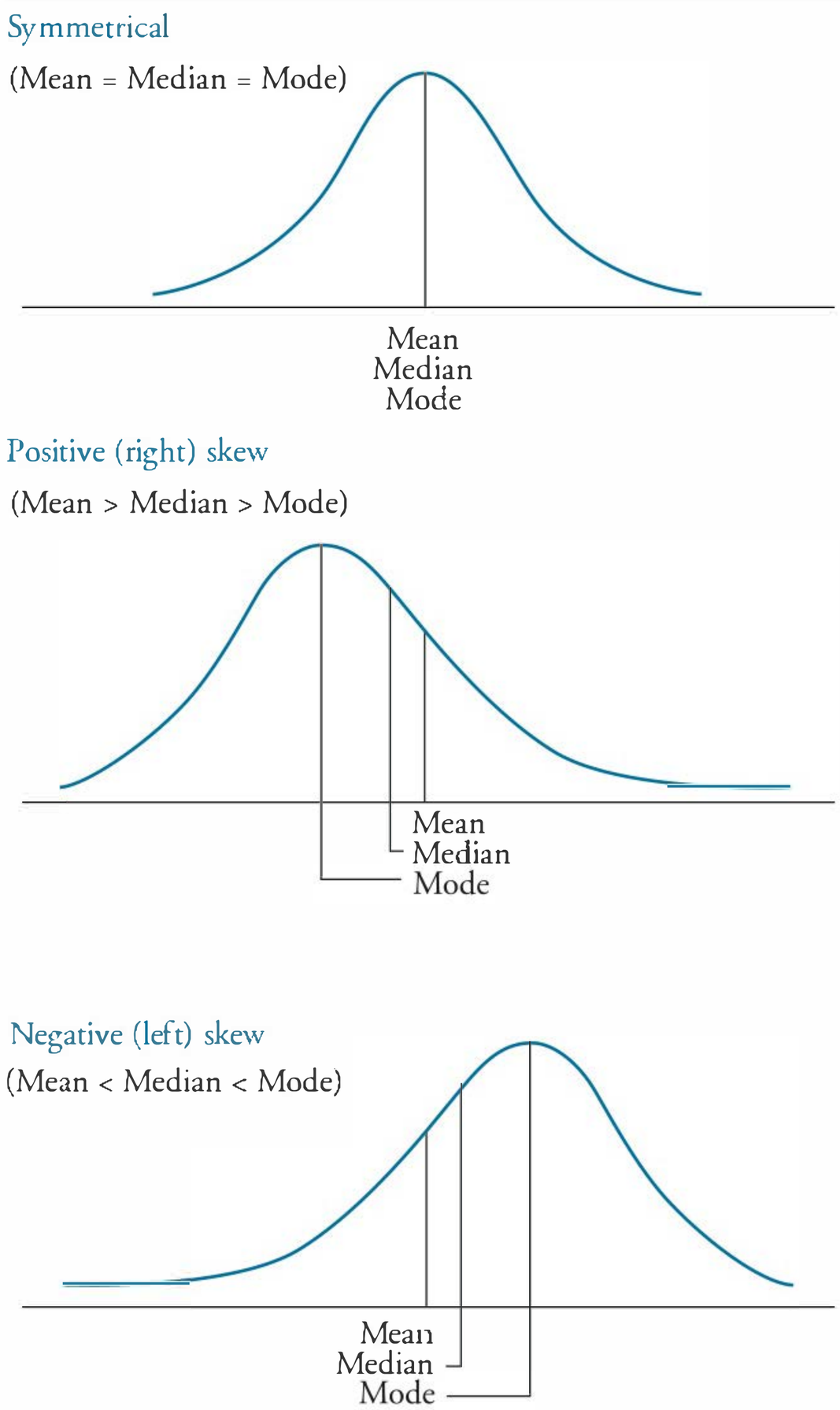
Skewness affects the location of the mean, median, and mode of a distribution.

- 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).

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

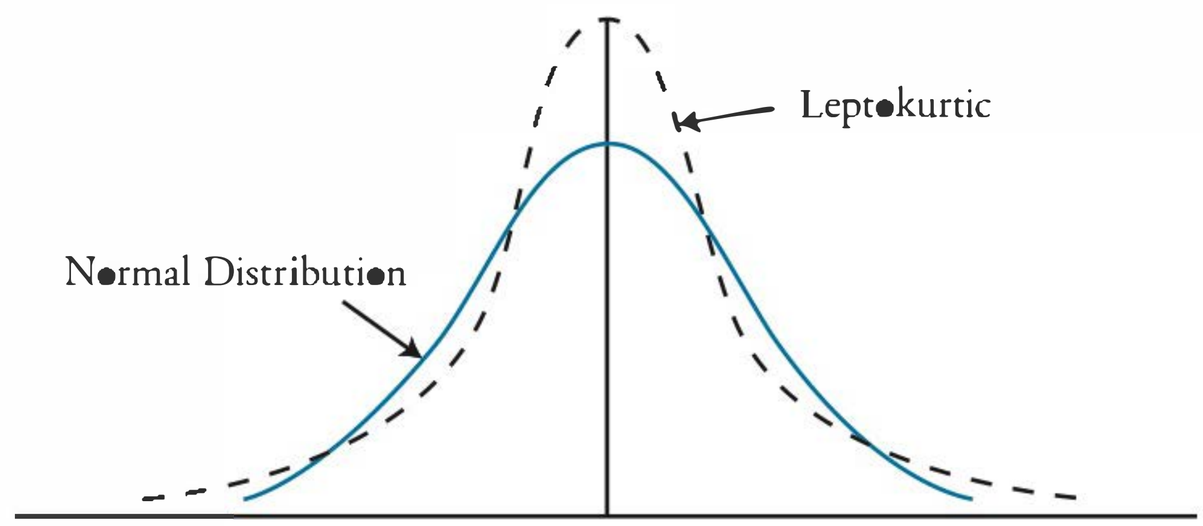
As indicated in Figure 4, a leptokurtic return distribution will have more returns clustered around the mean and more returns with large deviations from the mean (fatter tails).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 $\left[E\left(\bar{x}\right) = \mu\right]$, 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).

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

$$Var(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$$
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:

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

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

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

4. Given the probability matrix above, 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.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

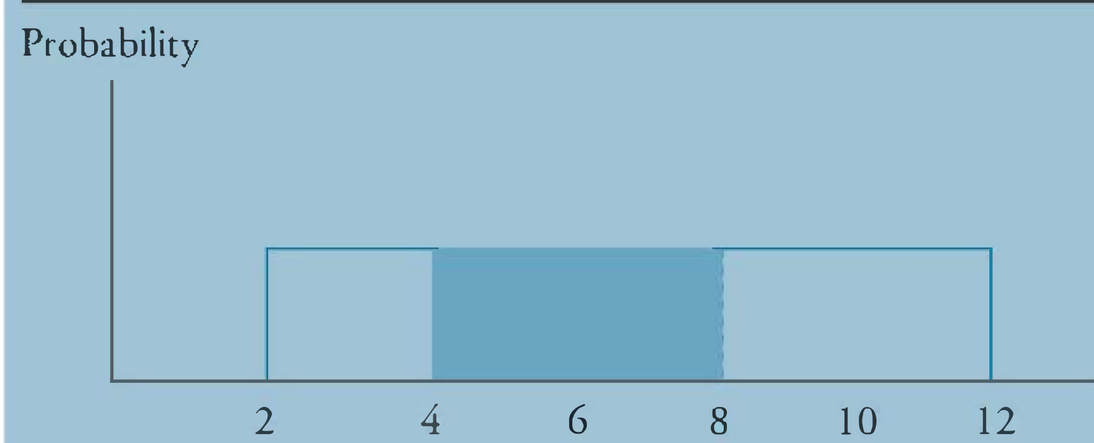
X is uniformly distributed between 2 and 12. Calculate the probability that *X* will be between 4 and 8.

Answer:

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

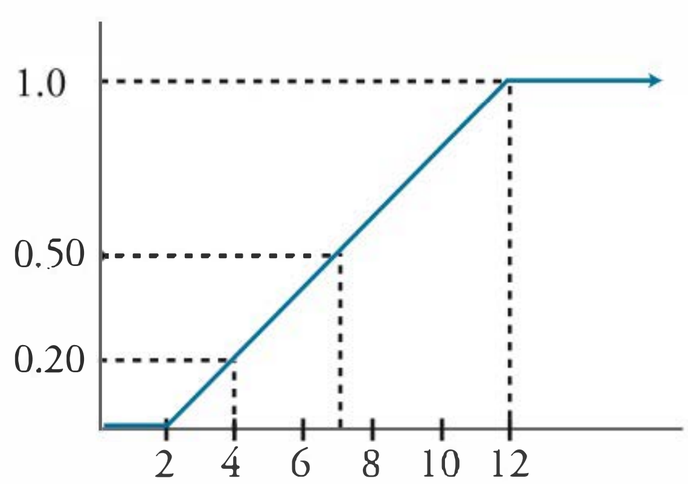
The figure below illustrates this continuous uniform distribution. Note that the area bounded by 4 and 8 is 40% of the total probability between 2 and 12 (which is 100%).

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 above 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:

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

FRM官方交流群 392417118

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 above.

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).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

$$Var(X) = np(1 - p) = 5(0.67)(0.33) = 1.106$$

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

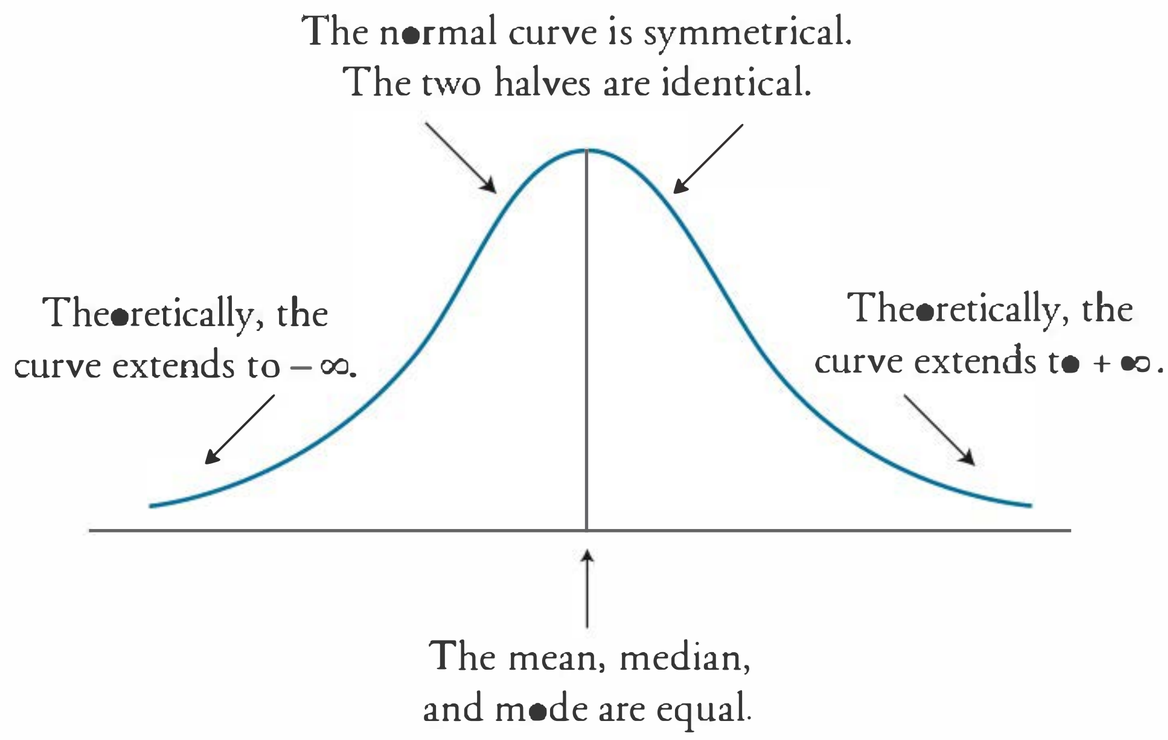
- It is completely described by its mean, μ, and variance, σ^2, stated as $X \sim N(\mu, \sigma^2)$. In words, this says that "*X* is normally distributed with mean μ and variance σ^2."
- Skewness = 0, meaning the normal distribution is symmetric about its mean, so that $P(X \leq \mu) = P(\mu \leq X) = 0.5$, 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

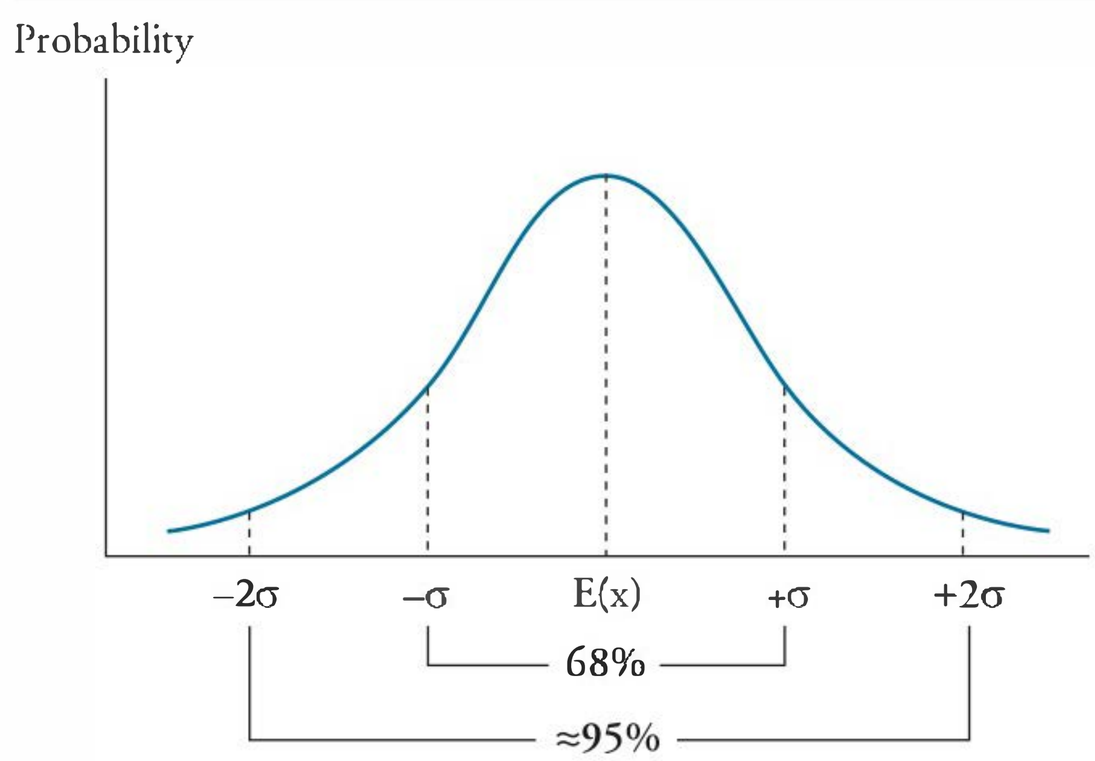
Figure 2: Normal Distribution Probability Density Function



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 $\bar{X} + 2.58s$.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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., $N\sim(0,1)$]. 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.

最新视频课程
QQ: 2211892999

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 below).

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.

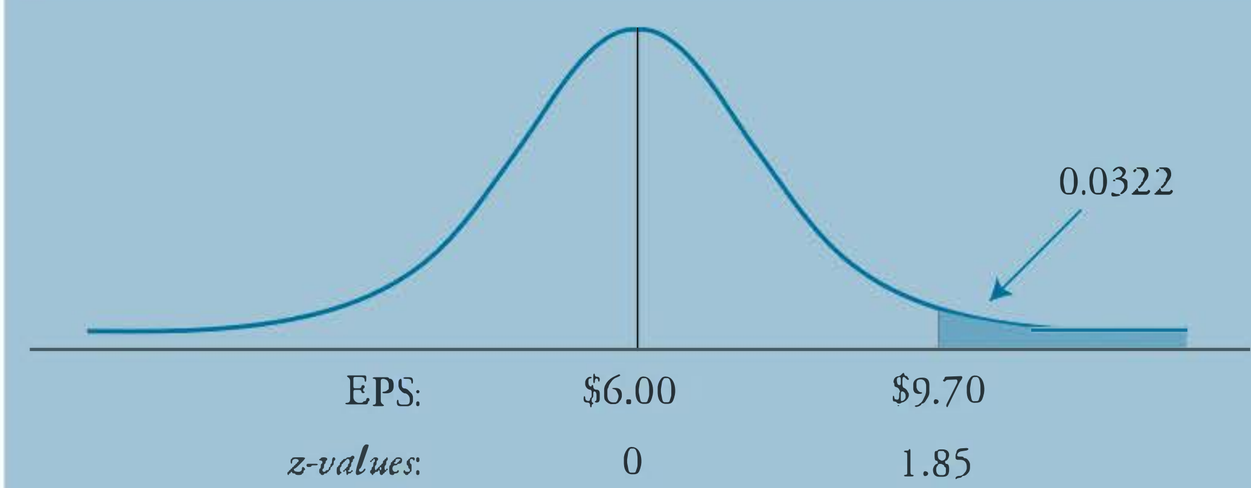
©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 below, 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:

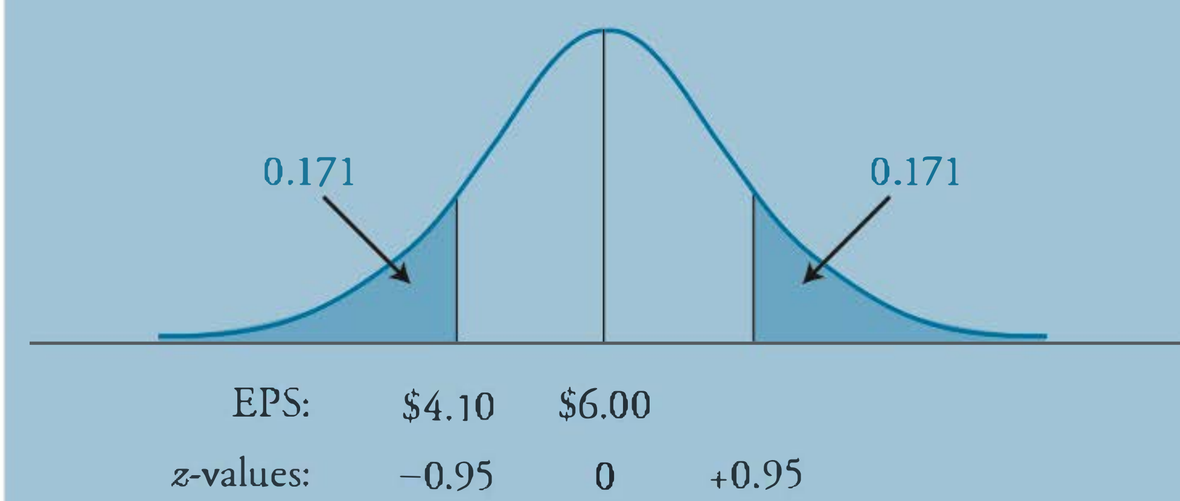
$$z = \frac{(\$4.10 - \$6)}{2} = -0.95,$$

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%.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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

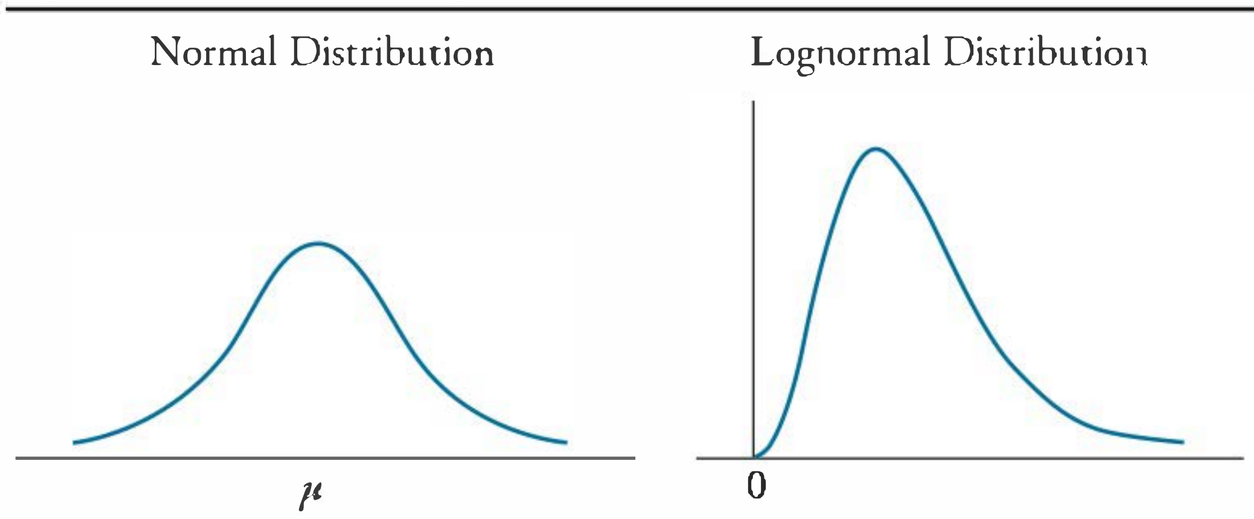
The lognormal distribution is generated by the function e^x, where x is normally distributed. Since the natural logarithm, ln, of e^x is x, the logarithms of lognormally distributed random variables are normally distributed, thus the name.

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

FRM官方交流群 392417118

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 i.i.d. random variables.

LO 17.3: Describe independent and identically distributed (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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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

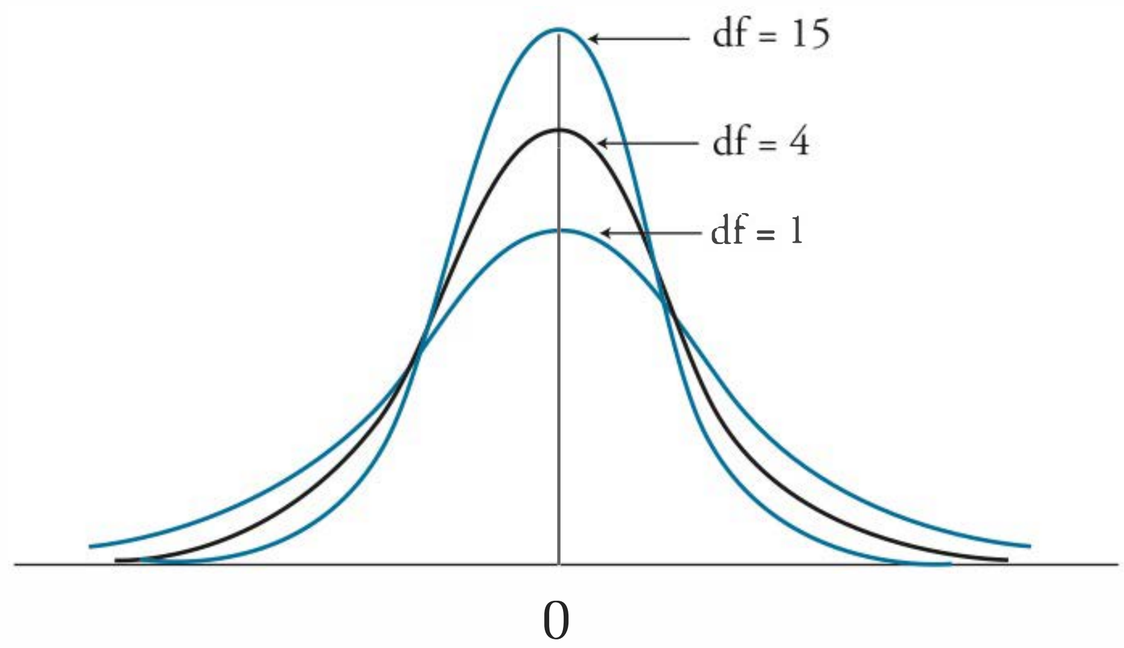
Figure 7 illustrates the different shapes of the *t*-distribution associated with different degrees of freedom. The tendency is for the *t*-distribution to look more and more like the normal

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

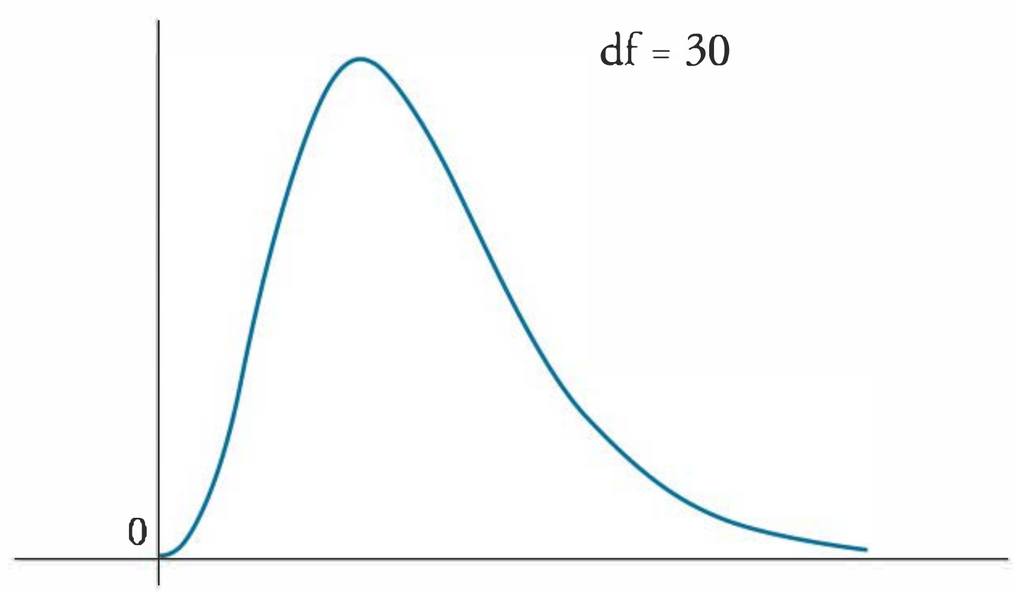
Figure 7: *t*-Distributions for Different Degrees of Freedom (df)



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-square 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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

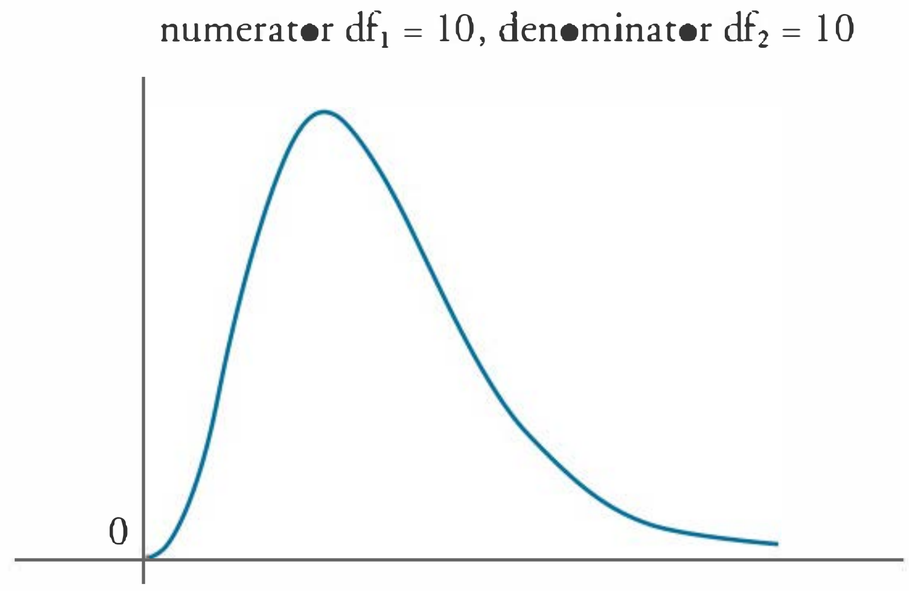
- The *F*-distribution approaches the normal distribution as the number of observations increases (just as with the *t*-distribution and chi-squared distribution).
- A random variable's *t*-value squared (t^2) with n – 1 degrees of freedom is *F*-distributed with 1 degree of freedom in the numerator 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 $\rightarrow \infty$

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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, $P(A \mid B)$. 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 above equation $[P(B \mid A) \times P(A)]$ 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:

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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%).

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 | 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.

FRM官方交流群 392417118

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 | 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 | 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 | E) × 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 | 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).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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?

FRM官方交流群 392417118

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 21.37% 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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 above, 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%.

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

FRM官方交流群 392417118

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%).

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

最新视频课程
QQ: 2211892999

FRM官方交流群 392417118

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%.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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%.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 *n*th 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:

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Thus, when applying the following variance property:

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

and assuming *a* and *c* are equal to 0.5, the variance of the sample average, $\mathrm{Var}(\overline{X})$, will be equal to $\frac{\sigma_X^2}{2}$. In more general terms, $\mathrm{Var}(\overline{X}) = \frac{\sigma_X^2}{n}$ 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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

FRM官方交流群 392417118

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 above, 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?

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

$$H_0: \mu = \mu_0 \text{ versus } H_A: \mu \neq \mu_0$$

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:

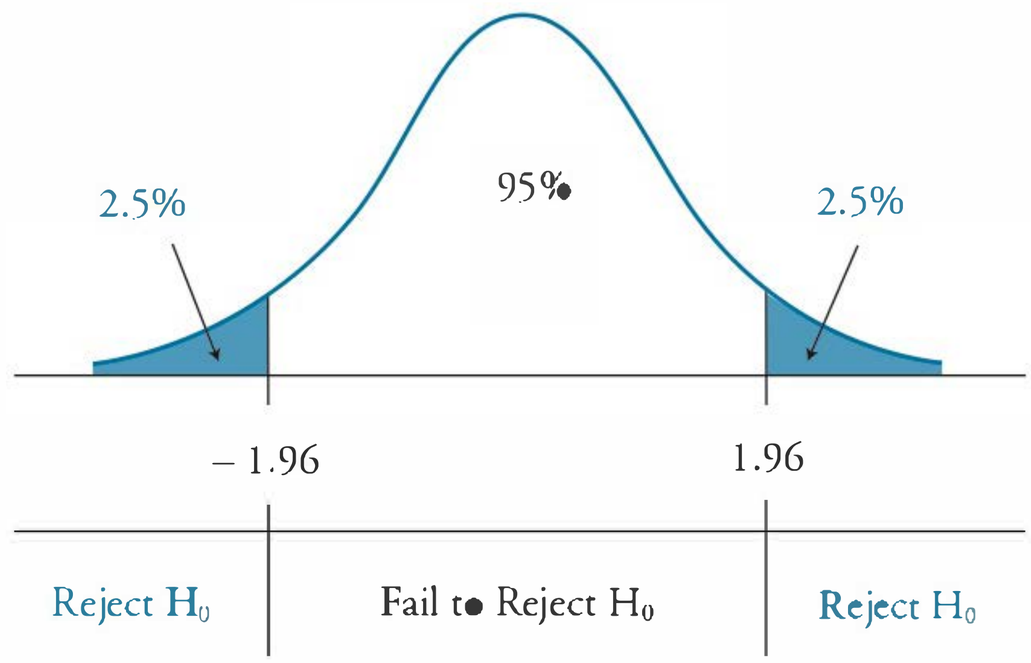
Reject H_0 if: test statistic < −1.96 or
test statistic > 1.96

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.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

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:

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

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:

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 $H_A: \mu > \mu_0$.

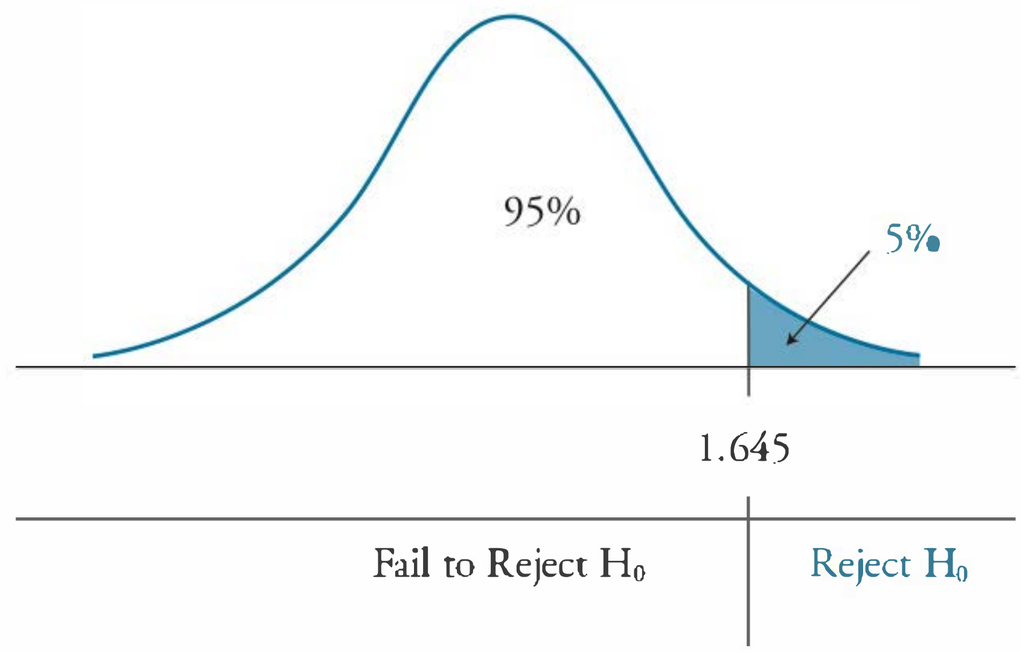
- If the calculated test statistic is greater than 1.645, we conclude that the sample statistic is sufficiently greater than the hypothesized value. In other words, we reject the null hypothesis.
- If the calculated test statistic is less than 1.645, we conclude that the sample statistic is not sufficiently 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

$$H_0: \mu \leq 0 \text{ versus } H_A: \mu > 0$$

The appropriate decision rule for this one-tailed z-test at a significance level of 5% is:

$$\text{Reject } H_0 \text{ if: test statistic} > 1.645$$

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

$$\left\{\left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix}-\left(\begin{matrix}\text{critical}\\\text{value}\end{matrix}\right)\left(\begin{matrix}\text{standard}\\\text{error}\end{matrix}\right)\right]\leq\begin{matrix}\text{population}\\\text{parameter}\end{matrix}\leq\left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix}+\left(\begin{matrix}\text{critical}\\\text{value}\end{matrix}\right)\left(\begin{matrix}\text{standard}\\\text{error}\end{matrix}\right)\right]\right\}$$

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 $\mu = 0$ should be rejected.

Answer:

Given a sample size of 250 with a standard deviation of 0.25%, the standard error can be computed as $s_{\bar{x}} = s/\sqrt{n} = 0.25/\sqrt{250} = 0.0158\%$.

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 $\mu = 0$ because 0 is not within the confidence interval.

Notice the similarity of this analysis with our test of whether $\mu = 0$. We rejected the hypothesis $\mu = 0$ because the sample mean of 0.1% is more than 1.96 standard errors from zero. Based on the 95% confidence interval, we reject $\mu = 0$ 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 above strategy, 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

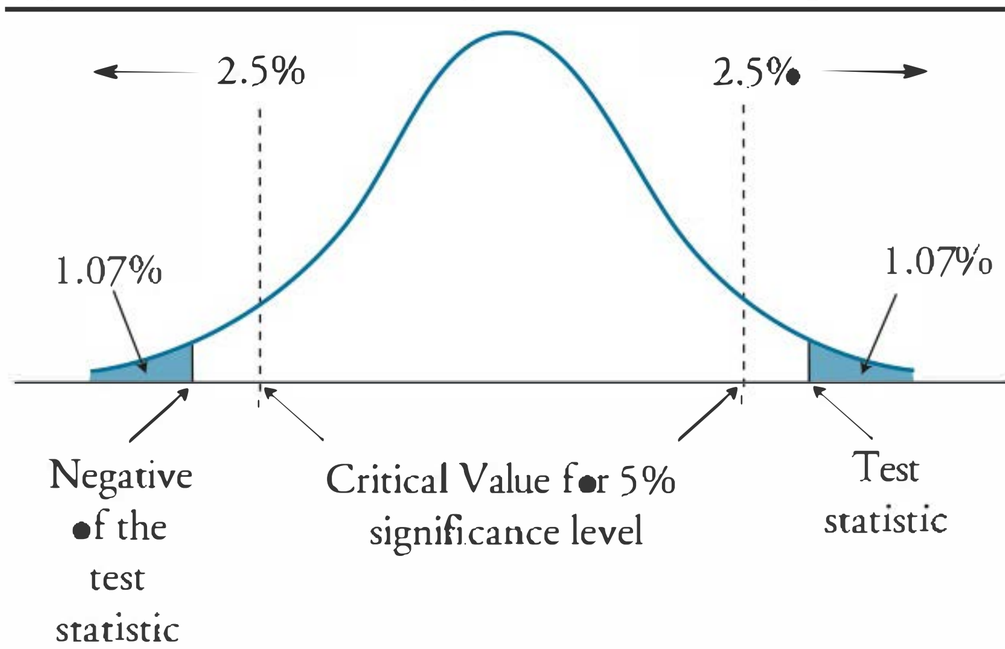
Any of these factors could make committing funds to a strategy unattractive, even though the statistical evidence of positive returns is highly significant. By the nature of statistical tests, a very large sample size can result in highly (statistically) significant results that are quite small in absolute 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 × 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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

$$z\text{-statistic} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

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.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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: $\mu = 2.5$ (The machine does not need an adjustment.)
H_A: $\mu \neq 2.5$ (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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

$$H_0: \sigma^2 = \sigma_0^2 \text{ versus } H_A: \sigma^2 \neq \sigma_0^2$$

The hypotheses for one-tailed tests are structured as:

$$H_0: \sigma^2 \leq \sigma_0^2 \text{ versus } H_A: \sigma^2 > \sigma_0^2, \text{ or}$$
$$H_0: \sigma^2 \geq \sigma_0^2 \text{ versus } H_A: \sigma^2 < \sigma_0^2$$

Hypothesis testing of the population variance requires the use of a chi-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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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

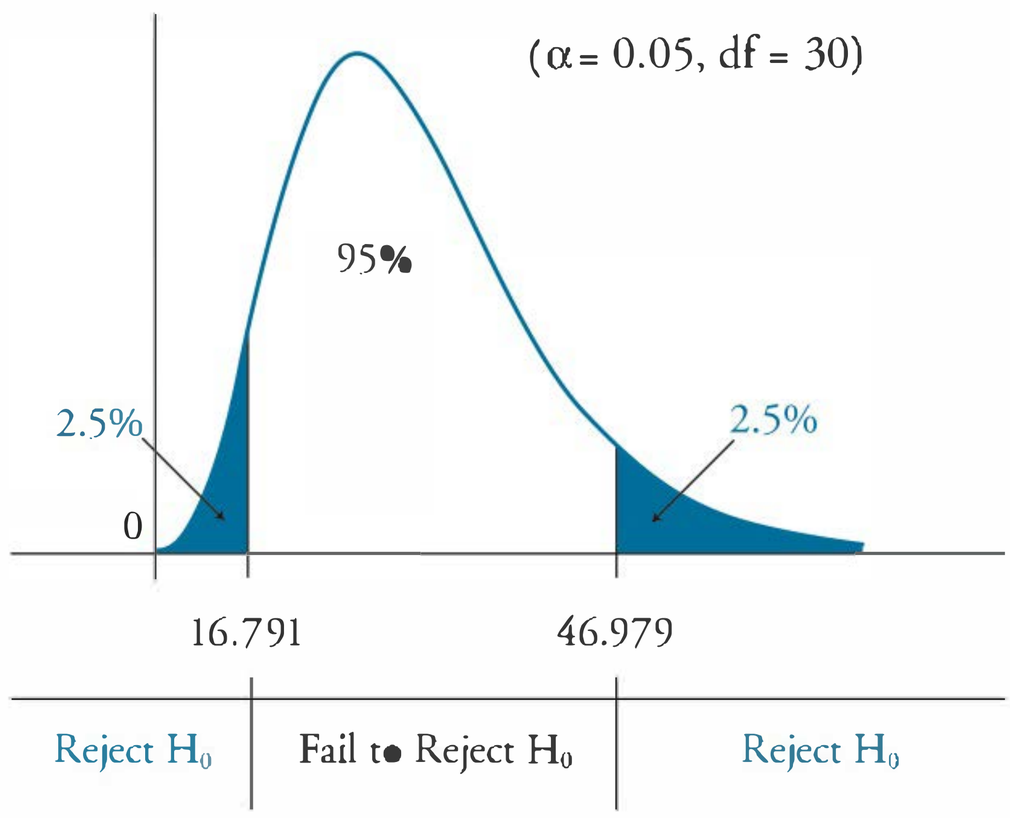


Figure 9: Chi-Squared Table

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

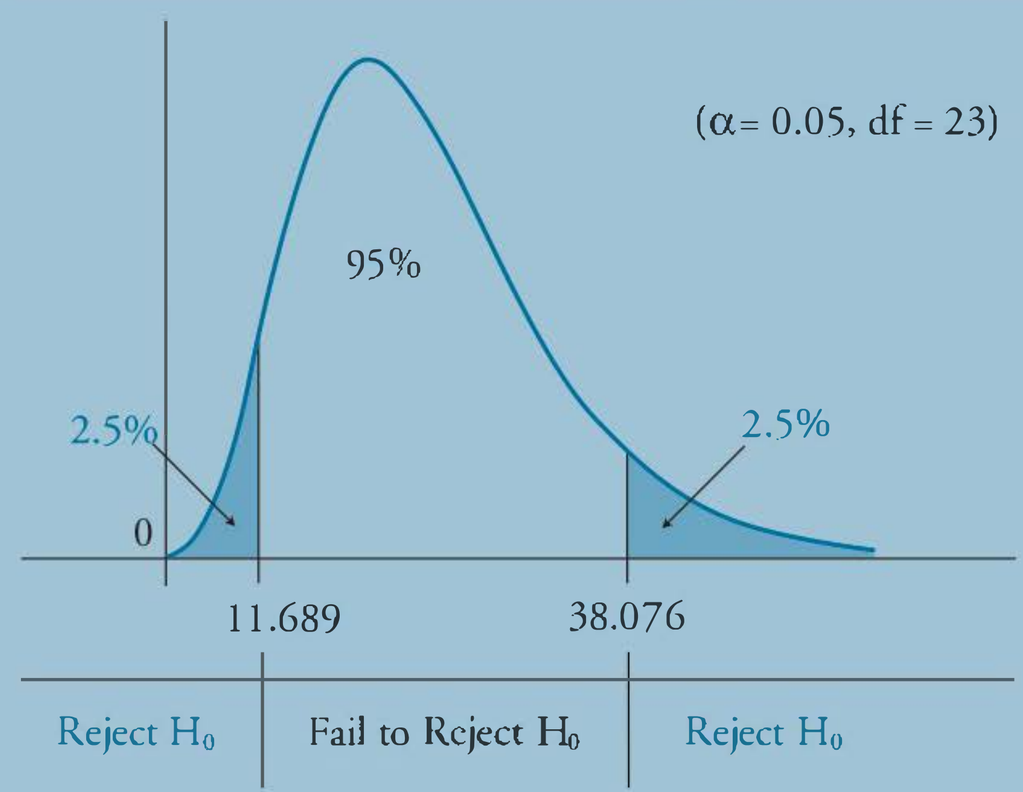
State the decision rule regarding the hypothesis. With a 24-month sample, there are 23 degrees of freedom. Using the table of chi-squared values at the back of this book, for 23 degrees of freedom and probabilities of 0.975 and 0.025, we find two critical values, 11.689 and 38.076. Thus, the decision rule is:

$$\text{Reject } H_0 \text{ if: } \chi^2 < 11.689, \text{ or } \chi^2 > 38.076$$

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ versus } H_A: \sigma_1^2 \neq \sigma_2^2$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

and the one-sided test structures can be specified as:

$$H_0\text{: } \sigma_1^2 \le \sigma_2^2 \text{ versus } H_A\text{: } \sigma_1^2 > \sigma_2^2\text{, or } H_0\text{: } \sigma_1^2 \ge \sigma_2^2 \text{ versus } H_A\text{: } \sigma_1^2 < \sigma_2^2$$

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

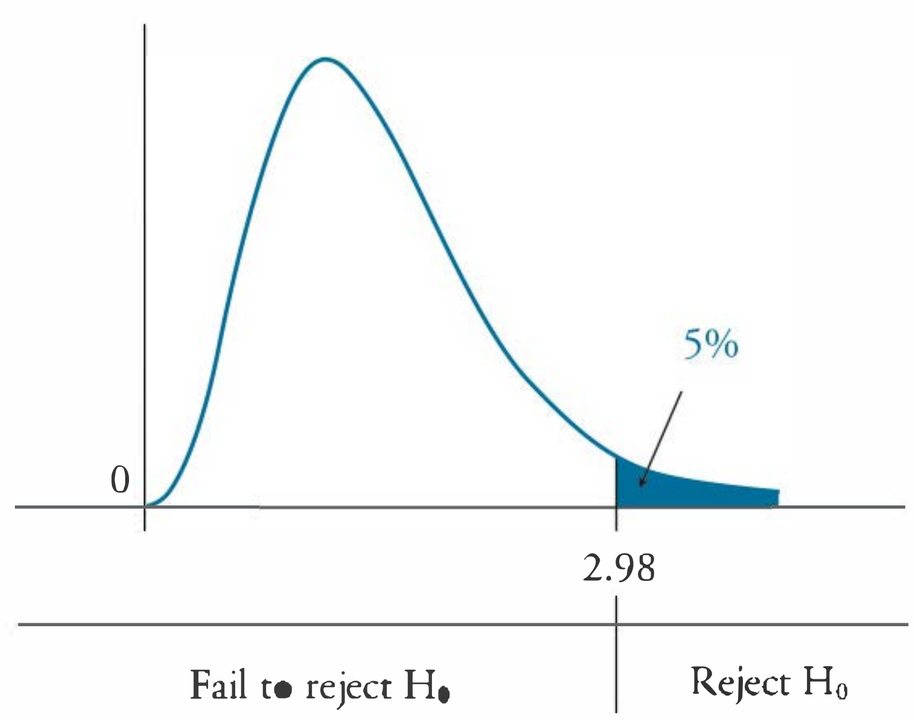
s_2^2 = variance of the sample of n_2 observations drawn from Population 2

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



©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

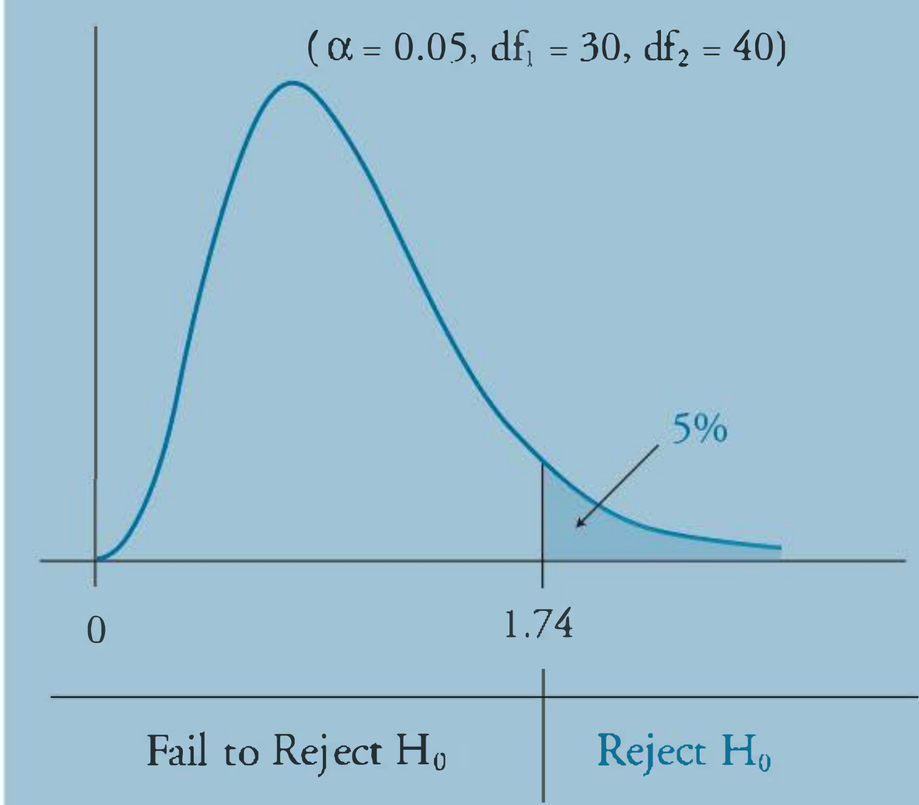
State the decision rule regarding the hypothesis. Using the sample sizes for the two industries, the critical *F*-value for our test is found to be 1.74. This value is obtained from the table of the *F*-distribution at the 5% level of significance with $df_1 = 30$ and $df_2 = 40$. Thus, if the computed *F*-statistic is greater than the critical value of 1.74, the null hypothesis is rejected. The decision rule, illustrated in the distribution below, can be stated as:

Reject H_0 if: $F > 1.74$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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?

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 19.1

$$\text{Population variance} = \sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}, \text{ where } \mu = \text{population mean and } N = \text{size}$$

$$\text{Sample variance} = s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}, \text{ where } \overline{X} = \text{sample mean and } n = \text{sample size}$$

The standard error of the sample mean is the standard deviation of the distribution of the sample means and is calculated as $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$, where σ, the population standard deviation, is known, and as $s_{\overline{x}} = \frac{s}{\sqrt{n}}$, 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:

- $\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ when the variance is known
- $\overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ when the variance is unknown

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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: $\mu <$ $145,000.
   B. H_A: $\mu \pm$ $145,000.
   C. H_A: $\mu \geq$ $145,000.
   D. H_A: $\mu >$ $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.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

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%.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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** $z = \dfrac{149{,}750 - 145{,}000}{24{,}000 / \sqrt{36}} = 1.1875.$

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:

$$s = \left[\frac{\left[(22-7.3)^2 + (5-7.3)^2 + (-7-7.3)^2 + (11-7.3)^2 + (2-7.3)^2 + (11-7.3)^2\right]}{6-1}\right]^{1/2}$$

$$= \left(96.3\%^2\right)^{1/2} = 9.8\%$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

Correlations and Copulas

Topic 20

Exam Focus

This topic examines correlation and covariance calculations and how covariance is used in exponentially weighted moving average (EWMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models. The later part of this topic defines copulas and distinguishes between several different types of copulas. For the exam, be able to calculate covariance using EWMA and GARCH(1,1) models. Also, understand how copulas are used to estimate correlations between variables. Finally, be able to explain how marginal distributions are mapped to known distributions to form copulas.

Correlation and Covariance

LO 20.1: Define correlation and covariance, differentiate between correlation and dependence.

Correlation and covariance refer to the co-movements of assets over time and measure the strength between the linear relationships of two variables. Correlation and covariance essentially measure the same relationship; however, correlation is standardized so the value is always between –1 and 1. This standardized measure is more convenient in risk analysis applications than covariance, which can have values between $-\infty$ and ∞. **Correlation** is mathematically determined by dividing the covariance between two random variables, cov(X,Y), by the product of their standard deviations, $\sigma_X\sigma_Y$.

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X\sigma_Y}$$

Multiplying each side of this equation by $\sigma_X\sigma_Y$ provides the formula for calculating **covariance:**

$$\text{cov}(X,Y) = \rho_{X,Y} \times \sigma_X\sigma_Y$$

In practice, it is necessary to first calculate the covariance between two random variables using the following equation and then solve for the standardized correlation.

$$\text{cov}(X,Y) = E\left[\left(X - E(X)\right) \times \left(Y - E(Y)\right)\right] = E(X,Y) - E(X) \times E(Y)$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

In this covariance equation, E(X) and E(Y) are the means or expected values of random variables *X* and *Y*, respectively. E(X,Y) is the expected value of the product of random variables *X* and *Y*.

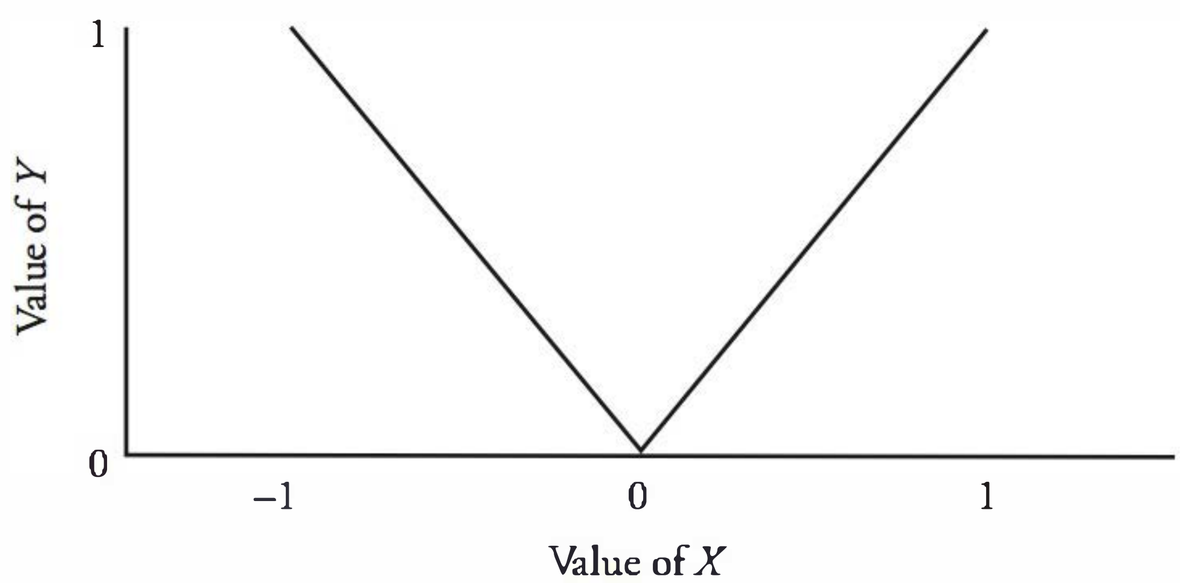
Variables are defined as independent variables if the knowledge of one variable does not impact the probability distribution for another variable. In other words, the conditional probability of V_2 given information regarding the probability distribution of V_1 is equal to the unconditional probability of V_2 as expressed in the following equation:

$$P(V_2 \mid V_1 = x) = P(V_2)$$

A correlation of zero between two variables *does not* imply that there is no dependence between the two variables. It simply implies that there is no linear relationship between the two variables, but the value of one variable can still have a nonlinear relationship with the other variable.

As an example, suppose variable *X* has three expected values of –1, 0, and 1 with an equal probability of occurrence, and variable *Y* has a value of 1 when variable *X* has a value of either –1 or 1. When variable *X* has a value of 0, then variable *Y* has a value of 0. This V-shaped relationship is illustrated in Figure 1.

Figure 1: Relationship between *X* and *Y*

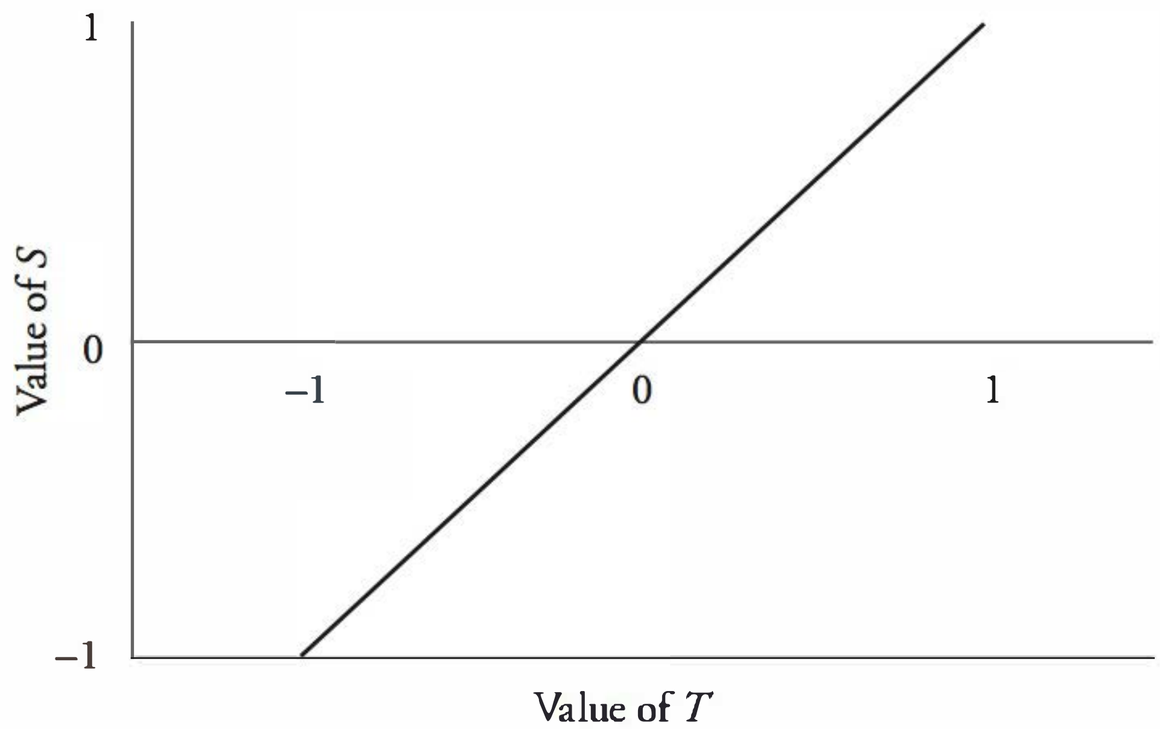


Also suppose that variables *S* and *T* are perfectly positively correlated and that variable *S* has three expected values of –1, 0, and 1 with an equal probability of occurrence. When variable *S* has a value of –1, 0, or 1, then variable *T* has a value of –1, 0, and 1, respectively. This relationship is illustrated in Figure 2.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Figure 2: Relationship between *S* and *T*



With the above information, we can now determine the correlation coefficient and dependency of these two pairs of variables. In this example, the coefficient of correlation between variables *X* and *Y* is zero, and the coefficient of correlation between variables *S* and *T* is one.

The coefficient of correlation is a statistical measure of linear dependency. If we know the value of *X*, it will change our expectations of the value or probability distribution of *Y*. Likewise, if we know the value of *Y*, it will change our expectations of the probability distribution of *X*. Clearly, there is a dependency between *X* and *Y*, as well as a dependency between *S* and *T*. A practical example of the V-shaped dependency in Figure 1 is with respect to financial derivatives that may have more value with large market movements in either direction.

Covariance Using EWMA and GARCH Models

LO 20.2: Calculate covariance using the EWMA and GARCH(1,1) models.

Professor's Note: EWMA and GARCH models for volatility will be discussed in greater detail in Topic 28.

EWMA Model

Covariance is a statistical measure that is calculated over historical time periods. Conventional wisdom suggests that more recent observations should carry more weight because they more accurately reflect the current market environment. The following equation calculates a new covariance on day *n* using an **exponentially weighted moving**

©2015 Kaplan, Inc.
FRM官方交流群 392417118

average (EWMA) model. This model is designed to vary the weight given to more recent observations (by adjusting λ).

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1-\lambda)X_{n-1}Y_{n-1}$$

where:
λ = the weight for the most recent covariance on day $n - 1$
X_{n-1} = the percentage change for variable X on day $n - 1$
Y_{n-1} = the percentage change for variable Y on day $n - 1$

Example: Calculating covariance using the EWMA model

Assume an analyst uses the EWMA model with $\lambda = 0.90$ to update correlation and covariance rates. The correlation estimate for two variables X and Y on day $n - 1$ is 0.7. In addition, the estimated standard deviations on day $n - 1$ for variables X and Y are 1.5% and 2%, respectively. Also, the percentage change on day $n - 1$ for variables X and Y are 2% and 1%, respectively. What is the updated estimate of the covariance rate and correlation between X and Y on day n?

Answer:

The estimated covariance rate between variables X and Y on day $n - 1$ can be calculated as:

$$\text{cov}(X,Y) = \rho_{X,Y} \times \sigma_X \sigma_Y = 0.7 \times 0.015 \times 0.02 = 0.00021$$

With this value, the EWMA model can update the covariance rate for day n.

$$\text{cov}_n = 0.9 \times 0.00021 + 0.1 \times 0.02 \times 0.01 = 0.000189 + 0.00002 = 0.000209$$

Note that the covariance of an asset with itself is equal to the variance of the asset ($\text{cov}(X,X) = \sigma_X^2$). Thus, the EWMA equation can also be used to estimate the new variances for variables X and Y. The modified equation for updating the variance of X becomes:

$$\sigma_{X,n}^2 = \lambda \sigma_{X,n-1}^2 + (1-\lambda)X_{n-1}^2$$

$$\sigma_{X,n}^2 = 0.9 \times 0.015^2 + 0.1 \times 0.02^2 = 0.0002025 + 0.00004 = 0.0002425$$

Similarly, the updated variance for variable Y is calculated as follows:

$$\sigma_{Y,n}^2 = 0.9 \times 0.02^2 + 0.1 \times 0.01^2 = 0.00036 + 0.00001 = 0.00037$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

The new standard deviation estimates for X and Y are found by taking the square root of their respective variances. The new volatility measure of X is:

$$\sigma_{X,n} = \sqrt{0.0002425} = 0.0155724$$

The new volatility measure of Y is:

$$\sigma_{Y,n} = \sqrt{0.00037} = 0.0192354$$

Therefore, the new correlation on day n can be found by dividing the updated covariance (cov_n) by the updated standard deviations for X and Y:

$$\frac{0.000209}{0.0155724 \times 0.0192354} = 0.6977$$

GARCH(1,1) Model

An alternative method for updating the covariance rate for two variables X and Y uses the **generalized autoregressive conditional heteroskedasticity (GARCH) model.** The GARCH(1,1) model for updating covariance rates is defined as follows:

$$cov_n = \omega + \alpha X_{n-1} Y_{n-1} + \beta cov_{n-1}$$

GARCH(1,1) applies a weight of α to the most recent observation on covariance ($X_{n-1}Y_{n-1}$) and a weight of β to the most recent covariance estimate (cov_{n-1}). In addition, a weight of ω is given to the long-term average covariance rate.

Professor's Note: The EWMA is a special case of GARCH(1,1), where $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$.

An alternative form for writing the GARCH(1,1) model is shown as follows:

$$cov_n = \gamma V_L + \alpha X_{n-1} Y_{n-1} + \beta cov_{n-1}$$

where:
γ = weight assigned to the long-term variance, V_L

In this equation, the three weights must equal 100% ($\gamma + \alpha + \beta = 1$). If α and β are known, the weight for the long-term variance, γ, can be determined as $1 - \alpha - \beta$. Therefore, the long-term average covariance rate must equal: $\omega / (1 - \alpha - \beta)$.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Example: Calculating covariance using the GARCH(1,1) model

Assume an analyst uses daily data to estimate a GARCH(1,1) model as follows:

$$\text{cov}_n = 0.000002 + 0.14X_{n-1}Y_{n-1} + 0.76\text{cov}_{n-1}$$

This implies $\alpha = 0.14$, $\beta = 0.76$, and $\omega = 0.000002$. The analyst also determines that the estimate of covariance on day $n - 1$ is 0.000324 and the most recent returns on X and Y are both 0.02. What is the updated estimate of covariance?

Answer:

The updated estimate of covariance on day n is 0.0304%, which is calculated as:

$$\begin{aligned}\text{cov}_n &= 0.000002 + \left(0.14 \times 0.02^2\right) + \left(0.76 \times 0.000324\right) \\ &= 0.000002 + 0.000056 + 0.000246 = 0.000304\end{aligned}$$

Evaluating Consistency for Covariances

LO 20.3: Apply the consistency condition to covariance.

A **variance-covariance matrix** can be constructed using the calculated estimates of variance and covariance rates for a set of variables. The diagonal of the matrix represents the variance rates where $i = j$. The covariance rates are all other elements of the matrix where $i \neq j$.

A matrix is known as *positive-semidefinite* if it is internally consistent. The following expression defines the necessary condition for an N × N variance-covariance matrix, Ω, to be internally consistent for all N × 1 vectors ω, where ω^T is the transpose of vector ω:

$$\omega^T \Omega \omega \geq 0$$

Variance and covariance rates are calculated using the same EWMA or GARCH model parameters to ensure that a positive-semidefinite model is constructed. For example, if a EWMA model uses $\lambda = 0.95$ for estimating variances, the same EWMA and λ should be used to estimate covariance rates.

When small changes are made to a small positive-semidefinite matrix such as a 3 × 3 matrix, the matrix will most likely remain positive-semidefinite. However, small changes to a large positive-semidefinite matrix such as 1,000 × 1,000 will most likely cause the matrix to no longer be positive-semidefinite.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

An example of a variance-covariance matrix that is not internally consistent is shown as follows:

$$\begin{pmatrix} 1 & 0 & 0.8 \\ 0 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}$$

Notice that the variances (i.e., diagonal of the matrix) are all equal to one. Therefore, the correlation for each pair of variables must equal the covariance for each pair of variables. This is true because the standard deviations are all equal to one. Thus, correlation is calculated as the covariance divided by one.

Also, notice that there is no correlation between the first and second variables. However, there is a strong correlation between the first and third variables as well as the second and third variables. This is very unusual to have one pair with no correlation while the other two pairs have high correlations. If we transpose a vector such that $\omega^T = (1, 1, -1)$, we would find that this variance-covariance matrix is not internally consistent since $\omega^T \Omega \omega \geq 0$ is not satisfied.

Another method for testing for consistency is to evaluate the following expression:

$$\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} \leq 1$$

We can substitute data from the above variance-covariance matrix into this expression because all covariances are also correlation coefficients. When computing the formula, we would determine that the left side of the expression is actually greater than the right side, indicating that the matrix is not internally consistent.

$$0^2 + 0.8^2 + 0.8^2 - 2 \times 0 \times 0.8 \times 0.8 = 1.28$$
$$1.28 > 1$$

Generating Samples

LO 20.4: Describe the procedure of generating samples from a bivariate normal distribution.

Suppose there is a bivariate normal distribution with two variables, *X* and *Y*. Variable *X* is known and the value of variable *Y* is conditional on the value of variable *X*. If variables *X* and *Y* have a bivariate normal distribution, then the expected value of variable *Y* is normally distributed with a mean of:

$$\mu_Y + \rho_{XY} \times \sigma_Y \times \frac{X - \mu_X}{\sigma_X}$$

©2015 Kaplan, Inc.
FRM官方交流群 392417118

and a standard deviation of:

$$\sigma_Y\sqrt{1-\rho_{XY}^2}$$

The means, μ_X and μ_Y, of variables X and Y are both unconditional means. The standard deviations of variables X and Y are both unconditional standard deviations. Also note that the expected value of Y is linearly dependent on the conditional value of X.

The following procedure is used to generate two sample sets of variables from a bivariate normal distribution.

Step 1: Independent samples Z_X and Z_Y are obtained from a univariate standardized normal distribution. Microsoft Excel® and other software programming languages have routines for sampling random observations from a normal distribution. For example, this is done in Excel with the formula = NORMSINV(RAND()).

Step 2: Samples ε_X and ε_Y are then generated. The first sample of X variables is the same as the random sample from a univariate standardized normal distribution, $\varepsilon_X = Z_X$.

Step 3: The conditional sample of Y variables is determined as follows:

$$\varepsilon_Y = \rho_{XY}Z_X + Z_Y\sqrt{1-\rho_{XY}^2}$$

where:
ρ_{XY} = correlation between variables X and Y in the bivariate normal distribution

Factor Models

LO 20.5: Describe properties of correlations between normally distributed variables when using a one-factor model.

A **factor model** can be used to define correlations between normally distributed variables. The following equation is a one-factor model where each U_i has a component dependent on one common factor (F) in addition to another component (Z_i) that is uncorrelated with other variables.

$$U_i = \alpha_i F + \sqrt{1-\alpha_i^2}Z_i$$

Between normally distributed variables, one-factor models are structured as follows:

- Every U_i has a standard normal distribution (mean = 0, standard deviation = 1).
- The constant α_i is between –1 and 1.
- F and Z_i have standard normal distributions and are uncorrelated with each other.
- Every Z_i is uncorrelated with each other.
- All correlations between U_i and U_j result from their dependence on a common factor, F.

There are two major advantages of the structure of one-factor models. First, the covariance matrix for a one-factor model is positive-semidefinite. Second, the number of correlations between variables is greatly reduced. Without assuming a one-factor model, the correlations of each variable must be computed. If there are N variables, this would

FRM官方交流群 392417118

require [N × (N – 1)] / 2 calculations. However, the one-factor model only requires *N* estimates for correlations, where each of the *N* variables is correlated with one factor, *F*. The most well-known one factor model in finance is the *capital asset pricing model* (CAPM). Under the CAPM, each asset return has a systematic component (measured by beta) that is correlated with the market portfolio return. Each asset return also has a nonsystematic (or idiosyncratic) component that is independent of the return on other stocks and the market.

Copulas

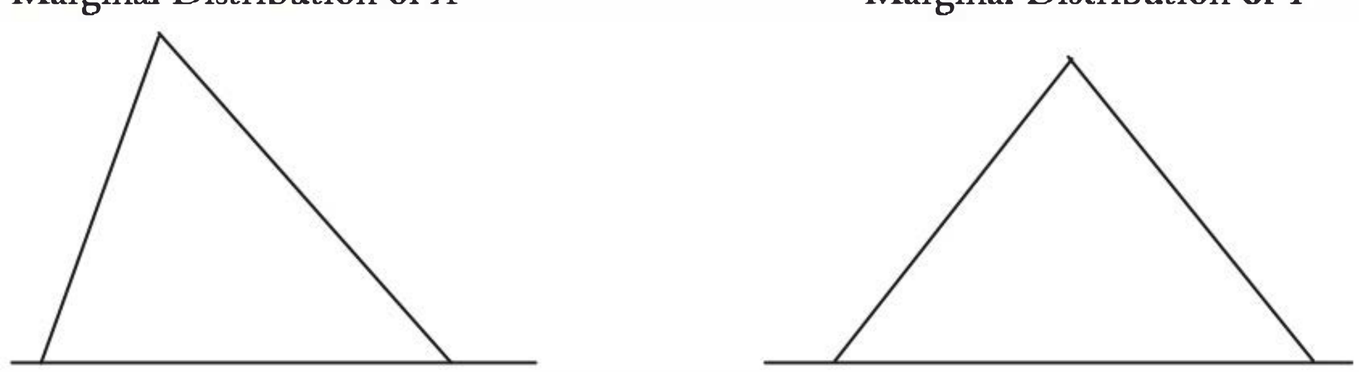
LO 20.6: Define copula and describe the key properties of copulas and copula correlation.

Suppose we have two **marginal distributions** of expected values for variables *X* and *Y*. The marginal distribution of variable *X* is its distribution with no knowledge of variable *Y*. The marginal distribution of variable *Y* is its distribution with no knowledge of variable *X*. If both distributions are normal, then we can assume the joint distribution of the variables is bivariate normal. However, if the marginal distributions are not normal, then a copula is necessary to define the correlation between these two variables.

A **copula** creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions. This is accomplished by mapping the marginal distributions to a new known distribution. For example, a Gaussian copula (discussed in LO 20.8) maps the marginal distribution of each variable to the standard normal distribution, which, by definition, has a mean of zero and a standard deviation of one. The mapping of each variable to the new distribution is done based on percentiles.

Suppose we have two triangular marginal distributions for two variables *X* and *Y* as illustrated in Figure 3.

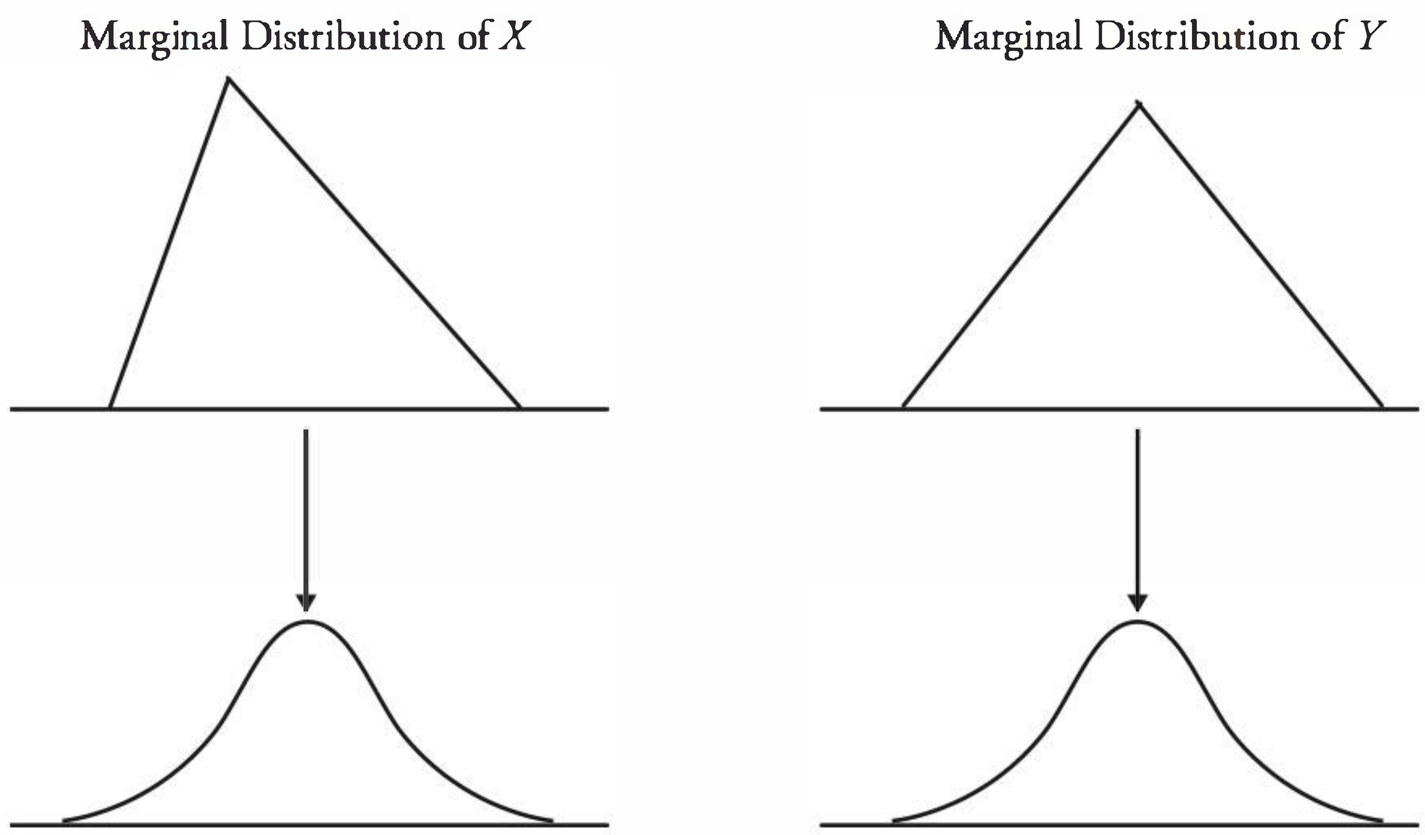
Figure 3: Marginal Distributions



These two triangular marginal distributions for *X* and *Y* are preserved by mapping them to a known joint distribution. Figure 4 illustrates how **a copula correlation** is created.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Figure 4: Mapping Variables to Standard Normal Distributions



The key property of a copula correlation model is the *preservation of the original marginal distributions while defining a correlation between them*. A correlation copula is created by converting two distributions that may be unusual or have unique shapes and mapping them to known distributions with well-defined properties, such as the normal distribution. As mentioned, this is done by mapping on a percentile-to-percentile basis.

For example, the 5th percentile observation for the variable X marginal distribution is mapped to the 5th percentile point on the U_X standard normal distribution. The 5th percentile will have a value of –1.645. This is repeated for each observation on a percentile-to-percentile basis. The value that represents the 95th percentile of the X marginal distribution will have a value mapped to the 95th percentile of the U_X standard normal distribution and will have a value of +1.645. Likewise, every observation on the variable Y distribution is mapped to the corresponding percentile on the U_Y standard normal distribution. The new distribution is now a multivariate normal distribution.

Both U_X and U_Y are now normal distributions. If we make the assumption that the two distributions are joint bivariate normal distributions, then a correlation structure can be defined between the two variables. The triangular structures are not well-behaved structures. Therefore, it is difficult to define a relationship between the two variables. However, the normal distribution is a well-behaved distribution. Therefore, using a copula is a way to indirectly define a correlation structure between two variables when it is not possible to directly define correlation.

As mentioned, the correlation between U_X and U_Y is referred to as the copula correlation. The conditional mean of U_Y is linearly dependent on U_X, and the conditional standard deviation of U_Y is constant because the two distributions are bivariate normal.

For example, suppose the correlation between U_X and U_Y is 0.5. A partial table of the joint probability distribution between variables X and Y when the values of X and Y are 0.1, 0.2, and 0.3 is illustrated in Figure 5.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Figure 5: Partial Cumulative Joint Probability Distribution

	Variable Y		
Variable X	0.1	0.2	0.3
0.1	0.006	0.017	0.028
0.2	0.013	0.043	0.081
0.3	0.017	0.061	0.124

Now assume that the variable X under the original distribution had a value of 0.1 at the 5th percentile with a corresponding U_X value of –1.645. Also assume that the variable Y under the original distribution had a value of 0.1 with a corresponding value of –2.05. The joint probability that $U_X < -1.645$ and $U_Y < -2.05$ can be determined as 0.006 based on the row and column in Figure 5 that corresponds to a 0.1 value for both variables X and Y.

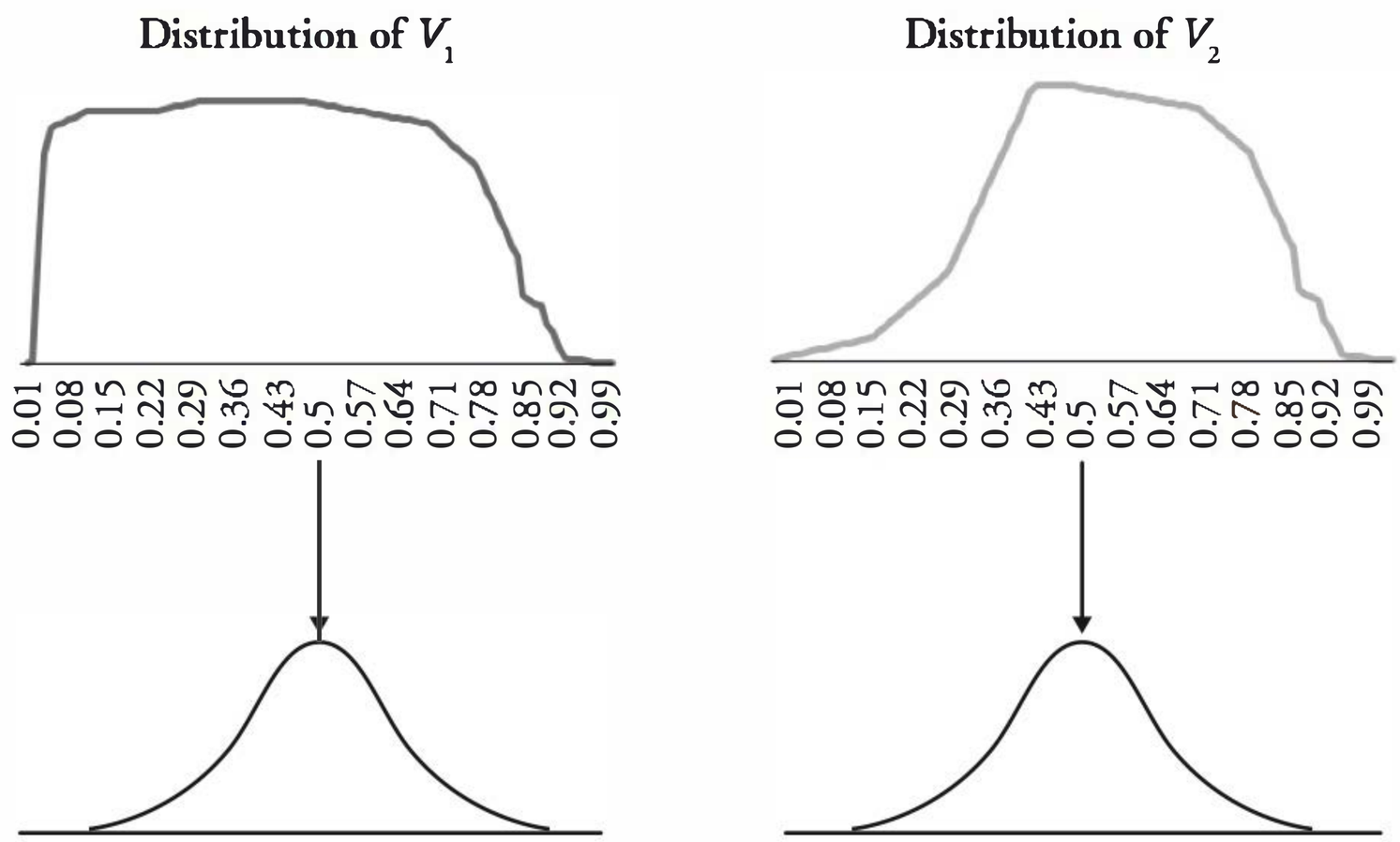
Types of Copulas

LO 20.8: Describe the Gaussian copula, Student's t-copula, multivariate copula, and one factor copula.

A **Gaussian copula** maps the marginal distribution of each variable to the standard normal distribution. The mapping of each variable to the new distribution is done based on percentiles. Figure 6 illustrates that V_1 and V_2 have unique marginal distributions. The observations of each distribution is mapped to the standard normal distribution on a percentile-to-percentile basis to create a Gaussian copula as follows:

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Figure 6: Mapping Gaussian Copula to Standard Normal Distribution



Other types of copulas are created by mapping to other well-known distributions. The **Student's *t*-copula** is similar to the Gaussian copula. However, variables are mapped to distributions of U_1 and U_2 that have a bivariate Student's *t*-distribution rather than a normal distribution.

The following procedure is used to create a Student's *t*-copula assuming a bivariate Student's *t*-distribution with *f* degrees of freedom and correlation ρ.

Step 1: Obtain values of χ by sampling from the inverse chi-squared distribution with *f* degrees of freedom.

Step 2: Obtain values by sampling from a bivariate normal distribution with correlation ρ.

Step 3: Multiply $\sqrt{f / \chi}$ by the normally distributed samples.

A **multivariate copula** is used to define a correlation structure for more than two variables. Suppose the marginal distributions are known for *N* variables: $V_1, V_2, \ldots, V_N$. Distribution V_i for each *i* variable is mapped to a standard normal distribution, U_i. Thus, the correlation structure for all variables is now based on a multivariate normal distribution.

Factor copula models are often used to define the correlation structure in multivariate copula models. The nature of the dependence between the variables is impacted by the choice of the U_i distribution. The following equation defines a **one-factor copula** model where *F* and Z_i are standard normal distributions:

$$U_i = \alpha_i F + \sqrt{1-\alpha_i^2}\, Z_i$$

The U_i distribution has a multivariate Student's *t*-distribution if Z_i and *F* are assumed to have a normal distribution and a Student's *t*-distribution, respectively. The choice of U_i

©2015 Kaplan, Inc.

FRM官方交流群 392417118

determines the dependency of the U variables, which also defines the covariance copula for the V variables.

A practical example of how a one-factor copula model is used is in calculating the value at risk (VaR) for loan portfolios. A risk manager assumes a one-factor copula model maps the default probability distributions for different loans. The percentiles of the one-factor distribution are then used to determine the number of defaults for a large portfolio.

Tail Dependence

LO 20.7: Explain tail dependence.

There is greater **tail dependence** in a bivariate Student's t-distribution than a bivariate normal distribution. In other words, it is more common for two variables to have the same tail values at the same time using the bivariate Student's t-distribution. During a financial crisis or some other extreme market condition, it is common for assets to be highly correlated and exhibit large losses at the same time. This suggests that the Student's t-copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in the distribution tails at the same time.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Key Concepts

LO 20.1

Correlation and covariance measure the strength between the linear relationship of two variables as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

A correlation of zero between two variables does not imply that there is no dependence between the two variables.

LO 20.2

The formula for calculating a new covariance on day *n* using an exponentially weighted moving average (EWMA) model is:

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1-\lambda) X_{n-1} Y_{n-1}$$

GARCH(1,1) applies a weight of α to the most recent observation on covariance ($X_{n-1}Y_{n-1}$), a weight of β to the most recent covariance estimate (cov_{n-1}), and a weight of ω to the long-term average covariance rate as follows:

$$\text{cov}_n = \omega + \alpha X_{n-1} Y_{n-1} + \beta \text{cov}_{n-1}$$

LO 20.3

A matrix is positive-semidefinite if it is internally consistent. The following expression defines the necessary condition for an N × N variance-covariance matrix, Ω, to be internally consistent for all N × 1 vectors ω, where ω^T is the transpose of vector ω:

$$\omega^T \Omega \omega \geq 0$$

LO 20.4

Independent samples of two variables Z_X and Z_Y can be generated from a univariate standardized normal distribution. The conditional sample of *Y* variables for a bivariate normal distribution is then generated as:

$$\varepsilon_Y = \rho_{XY} Z_X + Z_Y \sqrt{1 - \rho_{XY}^2}$$

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

LO 20.5

The covariance matrix for a one-factor model is positive-semidefinite. Also, the one-factor model only requires N estimates for correlations, where each of the N variables is correlated with one factor, F.

LO 20.6

A copula creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions.

LO 20.7

The Student's t-copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in tails at the same time.

LO 20.8

A Gaussian copula maps the marginal distribution of each variable to the standard normal distribution. The Student's t-copula maps variables to distributions of U_1 and U_2 that have a bivariate Student's t-distribution. The multivariate copula defines a correlation structure for three or more variables. The choice of U_i determines the dependency of the U variables in a one-factor copula model, which also defines the covariance copula for the V variables.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checkers

1. Suppose an analyst uses the EWMA model with $\lambda = 0.95$ to update correlation and covariance rates. The observed percentage change on day $n - 1$ for variables X and Y are 2.0% and 1.0%, respectively. The correlation estimate based on historical data for two variables X and Y on day $n - 1$ is 0.52. In addition, the estimated standard deviations on day $n - 1$ for variables X and Y are 1.4% and 1.8%, respectively. What is the new estimate of the correlation between X and Y on day n?
   A. 0.14.
   B. 0.42.
   C. 0.53.
   D. 0.68.

2. An equity analyst is concerned about satisfying the consistency condition for estimating new covariance rates. Which of the following procedures will most likely result in a positive-semidefinite matrix?
   A. The analyst uses an EWMA model with $\lambda = 0.95$ to update variances and a GARCH(1,1) model with $\lambda = 0.95$ to update the covariance rates for a 1,000 × 1,000 variance-covariance matrix.
   B. The analyst uses an EWMA model with $\lambda = 0.90$ to update variances and an EWMA model with $\lambda = 0.90$ to update the covariance rates for a 3 × 3 variance-covariance matrix.
   C. The analyst uses a GARCH(1,1) model with $\lambda = 0.95$ to update variances and a GARCH(1,1) model with $\lambda = 0.90$ to update the covariance rates for a 1,000 × 1,000 variance-covariance matrix.
   D. The analyst uses an EWMA model with $\lambda = 0.90$ to update variances and a GARCH(1,1) model with $\lambda = 0.90$ to update the covariance rates for a 3 × 3 variance-covariance matrix.

3. Suppose two samples, Z_X and Z_Y, are generated from a bivariate normal distribution. If variable Y is conditional on variable X, which of the following statements regarding these two samples is incorrect?
   A. The expected value of Y has a nonlinear relationship with all values of X.
   B. The mean and standard deviations for sample Z_X are unconditional.
   C. The value of variable Y is normally distributed.
   D. The conditional sample of Y variables is determined by:

$$\varepsilon_Y = \rho_{XY} Z_X + Z_Y \sqrt{1 - \rho_{XY}^2}\ .$$

FRM官方交流群 392417118

4. Which of the following statements is most reflective of a characteristic of one-factor models between multivariate normally distributed variables? The one-factor model is shown as follows:

$$U_i = \alpha_i F + \sqrt{1-\alpha_i^2} Z_i$$

A. Each U_i has a component dependent on one common factor (F) in addition to another component (Z_i) that is uncorrelated with other variables.
B. F and Z_i must both have Student's t-distributions.
C. The covariance matrix for a one-factor model is not positive-semidefinite.
D. The number of calculations for estimating correlations is equal to $[N \times (N - 1)] / 2$.

5. Suppose a risk manager wishes to create a correlation copula to estimate the risk of loan defaults during a financial crisis. Which type of copula will most accurately measure tail risk?
A. Gaussian copula.
B. Student's t-copula.
C. Gaussian one-factor copula.
D. Standard normal copula.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checker Answers

1. C First, calculate the estimated covariance rate between variables *X* and *Y* on day $n-1$ as:

$$\text{cov}(X,Y) = \rho_{X,Y} \times \sigma_X \sigma_Y = 0.52 \times 0.014 \times 0.018 = 0.00013$$

The EWMA model is then used to update the covariance rate for day *n*:

$$\text{cov}_n = 0.95 \times 0.00013 + 0.05 \times 0.02 \times 0.01 = 0.0001235 + 0.00001 = 0.0001335$$

The updated variance of *X* is:

$$\sigma^2_{X,n} = 0.95 \times 0.014^2 + 0.05 \times 0.02^2 = 0.0001862 + 0.00002 = 0.0002062$$

The new volatility measure of *X* is then:

$$\sigma^2_{X,n} = \sqrt{0.0002062} = 0.0143597$$

The updated variance for variable *Y* is:

$$\sigma^2_{Y,n} = 0.95 \times 0.018^2 + 0.05 \times 0.01^2 = 0.0003078 + 0.000005 = 0.0003128$$

The new volatility measure of *Y* is then:

$$\sigma^2_{Y,n} = \sqrt{0.0003128} = 0.01768615$$

The new correlation is found by dividing the new cov_n by the new standard deviations for *X* and *Y* as follows:

$$\frac{0.0001335}{0.0143597 \times 0.0176862} = 0.5257$$

2. B A matrix is positive-semidefinite if it is internally consistent. Variance and covariance rates must be calculated using the same EWMA or GARCH model and parameters to ensure that a positive-semidefinite model is constructed. For example, if an EWMA model is used with $\lambda = 0.90$ for estimating variances, the same EWMA model and λ should be used to estimate covariance rates.

3. A Both samples are normally distributed. The expected value of variable *Y* is normally distributed with a mean of:

$$\mu_Y + \rho_{XY} \times \sigma_Y \times \frac{X - \mu_X}{\sigma_X}$$

and a standard deviation of:

$$\sigma_Y \sqrt{1 - \rho^2_{XY}}$$

The expected value of *Y* is therefore linearly dependent on the conditional value of *X*.

4. A Each U_i has a component dependent on one common factor (*F*) in addition to another component (Z_i) that is uncorrelated with other variables. *F* and Z_i have standard normal distributions and are uncorrelated with each other. The covariance matrix for a one-factor model is positive-semidefinite and the one-factor model only requires *N* estimates for correlations, where each of the *N* variables is correlated with one factor, *F*.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

5. **B** There is greater *tail dependence* in a bivariate Student's t-distribution than a bivariate normal distribution. This suggests that the Student's t-copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in tails at the same time.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 21

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 21.1: 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.

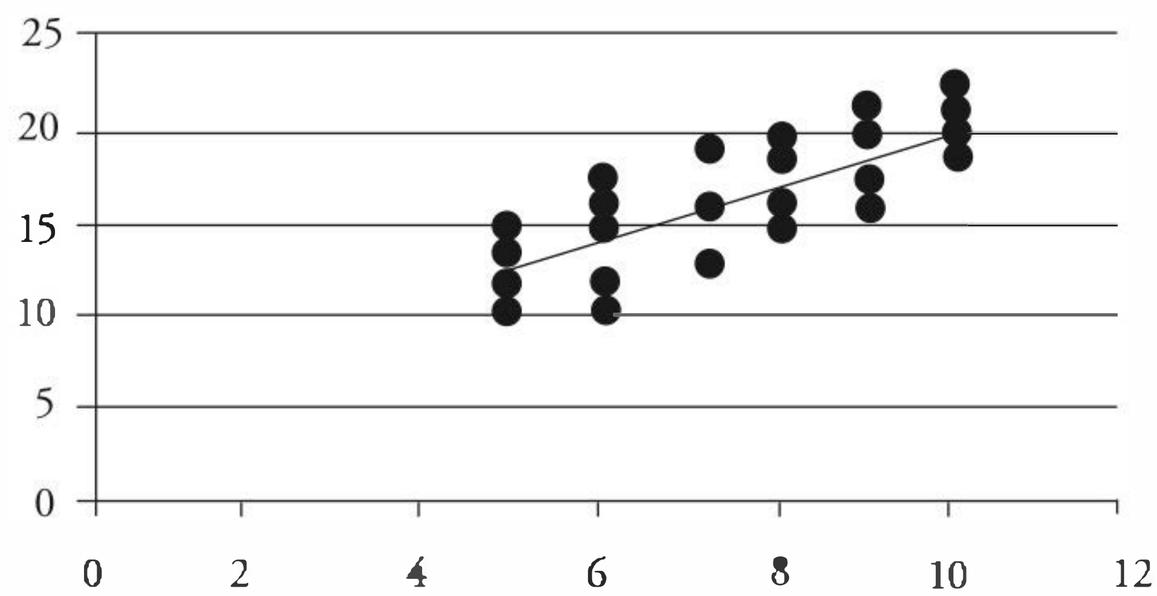
©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 21.2: 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:

$$E(\text{return} \mid \text{lockup period}) = B_0 + B_1 \times (\text{lockup period})$$

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 21.3: 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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 21.4: 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Ordinary Least Squares Regression

LO 21.5: 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:

$$\text{minimize } \Sigma e_i^2 = \Sigma[Y_i - (b_0 + b_1 \times X_i)]^2$$

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, $\overline{X}, \overline{Y}$).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Assumptions Underlying Linear Regression

LO 21.6: 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: $Var(\varepsilon_i|X_i) = \sigma^2$.
- No serial correlation of the error terms exists [i.e., $Corr(\varepsilon_i, \varepsilon_{i+j}) = 0$ 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 21.7: 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 21.8: 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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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 21.9: 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 21.10: 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

$$\text{Total sum of squares} = \text{explained sum of squares} + \text{sum of squared residuals}$$

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y_i - \hat{Y})^2$$

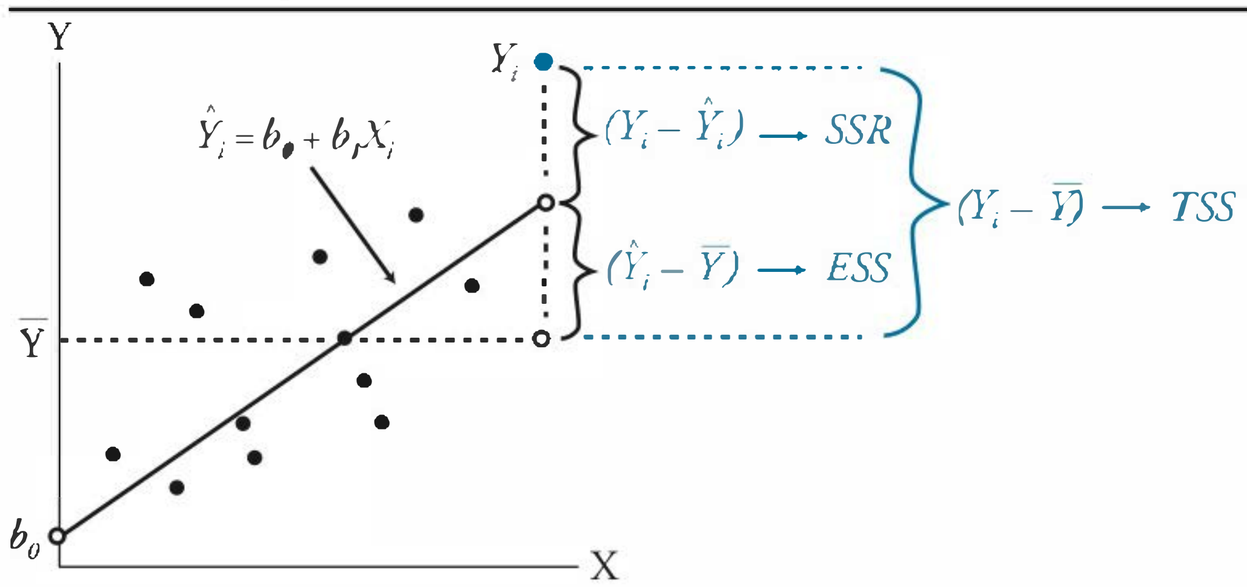
$$\text{TSS} = \text{ESS} + \text{SSR}$$

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Figure 4: Components of the Total Variation



The coefficient of determination can be calculated as follows:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

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., $\bar{Y}$ = 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	$\sum(Y_i - \bar{Y})^2$	$\hat{Y}_i$	$\sum(Y_i - \hat{Y}_i)^2$
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:

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{20}{90} = 0.778$$

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Key Concepts

LO 21.1

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 21.2

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 21.3

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 21.4

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 21.5

Ordinary least squares estimation is a process that estimates the population parameters B_i with corresponding values for b_i that minimize $\Sigma e_i^2 = \Sigma[Y_i - (b_0 + b_1 \times X_i)]^2$. The formulas for the coefficients are:

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{Cov(X,Y)}{Var(X)}$$

$$b_0 = \overline{Y} - b_1\overline{X}$$

LO 21.6

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

LO 21.7

OLS estimators are used widely in practice. In addition to practical benefits, OLS estimators exhibit desirable properties of an estimator.

LO 21.8

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 21.9

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 21.10

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 22

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 22.1: 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 ± 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 22.3: 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:

$$\text{Reject } H_0 \text{ if } t > +t_{critical} \text{ or } t < -t_{critical}$$

Rejection of the null means that the slope coefficient is *different* from the hypothesized value of B_1.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

$$H_0: B_1 = 0 \text{ versus } H_A: B_1 \neq 0$$

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 22.2: 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.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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%.

FRM官方交流群 392417118

Dummy Variables

Observations for most independent variables (e.g., firm size, level of GDP, and interest rates) can take on a wide range of values. However, there are occasions when the independent variable is binary in nature—it is either "on" or "off." Independent variables that fall into this category are called **dummy variables** and are often used to quantify the impact of qualitative events.

Dummy variables are assigned a value of "0" or "1." For example, in a time series regression of monthly stock returns, you could employ a "January" dummy variable that would take on the value of "1" if a stock return occurred in January and "0" if it occurred in any other month. The purpose of including the January dummy variable would be to see if stock returns in January were significantly different than stock returns in all other months of the year. Many "January Effect" anomaly studies employ this type of regression methodology.

The estimated regression coefficient for dummy variables indicates the difference in the dependent variable for the category represented by the dummy variable and the average value of the dependent variable for all classes except the dummy variable class. For example, testing the slope coefficient for the January dummy variable would indicate whether, and by how much, security returns are different in January as compared to the other months.

What is Heteroskedasticity?

LO 22.4: Evaluate the implications of homoskedasticity and heteroskedasticity.

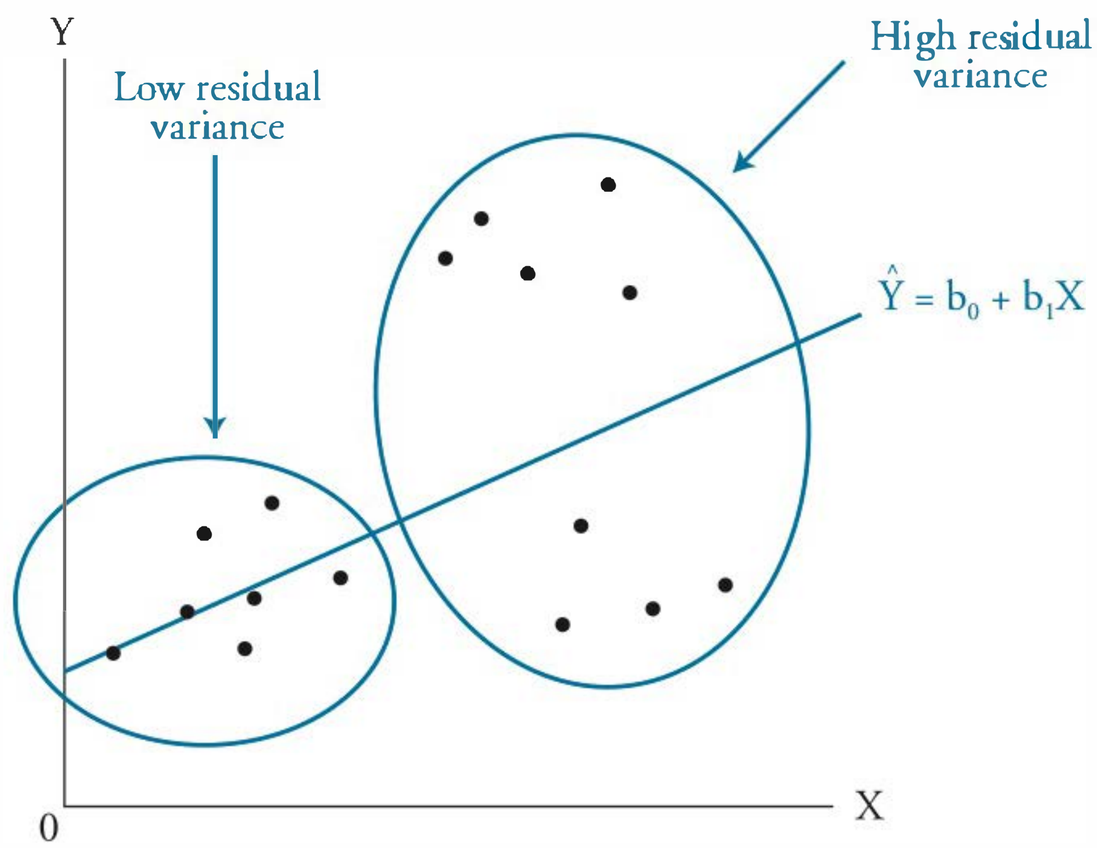
If the variance of the residuals is constant across all observations in the sample, the regression is said to be **homoskedastic.** When the opposite is true, the regression exhibits **heteroskedasticity**, which occurs when the variance of the residuals is not the same across all observations in the sample. This happens when there are subsamples that are more spread out than the rest of the sample.

Unconditional heteroskedasticity occurs when the heteroskedasticity is not related to the level of the independent variables, which means that it doesn't systematically increase or decrease with changes in the value of the independent variable(s). While this is a violation of the equal variance assumption, *it usually causes no major problems with the regression.*

Conditional heteroskedasticity is heteroskedasticity that is related to the level of (i.e., conditional on) the independent variable. For example, conditional heteroskedasticity exists if the variance of the residual term increases as the value of the independent variable increases, as shown in Figure 1. Notice in this figure that the residual variance associated with the larger values of the independent variable, X, is larger than the residual variance associated with the smaller values of X. Conditional heteroskedasticity *does create significant problems for statistical inference.*

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Figure 1: Conditional Heteroskedasticity



Effect of Heteroskedasticity on Regression Analysis

There are several effects of heteroskedasticity you need to be aware of:

- The standard errors are usually unreliable estimates.
- The coefficient estimates (the b_j) aren't affected.
- If the standard errors are too small, but the coefficient estimates themselves are not affected, the *t*-statistics will be too large and the null hypothesis of no statistical significance is rejected too often. The opposite will be true if the standard errors are too large.

Detecting Heteroskedasticity

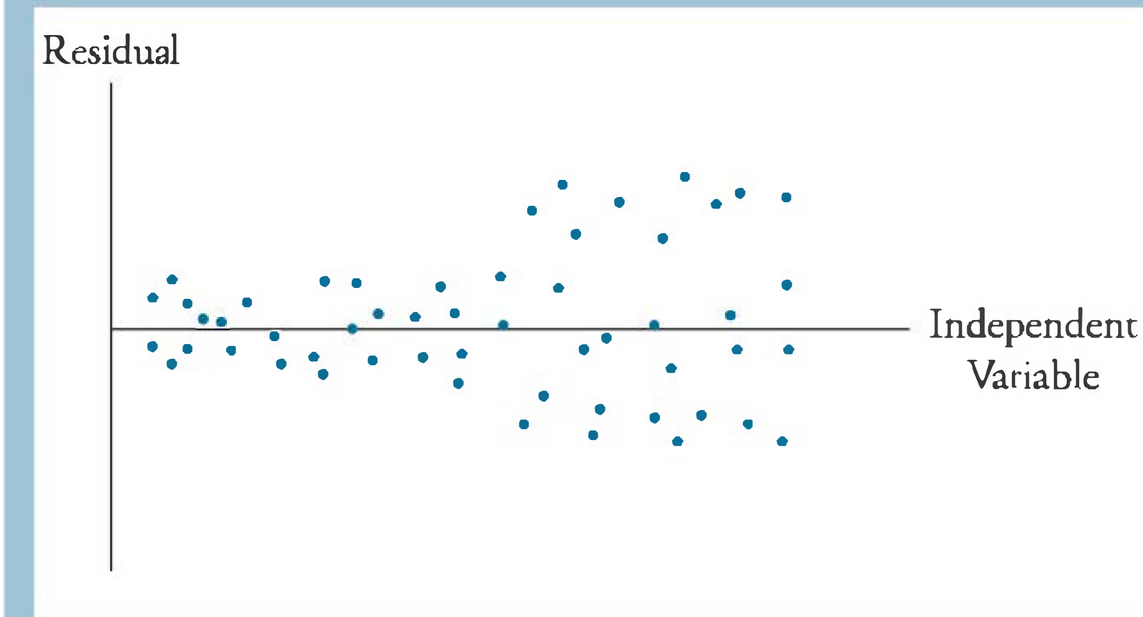
As was shown in Figure 1, a scatter plot of the residuals versus one of the independent variables can reveal patterns among observations.

FRM官方交流群 392417118

Example: Detecting heteroskedasticity with a residual plot

You have been studying the monthly returns of a mutual fund over the past five years, hoping to draw conclusions about the fund's average performance. You calculate the mean return, the standard deviation, and the portfolio's beta by regressing the fund's returns on S&P 500 index returns (the independent variable). The standard deviation of returns and the fund's beta don't seem to fit the firm's stated risk profile. For your analysis, you have prepared a scatter plot of the error terms (actual return – predicted return) for the regression using five years of returns, as shown in the following figure. Determine whether the residual plot indicates that there may be a problem with the data.

Residual Plot



Answer:

The residual plot in the previous figure indicates the presence of conditional heteroskedasticity. Notice how the variation in the regression residuals increases as the independent variable increases. This indicates that the variance of the fund's returns about the mean is related to the level of the independent variable.

Correcting Heteroskedasticity

Heteroskedasticity is not easy to correct, and the details of the available techniques are beyond the scope of the FRM curriculum. The most common remedy, however, is to calculate **robust standard errors**. These robust standard errors are used to recalculate the t-statistics using the original regression coefficients. On the exam, use robust standard errors to calculate t-statistics if there is evidence of heteroskedasticity. By default, many statistical software packages apply *homoskedastic* standard errors unless the user specifies otherwise.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

The Gauss-Markov Theorem

LO 22.5: Determine the conditions under which the OLS is the best linear conditionally unbiased estimator.

LO 22.6: Explain the Gauss-Markov Theorem and its limitations, and alternatives to the OLS.

The **Gauss-Markov theorem** says that if the linear regression model assumptions are true and the regression errors display homoskedasticity, then the OLS estimators have the following properties.

1. The OLS estimated coefficients have the minimum variance compared to other methods of estimating the coefficients (i.e., they are the most precise).

2. The OLS estimated coefficients are based on linear functions.

3. The OLS estimated coefficients are unbiased, which means that in repeated sampling the averages of the coefficients from the sample will be distributed around the true population parameters [i.e., $E(b_0) = B_0$ and $E(b_1) = B_1$].

4. The OLS estimate of the variance of the errors is unbiased [i.e., $E(\hat{\sigma}^2) = \sigma^2$].

The acronym for these properties is "BLUE," which indicates that OLS estimators are the best linear unbiased estimators.

One limitation of the Gauss-Markov theorem is that its conditions may not hold in practice, particularly when the error terms are heteroskedastic, which is sometimes observed in economic data. Another limitation is that alternative estimators, which are not linear or unbiased, may be more efficient than OLS estimators. Examples of these alternative estimators include: the weighted least squares estimator (which can produce an estimator with a smaller variance—to combat heteroskedastic errors) and the least absolute deviations estimator (which is less sensitive to extreme outliers given that rare outliers exist in the data).

Small Sample Sizes

LO 22.7: Apply and interpret the t-statistic when the sample size is small.

The central limit theorem is important when analyzing OLS results because it allows for the use of the *t*-distribution when conducting hypothesis testing on regression coefficients. This is possible because the central limit theorem says that the means of individual samples will be normally distributed when the sample size is large. However, if the sample size is small, the distribution of a *t*-statistic becomes more complicated to interpret.

In order to analyze a regression coefficient *t*-statistic when the sample size is small, we must assume the assumptions underlying linear regression hold. In particular, in order to apply and interpret the *t*-statistic, error terms must be homoskedastic (i.e., constant variance

©2015 Kaplan, Inc.

FRM官方交流群 392417118

of error terms) and the error terms must be normally distributed. If this is the case, the *t*-statistic can be computed using the default standard error (i.e., the homoskedasticity-only standard error), and it follows a *t*-distribution with n – 2 degrees of freedom.

In practice, it is rare to assume that error terms have a constant variance and are normally distributed. However, it is generally the case that sample sizes are large enough to apply the central limit theorem meaning that we can calculate *t*-statistics using homoskedasticity-only standard errors. In other words, with a large sample size, differences between the *t*-distribution and the standard normal distribution can be ignored.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 22.1

The confidence interval for the regression coefficient, B_1, is calculated as:

$$b_1 \pm (t_c \times s_{b_1}), \text{ or } \left[b_1 - (t_c \times s_{b_1}) < B_1 < b_1 + (t_c \times s_{b_1})\right]$$

LO 22.2

The *p*-value is the smallest level of significance for which the null hypothesis can be rejected. Interpreting the *p*-value offers an alternative approach when testing for statistical significance.

LO 22.3

A *t*-test with n – 2 degrees of freedom is used to conduct hypothesis tests of the estimated regression parameters:

$$t = \frac{b_1 - B_1}{s_{b_1}}$$

A predicted value of the dependent variable, $\hat{Y}$, is determined by inserting the predicted value of the independent variable, X_p, in the regression equation and calculating $\hat{Y}_p = b_0 + b_1 X_p$.

The confidence interval for a predicted *Y*-value is $\left[\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)\right]$, where s_f is the standard error of the forecast.

Qualitative independent variables (dummy variables) capture the effect of a binary independent variable:

- Slope coefficient is interpreted as the change in the dependent variable for the case when the dummy variable is one.
- Use one less dummy variable than the number of categories.

LO 22.4

Homoskedasticity refers to the condition of constant variance of the residuals. Heteroskedasticity refers to a violation of this assumption.

The effects of heteroskedasticity are as follows:

- The standard errors are usually unreliable estimates.
- The coefficient estimates (the b_j) aren't affected.
- If the standard errors are too small, but the coefficient estimates themselves are not affected, the *t*-statistics will be too large and the null hypothesis of no statistical significance is rejected too often. The opposite will be true if the standard errors are too large.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

LO 22.5

The Gauss-Markov theorem says that if linear regression assumptions are true, then OLS estimators are the best linear unbiased estimators.

LO 22.6

The limitations of the Gauss-Markov theorem are that its conditions may not hold in practice and alternative estimators may be more efficient. Examples of alternative estimators include the weighted least squares estimator and the least absolute deviations estimator.

LO 22.7

In order to interpret *t*-statistics of regression coefficients when a sample size is small, we must assume the assumptions underlying linear regression hold. In practice, it is generally the case that sample sizes are large, meaning that *t*-statistics can be computed using homoskedasticity-only standard errors.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checkers

1. What is the appropriate alternative hypothesis to test the statistical significance of the intercept term in the following regression?

$Y = a_1 + a_2(X) + \varepsilon$

A. $H_A: a_1 \neq 0$.
B. $H_A: a_1 > 0$.
C. $H_A: a_2 \neq 0$.
D. $H_A: a_2 > 0$.

Use the following information for Questions 2 through 4.

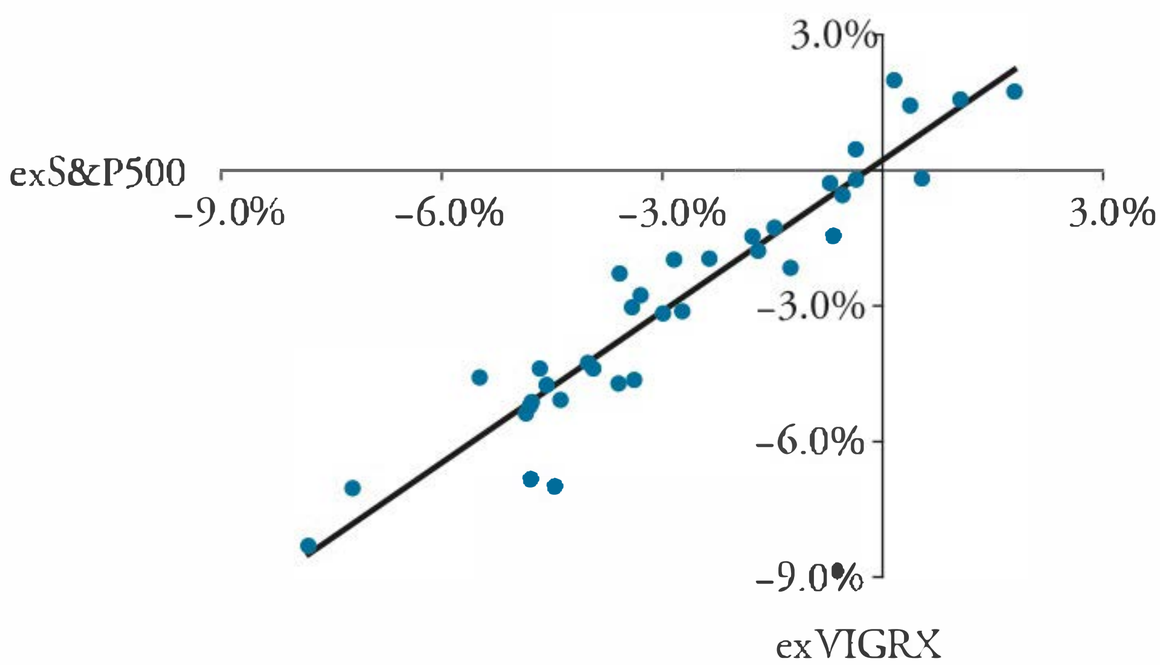
Bill Coldplay is analyzing the performance of the Vanguard Growth Index Fund (VIGRX) over the past three years. The fund employs a passive management investment approach designed to track the performance of the MSCI US Prime Market Growth index, a broadly diversified index of growth stocks of large U.S. companies.

Coldplay estimates a regression using excess monthly returns on VIGRX (exVIGRX) as the dependent variable and excess monthly returns on the S&P 500 index (exS&P) as the independent variable. The data are expressed in decimal terms (e.g., 0.03, not 3%).

$$exVIGRX_t = b_0 + b_1(exS\&P_t) + \varepsilon_t$$

A scatter plot of excess returns for both return series from June 2004 to May 2007 are shown in the following figure.

Analysis of Large Cap Growth Fund



FRM官方交流群 392417118

Results from that analysis are presented in the following figures.

Coefficient	*Coefficient Estimate*	*Standard Error*
b_0	0.0023	0.0022
b_1	1.1163	0.0624

Source of Variation	*Sum of Squares*
Explained	0.0228
Residual	0.0024

2. The 90% confidence interval for b_0 is closest to:
   A. –0.0014 to +0.0060.
   B. –0.0006 to +0.0052.
   C. +0.0001 to +0.0045.
   D. –0.0006 to +0.0045.

3. Are the intercept term and the slope coefficient statistically significantly different from zero at the 5% significance level?

	Intercept term significant?	Slope coefficient significant?
A.	Yes	Yes
B.	Yes	No
C.	No	Yes
D.	No	No

4. Coldplay would like to test the following hypothesis: H_0: $B_1 \leq 1$ vs. H_A: $B_1 > 1$ at the 1% significance level. The calculated *t*-statistic and the appropriate conclusion are:

	Calculated *t*-statistic	Appropriate conclusion
A.	1.86	Reject H_0
B.	1.86	Fail to reject H_0
C.	2.44	Reject H_0
D.	2.44	Fail to reject H_0

5. Consider the following statement: In a simple linear regression, the appropriate degrees of freedom for the critical *t*-value used to calculate a confidence interval around both a parameter estimate and a predicted Y-value is the same as the number of observations minus two. The statement is:
   A. justified.
   B. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a parameter estimate is the number of observations.
   C. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a predicted Y-value is the number of observations.
   D. not justified, because the appropriate degrees of freedom used to calculate a confidence interval depends on the explained sum of squares.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checker Answers

1. A In this regression, a_1 is the intercept term. To test the statistical significance means to test the null hypothesis that a_1 is equal to zero versus the alternative that it is not equal to zero.

2. A Note that there are 36 monthly observations from June 2004 to May 2007, so n = 36. The critical two-tailed 10% *t*-value with 34 (n – 2 = 36 – 2 = 34) degrees of freedom is approximately 1.69. Therefore, the 90% confidence interval for b_0 (the intercept term) is 0.0023 +/– (0.0022)(1.69), or –0.0014 to +0.0060.

3. C The critical two-tailed 5% *t*-value with 34 degrees of freedom is approximately 2.03. The calculated *t*-statistics for the intercept term and slope coefficient are, respectively, 0.0023 / 0.0022 = 1.05 and 1.1163 / 0.0624 = 17.9. Therefore, the intercept term is not statistically different from zero at the 5% significance level, while the slope coefficient is.

4. B Notice that this is a one-tailed test. The critical one-tailed 1% *t*-value with 34 degrees of freedom is approximately 2.44. The calculated *t*-statistic for the slope coefficient is (1.1163 – 1) / 0.0624 = 1.86. Therefore, the slope coefficient is not statistically different from one at the 1% significance level, and Coldplay should fail to reject the null hypothesis.

5. A In simple linear regression, the appropriate degrees of freedom for both confidence intervals is the number of observations in the sample (*n*) minus two.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 Multiple Regressors

Topic 23

Exam Focus

Multiple regression is, in many ways, simply an extension of regression with a single regressor. The coefficient of determination, t-statistics, and standard errors of the coefficients are interpreted in the same fashion. There are some differences, however; namely that the formulas for the coefficients and standard errors are more complicated. The slope coefficients are called partial slope coefficients because they measure the effect of changing one independent variable, assuming the others are held constant. For the exam, understand the implications of omitting relevant independent variables from the model, the adjustment to the coefficient of determination when adding additional variables, and the effect that heteroskedasticity and multicollinearity have on regression results.

Omitted Variable Bias

LO 23.1: Define and interpret omitted variable bias, and describe the methods for addressing this bias.

Omitting relevant factors from an ordinary least squares (OLS) regression can produce misleading or biased results. **Omitted variable bias** is present when two conditions are met: (1) the omitted variable is correlated with the movement of the independent variable in the model, and (2) the omitted variable is a determinant of the dependent variable. When relevant variables are absence from a linear regression model, the results will likely lead to incorrect conclusions as the OLS estimators may not accurately portray the actual data.

Omitted variable bias violates the assumptions of OLS regression when the omitted variable is in fact correlated with current independent (explanatory) variable(s). The reason for this violation is because omitted factors that partially describe the movement of the dependent variable will become part of the regression's error term since they are not properly identified within the model. If the omitted variable is correlated with the regression's slope coefficient, then the error term will also be correlated with the slope coefficient. Recall, that according to the assumptions of linear regression, the independent variable must be uncorrelated with the error term.

The issue of omitted variable bias occurs regardless of the size of the sample and will make OLS estimators inconsistent. The correlation between the omitted variable and the independent variable will determine the size of the bias (i.e., a larger correlation will lead to a larger bias) and the direction of the bias (i.e., whether the correlation is positive or negative). In addition, this bias can also have a dramatic effect on the test statistics used to determine whether the independent variables are statistically significant.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Testing for omitted variable bias would check to see if the two conditions addressed earlier are present. If a bias is found, it can be addressed by dividing data into groups and examining one factor at a time while holding other factors constant. However, in order to understand the full effects of all relevant independent variables on the dependent variable, we need to utilize multiple independent coefficients in our model. Multiple regression analysis is therefore used to eliminate omitted variable bias since it can estimate the effect of one independent variable on the dependent variable while holding all other variables constant.

Multiple Regression Basics

LO 23.2: Distinguish between single and multiple regression.

Multiple regression is regression analysis with more than one independent variable. It is used to quantify the influence of two or more independent variables on a dependent variable. For instance, simple (or univariate) linear regression explains the variation in stock returns in terms of the variation in systematic risk as measured by beta. With multiple regression, stock returns can be regressed against beta and against additional variables, such as firm size, equity, and industry classification, that might influence returns.

The general multiple linear regression model is:

$$Y_i = B_0 + B_1X_{1i} + B_2X_{2i} + \ldots + B_kX_{ki} + \varepsilon_i$$

where:

Y_i = *i*th observation of the dependent variable Y, i = 1, 2, …, n
X_j = independent variables, j = 1, 2, …, k
X_{ji} = *i*th observation of the *j*th independent variable
B_0 = intercept term
B_j = slope coefficient for each of the independent variables
ε_i = error term for the *i*th observation
n = number of observations
k = number of independent variables

LO 23.5: Describe the OLS estimator in a multiple regression.

The multiple regression methodology estimates the intercept and slope coefficients such that the sum of the squared error terms, $\sum_{i=1}^{n}\varepsilon_i^2$, is minimized. The estimators of these coefficients are known as **ordinary least squares** (OLS) **estimators**. The OLS estimators are typically found with statistical software, but can also be computed using calculus or a trial-and-error method. The result of this procedure is the following regression equation:

$$\hat{Y}_i = b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki}$$

where the lowercase b_i's indicate an estimate for the corresponding regression coefficient

The residual, e_i, is the difference between the observed value, Y_i, and the predicted value from the regression, $\hat{Y}_i$:

$$e_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki})$$

FRM官方交流群 392417118

LO 23.3: Interpret the slope coefficient in a multiple regression.

Let's illustrate multiple regression using research by Arnott and Asness (2003).[1] As part of their research, the authors test the hypothesis that future 10-year real earnings growth in the S&P 500 (EG10) can be explained by the trailing dividend payout ratio of the stocks in the index (PR) and the yield curve slope (YCS). YCS is calculated as the difference between the 10-year T-bond yield and the 3-month T-bill yield at the start of the period. All three variables are measured in percent.

Formulating the Multiple Regression Equation

The authors formulate the following regression equation using annual data (46 observations):

$$EG10 = B_0 + B_1PR + B_2YCS + \varepsilon$$

The results of this regression are shown in Figure 1.

Figure 1: Estimates for Regression of EG10 on PR and YCS

	Coefficient	*Standard Error*
Intercept	–11.6%	1.657%
PR	0.25	0.032
YCS	0.14	0.280

Interpreting the Multiple Regression Results

The interpretation of the estimated regression coefficients from a multiple regression is the same as in simple linear regression for the intercept term but significantly different for the slope coefficients:

- The **intercept term** is the value of the dependent variable when the independent variables are all equal to zero.
- Each slope coefficient is the estimated change in the dependent variable for a one-unit change in that independent variable, *holding the other independent variables constant.* That's why the slope coefficients in a multiple regression are sometimes called **partial slope coefficients.**

For example, in the real earnings growth example, we can make these interpretations:

- *Intercept term*: If the dividend payout ratio is zero and the slope of the yield curve is zero, we would expect the subsequent 10-year real earnings growth rate to be –11.6%.
- *PR coefficient*: If the payout ratio increases by 1%, we would expect the subsequent 10-year earnings growth rate to increase by 0.25%, *holding YCS constant.*
- *YCS coefficient*: If the yield curve slope increases by 1%, we would expect the subsequent 10-year earnings growth rate to increase by 0.14%, *holding PR constant.*

1. Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February): 70–87.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Let's discuss the interpretation of the multiple regression slope coefficients in more detail. Suppose we run a regression of the dependent variable *Y* on a single independent variable *X*1 and get the following result:

$$Y = 2.0 + 4.5X1$$

The appropriate interpretation of the estimated slope coefficient is that if *X*1 increases by 1 unit, we would expect *Y* to increase by 4.5 units.

Now suppose we add a second independent variable *X*2 to the regression and get the following result:

$$Y = 1.0 + 2.5X1 + 6.0X2$$

Notice that the estimated slope coefficient for *X*1 changed from 4.5 to 2.5 when we added *X*2 to the regression. We would expect this to happen most of the time when a second variable is added to the regression, unless *X*2 is uncorrelated with *X*1, because if *X*1 increases by 1 unit, then we would expect *X*2 to change as well. The multiple regression equation captures this relationship between *X*1 and *X*2 when predicting *Y*.

Now the interpretation of the estimated slope coefficient for *X*1 is that if *X*1 increases by 1 unit, we would expect *Y* to increase by 2.5 units, *holding X2 constant*.

LO 23.4: Describe homoskedasticity and heteroskedasticity in a multiple regression.

In multiple regression, homoskedasticity and heteroskedasticity are just extensions of their definitions discussed in the previous topic. **Homoskedasticity** refers to the condition that the variance of the error term is constant for all independent variables, *X*, from i = 1 to n: $\text{Var}(\varepsilon_i \mid X_i) = \sigma^2$. **Heteroskedasticity** means that the dispersion of the error terms varies over the sample. It may take the form of conditional heteroskedasticity, which says that the variance is a function of the independent variables.

Measures of Fit

LO 23.6: Calculate and interpret measures of fit in multiple regression.

The **standard error of the regression** (SER) measures the uncertainty about the accuracy of the predicted values of the dependent variable, $\hat{Y}_i = b_0 + b_1X_i$. Graphically, the relationship is stronger when the actual x,y data points lie closer to the regression line (i.e., the e_i are smaller).

Formally, SER is the standard deviation of the predicted values for the dependent variable about the regression line. Equivalently, it is the standard deviation of the error terms in the regression. SER is sometimes specified as s_e.

FRM官方交流群 392417118

Recall that regression minimizes the sum of the squared vertical distances between the predicted value and actual value for each observation (i.e., prediction errors). Also, recall that the sum of the squared prediction errors, $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$, is called the sum of squared residuals, SSR (not to be confused with SER). If the relationship between the variables in the regression is very strong (actual values are close to the line), the prediction errors, and the SSR, will be small. Thus, as shown in the following equations, the standard error of the regression is a function of the SSR:

$$\text{SER} = \sqrt{s_e^2} = \sqrt{\frac{\text{SSR}}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n}\left[Y_i - (b_0 + b_i X_i)\right]^2}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k-1}}$$

where:

n	= number of observations
k	= number of independent variables
$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$	= SSR = the sum of squared residuals
$\hat{Y}_i = b_0 + b_i X_i$	= a point on the regression line corresponding to a value of X_i. It is the expected (predicted) value of Y, given the estimated relation between X and Y.

Similar to the standard deviation for a single variable, SER measures the degree of variability of the actual Y-values relative to the estimated Y-values. The SER gauges the "fit" of the regression line. *The smaller the standard error, the better the fit.*

Coefficient of Determination, R^2

The multiple coefficient of determination, R^2, can be used to test the overall effectiveness of the entire set of independent variables in explaining the dependent variable. Its interpretation is similar to that for simple linear regression: the percentage of variation in the dependent variable that is *collectively* explained by all of the independent variables. For example, an R^2 of 0.63 indicates that the model, as a whole, explains 63% of the variation in the dependent variable.

R^2 is calculated the same way as in simple linear regression.

$$R^2 = \frac{\text{total variation} - \text{unexplained variation}}{\text{total variation}} = \frac{\text{TSS} - \text{SSR}}{\text{TSS}} = \frac{\text{explained variation}}{\text{total variation}} = \frac{\text{ESS}}{\text{TSS}}$$

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Adjusted R^2

Unfortunately, R^2 by itself *may not be a reliable measure of the explanatory power of the multiple regression model.* This is because R^2 almost always increases as independent variables are added to the model, even if the marginal contribution of the new variables is not statistically significant. Consequently, a relatively high R^2 may reflect the impact of a large set of independent variables rather than how well the set explains the dependent variable. This problem is often referred to as overestimating the regression.

To overcome the problem of overestimating the impact of additional variables on the explanatory power of a regression model, many researchers recommend adjusting R^2 for the number of independent variables. The *adjusted R^2* value is expressed as:

$$R_a^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) \times (1 - R^2)\right]$$

where:
n = number of observations
k = number of independent variables
R_a^2 = adjusted R^2

R_a^2 is less than or equal to R^2. So while adding a new independent variable to the model will increase R^2, it may either *increase or decrease* the R_a^2. If the new variable has only a small effect on R^2, the value of R_a^2 may decrease. In addition, R_a^2 may be less than zero if the R^2 is low enough.

Example: Calculating R^2 and adjusted R^2

An analyst runs a regression of monthly value-stock returns on five independent variables over 60 months. The total sum of squares for the regression is 460, and the sum of squared errors is 170. Calculate the R^2 and adjusted R^2.

Answer:

$$R^2 = \frac{460 - 170}{460} = 0.630 = 63.0\%$$

$$R_a^2 = 1 - \left[\left(\frac{60-1}{60-5-1}\right) \times (1 - 0.63)\right] = 0.596 = 59.6\%$$

The R^2 of 63% suggests that the five independent variables together explain 63% of the variation in monthly value-stock returns.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Example: Interpreting adjusted R^2

Suppose the analyst now adds four more independent variables to the regression, and the R^2 increases to 65.0%. **Identify** which model the analyst would most likely prefer.

Answer:

With nine independent variables, even though the R^2 has increased from 63% to 65%, the adjusted R^2 has decreased from 59.6% to 58.7%:

$$R_a^2 = 1 - \left[\left(\frac{60-1}{60-9-1}\right) \times (1-0.65)\right] = 0.587 = 58.7\%$$

The analyst would prefer the first model because the adjusted R^2 is higher and the model has five independent variables as opposed to nine.

Assumptions of Multiple Regression

LO 23.7: Explain the assumptions of the multiple linear regression model.

As with simple linear regression, most of the assumptions made with the multiple regression pertain to ε, the model's error term:

- A linear relationship exists between the dependent and independent variables. In other words, the model in LO 23.2 correctly describes the relationship.
- The independent variables are not random, and there is no exact linear relation between any two or more independent variables.
- The expected value of the error term, conditional on the independent variables, is zero [i.e., $E(\varepsilon | X_1, X_2, \ldots X_k) = 0$].
- The variance of the error terms is constant for all observations [i.e., $E(\varepsilon_i^2) = \sigma_\varepsilon^2$].
- The error term for one observation is not correlated with that of another observation [i.e., $E(\varepsilon_i \varepsilon_j) = 0, j \neq i$].
- The error term is normally distributed.

Multicollinearity

LO 23.8: Explain the concept of imperfect and perfect multicollinearity and their implications.

Multicollinearity refers to the condition when two or more of the independent variables, or linear combinations of the independent variables, in a multiple regression are highly correlated with each other. This condition distorts the standard error of the regression and the coefficient standard errors, leading to problems when conducting *t*-tests for statistical significance of parameters.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

The degree of correlation will determine the difference between perfect and imperfect multicollinearity. If one of the independent variables is a perfect linear combination of the other independent variables, then the model is said to exhibit **perfect multicollinearity**. In this case, it will not be possible to find the OLS estimators necessary for the regression results.

An important consideration when performing multiple regression with dummy variables is the choice of the number of dummy variables to include in the model. Whenever we want to distinguish between n classes, we must use $n - 1$ dummy variables. Otherwise, the regression assumption of no exact linear relationship between independent variables would be violated. In general, if every observation is linked to only one class, all dummy variables are included as regressors, and an intercept term exists, then the regression will exhibit perfect multicollinearity. This problem is known as the **dummy variable trap**. As mentioned, this issue can be avoided by excluding one of the dummy variables from the regression equation (i.e., $n - 1$ dummy variables). With this approach, the intercept term will represent the omitted class.

Imperfect multicollinearity arises when two or more independent variables are highly correlated, but less than perfectly correlated. When conducting regression analysis, we need to be cognizant of imperfect multicollinearity since OLS estimators will be computed, but the resulting coefficients may be improperly estimated. In general, when using the term multicollinearity, we are referring to the *imperfect case*, since this regression assumption violation requires detecting and correcting.

Effect of Multicollinearity on Regression Analysis

As a result of multicollinearity, there is a *greater probability that we will incorrectly conclude that a variable is not statistically significant* (e.g., a Type II error). Multicollinearity is likely to be present to some extent in most economic models. The issue is whether the multicollinearity has a significant effect on the regression results.

Detecting Multicollinearity

The most common way to detect multicollinearity is the situation where t-tests indicate that none of the individual coefficients is significantly different than zero, while the R^2 is high. This suggests that the variables together explain much of the variation in the dependent variable, but the individual independent variables do not. The only way this can happen is when the independent variables are highly correlated with each other, so while their common source of variation is explaining the dependent variable, the high degree of correlation also "washes out" the individual effects.

High correlation among independent variables is sometimes suggested as a sign of multicollinearity. In fact, as a general rule of thumb: If the absolute value of the sample correlation between any two independent variables in the regression is greater than 0.7, multicollinearity is a potential problem. However, this only works if there are exactly two independent variables. If there are more than two independent variables, while individual variables may not be highly correlated, linear combinations might be, leading to multicollinearity. High correlation among the independent variables suggests the possibility of multicollinearity, but low correlation among the independent variables *does not necessarily* indicate multicollinearity is *not* present.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Example: Detecting multicollinearity

Bob Watson runs a regression of mutual fund returns on average P/B, average P/E, and average market capitalization, with the following results:

Variable	*Coefficient*	*p-Value*
Average P/B	3.52	0.15
Average P/E	2.78	0.21
Market Cap	4.03	0.11
R^2	89.6%	

Determine whether or not multicollinearity is a problem in this regression.

Answer:

The R^2 is high, which suggests that the three variables as a group do an excellent job of explaining the variation in mutual fund returns. However, none of the independent variables individually is statistically significant to any reasonable degree, since the *p*-values are larger than 10%. This is a classic indication of multicollinearity.

Correcting Multicollinearity

The most common method to correct for multicollinearity is to omit one or more of the correlated independent variables. Unfortunately, it is not always an easy task to identify the variable(s) that are the source of the multicollinearity. There are statistical procedures that may help in this effort, like stepwise regression, which systematically remove variables from the regression until multicollinearity is minimized.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 23.1

Omitted variable bias is present when two conditions are met: (1) the omitted variable is correlated with the movement of the independent variable in the model, and (2) the omitted variable is a determinant of the dependent variable.

LO 23.2

The multiple regression equation specifies a dependent variable as a linear function of two or more independent variables:

$$Y_i = B_0 + B_1X_{1i} + B_2X_{2i} + \ldots + B_kX_{ki} + \varepsilon_i$$

The intercept term is the value of the dependent variable when the independent variables are equal to zero. Each slope coefficient is the estimated change in the dependent variable for a one-unit change in that independent variable, holding the other independent variables constant.

LO 23.3

In a multivariate regression, each slope coefficient is interpreted as a partial slope coefficient in that it measures the effect on the dependent variable from a change in the associated independent variable holding other things constant.

LO 23.4

Homoskedasticity means that the variance of error terms is constant for all independent variables, while heteroskedasticity means that the variance of error terms varies over the sample. Heteroskedasticity may take the form of conditional heteroskedasticity, which says that the variance is a function of the independent variables.

LO 23.5

Multiple regression estimates the intercept and slope coefficients such that the sum of the squared error terms is minimized. The estimators of these coefficients are known as ordinary least squares (OLS) estimators. The OLS estimators are typically found with statistical software.

最新视频课程 QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

LO 23.6

The standard error of the regression is the standard deviation of the predicted values for the dependent variable about the regression line:

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

The coefficient of determination, R^2, is the percentage of the variation in Y that is explained by the set of independent variables.

- R^2 increases as the number of independent variables increases—this can be a problem.
- The adjusted R^2 adjusts the R^2 for the number of independent variables.
- $R_a^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) \times (1 - R^2)\right]$

LO 23.7

Assumptions of multiple regression mostly pertain to the error term, ε_i.

- A linear relationship exists between the dependent and independent variables.
- The independent variables are not random, and there is no exact linear relation between any two or more independent variables.
- The expected value of the error term is zero.
- The variance of the error terms is constant.
- The error for one observation is not correlated with that of another observation.
- The error term is normally distributed.

LO 23.8

Perfect multicollinearity exists when one of the independent variables is a perfect linear combination of the other independent variable. Imperfect multicollinearity arises when two or more independent variables are highly correlated, but less than perfectly correlated.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checkers

Use the following table for Question 1.

Source	*Sum of Squares (SS)*
Explained	1,025
Residual	925

1. The total sum of squares (TSS) is closest to:
   A. 100.
   B. 1.108.
   C. 1,950.
   D. 0.9024.

Use the following information to answer Questions 2 and 3.

Multiple regression was used to explain stock returns using the following variables:

Dependent variable:

RET = annual stock returns (%)

Independent variables:

MKT = market capitalization = market capitalization / $1.0 million

IND = industry quartile ranking (IND = 4 is the highest ranking)

FORT = Fortune 500 firm, where {FORT = 1 if the stock is that of a Fortune 500 firm, FORT = 0 if not a Fortune 500 stock}

The regression results are presented in the tables below.

	Coefficient	*Standard Error*	*t-Statistic*	*p-Value*
Intercept	0.5220	1.2100	0.430	0.681
Market capitalization	0.0460	0.0150	3.090	0.021
Industry ranking	0.7102	0.2725	2.610	0.040
Fortune 500	0.9000	0.5281	1.700	0.139

2. Based on the results in the table, which of the following most accurately represents the regression equation?
   A. 0.43 + 3.09(MKT) + 2.61(IND) + 1.70(FORT).
   B. 0.681 + 0.021(MKT) + 0.04(IND) + 0.139(FORT).
   C. 0.522 + 0.0460(MKT) + 0.7102(IND) + 0.9(FORT).
   D. 1.21 + 0.015(MKT) + 0.2725(IND) + 0.5281(FORT).

©2015 Kaplan, Inc.

FRM官方交流群 392417118

3. The expected amount of the stock return attributable to it being a Fortune 500 stock is closest to:
   A. 0.522.
   B. 0.046.
   C. 0.710.
   D. 0.900.

4. Which of the following situations is not possible from the results of a multiple regression analysis with more than 50 observations?

	R^2	Adjusted R^2
A.	71%	69%
B.	83%	86%
C.	54%	12%
D.	10%	–2%

5. Assumptions underlying a multiple regression are most likely to include:
   A. The expected value of the error term is 0.00 < i < 1.00.
   B. Linear and non-linear relationships exist between the dependent and independent variables.
   C. The error for one observation is not correlated with that of another observation.
   D. The variance of the error terms is not constant for all observations.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checker Answers

1. **C** TSS = 1,025 + 925 = 1,950

2. **C** The coefficients column contains the regression parameters.

3. **D** The regression equation is 0.522 + 0.0460(MKT) + 0.7102(IND) + 0.9(FORT). The coefficient on FORT is the amount of the return attributable to the stock of a Fortune 500 firm.

4. **B** Adjusted R^2 must be less than or equal to R^2. Also, if R^2 is low enough and the number of independent variables is large, adjusted R^2 may be negative.

5. **C** Assumptions underlying a multiple regression include: the error for one observation is not correlated with that of another observation; the expected value of the error term is zero; a linear relationship exists between the dependent and independent variables; the variance of the error terms is constant.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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 Tests and Confidence Intervals in Multiple Regression

Topic 24

Exam Focus

This topic addresses methods for dealing with uncertainty in a multiple regression model. Hypothesis tests and confidence intervals for single- and multiple-regression coefficients will be discussed. For the exam, you should know how to use a *t*-test to assess the significance of the individual regression parameters and an *F*-test to assess the effectiveness of the model as a whole in explaining the dependent variable. Also, be able to identify the common model misspecifications. Focus on interpretation of the regression equation and the test statistics. Remember that most of the test and descriptive statistics discussed (e.g., *t*-stat, *F*-stat, and R^2) are provided in the output of statistical software. Hence, application and interpretation of these measurements are more likely than actual computations on the exam.

LO 24.1: Construct, apply, and interpret hypothesis tests and confidence intervals for a single coefficient in a multiple regression.

Hypothesis Testing of Regression Coefficients

As with simple linear regression, the magnitude of the coefficients in a multiple regression tells us nothing about the importance of the independent variable in explaining the dependent variable. Thus, we must conduct hypothesis testing on the estimated slope coefficients to determine if the independent variables make a significant contribution to explaining the variation in the dependent variable.

The *t*-statistic used to test the significance of the individual coefficients in a multiple regression is calculated using the same formula that is used with simple linear regression:

$$t = \frac{b_j - B_j}{s_{b_j}} = \frac{\text{estimated regression coefficient} - \text{hypothesized value}}{\text{coefficient standard error of } b_j}$$

The *t*-statistic has $n - k - 1$ degrees of freedom.

Professor's Note: An easy way to remember the number of degrees of freedom for this test is to recognize that "k" is the number of regression coefficients in the regression, and the "1" is for the intercept term. Therefore, the degrees of freedom is the number of observations minus k minus 1.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Determining Statistical Significance

The most common hypothesis test done on the regression coefficients is to test statistical significance, which means testing the null hypothesis that the coefficient is zero versus the alternative that it is not:

"testing statistical significance" $\Rightarrow H_0: b_j = 0$ versus $H_A: b_j \neq 0$

Example: Testing the statistical significance of a regression coefficient

Consider again, from the previous topic, the hypothesis that future 10-year real earnings growth in the S&P 500 (EG10) can be explained by the trailing dividend payout ratio of the stocks in the index (PR) and the yield curve slope (YCS). **Test** the statistical significance of the independent variable PR in the real earnings growth example at the 10% significance level. Assume that the number of observations is 46. The results of the regression are reproduced in the following figure.

Coefficient and Standard Error Estimates for Regression of EG10 on PR and YCS

	Coefficient	*Standard Error*
Intercept	–11.6%	1.657%
PR	0.25	0.032
YCS	0.14	0.280

Answer:

We are testing the following hypothesis:

$H_0: PR = 0$ versus $H_A: PR \neq 0$

The 10% two-tailed critical *t*-value with 46 – 2 – 1 = 43 degrees of freedom is approximately 1.68. We should reject the null hypothesis if the *t*-statistic is greater than 1.68 or less than –1.68.

The *t*-statistic is:

$$t = \frac{0.25}{0.032} = 7.8$$

Therefore, because the *t*-statistic of 7.8 is greater than the upper critical *t*-value of 1.68, we can reject the null hypothesis and conclude that the PR regression coefficient is statistically significantly different from zero at the 10% significance level.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Interpreting *p*-Values

The *p*-value is the smallest level of significance for which the null hypothesis can be rejected. An alternative method of doing hypothesis testing of the coefficients is to compare the *p*-value to the significance level:

- If the *p*-value is less than significance level, the null hypothesis can be rejected.
- If the *p*-value is greater than the significance level, the null hypothesis cannot be rejected.

Example: Interpreting *p*-values

Given the following regression results, **determine** which regression parameters for the independent variables are statistically significantly different from zero at the 1% significance level, assuming the sample size is 60.

Variable	*Coefficient*	*Standard Error*	*t-Statistic*	*p-Value*
Intercept	0.40	0.40	1.0	0.3215
X1	8.20	2.05	4.0	0.0002
X2	0.40	0.18	2.2	0.0319
X3	–1.80	0.56	–3.2	0.0022

Answer:

The independent variable is statistically significant if the *p*-value is less than 1%, or 0.01. Therefore *X*1 and *X*3 are statistically significantly different from zero.

Figure 1 shows the results of the *t*-tests for each of the regression coefficients of our 10-year earnings growth example, including the *p*-values.

Figure 1: Regression Results for Regression of EG10 on PR and YCS

	Coefficient	*Standard Error*	*t-statistic*	*p-value*
Intercept	–11.6%	1.657%	–7.0	< 0.0001
PR	0.25	0.032	7.8	< 0.0001
YCS	0.14	0.280	0.5	0.62

As we determined in a previous example, we can reject the null hypothesis and conclude that PR is statistically significant. We can also draw the same conclusion for the intercept term because –7.0 is less than the lower critical value of –1.68 (because it is a two-tailed test). However, we fail to reject the null hypothesis for YCS, so we cannot conclude that YCS has a statistically significant effect on the dependent variable, EG10, when PR is also included in the model. The *p*-values tell us exactly the same thing (as they always will): the

©2015 Kaplan, Inc.
FRM官方交流群 392417118

intercept term and PR are statistically significant at the 10% level because their *p*-values are less than 0.10, while YCS is not statistically significant because its *p*-value is greater than 0.10.

Other Tests of the Regression Coefficients

You should also be prepared to formulate one- and two-tailed tests in which the null hypothesis is that the coefficient is equal to some value other than zero, or that it is greater than or less than some value.

Example: Testing regression coefficients (two-tail test)

Using the data from Figure 1, test the null hypothesis that PR is equal to 0.20 versus the alternative that it is not equal to 0.20 using a 5% significance level.

Answer:

We are testing the following hypothesis:

$$H_0: PR = 0.20 \text{ versus } H_A: PR \neq 0.20$$

The 5% two-tailed critical *t*-value with 46 – 2 – 1 = 43 degrees of freedom is approximately 2.02. We should reject the null hypothesis if the *t*-statistic is greater than 2.02 or less than –2.02.

The *t*-statistic is:

$$t = \frac{0.25 - 0.20}{0.032} = 1.56$$

Therefore, because the *t*-statistic of 1.56 is between the upper and lower critical *t*-values of –2.02 and 2.02, we cannot reject the null hypothesis and must conclude that the PR regression coefficient is not statistically significantly different from 0.20 at the 5% significance level.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Example: Testing regression coefficients (one-tail test)

Using the data from Figure 1, **test** the null hypothesis that the intercept term is greater than or equal to –10.0% versus the alternative that it is less than –10.0% using a 1% significance level.

Answer:

We are testing the following hypothesis:

H_0: Intercept $\geq$ –10.0% versus H_A: Intercept < –10.0%

The 1% **one-tailed** critical *t*-value with 46 – 2 – 1 = 43 degrees of freedom is approximately 2.42. We should reject the null hypothesis if the *t*-statistic is less than –2.42.

The *t*-statistic is:

$$t = \frac{-11.6\% - (-10.0\%)}{1.657\%} = -0.96$$

Therefore, because the *t*-statistic of –0.96 is not less than –2.42, we cannot reject the null hypothesis.

Confidence Intervals for a Regression Coefficient

The confidence interval for a regression coefficient in multiple regression is calculated and interpreted the same way as it is in simple linear regression. For example, a 95% confidence interval is constructed as follows:

$$b_j \pm (t_c \times s_{b_j})$$

or

estimated regression coefficient ± (critical *t*-value)(coefficient standard error)

The critical *t*-value is a two-tailed value with n – k – 1 degrees of freedom and a 5% significance level, where *n* is the number of observations and *k* is the number of independent variables.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Example: Calculating a confidence interval for a regression coefficient

Calculate the 90% confidence interval for the estimated coefficient for the independent variable PR in the real earnings growth example.

Answer:

The critical *t*-value is 1.68, the same as we used in testing the statistical significance at the 10% significance level (which is the same thing as a 90% confidence level). The estimated slope coefficient is 0.25 and the standard error is 0.032. The 90% confidence interval is:

$$0.25 \pm (1.68)(0.032) = 0.25 \pm 0.054 = 0.196 \text{ to } 0.304$$

Professor's Note: Notice that because zero is not contained in the 90% confidence interval, we can conclude that the PR coefficient is statistically significant at the 10% level. Constructing a confidence interval and conducting a t-test with a null hypothesis of "equal to zero" will always result in the same conclusion regarding the statistical significance of the regression coefficient.

Predicting the Dependent Variable

We can use the regression equation to make predictions about the dependent variable *based on forecasted values of the independent variables.* The process is similar to forecasting with simple linear regression, only now we need predicted values for more than one independent variable. The predicted value of dependent variable *Y* is:

$$\hat{Y}_i = b_0 + b_1\hat{X}_{1i} + b_2\hat{X}_{2i} + \ldots + b_k\hat{X}_{ki}$$

where:

$\hat{Y}_i$ = the predicted value of the dependent variable

b_j = the estimated slope coefficient for the *j*th independent variable

$\hat{X}_{ji}$ = the forecast of the *j*th independent variable, j = 1, 2, …, *k*

Professor's Note: The prediction of the dependent variable uses the estimated intercept and all of the estimated slope coefficients, regardless of whether the estimated coefficients are statistically significantly different from zero. For example, suppose you estimate the following regression equation: $\hat{Y} = 6 + 2X_1 + 4X_2$, and you determine that only the first independent variable (X_1) is statistically significant (i.e., you rejected the null that $B_1 = 0$). To predict Y given forecasts of $X_1 = 0.6$ and $X_2 = 0.8$, you would use the complete model: $\hat{Y} = 6 + (2 \times 0.6) + (4 \times 0.8) = 10.4$. Alternatively, you could drop X2 and reestimate the model using just X1, but remember that the coefficient on X1 will likely change.

FRM官方交流群 392417118

Example: Calculating a predicted value for the dependent variable

An analyst would like to use the estimated regression equation from the previous example to **calculate** the predicted 10-year real earnings growth for the S&P 500, assuming the payout ratio of the index is 50%. He observes that the slope of the yield curve is currently 4%.

Answer:

$$\widehat{EG10} = -11.6\% + 0.25(50\%) + 0.14(4\%) = 1.46\%$$

The model predicts a 1.46% real earnings growth rate for the S&P 500, assuming a 50% payout ratio, when the slope of the yield curve is 4%.

Joint Hypothesis Testing

LO 24.2: Construct, apply, and interpret joint hypothesis tests and confidence intervals for multiple coefficients in a multiple regression.

LO 24.3: Interpret the F-statistic.

LO 24.5: Interpret confidence sets for multiple coefficients.

A joint hypothesis tests two or more coefficients at the same time. For example, we could develop a null hypothesis for a linear regression model with three independent variables that sets two of these coefficients equal to zero: H_0: $b_1 = 0$ and $b_2 = 0$ versus the alternative hypothesis that one of them is not equal to zero. That is, if just one of the equalities in this null hypothesis does not hold, we can reject the entire null hypothesis. Using a joint hypothesis test is preferred in certain scenarios since testing coefficients individually leads to a greater chance of rejecting the null hypothesis. For example, instead of comparing one t-statistic to its corresponding critical value in a joint hypothesis test, we are testing two t-statistics. Thus, we have an additional opportunity to reject the null. A robust method for applying joint hypothesis testing, especially when independent variables are correlated, is known as the *F*-statistic.

The *F*-Statistic

An *F*-test assesses how well the set of independent variables, as a group, explains the variation in the dependent variable. That is, the *F*-statistic is used to test whether *at least one* of the independent variables explains a significant portion of the variation of the dependent variable.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

For example, if there are four independent variables in the model, the hypotheses are structured as:

$$H_0: B_1 = B_2 = B_3 = B_4 = 0 \text{ versus } H_A: \text{at least one } B_j \neq 0$$

The *F*-statistic, *which is always a one-tailed test*, is calculated as:

$$\frac{ESS/k}{SSR/n-k-1}$$

where:
ESS = explained sum of squares
SSR = sum of squared residuals

Professor's Note: The explained sum of squares and the sum of squared residuals are found in an analysis of variance (ANOVA) table. We will analyze an ANOVA table from a multiple regression shortly.

To determine whether at least one of the coefficients is statistically significant, the calculated *F*-statistic is compared with the **one-tailed** critical *F*-value, F_c, at the appropriate level of significance. The degrees of freedom for the numerator and denominator are:

$$df_{numerator} = k$$

$$df_{denominator} = n - k - 1$$

where:
n = number of observations
k = number of independent variables

The decision rule for the *F*-test is:

Decision rule: reject H_0 if F (test-statistic) > F_c (critical value)

Rejection of the null hypothesis at a stated level of significance indicates that at least one of the coefficients is significantly different than zero, which is interpreted to mean that at least one of the independent variables in the regression model makes a significant contribution to the explanation of the dependent variable.

Professor's Note: It may have occurred to you that an easier way to test all of the coefficients simultaneously is to just conduct all of the individual t-tests and see how many of them you can reject. This is the wrong approach, however, because if you set the significance level for each t-test at 5%, for example, the significance level from testing them all simultaneously is NOT 5%, but rather some higher percentage. Just remember to use the F-test on the exam if you are asked to test all of the coefficients simultaneously.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Example: Calculating and interpreting the *F*-statistic

An analyst runs a regression of monthly value-stock returns on five independent variables over 60 months. The total sum of squares is 460, and the sum of squared residuals is 170. Test the null hypothesis at the 5% significance level (95% confidence) that all five of the independent variables are equal to zero.

Answer:

The null and alternative hypotheses are:

$$H_0: B_1 = B_2 = B_3 = B_4 = B_5 = 0 \text{ versus } H_A: \text{at least one } B_j \neq 0$$

$$\text{ESS} = \text{TSS} - \text{SSR} = 460 - 170 = 290$$

$$F = \frac{58.0}{3.15} = 18.41$$

The critical *F*-value for 5 and 54 degrees of freedom at a 5% significance level is approximately 2.40. Remember, it's a **one-tailed** test, so we use the 5% *F*-table! Therefore, we can reject the null hypothesis and conclude that at least one of the five independent variables is significantly different than zero.

Professor's Note: When testing the hypothesis that all the regression coefficients are simultaneously equal to zero, the F-test is always a one-tailed test, despite the fact that it looks like it should be a two-tailed test because there is an equal sign in the null hypothesis.

Interpreting Regression Results

Just as in simple linear regression, the variability of the dependent variable or **total sum of squares** (TSS) can be broken down into **explained sum of squares** (ESS) and **sum of squared residuals** (SSR). As shown previously, the coefficient of determination is:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum\left(\hat{Y} - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\text{SSR}}{\text{TSS}} = 1 - \frac{\sum e_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

Regression results usually provide R^2 and a host of other measures. However, it is useful to know how to compute R^2 from other parts of the results. Figure 2 is an ANOVA table of the results of a regression of hedge fund returns on lockup period and years of experience of the manager. In the ANOVA table, the value of 90 represents TSS, the ESS equals 84.057, and the SSR is 5.943. Although the output results provide the value $R^2 = 0.934$, it can also be computed using TSS, ESS, and SSR like so:

$$R^2 = \frac{84.057}{90} = 1 - \frac{5.943}{90} = 0.934$$

©2015 Kaplan, Inc.
FRM官方交流群 392417118

The **coefficient of multiple correlation** is simply the square root of *R*-squared. In the case of a multiple regression, the coefficient of multiple correlation is always positive.

Figure 2: ANOVA Table

R-squared	0.934					
Adj *R*-squared	0.890					
Standard error	1.407					
Observations	6					
	Degrees of Freedom	*SS*	*MS*	*F*		
Explained	2	84.057	42.029	21.217		
Residual	3	5.943	1.981			
Total	5	90				
Variables	*Coeff*	*Std Error*	*t-stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	−4.4511	3.299	−1.349	0.270	−14.950	6.048
Lockup	2.057	0.337	6.103	0.009	0.984	3.130
Experience	2.008	0.754	2.664	0.076	−0.391	4.407

The results in Figure 2 produce the following equation:

$$\hat{Y}_i = -4.451 + 2.057 \times X_{1i} + 2.008 \times X_{2i}$$

This equation tells us that holding other variables constant, increasing the lockup period will increase the expected return of a hedge fund by 2.057%. Also, holding other variables constant, increasing the manager's experience one year will increase the expected return of a hedge fund by 2.008%. A hedge fund with an inexperienced manager and no lockup period will earn a negative return of −4.451%.

The ANOVA table outputs the standard errors, *t*-statistics, probability values (p-values), and confidence intervals for the estimated coefficients. These can be used in a hypothesis test for each coefficient. For example, for the independent variable experience (b_2), the output indicates that the standard error is $se(b_2) = 0.754$, which yields a *t*-statistic of: 2.008 / 0.754 = 2.664. The critical *t*-value at a 5% level of significance is $t_{0.025} = 3.182$. Thus, a hypothesis stating that the number of years of experience is not related to returns could not be rejected. In other words, the result is to not reject the null hypothesis that $B_2 = 0$. This is also seen with the provided confidence interval. Upper and lower limits of the confidence interval can be found in the ANOVA results.

$$[b_2 - t_{\alpha/2} \times se(b_2)] < B_2 < [b_2 + t_{\alpha/2} \times se(b_2)]$$

$$(2.008 - 3.182 \times 0.754) < B_2 < (2.008 + 3.182 \times 0.754)$$

$$-0.391 < B_2 < 4.407$$

Since the confidence interval contains the value zero, then the null hypothesis: H_0: $B_2 = 0$ cannot be rejected in a two-tailed test at the 5% level of significance. Figure 2 provides a third way of performing a hypothesis test by providing a p-value. The **p-value** indicates the

©2015 Kaplan, Inc.

FRM官方交流群 392417118

minimum level of significance at which the two-tailed hypothesis test can be rejected. In this case, the p-value is 0.076 (i.e., 7.6%), which is greater than 5%.

The statistics for b_1 indicate that a null hypothesis can be rejected at a 5% level using a two-tailed test. The *t*-statistic is 6.103, and the confidence interval is 0.984 to 3.13. The p-value of 0.9% is less than 5%.

The statistics in the ANOVA table also allow for the testing of the joint hypothesis that both slope coefficients equal zero.

H_0: $B_1 = B_2 = 0$

H_A: $B_1 \neq 0$ or $B_2 \neq 0$

The test statistic in this case is the *F*-statistic where the degrees of freedom are indicated by two numbers: the number of slope coefficients (2) and the sample size minus the number of slope coefficients minus one (6 – 2 – 1 = 3). The *F*-statistic given the hedge fund data can be calculated as follows:

$$F = \frac{ESS/df}{SSR/df} = \frac{84.057/2}{5.943/3} = \frac{42.029}{1.981} = 21.217$$

The critical *F*-statistic at a 5% significance level is $F_{0.05} = 9.55$. Since the value from the regression results is greater than that value: F = 21.217 > 9.55, a researcher would reject the null hypothesis: H_0: $B_1 = B_2 = 0$. It should be noted that rejecting the null hypothesis indicates one or both of the coefficients are significant.

Specification Bias

Specification bias refers to how the slope coefficient and other statistics for a given independent variable are usually different in a simple regression when compared to those of the same variable when included in a multiple regression. To illustrate this point, the following three OLS results correspond to a two-variable regression using only the indicated independent variable and the results for a three-variable:

$$\hat{Y}_i = 1 + 2 \times (\text{lockup})_i \quad t = 3.742$$

$$\hat{Y}_i = 11.714 + 1.714 \times (\text{experience})_i \quad t = 2.386$$

$$\hat{Y}_i = -4.451 + 2.057 \times (\text{lockup})_i + 2.008 \times (\text{experience})_i \quad t = 6.103 \quad t = 2.664$$

Specification bias is indicated by the extent to which the coefficient for each independent variable is different when compared across equations (e.g., for lockup, the slope is 2 in the two-variable equation, and the slope is 2.057 in the multivariate regression). This is because in the two-variable regression, the slope coefficient includes the effect of the included independent variable in the equation and, to some extent, the indirect effect of the excluded

©2015 Kaplan, Inc.
FRM官方交流群 392417118

variable(s). In this case, the bias for the coefficient on the lockup coefficient was not large because the experience variable was not significant as indicated in its two-variable regression ($t = 2.386 < t_{0.025} = 2.78$) and was not significant in the multivariable regression either.

R^2 and Adjusted R^2

LO 24.7: Interpret the R^2 and adjusted-R^2 in a multiple regression.

To further analyze the importance of an added variable to a regression, we can compute an adjusted coefficient of determination, or **adjusted R^2**. The reason adjusted R^2 is important is because, mathematically speaking, the coefficient of determination, R^2, must go up if a variable with any explanatory power is added to the regression, even if the marginal contribution of the new variables is not statistically significant. Consequently, a relatively high R^2 may reflect the impact of a large set of independent variables rather than how well the set explains the dependent variable. This problem is often referred to as overestimating the regression.

When computing both the R^2 and the adjusted R^2, there are a few pitfalls to acknowledge, which could lead to invalid conclusions.

1. If adding an additional independent variable to the regression improves the R^2, this variable is not necessary statistically significant.

2. The R^2 measure may be spurious, meaning that the independent variables may show a high R^2; however, they are not the exact cause of the movement in the dependent variable.

3. If the R^2 is high, we cannot assume that we have found all relevant independent variables. Omitted variables may still exist, which would improve the regression results further.

4. The R^2 measure does not provide evidence that the most or least appropriate independent variables have been selected. Many factors go into finding the most robust regression model, including omitted variable analysis, economic theory, and the quality of data being used to generate the model.

Restricted vs. Unrestricted Least Squares Models

A restricted least squares regression imposes a value on one or more coefficients with the goal of analyzing if the restriction is significant. To explain this concept, it is useful to note that there is an implied restriction in each of the two variable regressions:

$$\hat{Y}_i = b_0 + b_{lockup} \times (lockup)_i$$
$$\hat{Y}_i = b_0 + b_{experience} \times (experience)_i$$

In essence, each of the two-variable regressions is a restricted regression where the coefficient on the omitted variable is restricted to zero. To help illustrate the concept, the more elaborate subscripts have been used in these expressions. Using the indicated notation, the first specification that only includes "lockup" is restricting $b_{experience}$ to 0. In the unrestricted

FRM官方交流群 392417118

multivariable regression, both b_{lockup} and $b_{experience}$ are allowed to assume the values that minimize the SSR. The R^2 from the restricted regression is called a **restricted R^2** or R_r^2. For comparison, the **unrestricted R^2** from the specification that includes both independent variables is given the notation R_{ur}^2, and both are included in an *F*-statistic that can test if the restriction is significant or not:

$$F = \frac{(R_{ur}^2 - R_r^2)/m}{(1 - R_{ur}^2)/(n - k_{ur} - 1)}$$

The symbol "*m*" refers to the number of restrictions, which in the example discussed would be equal to one. This *F*-stat is known as the **homoskedasticity-only *F*-statistic** since it can only be derived from R^2 when the error terms display homoskedasticity. An alternative formula for computing this *F*-stat is to use the sum of squared residuals in place of the R^2:

$$F = \frac{(SSR_{ur} - SSR_r)/m}{SSR_{ur}/(n - k_{ur} - 1)}$$

In the event that the error terms are not homoskedastic, a hetroskedasticity-robust *F*-stat would be applied. This statistic is used more frequently in practice; however, as the sample size, *n*, increases, these two types of *F*-statistics will converge.

LO 24.4: Interpret tests of a single restriction involving multiple coefficients.

With the *F*-statistic, we constructed a null hypothesis that tested multiple coefficients being equal to zero. However, what if we wanted to test whether one coefficient was equal to another such that: H_0: $b_1 = b_2$? The alternative hypothesis in this scenario would be that the two are not equal to each other. Hypothesis tests of single restrictions involving multiple coefficients requires the use of statistical software packages, but we will examine the methodology of two different approaches.

The first approach is to directly test the restriction stated in the null. Some statistical packages can test this restriction and output a corresponding *F*-stat. This is the easier of the two methods; however, a second method will need to be applied if your statistical package cannot directly test the restriction.

The second approach transforms the regression and uses the null hypothesis as an assumption to simplify the regression model. For example, in a regression with two independent variables: $Y_i = B_0 + B_1X_{1i} + B_2X_{2i} + \varepsilon_i$, we can add and subtract B_2X_{1i} to ultimately transform the regression to: $B_0 + (B_1 - B_2)X_{1i} + B_2(X_{1i} + X_{2i}) + \varepsilon_i$. One of the coefficients will drop out in this equation when assuming that the null hypothesis of $B_1 = B_2$ is valid. We can remove the second term from our regression equation so that: $B_0 + B_2(X_{1i} + X_{2i}) + \varepsilon_i$. We observe that the null hypothesis test changes from a single restriction involving multiple coefficients to a single restriction on just one coefficient.

Professor's Note: Remember that this process is typically done with statistical software packages, so on the exam, you would simply be asked to describe and/or interpret these tests.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Model Misspecification

LO 24.6: Identify examples of omitted variable bias in multiple regressions.

Recall from the previous topic that omitting relevant factors from a regression can produce misleading or biased results. Similar to simple linear regression, omitted variable bias in multiple regressions will result if the following two conditions occur:

- The omitted variable is a determinant of the dependent variable.
- The omitted variable is correlated with *at least* one of the independent variables.

As an example of omitted variable bias, consider a regression in which we're trying to predict monthly returns on portfolios of stocks (R) using three independent variables: portfolio beta (B), the natural log of market capitalization (lnM), and the natural log of the price-to-book ratio ln(PB). The correct specification of this model is as follows:

$$R = b_0 + b_1B + b_2\ln M + b_3\ln PB + \varepsilon$$

Now suppose we did not include lnM in the regression model:

$$R = a_0 + a_1B + a_2\ln PB + \varepsilon$$

If lnM is correlated with any of the remaining independent variables (B or lnPB), then the error term is also correlated with the same independent variables, and the resulting regression coefficients are biased and inconsistent. That means our hypothesis tests and predictions using the model will be unreliable.

最新视频课程 QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 24.1

A *t*-test is used for hypothesis testing of regression parameter estimates:

$$t = \frac{b_j - B_j}{s_{b_j}}, \text{ with } n - k - 1 \text{ degrees of freedom}$$

Testing for statistical significance means testing H_0: $B_j = 0$ vs. H_A: $B_j \neq 0$.

LO 24.2

The confidence interval for regression coefficient is:

estimated regression coefficient ± (critical t-value)(coefficient standard error)

The value of dependent variable Y is predicted as:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k$$

LO 24.3

The *F*-distributed test statistic can be used to test the significance of all (or any subset of) the independent variables (i.e., the overall fit of the model) using a one-tailed test:

$$F = \frac{ESS/k}{SSR/[n-k-1]} \text{ with } k \text{ and } n - k - 1 \text{ degrees of freedom}$$

LO 24.4

Hypothesis tests of single restrictions involving multiple coefficients requires the use of statistical software packages.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

LO 24.5

The ANOVA table outputs the standard errors, t-statistics, probability values (*p*-values), and confidence intervals for the estimated coefficients.

Upper and lower limits of the confidence interval can be found in the ANOVA results.

$$[b_2 - t_{\alpha/2} \times se(b_2)] < B_2 < [b_2 + t_{\alpha/2} \times se(b_2)]$$

The statistics in the ANOVA table also allow for the testing of the joint hypothesis that both slope coefficients equal zero.

H_0: $B_1 = B_2 = 0$

H_A: $B_1 \neq 0$ or $B_2 \neq 0$

The test statistic in this case is the *F*-statistic.

LO 24.6

Omitting a relevant independent variable in a multiple regression results in regression coefficients that are biased and inconsistent, which means we would not have any confidence in our hypothesis tests of the coefficients or in the predictions of the model.

LO 24.7

Restricted least squares models restrict one or more of the coefficients to equal a given value and compare the R^2 of the restricted model to that of the unrestricted model where the coefficients are not restricted. An *F*-statistic can test if there is a significant difference between the restricted and unrestricted R^2.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checkers

Use the following table for Question 1.

Source	*Sum of Squares (SS)*	*Degrees of Freedom*
Explained	1,025	5
Residual	925	25

1. The R^2 and the *F*-statistic, respectively, are closest to:

	R^2	*F*-statistic
A.	53%	1.1
B.	47%	1.1
C.	53%	5.5
D.	47%	5.5

Use the following information to answer Question 2.

An analyst calculates the sum of squared residuals and total sum of squares from a multiple regression with four independent variables to be 4,320 and 9,105, respectively. There are 65 observations in the sample.

2. The critical *F*-value for testing $H_0 = B_1 = B_2 = B_3 = B_4 = 0$ vs. H_A: at least one $B_j \neq 0$ at the 5% significance level is closest to:
   A. 2.37.
   B. 2.53.
   C. 2.76.
   D. 3.24.

3. When interpreting the R^2 and adjusted R^2 measures for a multiple regression, which of the following statements incorrectly reflects a pitfall that could lead to invalid conclusions?
   A. The R^2 measure does not provide evidence that the most or least appropriate independent variables have been selected.
   B. If the R^2 is high, we have to assume that we have found all relevant independent variables.
   C. If adding an additional independent variable to the regression improves the R^2, this variable is not necessarily statistically significant.
   D. The R^2 measure may be spurious, meaning that the independent variables may show a high R^2; however, they are not the exact cause of the movement in the dependent variable.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Use the following information for Questions 4 and 5.

Phil Ohlmer estimates a cross sectional regression in order to predict price to earnings ratios (P/E) with fundamental variables that are related to P/E, including dividend payout ratio (DPO), growth rate (G), and beta (B). In addition, all 50 stocks in the sample come from two industries, electric utilities or biotechnology. He defines the following dummy variable:

IND = 0 if the stock is in the electric utilities industry, or
= 1 if the stock is in the biotechnology industry

The results of his regression are shown in the following table.

Variable	*Coefficient*	*t-Statistic*
Intercept	6.75	3.89*
IND	8.00	4.50*
DPO	4.00	1.86
G	12.35	2.43*
B	–0.50	1.46

*significant at the 5% level

4. Based on these results, it would be most appropriate to conclude that:
   A. biotechnology industry PEs are statistically significantly larger than electric utilities industry PEs.
   B. electric utilities PEs are statistically significantly larger than biotechnology industry PEs, holding DPO, G, and B constant.
   C. biotechnology industry PEs are statistically significantly larger than electric utilities industry PEs, holding DPO, G, and B constant.
   D. the dummy variable does not display statistical significance.

5. Ohlmer is valuing a biotechnology stock with a dividend payout ratio of 0.00, a beta of 1.50, and an expected earnings growth rate of 0.14. The predicted P/E on the basis of the values of the explanatory variables for the company is closest to:
   A. 7.7.
   B. 15.7.
   C. 17.2.
   D. 11.3.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checker Answers

1. C $R^2 = \frac{ESS}{TSS} = \frac{1{,}025}{1{,}950} = 53\%$ $F = \frac{\frac{ESS}{df}}{\frac{SSR}{df}} = \frac{\frac{1{,}025}{5}}{\frac{925}{25}} = \frac{205}{37} = 5.5$

2. B This is a one-tailed test, so the critical *F*-value at the 5% significance level with 4 and 60 degrees of freedom is approximately 2.53.

3. B If the R^2 is high, we *cannot* assume that we have found all relevant independent variables. Omitted variables may still exist, which would improve the regression results further.

4. C The *t*-statistic tests the null that industry PEs are equal. The dummy variable is significant and positive, and the dummy variable is defined as being equal to one for biotechnology stocks, which means that biotechnology PEs are statistically significantly larger than electric utility PEs. Remember, however, this is only accurate if we hold the other independent variables in the model constant.

5. B Note that IND = 1 because the stock is in the biotech industry. Predicted P/E = 6.75 + (8.00×1) + (4.00×0.00) + (12.35×0.14) – (0.50×1.5) = 15.7.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

Modeling and Forecasting Trend

Topic 25

Exam Focus

The focus of this topic is on selecting the best forecasting model to estimate a trend using four common selection criteria. For the exam, you should understand how mean squared error (MSE) is calculated and how adjusting for degrees of freedom, *k*, is accomplished with the unbiased MSE (or s^2), Akaike information criterion (AIC), and Schwarz information criterion (SIC). Also, be able to explain how these three selection tools compare based on penalty factors and the consistency property.

Mean Squared Error

LO 25.1: Define mean squared error (MSE) and explain the implications of MSE in model selection.

Mean squared error (MSE) is a statistical measure computed as the sum of squared residuals divided by the total number of observations in the sample.

$$MSE = \frac{\sum_{t=1}^{T} e_t^2}{T}$$

where:

T = total sample size

e_t = $y_t - \hat{y}_t$ (the residual for observation *t* or difference between the observed and expected observation)

y_t = $\hat{\beta}_0 + \hat{\beta}_1 TIME_t$ (i.e., a regression model)

The MSE is based on *in-sample* data. The regression model with the smallest MSE is also the model with the smallest sum of squared residuals. The residuals are calculated as the difference between the actual value observed and the predicted value based on the regression model. Scaling the sum of squared residuals by 1 / T does not change the ranking of the models based on squared residuals.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

MSE is closely related to the coefficient of determination (R^2). Notice in the R^2 equation that the numerator is simply the sum of squared residuals (SSR), which is identical to the MSE numerator.

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

The denominator in the R^2 calculation is the sum of the difference of observations from the mean. Notice that we subtract the second term from one in the R^2 calculation. Thus, the regression model with the smallest MSE is also the one that has the largest R^2.

Model selection is one of the most important criteria in forecasting data. Unfortunately, selecting the best model based on the highest R^2 or smallest MSE is not effective in producing good *out-of-sample* forecasting models. A better methodology to select the best forecasting model is to find the model with the smallest out-of-sample, one-step-ahead MSE.

Reducing MSE Bias

LO 25.2: Explain how to reduce the bias associated with MSE and similar measures.

The use of in-sample MSE to estimate out-of-sample MSE is not very effective because in-sample MSE cannot increase when more variables are included in the forecasting model. Thus, MSE will have a downward bias when predicting out-of-sample error variance. Selection criteria differ based on the penalty imposed when the number of parameter estimates is increased in the regression model. One way to reduce the bias associated with MSE is to impose a penalty on the degrees of freedom, *k*. The s^2 **measure** is an unbiased estimate of the MSE because it corrects for degrees of freedom as follows:

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - k}$$

As more variables are included in a regression equation, the model is at greater risk of over-fitting the in-sample data. This problem is also often referred to as **data mining**. The problem with data mining is that the regression model does a very good job of explaining the sample data but does a poor job of forecasting out-of-sample data. As more parameters are introduced to a regression model, it will explain the data better, but may be worse at forecasting out-of-sample data.

Therefore, it is important to adjust for the number of variables or parameters used in a regression model because increasing the number of parameters will not necessarily improve the forecasting model. The degrees of freedom penalty rises with more parameters, but the MSE could fall. Thus, the best model is selected based on the smallest unbiased MSE, or s^2.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

The unbiased MSE estimate, s^2, will rank models in the same way as the adjusted R^2 measure. Adjusted R^2 using the s^2 estimate can be computed as follows:

$$\bar{R}^2 = 1 - \frac{s^2}{\sum_{t=1}^{T} \frac{(y_t - \bar{y})^2}{T-1}}$$

Notice that the denominator in this equation is based only on the data used in the regression. Therefore, it will be a constant number and the model with the highest adjusted R^2 will also have the smallest s^2. Thus, the s^2 and adjusted R^2 criteria will always rank forecasting models equivalently.

Model Selection Criteria

LO 25.3: Compare and evaluate model selection criteria, including s^2, the Akaike information criterion (AIC), and the Schwarz information criterion (SIC).

As mentioned, selection criteria are often compared based on a penalty factor. The unbiased MSE estimate, s^2, defined earlier, can be re-written (by multiplying T to the numerator and denominator) to highlight the penalty for degrees of freedom. In the following equation, the first term (T / T – k) can be thought of as the **penalty factor**.

$$s^2 = \left(\frac{T}{T-k}\right)\frac{\sum_{t=1}^{T} e_t^2}{T}$$

This notation is useful when comparing different selection criteria because it takes the form of a *penalty factor times the MSE*. The **Akaike information criterion** (AIC) and the **Schwarz information criterion** (SIC) use different penalty factors as follows:

$$AIC = e^{\left(\frac{2k}{T}\right)}\frac{\sum_{t=1}^{T} e_t^2}{T}$$

$$SIC = T^{\left(\frac{k}{T}\right)}\frac{\sum_{t=1}^{T} e_t^2}{T}$$

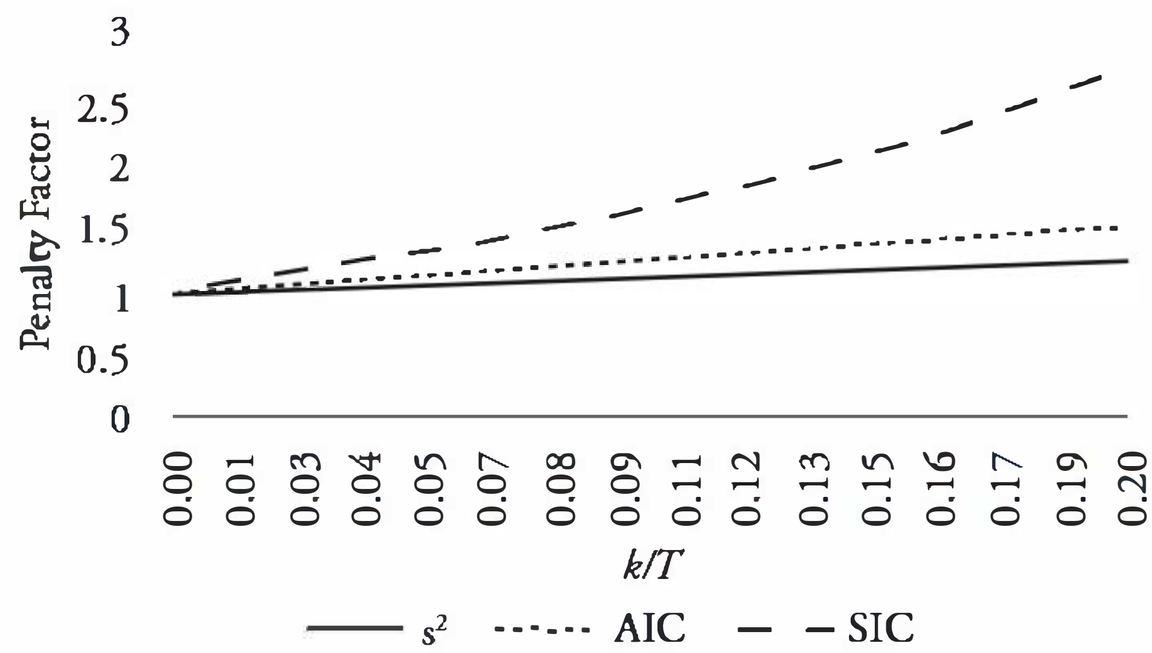
Note that the penalty factors for s^2, AIC, and SIC are (T / T – k), $e^{(2k/T)}$, and $T^{(k/T)}$, respectively.

Suppose an analyst runs a forecasting model with a total sample size of 150. Figure 1 illustrates the change in penalty factors for the s^2, AIC, and SIC as the degrees of freedom to total sample size (k / T) changes from 0 to 0.20. The s^2 penalty factor is the flattest line with a slow increase in penalty as k / T increases. The AIC penalty factor increases

FRM官方交流群 392417118

at a slightly higher rate than the s^2 penalty factor, and the SIC penalty factor increases exponentially at an increasing rate and, therefore, has the highest penalty factor.

Figure 1: Penalty Factor for s^2, AIC, and SIC



Evaluating Consistency

LO 25.4: Explain the necessary conditions for a model selection criterion to demonstrate consistency.

Consistency is a key property that is used to compare different selection criteria. Two conditions are required for a model selection criteria to be considered consistent based on whether the *true* model is included among the regression models being considered.

- When the *true* model or *data generating process* (DGP) is one of the defined regression models, then the probability of selecting the *true* model approaches one as the sample size increases.
- When the *true* model is *not* one of the defined regression models being considered, then the probability of selecting the *best approximation* model approaches one as the sample size increases.

Because we live in a very complex world, almost all economic and financial models have assumptions that simplify this complex environment. Thus, the reality is that the second condition of consistency is more relevant. All of our models are most likely false so, therefore, we are seeking the best approximation.

So how do our selection criteria fair based on consistency? MSE does not penalize for degrees of freedom and therefore is not consistent. The unbiased MSE, s^2, adjusts MSE for degrees of freedom, but the adjustment is too small for consistency. Figure 1 illustrated that AIC has a larger penalty factor than s^2. However, with large sample sizes the AIC tends to select models that have too many variables or parameters. This suggests that the penalty factor for degrees of freedom is still not large enough. The most consistent selection criteria with the greatest penalty factor for degrees of freedom is the SIC.

While the SIC is considered the most consistent criteria, the AIC is still a useful measure. If we consider the fact that the true model may be much more complicated than the models under consideration, then the AIC measure should be examined. *Asymptotic efficiency* is the

©2015 Kaplan, Inc.
FRM官方交流群 392417118

property that chooses a regression model with one-step-ahead forecast error variances closest to the variance of the true model. Interestingly, the AIC is asymptotically efficient and the SIC is not asymptotically efficient.

In conclusion, choosing the best forecasting model is an important task and we have discussed four key selection criteria. Adjusting for the degrees of freedom is extremely important and the SIC is the best selection criteria because it is consistent and also has the highest penalty factor. The AIC is also an important measure that is often considered in addition to SIC.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 25.1

Mean squared error (MSE) is a statistical measure computed as the sum of squared residuals (SSR) divided by the number of observations in a regression model:

$$MSE = \frac{\sum_{t=1}^{T} e_t^2}{T}$$

LO 25.2

The unbiased MSE, s^2, adjusts for the degrees of freedom, *k*, in the denominator as follows:

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T-k}$$

LO 25.3

The penalty factors for s^2, Akaike information criterion (AIC), and Schwarz information criterion (SIC) are $(T / T - k)$, $e^{(2k/T)}$, and $T^{(k/T)}$, respectively. SIC has the largest penalty factor.

LO 25.4

A selection criteria is considered to be consistent if the following two conditions are met:

- When the true model or data-generating process (DGP) is one of the defined regression models under consideration, then the probability of selecting the true model approaches one as the sample size increases.
- When the true model is not one of the defined regression models being considered, then the probability of selecting the best approximation model approaches one as the sample size increases.

The SIC is the most consistent selection criteria.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checkers

1. Which of the following statements best describes the effectiveness of the mean squared error (MSE) and R^2 criteria in selecting the best out-of-sample model?
   A. R^2 is the preferred criteria over MSE and will rank forecasting regression models more effectively than MSE.
   B. MSE is the preferred criteria over R^2 and will rank forecasting regression models more effectively than R^2.
   C. MSE is the preferred criteria over adjusted R^2 and will rank forecasting regression models more effectively than adjusted R^2.
   D. MSE and R^2 will rank forecasting models the same.

2. Richard Frank, FRM, is running a regression model to forecast in-sample data. He is concerned about data mining and over-fitting the data. Which of the following criteria provides the highest penalty factor based on degrees of freedom?
   A. Mean squared error (MSE).
   B. Unbiased mean squared error (s^2).
   C. Akaike information criterion (AIC).
   D. Schwarz information criterion (SIC).

3. Which of the following statements does not accurately describe the mean squared error (MSE) statistical measure?
   A. The regression model with the smallest MSE is also the model with the smallest sum of squared residuals.
   B. Scaling the sum of squared residuals by 1 / T changes the ranking of the models based on squared residuals.
   C. The residuals in the numerator of the MSE calculation are defined as the difference between the actual value observed and the predicted value based on the regression model.
   D. The best regression model based on minimizing the MSE will also be the one that maximizes R^2.

4. Sally Morgan, a junior analyst, is identifying a forecasting model based on a number of industry factors, company factors, and leading market indicators. She decides to choose the model with the highest R^2 measure because she knows this is a goodness-of-fit measure for selecting regression models. Morgan chooses a model with a very large number of parameters. How will Morgan's supervisor, Jessica Bolt, most likely respond to Morgan's choice of models? Bolt will:
   A. agree with Morgan as R^2 is the best goodness-of-fit measure available.
   B. agree with Morgan as R^2 is a common acceptable statistical measure and maximizing R^2 is the same as minimizing MSE.
   C. disagree with Morgan because MSE is a better measure than R^2 for selecting forecasting models.
   D. disagree with Morgan because R^2 is a biased measure.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

5. When selecting the best forecasting model among possible regression models, the property of consistency is desired. Which of the following statements most accurately describes a required condition for a model to be considered consistent?
   A. When the true model is one of the defined regression models under consideration, then the probability of selecting the best approximation model approaches one with a very large sample size.
   B. When the true model is one of the defined regression models under consideration, then the probability of selecting the true model approaches one with a very small sample size.
   C. When the true model is not one of the defined regression models being considered, then the probability of selecting the best approximation model approaches one as the sample size increases.
   D. When the true model is not one of the defined regression models being considered, the choice of the model selected is irrelevant and cannot be determined.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checker Answers

1. **D** MSE and R^2 will rank forecasting models the same. Neither criteria adjusts for the number of parameters and, therefore, both are based on similar measures of the sum of squared residuals for the regression model.

2. **D** The Schwarz information criterion (SIC) has the highest penalty factor. The mean squared error (MSE) does not penalize the regression model based on the increased number of parameters, k. The penalty factors for s^2, AIC, and SIC are $(T / T - k)$, $e^{(2k / T)}$, and $T^{(k / T)}$, respectively. Thus, SIC has the greatest penalty factor.

3. **B** Scaling the sum of squared residuals by 1 / T in the MSE statistic does *not* change the ranking of the models based on squared residuals. The rankings will be the same.

4. **D** The model selected by Morgan is at greater risk of over-fitting the in-sample data. It is important to adjust for the number of variables or parameters used in a regression model. The best model should be selected based on the smallest unbiased MSE, or s^2.

5. **C** A selection criteria is considered to be consistent if the following two conditions are met: (1) when the true model is not one of the defined regression models being considered, then the probability of selecting the best approximation model approaches one as the sample size increases and (2) when the true model or data-generating-process (DGP) is one of the defined regression models under consideration, then the probability of selecting the true model approaches one as the sample size increases.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

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:

Characterizing Cycles

Topic 26

Exam Focus

This topic focuses on ways to characterize a cycle in forecasting models. Along with seasonal and trend components, cycles constitute an essential third component in a forecasting model. Cyclicality captures the dynamics of a data series *outside* of trend or seasonal data. Thus, the complexity of cyclical dynamics demands a more robust forecasting model. For the exam, understand the concept of covariance stationary and the requirements for a time series to exhibit covariance stationarity. Also, be able to define a white noise process and know how a lag operator works. The concepts introduced in this topic serve as a foundation for the material in the next topic on modeling cycles.

Covariance Stationary

LO 26.1: Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function and autoregression.

A **time series** is a set of observations for a variable over successive periods of time (e.g., monthly stock market returns for the past 10 years). The series has a **trend** if a consistent pattern can be seen by plotting the data (i.e., the individual observations) on a graph. For example, a seasonal trend in sales data is easily detected by plotting the data and noting the significant jump in sales during the same month(s) each year. To forecast a time series, one needs to understand and characterize its structure. The following terminology relates to modeling data interrelationships and stability over time.

- **Autoregression** refers to the process of regressing a variable on lagged or past values of itself. As you will see in the next topic, when the dependent variable for a time series is regressed against one or more lagged values of itself, the resultant model is called as an **autoregressive (AR) model**. For example, the sales for a firm could be regressed against the sales for the firm in the previous month. Thus, in an autoregressive time series, past values of a variable are used to predict the current (and hence future) value of the variable.
- A time series is **covariance stationary** if its mean, variance, and covariances with lagged and leading values do not change over time. Covariance stationarity is a requirement for using AR models.
- **Autocovariance function** refers to the tool used to quantify stability of the covariance structure. Its importance lies in its ability to summarize cyclical dynamics in a series that is covariance stationary.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

- **Autocorrelation function** refers to the degree of correlation and interdependency between data points in a time series. It recognizes the fact that correlations lend themselves to clearer interpretation than covariances. Recall that the degree of correlation is measured on a continuum from –1 to 1, whereas degrees of covariance employ a much wider range, which can be unwieldy in determining levels of association.
- **Partial autocorrelation function** refers to the partial correlation and interdependency between data in a time series that measures the association between data in a series after controlling for the effects of lagged observations.

LO 26.2: Describe the requirements for a series to be covariance stationary.

A time series is covariance stationary if it satisfies the following three conditions:

1. *Constant and finite expected value.* The expected value of the time series is constant over time.

2. *Constant and finite variance.* The time series volatility around its mean (i.e., the distribution of the individual observations around the mean) does not change over time.

3. *Constant and finite covariance between values at any given lag.* The covariance of the time series with leading or lagged values of itself is constant.

LO 26.3: Explain the implications of working with models that are not covariance stationary.

Requirements for covariance stationarity of a time series, though strict in appearance, make allowances for many series that are not covariance stationary. This is achieved by working with models that provide special treatment to trend and seasonality components that are stationary, which allows the remaining, or residual, cyclical component to be covariance stationary.

Note that forecasting models whose "probabilistic nature" changes (i.e., lacks covariance stationarity) would not lend themselves well to predicting the future. Such a trait would make the process of characterizing a cycle difficult, if not impossible. However, a nonstationary series can be transformed to appear covariance stationary by using transformed data, such as growth rates.

White Noise

LO 26.4: Define white noise, describe independent white noise and normal (Gaussian) white noise.

LO 26.5: Explain the characteristics of the dynamic structure of white noise.

A time series process with a zero mean, constant variance, and no serial correlation is referred to as a **white noise** process (or *zero-mean white noise*). This is the simplest type of time series process and it is used as a fundamental building block for more complex time

©2015 Kaplan, Inc.

FRM官方交流群 392417118

series processes. Even though a white noise process is serially uncorrelated, it may not be serially independent or normally distributed.

Variants of a white noise process include independent white noise and normal white noise. A time series process that exhibits both serial independence and a lack of serial correlation is referred to as **independent white noise** (or *strong white noise*). A time series process that exhibits serial independence, is serially uncorrelated, and is normally distributed is referred to as **normal white noise** (or *Gaussian white noise*).

The dynamic structure of a white noise process includes the following characteristics:

- The unconditional mean and variance must be constant for any covariance stationary process.
- The lack of any correlation in white noise means that all autocovariances and autocorrelations are zero beyond displacement zero (displacement refers to the distance of a moving body from a central point). This same result holds for the partial autocorrelation function of white noise.
- Both conditional and unconditional means and variances are the same for an independent white noise process (i.e., they lack any forecastable dynamics).
- Events in a white noise process exhibit no correlation between the past and present.

Lag Operators

LO 26.6: Explain how a lag operator works.

A **lag operator** quantifies how a time series evolves by lagging a data series. It enables a model to express how past data links to the present and how present data links to the future. For example, a lag operator, L, operates on series, y_t, by lagging it as follows:

$$Ly_t = y_{t-1}$$

Another example of a common lag operator is a *first-difference operator* (Δ), which applies a polynomial in the lag operator as follows:

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$$

A key component of an operator is the **distributed lag**, which is a weighted sum of present and past values in a data series, achieved by lagging present values upon past values.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Wold's Representation Theorem

LO 26.7: Describe Wold's theorem.

LO 26.8: Define a general linear process.

LO 26.9: Relate rational distributed lags to Wold's theorem.

Wold's representation theorem is a model for the covariance stationary residual (i.e., a model that is constructed after making provisions for trends and seasonal components). Thus, the theorem enables the selection of the correct model to evaluate the evolution of covariance stationarity. Wold's representation utilizes an infinite number of distributed lags, where the one-step-ahead forecasted error terms are known as *innovations*.

The **general linear process** is a component in the creation of forecasting models in a covariance stationary time series. It uses Wold's representation to express innovations that capture an evolving information set. These evolving information sets move the conditional mean over time (recall that a requirement of stationarity is a constant unconditional mean). Thus, it can model the dynamics of a times series process that is outside of covariance stationarity (i.e., unstable).

As mentioned, applying Wold's representation requires an infinite number of distributed lags. However, it is not practical to model an infinite number of parameters. Therefore, we need to restate this lag model as infinite polynomials in the lag operator because infinite polynomials do not necessarily contain an infinite number of parameters. Infinite polynomials that are a ratio of finite-order polynomials are known as **rational polynomials**. The distributed lags constructed from these rational polynomials are known as **rational distributed lags**. With these lags, we can approximate Wold's representation. In the next topic, we'll examine the properties of an autoregressive moving average (ARMA) process, which is a practical approximation for Wold's representation.

Estimating the Mean and Autocorrelation Functions

LO 26.10: Calculate the sample mean and sample autocorrelation, and describe the Box-Pierce Q-statistic and the Ljung-Box Q-statistic.

LO 26.11: Describe sample partial autocorrelation.

Sample data for a time series forms the basis for estimating the sample mean and sample autocorrelation of a covariance stationary series. With these estimated parameters, an analyst can study the dynamics that underpin the dataset and find a model that best fits the data. Sample data can be used to estimate the sample mean and the sample autocorrelation.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

The **sample mean** is an approximation of the mean of the population and can be used to estimate the autocorrelation function. The sample mean, given a sample size of *T*, is computed as follows:

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

The **sample autocorrelation** estimates the degree to which white noise characterizes a series of data. Recall that for a time series to be classified as a white noise process, all autocorrelations must be zero in the population dataset. The sample autocorrelation, as a function of displacement τ, is computed as follows:

$$\hat{\rho}(\tau) = \frac{\sum_{t=\tau+1}^{T}\left[(y_t - \bar{y})(y_{t-\tau} - \bar{y})\right]}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Similar to sample autocorrelation, the **sample partial autocorrelation** can also be used to determine whether a time series exhibits white noise. It differs from sample autocorrelation in that it performs linear regression on a finite or feasible data series. However, the outcome of sample partial autocorrelation is typically identical to that achieved through sample autocorrelation. Sample partial autocorrelations usually plot within two-standard-error bands (i.e., 95% confidence interval) when the time series is white noise.

A **Q-statistic** can be used to measure the degree to which autocorrelations vary from zero and whether white noise is present in a dataset. This can be done by evaluating the overall statistical significance of the autocorrelations. This statistical measure is approximately chi-squared distributed with *m* degrees of freedom in large samples under the null hypothesis of no autocorrelations.

The **Box-Pierce Q-statistic** reflects the absolute magnitudes of the correlations, because it sums the squared autocorrelations. Thus, the signs do not cancel each other out, and large positive or negative autocorrelation coefficients will result in large Q-statistics. The **Ljung-Box Q-statistic** is similar to the Box-Pierce Q-statistic except that it replaces the sum of squared autocorrelations with a weighted sum of squared autocorrelations. For large sample sizes, weights for both statistics are roughly equal.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Key Concepts

LO 26.1

The terms covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression relate to the degree of data interrelationships and their stability. A time series is covariance stationary if its mean, variance, and covariances with lagged and leading values do not change over time.

LO 26.2

A time series is covariance stationary if it satisfies the following three conditions: (1) constant and finite expected value, (2) constant and finite variance, and (3) constant and finite covariance between values at any given lag.

LO 26.3

Models that lack covariance stationarity are unstable and do not lend themselves to meaningful forecasting.

LO 26.4

A time series process with a zero mean, constant variance, and no serial correlation is referred to as white noise. This is the simplest type of time series process and is used as a building block for more complex time series processes.

LO 26.5

The lack of any correlation in a white noise process means that all autocovariances and autocorrelations are zero beyond displacement zero. The past is not correlated with the present which, in turn, is not correlated with the future.

LO 26.6

A lag operator enables a forecasting model to express how past data links to the present and how present data links to the future.

LO 26.7

Wold's representation theorem evaluates covariance stationarity as a prerequisite for time series modeling. It utilizes an infinite number of distributed lags.

LO 26.8

The general linear process is intended to capture an information set that evolves.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

LO 26.9

The distributed lags constructed from rational polynomials are known as rational distributed lags. With these lags, Wold's representation can be approximated.

LO 26.10

Understanding the degree of data correlation and dynamics that underpin the dataset is critical to the characterization of a cycle. If white noise is present, then there should be no forecastable events. Q-statistics further refine the measurement of the degree to which autocorrelations vary from zero and whether white noise is present in the dataset.

LO 26.11

Sample partial autocorrelation is a somewhat simplified version of sample autocorrelation in that it uses a finite data series.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checkers

1. All of the following traits characterize the covariance stationarity of a time series process, except:
   A. stability of the mean.
   B. stability of the covariance structure.
   C. a nonconstant variance in the time series.
   D. stability of the autocorrelation.

2. Which of the following features correctly characterizes a white noise process?
   A. Conditional mean in the dataset.
   B. Minimal variance.
   C. No correlation between data points.
   D. Partial autocorrelations are greater than zero.

3. Which of the following statements is most likely correct regarding lag operators? Lag operators:
   A. consider only infinite-order polynomials.
   B. quantify how a time series evolves by lagging a data series.
   C. are of limited use in modeling a time series.
   D. only use lagged future values.

4. Regarding Q-statistics, the Box-Pierce and Ljung-Box Q-statistics:
   A. produce different results.
   B. are more accurate for smaller datasets.
   C. essentially yield the same result.
   D. both use an unweighted sum of squared autocorrelations.

5. Regarding sample partial autocorrelations, which of the following statements is true? A sample partial autocorrelation:
   A. is identical to sample autocorrelation.
   B. differs from sample autocorrelation in the size of the dataset to which it applies.
   C. utilizes non-linear regressions.
   D. typically falls within a one-standard-error band.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checker Answers

1. C The time series volatility around its mean (i.e., the distribution of the individual observations around the mean) does not change over time.

2. C The lack of any correlation in white noise means that all autocovariances and autocorrelations are zero.

3. B Lag operators may use finite-order polynomials and are an essential tool to model a time series. They quantify how a time series evolves by typically lagging present values upon past values.

4. C Both Q-statistics typically arrive at the same result. The Ljung-Box statistic works better with smaller samples of data and replaces the sum of squared autocorrelations in the Box-Pierce statistic with a weighted sum of squared autocorrelations.

5. B The linear regression that is part of the sample partial autocorrelation process takes place on a feasible data sample, which differs from the infinite data sample for partial autocorrelations. Sample partial autocorrelations should fall within two-standard-error (standard deviation) bands.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

Modeling Cycles: MA, AR, and ARMA Models

Topic 27

Exam Focus

Moving average (MA) processes can be used to capture the relationship between a time series variable and its current and lagged random shocks. This is useful for researchers if an event is mostly described by random shocks. However, it becomes even more useful when it is transformed into an autoregressive representation. An autoregressive (AR) process attempts to capture how a time series variable's lagged observations of itself combine with random shocks to forecast a variable. Sometimes forecasters need a combination of these two concepts to improve the usefulness of a forecasting model, which results in an autoregressive moving average model (ARMA). For the exam, understand the properties of an MA(1) process and an AR(1) process and how they can be broaden to incorporate additional lag operators. Also, be able to describe an ARMA process and understand its applications.

First-Order Moving Average Process

LO 27.1: Describe the properties of the first-order moving average (MA(1)) process, and distinguish between autoregressive representation and moving average representation.

Conceptually, a moving average process is a linear regression of the current values of a time series against both the current and previous unobserved white noise error terms, which are random shocks. The first-order moving average [MA(1)] process has a mean of zero and a constant variance and can be defined as:

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

where:

y_t = the time series variable being estimated
ε_t = current random white noise shock
ε_{t-1} = one-period lagged random white noise shock
θ = coefficient for the lagged random shock

The **MA(1) process** is considered to be first-order because it only has one lagged error term (ε_{t-1}). This yields a very short-term memory because it only incorporates what happens one period ago. If we ignore the lagged error term for a moment and assume that $\varepsilon_t > 0$, then $y_t > 0$. This is equivalent to saying that a positive error term will yield a positive dependent variable (y_t). When adding back the lagged error term, we are now saying that the dependent variable is impacted by not only the current error term, but also the previous

©2015 Kaplan, Inc.

FRM官方交流群 392417118

period's unobserved error term, which is amplified by a coefficient (θ). Consider an example using daily demand for ice cream (y_t) to better understand how this works:

$$y_t = \varepsilon_t + 0.3\varepsilon_{t-1}$$

In this equation, the error term (ε_t) is the daily change in temperature. Using only the current period's error term (ε_t), if the daily change in temperature is positive, then we would estimate that daily demand for ice cream would also be positive. But, if the daily change yesterday (ε_{t-1}) was also positive, then we would expect an amplified impact on our daily demand for ice cream by a factor of 0.3.

One key feature of moving average processes is called the *autocorrelation (ρ) cutoff*. We would compute the autocorrelation using the following formula:

$$\rho_1 = \frac{\theta_1}{1+\theta_1^2}; \text{ where } \rho_\tau = 0 \text{ for } \tau > 1$$

Using the previous example of estimating daily demand for ice cream with $\theta = 0.3$, we would compute the autocorrelation to be 0.2752 as follows:

$$0.2752 = \frac{0.3}{1+0.3^2}$$

For any value beyond the first lagged error term, the autocorrelation will be zero in an MA(1) process. This is important because it is one condition of being covariance stationary (i.e., mean = 0, variance = σ^2), which is a condition of this process being a useful estimator.

It is also important to note that this **moving average representation** has both a current random shock (ε_t) and a lagged unobservable shock (ε_{t-1}) on the independent side of this equation. This presents a problem for forecasting in the real world because it does not incorporate observable shocks. The solution for this problem is known as an **autoregressive representation** where the MA(1) process formula is inverted so we have a lagged shock and a lagged value of the time series itself. The condition for inverting an MA(1) process is $|\theta| < 1$. The autoregressive representation, which is an algebraic rearrangement of the MA(1) process formula, is expressed in the following formula:

$$\varepsilon_t = y_t - \theta\varepsilon_{t-1}$$

This process of inversion enables the forecaster to express current observables in terms of past observables.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

MA(q) Process

LO 27.2: Describe the properties of a general finite-order process of order q (MA(q)) process.

The MA(1) process is a subset of a much larger picture. Forecasters can broaden their horizon to a finite-order moving average process of order *q*, which essentially adds lag operators out to the qth observation and potentially improves on the MA(1) process. The MA(q) **process** is expressed in the following formula:

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}$$

where:

y_t = the time series variable being estimated
ε_t = current random white noise shock
ε_{t-1} = one-period lagged random white noise shock
ε_{t-q} = q^{th}-period lagged random white noise shock
θ = coefficients for the lagged random shocks

The MA(q) process theoretically captures complex patterns in greater detail, which can potentially provide for more robust forecasting. This also lengthens the memory from one period to the qth period. Returning to the previous example, using the demand for ice cream, a forecaster could use not only the current and previous day's changes in temperature to predict ice cream demand, but also the entire previous week's demand to enhance the informational value of the estimation.

Just as the MA(1) process exhibits autocorrelation cutoff after the first lagged error term, the MA(q) process experiences autocorrelation cutoff after the qth lagged error term. Again, this is important because covariance stationarity is essential to the predictive ability of the model.

First-Order Autoregressive Process

LO 27.3: Describe the properties of the first-order autoregressive (AR(1)) process, and define and explain the Yule-Walker equation.

We have seen that when a moving average process is inverted it becomes an autoregressive representation, and is, therefore, more useful because it expresses the current observables in terms of past observables. An autoregressive process does not need to be inverted because it is already in the more favorable rearrangement, and is, therefore, capable of capturing a more robust relationship compared to the unadjusted moving average process. The **first-order autoregressive [AR(1)] process** must also have a mean of zero and a constant variance.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

It is specified in the form of a variable regressed against itself in a lagged form. This relationship can be shown in the following formula:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where:
y_t = the time series variable being estimated
y_{t-1} = one-period lagged observation of the variable being estimated
ε_t = current random white noise shock
ϕ = coefficient for the lagged observation of the variable being estimated

Just like the moving average process, the predictive ability of this model hinges on it being covariance stationary. In order for an AR(1) process to be covariance stationary, the absolute value of the coefficient on the lagged operator must be less than one (i.e., $|\phi| < 1$).

Using our previous example of daily demand for ice cream, we would forecast our current period daily demand (y_t) as a function of a coefficient (ϕ) multiplied by our lagged daily demand for ice cream (y_{t-1}) and then add a random error shock (ε_t). This process enables us to use a past observed variable to predict a current observed variable.

In order to estimate the autoregressive parameters, such as the coefficient (ϕ), forecasters need to accurately estimate the autocovariance of the data series. The **Yule-Walker equation** is used for this purpose. When using the Yule-Walker concept to solve for the autocorrelations of an AR(1) process, we use the following relationship:

$$\rho_t = \phi^t \text{ for } t = 0,1,2,...$$

The Yule-Walker equation is used to reinforce a very important distinction between autoregressive processes and moving average processes. Recall that moving average processes exhibit autocorrelation cutoff, which means the autocorrelations are essentially zero beyond the order of the process [an MA(1) process shows autocorrelation cutoff after time 1]. The significance of the Yule-Walker equation is that for autoregressive processes, the autocorrelation decays vary gradually. Consider an AR(1) process that is specified using the following formula:

$$y_t = 0.65y_{t-1} + \varepsilon_t$$

The coefficient (ϕ) is equal to 0.65, and using the concept derived from the Yule-Walker equation, the first-period autocorrelation is 0.65 (i.e., 0.65^1), the second-period autocorrelation is 0.4225 (i.e., 0.65^2), and so on for the remaining autocorrelations.

Professor's Note: While autocorrelation cutoff is a hallmark of moving average processes, a gradual decay in autocorrelations is a sure sign that a forecaster is dealing with an autoregressive process.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

It should also be noted that if the coefficient (ϕ) were to be a negative number, perhaps –0.65, then the decay would still occur but the graph would oscillate between negative and positive numbers. This is true because $-0.65^3 = -0.2746$, $-0.65^4 = 0.1785$, and $-0.65^5 = -0.1160$. You would still notice the absolute value decaying, but the actual autocorrelations would alternate between positive and negative numbers over time.

AR(p) Process

LO 27.4: Describe the properties of a general p^{th} order autoregressive (AR(p)) process.

Just as the MA(1) process was described as a subset of the much broader MA(q) process, so is the relationship between the AR(1) process and the AR(p) process. The AR(p) process expands the AR(1) process out to the pth observation as seen in the following formula:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

where:
y_t = the time series variable being estimated
y_{t-1} = one-period lagged observation of the variable being estimated
y_{t-p} = p^{th}-period lagged observation of the variable being estimated
ε_t = current random white noise shock
ϕ = coefficients for the lagged observations of the variable being estimated

The AR(p) process is also covariance stationary if $|\phi| < 1$ and it exhibits the same decay in autocorrelations that was found in the AR(1) process. However, while an AR(1) process only evidences oscillation in its autocorrelations (switching from positive to negative) when the coefficient is negative, an AR(p) process will naturally oscillate as it has multiple coefficients interacting with each other.

Autoregressive Moving Average Process

LO 27.5: Define and describe the properties of the autoregressive moving average (ARMA) process.

So far, we have examined moving average processes and autoregressive processes assuming they interact independently of each other. While this may be the case, it is possible for a time series to show signs of both processes and theoretically capture a still richer relationship. For example, stock prices might show evidence of being influenced by both unobserved shocks (the moving average component) and their own lagged behavior (the

©2015 Kaplan, Inc.

FRM官方交流群 392417118

autoregressive component). This more complex relationship is called an **autoregressive moving average (ARMA)** process and is expressed by the following formula:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

where:
y_t = the time series variable being estimated
ϕ = coefficient for the lagged observations of the variable being estimated
y_{t-1} = one-period lagged observation of the variable being estimated
ε_t = current random white noise shock
θ = coefficient for the lagged random shocks
ε_{t-1} = one-period lagged random white noise shock

You can see that the ARMA formula merges the concepts of an AR process and an MA process. In order for the ARMA process to be covariance stationary, which is important for forecasting, we must still observe $|\theta| < 1$. Just as with the AR process, the autocorrelations in an ARMA process will also decay gradually for essentially the same reasons.

Consider an example regarding sales of an item (y_t) and a random shock of advertising (ε_t). We could attempt to forecast sales for this item as a function of the previous period's sales (y_{t-1}), the current level of advertising (ε_t), and the one-period lagged level of advertising (ε_{t-1}). It makes intuitive sense that sales in the current period could be affected by both past sales and by random shocks, such as advertising. Another possible random shock for sales could be the seasonal effects of weather conditions.

Professor's Note: Just as moving average models can be extrapolated to the qth observation and autoregressive models can be taken out to the pth observation, ARMA models can be used in the format of an ARMA(p,q) model. For example, an ARMA(3,1) model means 3 lagged operators in the AR portion of the formula and 1 lagged operator on the MA portion. This flexibility provides the highest possible set of combinations for time series forecasting of the three models discussed in this topic.

Application of AR and ARMA Processes

LO 27.6: Describe the application of AR and ARMA processes.

A forecaster might begin by plotting the autocorrelations for a data series and find that the autocorrelations decay gradually rather than cut off abruptly. In this case, the forecaster should rule out using a moving average process. If the autocorrelations instead decay gradually, he should consider specifying either an autoregressive (AR) process or an autoregressive moving average (ARMA) process. The forecaster should especially consider these alternatives if he notices periodic spikes in the autocorrelations as they are gradually decaying. For example, if every 12th autocorrelation jumps upward, this observation indicates a possible seasonality effect in the data and would heavily point toward using either an AR or ARMA model.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Another way of looking at model applications is to test various models using regression results. It is easiest to see the differences using data that follows some pattern of seasonality, such as employment data. In the real world, a moving average process would not specify a very robust model, and autocorrelations would decay gradually, so forecasters would be wise to consider both AR models and ARMA models for employment data.

We could begin with a base AR(2) model that adds in a constant value (μ) if all other values are zero. This is shown in the following generic formula:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Applying actual coefficients, our real AR(2) model might look something like:

$$y_t = 101.2413 + 1.4388 y_{t-1} - 0.4765 y_{t-2} + \varepsilon_t$$

We could also try to forecast our seasonally impacted employment data with an ARMA(3,1) model, which might look like the following formula:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \theta \varepsilon_{t-1} + \varepsilon_t$$

Applying actual coefficients our real ARMA(3,1) model might look something like:

$$y_t = 101.1378 + 0.5004 y_{t-1} + 0.8722 y_{t-2} - 0.4434 y_{t-3} + 0.9709 \varepsilon_{t-1} + \varepsilon_t$$

In practice, researchers would attempt to determine whether the AR(2) model or the ARMA(3,1) model provides a better prediction for the seasonally impacted data series.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 27.1

The first-order moving average process enables forecasters to consider the likely current effect on a dependent variable of current and lagged white noise error terms. While this is a useful process, it is most useful when inverted as an autoregressive representation so that current observables can be explained in terms of past observables.

LO 27.2

While the first-order moving average process does provide useful information for forecasting, the *q*th-order moving average process allows for a richer analysis because it incorporates significantly more lagged error terms all the way out to the order of *q*.

LO 27.3

The first-order autoregressive process incorporates the benefits of an inverted MA(1) process. Specifically, the AR(1) process seeks to explain the dependent variable in terms of a lagged observation of itself and an error term. This is a better forecasting tool if the autocorrelations decay gradually rather than cut off immediately after the first observation with a first-order process.

LO 27.4

The *p*th-order autoregressive process adds additional lagged observations of the dependent variable and enhances the informational value relative to an AR(1) process in much the same way that an MA(q) process adds a richer explanation to the MA(1) process.

LO 27.5

The autoregressive moving average (ARMA) process has the potential to capture more robust relationships. The ARMA process incorporates the lagged error elements of the moving average process and the lagged observations of the dependent variable from the autoregressive process.

LO 27.6

Both autoregressive (AR) and autoregressive moving average (ARMA) processes can be applied to time series data that show signs of seasonality. Seasonality is most apparent when the autocorrelations for a data series do not abruptly cut off, but rather decay gradually with periodic spikes.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checkers

1. In practice, the moving average representation of a first-order moving average [MA(1)] process presents a problem. Which of the following statements represents that problem and how can it be resolved? The problem is that a moving average representation of an MA(1) process:
   A. does not incorporate observable shocks, so the solution is to use a moving average representation.
   B. incorporates only observable shocks, so the solution is to use a moving average representation.
   C. does not incorporate observable shocks, so the solution is to use an autoregressive representation.
   D. incorporates only observable shocks, so the solution is to use an autoregressive representation.

2. Which of the following statements is a key differentiator between a moving average (MA) representation and an autoregressive (AR) process?
   A. A moving average representation shows evidence of autocorrelation cutoff.
   B. An autoregressive process shows evidence of autocorrelation cutoff.
   C. An unadjusted moving average process shows evidence of gradual autocorrelation decay.
   D. An autoregressive process is never covariance stationary.

3. The purpose of a qth-order moving average process is to:
   A. add exactly two additional lagged variables to the original specification.
   B. add a second error term to an MA(1) process.
   C. invert the moving average process to make the formula more useful.
   D. add as many additional lagged variables as needed to more robustly estimate the data series.

4. Which of the following statements about an autoregressive moving average (ARMA) process is correct?
   I. It involves autocorrelations that decay gradually.
   II. It combines the lagged unobservable random shock of the MA process with the observed lagged time series of the AR process.
   A. I only.
   B. II only.
   C. Both I and II.
   D. Neither I nor II.

5. Which of the following statements is correct regarding the usefulness of an autoregressive (AR) process and an autoregressive moving average (ARMA) process when modeling seasonal data?
   I. They both include lagged terms and, therefore, can better capture a relationship in motion.
   II. They both specialize in capturing only the random movements in time series data.
   A. I only
   B. II only.
   C. Both I and II.
   D. Neither I nor II.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checker Answers

1. C The problem with a moving average representation of an MA(1) process is that it attempts to estimate a variable in terms of unobservable white noise random shocks. If the formula is inverted into an autoregressive representation, then it becomes more useful for estimation because an observable item is now being used.

2. A A key difference between a moving average (MA) representation and an autoregressive (AR) process is that the MA process shows autocorrelation cutoff while an AR process shows a gradual decay in autocorrelations.

3. D The whole point of using more independent variables in a qth-order moving average process is to capture a better estimation of the dependent variable. More lagged operators often provide a more robust estimation.

4. C The autoregressive moving average (ARMA) process is important because its autocorrelations decay gradually and because it captures a more robust picture of a variable being estimated by including both lagged random shocks and lagged observations of the variable being estimated. The ARMA model merges the lagged random shocks from the MA process and the lagged time series variables from the AR process.

5. A Both autoregressive (AR) models and autoregressive moving average (ARMA) models are good at forecasting with seasonal patterns because they both involve lagged observable variables, which are best for capturing a relationship in motion. It is the moving average representation that is best at capturing only random movements.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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:

Estimating Volatilities and Correlations for Risk Management

Topic 28

Exam Focus

Traditionally, volatility has been synonymous with risk, and low correlation has been synonymous with risk reduction through diversification. The accurate estimation of volatility and correlation is crucial to understanding potential risk exposure. This topic pertains to methods that employ historical data when generating estimates of volatility and correlation. The more simplistic models tend to generate estimates assuming variance and correlation remain constant over short time periods. Conversely, the more sophisticated models account for variations over time.

Estimating Volatility

LO 28.1: Explain how various weighting schemes can be used in estimating volatility.

The traditional measure of volatility requires first a measure of change in asset value from period to period. The calculation of a continuously compounded return over successive days is:

$$u_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$$

where:
S_i = asset price at time i

The mean return of these individual returns is calculated as:

$$\bar{u} = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}$$

where:
m = number of observations leading up to the present period

If we assume that the mean return is zero, which would be true when the mean is small compared to the variability, we obtain the maximum likelihood estimator of variance:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

In simplest terms, historical data is used to generate returns in an asset-pricing series. This historical return information is then used to generate a volatility parameter, which can be used to infer expected realizations of risk. However, the straightforward approaches just presented weight each observation equally in that more distant past returns have the same influence on estimated volatility as observations that are more recent. If the goal is to estimate the current level of volatility, we may want to weight recent data more heavily. There are various weighting schemes, which can all essentially be represented as:

$$\sigma_n^2 = \sum_{i=1}^{m} \alpha_i u_{n-i}^2$$

where:
α_i = weight on the return i days ago

The weights (the α's) must sum to one, and if the objective is to generate a greater influence on recent observations, then the α's will decline in value for older observations.

One extension to this weighting scheme is to assume a long-run variance level in addition to the weighted squared return observations. The most frequently used model is an **autoregressive conditional heteroskedasticity model,** ARCH(m), which can be represented by:

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2 \text{ with } \gamma + \sum \alpha_i = 1 \text{ so that}$$

$$\sigma_n^2 = \omega + \sum_{i=1}^{m} \alpha_i u_{n-i}^2$$

where:
$\omega = \gamma V_L$ (long-run variance weighted by the parameter γ)

Therefore, the volatility estimate is a function of a long-run variance level and a series of squared return observations, whose influence declines the older the observation is in the time series of the data.

The Exponentially Weighted Moving Average Model

LO 28.2: Apply the exponentially weighted moving average (EWMA) model to estimate volatility.

LO 28.4: Explain the weights in the EWMA and GARCH(1,1) models.

The **exponentially weighted moving average** (EWMA) model is a specific case of the general weighting model presented in the previous section. The main difference is that the weights are assumed to decline exponentially back through time. This assumption results in a specific relationship for variance in the model:

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2$$

where:
λ = weight on previous volatility estimate (λ between zero and one)

©2015 Kaplan, Inc.
FRM官方交流群 392417118

The simplest interpretation of the EWMA model is that the day-n volatility estimate is calculated as a function of the volatility calculated as of day $n - 1$ and the most recent squared return. Depending on the weighting term λ, which ranges between zero and one, the previous volatility and most recent squared returns will have differential impacts. High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

Example: EWMA model

The decay factor in an exponentially weighted moving average model is estimated to be 0.94 for daily data. Daily volatility is estimated to be 1%, and today's stock market return is 2%. What is the new estimate of volatility using the EWMA model?

Answer:

$$\sigma_n^2 = 0.94 \times 0.01^2 + (1 - 0.94) \times 0.02^2 = 0.000118$$

$$\sigma_n = \sqrt{0.000118} = 1.086\%$$

One benefit of the EWMA is that it requires few data points. Specifically, all we need to calculate the variance is the current estimate of the variance and the most recent squared return. The current estimate of variance will then feed into the next period's estimate, as will this period's squared return. Technically, the only "new" piece of information for the volatility calculation will be that attributed to the squared return.

The GARCH(1,1) Model

LO 28.3: Describe the generalized autoregressive conditional heteroskedasticity (GARCH(p,q)) model for estimating volatility and its properties:

- **Calculate volatility using the GARCH(1,1) model**
- **Explain mean reversion and how it is captured in the GARCH(1,1) model**

One of the most popular methods of estimating volatility is the **generalized autoregressive conditional heteroskedastic** (GARCH)(1,1) model. A GARCH(1,1) model not only incorporates the most recent estimates of variance and squared return, but also a variable that accounts for a long-run average level of variance.

Professor's Note: In the GARCH(p,q) notation, the p stands for the number of lagged terms on historical returns squared, and the q stands for the number of lagged terms on historical volatility.

FRM官方交流群 392417118

The best way to describe a GARCH(1,1) model is to take a look at the formula representing its determination of variance, which can be shown as:

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

where:

α = weighting on the previous period's return

β = weighting on the previous volatility estimate

ω = weighted long-run variance $= \gamma V_L$

V_L = long-run average variance $= \dfrac{\omega}{1-\alpha-\beta}$

$\alpha + \beta + \gamma = 1$

$\alpha + \beta < 1$ for stability so that γ is not negative

The EWMA is nothing other than a special case of a GARCH(1,1) volatility process, with $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$. Similar to the EWMA model, β represents the exponential decay rate of information. The GARCH(1,1) model adds to the information generated by the EWMA model in that it also assigns a weighting to the average long-run variance estimate. An additional characteristic of a GARCH(1,1) estimate is the implicit assumption that variance tends to revert to a long-term average level. Recognition of a mean-reverting characteristic in volatility is an important feature when pricing derivative securities such as options.

Example: GARCH(1,1) model

The parameters of a generalized autoregressive conditional heteroskedastic (GARCH)(1,1) model are $\omega = 0.000003$, $\alpha = 0.04$, and $\beta = 0.92$. If daily volatility is estimated to be 1%, and today's stock market return is 2%, what is the new estimate of volatility using the GARCH(1,1) model, and what is the implied long-run volatility level?

Answer:

$$\sigma_n^2 = 0.000003 + 0.04 \times 0.02^2 + 0.92 \times 0.01^2 = 0.000111$$

$$\sigma_n = \sqrt{0.000111} = 1.054\%$$

$$\text{long-run average variance} = \frac{\omega}{(1-\alpha-\beta)} = \frac{0.000003}{(1-0.04-0.92)} = 0.000075$$

$$\bar{\sigma} = \sqrt{0.000075} = 0.866\% = \text{long-run volatility}$$

Mean Reversion

Empirical data indicates that volatility exhibits a mean-reverting characteristic. Given that stylized fact, a GARCH model tends to display a better theoretical justification than the EWMA model. The method for estimating the GARCH parameters (or weights), however, often generates outcomes that are not consistent with the model's assumptions. Specifically,

©2015 Kaplan, Inc.
FRM官方交流群 392417118

the sum of the weights of α and β are sometimes greater than one, which causes instability in the volatility estimation. In this case, the analyst must resort to using an EWMA model.

The sum $\alpha + \beta$ is called the **persistence**, and if the model is to be stationary over time (with reversion to the mean), the sum must be less than one. The persistence describes the rate at which the volatility will revert to its long-term value following a large movement. The higher the persistence (given that it is less than one), the longer it will take to revert to the mean following a shock or large movement. A persistence of one means that there is no reversion, and with each change in volatility, a new level is attained.

Estimation and Performance of GARCH Models

As was previously mentioned, one way to estimate volatility (e.g., variance) is to use a **maximum likelihood estimator**. Maximum likelihood estimators select values of model parameters that maximize the likelihood that the observed data will occur in a sample. Any variable of interest can be estimated via the maximum likelihood method, which requires formulating an expression or function for the underlying probability distribution of the data and then searching for the parameters that maximize the value generated by the expression. One important consideration relates to which distribution is chosen when calculating probability. The most popular is the normal distribution, but normally distributed data are not often found in financial markets.

GARCH models are estimated using maximum likelihood techniques. The estimation process begins with a guess of the model's parameters. Then a calculation of the likelihood function based on those parameter estimates is made. The parameters are then slightly adjusted until the likelihood function fails to increase, at which time the estimation process assumes it has maximized the function and stops. The values of the parameters at the point of maximum value in the likelihood function are then used to estimate GARCH model volatility.

LO 28.5: Explain how GARCH models perform in volatility forecasting.

LO 28.6: Describe the volatility term structure and the impact of volatility changes.

One of the useful features of GARCH models is that they do a very good job at modeling volatility clustering when periods of high volatility tend to be followed by other periods of high volatility and periods of low volatility tend to be followed by subsequent periods of low volatility. Thus, there is autocorrelation in u_i^2. If GARCH models do a good job of explaining volatility changes, there should be very little autocorrelation in u_i^2 / σ_i^2. GARCH models appear to do a very good job of explaining volatility.

The question then arises, if GARCH models do a good job at explaining past volatility, how well do they forecast future volatility? The simple answer to this question is that GARCH models do a fine job at forecasting volatility from a volatility term structure perspective (e.g., estimates of volatility given time to expiration for options). Even though the actual volatility term structure figures are somewhat different from those forecasted by GARCH models, GARCH-generated volatility data does an excellent job in predicting

©2015 Kaplan, Inc.

FRM官方交流群 392417118

how the volatility term structure responds to changes in volatility. This modeling tool is quite frequently used by financial institutions when estimating exposure to various option positions.

Estimating Correlations

LO 28.7: Describe how correlations and covariances are calculated, and explain the consistency condition for covariances.

So far, our discussion has focused on estimation of volatility. Generating accurate correlation estimates is also important in formulating portfolio risk exposures.

Correlation is a standardized measure of the covariance between two asset returns. The particular standardization is done with each asset's volatility estimates, which are subject to all the estimation issues previously discussed. The formula for calculating the correlation of returns between Asset X and Asset Y is:

$$\frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

Covariance can be calculated by the following formula:

$$\text{cov}_n = \frac{1}{m}\sum_{i=1}^{m} x_{n-i} y_{n-i}$$

where:
x_{n-i} = Asset X return in period $n-i$
y_{n-i} = Asset Y return in period $n-i$
m = number of observations leading up to the present period

Just like the straightforward estimation of variance, this covariance estimator assumes that past observations receive the same weighting. One alternative is to apply differential weightings to the past observations and create EWMA-type and GARCH-type models to take into account the time variability of covariance. The main factor to consider when calculating covariance, however, is to maintain consistency in the calculation process. That is, if one uses a particular weighting scheme (e.g., EWMA or GARCH) for variance calculations, one should use the same weighting scheme and procedure for the covariance calculations. Otherwise, the estimates are inconsistent with each other.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Key Concepts

LO 28.1

Historical price data is used to generate return estimates, which are then used to estimate volatility. Traditional volatility estimation methods weight past information equally across time. Weighting schemes can be used to weight recent information more heavily than distant data.

LO 28.2

The EWMA model generates volatility estimates based on weightings of the last estimate of volatility and the latest current price change information. The objective is to account for previous volatility estimates, as well as to account for the latest return information.

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

where:
λ = weight on previous volatility estimate (λ between zero and one)

LO 28.3

GARCH(1,1) models not only incorporate the most recent estimates of volatility and return, but also incorporate a long-run average level of variance.

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

where:

α = weighting on the previous period's return

β = weighting on the previous volatility estimate

ω = weighted long-run variance $= \gamma V_L$

V_L = long-run average variance $= \dfrac{\omega}{1-\alpha-\beta}$

$\alpha + \beta + \gamma = 1$

$\alpha + \beta < 1$ for stability so that γ is not negative

GARCH(1,1) estimates of volatility have a better theoretical justification than the EWMA model. In the event that model parameter estimates indicate instability, however, EWMA volatility estimates may be used.

LO 28.4

The EWMA is nothing other than a special case of a GARCH(1,1) volatility process, with $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$. Similar to the EWMA model, β in the GARCH(1,1) equation represents the exponential decay rate of information. The GARCH(1,1) model adds to the information generated by the EWMA model in that it also assigns a weighting to the average long-run variance estimate.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

LO 28.5

GARCH models do a very good job at modeling volatility clustering when periods of high volatility tend to be followed by other periods of high volatility and periods of low volatility tend to be followed by subsequent periods of low volatility.

LO 28.6

When forecasting future volatility, GARCH-generated volatility data does an excellent job in predicting the volatility term structure (i.e., differing volatilities for options given differing maturities). This modeling tool is quite frequently used by financial institutions when estimating exposure to various option positions.

LO 28.7

Correlation and covariance estimates provide additional information when formulating portfolio risk exposures. To maintain consistency in the estimation process, covariance parameters should be estimated in the same manner as that used to estimate volatility.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Concept Checkers

1. The parameters of a generalized autoregressive conditional heteroskedastic (GARCH)(1,1) model are $\omega = 0.00003$, $\alpha = 0.04$, and $\beta = 0.92$. If daily volatility is estimated to be 1.5%, and today's stock market return is 0.8%, what is the new estimate of the standard deviation?
   A. 1.68%.
   B. 1.55%.
   C. 1.45%.
   D. 2.74%.

2. The λ of an exponentially weighted moving average (EWMA) model is estimated to be 0.9. Daily standard deviation is estimated to be 1.5%, and today's stock market return is 0.8%. What is the new estimate of the standard deviation?
   A. 1.68%.
   B. 1.55%.
   C. 1.45%.
   D. 2.74%.

3. The parameters of a GARCH(1,1) model are $\omega = 0.00003$, $\alpha = 0.04$, and $\beta = 0.92$. These figures imply a long-run daily standard deviation of:
   A. 1.68%.
   B. 1.55%.
   C. 1.45%.
   D. 2.74%.

4. GARCH(1,1) models can only be used to estimate volatility in the case where:
   A. $\alpha + \beta > 0$.
   B. $\alpha + \beta < 1$.
   C. $\alpha > \beta$.
   D. $\alpha < \beta$.

5. If the volatilities of two variables are estimated using a GARCH(1,1) model, which of the following models used to estimate covariance will generate a consistent correlation estimate between the two variables?
   A. GARCH(1,1) model.
   B. EWMA model.
   C. Unweighted historical volatility model.
   D. Geometrically weighted historical volatility model.

最新视频课程 QQ: 2211892999

FRM官方交流群 392417118

Concept Checker Answers

1. **B** $\sigma_n^2 = 0.00003 + (0.008)^2 \times 0.04 + (0.015)^2 \times 0.92 = 0.00023956$

   $\sigma_n = \sqrt{0.00023956} = 0.0155 = 1.55\%$

2. **C** $\sigma_n^2 = 0.9 \times (0.015)^2 + (1 - 0.9) \times (0.008)^2 = 0.0002089$

   $\sigma_n = \sqrt{0.0002089} = 0.0145 = 1.45\%$

3. **D** The long-run variance rate can be estimated by dividing the ω of a GARCH(1,1) model by $1 - \alpha - \beta$. This yields 0.00003 / (1 – 0.04 – 0.92) = 0.00075; long-run standard deviation = $\sqrt{0.00075} = 0.0274 = 2.74\%$.

4. **B** Stable GARCH(1,1) models require $\alpha + \beta < 1$; otherwise the model is unstable.

5. **A** To maintain consistency, covariance should be estimated using the same method as that used for the volatility estimation. Since the volatility of each variable was estimated using the GARCH(1,1) model, the covariance estimate should also use the same technique.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

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:

Simulation Methods

Topic 29

Exam Focus

Simulation methods model uncertainty by generating random inputs that are assumed to follow an appropriate probability distribution. This topic discusses the basic steps for conducting a Monte Carlo simulation and compares this simulation method to the bootstrapping technique. For the exam, be able to explain ways to reduce Monte Carlo sampling error, including the use of antithetic and control variates. Also, understand the pseudo-random number generation method and the benefits of reusing sets of random number draws in Monte Carlo experiments. Finally, be able to describe the advantages and disadvantages of the bootstrapping technique in comparison to the traditional Monte Carlo approach.

Monte Carlo Simulation

LO 29.1: Describe the basic steps to conduct a Monte Carlo simulation.

Monte Carlo simulations are often used to model complex problems or to estimate variables when there are small sample sizes. A few practical finance applications of Monte Carlo simulations are: pricing exotic options, estimating the impact to financial markets of changes in macroeconomic variables, and examining capital requirements under stress-test scenarios.

There are four basic steps required to conduct a Monte Carlo simulation.

Step 1: Specify the data generating process (DGP)
Step 2: Estimate an unknown variable or parameter
Step 3: Save the estimate from step 2
Step 4: Go back to step 1 and repeat this process N times

The first step of conducting a simulation requires generating random inputs that are assumed to follow a specific probability distribution. The DGP could be a simple time series model or a more complex full structural model that requires multiple DGPs.

The second step of the simulation generates scenarios or trials based on randomly generated inputs drawn from a pre-specified probability distribution. The most common probability distribution used is the standard normal distribution. However, Student's t distribution is often used if the user believes it is a better fit for the data. A well-defined simulation model requires the generation of variables that follow appropriate probability distributions.

The last two steps in the simulation process allow for data analysis related to the properties of the probability distributions of the output variables. In other words, rather than making

©2015 Kaplan, Inc.

FRM官方交流群 392417118

just one output estimate for a problem, the model generates a probability distribution of estimates. This provides the user with a better understanding of the range of possible outcomes. The quantity *N* in step four is the number of times the simulation is repeated. This is referred to as the number of replications or iterations and is typically 1,000 to 10,000 times depending on how costly it is to generate the sample size.

For example, suppose we are managing an investment portfolio and desire to estimate the ending capital in the portfolio in one year, C_1. The initial capital investment, C_0, is $100 invested in the Standard & Poor's 500 index (S&P 500). The return is a random variable that depends on how the market performs over the next year.

If we assume the return over the next year is equal to a historical mean return, we can calculate one point estimate of the ending capital based on the equation: $C_1 = C_0(1 + r)$. The return over the next period is a random variable, and a simulation model estimates multiple scenarios to represent future returns based on a probability distribution of possible outcomes. The output variable is an estimate of an ending amount of capital that is also a random variable. The simulation model allows us to visualize the output and analyze the probability distribution of the ending capital amounts generated by the model.

Reducing Monte Carlo Sampling Error

LO 29.2: Describe ways to reduce Monte Carlo sampling error.

The sampling variation for a Monte Carlo simulation is quantified as the standard error estimate. The standard error of the true expected value is computed as $s / \sqrt{N}$, where *s* is the standard deviation of the output variables and *N* is the number of scenarios or replications in the simulation. Based on this equation, it intuitively follows that in order to reduce the standard error estimate by a factor of 10, the analyst must increase *N* by a factor of 100. (Because the square root of 100 is 10, if we increase the sample size 100 times it will reduce the standard error estimate by dividing by 10.)

Suppose we continue the illustration from the previous example and run a simulation to estimate the ending capital amount for an initial investment portfolio of $100. The number of replications is initially 100 (i.e., N = 100), resulting in a mean ending capital of $110 and a standard deviation of $14.798. For this example, the standard error estimate is computed as $1.4798 (i.e., $14.798 / 10). Now, suppose we want to increase the accuracy by reducing the standard error estimate. How can we increase the accuracy of the simulation?

The accuracy of simulations depends on the standard deviation and the number of scenarios run. We cannot control the standard deviation, but we can control the number of replications. Assume we rerun the previous simulation with 400 replications that result in the same mean ending capital of $110, and the standard deviation remains at $14.798. The standard error estimate for the simulation with 400 replications is then $0.7399 (i.e., 14.798 / 20). With four times the number of scenarios (4 × N, or 400, in this example) the standard error estimate is cut in half to $0.7399. In other words, quadrupling the number of scenarios will improve the accuracy twofold.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

However, increasing the number of generated scenarios can become costly for more complex multi-period simulations. Variance reduction techniques offer an alternative way to reduce the sampling error of a Monte Carlo simulation. The two most commonly used techniques for reducing the standard error estimate are antithetic variates and control variates.

Antithetic Variates

LO 29.3: Explain how to use antithetic variate technique to reduce Monte Carlo sampling error.

One reason sampling error occurs is because there are often a wide range of possible outcomes for a particular experiment or problem. Thus, in order to replicate the entire range of possible outcomes the sampling sets must be recreated numerous times. However, increasing the number of samples drawn may be too costly and time consuming. As an alternative approach, the **antithetic variate technique** can reduce Monte Carlo sampling error by rerunning the simulation using a *complement* set of the original set of random variables.

If the original set of random draws is denoted u_t for each replication, then the simulation is rerun with the complement set of random numbers denoted $-u_t$. By definition, the use of antithetic variates results in a lower covariance and variance, because the two sets are perfectly negatively correlated [i.e., $\text{corr}(u_t, -u_t) = -1$]. The following example illustrates how the standard error for a Monte Carlo simulation is reduced by using the antithetic variate technique.

First, consider a simulation of two sets that does not use the antithetic variate technique. Suppose the average parameter estimate is determined by two Monte Carlo simulations using different random sample sets. The average output parameter value, $\bar{x}$, for the two simulations using different random sample replications is simply calculated as:

$$\bar{x} = (x_1 + x_2)/2$$

Where x_1 and x_2 are the average output parameter values for simulation sets 1 and 2, respectively.

Next, we can calculate the variance of the average of the two sets as follows:

$$\text{var}(\bar{x}) = \frac{\text{var}(x_1) + \text{var}(x_2) + 2\,\text{cov}(x_1, x_2)}{4}$$

Without using antithetic variates, the two sets of Monte Carlo replications are independent. Thus, the covariance will be zero and the variance of $\bar{x}$ is simply reduced to the following:

$$\text{var}(\bar{x}) = \frac{\text{var}(x_1) + \text{var}(x_2)}{4}$$

The use of antithetic variates results in a negative covariance between the original random draws and their complements (i.e., antithetic variates). Thus, the use of antithetic variates

FRM官方交流群 392417118

causes the error terms to be independent for the two sets, which results in a negative covariance term in the variance equation. This negative relationship means that the Monte Carlo sampling error must always be smaller using this approach.

Control Variates

LO 29.4: Explain how to use control variates to reduce Monte Carlo sampling error and when it is effective.

The **control variate technique** is a widely used method to reduce the sampling error in Monte Carlo simulations. A control variate involves replacing a variable x (under simulation) that has unknown properties with a similar variable y that has known properties.

Suppose two separate simulations are conducted on variable x with unknown properties and control variable y with known properties using the same set of random numbers. Also assume that the Monte Carlo simulation estimated variables for x and y are denoted as $\hat{x}$ and $\hat{y}$, respectively. The original estimate x can be redefined as x* as follows:

$$x^* = y + (\hat{x} - \hat{y})$$

The new x* variable estimate will have a smaller sampling error than the original x variable if the control statistic and statistic of interest are highly correlated. The Monte Carlo results for the new x* variable are assumed to have similar properties to the known y control variable.

The following mathematical equations help illustrate the condition that is necessary to reduce the sampling error using control variates. Consider taking the variance of both sides of the equation that defines the new variable such that:

$$\text{var}(x^*) = \text{var}[y + (\hat{x} - \hat{y})]$$

The control variable y does not have a sampling error because it has known properties. Thus, the var(y) equals zero. Now, the variance of the remaining two variables can be rewritten as follows:

$$\text{var}(x^*) = \text{var}(\hat{x}) + \text{var}(\hat{y}) - 2\,\text{cov}(\hat{x}, \hat{y})$$

The control variate method will only reduce the sampling error in Monte Carlo simulations if var(x*) is less than var($\hat{x}$). Another way of expressing this condition is as follows:

$$\text{var}(\hat{y}) - 2\,\text{cov}(\hat{x}, \hat{y}) < 0$$

This relationship can be simplified as follows:

$$\text{cov}(\hat{x}, \hat{y}) > \frac{\text{var}(\hat{y})}{2}$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

The covariance can be converted to correlation by dividing both sides of the previous inequality by the product of the standard deviations as follows:

$$\text{corr}(\hat{x},\hat{y}) > \frac{1}{2}\sqrt{\frac{\text{var}(\hat{y})}{\text{var}(\hat{x})}}$$

A practical financial example of applying control variates is the use of Monte Carlo simulations in pricing Asian options (which will be discussed in Book 4). An Asian option is priced based on the average value of the underlying asset over the lifespan of the option. The use of a similar derivative, such as a European option, with known statistical properties can be used as a control variate. The price of the European option, P_{BS}, is determined by the Black-Scholes-Merton option pricing model. Next, simulated prices are determined for the Asian option and the European option and denoted P_A and P_{BS}^*, respectively. The new estimate of the Asian option price, P_A^*, could then be determined based on the following equation:

$$P_A^* = (P_A - P_{BS}) + P_{BS}^*$$

Reusing Sets of Random Numbers

LO 29.5: Describe the benefits of reusing sets of random number draws across Monte Carlo experiments and how to reuse them.

Reusing sets of random number draws across Monte Carlo experiments reduces the estimate variability across experiments by using the same set of random numbers for each simulation. Normally, a user would not desire to reuse the same random draws. However, in certain situations this technique is useful. Two examples of reusing sets of random numbers are for testing the power of the Dickey-Fuller test (used to determine whether a time series is covariance stationary) or for different experiments with options using time series data.

Dickey-Fuller (DF) test. Suppose an analyst wants to examine the DF test for sample sizes of 1,000 to test whether or not a particular market follows a random walk or contains a drift element. The analyst could reuse the same set of standard normal random variables for each simulation run while testing with different DF parameters. Using the same set of random numbers for each Monte Carlo experiment reduces the sampling variation across experiments. In this case, the sampling variability is reduced, but the accuracy of the actual estimates is not increased.

Different experiments. Another example where reusing sample data is useful is in testing differences among options. For example, suppose an analyst is examining option prices that are similar in all aspects except for time to maturity. The analyst could simulate a long time series of random draws and then split this longer time series into shorter time frames. A six-month time series of data could be subdivided into three sets of two-month maturity options or six sets of one-month maturity options. Using the same random number data set reduces the variability of simulated option prices across maturities.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Bootstrapping Method

LO 29.6: Describe the bootstrapping method and its advantage over Monte Carlo simulation.

Another way to generate random numbers is the bootstrapping method. The bootstrapping approach draws random return data from a sample of historical data. Under traditional Monte Carlo simulation, data sets are created by selecting random variables drawn from a pre-determined probability distribution. The bootstrapping method uses actual historical data instead of random data from a probability distribution. In addition, bootstrapping repeatedly draws data from a historical data set and replaces the data so it can be drawn again.

For example, suppose an analyst uses the bootstrapping method to estimate parameter θ. The analyst begins by obtaining sample historical data over a specific time period. This historical data is denoted:

$$y = y_1, y_2, \ldots, y_T$$

The statistical properties of parameter $\hat{\theta}_T$ are then estimated based on the bootstrapping sample data. The analyst creates N samples of T variables with replacement from the original y data sample. The parameter estimate $\hat{\theta}$ is calculated for every sample to create N estimates. In other words, the samples that are drawn are not totally random, but are drawn from a pre-determined historical sample set y. The statistical properties of this sample of $\hat{\theta}$ estimates are then analyzed.

An obvious advantage of the bootstrapping approach is that no assumptions are made regarding the true distribution of the parameter estimate that is being examined. This implies that it can include extreme events that have occurred in the past (e.g., during a financial crisis). Inclusion of outliers will produce a distribution that has fatter tails than the normal distribution, which allows for a more realistic view of actual return data. Thus, the bootstrapping methodology generates a collection of data sets with approximately the same distribution properties as the original data. However, any dependency of variables or autocorrelations in the original data set will no longer be present, because variables are not drawn in the same sequence as the original data set.

The following example describes how bootstrapping is used with a regression model. Assume that the bootstrapping approach is used to re-sample data with respect to the following standard regression model:

$$y = u + X\beta$$

The first step of the bootstrapping approach is to generate a sample size T of the historical data by drawing samples with replacement that take all related data corresponding to each observation y_i. In other words, for the 21st data observation, y_{21}, the approach takes this estimate along with all values of the explanatory variables for the 21st observation.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Next the coefficient matrix, $\hat{\beta}^*$, is estimated for this bootstrap sample. This process is then repeated a total of N times. Every time data is resampled, a sample size of T is generated from the original sample data with replacement and a coefficient matrix is estimated. This results in a set of N coefficient vectors that will all be unique, and a distribution of estimates is created for each coefficient.

This bootstrapping approach has a methodological problem resulting from sampling from regressors rather than using a fixed estimate in repeated samples. To correct for this problem, the approach can be slightly modified where re-sampling occurs with the residuals. Thus, the first step would be to sample actual data, estimate the value $\hat{y}$ and calculate the residuals, $\hat{u}$. The coefficient vector is then created using a modified dependent variable that is the sum of the fitted values and the bootstrap residuals $\hat{u}^*$ as follows:

$$y^* = \hat{y} + \hat{u}^*$$

LO 29.8: Describe situations where the bootstrapping method is ineffective.

Two situations that cause the bootstrapping method to be ineffective are *outliers* in the data and *non-independent data.*

If outliers exist in the data, the inferences drawn from parameter estimates may not be accurate depending on how many times the outliers are included in the bootstrapped sample. Because replacement is used in the bootstrap method, outliers could be drawn more often, causing the bootstrap distribution to have fatter tails. Alternatively, not drawing the outlier in the bootstrapped sample may lead to the opposite conclusions regarding the parameter estimate statistical properties. Recall that a major advantage of the bootstrapping approach over traditional approaches is that it does not require any assumptions of the probability distribution of the sampled data. Thus, the best way to mitigate this issue is to have a large number of replications.

If autocorrelation exists in the original sample data, then the original historical data are not independent of one another. A technique known as a *moving block bootstrap* is used to overcome the problem of autocorrelation. Blocks of data are examined at one time in order to preserve the original data dependency.

Random Number Generation

LO 29.7: 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.

A good random number generator has the ability to reproduce a random sequence and analyze characteristics of random numbers. Simulation software programs are able to reproduce the same sequence of iterations by starting sequences with a seed random number. The algorithms used to generate these random sequences are referred to as **pseudo-random number generators.** These number generators are advantageous because risk managers can improve models by reducing the estimate variance or debugging computer

FRM官方交流群 392417118

codes if the same sequence of random numbers is reproduced when programming the model.

A very common pseudo-random number generator is one that generates random number sequences uniformly distributed between 0 and 1. Each number has an equal probability of being drawn from this uniform (0,1) distribution. Numbers can be drawn from a discrete or continuous distribution. The term *pseudo* implies that these computer-generated numbers are *not truly random*, because they are actually generated from a formula. For example, suppose random numbers are generated from a continuous uniform (0,1) distribution based on the following formula:

$$y_{i+1} = (ay_i + c) \text{ modulo } m, \; i = 0,1,2,\ldots,T$$

In the above formula, T is the total number of random numbers drawn, y_0 is the initial value of y, which is referred to as the **seed**, a is a constant multiplier, and c is an incremental value. The statement "modulo m" in the above formula refers to modulo operator, which is a clocklike process where the generator returns to 1 when the value m is reached.

In order to run a simulation, the user must first define the initial seed value, y_0. The choice of seed value will influence the properties of the random number distribution that is generated. The effect is strongest for the early draws in a series, but eventually the impact fades away. Therefore, the best way to control for this problem is to generate a very large number of observations and then discard the earliest observations.

For example, if a user requires 800 observations, then 1,000 random numbers are generated and the first 200 are eliminated from the sample. This ensures that the statistical properties of the sample reflect those of true random numbers that are not based on a pre-specified formula. Eventually random number sequences will repeat. Therefore, a good random number generator uses sequences with long cycles that require numerous iterations before a sequence is repeated.

Disadvantages of Simulation Approaches

LO 29.9: Describe disadvantages of the simulation approach to financial problem solving.

Disadvantages of the simulation approach to financial problem solving include:

- High computation costs
- Results are imprecise
- Results are difficult to replicate
- Results are experiment-specific

Some problems may require a large number of replications to obtain more accurate results. If estimated parameters are complex, the computations may take an extremely long time to run. Computer processor times have improved exponentially. However, the complexity of markets and issues that are examined have also become increasingly complex, leading to *high computation costs.*

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Imprecise results may be present even with a very large number of simulation iterations when the assumptions of model inputs or the data generating process are unrealistic. A common mis-specified model assumption is related to the underlying probability distribution of inputs. For example, option prices are typically fat-tailed, but a model could erroneously draw option prices from a normal distribution. This would lead to inaccurate results regardless of the number of replications.

In practice, users seldom use a defined seed for the start of random draws in simulations. Without the use of an initial seed, it is *not possible to replicate results* from previous experiments. The best way to overcome this problem and reduce the variation of results is to use a very large number of replications. Thus, it is common to use at least 10,000 replications in Monte Carlo simulations if it is computationally cost-effective.

Simulation results are experiment-specific because financial problems are analyzed based on a specific data generating process and set of equations. If alternate assumptions are made in the equations or data generating process, the results may differ substantially.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Key Concepts

LO 29.1

The basic steps of a Monte Carlo simulation are: (1) specify the data generating process (DGP), (2) estimate an unknown variable, (3) save the estimate from step 2, and (4) go back to step 1 and repeat this process *N* times.

LO 29.2

The standard error estimate of a Monte Carlo simulation, $s/\sqrt{N}$, can be reduced by a factor of 10 by increasing *N* by a factor of 100.

LO 29.3

The antithetic variate technique reduces Monte Carlo sampling error by rerunning the simulation using a complement set of the original set of random variables.

LO 29.4

The control variate technique replaces a variable x that has unknown properties in a Monte Carlo simulation with a similar variable y that has known properties. The new x^* variable estimate will have a smaller sampling error than the original x variable if the control statistic and statistic of interest are highly correlated.

LO 29.5

Reusing sets of random number draws across Monte Carlo experiments reduces the estimate variability across experiments.

LO 29.6

Bootstrapping simulations repeatedly draw data from historical data sets and replace the data so it can be re-drawn. The bootstrapping technique requires no assumptions with respect to the true distribution of the parameter estimates.

LO 29.7

Pseudo-random numbers are not truly random, because they are actually generated from a formula. The choice of the initial seed value influences the properties of the random number distribution that is generated. Thus, when using a seed value, increasing the number of replications and eliminating early estimates from the sample can mitigate any biases.

LO 29.8

The bootstrapping method is ineffective when there are outliers in the data or when the data is non-independent.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

LO 29.9

Disadvantages of the simulation approach to financial problem solving include: high computation costs, imprecise results, difficulty with replicating results, and experiment-specific results.

最新视频课程
QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checkers

1. Suppose an analyst is concerned about Monte Carlo sampling error. Based on an initial Monte Carlo simulation with 100 replications, the results indicated a standard deviation of 12.64. The simulation was rerun with 900 replications and the standard deviation remained at 12.64. What are the standard error estimates for the simulations with 100 replications and 900 replications, respectively?

	N = 100	N = 900
A.	0.126	0.014
B.	0.126	0.140
C.	1.264	0.421
D.	1.264	0.214

2. A concern for Monte Carlo simulations is the size of the sampling error. One way to reduce the sampling error is to use the antithetic variate technique. Which of the following statements best describe this technique?
   A. The simulation is rerun using a complement set of the original set of random variables.
   B. The number of replications is increased significantly to reduce sampling error.
   C. Sample data is replaced after every replication to ensure it has an equal probability of being redrawn.
   D. The data generating process is approximated by redefining the unknown variable with a variable that has known properties.

3. Suppose an analyst is testing the robustness of the Dickey-Fuller test by changing the drift parameter for several different experiments. Reusing sets of random number draws across Monte Carlo experiments will most likely result in:
   A. increasing the accuracy of the drift estimates for each experiment.
   B. increasing the sampling variance across experiments.
   C. reducing the accuracy of the drift estimates for each experiment.
   D. reducing the sampling variance across experiments.

4. Suppose a pseudo-random number generator is used that generates random number sequences uniformly and continuously distributed between 0 and 1. An analyst begins by defining the initial seed value for the number generator process. The analyst knows that the choice of seed value will influence the properties of the generated random number distribution. The best way to reduce this problem is by using a:
   A. large number of replications and discarding the outliers.
   B. large number of replications and discarding the earliest draws.
   C. small seed or initial value.
   D. large seed or initial value.

5. Monte Carlo simulation is a widely used technique in solving economic and financial problems. Which of the following statements is not a limitation of the Monte Carlo technique when solving problems of this nature?
   A. High computational costs arise with complex problems.
   B. Simulation results are experiment-specific because financial problems are analyzed based on a specific data generating process and set of equations.
   C. Results of most Monte Carlo experiments are difficult to replicate.
   D. If the input variables have fat tails, Monte Carlo simulations are not relevant because it always draws random variables from a normally distributed population.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Concept Checker Answers

1. **C** The standard error is determined by dividing the standard deviation by the square root of the number of replications $s / \sqrt{N}$. The standard error estimate for the first simulation of 100 replications is 1.264 (i.e., 12.64 / 10). With 900 replications, the standard error estimate is reduced to 0.4213 (i.e., 12.64 / 30).

2. **A** The antithetic variate technique reduces Monte Carlo sampling error by rerunning the simulation using a complement set of the original set of random variables.

3. **D** Using the same set of random numbers for each Monte Carlo experiment reduces the sampling variation across experiments. Although the sampling variability is reduced, the accuracy of the actual estimates in each case is not influenced.

4. **B** The best way to control for this problem is to generate a very large number of observations and then discard the earliest observations. This ensures that the statistical properties of the sample reflect those of true random numbers that are not based on a pre-specified formula.

5. **D** A disadvantage of Monte Carlo simulation is that imprecise results may be present when the assumptions of model inputs or data generating process are unrealistic. The distribution of input variables does not need to be the normal distribution. The problem arises when a variable in the real world is fat-tailed, but a model could erroneously draw option prices from a normal distribution.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Self-Test: Quantitative Analysis

10 Questions: 24 Minutes

1. Given the following probability data for the return on the market and the return on Best Oil, calculate the covariance of returns between Best Oil and the market.

Probability Matrix

	R_{Best} = 20%	R_{Best} = 10%	R_{Best} = 5%
R_{Mkt} = 15%	40%	0	0
R_{Mkt} = 10%	0	20%	0
R_{Mkt} = 0%	0	0	40%

A. 44.0.
B. 12.0.
C. 2.8.
D. 22.5.

2. Rob Conniff has encountered a difficult section on a multiple-choice exam. There are five questions in this section and each question has three equally likely answer choices. Which of the following amounts is closest to the probability that he will get three or more questions correct by randomly guessing?
A. 4.5%.
B. 16.5%.
C. 21.0%.
D. 79.0%.

3. You are forecasting the sales of a building materials supplier by assessing the expansion plans of its largest customer, a homebuilder. You estimate the probability that the customer will increase its orders for building materials to 25%. If the customer does increase its orders, you estimate the probability that the homebuilder will start a new development at 70%. If the customer does not increase its orders from this supplier, you estimate only a 20% chance that it will start the new development. Later, you find out that the homebuilder will start the new development. In light of this new information, what is your new (updated) probability that the builder will increase its orders from this supplier?
A. 17.50%.
B. 32.55%.
C. 53.85%.
D. 60.00%.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

4. In performing hypothesis testing as a quantitative analyst, you have recently encountered some unsatisfactory results. You consult your boss and he suggests that you consider increasing the significance level in your testing activities. Which of the following outcomes would most likely occur with such an increase?
   A. Increased probability of making a Type I error.
   B. Increased probability of making a Type I or II error.
   C. Decreased probability of making a Type I error.
   D. Decreased probability of making a Type I or II error.

Use the following information to answer Question 5.

An analyst is given the data in the following table for a regression of the annual sales for Company XYZ, a maker of paper products, on paper product industry sales.

Parameters	*Coefficient*	*Standard Error of the Coefficient*
Intercept	–94.88	32.97
Slope (industry sales)	0.2796	0.0363

The correlation between company and industry sales is 0.9757. The regression was based on five observations.

5. Which of the following is closest to the value and reports the most likely interpretation of the R^2 for this regression? The R^2 is:
   A. 0.048, indicating that the variability of industry sales explains about 4.8% of the variability of company sales.
   B. 0.048, indicating that the variability of company sales explains about 4.8% of the variability of industry sales.
   C. 0.952, indicating that the variability of industry sales explains about 95.2% of the variability of company sales.
   D. 0.952, indicating that the variability of company sales explains about 95.2% of the variability of industry sales.

Use the following information to answer Questions 6 through 8.

Theresa Miller is attempting to forecast sales for Alton Industries based on a multiple regression model. The model Miller estimates is:

$$\text{sales} = b_0 + (b_1 \times \text{DOL}) + (b_2 \times \text{IP}) + (b_3 \times \text{GDP}) + \varepsilon_t$$

where:
sales = change in sales adjusted for inflation
DOL = change in the real value of the $ (rates measured in €/$)
IP = change in industrial production adjusted for inflation (millions of $)
GDP = change in inflation-adjusted GDP (millions of $)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

All changes in variables are in percentage terms.

Miller runs the regression using monthly data for the prior 180 months. The model estimates (with coefficient standard errors in parentheses) are:

$$\text{sales} = \underset{(5.4)}{10.2} + \underset{(3.5)}{(5.6 \times \text{DOL})} + \underset{(4.2)}{(6.3 \times \text{IP})} + \underset{(5.3)}{(9.2 \times \text{GDP})}$$

The sum of squared residuals (SSR) is 145.6 and the total sum of squares (TSS) is 357.2.

Figure 1: Partial Student's *t*-distribution (one-tailed probabilities)

df	p = 0.10	p =0.05	p = 0.025	p = 0.01	p = 0.005
170	1.287	1.654	1.974	2.348	2.605
176	1.286	1.654	1.974	2.348	2.604
180	1.286	1.653	1.973	2.347	2.603

Figure 2: Partial *F*-Table critical values for right-hand tail area equal to 0.05

	df 1 = 1	df 1 = 3	df 1 = 5
df 2 = 170	3.90	2.66	2.27
df 2 = 176	3.89	2.66	2.27
df 2 = 180	3.89	2.65	2.26

Figure 3: Partial *F*-Table critical values for right-hand tail area equal to 0.025

	df 1 = 1	df 1 = 3	df 1 = 5
df 2 = 170	5.11	3.19	2.64
df 2 = 176	5.11	3.19	2.64
df 2 = 180	5.11	3.19	2.64

6. The unadjusted R^2 and the standard error of the regression (SER) are closest to:

	R^2	SER
A.	59.2%	1.425
B.	59.2%	0.910
C.	40.8%	0.910
D.	40.8%	1.425

©2015 Kaplan, Inc.
FRM官方交流群 392417118

7. The appropriate decision with regard to the *F*-statistic for testing the null hypothesis that all of the independent variables are simultaneously equal to zero at the 5% significance level is to:
   A. reject the null hypothesis because the *F*-statistic is larger than the critical *F*-value of 3.19.
   B. fail to reject the null hypothesis because the *F*-statistic is smaller than the critical *F*-value of 3.19.
   C. reject the null hypothesis because the *F*-statistic is larger than the critical *F*-value of 2.66.
   D. fail to reject the null hypothesis because the *F*-statistic is smaller than the critical *F*-value of 2.66.

8. What is the width of the 99% confidence interval for GDP, and is zero in that 99% confidence interval?

	Width of 99% CI	Zero in interval
A.	13.8	Yes
B.	3.8	No
C.	27.6	Yes
D.	27.6	No

9. The GTEC Corporation uses an exponentially weighted moving average (EWMA) model with a decay factor of 0.75 to model the daily volatility of a stock. The current estimate of daily volatility 1.8%. The closing price of the stock was $38 yesterday and $35 today. Using continuously compounded returns, what is the updated estimate of volatility?
   A. 5.39%.
   B. 4.39%.
   C. 3.39%.
   D. 2.39%.

10. A risk manager estimates the daily variance using a GARCH(1,1) model on daily returns (r_t):

$$h_t = \alpha_0 + \alpha_1 r^2_{t-1} + \beta h_{t-1}$$

The model parameter values are:

$\alpha_0 = 0.0000008$
$\alpha_1 = 0.050$
$\beta = 0.93$

Using the model, what is the long-run annualized volatility estimate (assuming 252 trading days in a year and that volatility increases by the square root of time)?
   A. 0.52%.
   B. 0.63%.
   C. 9.89%.
   D. 10.04%.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Self-Test Answers: Quantitative Analysis

1. A $E(R_{Best}) = 0.4(20\%) + 0.2(10\%) + 0.4(5\%) = 12\%$

$E(R_{Mkt}) = 0.4(15\%) + 0.2(10\%) + 0.4(0\%) = 8\%$

$Cov(R_{Best}, R_{Mkt}) = 0.4(20\% - 12\%)(15\% - 8\%)$

$+ \quad 0.2(10\% - 12\%)(10\% - 8\%)$

$+ \quad 0.4(5\% - 12\%)(0\% - 8\%)$

$= \quad 0.4(8)(7) + 0.2(-2)(2) + 0.4(-7)(-8) = 44$

The units of covariance (like variance) are percent squared here. We used whole number percents in the calculations and got 44; if we had used decimals, we would have gotten 0.0044.

(See Topic 16)

2 C The number of questions correct would follow a binomial distribution. Probability of success is 1/3 and the number of trials is 5. The probability of getting three or more questions correct is the sum of the following:

$P(3) = 10 \times (1/3)^3 \times (2/3)^2 = 0.1646$

$P(4) = 5 \times (1/3)^4 \times (2/3)^1 = 0.0412$

$P(5) = 1 \times (1/3)^5 \times (2/3)^0 = 0.0041$

$0.1646 + 0.0412 + 0.0041 = 21.0\%$

(See Topic 17)

3. C The prior probability that the builder will increase its orders is 25%.

P(increase) = 0.25

P(no increase) = 0.75

There are four possible outcomes:
- Builder increases its orders and starts new development.
- Builder increases its orders and does not start new development.
- Builder does not increase its orders and starts new development.
- Builder does not increase its orders and does not start new development.

The probabilities of each outcome are as follows:
- P(increase and development) = (0.25)(0.70) = 0.175.
- P(increase and no development) = (0.25)(0.30) = 0.075.
- P(no increase and development) = (0.75)(0.20) = 0.15.
- P(no increase and no development) = (0.75)(0.80) = 0.60.

©2015 Kaplan, Inc.

FRM官方交流群 392417118

We want to update the probability of an increase in orders, given the new information that the builder is starting the development. We can apply Bayes' formula:

$$P(\text{increase} \mid \text{development}) = \frac{P(\text{development} \mid \text{increase}) \times P(\text{increase})}{P(\text{development})}$$

From our assumptions, P(development | increase) = 0.70, and P(increase) = 0.25, so the numerator is (0.70)(0.25) = 0.175.

P(development) is the sum of P(development and increase) and P(development and no increase).

P(development) = 0.175 + 0.15 = 0.325

$$\text{Thus, } P(\text{increase} \mid \text{development}) = \frac{(0.7)\times(0.25)}{0.175+0.15} = \frac{0.175}{0.325} = 0.5385, \text{ or } 53.85\%$$

(See Topic 18)

4. **A** An increase in the significance level (from 1% to 5%, for example) means that a researcher is more likely to reject the null hypothesis since the critical value will be lower. Therefore, there is a greater probability of making a Type I error (rejecting the null hypothesis when it is actually true).

   (See Topic 19)

5. **C** The R^2 is computed as the correlation squared: $(0.9757)^2 = 0.952$.

   The interpretation of this R^2 is that 95.2% of the variation in Company XYZ's sales is explained by the variation in industry sales. Answer D is incorrect because it is the independent variable (industry sales) that explains the variation in the dependent variable (company sales). This interpretation is based on the economic reasoning used in constructing the regression model.

   (See Topic 21)

6. **B** $$\text{SER} = \sqrt{\frac{145.6}{180-3-1}} = 0.910$$

   $$\text{unadjusted } R^2 = \frac{357.2-145.6}{357.2} = 0.592$$

   (See Topic 23)

7. **C** ESS = 357.2 – 145.6 = 211.6, *F*-statistic = (211.6 / 3) / (145.6 / 176) = 85.3. The critical value for a one-tailed 5% *F*-test with 3 and 176 degrees of freedom is 2.66. Because the *F*-statistic is greater than the critical *F*-value, the null hypothesis that all of the independent variables are simultaneously equal to zero should be rejected.

   (See Topic 24)

8. **C** The confidence interval is 9.2 +/– (5.3 × 2.604), where 2.604 is the two-tailed 1% *t*-statistic with 176 degrees of freedom (which is the same as a one-tailed 0.5% *t*-statistic with 176 degrees of freedom). The interval is -4.6 to 23.0, which has a width of 27.6 and zero is in that interval.

   (See Topic 24)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

9. **B** Updated volatility estimate = $[\lambda \times (\text{volatility}_{t-1})^2 + (1 - \lambda) \times (\text{current return})^2]^{0.5}$

Current return = ln(price today / price yesterday)

ln(35/38) = –8.223%

Updated volatility estimate = $[0.75 \times (0.018)^2 + 0.25 \times (-0.08223)^2]^{0.5}$

= $[0.000243 + 0.001690443]^{0.5}$

= 4.39%

(See Topic 28)

10. **D** Remember that when questions ask for volatility, they are referring to the standard deviation.

We first calculate the daily variance, which then needs to be adjusted to an annualized variance and finally we can take the square root to find the annualized volatility (standard deviation).

Long-run daily variance = $\alpha_0 / (1 - \alpha_1 - \beta)$

= 0.0000008 / (1 – 0.05 – 0.93) = 0.00004

Long-run daily standard deviation = $\sqrt{\text{variance}} = \sqrt{0.00004}$ = 0.6325%

Annualized standard deviation = daily standard deviation $\times \sqrt{\text{time}}$

= $0.6325\% \times \sqrt{252}$ = 10.04%

(See Topic 28)

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Formulas

Topic 15

joint probability: $P(AB) = P(A \mid B) \times P(B)$

conditional probability: $P(A \mid B) = \dfrac{P(AB)}{P(B)}$

independent events: $P(A \mid B) = P(A)$

Topic 16

expected value: $E(X) = \Sigma P(x_i)x_i$

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

covariance: $\mathrm{Cov}(R_i,R_j) = E\{[R_i - E(R_i)][R_j - E(R_j)]\}$

correlation: $\mathrm{Corr}(R_i,R_j) = \dfrac{\mathrm{Cov}(R_i,R_j)}{\sigma(R_i)\sigma(R_j)}$

portfolio variance: $\mathrm{Var}(R_P) = w_A{}^2\sigma^2(R_A) + w_B{}^2\sigma^2(R_B) + 2w_Aw_B\sigma(R_A)\sigma(R_B)\rho(R_A,R_B)$

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

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

Topic 17

Poisson distribution: $P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

binomial probability function: (number of ways to choose x from n) $p^x(1-p)^{n-x}$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

expected value of a binomial random variable: $E(X) = np$

variance of a binomial random variable: $np(1 - p) = npq$

uniform distribution range: $P(x_1 \leq X \leq x_2) = (x_2 - x_1)/(b - a)$

mean of uniform distribution: $E(x) = \frac{a+b}{2}$

variance of uniform distribution: $Var(x) = \frac{(b-a)^2}{12}$

Topic 18

Bayes' theorem: $P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$

Topic 19

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

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

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

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

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

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

©2015 Kaplan, Inc.

FRM官方交流群 392417118

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

sample correlation coefficient: $r_{XY} = \frac{\text{Cov}(X,Y)}{(s_X)(s_Y)}$

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

sampling error of the mean = sample mean – population mean = $\overline{x} - \mu$

standard error of the sample mean: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$

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

$$F\text{-test} = \frac{s_1^2}{s_2^2}$$

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

confidence interval:

$$\left\{\left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix} - \begin{pmatrix}\text{critical}\\\text{value}\end{pmatrix}\begin{pmatrix}\text{standard}\\\text{error}\end{pmatrix}\right] < \begin{matrix}\text{population}\\\text{parameter}\end{matrix} < \left[\begin{matrix}\text{sample}\\\text{statistic}\end{matrix} + \begin{pmatrix}\text{critical}\\\text{value}\end{pmatrix}\begin{pmatrix}\text{standard}\\\text{error}\end{pmatrix}\right]\right\}$$

t-statistic: $t_{n-1} = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$

$$z\text{-statistic} = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Topic 20

exponentially weighted moving average (EWMA) model (covariance):

$$\text{cov}_n = \lambda\,\text{cov}_{n-1} + (1-\lambda)X_{n-1}Y_{n-1}$$

where:
λ = the weight for the most recent covariance on day $n-1$
X_{n-1} = the percentage change for variable X on day $n-1$
Y_{n-1} = the percentage change for variable Y on day $n-1$

GARCH(1,1) model (covariance): $\text{cov}_n = \omega + \alpha X_{n-1}Y_{n-1} + \beta\,\text{cov}_{n-1}$

covariance consistency condition: $\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} \le 1$

factor model: $U_i = \alpha_i F + \sqrt{1-\alpha_i^2}\,Z_i$

Topic 21

sample regression function: $Y_i = b_0 + b_1 \times X_i + e_i$

residual: $e_i = Y_i - (b_0 + b_1 \times X_i)$

regression slope coefficient: $b_1 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \dfrac{\text{Cov}(X,Y)}{\text{Var}(X)}$

regression intercept: $b_0 = \bar{Y} - b_1\bar{X}$

where:
$\bar{Y}$ = mean of Y
$\bar{X}$ = mean of X

sum of squared residuals (SSR) = $\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$

total sum of squares = explained sum of squares + sum of squared residuals

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y_i - \hat{Y})^2$$

$$\text{TSS} = \text{ESS} + \text{SSR}$$

©2015 Kaplan, Inc.
FRM官方交流群 392417118

coefficient of determination:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

Topic 23

standard error of the regression: $SER = \sqrt{s_\varepsilon^2} = \sqrt{\frac{SSR}{n-k-1}}$

F-statistic = (ESS / df) / (SSR / df)

adjusted $R^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1}$

Topic 24

homoskedasticity-only *F*-statistic: $F = \frac{(R_{ur}^2 - R_r^2)/m}{(1 - R_{ur}^2)/(n - k_{ur} - 1)}$

Topic 25

mean squared error (MSE): $MSE = \frac{\sum_{t=1}^{T} e_t^2}{T}$

unbiased mean squared error (s^2): $s^2 = \left(\frac{T}{T-k}\right)\frac{\sum_{t=1}^{T} e_t^2}{T}$

Akaike information criterion (AIC): $AIC = e^{\left(\frac{2k}{T}\right)}\frac{\sum_{t=1}^{T} e_t^2}{T}$

Schwarz information criterion (SIC): $SIC = T^{\left(\frac{k}{T}\right)}\frac{\sum_{t=1}^{T} e_t^2}{T}$

Topic 26

first-difference operator: $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Topic 27

first-order moving average [MA(1)] process:

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

where:
y_t = the time series variable being estimated
ε_t = current random white noise shock
ε_{t-1} = one-period lagged random white noise shock
θ = coefficient for the lagged random shock

MA(q) process:

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}$$

where:
y_t = the time series variable being estimated
ε_t = current random white noise shock
ε_{t-1} = one-period lagged random white noise shock
ε_{t-q} = q^{th}-period lagged random white noise shock
θ = coefficients for the lagged random shocks

first-order autoregressive [AR(1)] process:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where:
y_t = the time series variable being estimated
y_{t-1} = one-period lagged observation of the variable being estimated
ε_t = current random white noise shock
ϕ = coefficient for the lagged observation of the variable being estimated

Yule-Walker equation: $\rho_t = \phi^t$ for $t = 0,1,2,\ldots$

AR(p) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

where:
y_t = the time series variable being estimated
y_{t-1} = one-period lagged observation of the variable being estimated
y_{t-p} = p^{th}-period lagged observation of the variable being estimated
ε_t = current random white noise shock
ϕ = coefficients for the lagged observations of the variable being estimated

©2015 Kaplan, Inc.
FRM官方交流群 392417118

autoregressive moving average (ARMA) process:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

where:
y_t = the time series variable being estimated
ϕ = coefficient for the lagged observations of the variable being estimated
y_{t-1} = one-period lagged observation of the variable being estimated
ε_t = current random white noise shock
θ = coefficient for the lagged random shocks
ε_{t-1} = one-period lagged random white noise shock

Topic 28

continuously compounded return: $u_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$

exponentially weighted moving average (EWMA) model (volatility):

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

where:
λ = weight on previous volatility estimate (λ between zero and one)

GARCH(1,1) model (volatility):

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

where:
α = weighting on the previous period's return
β = weighting on the previous volatility estimate
ω = weighted long-run variance $= \gamma V_L$
V_L = long-run average variance $= \frac{\omega}{1-\alpha-\beta}$
$\alpha + \beta + \gamma = 1$
$\alpha + \beta < 1$ for stability so that γ is not negative

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Using the Cumulative Z-Table

Probability Example

Assume that the annual earnings per share (EPS) for a large sample of firms is normally distributed with a mean of \$5.00 and a standard deviation of \$1.50. What is the approximate probability of an observed EPS value falling between \$3.00 and \$7.25?

If EPS = x = \$7.25, then $z = (x - \mu)/\sigma$ = (\$7.25 – \$5.00)/\$1.50 = +1.50

If EPS = x = \$3.00, then $z = (x - \mu)/\sigma$ = (\$3.00 – \$5.00)/\$1.50 = –1.33

For z-value of 1.50: Use the row headed 1.5 and the column headed 0 to find the value 0.9332. This represents the area under the curve to the left of the critical value 1.50.

For z-value of –1.33: Use the row headed 1.3 and the column headed 3 to find the value 0.9082. This represents the area under the curve to the left of the critical value +1.33. The area to the left of –1.33 is 1 – 0.9082 = 0.0918.

The area between these critical values is 0.9332 – 0.0918 = 0.8414, or 84.14%.

Hypothesis Testing—One-Tailed Test Example

A sample of a stock's returns on 36 non-consecutive days results in a mean return of 2.0%. Assume the population standard deviation is 20.0%. Can we say with 95% confidence that the mean return is greater than 0%?

H_0: $\mu \leq 0.0\%$, H_A: $\mu > 0.0\%$. The test statistic = *z*-statistic = $\dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
= (2.0 – 0.0) / (20.0 / 6) = 0.60.

The significance level = 1.0 – 0.95 = 0.05, or 5%.

Since this is a one-tailed test with an alpha of 0.05, we need to find the value 0.95 in the cumulative *z*-table. The closest value is 0.9505, with a corresponding critical *z*-value of 1.65. Since the test statistic is less than the critical value, we fail to reject H_0.

Hypothesis Testing—Two-Tailed Test Example

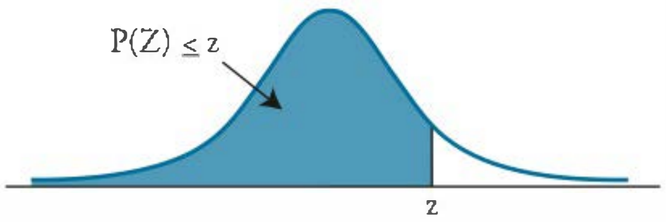
Using the same assumptions as before, suppose that the analyst now wants to determine if he can say with 99% confidence that the stock's return is not equal to 0.0%.

H_0: $\mu = 0.0\%$, H_A: $\mu \neq 0.0\%$. The test statistic (*z*-value) = (2.0 – 0.0) / (20.0 / 6) = 0.60.
The significance level = 1.0 – 0.99 = 0.01, or 1%.

Since this is a two-tailed test with an alpha of 0.01, there is a 0.005 rejection region in both tails. Thus, we need to find the value 0.995 (1.0 – 0.005) in the table. The closest value is 0.9951, which corresponds to a critical *z*-value of 2.58. Since the test statistic is less than the critical value, we fail to reject H_0 and conclude that the stock's return equals 0.0%.

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Cumulative *Z*-Table



$P(Z \leq z) = N(z)$ for $z \geq 0$

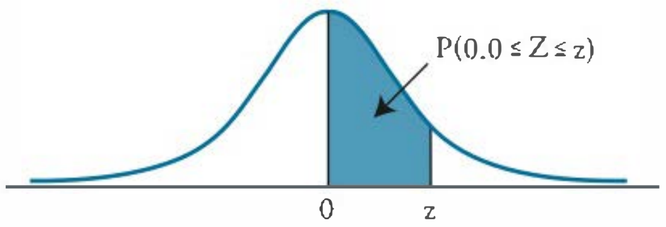
$P(Z \leq -z) = 1 - N(z)$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Alternative *Z*-Table



$P(Z \leq z) = N(z)$ for $z \geq 0$

$P(Z \leq -z) = 1 - N(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3356	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4939	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Student's T-Distribution

Level of Significance for One-Tailed Test						
df	0.100	0.050	0.025	0.01	0.005	0.0005
Level of Significance for Two-Tailed Test						
df	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.294
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

©2015 Kaplan, Inc.

FRM官方交流群 392417118

F-Table at 5%

Critical values of the *F*-distribution at a 5% level of significance

Degrees of freedom for the numerator along top row
Degrees of freedom for the denominator along side row

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	6.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

©2015 Kaplan, Inc.
FRM官方交流群 392417118

F-Table at 2.5%

Critical values of the *F*-distribution at a 2.5% level of significance

Degrees of freedom for the numerator along top row
Degrees of freedom for the denominator along side row

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	648	799	864	900	922	937	948	957	963	969	977	985	993	997	1001	1006
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.41	4.36	4.31
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Chi-Squared Table

Values of χ^2 (Degrees of Freedom, Level of Significance)
Probability in Right Tail

Degrees of Freedom	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000157	0.000982	0.003932	0.0158	2.706	3.841	5.024	6.635	7.879
2	0.020100	0.050636	0.102586	0.2107	4.605	5.991	7.378	9.210	10.597
3	0.1148	0.2158	0.3518	0.5844	6.251	7.815	9.348	11.345	12.838
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170

©2015 Kaplan, Inc.
FRM官方交流群 392417118

Index

A

addition rule 20
adjusted R^2 181, 201
Akaike information criterion 211
alternative hypothesis 101
annuity 5
ANOVA table 198
antithetic variate technique 249
arithmetic means 30
asymptotic efficiency 212
autocorrelation function 219
autocovariance function 218
autoregression 218
autoregressive (AR) model 218
autoregressive conditional heteroskedasticity model 238
autoregressive moving average process 231
autoregressive representation 228

B

backtesting 121
Bayes' theorem 75
Bernoulli distribution 55
best linear unbiased estimator (BLUE) 48, 169
binomial distribution 55
binomial random variable 55
Bootstrapping Method 252
Box-Pierce Q-statistic 222

C

central limit theorem 66, 153
central moments 43
central tendency 29
Chebyshev's inequality 120
chi-squared distribution 68
chi-squared test 114
coefficient of determination 154, 180
coefficient of multiple correlation 199
cokurtosis 46
compounding 2
compound interest 1
conditional heteroskedasticity 166
conditional probability 18, 75
confidence interval 60, 96, 164
  regression coefficient 161, 194
consistency 212
consistent estimator 48, 153
continuous random variable 14
continuous uniform distribution 53
control variate technique 250
copula 136
copula correlation 136
correlation 40, 128, 156, 242
coskewness 46
cost of capital 4
covariance 38, 128, 242
covariance stationary 218
critical value 101
cumulative distribution function 15

D

data generating process 212, 247
data mining 210
decision rule 103
default risk 3
degrees of freedom 190, 194
dependent events 19
dependent variable 147
  predicting 195
descriptive statistics 29
discount factor 4
discounting 2
discount rates 3
discrete probability function 16
discrete random variable 13
discrete uniform random variable 16
distributed lag 220
dummy variables 166
dummy variable trap 183

E

economic significance 109
efficient estimator 48
error term 149
estimator 48
event 13
EWMA model 131
exceedance 121
exception 121
excess kurtosis 44, 46
exhaustive events 13
expectations 34
expected value 34
explained sum of squares 154, 198
explained variable 147
explanatory variable 147

©2015 Kaplan, Inc.

FRM官方交流群 392417118

exponentially weighted moving average 130, 238

F

factor model 135
F-distribution 69
first-order autoregressive process 229
frequentist approach 80
F-statistic 196
F-test 117
future value 1

G

GARCH model 132, 239
Gaussian copula 138
Gauss-Markov theorem 169
general linear process 221
geometric mean 33

H

heteroskedasticity 166, 179
homoskedasticity 166, 179
homoskedasticity-only *F*-statistic 202
hypothesis 100
hypothesis testing 162

I

independent and identically distributed (i.i.d.) random variables 66
independent events 19
independent variable 147
independent white noise 220
inferential statistics 29
innovations 221
intercept 149, 178
inverse cumulative distribution function 16

J

joint probability 18, 21

K

kurtosis 43, 45

L

lag operator 220
leptokurtic distributions 45
level of significance 96
liquidity risk 3
Ljung-Box Q-statistic 222
lognormal distribution 65

M

marginal distributions 136
maturity risk 3
maximum likelihood estimator 241
mean 42
mean reversion 240
mean squared error 209
measures of fit 179
median 31
mesokurtic distributions 45
mixture distributions 70
mode 32
Monte Carlo Simulation 247
moving average process 227
moving average representation 228
moving block bootstrap 253
multicollinearity 182
multiple regression 177
  analysis 177
  formulation 178
  interpretation 178
multiplication rule 18
multivariate copula 139
mutually exclusive events 13, 20

N

negative skew 44
noise component 149
nominal risk-free rate 3
non-independent data 253
nonparametric distributions 53
normal distribution 59
normal white noise 220
null hypothesis 101

O

OLS estimators 177
omitted variable bias 176
one-factor copula 139
one-tailed test 102, 194
opportunity cost 3, 4
ordinary least squares 151
outcome 13
outliers 44, 253

P

parameter 147
parametric distributions 53
partial autocorrelation function 219
partial slope coefficients 178
penalty factor 211
perpetuity 6
persistence 241

©2015 Kaplan, Inc.

FRM官方交流群 392417118

platykurtic distributions 45
point estimate 48, 96
Poisson distribution 58
population 29
population mean 30, 90
population variance 91
positive-semidefinite 133
positive skew 44
posterior probabilities 82
power of a test 107
predicted values 164
present value 1
present value factor 4
price relatives 66
probability density function 15
probability distribution 13
probability function 13
probability matrix 23
pseudo-random number generators 253
p-value 110, 163, 192, 199

Q

Q-statistic 222

R

R^2, adjusted 181, 201
R^2, coefficient of determination 154, 180
Random Number Generation 253
random variable 13
rational distributed lags 221
rational polynomials 221
real risk-free rate 3
regression analysis 147
  heteroskedasticity 167
  multicollinearity 183
regression coefficient 148
  confidence interval 161, 194
  hypothesis testing 190
  t-test 162
required rate of return 3, 4
residual 150
residual plot 168
restricted R^2 202
robust standard errors 168

S

s^2 measure 210
sample 29
sample autocorrelation 222
sample covariance 95
sample mean 30, 90, 222
sample partial autocorrelation 222
sample regression function 149
sample standard deviation 93
sample variance 92
sampling distribution 89
scatter plot 41, 147
Schwarz information criterion 211
seed 254
skewness 43, 44
slope coefficient 149
standard deviation 37
standard error 90, 93
standard error of the forecast 165
standard error of the regression 156, 179
statistical significance 109, 191
Student's *t*-copula 139
Student's *t*-distribution 66
sum of squared residuals 153, 198
symmetrical distributions 44

T

tail dependence 140
t-distribution 66
test statistic 101
time series 218
time value of money 1
total sum of squares 154, 198
trend 218
t-test 110, 162
two-tailed test 102, 193
Type I error 107
Type II error 107, 183

U

unbiased estimator 48, 153
unconditional heteroskedasticity 166
unconditional probability 18, 75
uniform distribution 53
unrestricted R^2 202

V

variance 37, 43
variance-covariance matrix 133
Venn diagram 20

W

white noise 219
Wold's representation theorem 221

Y

Yule-Walker equation 230

Z

z-distribution 61
z-test 111
z-value 61

最新视频课程
QQ: 2211892999

©2015 Kaplan, Inc.

FRM官方交流群 392417118

Required Disclaimers:

CFA Institute does not endorse, promote, or warrant the accuracy or quality of the products or services offered by Kaplan Schweser. CFA Institute, CFA®, and Chartered Financial Analyst® are trademarks owned by CFA Institute.

Certified Financial Planner Board of Standards Inc. owns the certification marks CFP®, CERTIFIED FINANCIAL PLANNER™, and federally registered CFP (with flame design) in the U.S., which it awards to individuals who successfully complete initial and ongoing certification requirements. Kaplan University does not certify individuals to use the CFP®, CERTIFIED FINANCIAL PLANNER™, and CFP (with flame design) certification marks. CFP® certification is granted only by Certified Financial Planner Board of Standards Inc. to those persons who, in addition to completing an educational requirement such as this CFP® Board-Registered Program, have met its ethics, experience, and examination requirements.

Kaplan Schweser and Kaplan University are review course providers for the CFP® Certification Examination administered by Certified Financial Planner Board of Standards Inc. CFP Board does not endorse any review course or receive financial remuneration from review course providers.

GARP® does not endorse, promote, review, or warrant the accuracy of the products or services offered by Kaplan Schweser of FRM® related information, nor does it endorse any pass rates claimed by the provider. Further, GARP® is not responsible for any fees or costs paid by the user to Kaplan Schweser, nor is GARP® responsible for any fees or costs of any person or entity providing any services to Kaplan Schweser. FRM®, GARP®, and Global Association of Risk Professionals™ are trademarks owned by the Global Association of Risk Professionals, Inc.

CAIAA does not endorse, promote, review or warrant the accuracy of the products or services offered by Kaplan Schweser, nor does it endorse any pass rates claimed by the provider. CAIAA is not responsible for any fees or costs paid by the user to Kaplan Schweser nor is CAIAA responsible for any fees or costs of any person or entity providing any services to Kaplan Schweser. CAIA®, CAIA Association®, Chartered Alternative Investment Analyst℠, and Chartered Alternative Investment Analyst Association® are service marks and trademarks owned by CHARTERED ALTERNATIVE INVESTMENT ANALYST ASSOCIATION, INC., a Massachusetts non-profit corporation with its principal place of business at Amherst, Massachusetts, and are used by permission.

FRM官方交流群 392417118