

Quantitative Tools for Economics and Business

By Tim Murray, PhD

VERSION: SEPTEMBER 2023

Table of Contents

| | | |
|-----------|---|-----------|
| 1 | Introduction | 6 |
| I | Mathematical Concepts | 7 |
| 2 | Review of Algebra | 8 |
| 2.1 | Exponents | 8 |
| 2.2 | Logarithms | 9 |
| 2.3 | Algebra Rules and Operations | 10 |
| 2.4 | Linear Equations in One Variable | 13 |
| 2.5 | Linear Equation with Two Variables | 14 |
| 2.6 | Quadratic Equations | 15 |
| 2.7 | Rational Equations | 17 |
| 2.8 | Systems of Equations | 18 |
| 3 | Introduction to Differential Calculus | 20 |
| 3.1 | The Derivative | 20 |
| 3.2 | Calculating Derivatives | 22 |
| 3.2.1 | Functions with Exponents | 22 |
| 3.2.2 | The Sum and Difference Rules | 24 |
| 3.2.3 | The Product Rule | 24 |
| 3.2.4 | The Quotient Rule | 25 |
| 3.2.5 | The Chain Rule | 26 |
| 3.2.6 | Derivatives of Different Types of Functions | 27 |
| 3.3 | Finding Minima and Maxima using Derivatives | 28 |
| 3.3.1 | First Derivative Test | 29 |
| 3.3.2 | Second Derivative Test | 30 |
| 3.4 | Multivariate Calculus: Partial Derivatives | 31 |
| 4 | Constrained Optimization | 33 |
| 4.1 | The Method of Lagrange Multipliers | 33 |
| II | Solving Economic Models | 35 |
| 5 | Supply, Demand, and Elasticity | 36 |
| 5.1 | Demand | 36 |
| 5.2 | Supply | 37 |
| 5.3 | Defining a Market | 37 |
| 5.4 | Government Intervention | 40 |
| 5.4.1 | Price Controls | 41 |
| 5.4.2 | Taxation | 42 |
| 5.4.3 | Government Subsidies | 44 |

| | | |
|------------|---|-----------|
| 5.5 | Nonlinear Markets | 46 |
| 5.6 | Elasticity of Demand | 46 |
| 6 | Utility Maximization | 49 |
| 6.1 | Utility | 49 |
| 6.1.1 | Indifference Curves | 50 |
| 6.1.2 | Marginal Rate of Substitution | 51 |
| 6.2 | The Budget Constraint | 52 |
| 6.3 | Utility Maximization | 53 |
| | Appendix 6.A Appendix: Constrained Optimization | 56 |
| 7 | Profit Maximization | 58 |
| 7.1 | The Production Function | 58 |
| 7.2 | Marginal Products | 58 |
| 7.2.1 | Marginal Product of Labor | 59 |
| 7.2.2 | Marginal Product of Capital | 59 |
| 7.3 | Profit | 60 |
| 7.4 | Profit Maximization | 60 |
| III | Probability and Statistics | 63 |
| 8 | Probability and Counting Rules | 64 |
| 8.1 | Single Events | 64 |
| 8.2 | Multiple Events | 65 |
| 8.2.1 | The Counting Rule | 65 |
| 8.2.2 | Combinations and Permutations | 65 |
| 8.3 | Probability with Multiple Events | 68 |
| 8.4 | Joint Probability | 70 |
| 8.4.1 | The Intersection of two Events | 71 |
| 8.4.2 | The Union of two Events | 72 |
| 8.4.3 | Bayes Rule | 73 |
| 9 | Data Visualization | 76 |
| 9.1 | Collecting Data | 76 |
| 9.2 | Visualizing Data | 77 |
| 9.2.1 | Bar Graphs | 77 |
| 9.2.2 | Categorical Variables | 78 |
| 9.2.3 | Pie Chart | 79 |
| 9.2.4 | Histogram | 80 |
| 9.2.5 | Line Graph | 80 |
| 9.2.6 | Advanced Graphs | 82 |

| | |
|--|------------|
| 10 Descriptive Statistics | 83 |
| 10.1 Descriptive Statistics of Population Data | 83 |
| 10.1.1 Measures of Central Tendency | 83 |
| 10.1.2 Mode | 83 |
| 10.1.3 Measures of Variability | 84 |
| 10.1.4 Interquartile Range | 84 |
| 10.1.5 Coefficient of Variation | 85 |
| 10.2 Descriptive Statistics of a Sample | 86 |
| 10.2.1 Measures of Variability | 86 |
| 10.2.2 Coefficient of Variation | 87 |
| 10.2.3 Boxplot | 87 |
| 10.3 Relationship between two Variables | 87 |
| 10.3.1 Covariance | 88 |
| 10.3.2 Correlation | 88 |
| 10.4 Mean-Variance Analysis | 89 |
| 11 Probability Distributions | 90 |
| 11.1 Discrete Random Variables | 90 |
| 11.1.1 Probability Mass Function | 90 |
| 11.1.2 Cumulative Density Function | 91 |
| 11.1.3 Expected Value of a Discrete Variable | 92 |
| 11.2 Continuous Random Variables | 93 |
| 11.2.1 Probability Density Function | 93 |
| 11.2.2 Cumulative Density Function | 94 |
| 11.2.3 Expected Value of a Continuous Variable | 95 |
| 11.3 Common Probability Distributions | 96 |
| 11.3.1 The Normal Distribution | 96 |
| 11.3.2 Poisson Distribution | 101 |
| 12 Hypothesis Testing | 104 |
| 12.1 The T-Test | 104 |
| 12.1.1 One Sample T-Test | 105 |
| 12.1.2 Two-Sample T-Test: Independent Sampling | 108 |
| 12.2 The Z-Test | 109 |
| 12.3 P-Values | 110 |
| 12.4 Type I and Type II Errors | 111 |
| 12.4.1 Type I Error | 111 |
| 12.4.2 Type II Error | 111 |
| 12.5 Confidence Intervals | 111 |
| 12.5.1 Confidence Interval when Population Standard Deviation is Known | 111 |
| 12.5.2 Confidence Interval when Population Standard Deviation is Unknown | 112 |

| | |
|---|------------|
| IV Introduction to Finance | 114 |
| 13 The Time Value of Money | 115 |
| 13.1 Future Value | 115 |
| 13.1.1 Simple Interest | 115 |
| 13.1.2 Compound Interest | 116 |
| 13.2 Present Value | 117 |
| 13.3 Annuities | 118 |
| 13.3.1 Future Value of an Ordinary Annuity | 118 |
| 13.3.2 Present Value of an Ordinary Annuity | 119 |
| 13.4 Amortized Loans | 120 |

INTRODUCTION

In economics and business, we study how people and businesses make decisions. It is not possible to observe and study every person, so we use mathematical models to approximate the behavior we observe. We explore whether these models are correct by collecting data and using statistics to evaluate the theory. If the statistical analysis contradicts the models, professional researchers seek explanations for why the models are wrong and look to build extensions to the models.

Economics and business models use many quantitative tools to help us better understand the world we live in. These tools include algebra, calculus, and statistics, which are applied in economics, finance, operations management, marketing, and more. This book covers basic algebra rules, an introduction to calculus, a review of statistics, data visualization, and an introduction to finance. This book is not intended to be a substitute for algebra, calculus, statistics, and finance courses, but can serve as a book used in an undergraduate quantitative tools course, mathematical economics, or statistics course for economics and business majors, and a book that can be used to reference as students take intermediate and upper-level undergraduate electives and as review and preparation for master's and Ph.D. level economics and business courses.

Part I

Mathematical Concepts

REVIEW OF ALGEBRA

Many of the theories encountered in economics and business rely on mathematical models to help us understand the complexities seen in the real world. Algebra uses alphabetical letters and Greek symbols to represent variables and is the study of how to manipulate these variables. This chapter will cover the algebraic rules and properties that are necessary to solve economics and business models.

2.1 Exponents

An exponent is a number that represents how many times a number is multiplied times itself. In the expression b^n , b is the base and n is the exponent. This means the base, b is multiplied together n times. When encountering exponents, it is common to say b raised to the power of n .

Example

$$5^3 = 5 \cdot 5 \cdot 5 = 125$$

$$10^6 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 1,000,000$$

$$x^4 = x \cdot x \cdot x \cdot x$$

There are a set of rules that are necessary to solve algebraic problems when there are exponents.

Exponent of One

Anything raised to the power of 1 is always equal to itself.

$$\text{Rule: } x^1 = x \quad \text{Example: } 6^1 = 6$$

Exponent of Zero

Anything raised to the power of 0 is always equal to 1.

$$\text{Rule: } x^0 = 1 \quad \text{Example: } 6^0 = 1$$

Negative Exponents

A negative exponent represents the inverse. This means that the base could be expressed on the opposite side of the fraction with a positive exponent (i.e., if the base is in the numerator, it could be expressed in the denominator with a positive exponent and vice versa).

$$\text{Rule: } x^{-a} = \frac{1}{x^a} \text{ or } \frac{1}{x^{-a}} = x^a \quad \text{Example: } K^{-2} = \frac{1}{K^2} \text{ or } \frac{1}{K^{-2}} = K^2$$

Division with Exponents

$$\text{Rule: } \frac{x^a}{x^b} = x^{a-b} \quad \text{Example: } \frac{x^5}{x^2} = x^{5-2} = x^3$$

Multiplication with Exponents

$$\text{Rule: } x^a x^b = x^{a+b} \quad \text{Example: } x^2 x^3 = x^{2+3} = x^5$$

Power of a Fraction

$$\text{Rule: } \left(\frac{x}{y}\right)^a = \frac{x^a}{y^a} \quad \text{Example: } \left(\frac{x}{y}\right)^5 = \frac{x^5}{y^5}$$

Power of a Power

$$\text{Rule: } (x^a)^b = x^{a \cdot b} \quad \text{Example: } (x^2)^3 = x^{2 \cdot 3} = x^6$$

Exponents with Fractions

$$\text{Rule: } x^{a/b} = \sqrt[b]{x^a} \quad \text{Example: } x^{1/2} = \sqrt{x} \text{ or } x^{2/3} = \sqrt[3]{x^2}$$

2.2 Logarithms

A logarithm (typically abbreviated log) is the inverse of an exponent. In other words, it is the power to which a base must be raised to yield a specific number.

Logarithm Rule

$$\text{Rule: If } b^n = x, \text{ then } x = \log_b n \quad \text{Example: If } 2^3 = 8, \text{ then } \log_2 8 = 3$$

The natural logarithm (typically called the natural log) is commonly used in economics and business. The natural logarithm is a logarithm with a base of the mathematical constant e , which is approximately equal to 2.718. Therefore, if $e^n = x$ then $\log_e x = n$. $\log_e x = n$ is generally written as $\ln(x) = n$ for clarity. There are a set of rules for natural logarithms that are necessary when solving algebraic problems.

Natural Log of One

The natural log of 1 is always equal to 0.

$$\text{Rule: } \ln(1) = 0$$

Natural Log of Zero

The natural log of 0 does not exist and is always undefined.

| |
|--|
| Rule: $\ln(0) = \text{undefined}$ |
| Natural Log of e |
| Rule: $\ln(e) = 1$ |
| Natural Log Product Rule |
| Rule: $\ln(x \cdot y) = \ln(x) + \ln(y)$ Example: $\ln(x^2y^3) = \ln(x^2) + \ln(y^3)$ |
| Natural Log Quotient Rule |
| Rule: $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$ Example: $\ln\left(\frac{x^2}{y^3}\right) = \ln(x^2) - \ln(y^3)$ |
| Natural Log Power Rule |
| Rule: $\ln(x^a) = a \ln(x)$ Example: $\ln(x^2) = 2 \ln(x)$ |
| Natural Log Reciprocal Rule |
| Rule: $\ln\left(\frac{1}{x}\right) = -\ln(x)$ Example: $\ln\left(\frac{1}{3}\right) = -\ln(3)$ |
| Inverse of the Natural Log |
| Rule: $\ln(e^x) = x$ and $e^{\ln(x)} = x$ |

2.3 Algebra Rules and Operations

Economics and business models require solving and simplifying algebraic expressions and equations. When simplifying an algebraic expression or solving an equation, there are some rules, conventions, and “tricks” that can be helpful and are necessary.

Factoring

If multiple terms are added or subtracted together, a common term can be factored out. Suppose you have the following expression:

$$2x^3 + 4x^2 + 10x =$$

Each term in this equation has an x term and a term that can be multiplied by 2. Therefore, $2x$ can be factored out of this expression:

$$2x(x^2 + 2x + 5)$$

Distributive Property

When two terms are separated with parentheses, we can multiply them together by distributing the term outside of the parenthesis to each term inside the parentheses. This is the opposite of what we just did with factoring. We can use our previous example:

$$2x(x^2 + 2x + 5) = 0$$

$$2x^3 + 4x^2 + 10x = 0$$

If two expressions are multiplied together by parentheses, then we can use the “FOIL” method. Suppose we have the following expression:

$$(x + 1)(x + 2)$$

The “FOIL” method says we multiply the first terms of each expression together, the outside terms together, the inside terms together, and the last terms together.

$$(x + 1)(x + 2)$$

$$(x + 1)(x + 2) = x^2 + 2x + x + 3$$

$$(x + 1)(x + 2) = x^2 + 3x + 3$$

If an expression in parenthesis is raised to a power then you cannot distribute:

$$2x(x + 1)^2 \neq (2x^2 + 2x)^2$$

This is because $(x + 1)^2 = (x + 1)(x + 1)$, so there are technically three terms in this expression:

$$2x(x + 1)^2 = 2x(x + 1)(x + 1)$$

Once you expand the term raised to a power, then you can use the distributive property and “FOIL” method:

$$(2x^2 + 2x)(x + 1) = 2x^3 + 2x^2 + 2x^2 + 1$$

$$2x^3 + 4x^2 + 1$$

Multiplying by “1”

Anything divided by itself is equal to one: $\frac{3x}{3x} = 1$.

When trying to find a common denominator, we can use this principle. Suppose you have the following expression:

$$\frac{1}{x} + \frac{2}{x^2}$$

To simplify this expression, we need to find a common denominator, which in this example, is x^2 . To make both terms have a denominator of x^2 , we can multiply the first term by $\frac{x}{x}$, which equals 1.

$$\left(\frac{x}{x}\right) \frac{1}{x} + \frac{2}{x^2}$$

$$\frac{x}{x^2} + \frac{2}{x^2}$$

$$\frac{x + 2}{x^2}$$

Division Property of Equality

If two expressions are equal to each other and you divide both sides by the same number that is not equal to zero, the resulting expressions will also be equivalent. So if $a = b$ then

$$\frac{a}{c} = \frac{b}{c} \text{ as long as } c \neq 0.$$

$$3x = 24$$

$$\frac{3}{3}x = \frac{24}{3}$$

$$x = \frac{24}{3}$$

$$x = 8$$

Multiplication Property of Equality

If two expressions are equal to each other and you multiply both sides by the same number, the resulting expressions will also be equivalent. So, if $a = b$ then $ac = bc$.

$$\frac{10}{x} = 5$$

$$x \left(\frac{10}{x} \right) = 5(x)$$

$$10 = 5x$$

$$x = \frac{10}{5}$$

$$x = 2$$

Cross Multiplication

When a fraction is equal to another fraction, it can be solved by cross multiplication

$$\frac{3}{x} = \frac{x}{3}$$

$$\frac{3}{x} \times \frac{x}{3}$$

$$9 = x^2$$

$$\sqrt{9} = x$$

$$x = 3$$

Fraction Divided by Another Fraction

When a fraction is divided by another fraction, the expression can be simplified by multiplying the numerator by the reciprocal (inverse) of the denominator

$$\frac{a/b}{c/d} = \frac{a}{b} \times \frac{d}{c}$$

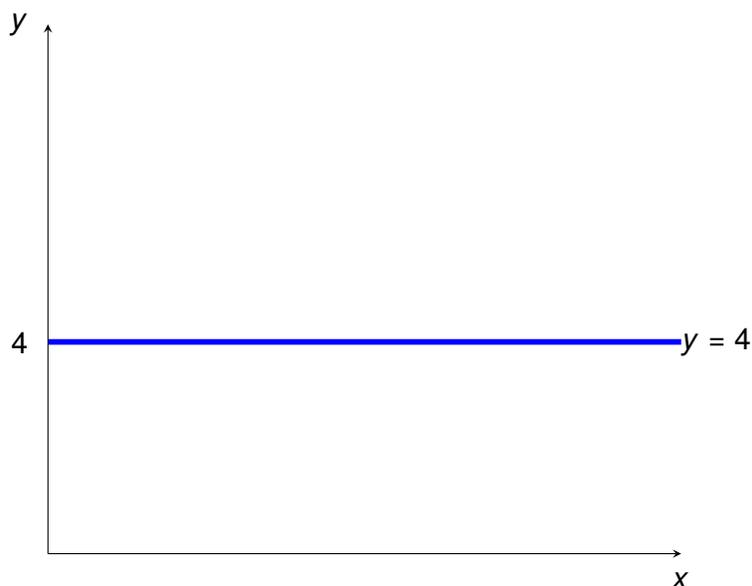
$$\frac{\left(\frac{1}{2}\right)}{\left(\frac{5}{7}\right)} = \frac{1}{2} \times \frac{7}{5} = \frac{7}{10}$$

$$\frac{\left(\frac{x^2}{y^2}\right)}{\left(\frac{2x}{y}\right)} = \frac{x^2}{y^2} \times \frac{y}{2x} = \frac{x}{2y}$$

2.4 Linear Equations in One Variable

A linear equation of one variable is an equation of a straight line. Figure 2.1 shows an example of a linear equation in one variable where $y = 4$.

Figure 2.1: Linear Equation with One Variable



In this equation, $y = 4$ for every value of x . If we have a linear equation with one variable, we can always find the solution to that equation, even if the variable shows up multiple times.

Example

$$5x + 2 = 3x - 6$$

$$2x = -8$$

$$x = -4$$

We could have a more complex equation where we have to apply the distributive property.

Example

$$4(x - 3) + 12 = 15 - 5(x + 6)$$

$$4x - 12 + 12 = 15 - 5x - 30$$

$$4x = -15 - 5x$$

$$9x = -15$$

$$x = -\frac{15}{9} = -\frac{5}{3}$$

2.5 Linear Equation with Two Variables

An equation is linear of two variables if it can be written as $ax + by + c = 0$. This expression can be rearranged using algebra rules and operations to get the equation of a line.

$$ax + by + c = 0 \tag{2.1}$$

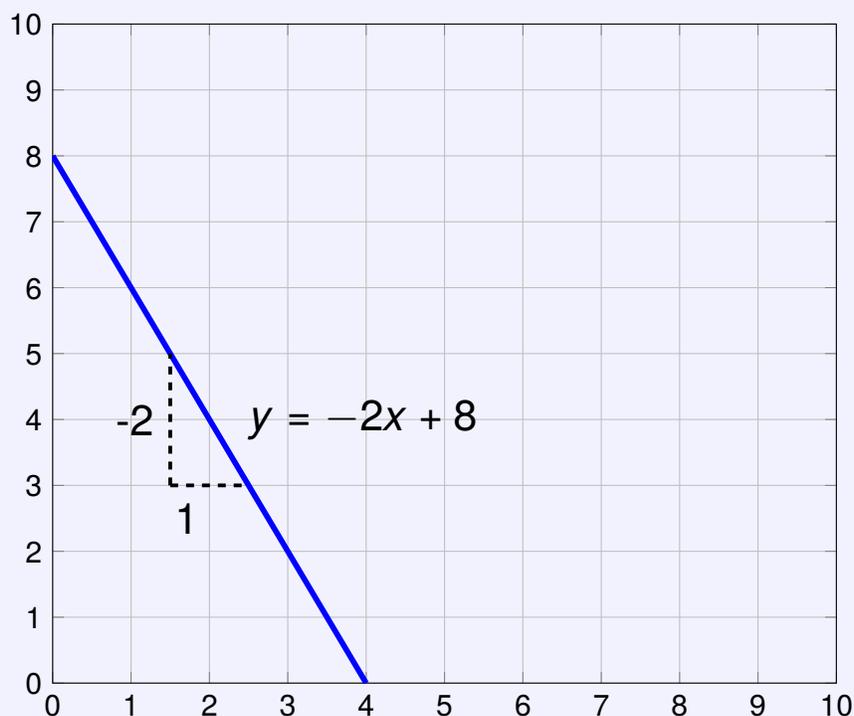
$$by = ax - c \tag{2.2}$$

$$y = \frac{a}{b}x - \frac{c}{b} \tag{2.3}$$

The equation of a line is more commonly expressed as $y = mx + b$, where m is the slope of the line and b is y-intercept. Slope measures the rate of change, or, as x changes, how much does y change.

Example

$$y = -2x + 8$$



If we plug in any point for y , we can find out what the x value would be on that line, and vice versa. Being able to plug in a value for x or y and solve for the other variable is a very important skill.

Example: If $y = 6$, what is x ?

$$6 = -2x + 8$$

$$6 - 8 = -2x$$

$$-2 = -2x$$

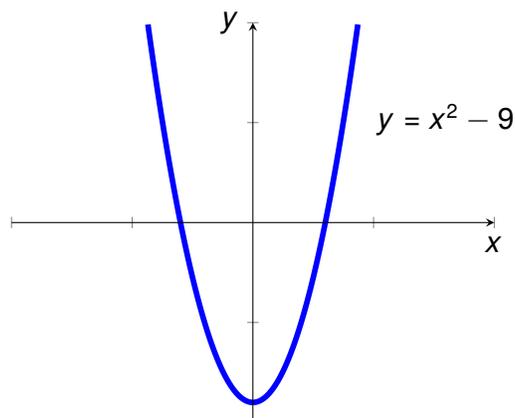
$$x = \frac{-2}{-2} = 1$$

2.6 Quadratic Equations

A quadratic equation contains a variable that is raised to the power of 2 (e.g., x^2) and takes the form of $ax^2 + bx + c = 0$. Figure 2.2 shows a graphical example of a quadratic equation.

When solving a quadratic equation, we want to find what the values of x when $y = 0$. In other words, we want to find the two x values where the graph crosses the x -axis. Depending on the specific equation and shape of the graph, there could be 0, 1, or 2 solutions to a quadratic equation

Figure 2.2: Graphical Example of a Quadratic Equation



If a quadratic equation does not have a linear term (e.g., $ax^2 + c = 0$), like the graph shown in Figure 2.2, then it can be solved using the square root property:

Square Root Property of Quadratic Equations

$$\text{If } x^2 = k, \text{ then } x = \pm\sqrt{k}$$

Example

Using the equation from Figure 2.2, we can see that this graph crosses the x -axis at two points, one point is negative and one point is positive. Since $y = x^2 - 9$ does not have a linear term, we can solve this using the square root property.

$$x^2 - 9 = 0$$

$$x^2 = 9$$

$$x = \pm\sqrt{9}$$

$$x = 3 \text{ and } x = -3$$

If a quadratic equation has a linear term (e.g., $ax^2 + bx + c = 0$), then there are a few ways to solve this equation: factoring, grouping, or the quadratic formula. No matter the type of quadratic equation, it can *always* be solved using the quadratic formula.

The Quadratic Formula

$$\text{If } ax^2 + bx + c = 0, \text{ then } x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Example

$$4x^2 - 6x + 2 = 0$$

$$a = 4 \quad b = -6 \quad c = 2$$

$$x = \frac{6 \pm \sqrt{(-6)^2 - 4(4)(2)}}{2(4)}$$

$$x = 1 \text{ and } x = \frac{1}{2}$$

If a quadratic equation does not cross the x-axis, then there are no real solutions to the problem. If you try to solve for the roots of one of these equations using the quadratic formula, you will get the square root of a negative number. The solution to those problems involve imaginary numbers, which are not something encountered in economics and business and thus are beyond the scope of this book.

2.7 Rational Equations

A rational equation is an equation, which typically has fractions, that can be converted into a linear equation with some algebraic manipulation. In order to add or subtract fractions, they must have the same denominator. We can use algebraic rules from section 2.3 to find a common denominator and solve a rational equation.

Example

$$\frac{7}{2x} - \frac{5}{3x} = \frac{22}{3}$$

To solve this problem, we need to find a common denominator. One trick to find this is to multiply the terms that are different, together. For example, $2 \times 3 = 6$. So, we can use $6x$ as the common denominator.

We can use the “Multiplying by ‘1’” rule to get a common denominator:

$$\left(\frac{3}{3}\right) \frac{7}{2x} - \left(\frac{2}{2}\right) \frac{5}{3x} = \frac{22}{3}$$

$$\frac{21}{6x} - \frac{10}{6x} = \frac{22}{3}$$

$$\frac{11}{6x} = \frac{22}{3}$$

From here, we can use cross multiplication to get a linear equation:

$$132x = 33$$

Now, we can solve for x :

$$x = \frac{33}{132} = \frac{1}{4}$$

2.8 Systems of Equations

Suppose you have multiple equations and multiple unknown variables and you want to figure the solutions for each of those variables. This is called a system of equations. We are going to assume that the number of equations and the number of unknown variables are the same (if they are not, it gets more complicated). There are three ways to solve a system of equations:

- The substitution method
 - Rearrange one equation in terms of a single variable
 - Substitute the rearranged equation into the other equation
- The elimination method
 - Rearrange both equations so that one variable is by itself (same variable for both equations)
 - Set them equal to each other
- Graphing

Example

Suppose you have the following system of equations:

$$y - 2 = 5x$$

$$y = 10x - 4$$

Substitution Method

Since $y = 10x - 4$, substitute this equation into the other one:

$$10x - 4 - 2 = 5x$$

Now there is a linear equation of one variable.

$$10x - 6 = 5x$$

$$5x = 6$$

$$x = \frac{6}{5}$$

Plug $x = \frac{6}{5}$ into either equation to solve for y

$$y = 10 \left(\frac{6}{5} \right) - 4$$

$$y = \frac{60}{5} - 4 = 8$$

Elimination Method

Rearrange both equations in terms of y :

$$y = 5x + 2$$

$$y = 10x - 4$$

Replace y in the first equation with the y in the second equation to create an equation that is all of one variable, x , then you can solve for x :

$$10x - 4 = 5x + 2$$

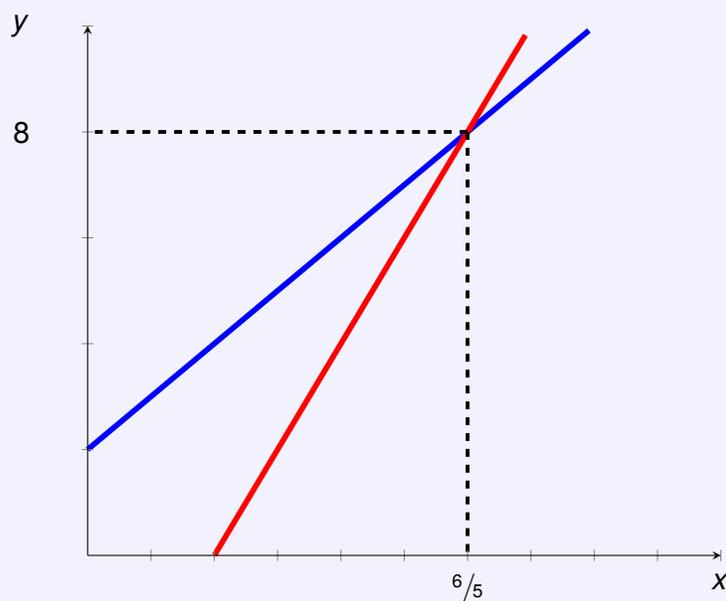
$$10x - 5x = 2 + 4$$

$$5x = 6$$

$$x = \frac{6}{5}$$

$$y = \frac{60}{5} - 4 = 8$$

Graphing



INTRODUCTION TO DIFFERENTIAL CALCULUS

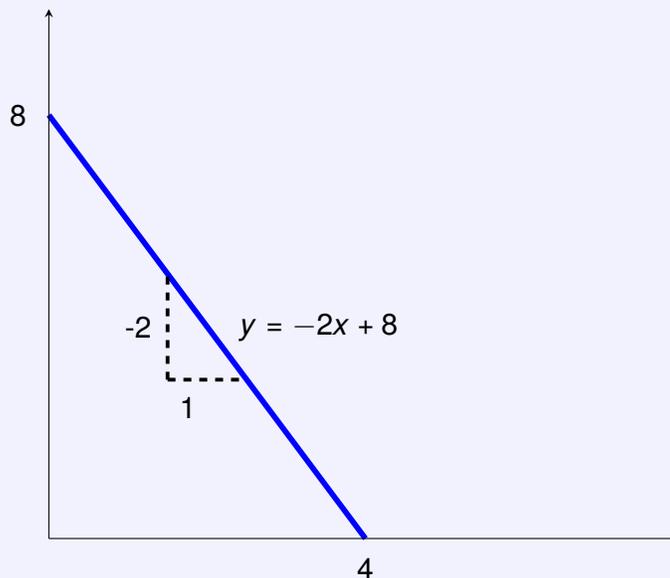
Differential calculus studies the rate at which quantities change. In economics and business, differential calculus is used to see how the quantity of one variable changes with respect to another variable. For example, we may want to know how the number of workers a company hires changes as the wage changes. The tool we use to answer that question is called a derivative.

3.1 The Derivative

From the previous section, we know that the equation of a line is $y = mx + b$ where m is the slope. This means that as x changes by 1, y changes by m . For example, if $y = -2x + 8$, the slope is -2.

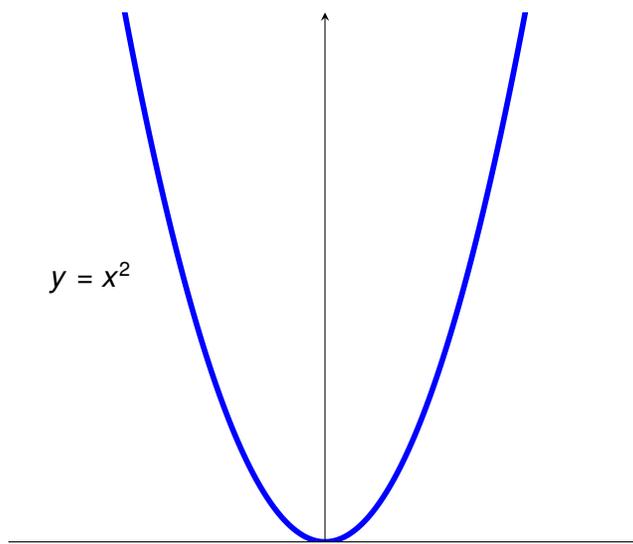
Example

Suppose that the equation of a line is $y = -2x + 8$. The slope of this line is -2. For every value of x , the slope of the line is -2.



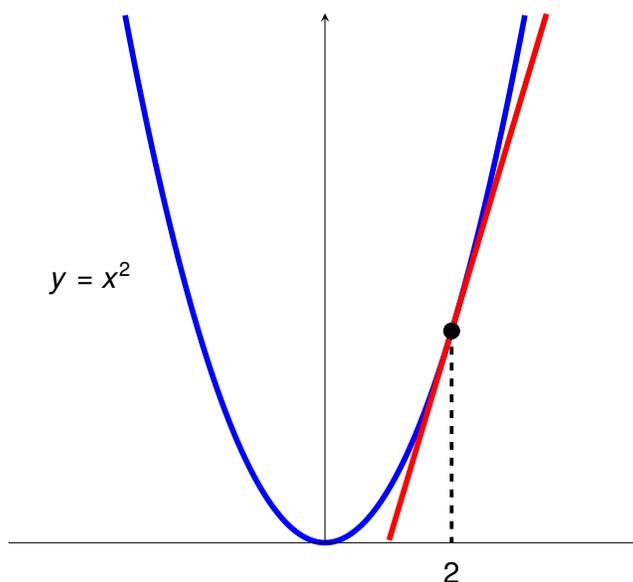
When an equation is linear, determining the rate of change is simple because it is the slope and the slope is constant. When an equation is not linear, determining the slope is not as simple. Let's consider the equation $y = x^2$. In this example, the slope when $x = 2$ is not the same as when $x = 5$. This means that an increase in x does not have the same impact on y as x

increases.



Since the slope of $y = x^2$ is different for every value of x , we need a method to determine the slope at each value of x . The derivative of a function yields an equation that we can use to calculate the slope of a function for any value of x . A derivative measures the instantaneous rate of change in y as x changes at a given point.

A way to visualize the derivative of a function at a specific value of x is to find the slope of a tangent line at x . A tangent line is a line that intersects the function at only one point without crossing the function. In our example of $y = x^2$, if we want to know the derivative when $x = 2$, then we would need to find the slope of the line that is tangent to $y = x^2$ when $x = 2$.



For any function, we can find the slope of the tangent line at any point. The formal definition and proof of a derivative can be found in the appendix to this chapter, but it is not necessary as a practitioner to calculating and interpreting a derivative. The next section will cover the methods to calculating the types of derivatives you will encounter in economics and business (note: this is a subset of the types of derivatives you will encounter in a typical calculus course).

3.2 Calculating Derivatives

For a function $y = f(x)$, then the notation used for a derivative is $\frac{dy}{dx}$.¹ $\frac{dy}{dx}$ means the derivative of dependent variable y with respect to independent variable x . This section will cover the methods for taking derivatives of common functions encountered in economics and business.

3.2.1 Functions with Exponents

For functions that have a variable with an exponent (e.g., $y = x^3$), the derivative can be calculated using the Power Rule.

The Power Rule

If the function is $y = x^n$ then the derivative is $\frac{dy}{dx} = nx^{n-1}$

Example 1

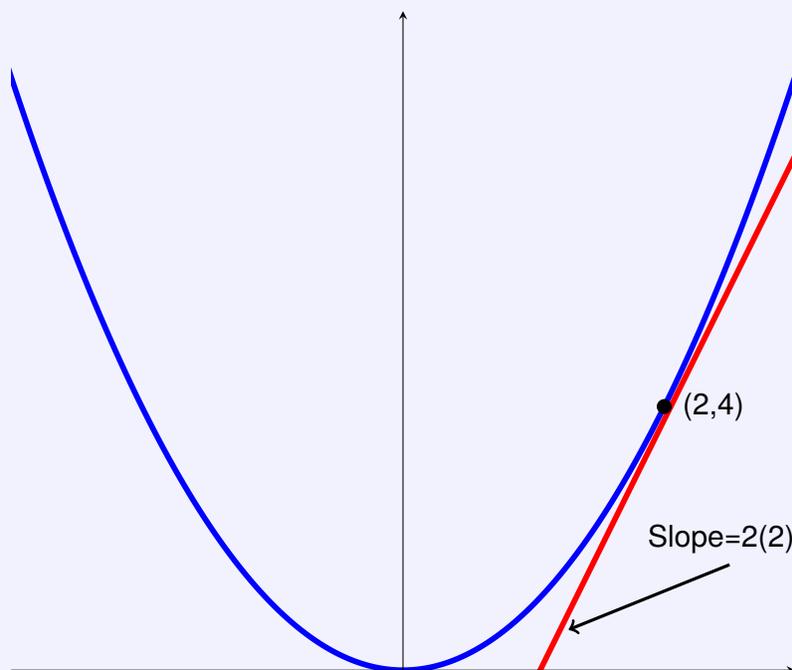
If $y = x^2$, then $n = 2$ in for this function.

This means the derivative, $\frac{dy}{dx} = 2x^{2-1} = 2x$

So what does this mean? The derivative gives us an equation that tells us what the slope (rate of change) is for any value of x .

For example, if $x = 2$, then the slope at that point = $2(2) = 4$ (see this on the graph below).

¹ In some books, $f'(x)$ or $f_x(x)$ are used as notations for the derivative



Using the derivative, we can calculate the slope for other values of x as well. If $x = -3$, then the slope $= 2(-3) = -6$.

Example 2

If $y = x^3$, the $n = 3$

$$\frac{dy}{dx} = 3x^{3-1} = 3x^2$$

Example 3

If $y = x$, this is the same as $y = x^1$, so $n = 1$

$$\frac{dy}{dx} = 1x^{1-1} = x^0 = 1$$

Example 4

If $y = 3$, this is the same as $y = 3x^0$, so $n = 0$

$$\frac{dy}{dx} = (0 \times 3)x^{0-1} = 0$$

If you think of this graphically, it is a horizontal line. For each value of x , y does not change. For any “constant value” (i.e., where x^0), then $\frac{dy}{dx}$ *always* equals 0.

It is often common to have functions with exponents that are multiplied by a constant number. For these functions, we need to expand the power rule.

The Expanded Power Rule

If the function is $y = cx^n$, where c is any constant number, then the derivative is

$$\frac{dy}{dx} = (c \times n)x^{n-1}$$

Example

If $y = 5x^3$, then $n = 3$, and $c = 5$

$$\text{So, } \frac{dy}{dx} = (5 \times 3)x^{3-1} = 15x^2$$

3.2.2 The Sum and Difference Rules

When a function has terms that are added or subtracted together, the derivative can be found using the sum or the difference rule. With these rule, the derivative of each “part” is taken independently and the addition or subtraction sign remains unchanged.

The Sum/Difference Rule

$$\text{If } y = f(x) + g(x) \text{ then } \frac{dy}{dx} = \frac{df(x)}{dx} + \frac{dg(x)}{dx}$$

Example

Suppose $y = 5x^3 + 6x$

$$\text{Part 1: } f(x) = 5x^3, \text{ so } \frac{df(x)}{dx} = 15x^2$$

$$\text{Part 2: } g(x) = 6x, \text{ so } \frac{dg(x)}{dx} = 6$$

$$\text{Therefore, } \frac{dy}{dx} = 15x^2 + 6$$

3.2.3 The Product Rule

When a function is the combination of two functions that are multiplied together (e.g., $y = (x^3 + 1)(x^2 - 2)$), the derivative can be found using the product rule.

The Product Rule

$$\text{If } y = f(x)g(x), \text{ then } \frac{dy}{dx} = f(x)\frac{dg(x)}{dx} + g(x)\frac{df(x)}{dx}$$

In English: $\frac{dy}{dx} = (\text{First Function} \times \text{Derivative of the Second Function}) + \text{Second Function} \times \text{Derivative of the First Function}$

Example

Suppose $y = (x^3 + 1)(x^2 - 2)$

Function 1: $x^3 + 1$

Function 2: $x^2 - 2$

Derivative of Function 1: $3x^2$

Derivative of Function 2: $2x$

$$\frac{dy}{dx} = (x^3 + 1) \cdot 2x + (x^2 - 2) \cdot 3x^2$$

We can simplify this derivative:

$$\frac{dy}{dx} = 5x^4 - 6x^2 + 2x$$

$$\frac{dy}{dx} = x(5x^3 - 6x + 2)$$

3.2.4 The Quotient Rule

If a function is a combination of two functions divided by each other (e.g., $y = \frac{5x + 1}{x^2}$), we need to use the quotient rule to find the derivative.

The Quotient Rule

$$\text{If } y = \frac{f(x)}{g(x)}, \text{ then } \frac{dy}{dx} = \frac{g(x) \cdot \frac{df(x)}{dx} - f(x) \cdot \frac{dg(x)}{dx}}{g(x)^2}$$

In English: $\frac{dy}{dx} =$

$$\frac{(\text{Function in Denominator}) \times (\text{Derivative of the Numerator}) - (\text{Function in numerator}) \times (\text{Derivative of the Denominator})}{(\text{Function in the Denominator})^2}$$

A helpful mnemonic that can be used to remember the quotient rule is “low d’high minus high d’low all over low low.”

Example

Suppose $y = \frac{5x + 1}{x^2}$

Numerator: $5x + 1$

Denominator: x^2

Derivative of the Numerator: 5

Derivative of the Denominator: $2x$

$$\frac{dy}{dx} = \frac{(x^2 \cdot 5) - [(5x + 1) \cdot 2x]}{x^2 \cdot x^2}$$

We can simplify this derivative:

$$\frac{dy}{dx} = \frac{-5x^2 - (10x^2 - 2x)}{x^4}$$

$$\frac{dy}{dx} = \frac{-(5x + 2)}{x^3}$$

3.2.5 The Chain Rule

Suppose you want to take the derivative of a function inside of another function (e.g., $y = (x^3 + x^2)^2$). To find this derivative, we will need to use the chain rule.

The Chain Rule

$$\text{If } y = f(g(x)), \text{ then } \frac{dy}{dx} = \frac{df(x)}{dx} \times \frac{dg(x)}{dx}$$

In English: $\frac{dy}{dx} = (\text{Derivative of the Inside Function}) \times (\text{Derivative of the Outside Function while Ignoring the Inside Function})$

Example 1

Suppose $y = (x^3 + x^2)^2$

Step 1: Define the inside function: $g(x) = x^3 + x^2$

Step 2: Take the derivative of the inside function: $\frac{dg(x)}{dx} = 3x^2 + 2x$

Step 3: Define the outside function by replacing the inside function with x : $y = (z)^2$ where $z = x^3 + x^2$

Step 4: Take the derivative of the outside function as defined in Step 3: $2z$

Step 5: Apply the chain rule: $(3x^2 + 2x) \times 2z$

Step 6: Substitute in for z : $(3x^2 + 2x) \times 2(x^3 + x^2)$

Step 7: Simplify: $\frac{dy}{dx} = 2x^3(3x^2 + 5x + 2)$

Example 2

Suppose $y = \ln(x^3)$

Step 1: Define the inside function: $g(x) = x^3$

Step 2: Take the derivative of the inside function: $\frac{dg(x)}{dx} = 3x^2$

Step 3: Define the outside function by replacing the inside function with x : $y = \ln(z)$ where $z = x^3$

Step 4: Take the derivative of the outside function as defined in Step 3: $\frac{1}{z}$

Step 5: Apply the chain rule: $\left(\frac{1}{z}\right) \times 3x^2$

Step 6: Substitute in for z : $\left(\frac{1}{x^3}\right) \times 3x^2$

Step 7: Simplify: $\frac{dy}{dx} = \frac{3x^2}{x^3} = \frac{3}{x}$

3.2.6 Derivatives of Different Types of Functions

Up until this point, we have been discussing functions with exponents. There are other types of functions, such as logarithms, power functions, and trigonometric functions, that also have derivatives. Trigonometric functions are rarely used in economics and business models, so this book will not cover them. However, it is not uncommon to see natural logarithms or power functions.

Derivative of a Natural Log

If $y = c \ln(x)$, where c is any constant, then $\frac{dy}{dx} = \frac{c}{x}$

Example

Suppose $y = 5 \ln(x)$

$$\frac{dy}{dx} = \frac{5}{x}$$

Derivative of a Power Function

If $y = c^x$, where c is any constant, then $\frac{dy}{dx} = c^x \ln(c)$

Example 1

Suppose $y = 6^x$

$$\frac{dy}{dx} = 6^x \ln(6)$$

Example 2

Suppose $y = e^x$

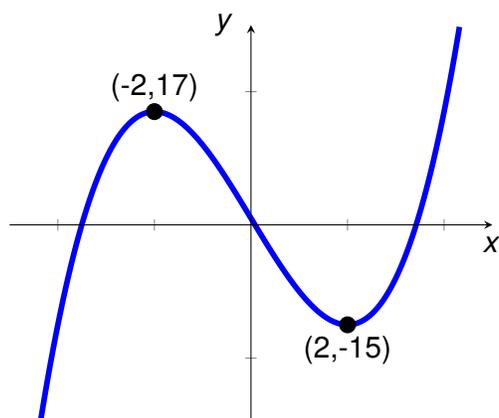
$$\frac{dy}{dx} = e^x \ln(e) = e^x$$

3.3 Finding Minima and Maxima using Derivatives

We often look to find minimum or maximum values in economics and business models. For example, a company may want to know the number of workers it can hire to maximize profits or how much capital to use that minimize costs. We can use differential calculus to find these points on any function without graphing.

Let's use $y = x^3 - 12x + 1$ as an example. Figure 3.1 shows this function graphically. This function has a maximum at the point $(-2, 17)$ and a minimum at the point $(2, -15)$.

Figure 3.1: Finding Minima and Maxima



We can also find these points using derivatives. Recall, that a derivative tells us the slope at any given point. At a maximum point, the slope of a function changes from increasing in y to decreasing in y , and at a minimum point, the slope changes from decreasing in y to increasing in y . At the point where the slope changes direction, the tangent line would be a vertical line

with a slope of zero. The points of a function where the slope is equal to zero are called critical points. We can find all the critical points of a function by finding all the values of x that make the derivative equal to zero.

Example

If $y = x^3 - 12x + 1$, then $\frac{dy}{dx} = 3x^2 - 12$

Set $\frac{dy}{dx} = 0$, and solve for x :

$$3x^2 - 12 = 0$$

$$3x^2 = 12$$

$$x^2 = 4$$

$$x = \sqrt{4}$$

$$x = 2 \text{ and } x = -2$$

To determine if a critical point is a minima or a maxima, we need to use either the first or second derivative test.

3.3.1 First Derivative Test

We know that each critical point is an “inflection” point, where the function changes from increasing to decreasing (a maxima) or decreasing to increasing (a minima). To conduct the first derivative test, we are going to split the function into segments before and after each critical point. In each one of these segments, we are going to determine if the function is increasing or decreasing.

First Derivative Test

If $\frac{dy}{dx} > 0$, then the function is increasing.

If $\frac{dy}{dx} < 0$, then the function is decreasing.

Example

We found that for the function $y = x^3 - 12x + 1$, it has two critical points at $x = 2$ and $x = -2$.

Therefore, there are three segments to this function: $x = -\infty$ to $x = -2$, $x = -2$ to $x = 2$, and $x = 2$ to $x = \infty$

Pick a point in each segment, and plug it into $\frac{dy}{dx}$ to determine the sign:

| Segment | Point | Derivative |
|---------------------------|-------|-------------------------------------|
| $x = -\infty$ to $x = -2$ | -3 | $\frac{dy}{dx} = 3(-3)^2 - 12 = 15$ |
| $x = -2$ to $x = 2$ | 0 | $\frac{dy}{dx} = 3(0)^2 - 12 = -12$ |
| $x = 2$ to $x = \infty$ | 3 | $\frac{dy}{dx} = 3(3)^2 - 12 = 15$ |

Create the following table, with one row for each segment:

| Segment | Sign of $\frac{dy}{dx}$ | Increasing/Decreasing |
|---------------------------|-------------------------|-----------------------|
| $x = -\infty$ to $x = -2$ | Positive | Increasing |
| $x = -2$ to $x = 2$ | Negative | Decreasing |
| $x = 2$ to $x = \infty$ | Positive | Increasing |

Because the function switches from increasing to decreasing when $x = -2$, this means that the critical point is a maximum.

Because the function switches from decreasing to increasing when $x = 2$, this means that the critical point is a minimum.

3.3.2 Second Derivative Test

A second method, and more common method, to finding the minima and maxima of a function is the second derivative test. Before we can discuss the test, we need to introduce the second derivative. The second derivative is the derivative of the first derivative. The second derivative has the following notation: $\frac{d^2y}{dx^2}$.²

We can evaluate the second derivative at each of the critical points to determine if they are a minima or maxima.

Second Derivative Test

If $\frac{d^2y}{dx^2} > 0$, then the critical point is a minimum.

² In other books, $f''(x)$ or $f_{xx}(x)$ are used as notations for the second derivative.

If $\frac{d^2y}{dx^2} < 0$, then the critical point is a maximum.

If $\frac{d^2y}{dx^2} = 0$ or if $\frac{d^2y}{dx^2}$, then the second derivative test is inconclusive and we need to use the first derivative test.

Example

Using our previous example $y = x^3 - 12x + 1$, we know that $\frac{dy}{dx} = 3x^2 - 12$ and that there are two critical points: $x = 2$ and $x = -2$.

The second derivative is the derivative of the derivative: $\frac{d^2y}{dx^2} = 6x$.

We can evaluate $\frac{d^2y}{dx^2}$ at each critical point:

When $x = 2$, $\frac{d^2y}{dx^2} = 6(2) = 12$. $\frac{d^2y}{dx^2} > 0$, therefore there is a minimum at $x = 2$.

When $x = -2$, $\frac{d^2y}{dx^2} = 6(-2) = -12$. $\frac{d^2y}{dx^2} < 0$, therefore there is a maximum at $x = -2$.

3.4 Multivariate Calculus: Partial Derivatives

So far, all the equations we have looked at have had y as the dependent variable and x as the independent variable. However, it is quite common in economics and business models to encounter multiple independent variables that influence the dependent variable. For example, the price of a good or service may depend on the quantity demanded of that good or service as well as the quantity demanded of complements and substitutes, or, the profit of a firm may depend on the quantity of multiple inputs (capital and labor). Examples of functions in economics and business models may look like $y = x^2 + 2z^3$, $Q = L^{1/3}K^{2/3}$, or $L = 10y + 2r$.

In order to determine how a change in an independent variable affects the dependent variable *ceteris paribus* (all other independent variables remaining constant), we need to introduce the partial derivative. The rules of partial derivatives are the same as a derivative and the dependent variables we are not interested in will be treated as if they are a constant. The notation for a partial derivative is $\frac{\partial y}{\partial x}$. The interpretation is the partial derivative of the independent variable y with respect to the dependent variable x , while all other dependent variables are held constant.

Example

Suppose that we have a function: $y = x^{1/3}z^{2/3}$

First, lets find how y changes as x changes $\left(\frac{\partial y}{\partial x}\right)$:

- We are going to treat z as if it is a constant
- To do this, define c to be the variable that has z : $c = z^{2/3}$
- Substitute c into y : $y = cx^{1/3}$
- Now we have a simple derivative like we did earlier
- $\frac{\partial y}{\partial x} = \left(\frac{1}{3} \times c\right) x^{(1/3)-1} = (c)\frac{1}{3}x^{-2/3}$
- Substitute c back into y and simplify: $\frac{\partial y}{\partial z} = \frac{1}{3}x^{-2/3}z^{2/3}$

Now, lets find out how y changes as z changes $\left(\frac{\partial y}{\partial z}\right)$:

- We are going to treat x as if it is a constant
- Define c to be the variable that has x : $c = x^{1/3}$
- Substitute c into y : $y = cz^{2/3}$
- Now we have a simple derivative:
- $\frac{\partial y}{\partial z} = \left(\frac{2}{3} \times c\right) z^{(2/3)-1} = (c)\frac{2}{3}z^{-1/3}$
- Substitute c back into y and simplify: $\frac{\partial y}{\partial z} = \frac{2}{3}z^{-1/3}x^{1/3}$

CONSTRAINED OPTIMIZATION

Constrained optimization is the process of finding the minimum or maximum values of a function that is subject to one or more constraints.

4.1 The Method of Lagrange Multipliers

Suppose that you want to maximize $f(x, y)$ subject to the constraint $g(x, y) = c$. The notation that is used to show a constrained optimization is:

$$\begin{aligned} \max \quad & f(x, y) \\ \text{s.t.} \quad & g(x, y) = c \end{aligned} \quad (4.1)$$

The Method of Lagrange Multipliers is the mathematical technique to evaluate this constrained optimization problem. To do this, we must establish the Lagrangian, \mathcal{L} :

$$\mathcal{L} = f(x, y) + \lambda[c - g(x, y)] \quad (4.2)$$

λ in equation 4.2 is called the Lagrange multiplier. The Lagrange multiplier measures the increase in $f(x, y)$ that is obtained by relaxing the constraint. In other words, if we increased c by one unit, how much would $f(x, y)$ increase? There are economic meanings to λ that will be discussed in future sections and other economics courses that use this method.

To solve the Lagrangian, we will need to find three partial derivatives: $\frac{\partial \mathcal{L}}{\partial x}$, $\frac{\partial \mathcal{L}}{\partial y}$, $\frac{\partial \mathcal{L}}{\partial \lambda}$ and set them equal to zero.

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f(x, y)}{\partial x} + \lambda \frac{\partial g(x, y)}{\partial x} = 0 \quad (4.3)$$

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f(x, y)}{\partial y} + \lambda \frac{\partial g(x, y)}{\partial y} = 0 \quad (4.4)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = c - g(x, y) = 0 \quad (4.5)$$

Equations 4.3, 4.4, and 4.5 form a system of three equations with three unknown variables, x , y , and λ . We can solve this system of equations for x^* , y^* , and λ^* .

Example

We want to find the values of x and y that maximize $f(x, y) = xy$ subject to the constraint $x + y = 1$.

$$\begin{aligned} \max \quad & xy \\ \text{s.t.} \quad & x + y = 1 \end{aligned} \tag{4.6}$$

The first step is to set up the Lagrangian.

$$\mathcal{L} = xy + \lambda[1 - (x + y)] \tag{4.7}$$

Next, we need to find $\frac{\partial \mathcal{L}}{\partial x}$, $\frac{\partial \mathcal{L}}{\partial y}$, $\frac{\partial \mathcal{L}}{\partial \lambda}$ and set them equal to zero.

$$\frac{\partial \mathcal{L}}{\partial x} = x - \lambda = 0 \tag{4.8}$$

$$\frac{\partial \mathcal{L}}{\partial y} = y - \lambda = 0 \tag{4.9}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - x - y = 0 \tag{4.10}$$

We can rearrange equations 4.8 and 4.9 by solving them for λ and using the elimination method:

$$\lambda = x \tag{4.11}$$

$$\lambda = y \tag{4.12}$$

Therefore,

$$x = y \tag{4.13}$$

We can substitute equation 4.13 into equation 4.10 and solve for y :

$$1 - y - y = 0 \tag{4.14}$$

$$1 = 2y \tag{4.15}$$

$$y^* = \frac{1}{2} \tag{4.16}$$

We can substitute y^* into equation 4.13 to solve for x^* :

$$x^* = \frac{1}{2} \tag{4.17}$$

To find λ^* , we can substitute x^* and y^* into equations 4.8 or 4.9:

$$\lambda^* = \frac{1}{2} \tag{4.18}$$

The solution to this constrained optimization problem is $x^* = \frac{1}{2}$, $y^* = \frac{1}{2}$, and $\lambda^* = \frac{1}{2}$

Part II

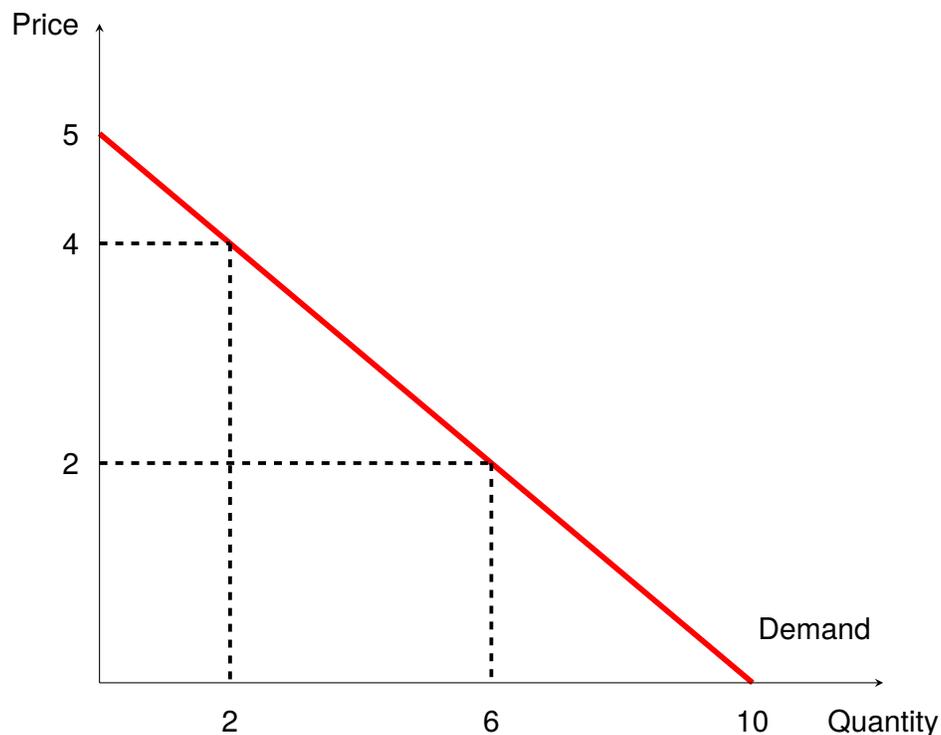
Solving Economic Models

SUPPLY, DEMAND, AND ELASTICITY

A market consists of people who buy goods and services (buyers) and firms who sell goods and services (sellers). Buyers determine demand for a good or service, and sellers determine supply. Markets can vary. Some markets are highly organized, others are not. Some markets have many buyers and sellers; others do not. You will explore these nuances in greater detail in a course on principles and intermediate microeconomics. Here, we will focus on solving and understanding market outcomes in a perfectly competitive market. This will allow us to determine how prices are determined and the impact of government intervention (e.g., taxes, subsidies, and regulation) on prices and output.

5.1 Demand

Demand is the relationship between the price of a good or service and the amount that that good or service is purchased (i.e., quantity demanded). The Law of Demand states that as the price of a good or service increases, the quantity demanded will decrease and vice versa.



Example

$$Q_D = 10 - 2P$$

If $P = 2$, then $Q_D = 10 - 2(2) = 6$

If $P = 4$, then $Q_D = 10 - 2(4) = 2$

If we want to use the demand equation to show it graphically, we need to rearrange it:

$$Q_D = 10 - 2P \quad (5.1)$$

$$2P = 10 - Q_D \quad (5.2)$$

$$P = 5 - \frac{1}{2}Q_D \quad (5.3)$$

This rearranged equation is called inverse demand function. This is because the price is on the vertical axis and the quantity demanded is on the horizontal axis.

5.2 Supply

Supply is the relationship between the price of a good or service and the amount that the company produces (i.e., quantity supplied). The Law of Supply states that as the price of a good or service increases, the quantity supplied will increase, and vice versa.

Example

$$Q_S = 4 + 2P$$

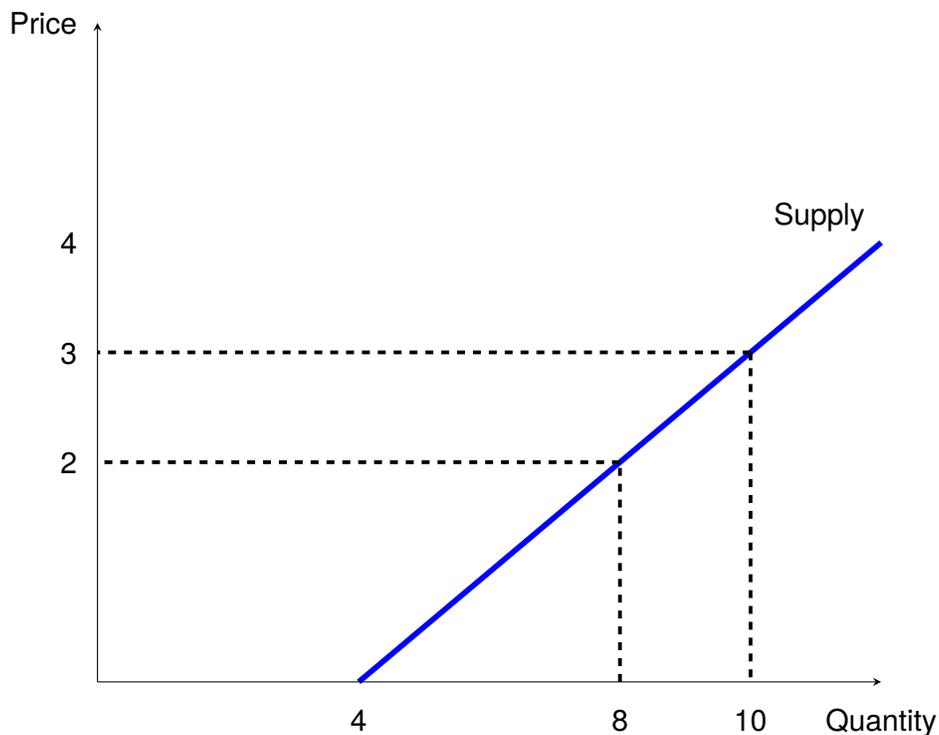
If $P = 2$, then $Q_S = 4 + 2(2) = 8$

If $P = 4$, then $Q_S = 4 + 2(4) = 12$

If we want to use the supply equation to show it graphically, we need to rearrange it the same way as we did with demand, which would be called the inverse supply function.

5.3 Defining a Market

A market is where the exchange of goods and services occurs as a result of buyers and sellers being in contact with one another, either directly or through mediating agents or institutions.



Basically, a market is a place where things are bought and sold. A market is characterized by a system of equations for Q_S and Q_D . From our examples, we have:

$$Q_D = 10 - 2P \quad (5.4)$$

$$Q_S = 4 + 2P \quad (5.5)$$

We can view this system of equations graphically:

A market is considered in equilibrium when market supply and demand balance each other and, as a result, prices become stable. When the market is in equilibrium, $Q_D = Q_S$. The solution of the system of equations, Q_D and Q_S , gives us the resulting stable price and the quantity of that good or service that exists in the market.

Example

$$Q_D = 10 - 2P$$

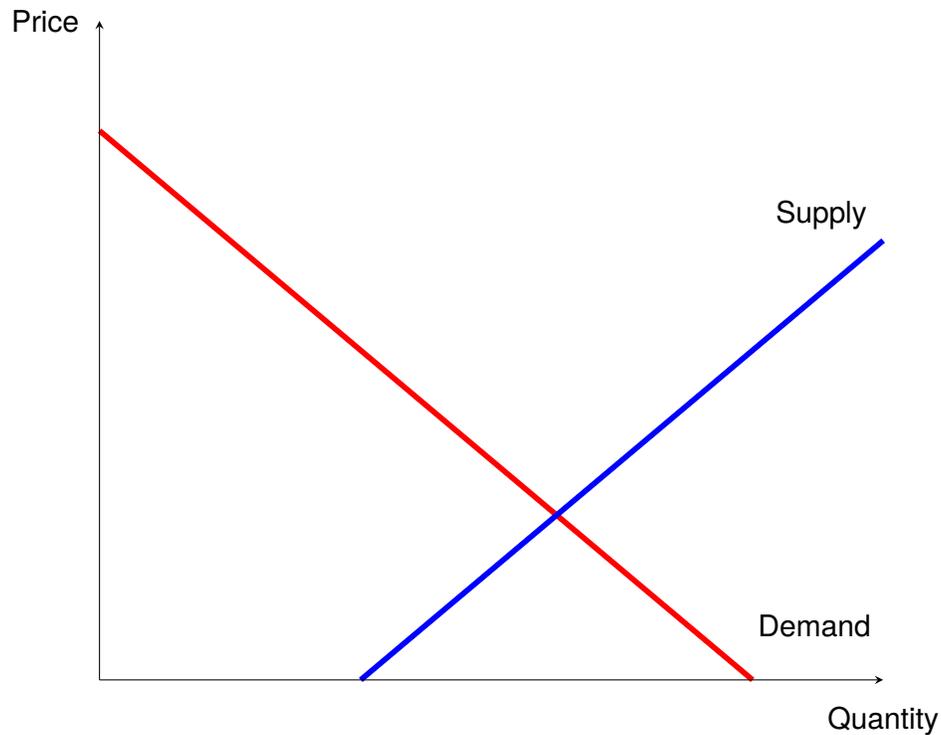
$$Q_S = 4 + 2P$$

Set $Q_D = Q_S$, solve for the equilibrium price, P^* :

$$4 + 2P = 10 - 2P$$

$$4P = 6$$

$$P^* = \frac{6}{4} = 1.5$$



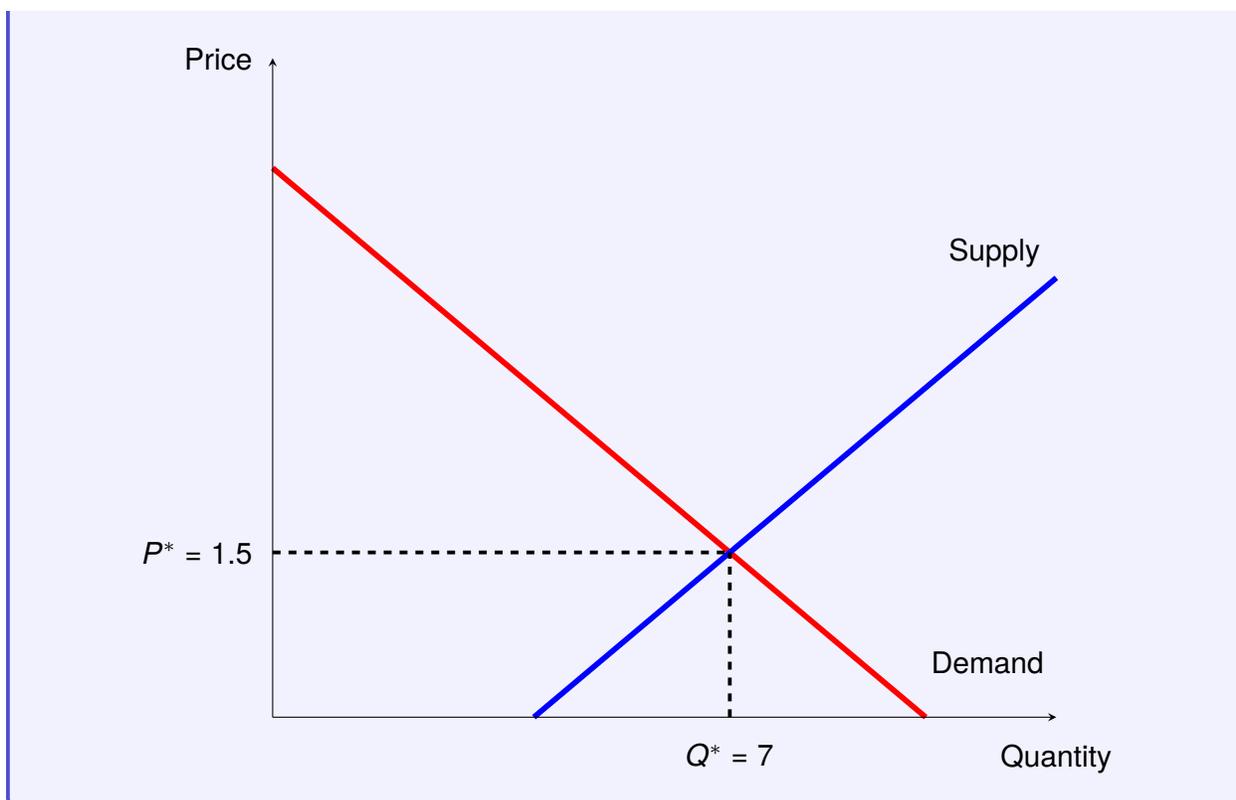
Substitute P^* into Q_S or Q_D to solve for the equilibrium quantity, Q^* (you will get the same answer regardless of which equation you plug it in to:

$$Q_D - 2(1.5) = 7$$

$$Q_S = 4 + 2(1.5) = 7$$

$$Q^* = 7$$

If we look at the graph, we can see visually that this is the point where the supply and demand intersect



If $Q_D \neq Q_S$, then the market is not in equilibrium. There are several ways this can occur. Two main reasons are, one, if the market is influenced by government policy (e.g., sales taxes, subsidies, or price controls). Two, if the market has an externality, An externality is when a company is operating on the wrong supply or demand curve because they are not considering the benefits or harm to society as the equilibrium price and quantity are determined. You will discuss government policy and externalities in greater detail in other courses, but here we will look at the math and intuition behind government policy.

5.4 Government Intervention

There are four main ways in which the government can intervene in a market:

1. A price floor is where the government believes that the market price, P^* , is too low. A price floor is a type of price control that sets the minimum price a company can charge. The best example of a price floor in practice is the minimum wage.
2. A price ceiling is where the government believes that the market price, P^* , is too high. A price ceiling is a type of price control that sets the maximum price that a company can charge.
3. Sales taxes can either be a tax that is levied on the total amount of a good or service that is consumed *or* as a percentage amount. Sales taxes can be used to raise revenue for the government or to reduce the quantity of a good or service that is consumed. This could be because the good is harmful to you (e.g., cigarettes). This could be because

the production of the good is harmful to society and is used to correct an externality (e.g., reduce pollution).

4. A subsidy is a direct payment or tax credit given to a company by the government.

5.4.1 Price Controls

Figure 5.1: Graphical Example of a Price Ceiling

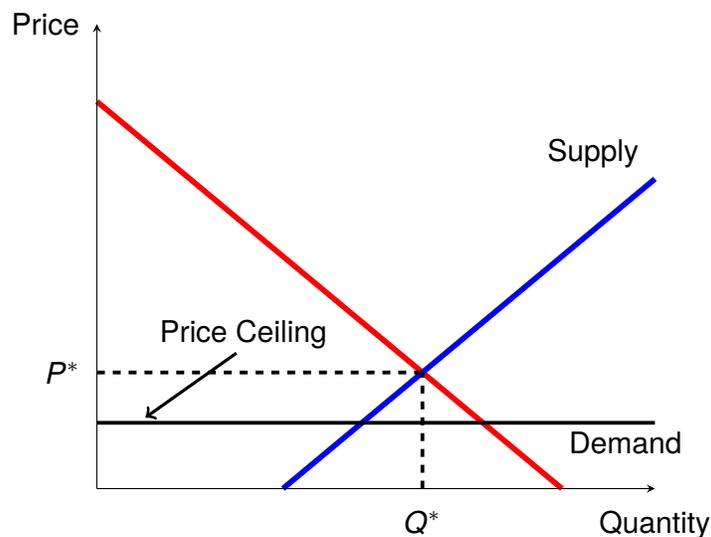
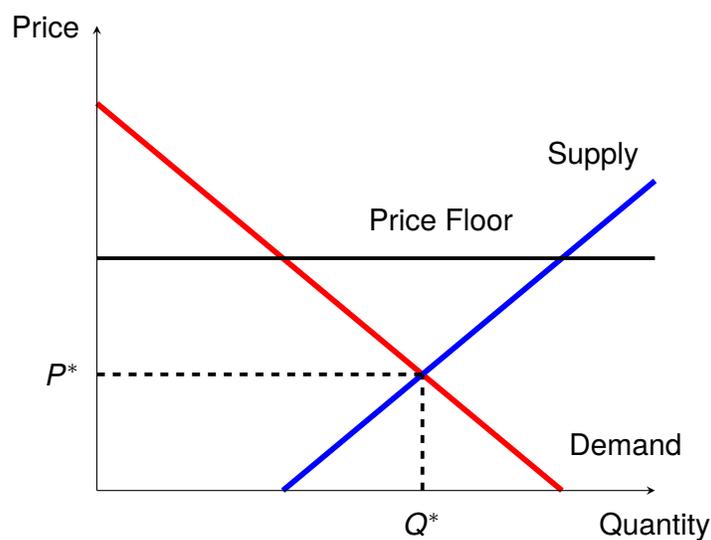


Figure 5.2: Graphical Example of a Price Floor



A price control is when the government passes a law that limits how high or how low a price can go or how fast and large a price can increase over a given time interval. A price ceiling and a price floor are examples of price controls. You will explore these in greater depth in other classes, but the mechanics of how a price control affects market outcomes can be seen in the following example of a price ceiling.

Example

Suppose the market for renting an apartment in Lexington is defined by the following:

$$Q_D = 5000 - P$$

$$Q_S = P$$

Find the equilibrium price and quantity:

$$P = 5000 - P$$

$$2P = 5000$$

$$P^* = 2,500$$

$$Q^* = 2,500$$

Suppose the government thinks that \$2,500 is too high for the price of rent and institutes rent control (or a price ceiling) stating the maximum price of rent is \$2,000. How does this change Q_S and Q_D ?

To find out, plug the value of the price ceiling into Q_D and Q_S and compare:

$$Q_D = 5000 - 2000 = 3000$$

$$Q_S = 2000$$

When the market is in equilibrium, $Q_D = Q_S = 2500$. With the price ceiling, $Q_D > Q_S$.

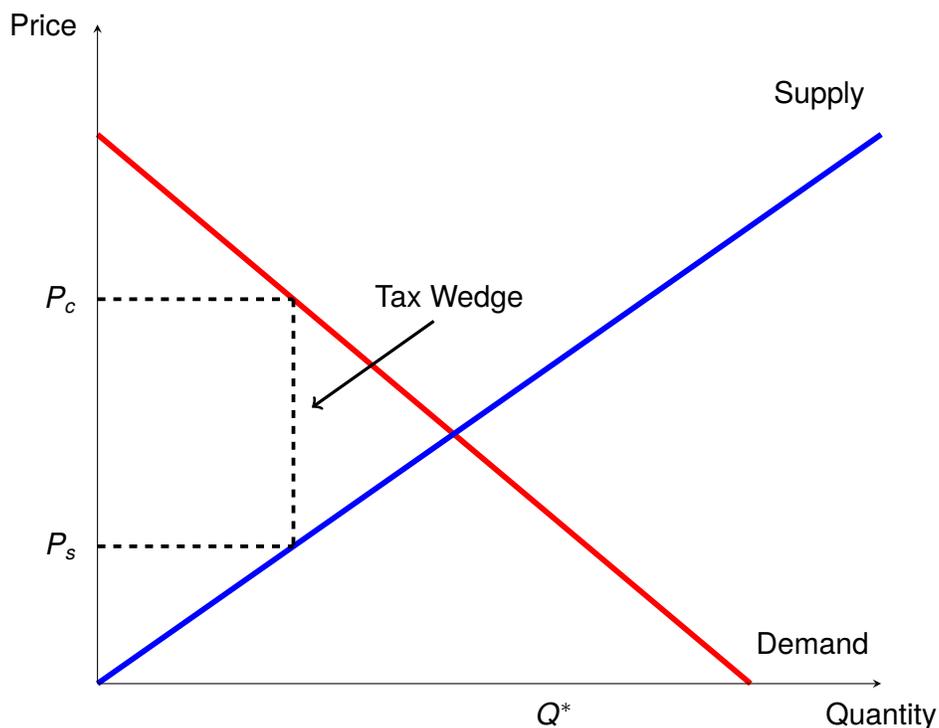
Because the price is lower, more people will want to rent apartments and landlords will be less likely to rent their property. To see the gap in Q_S and Q_D : $Q_S - Q_D = -1000$. This means that there is a shortage of 1,000 apartments.

This was an example of a price ceiling. If there is a price floor, it will have the opposite effect: $Q_D < Q_S$ and $Q_S - Q_D$ will be positive. In this instance, there will be a surplus.

5.4.2 Taxation

We have all gone to the store to find that the final price we pay for something is not the price we saw on the shelf. This is due to a sales tax. When the government institutes a sales tax, it puts a “wedge” between supply and demand. This “wedge” is equal to the value of the tax.

Figure 5.3: Graphical Example of a Sales Tax



Based on supply and demand for each market, the tax wedge determines the price consumers pay at the store, P_c . Consumers pay P_c to the store, then the store pays the tax, t , to the government and is left with P_s . So when a sales tax is levied, the price you (consumers) pay is different from the price companies receive. In practice, sales taxes can be applied in two ways:

1. A quantity tax is a tax levied on the amount of a good consumed. We can relate P_c to P_s using a simple equation: $P_c = P_s + t$
2. A value tax is a tax levied on the expenditure of a good as a percent of the price. The tax, t , is expressed as a percent (τ) of P_s : $t = \tau P_s$. So, $P_c = P_s + \tau P_s$, or $P_c = (1 + \tau)P_s$.

Example

Suppose the market for ice cream is defined by:

$$Q_D = 10 - 2P$$

$$Q_S = 1 + \frac{1}{2}P$$

First, let's find the equilibrium price, P^*

$$10 - 2P = 1 + \frac{1}{2}P$$

$$9 = 2.5P$$

$$P^* = 3.60$$

The state of Virginia imposes a 3% (τ) tax on food. Find P_c and P_s .

Step 1: Set up the tax equation relating P_c to P_s

$$P_c = (1 + .03)P_s$$

Step 2: We want to solve this system of equations where there is only one price. We use P_c in the demand equation and P_s in the supply equation (*Note:* In equilibrium, $P_c = P_s$).

Substitute the tax equation into the demand equation to get one equation in terms of P_s :

$$10 - 2[(1 + .03)P_s] = 1 + \frac{1}{2}P_s$$

$$10 - 2.06P_s = 1 + \frac{1}{2}P_s$$

$$9 = 2.56P_s$$

$$P_s = 3.5156$$

To get P_c , plug P_s into the tax equation:

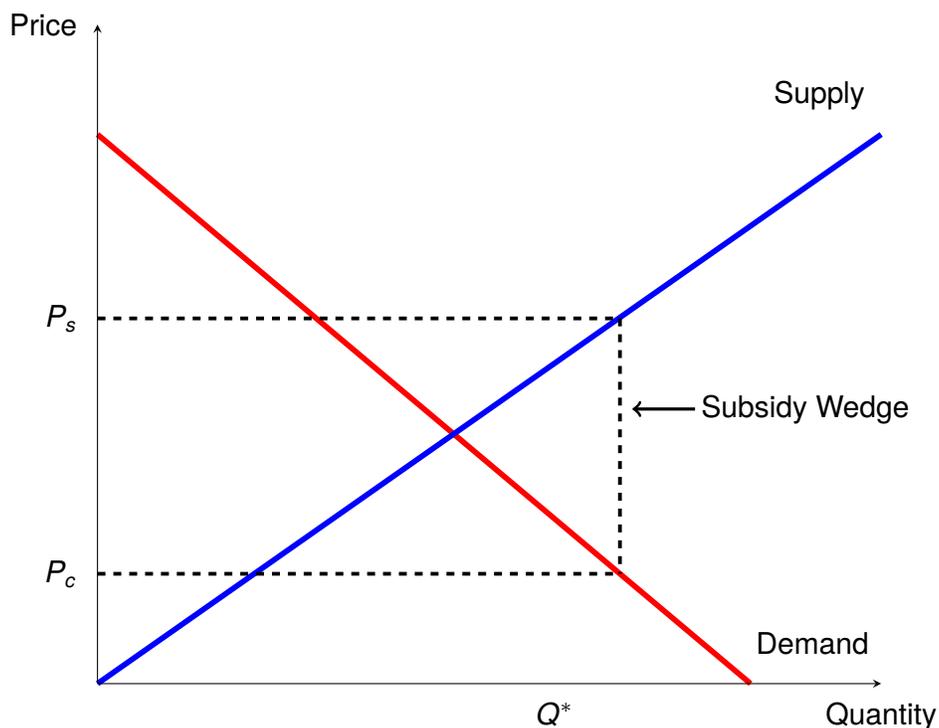
$$P_c = (1.03)(3.5156) = 3.62109$$

You will notice that $P_c > P_s$ and that P^* falls in between the two, just like we see graphically.

5.4.3 Government Subsidies

A subsidy is a payment by the government to a company to incentivize production. A subsidy works similarly to a tax, except that with a subsidy, the company receives a higher price than the consumer pays (the subsidy reduces the amount the consumer pays by s dollars per unit).

Figure 5.4: Graphical Example of a Subsidy



Example

Suppose we have the same market for ice cream, but the government decides that it wants to encourage more people to eat ice cream by lowering the price by providing a \$1 subsidy per unit. Find P_c and P_s .

Step 1: Set up the subsidy equation relating P_c to P_s

$$P_c = P_s - 1$$

Step 2: We want to solve this system of equations where there is only one price. We use P_c in the demand equation and P_s in the supply equation (*Note:* In equilibrium, $P_c = P_s$).

Substitute the subsidy equation into the demand equation:

$$10 - 2[P_s - 1] = 1 + \frac{1}{2}P_s$$

$$11 = 2.5P_s$$

$$P_s = 4.40$$

To get P_c , plug P_s into the subsidy equation:

$$P_c = 4.40 - 1 = 3.40$$

You will notice that $P_c < P_s$ and that P^* falls in between the two, just like we see graphically.

5.5 Nonlinear Markets

In practice, supply and demand are not always linear relationships. We can easily solve systems of equations that are linear and quadratic in nature. Higher-powered relationships exist but are more easily solved using computer software. To solve systems of equations that are quadratic in nature, we *may* need to use the quadratic formula.

Example

$$Q_D = 5 - P^2$$

$$Q_S = P^2 + 2P$$

Solve for P^* :

$$5 - P^2 = P^2 + 2P$$

$$2P^2 + 2P - 5 = 0$$

Using the quadratic formula:

$$\frac{-2 \pm \sqrt{2^2 - 4(2)(-5)}}{2(2)}$$

We get that $P = 1.16$ and $P = -2.16$. It is not possible to have a negative price, so $P^* = 1.16$

5.6 Elasticity of Demand

We are often interested in how consumers will respond if the price of a good or service changes. A manager may want to know how an increase in price or a sale will affect the quantity sold. The government may be interested in how a sales tax will affect the quantity that is sold or how it may impact tax revenue.

Suppose $Q_D = 10 - 2P$. If $P = 4$, then $Q_D = 2$. If P decreases to 3, then $Q_D = 4$. This can be seen graphically in Figure 5.5. Suppose that instead we have the following demand curve:

$Q_D = 12 - \frac{1}{3}P$. If $P = 4$, then $Q_D = 2.67$. If P decreases to 3, then $Q_D = 3$. This can be seen graphically in Figure 5.6.

In Figure 5.6, Q_D is less sensitive to a price change than in Figure 5.5. Figure 5.5, a \$1 decrease in price led to a 2 unit increase in Q_D . Figure 5.6, a \$1 decrease in price led to a 0.33 unit increase in Q_D . We can formally measure this for any demand equation using elasticity of demand. The elasticity of demand is a measure of how responsive consumers are to a change in price.

Demand curves that are more “flat” are called elastic. This means that consumers are sensitive to changes in price. These are often luxury goods or goods that have many alternatives (substitutes). Demand curves that are more “steep” are called inelastic. This means that consumers are not very sensitive to price changes. These are often necessities, addictive products,

Figure 5.5: Change in Price leading to a Large Change in Quantity

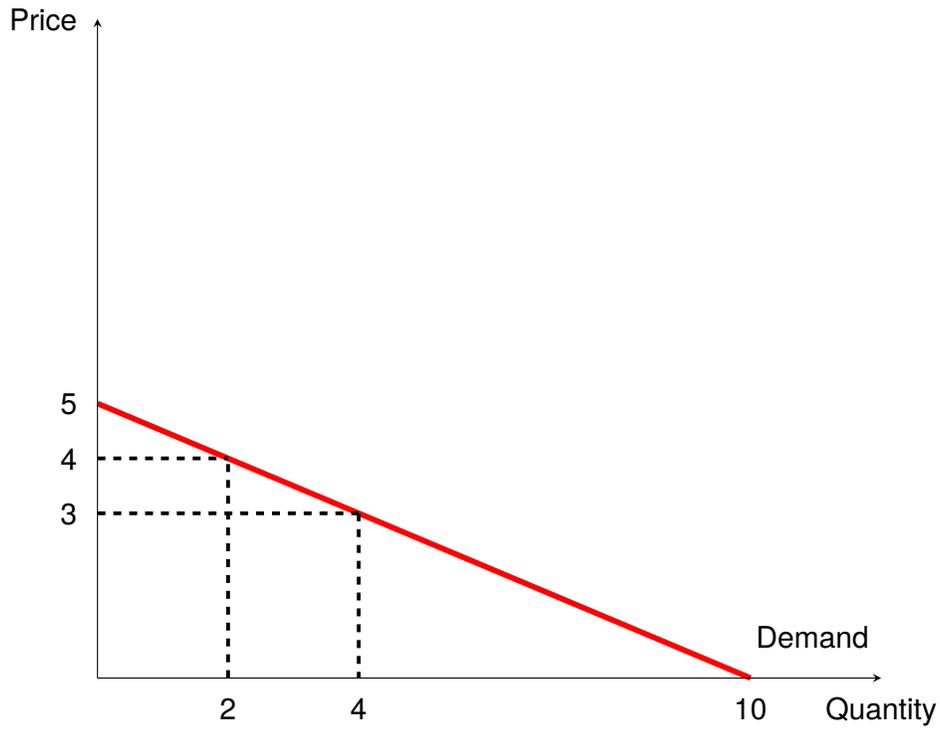
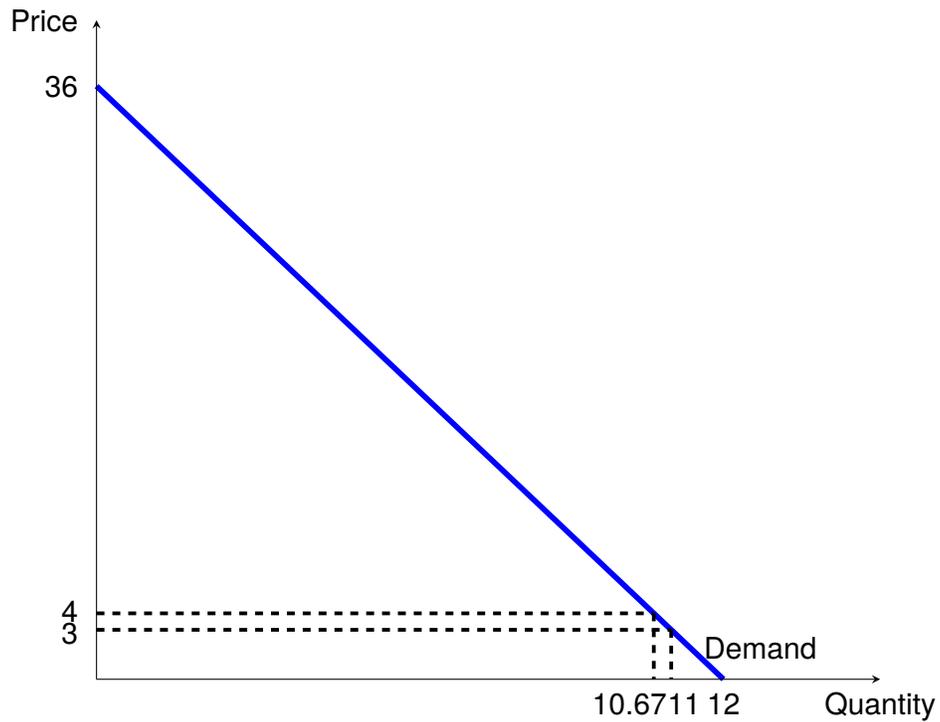


Figure 5.6: Change in Price leading to a Small Change in Quantity



or goods with no alternative. If Q_D and P change by the same amount, these demand curves are called unit elastic.

The elasticity of demand can be calculated by:

$$E_D = \frac{dQ}{dP} \times \frac{P}{Q} \quad (5.6)$$

The interpretation of E_D is the percent change in quantity demanded resulting from a 1% increase in the price. The result you will get will always be a negative number because the demand curve is downward sloping.

If $|E_D| > 1$, the good or service is elastic. When the demand is elastic, for every 1% increase in P , Q_D decreases by more than 1%. If $|E_D| < 1$, the good or service is inelastic. When the demand is inelastic, for every 1% increase in P , Q_D decreases by less than 1%. If $|E_D| = 1$, the good or service is unit elastic. When the demand is unit elastic, for every 1% increase in P , Q_D decreases exactly by 1%.

Example

If the demand for paper cups is given by: $Q_D = 10 - 2P$, what is the elasticity of demand if the price of paper cups is \$2?

Step 1: Find the derivative:

$$\frac{dQ}{dP} = -2$$

Step 2: We know that $P = 2$. We can plug this into the demand equation to find Q_D :

$$Q_D = 10 - 2(2) = 6$$

Step 3: Calculate the elasticity:

$$E_D = -2 \times \frac{2}{6} = -\frac{2}{3} = -0.67$$

The interpretation of this would be that for every 1% increase in P , Q_D decreases by 0.67%. Because $|E_D| < 1$ when $P = 2$, paper cups are inelastic.

UTILITY MAXIMIZATION

Economics is the study of how people make choices when faced with constraints. People face three constraints: money, time, and information. The constraints of time and money will be explored in more detail in principles/intermediate microeconomics and labor economics courses, and information constraints will be explored in game theory and behavioral economics courses. In this book, we will discuss the mathematical techniques used to model how people make choices when faced with constraints, with a focus on financial constraints.

6.1 Utility

Every person has their own unique set of preferences. These preferences determine how we choose to spend our money. In general, the more we consume the things we like, the more “happy” we are. In economics, utility is a measure of the level of “happiness” we get from consuming goods and services. People generally make choices that make them as happy as possible; therefore, we try to obtain the highest level of utility possible.

Since each person has different preferences, the level of “happiness” one person gets consuming something may be very different than the level of satisfaction someone else gets. We can mathematically model the relationship between our individual preferences and the level of utility with a utility function. In its most simplistic form, a utility function, $U = f(x, y)$ where x is the quantity of one good consumed and y represents the quantity of a second good consumed. For a utility function to represent a person’s preferences, there are certain properties that must be true:

1. **Preferences are complete.** It cannot be true that item A is preferred to item B *and* that you are indifferent between item A and item B, only one can be true.
2. **Preferences are transitive.** If A is preferred to B and B is preferred to C, then A must be preferred to C. This means that consumers are consistent in their preferences.
3. **Preferences are continuous.** If the function is not continuous, we cannot take a derivative, and we will have a demand curve that is not well-behaved.
4. **Preferences are monotonically increasing.** This means that we always prefer more utility to less utility.

There are several types of utility functions that satisfy the above properties and are commonly seen in economic models.

Commonly Used Utility Functions

- **Cobb-Douglass**

$$U = X^\alpha Y^{1-\alpha} \quad \text{Example: } U = X^{1/2} Y^{1/2}$$

- **Constant Elasticity of Substitution (CES)**

$$U = (\alpha X^\rho + (1 - \alpha) Y^\rho)^{\frac{1}{\rho}} \quad \text{Example: } U = \left(\frac{1}{3}X^2 + \frac{2}{3}Y^2\right)^{\frac{1}{2}}$$

- **Constant Relative Risk Aversion (CRRA)**

$$U = \begin{cases} \frac{X_i^{1-\theta}}{1-\theta} & \text{if } \theta > 0, \theta \neq 1 \\ \ln(X_i) & \text{if } \theta = 1 \end{cases}$$

6.1.1 Indifference Curves

Preferences can be shown graphically using indifference curves. An indifference curve is a graph showing the combination of two goods that give the consumer equal satisfaction and utility. A consumer would be indifferent to each point on an indifference curve because all points give them the same utility. The equation that is used to graph an indifference curve can be obtained by rearranging the utility function. If $U = X^{1/2} Y^{1/2}$, then the indifference curve is $Y = \frac{U^2}{X}$, where U is the “fixed” level of utility.

Example

Suppose a person’s preferences are defined by $U = X^{1/2} Y^{1/2}$. In order to graph this function, we need to rearrange it:

$$Y^{1/2} = \frac{U}{X^{1/2}}$$

$$Y = \frac{U^2}{X}$$

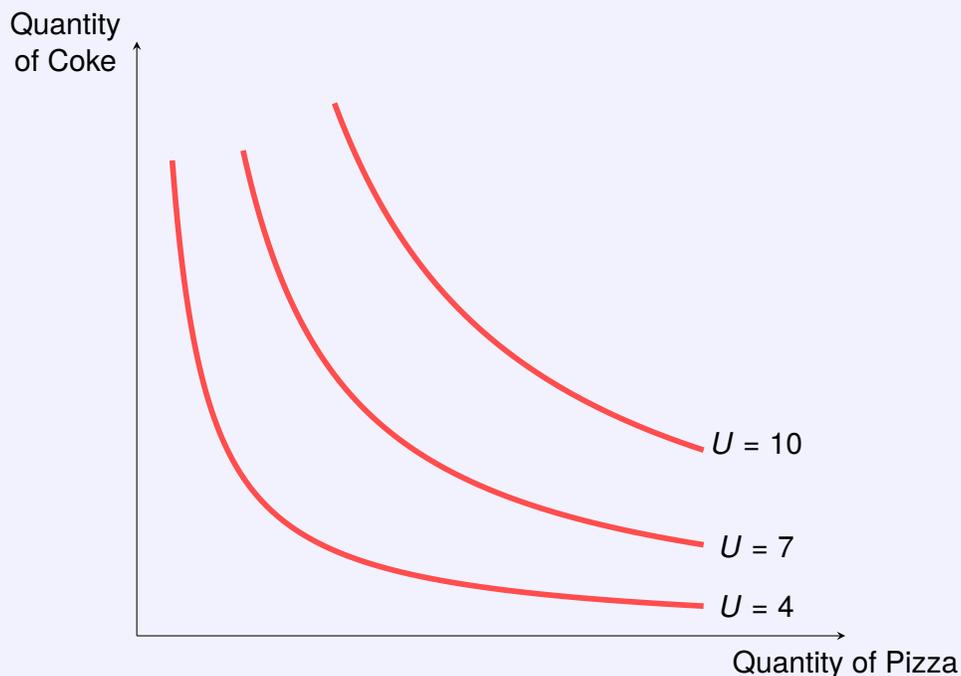
U represents the level of utility. Therefore, this will be a number. For each level of utility, there will be a different indifference curve:

$$\text{If } U = 4, \text{ then } Y = \frac{16}{X}$$

$$\text{If } U = 7, \text{ then } Y = \frac{49}{X}$$

$$\text{If } U = 10, \text{ then } Y = \frac{100}{X}$$

We can graph each of these indifference curves:



If all of the properties of the utility function are satisfied, then the indifference curves will never intersect and will be parallel to each other. Since each indifference curve represents a fixed level of utility and preferences are monotonically increasing, we would prefer to be on a higher indifference curve.

6.1.2 Marginal Rate of Substitution

The marginal rate of substitution (MRS) is the amount of a good that a consumer is willing to give up of one good to consume another good, while keeping utility constant. The MRS can be calculated by:

$$MRS = \frac{\text{Marginal Utility of Good X}}{\text{Marginal Utility of Good Y}} \quad (6.1)$$

Marginal utility can be calculated by:

$$\text{Marginal Utility}_x = \frac{\partial U}{\partial X} \quad \text{Marginal Utility}_y = \frac{\partial U}{\partial Y} \quad (6.2)$$

If we combine equations 6.1 and 6.2, we get:

$$MRS = \frac{\left(\frac{\partial U}{\partial X}\right)}{\left(\frac{\partial U}{\partial Y}\right)} \quad (6.3)$$

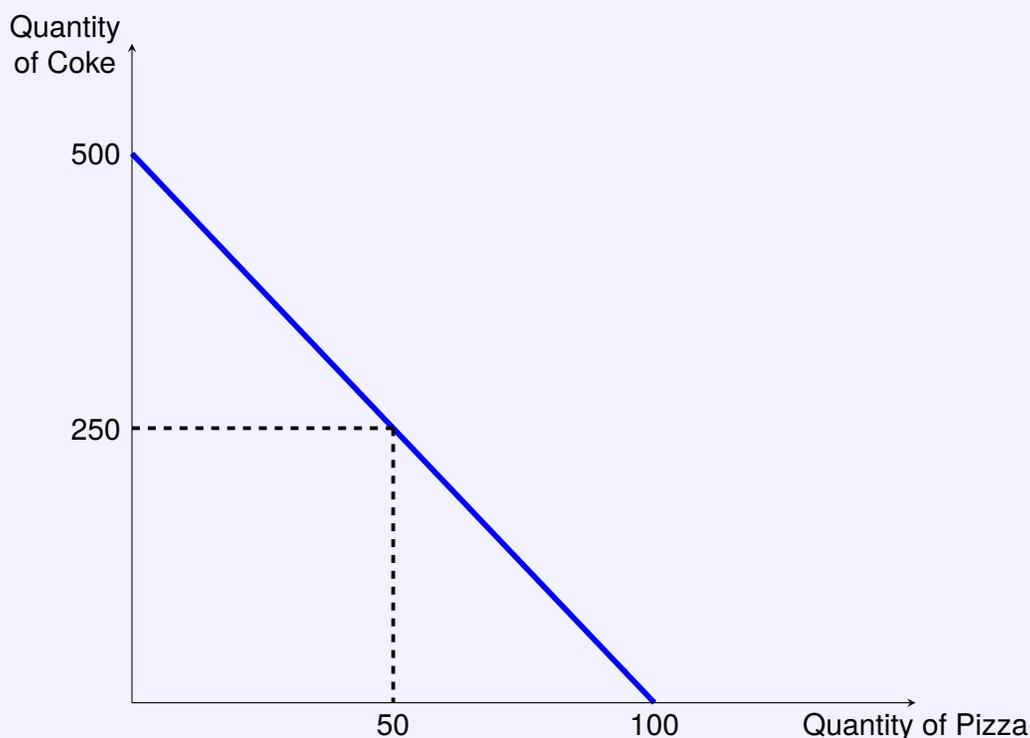
6.2 The Budget Constraint

The budget constraint shows the combinations of goods and services that you can buy with your available income. The easiest way to see the construction of a budget constraint is through an example.

Example

Suppose that you have an income of \$1,000 and you are at the store and want to buy some frozen pizza and bottles of Coke for a party you are going to have. The price of pizza is \$10 and the price of Coke is \$2.

If you spend all your money on pizza, then you could buy $\frac{1000}{10} = 100$ pizzas. If you spend all your money on Coke, you could purchase $\frac{1000}{2} = 500$ bottles of Coke. Alternatively, you could split your money and buy some pizza and Coke. We can see the budget constraint graphically:



We can also express this as a mathematical relationship:

$$(\text{Price of Coke} \times \text{Quantity of Coke}) + (\text{Price of Pizza} \times \text{Quantity of Pizza}) = \text{Income}$$

In the budget constraint, the left-hand side must equal the right-hand side. This equation will give you every possible combination of Coke and pizza you could buy for the given prices and income. So from our example, the budget constraint would be:

$$(\$2 \times \text{Quantity of Coke}) + (\$10 \times \text{Quantity of Pizza}) = \$1000$$

More generally, we can write the budget constraint as follows:

$$I = P_x X + P_y Y$$

Where,

P_x = Price of good x

P_y = Price of good y

X = Quantity of good x

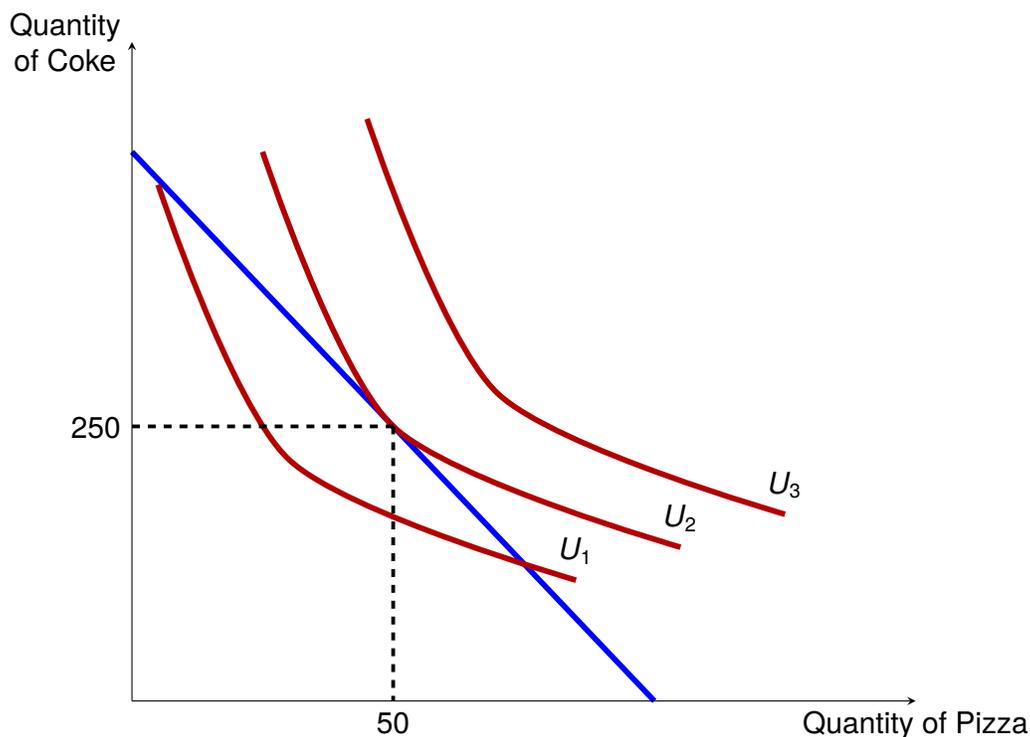
Y = Quantity of good y

I = Income

6.3 Utility Maximization

Because preferences are monotonically increasing, this means that we always want to have the highest possible utility. However, how high our utility can be is limited by our budget constraint. Utility maximization is the process of finding the optimal quantities for goods that give us the highest utility possible given our individual preferences and our budget constraint. We can see this graphically in Figure 6.1 by combining the budget constraint and indifference curves.

Figure 6.1: Graphical Example of Utility Maximization



We can also solve the following optimization problem for the quantities of goods x and Y mathematically:

$$\begin{aligned} \max \quad & U(X, Y) \\ \text{s.t.} \quad & P_1X + P_2Y = I \end{aligned} \tag{6.4}$$

Below are the steps required to solve this optimization problem.¹

- **Step 1:** Find the Marginal Utility for each good

- $MU_x = \frac{dU}{dX}$

- $MU_y = \frac{dU}{dY}$

- **Step 2:** Find the Marginal Rate of Substitution

- $MRS = \frac{MU_1}{MU_2} = \frac{dU}{dX} \bigg/ \frac{dU}{dY}$

- **Step 3:** Apply the Utility Maximization Rule

- $MRS = \frac{P_x}{P_y}$

- **Step 4:** Rearrange the utility maximization rule and insert it into the budget constraint to solve for X^* and Y^* . X^* and Y^* are the quantities of good X and good Y that maximize a persons utility.

- **Step 5:** Plug X^* and Y^* into the utility function to solve for the level of utility achieved by consuming X^* and Y^* .

Example

Becca has an income of \$100. She is trying to decide between buying pizza and fries. Pizza costs \$10 and fries cost \$5. Becca's preferences are defined by:

$$U = \ln(X) + \ln(Y)$$

Where X is the quantity of pizza and Y is the quantity of fries. What combination of pizza and fries maximizes her utility?

- Set up the budget constraint: $10X + 5Y = 100$
- Find MU_1 :

- $\frac{dU}{dX} = \frac{1}{X}$

¹ A more advanced method for solving utility maximization problems is to use the Method of Lagrange. An example of this can be found in the Appendix of this chapter.

- Find MU_2 :

- $\frac{dU}{dY} = \frac{1}{Y}$

- Find MRS :

- $MRS = \frac{1}{X} / \frac{1}{Y} = \frac{Y}{X}$

- Apply Utility Maximization Rule:

- $\frac{Y}{X} = \frac{10}{5}$

- Rearrange the equation in terms of either X or Y :

- $Y = \frac{10}{5}X$

- $X = \frac{5}{10}Y$

- Substitute into the budget constraint:

- $10 \left(\frac{5}{10}Y \right) + 5Y = 100$

- Solve for Y^* :

- $10Y = 100$

- $Y^* = \frac{100}{10} = 10$

- Plug Y^* into rearranged Utility Maximization Rule and solve for X^* :

- $10 = \frac{10}{5}X$

- $X^* = 5$

- Plug X^* and Y^* into U , solve for U^*

- $U^* = \ln(5) + \ln(10) = 3.91$

Utility maximization models are great for evaluating public policy. Suppose that we want to see the impact of a change in government policy on people's choices. We can construct a utility maximization model that reflects the current world and determine U^* . We can then make changes to the budget constraint based on the proposed policy change and evaluate if the policy change increases or decreases U^* and determine if a policy helps or hurts people with different

types of preferences (utility functions). These are types of concepts that you will explore in more detail in other courses.

6.A Appendix: Constrained Optimization

In Chapter 4 we introduced the Method of Lagrange Multipliers. We can use that method to solve utility maximization problems. Let's revisit the example used in this chapter:

$$\begin{aligned} \max \quad & \ln(X) + \ln(Y) \\ \text{s.t.} \quad & 10X + 5Y = 100 \end{aligned} \tag{6.5}$$

We can set up the Lagrangian:

$$\mathcal{L} = \ln(X) + \ln(Y) + \lambda[10 - (10X + 5Y)] \tag{6.6}$$

Next, we need to find $\frac{\partial \mathcal{L}}{\partial X}$, $\frac{\partial \mathcal{L}}{\partial Y}$, $\frac{\partial \mathcal{L}}{\partial \lambda}$ and set them equal to zero.

$$\frac{\partial \mathcal{L}}{\partial X} = \frac{1}{X} - 10\lambda = 0 \tag{6.7}$$

$$\frac{\partial \mathcal{L}}{\partial Y} = \frac{1}{Y} - 5\lambda = 0 \tag{6.8}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 100 - 10X - 5Y = 0 \tag{6.9}$$

We can use the elimination method on Equations 6.7 and 6.8:

$$\lambda = \frac{1}{10X} \tag{6.10}$$

$$\lambda = \frac{1}{5Y} \tag{6.11}$$

Therefore,

$$\frac{1}{5Y} = \frac{1}{10X} \tag{6.12}$$

Note, we can find the MRS from Equation 6.12 by rearranging to get $\frac{Y}{X} = \frac{10}{5}$. We can then continue to rearrange to solve for X .

$$10X = 5Y \tag{6.13}$$

$$X = \frac{1}{2}Y \tag{6.14}$$

We can substitute Equation 6.14 into Equation 6.9 and solve for y^* :

$$100 - 10 \left(\frac{1}{2} Y \right) - 5Y = 0 \quad (6.15)$$

$$100 = 10 \left(\frac{1}{2} Y \right) + 5Y \quad (6.16)$$

$$\frac{100}{10} = Y \quad (6.17)$$

$$Y^* = 10 \quad (6.18)$$

We can substitute Y^* into Equation 6.14 and solve for X^* :

$$X = \left(\frac{1}{2} \right) 10 \quad (6.19)$$

$$X^* = 5 \quad (6.20)$$

We can substitute X^* and Y^* into the utility function to solve for U^* :

$$U^* = \ln(5) + \ln(10) = 3.91 \quad (6.21)$$

Using the Method of Lagrange Multipliers, we get the same answer as we did in the example in the chapter. However, when using constrained optimization, we can also solve for λ^* :

$$\lambda = \frac{1}{10(5)} \quad (6.22)$$

$$\lambda^* = \frac{1}{50} = 0.02 \quad (6.23)$$

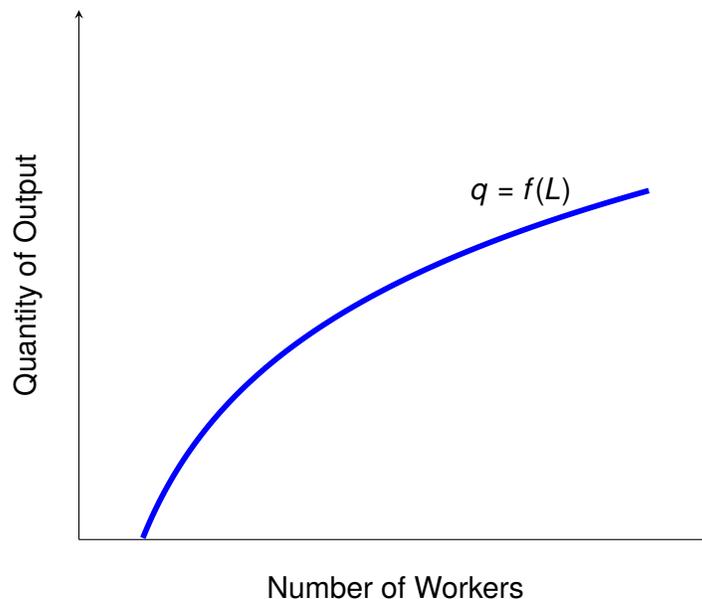
In economics, λ^* is called the shadow price. The shadow price is the amount of utility (U^*) that could be gained by increasing income, I , by one dollar.

PROFIT MAXIMIZATION

7.1 The Production Function

The production function shows the relationship between the inputs a company uses in production and the quantity of output it produces. An input is anything that a company uses to produce goods and services. Typical inputs include capital and labor. Output is the product a company sells. Figure 7.1 shows a graphical example of a production function.

Figure 7.1: A Production Function



The shape of the production function is concave, or it does not increase until infinity. Each additional input that is added has a smaller impact on output than the unit before it. This is called diminishing marginal returns. This means that a company cannot hire an infinite number of workers and produce an infinite amount of output. Because of this, companies need to determine the appropriate number of workers to hire and capital to use.

7.2 Marginal Products

A marginal product tells us what additional, or incremental, output each additional worker adds to the total output (q). We can calculate the marginal product for each input, labor (L) and capital (K).

7.2.1 Marginal Product of Labor

The marginal product of labor (MP_L) is the additional output that is added if the company were to hire one additional worker.

$$MP_L = \frac{\text{Change in Quantity}}{\text{Change in the Number of Workers}} \quad (7.1)$$

It is first helpful to visualize the MP_L through an example. Suppose that a pizza company is trying to evaluate how productive each of its employees is and has the following information on its output:

| Number of Workers | Quantity of Pizza | Marginal Product of Labor |
|-------------------|-------------------|---------------------------|
| 0 | 0 | – |
| 1 | 10 | $10 - 0 = 10$ |
| 2 | 15 | $15 - 10 = 5$ |
| 3 | 17 | $17 - 15 = 2$ |

We can formally express MP_L as follows:

$$\text{If } q = f(K, L) \text{ then } MP_L = \frac{dq}{dL} \quad (7.2)$$

Example

Lay's produces potato chips using workers and heavy machinery. It's output per day is defined by the following production function:

$$q = K^{1/2}L^{1/2}$$

Find the marginal product of labor.

$$q = K^{1/2}L^{1/2}$$

$$MP_L = \frac{dq}{dL} = \frac{1}{2}L^{-1/2}K^{1/2}$$

$$MP_L = \frac{1}{2} \left(\frac{K}{L} \right)^{1/2}$$

7.2.2 Marginal Product of Capital

Just like with labor, a company needs to know how many units of capital, K it should use. The marginal product of capital (MP_K) is the additional output that is added if the company were to

use one additional unit of capital. The MP_K can be calculated by:

$$\text{If } q = f(K, L) \text{ then } MP_L = \frac{dq}{dK} \quad (7.3)$$

7.3 Profit

Each product a company produces will have its own unique production function. The company will then sell the products it produces for a certain price, which will generate revenue for the company. Revenue is defined as *price* \times *quantity*, or pq . The company must also spend money on the inputs used to produce these products (costs). Firms have two costs: labor and capital. The cost of labor is the *wage* \times *number of workers*, or wL . The cost of capital is the *rental price* \times *units of capital*, or rK . Therefore, the cost of a firm is defined as $wL + rK$. Therefore, we can define the profit, π , of a company as the sum of its revenue minus the sum of its costs.

$$\pi = pq - (wL + rK) \quad (7.4)$$

Where,

p is the price of the product

q is the output of the firm

L is the units of labor (number of workers)

w is the wage rate

K is the units of capital

r is the rental rate of capital

The quantity of goods a company produces can be obtained from the company's production function: $q = f(K, L)$. Therefore, we can write revenue as $p \cdot f(K, L)$. Therefore, we can rewrite Equation 7.4 as:

$$\pi = pf(K, L) - wL - rK \quad (7.5)$$

7.4 Profit Maximization

The goal of any company is to generate the maximum profit possible. To do that, a firm must find the optimal level of inputs, L^* and K^* . To find the optimal level of inputs, we need to maximize the profit function with respect to each input by taking partial derivatives. Recall from earlier that we can set the derivative equal to zero to find maximum and minimum points. The profit maximization problem is set up as follows:

$$\max \pi = pf(K, L) - wL - rK \quad (7.6)$$

Which has the following first derivatives:

$$\frac{\partial \pi}{\partial L} = p \frac{\partial q}{\partial L} - w = 0 \quad (7.7)$$

$$\frac{\partial \pi}{\partial K} = p \frac{\partial q}{\partial K} - r = 0 \quad (7.8)$$

Recall that $\frac{\partial q}{\partial L}$ is the Marginal Product of Labor and $\frac{\partial q}{\partial K}$ is the Marginal Product of Capital. To find the profit-maximizing quantities of K and L , we need to find the Marginal Rate of Technical Substitution (MRTS). The MRTS is the amount by which capital and labor can be substituted for each other so that the output does not change.

$$MRTS = \frac{MP_L}{MP_K} = \frac{\partial q / \partial L}{\partial q / \partial K} \quad (7.9)$$

We then set the MRTS equal to the ratio of input prices:

$$\frac{MP_L}{MP_K} = \frac{w}{r} \quad (7.10)$$

From this we rearrange this equation for K or L . To solve the profit-maximizing quantity of labor, L^* , substitute K from the above step into $\frac{\partial q}{\partial L}$ and solve. To solve for K^* , substitute L into $\frac{\partial q}{\partial K}$ and solve. Once you have K^* or L^* , you can find the other by substituting it into $MRTS = \frac{w}{r}$ or doing both the previous steps. To ensure that K^* or L^* are profit maximizing quantities, we will check to make sure the sign of the second derivatives is negative

Example

Suppose that you have the following production function for a firm: $q = 10K^{1/3}L^{1/3}$. This company pays its workers \$15 per hour and rents its capital at \$15 per hour. It sells its product for \$20. Find the values of K and L that maximize the company's profit.

$$\begin{aligned} \max \pi &= pq - wL - rK \\ \text{s.t. } q &= f(K, L) \\ \max \pi &= 200K^{1/3}L^{1/3} - 15L - 15K \end{aligned}$$

Find the first derivatives:

$$\begin{aligned} \frac{\partial \pi}{\partial L} &= \frac{200}{3}K^{1/3}L^{-2/3} - 15 = 0 \\ \frac{\partial \pi}{\partial K} &= \frac{200}{3}K^{-2/3}L^{1/3} - 15 = 0 \end{aligned}$$

Find the Marginal Rate of Technical Substitution:

$$MRTS = \frac{MP_L}{MP_K} = \frac{\frac{1}{3}K^{1/3}L^{-2/3}}{\frac{1}{3}K^{-2/3}L^{1/3}} = \frac{K}{L}$$

Set MRTS equal to the ratio of the factor prices, $\frac{w}{r}$:

$$\frac{K}{L} = \frac{15}{15}$$

$$K = L$$

Substitute this into the first derivatives, solve for L^* :

$$\frac{200}{3}(L)^{1/3}L^{-2/3} = 15$$

$$\frac{200}{45} = L^{1/3}$$

$$L^* = 87.79$$

Plug L^* into $K = L$ to solve for K^* :

$$K^* = 87.79$$

Plug K^* and L^* into the production function to find q^* :

$$q^* = 10(87.79)^{1/3}(87.79)^{1/3} = 197.53$$

Plug q^* , K^* and L^* into the profit function to find π^*

$$\pi = 50(197.53) - 15(87.79) - 15(87.79) = 1316.87$$

Check the second derivatives to ensure both are a maxima:

$$\frac{\partial^2 \pi}{\partial L^2} = -\frac{400}{9}(87.79)^{1/3}(87.79)^{-5/3} = -0.114$$

$$\frac{\partial^2 \pi}{\partial K^2} = -\frac{400}{9}(87.79)^{-5/3}(87.79)^{1/3} = -0.114$$

Since both second derivatives are negative, K^* and L^* are profit maximizing values.

Part III

Probability and Statistics

PROBABILITY AND COUNTING RULES

Probability measures how likely something will happen. Every event that occurs has a probability assigned to it. The more likely an event is to occur, the higher the probability. The probability that an event can occur can be calculated as follows:

$$P(\text{event}) = \frac{\text{Number of Favorable Outcomes}}{\text{Number of Total Possible Outcomes}} \quad (8.1)$$

Where, $0 \leq P(\text{event}) \leq 1$. Additionally, the probability that an event does not occur can be calculated by:

$$\text{Prob}(\text{event does not occur}) = 1 - \text{Prob}(\text{event}) \quad (8.2)$$

8.1 Single Events

When only one action occurs, this is called a single event. Examples of single events include flipping a coin, rolling a die, or pulling one marble out of a bag.

Example 1

If you flip a coin, what is the probability of getting heads?

Number of favorable outcomes: 1

Number of total possible outcomes: 2

$$P(\text{heads}) = \frac{1}{2} = 0.50$$

Example 2

If you roll a single die, what is the probability of getting a 4?

Number of favorable outcomes: 1

Number of total possible outcomes: 6

$$P(4) = \frac{1}{6} = 0.167$$

Example 3

There are 5 marbles in a bag. 3 are blue, 1 is red, and 1 is green. What is the probability of picking a blue marble?

Number of favorable outcomes: 3

Number of total possible outcomes: 5

$$P(\text{Blue}) = \frac{3}{5} = 0.60$$

8.2 Multiple Events

If more than one action occurs at the same time, these are called multiple events. Examples of multiple events include rolling two dice, opening a combination lock, and winning the lottery. In order to calculate probability, it is necessary to count the number of favorable and possible outcomes. This section goes over the ways to count outcomes when there are multiple events.

8.2.1 The Counting Rule

The Counting Rule is a way to figure out the number of outcomes in a probability problem. To use the counting rule, multiply the outcomes of the events together to calculate the total number of outcomes.

Example 1

You take a survey with five “yes” or “no” answers. How many different ways could you complete the survey?

There are 5 questions and each question has 2 choices.

Number of total choices: $2 \times 2 \times 2 \times 2 \times 2 = 32$

Example 2

A company puts a code on each different product they sell. The code is made up of 3 numbers and 2 letters. How many different codes are possible?

There are 10 possible numbers: 0 – 9

There are 26 possible letters: A – Z

Number of total choices: $10 \times 10 \times 10 \times 26 \times 26 = 676,000$

8.2.2 Combinations and Permutations

When thinking about multiple events, we need to think about how to count the number of possible outcomes and the number of favorable outcomes. Permutations and combinations are two methods employed to count how many outcomes are possible in various situations. In English, we use the word “combination” whenever we are talking about a group of things together. However, with statistics, we can be more precise because sometimes the order in which events

occur matters. For example, a combination on a lock requires a specific order, whereas, to win the lottery, you just need to match the same numbers, they do not need to be in the same order. When the order *does not* matter, we are dealing with a combination. When the order *does* matter, we are dealing with a permutation.

8.2.2.1 Permutations with Repetition

A permutation with repetition is dealing with counting a group where order does matter and events are allowed to repeat themselves. An example of this is a combination lock.

Example

A combination lock has 4 dials that each have the numbers 0-9, so each dial has 10 options. Since each dial is independent of each other, we can get the number of possible combinations by:

$$10 \times 10 \times 10 \times 10 \text{ or } 10^4 = 10,000$$

This means that there are 10,000 total possible outcomes for the lock.

More generally, we can calculate the number of total possible outcomes for a permutation with repetition by the following.

$$n^r \tag{8.3}$$

Where n = is the number of possible outcomes of the first choice and r = is the number of times the first choice is repeated. Anytime we are trying to figure out the number of outcomes where order matters and repetition is allowed, we can use this formula.

8.2.2.2 Permutation without repetition

A permutation without repetition is dealing with the counting of a group where order matters but events are not allowed to repeat themselves.

Example

You have three classes that you need to take, math, science, and history. How many possible ways are there to arrange these classes?

- You start out with three choices: Math, Science, and History
- If you pick science to be your first class, then your are left to select between math and history for your second class
- If you pick history as your second class, then you only have one choice left for your third class: math

Using the counting rule, the number of possible ways to arrange these classes = $3 \times 2 \times 1 = 6$

We can express the number of possible outcomes for a permutation using a math tool called a factorial, which is designated by “!”

$$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

On the TI-83/84, this can be found on the MATH button on the PRB tab.

Sometimes we may have many options to choose from, but we are not going to pick them all. We can count the number of permutations if there is no repetition using the following:

$${}_n P_r = \frac{n!}{(n-r)!} \quad (8.4)$$

Where n = is the number of possible choices and r = is the number of selected choices. **NOTE:** if $n = r$, then the denominator = $0!$ and $0! = 1$.

Example

Suppose that you have 10 classes to choose from and that you only need to take 6. How many ways could you take those 6 classes?

There are 10 classes to choose from (n), you are choosing 6 of them (r)

$${}_{10} P_6 = \frac{10!}{(10-6)!} = 151,200$$

If you were to choose all 10 classes:

$${}_{10} P_{10} = \frac{10!}{(10-10)!} = \frac{10!}{0!} = 10! = 3,628,800$$

If $n = r$, then we have our first scenario, ${}_n P_r = n!$

8.2.2.3 Combinations without Repetition

A combination without repetition deals with counting a group where order does not matter, but events are not allowed to repeat. To calculate the number of possible outcomes for a combination without repetition:

$${}_n C_r = \frac{n!}{r!(n-r)!}$$

where,

n = the number of possible choices

r = the number of the choices selected

The best example of a combination without repetition is a lottery.

Example

Suppose a lottery has 48 ping pong balls numbered 1-48. 6 of them are drawn at random. If the 6 numbers drawn match the 6 numbers you picked, you win, even if the numbers are not in the same order. How many different combinations of ping pong balls are there?

There are 48 ping pong balls to choose from (n), 6 of them are chosen (r).

$${}_{48}C_6 = \frac{48!}{6!(48 - 6)!} = 12,271,512$$

There are 12,271,512 possible choices.

8.3 Probability with Multiple Events

It is helpful to remember that $P(event) = \frac{\text{Number of Favorable Outcomes}}{\text{Number of Total Possible Outcomes}}$. These problems require some thought as to the total number of possible and favorable outcomes. For each problem, we need to determine how to count these outcomes. Let's start with continuing with our lottery example.

Example 1

What is the probability of winning the lottery that has 48 ping-pong balls where 6 numbers are drawn?

Number of possible outcomes: This is a combination, ${}_{48}C_6 = 12,271,512$

Number of favorable outcomes: This is a simple event, 1

$$P(\text{Win Lottery}) = \frac{1}{{}_{48}C_6} = 0.0000000815 = 0.00000815\%$$

Example 2

If you throw two dice, what is the probability you do not throw any 4, 5, or 6's?

To count the number of possible outcomes, we will use the counting rule:

Number of outcomes for one die: 6

Number of die: 2

Total possible outcomes: $6 \times 6 = 36$

To count the number of favorable outcomes, we need to use the counting rule:

Each die has 3 favorable outcomes: 1,2, or 3

Since there are 2 dice, number of favorable outcomes is $3 \times 3 = 9$

$$P(\text{no 4, 5, or 6}) = \frac{9}{36} = \frac{1}{4} = 0.25$$

Example 3

Suppose you throw three dice, what is the probability you do not throw any 4, 5, or 6's?

To count the number of possible outcomes, we will use the counting rule:

Number of outcomes for one die: 6

Number of die: 3

Total possible outcomes: $6 \times 6 \times 6 = 216$

To count the number of favorable outcomes, we need to use the counting rule:

Each die has 3 favorable outcomes: 1, 2, or 3

Since there are 3 dice, the number of favorable outcomes is $3 \times 3 \times 3 = 27$

$$P(\text{no 4, 5, or 6}) = \frac{27}{216} = \frac{1}{8} = 0.125$$

Example 4

You have to select a 4-digit PIN for your ATM/debit card. What is the probability that if randomly selected, there are no repeated digits? There are 10 digits to choose from 0-9

To count the total number of possible outcomes, this is a permutation with repetition because each pin number can be repeated:

Number of possible outcomes for first choice (n): 10

Number of times this choice is repeated (r): 4

Number of possible outcomes: $10^4 = 10,000$

To count the number of events where there are no repeated digits, we need to use a permutation without repetition:

Number of possible choices (n): 10

Number of times this choice is repeated (r): 4

Number of favorable outcomes: ${}_{10}P_4 = \frac{10!}{(10-4)!} = 5,040$

$$P(4 \text{ Digit PIN without repeated numbers}) = \frac{5,040}{10,000} = 0.504$$

Example 5

Suppose there are 10 people in a room. What is the probability that at least 2 of them have a birthday that falls on the same day?

First, we want to calculate the probability of no repeated birthdays, then we want to calculate the opposite where there is a repeated birthday. This is a permutation because the order of days matters.

For the number of possible outcomes, people *could* have the same birthday, so this is a permutation with repetition:

Number of possible outcomes (number of days in a year) (n): 365

Number of times first choice is repeated (number of birthdays) (r): 10

Number of possible outcomes: 365^{10}

For the number of favorable outcomes, people all have to have different birthdays, so this is a permutation without repetition:

Number of possible choices (n): 365

Number of choices selected (r): 10

Number of Favorable Outcomes: ${}_{365}P_{10} = \frac{365!}{(365 - 10)!} = 3.71 \times 10^{25}$

$$P(\text{No Repeated Birthdays}) = \frac{{}_{365}P_{10}}{365^{10}} = 0.883$$

$$P(\text{Repeated Birthday}) = 1 - P(\text{No Repeated Birthdays})$$

$$P(\text{Repeated Birthday}) = 1 - 0.883 = 0.117$$

8.4 Joint Probability

When thinking about multiple events, we can visualize them with the sample space. The sample space contains all possible outcomes of a random experiment. A random experiment is a process that can be repeated and has a defined set of possible outcomes).

Suppose that there are two events, A and B that exist in the same sample space. There are two questions that are typically asked of events A and B:

1. What is the probability that event A *and* event B occur?
 - This is called the intersection of events A and B.
 - This is denoted by $P(A \cap B)$.

2. What is the probability that event A *or* event B occur?

- This is called the union of events A and B.
- This is denoted by $P(A \cup B)$.

8.4.1 The Intersection of two Events

The $P(A \cap B)$ can also be expressed as $P(A \text{ and } B)$. To find this probability, it depends on whether the events are independent or dependent.

1. Two events can occur, but one event does not depend on another event. These are called independent events.
2. Two events can occur, but the second event depends on the outcome of the first event. These are called dependent events.

8.4.1.1 Independent Events

Multiple events that occur, but neither event that occurs is dependent on the other event, are called independent events. To calculate the probability that two events occur together, we can use the counting rule:

$$P(A \cap B) = P(A \text{ and } B) = P(A) \times P(B) \quad (8.5)$$

Example

If you roll three dice, what is the probability of getting three fours?

$$P(\text{four}) = \frac{1}{6}$$

$$P(\text{three fours}) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216} = 0.00463$$

8.4.1.2 Dependent Events

Multiple events that occur, but an event that occurs depends on the outcome of a previous event, are called dependent events. To calculate the probability of event A occurring, then event B occurring conditional on event A having occurred, we can use the counting rule as follows:

$$P(A \cap B) = P(A \text{ and } B) = P(A) \times P(B|A) \quad (8.6)$$

The notation for this is $P(B|A)$, which means the probability that event A occurs conditional on event B having occurred.

Example

Suppose that you pick two cards out of a deck of cards. What is the probability that the first card is an ace and the second card is a king?

$$P(\text{Ace}) = \frac{4}{52}$$

Once the ace has been removed, there are only 51 cards left, so the total possible outcomes changes:

$$P(\text{King} | 1^{\text{st}} \text{ Card Ace}) = \frac{4}{51}$$

$$P(\text{Ace and King}) = \frac{4}{52} \times \frac{4}{51} = \frac{4}{663} = 0.006$$

8.4.2 The Union of two Events

To find the probability of the union of two events, or the probability that A or B occur, we can use the addition rule:

$$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B) - P(A \cap B) \quad (8.7)$$

As we know from the previous section, $P(A \cap B)$ (or $P(A \text{ and } B)$) depends on whether the events are independent or dependent.

Example

An analyst estimates the probability that company A goes out of business is 25% and the probability that company B goes out of business is 40%. The probability that both companies go out of business is 10%. What is the probability that at least one of the companies goes out of business?

This question is asking what is the probability that company A or company B will go out of business.

$$P(A \cup B) = P(A \text{ or } B) = 0.25 + 0.40 - 0.10 = 0.55 = 55\%$$

8.4.2.1 Mutually Exclusive Events

Mutually exclusive events are where there are two events, A and B, and only one of them occurs. In this instance, $P(A \cap B) = P(A \text{ and } B) = 0$. Therefore, to calculate the probability of mutually exclusive events, we use the following rule:

$$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B) \quad (8.8)$$

Example

If you pick the top card of a deck of cards, what is the probability that it is a king or a queen?

$$P(\text{king}) = \frac{4}{52}$$

$$P(\text{queen}) = \frac{4}{52}$$

$$P(\text{king or queen}) = \frac{4}{52} + \frac{4}{52} = \frac{2}{13} = 0.15$$

Note, in the previous example with company A and company B, those were not mutually exclusive events because both company A and company B could go out of business at the same time, whereas in this example, you could not pull both a king and a queen at the same time.

8.4.3 Bayes Rule

When considering conditional probabilities, it is really important to make sure that the statistic you are looking at is what you are actually interested in because $P(A|B) \neq P(B|A)$. Let's consider the following example:

Example

Suppose that you have 100 people. 50 are vaccinated. 50 are not. 14 people are sick, 2 of them are vaccinated.

You see the following headline: "14.3% of all those who are sick are vaccinated." What is this conditional probability?

This is the probability of being vaccinated if you are sick: $P(\text{vaccinated}|\text{sick}) = \frac{2}{14} = 0.1428$

However, there are more useful probabilities for this issue:

The probability of being sick if you are vaccinated:

$$P(\text{sick}|\text{vaccinated}) = \frac{2}{50} = 0.04$$

The probability of being sick if you are not vaccinated:

$$P(\text{sick}|\text{not vaccinated}) = \frac{12}{50} = 0.24$$

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

where,

$P(A)$ = The probability of event A occurring.

$P(B)$ = The probability that event B occurs.

$P(A|B)$ = The probability that event A occurs given that event B has occurred.

$P(B|A)$ = The probability that event B occurs, given that event A has occurred.

Bayes rule provides us with a way to find the relationship between $P(A|B)$ and $P(B|A)$. Sometimes we know $P(A|B)$ or $P(B|A)$ but not the other. Using Bayes Rule, as long as we know one, we can find the other. Let's look at our previous example:

Example

$$P(\text{sick}|\text{vaccinated}) = \frac{P(\text{sick})P(\text{vaccinated}|\text{sick})}{P(\text{vaccinated})} = \frac{\frac{14}{100} \times \frac{2}{14}}{\frac{50}{100}} = 0.04$$

Bayes rules also provide us with a way to update our beliefs based on the arrival of new, relevant pieces of evidence. For example, if we were trying to provide the probability that a given person has cancer, we would initially just say that it is whatever percent of the population has cancer. However, given additional evidence, such as the fact that the person is a smoker, we can update our probability, since the probability of having cancer is higher given that the person is a smoker. This allows us to utilize prior knowledge to improve our probability estimates.

Example

You have the following data from a hospital:

10% of your patients have liver disease ($P(A)$)

5% of your patients are an alcoholic ($P(B)$)

7% of those with liver disease are alcoholics ($P(B|A)$)

What is the probability of having liver disease if they are an alcoholic?

$$P(\text{Liver Disease}|\text{Alcoholic}) = \frac{P(\text{Liver Disease})P(\text{Alcoholic}|\text{Liver Disease})}{P(\text{Alcoholic})}$$

$$\text{Prob}(\text{Liver Disease}|\text{Alcoholic}) = .10 \times \frac{.07}{.05} = 0.14$$

- Baye's rule can be extended if there are multiple groups:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)}$$

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

where,

$P(A_i)$ = The probability of event A for group i occurring

$P(B|A_i)$ = The probability of event B conditional on being in group i occurring

Example

There is a test for an allergy to cats, but this test is not always right. For people who have an allergy to cats, the test correctly says “yes” 80% of the time. For people who do not have an allergy to cats, the test says “yes” 10% of the time (false positive). If 1% of the population has an allergy to cats and you take a test that says you have a cat allergy, what is the probability you actually have a cat allergy?

To solve this, we need to use Bayes Rule:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$

Group 1: Has allergy

Group 2: Does not have allergy

$$P(\text{Allergy}|\text{Pos}) = \frac{P(\text{Allergy})P(\text{Pos}|\text{Allergy})}{P(\text{Allergy})P(\text{Pos}|\text{Allergy}) + P(\text{No Allergy})P(\text{Pos}|\text{No Allergy})}$$

$$P(\text{Allergy}) = 0.01$$

$$P(\text{Positive}|\text{Allergy}) = 0.80$$

$$P(\text{Not Allergy}) = 1 - P(\text{Allergy}) = 0.99$$

$$P(\text{Positive}|\text{No Allergy}) = 0.10$$

$$P(\text{Allergy}|\text{Positive}) = \frac{(0.01)(0.80)}{(0.01)(0.80) + (0.99)(0.10)} = 0.0748$$

This means that if you test positive, there is a 7.48% chance that you are allergic to cats.

DATA VISUALIZATION

Statistics is the practice of collecting and analyzing data. There are two ways in which statistics can be applied. Descriptive statistics uses numerical and graphical analyses to summarize data, identify patterns in a data set, and present data in a manner that is easy to consume. Inferential statistics uses samples of data to make predictions and estimate parameters of a larger population.

Before we explore applications of statistics, there are several important definitions that are necessary when looking at data. In a data set, an observation is an object about which we collect data. An object could be a person, a business, a country, a transaction, etc. A variable is a characteristic or property of a specific observation. For example, we may have a variable that describes the income, age, gender, race, height, or weight of a person.

There are two types of variables that you will encounter when looking at a data set. The first type of variable is a quantitative variable, which takes numerical values for which adding and averaging makes sense. A quantitative variable can be either a discrete variable, which means that it can be counted and there are no in-between values (e.g., 0, 1, 2, 3, 4, etc.), or a continuous variable, which are variables that can take any value and are recorded as they naturally occur. Examples of discrete variables include the number of people attending an event, the number of people who are sick, and the number of shots a player takes in a game. Examples of continuous variables are height, weight, and speed.

The second type of variable is a qualitative variable (or sometimes called a categorical variable), which is a variable where something falls into different categories, and the number of times something occurs in each category can be counted. Examples of qualitative variables would be identifying an observation's gender (e.g., man or woman), marital status (e.g., married, single, divorced, widowed), or company they work for.

9.1 Collecting Data

Economists and data analysts analyze data to test the predictions of economic models, assess the impact of policy changes, and improve business operations. Data can be collected in several ways, including a designed experiment, an observational study, or a survey. The purpose of data collection is to have a representative sample of a larger population. It is often very difficult and expensive to collect data on every observation of a particular population. Let us take the United States population for example. There are ~ 330 million people living in the United States. It would be difficult to ask every person in the United States their opinion on an issue, but a representative sample is a subset of those who live in the United States that is large enough and has the same characteristics as the population. One great example of this is Gallup polling. Gallup asks many questions about public opinion in the United States using a representative sample. The easiest way to have a representative sample is to use a random sample. A random

sample is a subset of observations chosen from a larger population in which each observation is selected at random, where each selection occurs with the same probability. Other more complex types of sampling include stratified random sampling, cluster sampling, systemic sampling, and randomized response sampling. We will not cover the more advanced sampling methods in this book, but they are worth studying if you plan to pursue more advanced data collection.

Before we can analyze and use data to make inferences about a population, we must first know how to describe the data. The first way in which we describe the data is using descriptive statistics. First, we will talk about descriptive statistics of population-level data followed by sample data. The descriptive statistics of a population are called population parameter's.

9.2 Visualizing Data

Whether it is to see the distribution of data, compare categories of a variable, or tell a story, the adage “a picture is worth a thousand words” is very important when looking at data. Whether you are presenting research or a report for your job, a well-designed graph can often tell the entire story. The way you choose to visualize data can make or break someone’s perception of your work. In this chapter, we are going to discuss different ways to visualize different types of data, as well as best practices to make your graphs look neat and clean.

Depending on the type of data you have, different graphs may be appropriate.

9.2.1 Bar Graphs

A bar graph has a category on one axis and a value on the other axis. There are two instances where a bar graph can be useful. The categories can be on either the y- or x-axis. Some data may present better one direction or the other, you can easily graph both and decide which is best for your data.

9.2.1.1 Frequency Graph

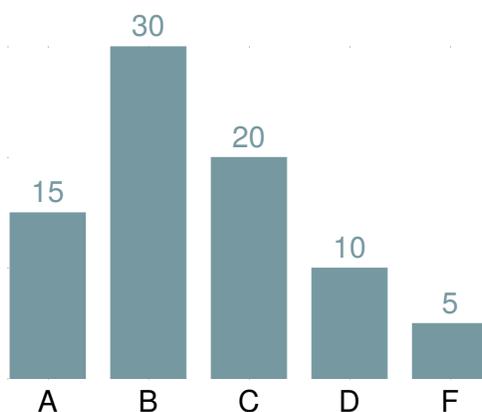
A bar graph can be used to graph the frequency of a discrete variable. This graph can be used to show the shape of the probability distribution of a discrete variable. Suppose that you have the following frequency table of grades in a class:

Table 9.1: Frequency Table for Grade Distribution

| | |
|---|----|
| A | 15 |
| B | 30 |
| C | 20 |
| D | 10 |
| F | 5 |

Using the data in Table 9.1, we can plot a bar chart:

Figure 9.1: Frequency Graph of Grade Distribution

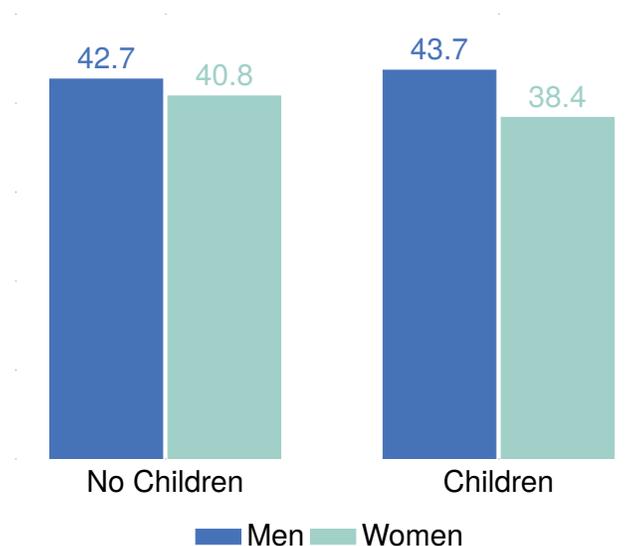


9.2.2 Categorical Variables

A categorical variable identifies specific categories that an observation can take (e.g., gender, race, age, religion, political affiliation, location, level of education, etc.). A bar graph can be used to graph another variable by category (e.g., income by gender, GPA by class year, population by state, etc.).

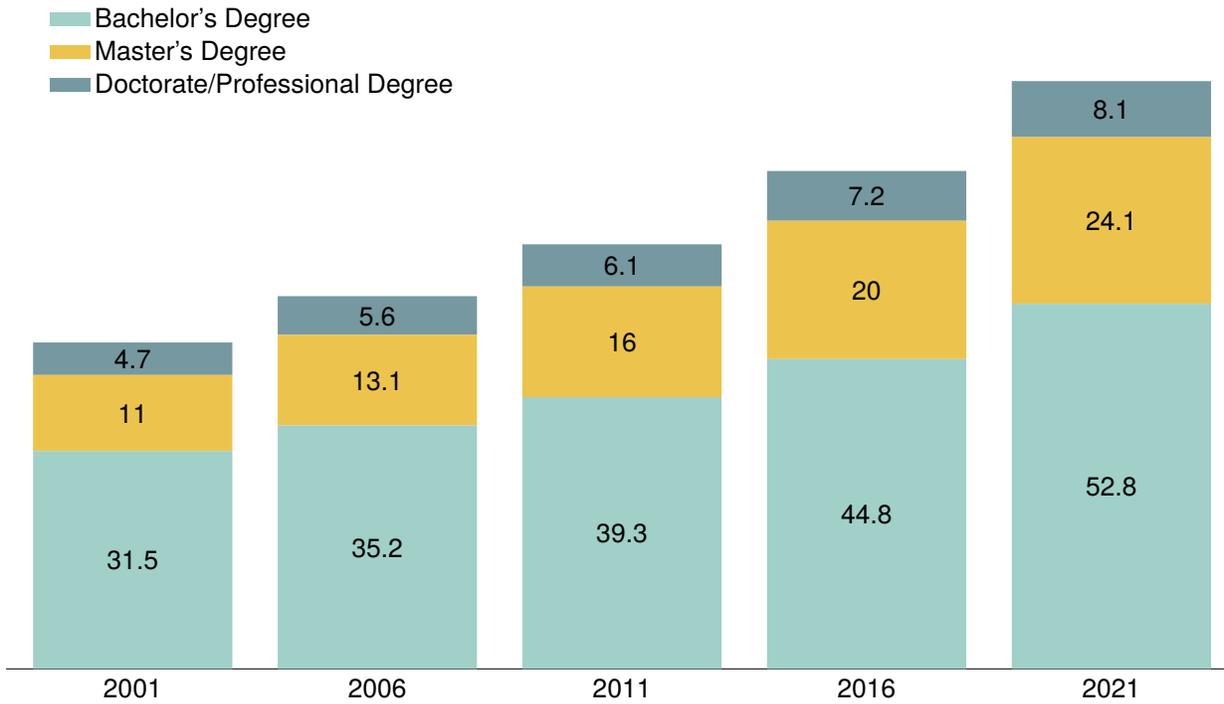
There are two ways in which a bar chart can be used when graphing categorical variables. Figure 9.2 shows a categorical bar chart showing the number of hours worked by men and women with and without children. Figure 9.3 shows a stacked bar chart showing the number of people who have different levels of education by year.

Figure 9.2: Hours Worked by Men and Women with and without Children, 2022



Source: Authors calculations using the Current Population Survey.

Figure 9.3: Population Breakdown by Higher Education (in Millions)

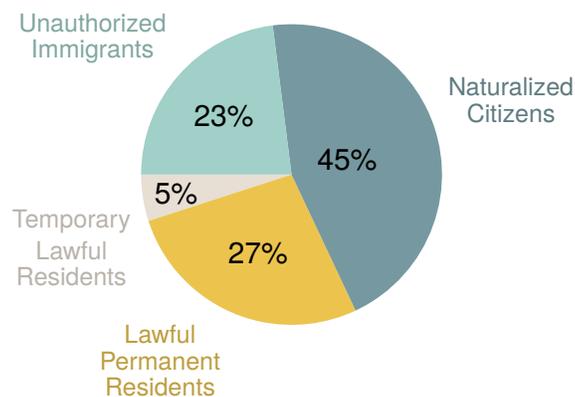


Source: Authors calculations using [US Census Bureau \(2022\)](#).

9.2.3 Pie Chart

A pie chart can be used to show the percentage of a specific variable that falls into different categories. Figure 9.4 shows the source of immigration in the United States.

Figure 9.4: Sources of Immigration in the United States, 2017

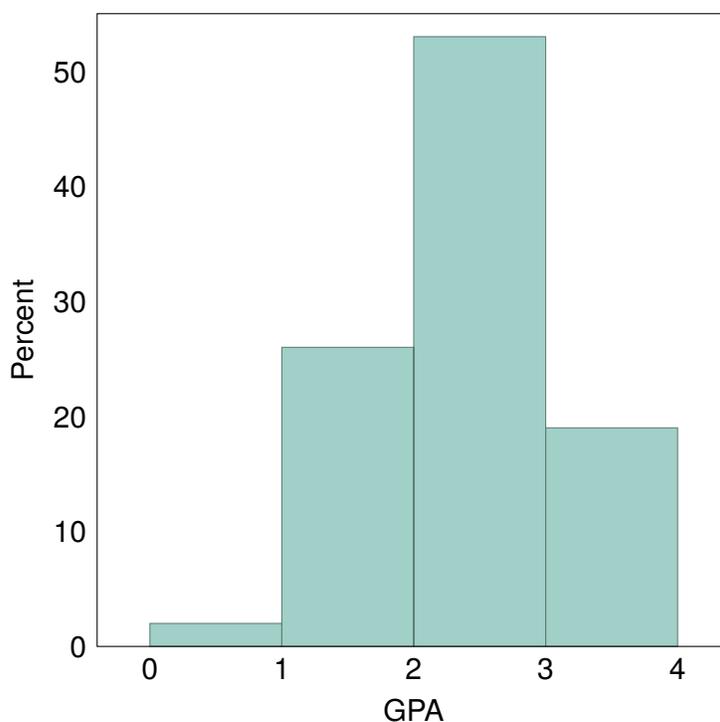


Source: [Passell and Cohn \(2019\)](#)

9.2.4 Histogram

A histogram can be used to graph the distribution of a continuous variable. The x-axis breaks the continuous variable into “bins” or “ranges” and displays the count or the percent of the observations for a specific variable that falls into each “bin.” Figure 9.5 shows a histogram of GPA's. GPA is a continuous variable that falls between 0 and 4. GPA is divided into four bins: 0.00-0.99, 1.00-1.99, 2.00-2.99, 3.00-4.00. The histogram shows the percentage of all grades that fall into each bin.

Figure 9.5: Histogram of GPA's

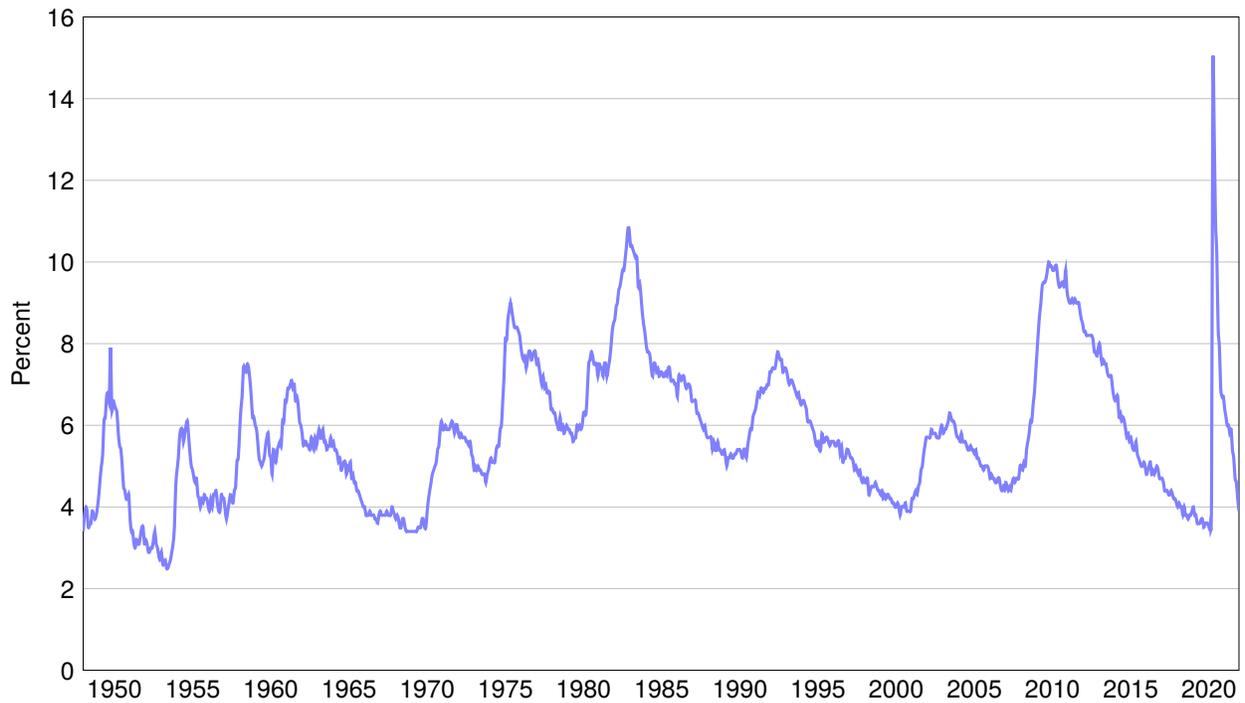


9.2.5 Line Graph

A line graph is used to graph the value of a variable over an interval of time. Figure 9.6 shows the unemployment rate in the United States from 1948-2021.

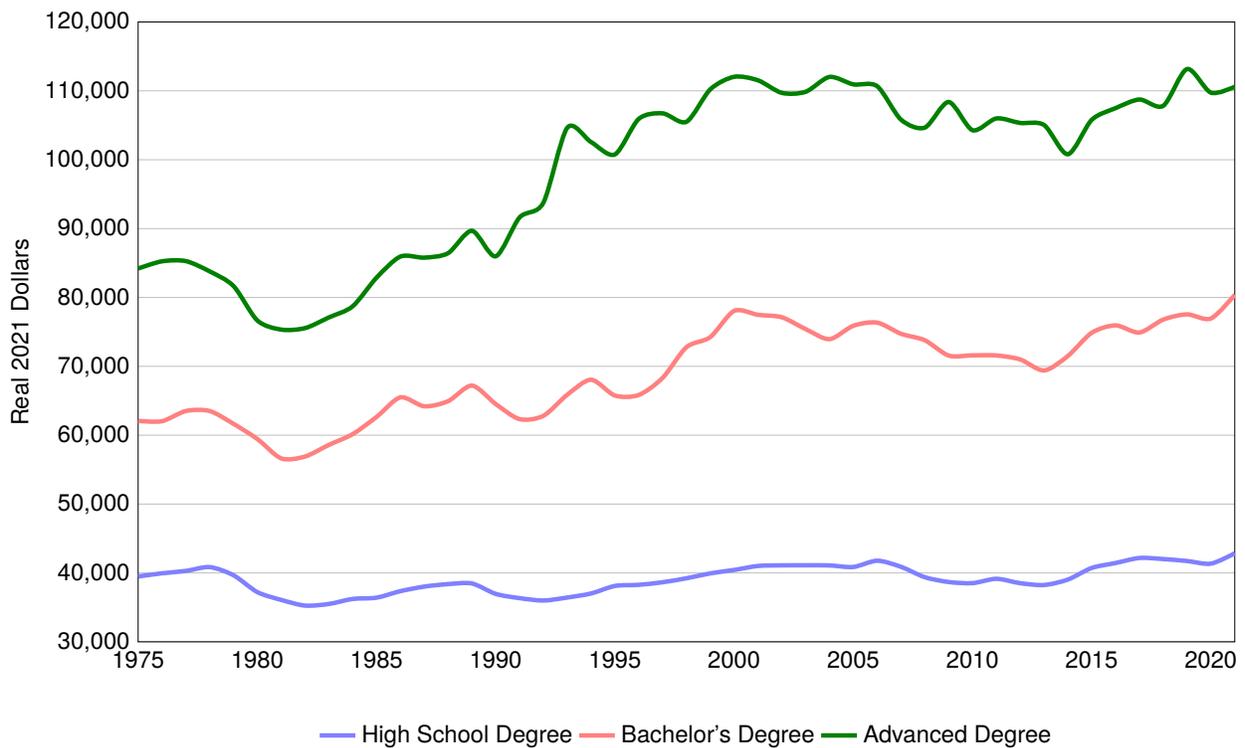
A line graph can also be used to graph the values of a variable over time by category. Figure ?? shows income by educational attainment.

Figure 9.6: Monthly Unemployment Rate, 1948–2021



Source: Bureau of Labor Statistics via FRED.

Figure 9.7: Mean Income Adjusted for Inflation by Education, 1975–2021

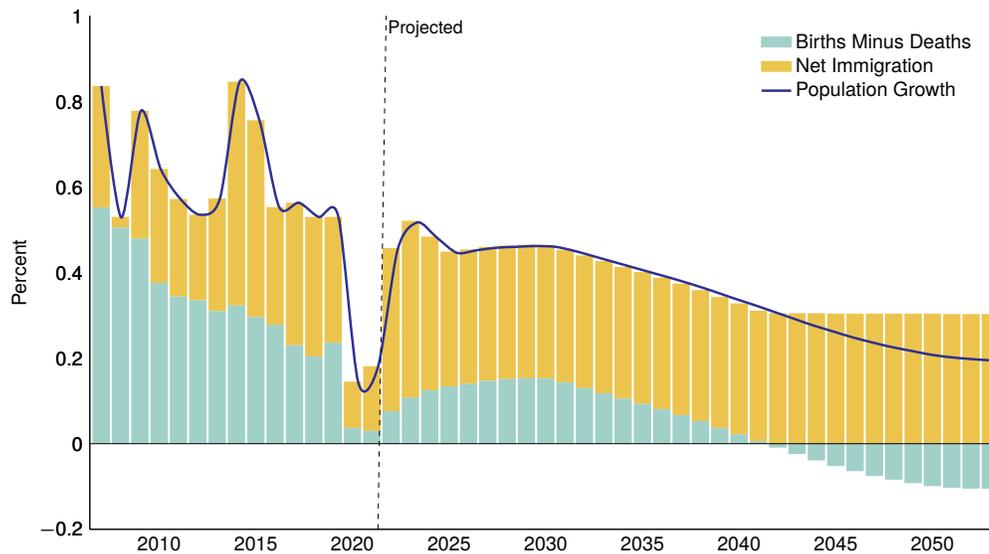


Source: Authors calculations using the [Current Population Survey \(2023\)](#) and the Bureau of Labor Statistics.

9.2.6 Advanced Graphs

Once you are comfortable with the graphing techniques in the previous section, you may want to explore combining multiple graph types, as in Figure 9.8.

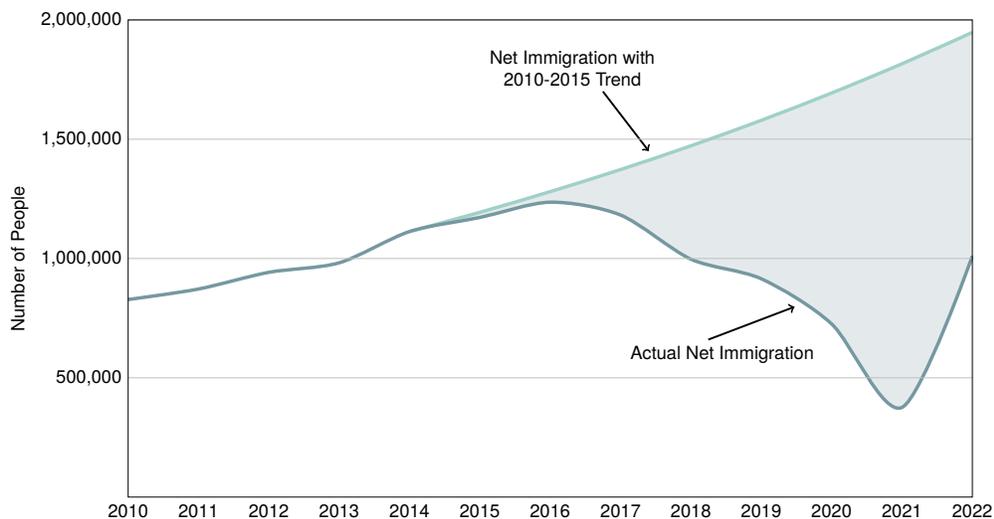
Figure 9.8: Population Growth and its Underlying Factors, 2007–2053



Source: Authors Calculations using [White House Council of Economic Advisors \(2023\)](#).

Or by using shading, arrows, and labels as in Figure 9.9.

Figure 9.9: Net Immigration into the United States, 2010-2022



Source: Authors calculations using [US Census Bureau \(2022\)](#).

DESCRIPTIVE STATISTICS

10.1 Descriptive Statistics of Population Data

10.1.1 Measures of Central Tendency

Measures of central tendency help to find the middle or “typical value” of a data set. There are two common measures of central tendency, the mean and the median.

10.1.1.1 Mean

The most common measure of central tendency is the mean (or average). The mean of a data set is the sum of the value of each observation divided by the number of observations. The mean of a population is expressed using the Greek letter “mu”

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

Where, N = the size of the population.

There are two instances where the mean can be misleading as a measure of central tendency. The first is if the data has outliers. An outlier is a value that is much larger or smaller than the rest of the data. The second is if the data is skewed (we will discuss this in more detail in the next chapter).

10.1.1.2 Median

The median is the middle value of a data set that has been ordered from the largest to the smallest. If there is an even number of observations, then we can calculate the median by taking the mean of the two middle scores. The median is less likely to be affected by outliers or skewed data. If the median and mean are not similar, it probably means that the data are skewed or that there are outliers (or both). The median of a population is expressed using the Greek letter “eta,” η .

10.1.2 Mode

The mode is the value that occurs most frequently in the population. It is possible for a variable to have multiple modes or no mode.

10.1.3 Measures of Variability

Measures of central tendency tell us where, on average, most of the data are. Measures of variability tell us how spread apart the data are. Two variables can have the same central tendency, but have various levels of variability. Combining these two measures can help paint a complete picture of the data.

10.1.4 Interquartile Range

The median tells us the middle value of a data set. That means that half of all the observations of a variable are less than the median and half are greater than the median. The median is also sometimes called the 50th percentile. A percentile is a metric that divides the data into two parts. There are two other common percentiles: the 25th and 75th percentiles. The 25th percentile is the value at which 25% of the values are below and 75% are above. The 75th percentile is the value at which 75% of the values are below and 25% are above. The 25th percentile is often referred to as the first quartile (or Q1) and the 75th percentile is often referred to as the third quartile (or Q3). However, there may be instances where other percentiles may be of interest. In general, the k -th percentile divides the data into two groups where k percent of the values are below and $1 - k$ are greater. The interquartile range (or IQR) is a measure of how spread out the middle half of the data are.

$$IQR = Q3 - Q1 \quad (10.1)$$

10.1.4.1 Variance

The variance is the average of squared deviations from the mean. A deviation from the mean distance an observation lies from the mean. The variance reflects the degree of spread of the data set; the more spread the data, the larger the variance. The variance of a population is denoted by using σ^2 and can be calculated by:

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N} = \frac{\sum_{i=1}^N (x_N - \mu)^2}{N}$$

10.1.4.2 Standard Deviation

The standard deviation is the average amount of variability in a sample. The standard deviation tells us, on average, how far each observation is from the mean. The larger the standard deviation, the more spread out the data is. The standard deviation of a population is denoted using σ and be calculated by taking the square root of the variance:

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^N (x_N - \mu)^2}{N}}$$

Example

Suppose you have the following data on returns from your stock investments for the last five years which represents the population of your investments:

| | |
|------|--------|
| 2020 | 31.10% |
| 2019 | -4.41% |
| 2018 | 21.94% |
| 2017 | 11.93% |
| 2016 | 1.31% |

Find the Mean, Median, Variance, and Standard Deviation.

Mean:

$$\mu = \frac{31.10 - 4.41 + 21.94 + 11.93 + 1.31}{5} = 12.37$$

Median: (Arrange from smallest to largest)

-4.41 1.31 11.93 21.94 31.10 → 11.93 is the median

Variance:

$$\sigma^2 = \frac{(31.10 - 12.37)^2 + (-4.41 - 12.37)^2 + (21.94 - 12.37)^2 + (1.31 - 12.37)^2}{5}$$

$$\sigma^2 = 169.3$$

Standard Deviation:

$$\sigma = \sqrt{\frac{(31.10 - 12.37)^2 + (-4.41 - 12.37)^2 + (21.94 - 12.37)^2 + (1.31 - 12.37)^2}{5}}$$

$$\sigma = 13.0$$

10.1.5 Coefficient of Variation

Sometimes we may want to compare the spread of the data between two variables with different units of measurement. The coefficient of variation (CV) is a measure that adjusts for the different levels of measurement and the magnitudes of the mean, making it a unitless statistic. The CV is calculated by dividing the standard deviation by its mean.

$$CV = \frac{\sigma}{\mu}$$

10.2 Descriptive Statistics of a Sample

Because it is often difficult to obtain data on a full population, we use a sample of data drawn from a larger population. The reason we can use a sample to make inferences on a larger population is the Law of Large Numbers. The Law of Large Numbers states that as a sample size grows, the mean of the sample will converge to the mean of the population. Therefore, as long as the sample size is large enough, the sample can be representative of the population. The appropriate sample size depends on the margin of error that you are willing to accept in your analysis.

We can calculate descriptive statistics of a sample just as we did with a population and call them sample statistic's. The notation for the population parameters differs from the sample statistics. These differences can be found in Table 10.1 showing the different notation.

Table 10.1: Notation for Population Parameters and Sample Statistics

| | Population Parameter | Sample Statistic |
|--------------------|----------------------|------------------|
| Mean | μ | \bar{x} |
| Median | η | M |
| Variance | σ^2 | s^2 |
| Standard Deviation | σ | s |
| Population Size | N | |
| Sample Size | | n |

10.2.0.1 Mean, Median, and Mode

The mean of a sample can be calculated using the same formula as a population, except that the sample size is used instead of the population size:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Where, n = the number of observations in the sample.

The median and mode of a sample are calculated the same way as the population.

10.2.1 Measures of Variability

The interquartile range is calculated in the same way for a sample as for the population. However, when we have a sample of data, we must make a small adjustment when calculating the variance and standard deviation.

10.2.1.1 Variance

The variance of a sample can be calculated by:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

The main difference between calculating the variance of a sample versus the variables of the population is the denominator. When calculating σ^2 , the denominator is N and when calculating s^2 , the denominator is $n - 1$. When calculating the variance of a sample, we use the mean of the sample. Since the mean of a sample is dependent on the sample itself, we could get a biased result when calculating the variance. By dividing by a smaller number in $n - 1$, this gives us a larger variance and corrects for any possible bias that could result from using a sample of data from a larger population.

10.2.1.2 Standard Deviation

The standard deviation of a sample can be calculated by:

$$s = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

10.2.2 Coefficient of Variation

The CV for a sample is calculated the same as for a population, dividing the sample standard deviation by the sample mean

$$CV = \frac{s}{\bar{x}} \quad (10.2)$$

10.2.3 Boxplot

A boxplot (sometimes called a box-and-whisker plot) is a graphical way to show measures of central tendency, variability, skewness, and outliers of the data. There are several steps in drawing a boxplot. First, we need to determine if there are any outliers in the data. A value is considered an outlier if it is 1.5 times the interquartile range. In other words, if a value is greater than $1.5 \times Q3$ or less than $1.5 \times Q1$, it is an outlier. If there are no outliers, then a boxplot is graphed as follows:

10.3 Relationship between two Variables

Sometimes we may have multiple variables in a data set and we want to understand the relationship between those two variables. This can be very useful when doing forecasting, machine learning, and data analysis. There are two measures that measure the relationship and dependence between two variables: covariance and correlation.

10.3.1 Covariance

Covariance, $cov(x, y)$, measures the directional relationship between two variables. If the covariance is positive, the variables move together (i.e., if one increases, the other increases). If covariance is negative, this means that the variables move in opposite directions (i.e., if one increases, the other decreases). If the covariance is zero, the two variables are not related. The covariance between two variables of a sample¹ can be calculated using:

$$cov(x, y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n - 1} \quad (10.3)$$

$$cov(x, y) = \frac{1}{n - 1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (10.4)$$

Covariance cannot be compared between two variables because they are not standardized, it can only tell us the direction in which two variables move.

10.3.2 Correlation

Correlation, ρ , measures the directional relationship and the strength of that relationship between two variables. Correlation is standardized, so its values fall between -1 and 1. If $\rho > 0$, then the two variables have a positive correlation and move in the same direction. If $\rho < 0$, then the two variables have a negative correlation and move in the opposite direction. The closer ρ gets to 1 or -1, the stronger the relationship is. If $\rho = 1$, then both variables move in exactly the same direction and in the same magnitudes. If $\rho = -1$, then both variables move in exactly the opposite direction with the same magnitudes. Correlation between two variables of a sample² can be calculated by:

$$\rho = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{(n - 1)s_x s_y} \quad (10.5)$$

$$\rho = \frac{1}{(n - 1)s_x s_y} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (10.6)$$

Where,

s_x is the sample standard deviation of x

s_y is the sample standard deviation of y

Substituting Equation 10.4 into Equation 10.6, we can get a more simplified equation to calculate correlation:

$$\rho = \frac{cov(x, y)}{s_x s_y} \quad (10.7)$$

¹ We can calculate the covariance between two variables of a population using the same equation except substituting $n - 1$ for N and σ for s .

² We can calculate the correlation between two variables of a population using the same equation except substituting $n - 1$ for N and σ for s .

10.4 Mean-Variance Analysis

Mean-variance analysis is the process of weighing the risk of an asset, expressed as variance, against its expected return. Generally, investments with a higher rate of return are typically more risky. A metric that can be used for mean-variance analysis is the Sharpe ratio. The Sharpe ratio tells us how well the return of an asset compensates for its risk. The Sharpe ratio measures the extra reward per unit of risk and can be calculated as follows:

$$\text{Sharpe Ratio} = \frac{\bar{x}_i - R_f}{s_i} \quad (10.8)$$

Where \bar{x}_i is the mean return of an investment, R_f is the mean return for a risk-free asset (such as a treasury bill), and s_i is the standard deviation for an investment.

PROBABILITY DISTRIBUTIONS

When it comes to real-world data and events, we often want to know how likely a specific value or event is to occur. We may want to know

- The probability that a child will grow to be more than 6 feet tall.
- The probability that a basketball player will score more than 20 points in a game.
- The probability that a fast food restaurant will serve more than 20 customers in an hour.
- The probability that a state will vote for a specific presidential candidate.

With real-world events, we need to find a way to characterize all the possible outcomes of an event. Before we can dive into that, we first need to go over some key terminology.

Table 11.1: Key Terminology for Probability Distributions

| | |
|------------------------------------|---|
| Random Variable | A value that characterizes the outcomes of a random occurrence. |
| Algebraic Variable | An algebraic variable is an unknown value that can be calculated using a mathematical equation. |
| Event | The set of possible values (or outcomes) of a random variable. |
| Probability Function | Describes the probability that each event occurs. |
| Cumulative Density Function | Function that evaluates the probability that a random variable will take a value less than or equal to some value k . |

As we discussed in Chapter 9, there are different types of variables, discrete and continuous. We will discuss how these terms apply to each type of variable.

11.1 Discrete Random Variables

11.1.1 Probability Mass Function

A discrete variable is countable. Examples of discrete variables include the number of people in line for a roller coaster, the size of people's shoes, the number of students in a class, the

number of shots attempted in a game, etc. The probability function for a discrete variable is called a probability mass function (PMF). The PMF of a discrete variable can take several forms: a mathematical function, a frequency table, or a histogram. The form of the PMF will depend on the type of event being analyzed.

The PMF shows all possible values of a discrete variable and the probabilities that each value occurs. The probability of each value occurring must fall between 0 and 1 ($0 \leq p_i \leq 1$). Furthermore, the sum of all the probabilities must sum up to 1 ($\sum p_i = 1$).

Example

When rolling a single die, there are six possible outcomes that can be distinctly counted and there are no “in-between” outcomes. This makes the values that can occur from rolling a single die discrete. We can show the PMF of rolling a single die using a frequency table:

| | | | | | | |
|-------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Outcome of Roll | 1 | 2 | 3 | 4 | 5 | 6 |
| Probability of Outcome | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

If we want to know the probability that one outcome will occur, we can calculate that from the PMF (e.g., $P[\text{roll} = 3] = \frac{1}{6}$).

Additionally, you will notice that all the probabilities of the PMF sum to 1:

$$\sum p_i = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

11.1.2 Cumulative Density Function

The cumulative density function (CDF) of a discrete variable shows the probability that an outcome is less than or equal to a value k ($P[\text{outcome}] \leq k$). The $P[\text{outcome}] > k = 1 - P[\text{outcome}] \leq k$, where the $P[\text{outcome}] \leq k$ is determined by the CDF. For a discrete variable, the CDF can take the form of a mathematical function, a frequency table, or a histogram. The CDF of a discrete variable can be found by summing the probabilities of the PDF.

Example

Building on our previous example, we can show the CDF for rolling a single die:

| | | | | | | |
|-------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Outcome of Roll | 1 | 2 | 3 | 4 | 5 | 6 |
| Probability of Outcome | $\frac{1}{6}$ | $\frac{2}{6}$ | $\frac{3}{6}$ | $\frac{4}{6}$ | $\frac{5}{6}$ | $\frac{6}{6}$ |

From the CDF, we can find the probability that a single die roll is less than or equal to 4 ($P[\text{roll} \leq 4]$).

From the CDF:

$$P[\text{roll} \leq 4] = \frac{4}{6} = \frac{2}{3}$$

If we wanted to find the probability that a single die roll is greater than 4:

$$P[\text{roll} > 4] = 1 - P[\text{roll} \leq 4]$$

$$P[\text{roll} > 4] = 1 - \frac{4}{6} = \frac{2}{6} = \frac{1}{3}$$

11.1.3 Expected Value of a Discrete Variable

The expected value (or the mean) of a discrete random variable can be interpreted as the anticipated outcome of a discrete random variable.

$$\mu = E[x] = p_1x_1 + p_2x_2 + \cdots + p_nx_n = \sum_{i=1}^n p_i x_i \quad (11.1)$$

Where,

p_i = the probability of a possible outcome

x_i = the result from the possible outcome

Example 1

The probability of winning the Powerball lottery is 0.000000342%.

Suppose that the jackpot is \$10 million and a Powerball ticket costs \$2. What would your expected earnings be?

There are two outcomes: you win, or you don't win. Using this, we can set up a probability table knowing that regardless of whether you win or don't, you still must pay \$2 for the lottery ticket:

| | Win | Don't Win |
|-------------|----------------------------------|------------------|
| Payoff | \$10,000,000 – \$2 = \$9,999,998 | \$0 – \$2 = –\$2 |
| Probability | 0.000000342% | 99.99999966% |

$$E[\text{Lottery Return}] = .00000000342(\$9,999,998) + 0.9999999966(-\$2) = -\$1.97$$

This means that your expected return from buying a lottery ticket is to lose \$1.97.

Example 2

Industry analysts have derived the following discrete probability distribution for stock market returns:

| Stock Market Returns | Strong | Average | Weak |
|----------------------|--------|---------|------|
| Return | 8% | 2% | -2% |
| Probability | 10% | 70% | 20% |

From this table we can calculate the expected value (or mean) for stock market returns:

$$\mu = E[\text{return}] = 0.10(0.08) + 0.70(0.02) + 0.20(-0.02) = 0.018 = 1.8\%$$

The expected value (or mean) of a discrete random variable tells us the most likely outcome. However, as we learned in Chapter 10, two variables can have the same mean, but they have a wide range of possible values. To understand how spread out the values of a discrete probability distribution are, we can calculate the variance (σ^2) and the standard deviation (σ) of a discrete random variable:

$$\sigma^2 = E[(x - \mu)^2] = \sum_{i=1}^n p_i(x_i - \mu)^2 \quad (11.2)$$

$$\sigma = \sqrt{\sigma^2} \quad (11.3)$$

Example

We can calculate the variance and standard deviation using the previous example of stock returns, where we found that $\mu = 1.8\%$.

$$\sigma^2 = .10(.08 - .018)^2 + .70(.02 - .018)^2 + .20(-.02 - .018)^2 = .000676 = .0676\%$$

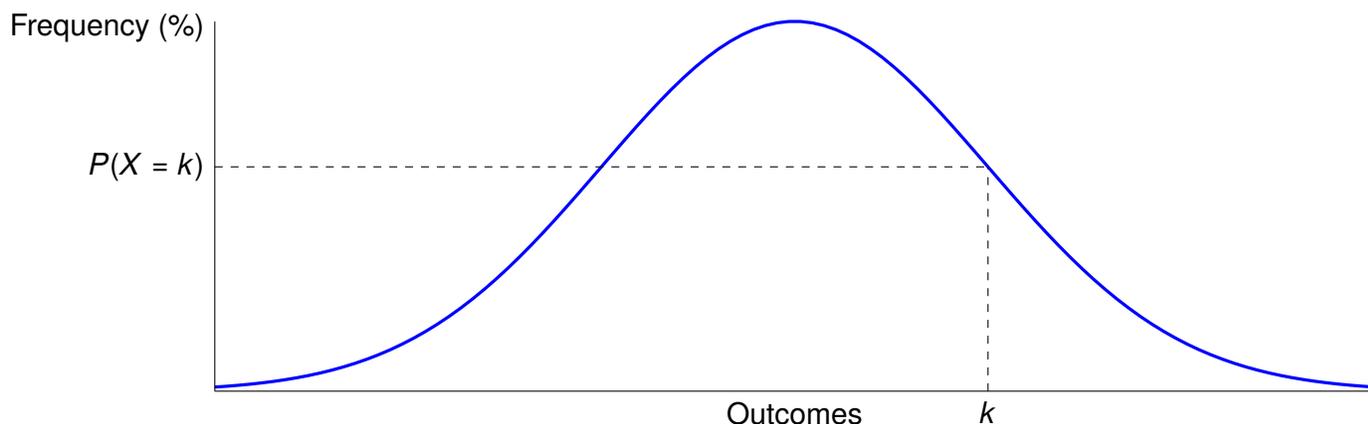
$$\sigma = \sqrt{.000676} = .026 = 2.6\%$$

11.2 Continuous Random Variables

11.2.1 Probability Density Function

A continuous variable can take any value. Examples of continuous variables include height, weight, distance, time, etc. The probability function for a continuous variable is called a probability density function (PDF). The PDF of a continuous variable is expressed as a mathematical function. Like the PMF, the PDF shows all possible values of a continuous variable and the probability that each value occurs. If we were to plot a graph of the PDF, we can visually see the outcomes and their probabilities.

Figure 11.1: PDF of a Continuous Random Variable



If we evaluate the PDF of a continuous variable, we can determine the probability that a specific value k will occur.

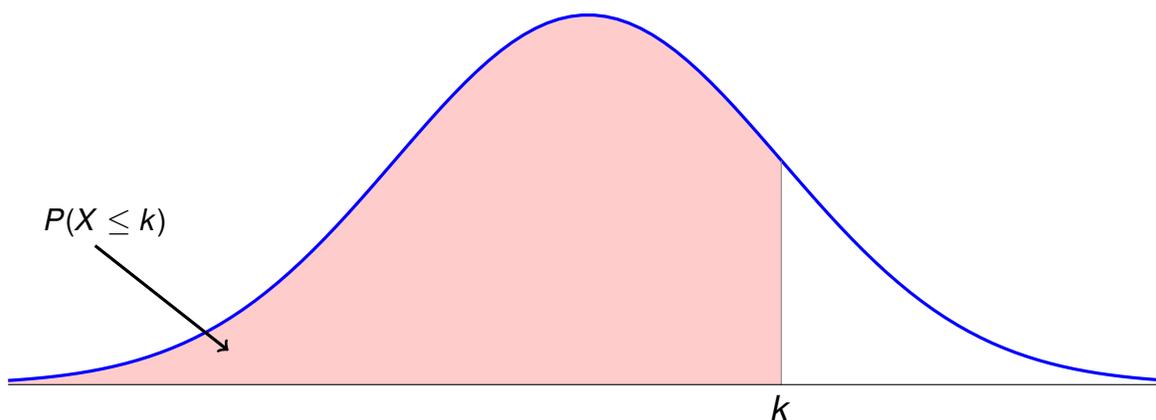
Unlike the PMF of a discrete function, there are infinite possible values that the PDF of a continuous can take. This can be demonstrated by looking at weight. There are infinite values between 100 pounds and 101 pounds (i.e., 100.0001, 100.0002, 100.0003, etc.). Since the PDF represents every possible outcome of a variable, the area under the PDF must sum to 1.

11.2.2 Cumulative Density Function

If we want to find the probability that variable X is less than or equal to k , we need to evaluate the CDF. Because the PDF of a continuous variable contains infinite values, we cannot calculate the CDF simply by summing the values, as we did for a discrete variable. To evaluate the PDF, you must calculate the area under the PDF to the left of the value k .

Similarly to the CDF of a discrete variable, the CDF of a continuous variable evaluated at a specific value of k , gives the probability that an outcome will be less than or equal to k , or $P(X \leq k)$.

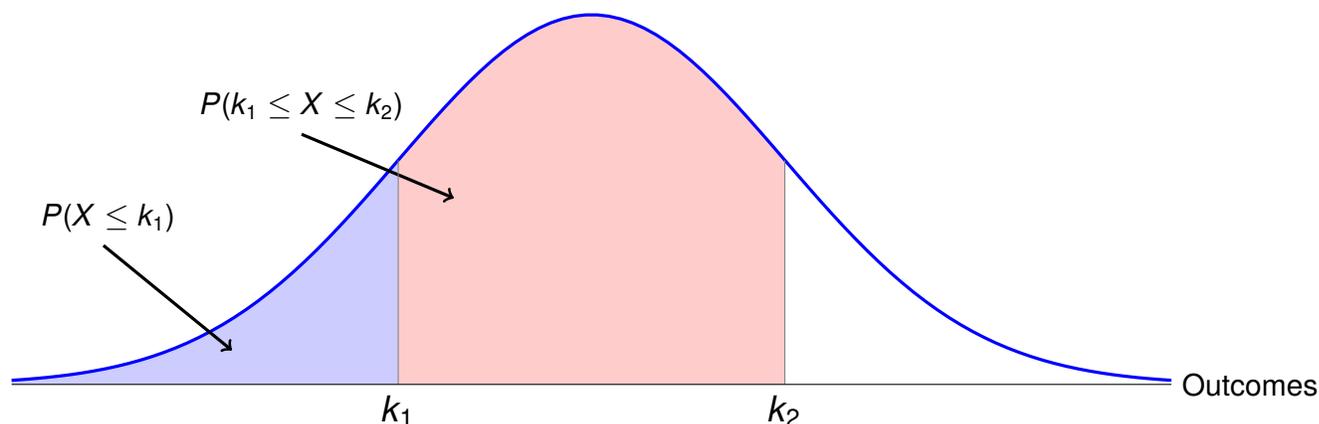
Figure 11.2: Evaluating the CDF of a Continuous Variable



Because the area under the PDF must sum up to 1, if we want the probability that the variable X is greater than k , we can find it by $P(X > k) = 1 - P(X \leq k)$, or 1 minus the CDF evaluated at the value of k .

Suppose that we want to find the probability that the variable X is between two values, k_1 and k_2 . First, evaluate the CDF at k_2 to find the probability that $P(X \leq k_2)$. Second, evaluate the CDF at k_1 to find the probability that $P(X \leq k_1)$. Last, subtract: $P(k_1 \leq X \leq k_2) = P(X \leq k_2) - P(X \leq k_1)$

Figure 11.3: Finding $P(k_1 \leq X \leq k_2)$ using the CDF



11.2.3 Expected Value of a Continuous Variable

Calculating the expected value (or mean) of a continuous variable requires the use of integration. We will go through how to calculate this by hand, but if integration is not required for your course, this can be calculated using a calculator or computer software.

The expected value (or mean) of a continuous random variable can be calculated by

$$\mu = E(x) = \int_a^b xf(x)dx \quad (11.4)$$

Where,

x is the variable of interest

$f(x)$ is the PDF of the continuous variable

a is the lower bound of the PDF

b is the upper bound of the PDF

The variance and standard deviation of a continuous random variable can be calculated by

$$\sigma^2 = E[(x - \mu)^2] = \int_a^b (x - \mu)^2 f(x)dx \quad (11.5)$$

$$\sigma = \sqrt{\sigma^2} \quad (11.6)$$

11.3 Common Probability Distributions

It is often helpful to graph a histogram for a continuous variable or a bar graph for a discrete variable. The shape of the graph will help determine the type of distribution of the data. Some probability distributions are well known with well-defined mathematical properties; others are not.

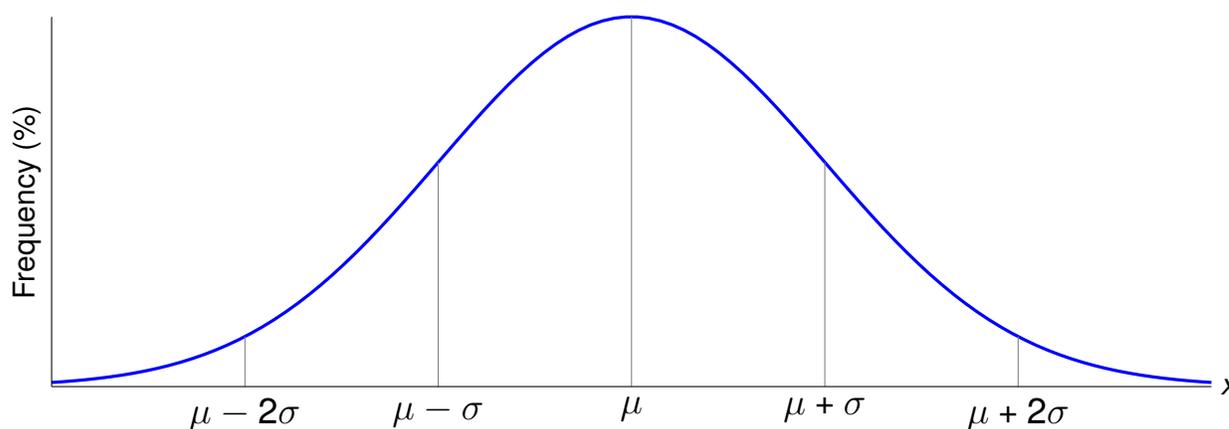
A distribution is symmetrical if the *mean* and *median* are similar. If the $mean > median$, then a distribution has a positive skew. A distribution with a positive skew is where the values are clustered towards the left side, with few values further out to the right side. Examples of data with a positive skew are income, house values, and the number of pets in a household. If the $mean < median$, then a distribution has a negative skew. A distribution with a negative skew is where the values are clustered toward the right side, with a few values closer to the left side. Examples of data with a negative skew are the age of death, GPA values, and stock market returns.

The Normal and Poisson distributions are two common distributions that are used in economics and business. Their properties will be discussed in detail in this section.

11.3.1 The Normal Distribution

The most common distribution seen in real-world data is the normal distribution. The normal distribution is often used in the natural and social sciences to represent random variables. Visually, the normal distribution has a bell-shaped curve, which can be seen in Figure 11.4.

Figure 11.4: The PDF of a Normal Distribution

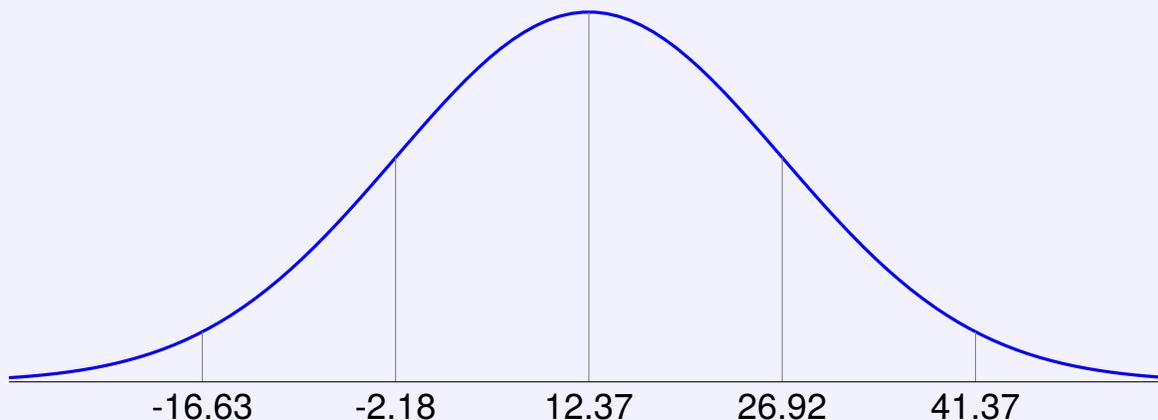


For a given variable X , if X is distributed normally, we use the following notation: $X \sim N(\mu, \sigma)$. The normal distribution has several nice properties that help us visualize the spread of the data. If the data are distributed “normally,” then 68.3% of all the data lies between ± 1 standard deviations from the mean and 95.5% of all the data lies between ± 2 standard deviations from the mean.

Example

Let us use data from our previous example using stock returns, where we found that $\mu = 12.37$ and $\sigma = 14.55$

$$\begin{aligned}\mu + \sigma &= 12.37 + 14.55 = 26.92 & \mu - \sigma &= 12.37 - 14.55 = -2.18 \\ \mu + 2\sigma &= 12.37 + 29.10 = 41.37 & \mu - 2\sigma &= 12.37 - 29.10 = -16.63\end{aligned}$$



This means that we can expect 68.3% of stock returns to fall between -2.18% and 26.92% and that 95.5% of stock returns will fall between -16.63% and 41.37%.

When data is distributed normally, the mean and median should have similar values. If we wanted to graph the PDF of the normal distribution, we would use the following equation, using the mean, μ , and the standard deviation, σ :

$$f(k) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-(k-\mu)^2/2\sigma^2} \quad (11.7)$$

We can use the PDF to find the probability that the variable X takes the value of k .

Example

Suppose that you have the height of a population of third-graders where the mean height is 37.0 inches with a standard deviation of 1.99. What is the probability that a third-grader is 39 inches?

$$\mu = 37.0$$

$$\sigma = 1.99$$

$$k = 39.0$$

$$f(k) = \frac{1}{1.99\sqrt{2\pi}} \cdot e^{-(39-37)^2/2(1.99)^2} = 0.12098$$

We can use the CDF of the normal distribution to find the probability that the variable X is

less than or equal to k , or $P(X \leq k)$. The CDF of the normal distribution is as follows:

$$\Phi(k) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^k e^{-(k-\mu)^2/2\sigma^2} \quad (11.8)$$

If we want to know the probability that X is greater than k , then: $P(X > k) = 1 - \Phi(k)$. You can also calculate the value of the CDF using a graphing calculator or Microsoft Excel if you are not familiar with integration.

Normal CDF with TI-83/84

Step 1: Press the 2nd key

Step 2: Press the VARS key

Step 3: Select option 2 or “normalcdf(”

The format you type in the calculator is: **normalcdf(lower bound, upper bound, μ , σ)**

Note: If you want $P(X \leq k)$, then the lower bound would be $-\infty$, which can be approximated on the calculator by using “-1e99” for the lower bound, which is the default option. If you want to find the $k_1 \leq X \leq k_2$, then k_1 would be the lower bound and k_2 the upper bound

Normal CDF with Microsoft Excel

=NORM.DIST(k , μ , σ , cumulative)

Where,

Note: If cumulative is set to “TRUE,” it returns the value of the CDF for a given value of k , if it is set to “FALSE,” it returns the value of the PDF for a given value of k .

Example 1

Using the same information as the previous example, what is the probability that the height of a third grader is less than or equal to 39, or $P(\text{height} \leq 39)$?

For TI-83/84

normalcdf(-1e99, 39, 37, 1.99)

For Microsoft Excel

=NORM.DIST(39, 37, 1.99, TRUE)

$$\Phi(39) = 0.84256$$

There is an 84% chance that the height of a third grader is less than or equal to 39 inches.

Example 2

The average price of gasoline in the United States is \$3.10 with a standard deviation of

\$0.60 and follows a normal distribution. What percent of gas stations charge between \$2.90 and \$3.25?

For TI-83/84

normalcdf(2.90, 3.25, 3.10, 0.60)

For Microsoft Excel

=NORM.DIST(3.25, 3.10, 0.60, TRUE) - NORM.DIST(2.90, 3.10, 0.60, TRUE)

$$\phi = 0.229$$

11.3.1.1 The Sampling Distribution

It is often the case that we do not have data on the entire population, but rather a sample of data that is drawn from a larger population. For example, a survey of public opinion in a state will only sample a few thousand people in the state, not all residents. When we have a sufficiently large sample of data, we can use the normal distribution to make inferences about the population. This is because of the Central Limit Theorem. The Central Limit Theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large samples of the population, then the distribution of the sample means will have a normal distribution, regardless of whether the true population is normal or not.

In order to use the normal PDF and CDF for a sample of data, we must use the standard error of the sample instead ($se(\bar{x})$) of the population standard deviation (σ). The standard error of a sample is defined as the population standard deviation divided by the square root of the number of observations:

$$se(\bar{x}) = \frac{\sigma}{\sqrt{n}} \quad (11.9)$$

Normal CDF of a Sampling Distribution with TI-83/84

Step 1: Press the 2nd key

Step 2: Press the VARS key

Step 3: Select option 2 or “normalcdf(”

The format you type in the calculator is: **normalcdf(lower bound, upper bound, μ , $se(\bar{x})$)**

Note: If you want $P(X \leq k)$, then the lower bound would be $-\infty$, which can be approximated on the calculator by using “-1e99” for the lower bound, which is the default option. If you want to find the $k_1 \leq X \leq k_2$, then k_1 would be the lower bound and k_2 the upper bound

Normal CDF of a Sampling Distribution with Microsoft Excel

=NORM.DIST(k , μ , $se(\bar{x})$, cumulative)

Where,

Note: If cumulative is set to “TRUE,” it returns the value of the CDF for a given value of k , if it is set to “FALSE,” it returns the value of the PDF for a given value of k .

Example 1

The average price of a house in a city is \$525,000 with a standard deviation of \$115,000. What is the probability that a random sample of 50 houses has a mean of less than \$500,000? For TI-83/84

normalcdf(-1e99, 500,000, 525,000, 115,000/ $\sqrt{50}$)

For Microsoft Excel

=NORM.DIST(500,000, 525,000, 115,000/ $\sqrt{50}$), TRUE)

$$\phi = 0.062$$

Example 2

What is the probability that a random sample of 100 houses has a mean greater than or equal to \$500,000?

For TI-83/84

1-normalcdf(-1e99, 500,000, 525,000, 115,000/ $\sqrt{100}$)

For Microsoft Excel

=1-NORM.DIST(500,000, 525,000, 115,000/ $\sqrt{100}$), TRUE)

$$1 - \phi = 0.985$$

11.3.1.2 The Sampling Distribution of the Population Proportion

In some instances, there are economic and business applications where you may be interested in the probability that something occurs in a population. For example, the CDC may be interested in the percentage of the population who has the flu; a bank may be interested in what percent of its clients are late on a loan payment; or a politician may care what percent of the population supports a specific issue. For these cases, we would be interested in the population proportion, p . Just like with the population mean, we often rely on a sample of data to make inferences about the population. The sample proportion is denoted by \bar{p} .

$$\bar{p} = \frac{X}{n} \quad (11.10)$$

Where X is a count of the favorable outcomes and n is the number of total outcomes in the sample.

Example

A sample of 1,000 students at a college are randomly tested for the flu and 102 are positive. What is the sample proportion?

$$\bar{p} = \frac{102}{1,000} = 0.102 = 10.2\%$$

The standard error of the sample proportion can be calculated as follows (note, p is the population proportion):

$$se(\bar{p}) = \sqrt{\frac{p(1-p)}{n}} \quad (11.11)$$

Normal CDF of a Sampling Distribution of a Proportion with TI-83/84

Step 1: Press the 2nd key

Step 2: Press the VARS key

Step 3: Select option 2 or “normalcdf(”

The format you type in the calculator is: **normalcdf(lower bound, upper bound, p , $se(\bar{p})$)**

Note: If you want $P(X \leq k)$, then the lower bound would be $-\infty$, which can be approximated on the calculator by using “-1e99” for the lower bound, which is the default option. If you want to find the $k_1 \leq X \leq k_2$, then k_1 would be the lower bound and k_2 the upper bound

Normal CDF of a Sampling Distribution of a Proportion with Microsoft Excel

=NORM.DIST(k , μ , $se(\bar{x})$, cumulative)

Where,

Note: If cumulative is set to “TRUE,” it returns the value of the CDF for a given value of k , if it is set to “FALSE,” it returns the value of the PDF for a given value of k .

11.3.2 Poisson Distribution

The Poisson probability distribution is a discrete probability distribution that expresses the probability that a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event. The Poisson distribution helps to predict the probability of certain events happening when you know how often the event has occurred. The Poisson distribution is often used by companies to make forecasts about the number of customers or sales on certain days, months, or seasons of the year. This is important to ensure that companies have the appropriate level of inventory on hand. Because the Poisson distribution is a discrete probability distribution, it has what is

called a Probability Mass Function (PMF), which operates similarly to the PDF of a continuous probability distribution. The PMF for the Poisson distribution is as follows:

$$f(k, \lambda) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \quad (11.12)$$

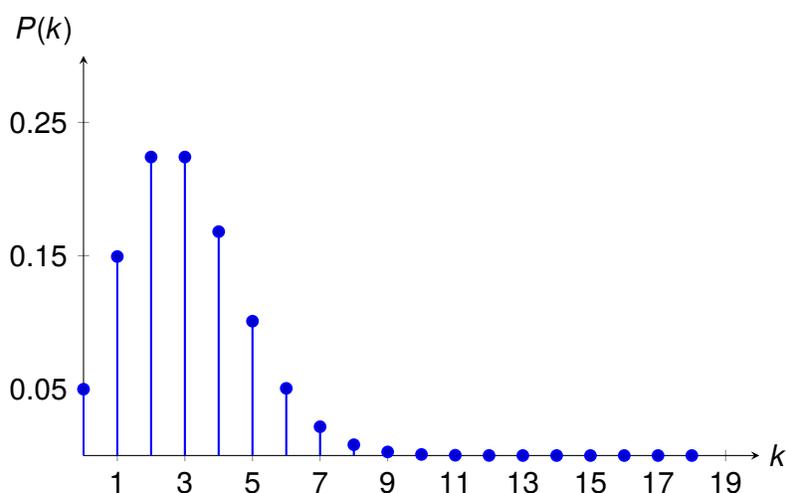
Where,

k is the number of occurrences

λ is the average rate

The PMF of the Poisson distributions tells us the probability that the variable X occurs k times, or $P(X = k)$. The notation for the Poisson distribution is $X \sim \text{Poisson}(\lambda)$

Figure 11.5: The PDF of a Poisson Distribution with $\lambda = 3$



Example

Leah's answering machine receives about six telephone calls between 8 a.m. and 10 a.m. What is the probability that Leah receives 3 calls in the next 15 minutes?

On average, Leah receives 6 telephone calls in two hours and we split each hour into 15-minute intervals.

There are 8 15-minute intervals in 2 hours, so that means Leah receives an average of $6/8 = 0.75$ calls every 15 minutes.

For this problem then, $k = 3$ and $\lambda = 0.75$

$$f(3, 0.75) = \frac{0.75^3 \cdot e^{-0.75}}{3!} = 0.03321$$

There is a 3.3% chance that Leah gets exactly 3 calls in the next 15 minutes.

If we wanted to find the probability that the discrete variable X is less than or equal to k , or $P(X \leq k)$, we can use the CDF of the Poisson distribution:

$$F(k, \lambda) = \sum_{i=1}^k \frac{\lambda^k \cdot e^{-\lambda}}{k!} \quad (11.13)$$

The CDF of the Poisson function essentially sums the probability of each k in the specified range. If we wanted to find the probability that X is greater than k , $P(X > k) = 1 - F(k, \lambda)$.

Normal CDF with TI-83/84

Step 1: Press the 2nd key

Step 2: Press the VARS key

Step 3: Select option D or "poissoncdf("

The format you type in the calculator is: **poissoncdf**(λ , x)

Normal CDF with Microsoft Excel

=POISSON.DIST(x, λ , cumulative) Where,

- x is the value of x
- λ is the mean
- cumulative is set to "TRUE" to return the value of the CDF, if set to "FALSE," it returns the value of the PDF for a given value of x

If the data follow a Poisson distribution, the mean and the variance of those data are equal λ ($\mu = \lambda$ and $\sigma^2 = \lambda$).

HYPOTHESIS TESTING

Hypothesis testing in statistics is a way to test if the results of an experiment or research produce meaningful results. You want to know if the results you have just happened by random chance. A valid experiment or research design should be able to be repeated and get similar results. Meaningful results are called statistically significant.

When collecting samples of data, we want to ask questions about the sample relative to the larger population. Some questions we may ask:

- Is the mean of the sample greater than the mean of the population?
- Is the mean of the sample different from zero?
- Is the mean for men different from the mean for women?

In order to answer questions like these, we first need to establish a testable hypothesis. There are two hypotheses that must be proposed to conduct a hypothesis test. The null hypothesis and the alternative hypothesis.

The null hypothesis, denoted by H_0 , proposes that there is no difference between certain characteristics of a population. The alternative hypothesis, denoted by H_1 , is a claim about the population that is contradictory to the null hypothesis. The goal of a hypothesis test is to determine if there is enough statistical evidence from the data to reject the null hypothesis to be able to accept the alternative hypothesis as true. One note of caution: if the hypothesis test fails to reject the null hypothesis, that does not mean that we accept the null hypothesis as true. Failing to reject the null hypothesis simply means that there is not enough evidence to accept the alternative hypothesis as true.

The statistical test you use to test the alternative hypothesis will depend on the distribution of your data and the question you are asking.

12.1 The T-Test

The T-Test is a hypothesis test that can be used to determine if the mean of a sample differs from a target value or if the means of two samples are different from each other if we do not know the standard deviation of the population. The T-Test is the most commonly used hypothesis test. The test determines if the differences are statistically significant or if they are due to random chance. The t-test can be used in two ways. A one-sample t-test is used to determine whether the unknown mean of a population differs from a specific target value if you have a random sample of data. A two-sample t-test is used to test if the means of two separate populations are equal if you have random samples of data.

12.1.1 One Sample T-Test

If you have a sample of data, we can use the t-test to determine if the mean of the population from which the sample is drawn is different from a specific target value. For example, suppose that you collect data on the prices of cigarettes before and after a change in the cigarette tax and you want to know if the tax caused an increase in prices. You don't have data on every store that sells cigarettes, only a few in the local area. You could use a t-test to determine whether the prices before and after the tax are statistically different or not.

When formulating the null and alternative hypotheses for the one sample t-test, it can be asked in two ways. The first way is to test to see if the mean is equal to or not equal to a specific target value. This is called a two-tailed test. The second way is to test if the mean is greater than or less than a specific target value. This is called a one-tailed test.

12.1.1.1 Two-Tailed Test

For a two-tailed t-test, we test to determine if the mean, \bar{x} , is statistically different from another value, m . For this two-tailed test, the null and alternative hypotheses are stated as:

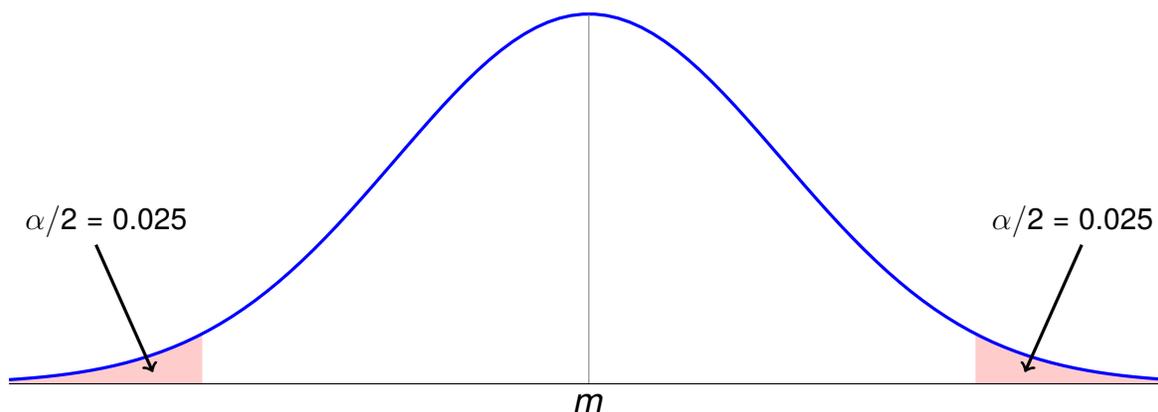
Null hypothesis: $H_0 : \bar{x} = m$

Alternative hypothesis: $H_1 : \bar{x} \neq m$

We use this test to determine whether we have collected enough data to determine whether the null hypothesis, H_0 , is false so that we can accept the alternative hypothesis, H_1 , as true.

Because we only have data on a sample of the population, we need to determine how sensitive our test is to the possibility of coming to an incorrect conclusion. We do this by determining a significance level, which is defined at α . The significance level is the probability of rejecting the null hypothesis when it is true. The significance level also defines how far the sample mean, \bar{x} , must be from m before we can reject the null hypothesis. The most common value of for a significance level is $\alpha = 0.05$, which means that we are conducting a hypothesis test where there is a 5% chance that we will incorrectly reject the null hypothesis (if there is a significance level of $\alpha = 0.05$, then you could say that the test has as $(1 - 0.05) = 0.95$, or 95% confidence level). The lower the value of α , the less likely the test is to incorrectly reject the null hypothesis. We use α to determine the rejection region in the normal distribution, which can be seen in Figure 12.1.

Figure 12.1: Two-Tailed T-Test Rejection Region



The rejection region is the percentage of the area under the curve of the normal distribution in which the mean of the sample, \bar{x} will fall if we reject the null hypothesis. For a two-tailed test, the rejection region lies in both tails of the distribution, where each tail has $\alpha/2$ percent of the total area. If the mean of the sample lies in the rejection region, then that means that there is enough evidence to reject the null hypothesis and accept the alternative hypothesis.

12.1.1.2 One-Tailed T-Test

The one-tailed t-test is similar to a two-tailed t-test, except that instead of testing to determine if the sample mean, \bar{x} , is equal to m , we are testing to determine if the sample mean is greater than or less than m . For this one-tailed test, the null and alternative hypotheses are stated as:

Null hypothesis: $H_0 : \bar{x} \leq m$

Alternative hypothesis: $H_1 : \bar{x} > m$

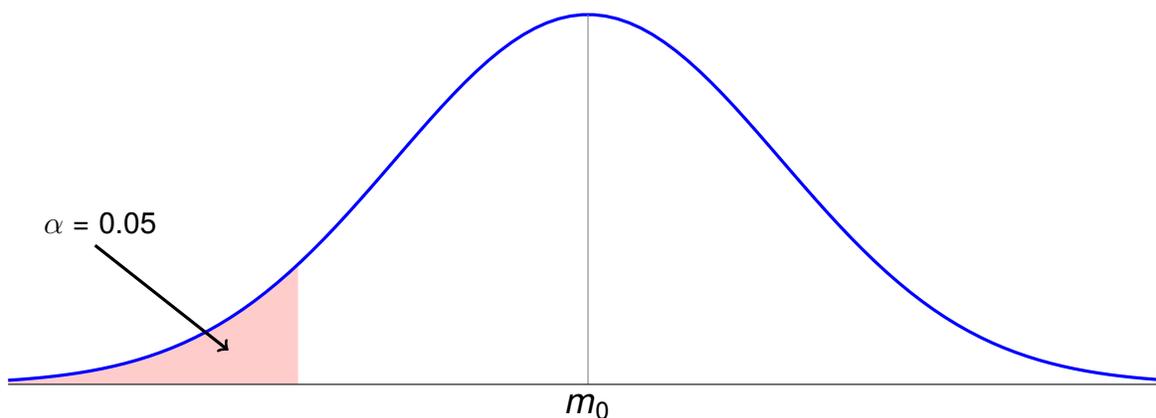
Or,

Null hypothesis: $H_0 : \bar{x} \geq m$

Alternative hypothesis: $H_1 : \bar{x} < m$

For a one-tailed test, the rejection region lies only on one tail of the normal distribution, accounting for α percent of the area under the curve. If the mean, \bar{x} , lies in the rejection region, then there is enough evidence to reject the null hypothesis and accept the alternative hypothesis.

Figure 12.2: One-Tailed T-Test Rejection Region



Steps to Conduct a One-Sample T-Test

- Step 1: Determine the value of comparison, m
- Step 2: State the null hypothesis, H_0
 - $H_0: \bar{x} = m$ or $H_0: \bar{x} \geq m$ or $H_0: \bar{x} \leq m$
- Step 3: State the alternative hypothesis, H_1

- $H_1: \bar{x} \neq m$ or $H_1: \bar{x} > m$ or $H_1: \bar{x} < m$
- Step 4: Determine the significance level, α
- Step 5: Determine the degrees of freedom
 - Degrees of freedom is an estimate of the number of independent pieces of information used to calculate the test
 - for the T-Test, degrees of freedom = $n - 1$
- Step 6: Calculate the T-Statistic
 - $t = \frac{\bar{x} - m}{s/\sqrt{n}} = \bar{x} - m \left(\frac{\sqrt{n}}{s} \right)$
 - \bar{x} is the mean of the sample
 - n is the number of observations/data points
 - s is the standard deviation of the sample
- Step 7: Determine the critical value from the table
- Step 8: Compare the T-Statistic to the critical value
 - If $|\text{T-Statistic}| > \text{critical value}$, then there is enough evidence to reject the null hypothesis and accept the alternative hypothesis
 - If $|\text{T-Statistic}| \leq \text{critical value}$, then we do not have enough evidence to reject the null hypothesis
 - We cannot claim that the null hypothesis is true, we did not test for that

Example

In the population, the average IQ is 100. A team of scientists is testing a new medication to see what kind of effect it has on intelligence. A sample of 30 people participate in a trial where the mean IQ is 140 with a standard deviation of 20. Did the medication of a significant effect on IQ with a 95% confidence level ($\alpha = 0.05$)?

- From the problem, we know that $\bar{x} = 140$, $n = 30$, and $\sigma = 20$
- Step 1: We are comparing the sample IQ to the national average, so $m = 100$
- Step 2: $H_0 : \bar{x} = 100$
- Step 3: $H_1 : \bar{x} \neq 100$
- Step 4: $\alpha = 0.05$
- Step 5: $n - 1 = 30 - 1 = 29$

- Step 6: $t = (140 - 100) \left(\frac{\sqrt{30}}{20} \right) = 10.95$
- Step 7: We are performing a two-tailed test, $\alpha = 0.05$, and degrees of freedom = 29. From the table, the critical value = 2.045
- Step 8: $|10.95| > 2.045$. Since $|t| >$ critical value, there is enough evidence to reject H_0 and say that the IQ of the sample is statistically different from the IQ of the national average.

Example2

The mean body weight of men in the United States is 195lb. A survey of 50 male cadets at VMI found that the average weight was 170lb with a standard deviation of 10lb. Is the weight of a male VMI cadet less than the population average? Assume $\alpha = 0.05$

- From the problem, we know that $\bar{x} = 170$, $n = 50$, and $s = 10$
- Step 1: $m = 195$
- Step 2: $H_0 : \bar{x} \leq 195$
- Step 3: $H_1 : \bar{x} > 100$
- Step 4: $\alpha = 0.05$
- Step 5: $n - 1 = 50 - 1 = 49$
- Step 6: $t = (170 - 195) \left(\frac{\sqrt{50}}{10} \right) = -17.68$
- Step 7: We are performing a one-tailed test $\alpha = 0.05$, and degrees of freedom = 49. From the table, the critical value = 1.684
- Step 8: $|17.68| > 1.684$. Since $|t| >$ critical value, there is enough evidence to reject H_0 and say that the mean weight of male cadets at VMI is statistically less than the population.

12.1.2 Two-Sample T-Test: Independent Sampling

If we want to compare the means of two samples that have been collected, we need to use a two-sample t-test. The samples are considered independent if they are randomly selected such that the observations of one group do not depend on the observations of the other group. Many studies have independent samples. Examples of groups that you may want to compare means are men and women, young and old, Republicans and Democrats, different races, different states, etc.

The steps for conducting a two-sample test with independent samples are the same as a

one sample test with small differences in how the null and alternative hypotheses are stated, the test statistic, and degrees of freedom.

Steps to Conduct a Two-Sample T-Test with Independent Samples

- Null hypothesis
 - H_0 : mean group 1 = mean group 2
- Alternative hypothesis
 - H_1 : mean group 1 \neq mean group 2
- Degrees of freedom
 - $n_1 + n_2 - 2$
- For a two-sample test, there will be two different samples, each sample will have its own mean, standard deviation, and number of observations. The t-statistic can be calculated by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{Z \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

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

- Where,
 - \bar{x}_1, \bar{x}_2 = are the means for each group.
 - s_1^2, s_2^2 = are the variances for each group.
 - n_1, n_2 = are the number of observations for each group.
- The Z term is what is called the pooled standard deviation
- Once you calculate the t-statistic, compare it to the critical value to determine whether you should reject the null hypothesis.

12.2 The Z-Test

The Z-Test is similar to the T-Test, but it is used when the population standard deviation is known and the sample size is greater than 30.

The steps for a one-sample Z-Test are as follows:

Steps to Conduct a One-Sample Z-Test

- Step 1: Determine the value of comparison, m
- Step 2: State the null hypothesis, H_0
 - $H_0: \bar{x} = m$ or $H_0: \bar{x} \geq m$ or $H_0: \bar{x} \leq m$
- Step 3: State the alternative hypothesis, H_1
 - $H_1: \bar{x} \neq m$ or $H_1: \bar{x} > m$ or $H_1: \bar{x} < m$
- Step 4: Determine the significance level, α
- Step 5: Calculate the Z-Statistic

$$\circ z = \frac{\bar{x} - m}{\sigma/\sqrt{n}} = \bar{x} - m \left(\frac{\sqrt{n}}{\sigma} \right)$$

\bar{x} is the mean of the sample

n is the number of observations/data points

s is the standard deviation of the sample

- Step 6: Find the appropriate critical value from the table:

| Confidence Level | Two-Tailed Test | One-Tailed Test |
|------------------|-----------------|-----------------|
| 0.90 | 1.64 | 1.28 |
| 0.95 | 1.96 | 1.65 |
| 0.99 | 2.58 | 2.33 |

Step 8: Compare the Z-Statistic to the critical value

- If $|Z\text{-Statistic}| > \text{critical value}$, then there is enough evidence to reject the null hypothesis and accept the alternative hypothesis
- If $|Z\text{-Statistic}| \leq \text{critical value}$, then we do not have enough evidence to reject the null hypothesis
 - We cannot claim that the null hypothesis is true, we did not test for that

12.3 P-Values

A p-value (probability value) is the probability that the null hypothesis is correct. P-values are expressed as a decimal between 0 and 1. Once the p-value has been calculated, we compare it to α . If the p-value $\leq \alpha$, then that means that there is enough evidence to reject the null

hypothesis.

P-values are typically calculated on a calculator or computer using the t-statistic but can be approximated manually using the t-table. To approximate a p-value using the t-table, find where the calculated t-statistic would be on the row for degrees of freedom. Find the two values of α the t-statistic falls between. The average of these two values is an approximation of the p-value.

12.4 Type I and Type II Errors

Because we are using a sample of data from a larger population, there is the possibility that we come to the wrong conclusion. When doing hypothesis tests, there are two possible errors we can make: Type I and Type II errors.

12.4.1 Type I Error

A Type I Error is where we reject the null hypothesis and support the alternative hypothesis when we should have failed to reject the null hypothesis. The significance level, α , is how likely a Type I Error will occur. As α increases, the probability of committing a Type I Error increases.

12.4.2 Type II Error

A Type II Error is where we fail to reject the null hypothesis, but the alternative hypothesis is actually true. As α decreases, the probability of committing a Type II Error increases. To balance between the two, the most common alpha is 0.05, but sometimes 0.10 and 0.01 are also used. If $\alpha = .05$, then there is a 5% chance of a Type I Error, or we are 95% confident that we are right.

12.5 Confidence Intervals

A confidence interval is the probability that a population parameter will fall between certain values. A confidence interval measures the degree of uncertainty in a sampling method. Confidence intervals are most often constructed with 95% or 99% confidence levels. In general, a confidence interval is expressed as follows:

$$\text{mean} \pm \text{margin of error} \quad (12.1)$$

The level of confidence can be expressed using α . $\alpha = 1 - \text{confidence level}$. The most common confidence level is 95%, or $\alpha = 0.05$. However, there are instances where a confidence level of 90% ($\alpha = 0.10$) or 99% ($\alpha = 0.01$) may be appropriate.

12.5.1 Confidence Interval when Population Standard Deviation is Known

If we know the population standard deviation, we can calculate a confidence interval of the sample as follows:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (12.2)$$

Where,

\bar{x} is the sample mean.

$z_{\alpha/2}$ is the z-value associated with a probability of $\frac{\alpha}{2}$.

σ is the standard deviation of the population.

n is the number of observations.

The z-value can be obtained from the z-table, but the following table can be useful:

Table 12.1: Common Z-Values

| α | z-value |
|----------|---------|
| 0.90 | 1.645 |
| 0.95 | 1.96 |
| 0.99 | 2.576 |

Example

A sample of 15 bottles of soda has a mean volume of 20oz. per bottle. Construct a 95% confidence interval for the mean volume of a bottle of soda if the population standard deviation is 0.05oz.

$$20 \pm 1.96 \cdot \frac{0.05}{\sqrt{15}}$$

$$20 \pm 1.96 \cdot .0129$$

$$20 \pm 0.253$$

This means we are 95% confident the population mean lies between 19.747oz. and 20.253oz.

12.5.2 Confidence Interval when Population Standard Deviation is Unknown

In most instances, we will not know the population standard deviation, σ , we will have the sample standard deviation, s . When this is the case, we cannot rely on z-values to calculate our confidence interval. Instead, we must just the t-statistic which comes from the t-distribution. The t-distribution is similar to the normal distribution in that it is bell-shaped and symmetrical around the mean. To calculate a confidence interval when σ is unknown, we use the following formula:

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (12.3)$$

Where,

\bar{x} is the sample mean.

$t_{\alpha/2}$ is the critical value of a two-tailed test from the t-table with confidence level α .

s is the standard deviation of the sample.

n is the number of observations.

We can find the critical value using the t-table for a given confidence level and degrees of freedom. Recall, that degrees of freedom is $n - 1$.

Example

Suppose you have a sample of 100 people who completed a survey where 55% said they support a specific presidential candidate. If the sample standard deviation is 15%, construct a 95% confidence interval.

First, we need to look up the critical value using the t-table for degrees of freedom = 99. For this problem, the critical value is 1.984.

$$55 \pm 1.984 \cdot \frac{15}{\sqrt{100}}$$
$$55 \pm 1.984 \cdot 1.5$$

$$55 \pm 2.976$$

This means we are 95% confident the population mean lies between 52.024% and 57.976%.

Part IV

Introduction to Finance

THE TIME VALUE OF MONEY

There is a time component to money. If you have \$100 and are at the bar where you can buy drinks and food today and someone says you should hold your \$100 and save it for the summer where you can use it for something else. You likely would say no, you are going to spend it now because that is what is most fun and “valuable” to you in the present. Economics teaches us that there is an “opportunity cost” associated with waiting to spend your money → If you wait to spend it, you are forgoing current utility (“satisfaction”).

That \$100 is worth more to you today than it would be in the future. If I wanted you to hold your money for the future, I would need to compensate you for the opportunity cost or lost utility. I could offer you the opportunity to put your money in an investment that earns interest with the promise that you will have more money in the future than you do today.

We cannot compare dollar values from different time periods. We either need to convert dollar values into terms of the present or at some point in the future. Present Value (PV) is stating a future dollar value in today’s terms when accounting for the interest you could earn. Future Value (FV) is stating today’s dollar value in terms of a future date when accounting for the interest you could earn.

13.1 Future Value

If we were to state dollars today in terms of a future period, we call this future value. The easiest way to think about future value is if you put money into an account and let it accumulate interest.

13.1.1 Simple Interest

Simple interest is where the interest payment is calculated on the *PV* and then applied for each period and does not consider that the amount in the account changes as a result of the interest. The interest payment can be calculated by:

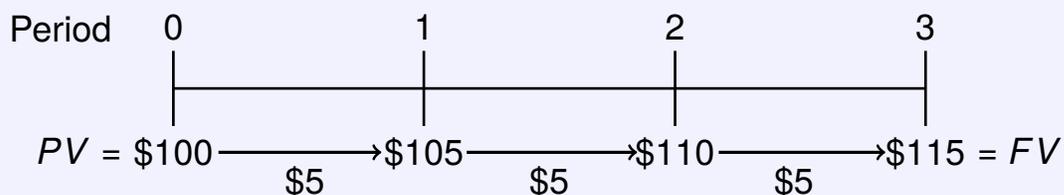
$$\text{Interest Payment} = PV \times \text{interest rate} \quad (13.1)$$

Example

Suppose you put \$100 in a bank account today that will earn 5% interest annually, how much money will you have after three years?

First, we need to calculate the interest payment:

$$\text{Interest} = (\$100)(.05) = \$5$$



We can see that the total simple interest payment is \$5 made 3 times. Therefore:

$$\text{Total : Interest} = (\$100)(.05)(3) = \$15$$

Using the previous example, we can construct a generalized equation to calculate the simple interest payment:

$$\text{Simple Interest} = PV_0 \times i \times n \quad (13.2)$$

Where,

PV_0 = the present value at time $t = 0$

i = the interest rate

n = the number of years

Using equation Equation 13.2, we can construct an equation for future value with simple interest:

$$FV_n = PV_0 + \text{Simple Interest} \quad (13.3)$$

$$FV_n = PV_0 + PV_0in \quad (13.4)$$

$$FV_n = PV_0(1 + in) \quad (13.5)$$

Using Equation 13.5, we can see that we get the solution from our example:

$$FV_3 = \$100[1 + (.05)(3)] = \$115$$

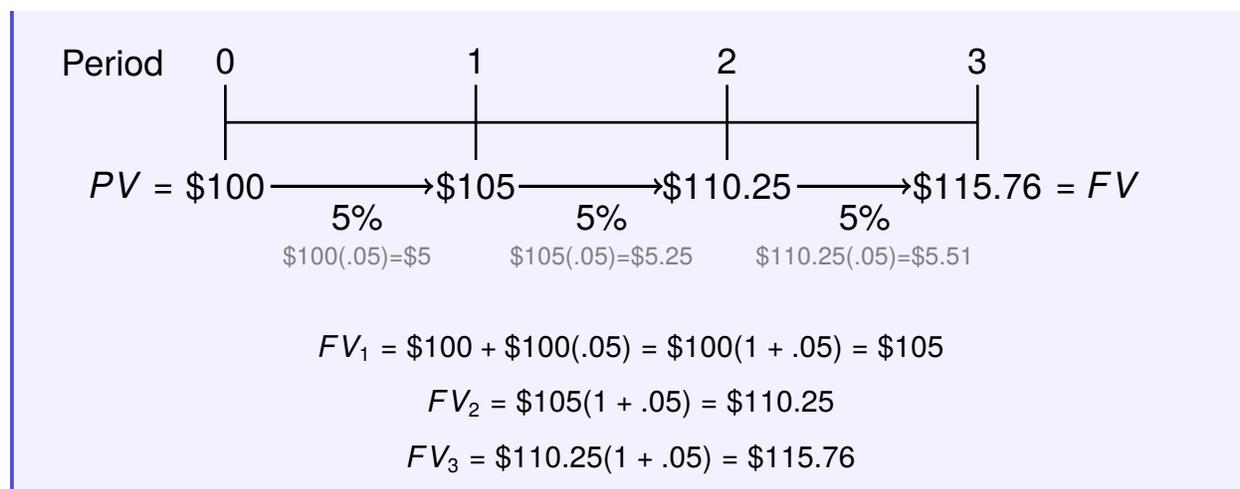
The most common application of simple interest is in auto loans.

13.1.2 Compound Interest

Compound interest is when the interest payment is updated to consider that the money stored in the account changes each period.

Example

Suppose you put \$100 in a bank account today that will earn 5% interest compounded annually, how much money will you have after three years?



From the example, we can see that:

$$FV_1 = PV_0(1 + i) \quad (13.6)$$

$$FV_2 = FV_1(1 + i) \quad (13.7)$$

$$FV_3 = FV_2(1 + i) \quad (13.8)$$

If we substitute Equation 13.6 into Equation 13.7, we get:

$$FV_2 = FV_1(1 + i)(1 + i) \quad (13.9)$$

Then if we substitute Equation 13.9 into Equation 13.8, we get:

$$FV_3 = PV_0(1 + i)(1 + i)(1 + i) \quad (13.10)$$

From this, we get a nice Equation that we can use to calculate Future Value when interest is compounded:

$$FV_n = PV_0(1 + i)^n \quad (13.11)$$

If we apply Equation 13.11 to our example, we get the same solution:

$$FV_3 = \$100(1 + .05)^3 = \$115.76$$

13.2 Present Value

Now that we have established how to calculate the future value of a present dollar amount, what if we are discussing a dollar amount in the future? Suppose that you are asked the following question: Would you prefer \$500 in 4 years or \$400 today? We cannot directly compare money in the future with money today. So we must find the present value of the future dollar amount. This is called discounting a future value into today's terms. We can obtain the present value by rearranging Equation 13.11:

$$PV = \frac{FV_n}{(1 + i)^n} \quad (13.12)$$

Example 1

If we want to compare \$500 in 4 years to \$400 today, need to find the present value of \$500. Assume that the interest rate is 5% compounded annually.

$$PV = \frac{\$500}{(1 + .05)^4} = \$411.35$$

This means that \$500 in 4 years is worth \$411.35 today, so that means that we would be better off taking the \$411.35 in 4 years than \$400 today. Why? Because if we were to take the \$400 and invest it at 5% compounded annually for 4 years, we would only have:

$$FV_4 = \$400(1.05)^4 = \$486.20$$

Example 2

Suppose you want to buy a car in 3 years that cost \$12,000. If you had an account that paid 7% interest compounded annually, how much would you need to put into that account today to have \$12,000 in 3 years?

$$PV = \frac{\$12,000}{(1 + .07)^3} = \$9,795.57$$

This means that if you put \$9,795.57 in an account today at 7% interest compounded annually, you would have \$12,000 in 3 years.

13.3 Annuities

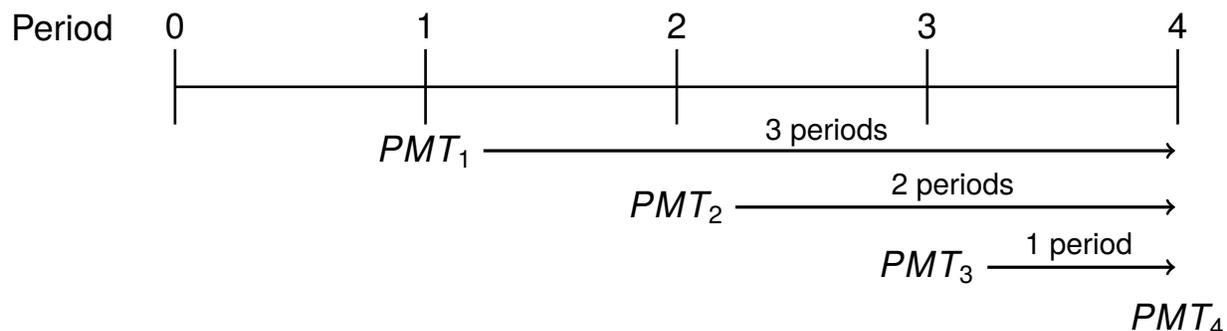
Present value and future value focus on one-time lump-sum payments. While these are interesting problems and sometimes encountered in the real world, it is more common where payments are made over time while collecting compound interest. If the payments made are identical, equal dollar amounts, for each period, this is called an annuity. An ordinary annuity is an annuity in which the payment is made at the end of each period¹. Examples of ordinary annuities are mortgage payments and student loan payments.

13.3.1 Future Value of an Ordinary Annuity

With an ordinary annuity, a deposit is made each period into the interest-bearing account. Each deposit will collect interest that is compounded. This can be seen in Figure 13.1.

¹ Another type of annuity is called an annuity due. These are annuities where payments are made at the beginning of the period. Examples of annuities due are insurance payments and rent payments. We will not discuss annuities due in this book, but you will encounter them in a finance class.

Figure 13.1: Timeline of an Ordinary Annuity



We can see that PMT_1 collects interest for 3 periods:

$$FV = PMT_1(1 + i)^3 \quad (13.13)$$

PMT_2 collects interest for 2 periods:

$$FV = PMT_2(1 + i)^2 \quad (13.14)$$

PMT_3 collects interest for 1 period:

$$FV = PMT_3(1 + i) \quad (13.15)$$

PMT_4 collects interest for 0 periods:

$$FV = PMT_4 \quad (13.16)$$

To find the FV of the annuity (or the value of the annuity after four years), we sum the FV's of each payment:

$$FVA_{ord} = PMT(1 + i)^3 + PMT(1 + i)^2 + PMT(1 + i) + PMT \quad (13.17)$$

This equation can be generalized to n periods:

$$FVA_{ord} = PMT + PMT(1 + i) + PMT(1 + i)^2 + PMT(1 + i)^3 + \dots + PMT(1 + i)^n \quad (13.18)$$

Fortunately for us, Equation 13.18 can be expressed as a simplified equation to calculate that can calculate FV of an ordinary annuity of any length n :

$$FVA_{ord} = \frac{PMT}{i} [(1 + i)^n - 1] \quad (13.19)$$

13.3.2 Present Value of an Ordinary Annuity

If we want to compare the value of an annuity with another investment, we need to convert it to the present value. The present value of an ordinary annuity can be calculated by:

$$PVA_{ord} = \frac{PMT}{1 + i} + \frac{PMT}{(1 + i)^2} + \frac{PMT}{(1 + i)^3} + \dots + \frac{PMT}{(1 + i)^n} \quad (13.20)$$

Which simplifies to:

$$PVA_{ord} = \frac{PMT}{i} \left[1 - \frac{1}{(1+i)^n} \right] \quad (13.21)$$

Example

Suppose that you win the Powerball lottery worth \$500,000,000. The \$500,000,000 payment assumes that you will take payments over time. Typically you will be offered the option to take the winnings as an annuity (say \$7,525,717 per year for 30 years, if $i = .05$), where you will receive the full value, or as a lump sum payment where you will receive a smaller amount. The lump sum (say it is \$110,500,000) is money you will get today, so it is expressed in the present value. The \$500,000,000 you will get through an annuity is expressed as a future value. If we want to compare the two, we will need to discount the value of the annuity back to today (e.g., calculate the present value of the annuity).

- Lump sum payment: $PV = \$110,500,000$

- Annuity: $PVA_{ord} = \frac{7,525,717}{.05} \left[1 - \frac{1}{(1.05)^{30}} \right] = \$115,688,716$

- This means that the annuity is worth more than the lump sum

13.4 Amortized Loans

When you borrow money (e.g., student loans, mortgages, business loans, etc.), you pay it back in installments over time. We can use the concept of annuities to understand how your monthly payment is calculated. When you borrow money, the initial amount you borrow is called the principal balance. As you make regular loan payments, you pay the principal and pay interest. Interest is usually expressed as an Annual Percentage Rate (APR). To find out what the regular payment is, we use the PVA_{ord} formula where the PVA = Principal, n = number of years, and i = the APR. Then we solve for PMT . We can identify how much of each regular payment goes towards the principal and interest by constructing an amortization schedule.

Example

Suppose you want to borrow \$15,000 to build an addition on your house. The loan is going to be repaid in three equal payments at the end of the next three years. The bank is offering you an APR of 8%.

To find the payment:

$$15,000 = \frac{PMT}{.08} \left[1 - \frac{1}{(1.08)^3} \right]$$

$$PMT = \frac{15,000}{2.5771} = \$5,820.50$$

Once we have the payment, we can construct the amortization schedule. For each payment, the interest gets paid first, then the remaining amount goes towards the principal.

| End of Year | Payment | Interest | Principal Repayment | Ending Balance |
|-------------|------------|----------------------------|-----------------------------------|--------------------------------------|
| 0 | | | | \$15,000 |
| 1 | \$5,280.50 | 15,000(.08) = 1,200 | 5280.50 – 1,200 = \$4,620.50 | 15,000 – 4,620.50 = \$10,379.50 |
| 2 | \$5820.50 | 10,379.50(.08) = 830.36 | 5,820.50 – 830.36 = \$4,990.14 | 10,379.50 – 4,990.14 = \$5,389.36 |
| 3 | \$5,820.40 | 5,389.36(0.8) = 431.15 | 5,820.50 – 431.15 = 5,389.36 | 5,389.36 – 5,389.36 = 0.00 |