DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
DEPARTMENT OF ECONOIMCS
COLORADO STATE UNIVERSITY
MATH CAMP

FALL, 2011

One of the best things you can do as a new graduate student in Agricultural and Resource Economics or Economics is to have a very solid understanding of basic single variable calculus (and its relation to basic economic concepts that you already know) and basic statistics.

In fact, in many ways, the primary difference between a graduate and undergraduate-level understanding of economic theory is simply a difference in the *language* we use to help us tell economic stories. Most likely, your undergraduate training resulted in a familiarity with both graphical and verbal analysis. For example, you likely used indifference curves to represent preferences in a two-good model, then added a budget constraint, and showed that the optimal bundle of consumption occurred where the marginal rate of substitution equaled the price ratio of the two goods. You could then change prices and trace out a demand curve, and put it together with a supply curve to determine market outcomes.

In graduate school, we use all of the same economic concepts….marginal rates of substitution, budget constraints, supply and demand, etc… However, in order to tell more complicated stories, we need to move beyond simple two-dimensional cases which we can graph. Enter mathematics, especially the tools of calculus. Think about it…we're always going on about marginal this and marginal that in economics, right? Well, marginal means incremental, and the partial derivative of a differentiable function gives us the incremental change in a function if one of its arguments changes. Although this is just the tip of the iceberg, it illustrates the potential power of the mathematical language to allow economics to represent and tell stories about all sorts of economic behavior. And, in essence, that's what we do as economists.

As such, the materials contained herein provide a review of the mathematical concepts that will likely prove to be exceedingly useful during your graduate career (and especially useful for ECON 501 and AREC/ECON 535/635). This includes material covered in Chapters 2-5 and parts of 17-18 in Simon, C.P. and L. Blume. 1994. *Mathematics for Economists*. New York: W.W. Norton & Co. If you haven't already, buy this book. Per student request, there is also a chapter briefly reviewing some statistical concepts useful in econometrics, taken from Dr. Bond's AREC.ECON 335 class. We suggest that you work through the notes and book chapters at your own pace, treating this material as supplemental reference as you begin your graduate career.

If you feel like you'd like to get a head start on many of the other mathematical techniques you will likely encounter throughout your program, please work through Lectures 2, 4, 5, and 6 in the expanded "A Review of Mathematics used in Agricultural and Resource Economics", available in the Documents box at: dare.colostate.edu/ grad/info/continuing.aspx.

2011 In-Class Math Camp Schedule

Instructor: Mr. Anthony Underwood, Department of Economics
Room: C 238 Clark
Text: Simon, C.P. and L. Blume. 1994. *Mathematics for Economists*. New York: W.W. Norton & Co.

Schedule: Monday, August 15, 1-3PM
Wednesday, August 17, 1-3PM
Thursday, August 18, 1-3 PM
Friday, August 19, 9-11AM & 1-3PM

Additional References in the Library

Binger, B.R. and E. Hoffman, 1988. *Microeconomics with Calculus*. Glenview, IL: Scott, Foresman.

Chiang, A.C., 1967. *Fundamental Methods of Mathematical Economics*. New York: McGraw Hill.

Hill, R.C., W.E. Griffiths, and G.C. Lin. 2008. *Principles of Econometrics, 3rd Ed.* John Wiley and Sons, or equivalent text.

Section I: Univariate Calculus and Economics
References: Simon and Blume, Ch 1-5

INTRODUCTION

- In economics, we study the allocation of scarce resources, and this scarcity leads to tradeoffs in allocation decisions
- Mathematics is a *language* that we use to describe these tradeoffs
- In our most basic problems, the mathematics allows us to represent:
    - Consumer preferences and choice on the demand side (through utility functions)
    - Technology, costs, and profits on the production side (through production, cost, and profit functions)
    - Behavior (through maximization/minimization)
- Mathematics is extremely powerful in that it is:
    - Logical
    - Concise
    - Formal

  and can be used to tell economic stories about how economic agents are predicted behave under a set of assumptions
- The mathematical formalism in economics can be traced back to the 19th century "Marginal Revolution", with significant contributions by Leon Walras, Augustin Cournot, W. Stanley Jevons, and Alfred Marshall.
- All of the graphical analysis you have seen in your undergraduate economics education (demand, supply, indifference curves, marginal benefits/costs, OLS, etc…) is developed from formal mathematical analysis
- Just like the progression from Principles of Micro to Intermediate Micro (where you "got behind the scenes" of the main concepts like demand and supply), the next step in Graduate School is to progress into a deeper, more formal (read: mathematical) understanding of the relationships and decisions of economic agents and the assumptions behind how we model them.

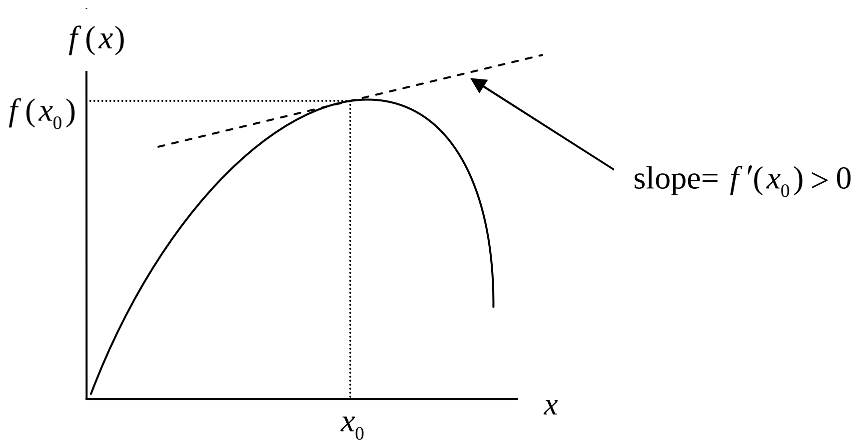
CANONICAL FUNCTIONAL FORMS

- A function is just like a "sausage machine"... you feed it one or more inputs (the arguments of the function, also called *exogenous* variables), and it returns one or more outputs (the value of the output once you feed in the inputs, or *endogenous* variables)
- A univariate function that maps from one (continuous) input into an output maps from the real line to the real line ($\Re^1 \to \Re^1$).
    - Canonical example: $y = f(x)$, where y is the dependent/endogenous variable and x is the independent/exogenous variable.
    - Specific example: $y = x^2$, where $f(x) = x^2$.
    - Economic examples:
        - A utility function which maps the consumption of a good x to an ordinal measure of preferences: $v = u(x)$, where if $u(x_1) > u(x_2)$, then x_1 is preferred to x_2.
        - A production function which maps the use of an input z into an output q: $q = f(z)$.

- Why use canonical forms?
  - Because they apply to a whole *class* of functions. In other words, we are representing *all* of the specific functions that admit the properties of the canonical function.

FIRST PARTIAL DERIVATIVES

- Recall that partial derivatives of a function approximate the change in the value of a function when the input changes by a very small amount.
- In other words, it is the *slope* of a function at the evaluation point, or the line just tangent to the function at a given evaluation point
- Formally, this derivative (if it exists) is defined by $\lim_{h_n \to 0} \frac{f(x_0 + h_n) - f(x_0)}{h_n}$.
- We want to know how to *interpret* a partial derivative and how to use them to *graph* economic relationships
- NOTATION: Let the partial derivative of a continuous, differentiable function $f(x)$ be denoted by as follows: $\frac{\partial f(x)}{\partial x} = f'(x) = f_x(x)$.
  - Specific example: $f(x) = x^2, \quad f'(x) = 2x$.



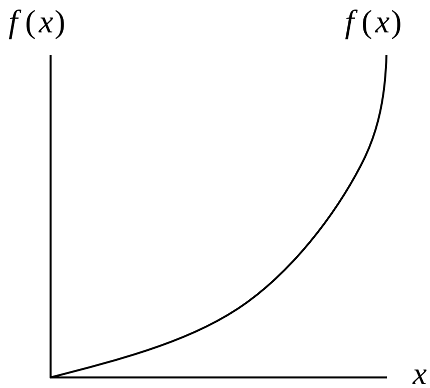
- In economics, the partial derivative plays a key role…it is the "*marginal*", or equivalently "*incremental*", effect of a change in the exogenous variable on the endogenous variable
  - Utility example: $u'(x)$ is how our utility changes with increases in the consumption of good x. This is called MARGINAL UTILITY. We would typically assume $u'(x) > 0$ for a "good", and $u'(x) < 0$ for a "bad", like pollution.

- Production example: $f'(z)$ tells us how production changes with increases in input z. This is called MARGINAL PRODUCT. We would typically assume $f'(z) \geq 0$.
- Note that in economic theory, it is often the case that a sign on a derivative is much more important than a magnitude

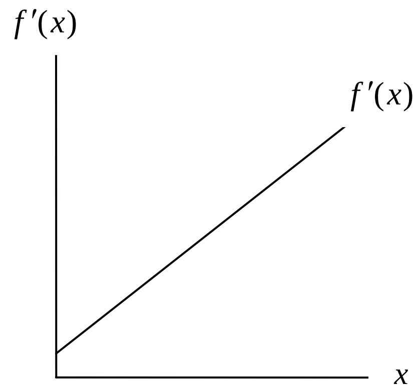
- PARTIAL DERIVATIVE = SLOPE OF A CURVE = "MARGINAL" IN ECONOMICS

SECOND PARTIAL DERIVATIVES AND GRAPHING

- In economics, we are usually concerned not only with first partial derivatives (see above), but second partial derivatives (the derivative of the first derivative).
- NOTATION: $\frac{\partial^2 f(x)}{\partial x^2} = f''(x) = f_{xx}(x)$.
- By natural extension, the second derivative is the "slope of the slope", or the slope of the marginal curve
- Economic examples:
  - $u''(x) < 0 \rightarrow$ decreasing marginal utility
  - $f''(z) < 0 \rightarrow$ decreasing marginal product
- As such, we can use the first and second derivatives of functions to provide a general graphical illustration of the "total" and "marginal" curves
  - The first partial derivative tells us the direction of the slope of the "total" curve and the level of the "marginal" curve
  - The second partial derivative tells us how the slope of the "total" curve changes and the slope of the "marginal" curve
- Example: Assume $f(x) \geq 0, f(0) = 0, f'(x) > 0, f''(x) > 0, \lim_{x\to\infty} f'(x) = \infty$.
  - The first assumption tells us that the value of the total curve $f(x)$ is always non-negative. The second assumption tells us that the value of $f(x)$ at $x = 0$ is zero.
  - The first partial derivative tells us that the total curve $f(x)$ is increasing in x and that the marginal curve $f'(x)$ is positive for all values of x.
  - The second partial derivative tells us that the slope of $f(x)$ is increasing with increases in x, and that the marginal curve $f'(x)$ has a positive slope.
  - The limit expression tells us that as x gets really large, the function $f(x)$ has an infinite slope, and the marginal curve $f'(x)$ has no limit.



The “Total” Curve

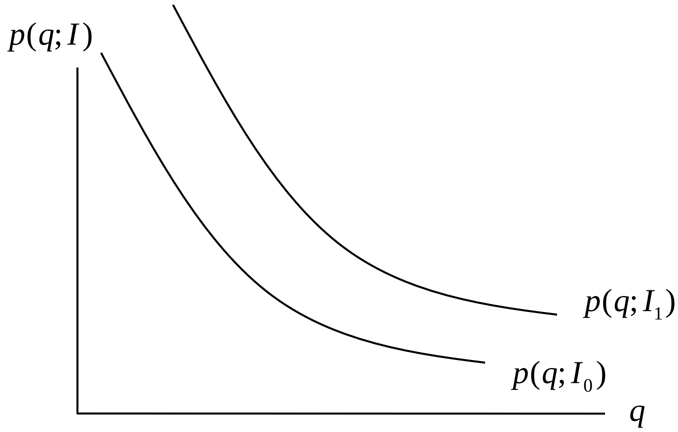


The “Marginal” Curve

- Note that without the third derivative, we cannot exactly represent how the slope of the marginal curve changes with changes in x.

A DERIVATIVE APPLICATION: ELASTICITIES

- Suppose we had a *demand function* $q(p;I)$, where q is quantity demanded of a certain good, p is the price, and I is a parameter representing income. Note that if we “inverted” this function to get $p(q;I)$, assuming we could, this is called the “*inverse demand function*”.
  - Specific example: $q = \alpha + \beta p + \gamma I$.
- From a derivative standpoint, the law of demand suggests $q'(p;I) < 0$, or that as price increases, quantity demanded goes down. If this good was a normal good, this implies $\frac{\partial q(p;I)}{\partial I} > 0$, or demand increases as income increases. If it was an inferior good, then $\frac{\partial q(p;I)}{\partial I} < 0$.
  - For our specific example, this implies $\beta < 0$, $\gamma > 0$ for a normal good, and $\gamma < 0$ for an inferior good.



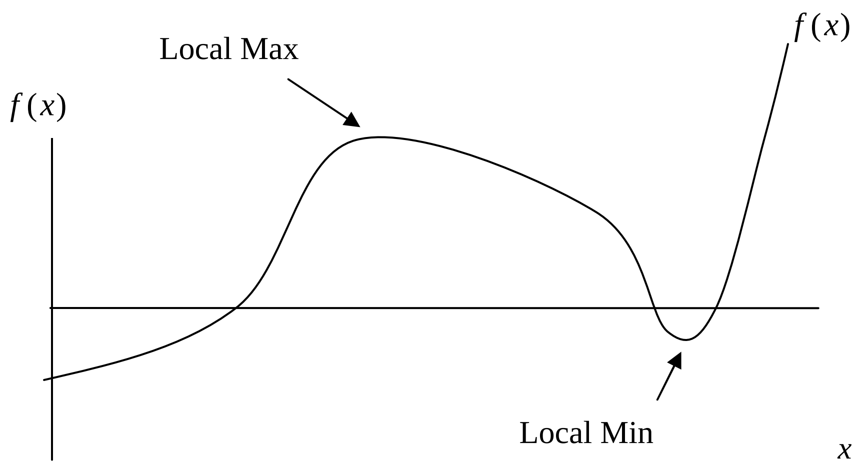
If $I_1 > I_0$, this is a normal good.

- Recall that economists like to summarize response to changes in a unit-free way…hence the idea of elasticities.
- Elasticities are defined by a formula such as: $\frac{\%\Delta Y}{\%\Delta X} = \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}$.
- Note that if we make the ΔX very small, then $\frac{\Delta Y}{\Delta X} \approx \frac{\partial Y}{\partial X}$.
- As such, our own-price and income elasticities of demand become:
  - Own Price: $\frac{p \cdot q'\left(p;I\right)}{q\left(p;I\right)} < 0$,
  - Income: $\frac{I}{q\left(p;I\right)} \frac{\partial q\left(p;I\right)}{\partial I} \underset{<}{\overset{>}{-}} 0$.

GLOBAL AND LOCAL MAXIMA AND MINIMA

- Mathematical models of economic phenomena often assume some sort of behavior on behalf of an agent; namely, maximization (of, say, utility) or minimization (of, say, total cost).
- Over a fixed interval (a closed set) of the range of x, a function will always have a (possibly not unique) boundary maximum and boundary minimum.
  - Example: On [0,1], the function $f(x) = 3x$ as a boundary maximum of $f(1) = 3$ and a boundary minimum of $f(0) = 0$.
- A maximum or minimum that occurs off of the endpoints of the domain is called an interior maximum or interior minimum.
- A critical point x_0 of the differentiable function $f(x)$ occurs when $f'(x_0) = 0$; in other words, at a point where the slope of the curve is equal to zero.
  - If $f'(x_0) = 0$ and $f''(x_0) < 0$, then x_0 is a max of $f(\cdot)$
  - If $f'(x_0) = 0$ and $f''(x_0) > 0$, then x_0 is a min of $f(\cdot)$

- If $f'(x_0)=0$ and $f''(x_0)=0,$ then x_0 can be a max, min, or neither of $f(\cdot)$



- A global maximum or minimum occurs at a point x_0 over the entire range of x when $f(x_0)$ is either greatest or smallest.
- The following situations allow us to find global maxima and minima easily:
  - If the domain of $f(\cdot)$ is an interval *I* (finite or infinite) in $\Re^1$, x_0 is a local maximum of $f(\cdot)$, and x_0 is the only critical point of $f(\cdot)$ on the interval, then x_0 is a global maximum of $f(\cdot)$ on *I*.
  - If $f(\cdot)$ is a twice-continuously differentiable function whose domain is an interval *I* and if $f''(\cdot)\neq 0\ \forall x\in I,$ then $f(\cdot)$ has at most one critical point in *I*. This critical point x_0, if it exists, is a global minimum if $f''(x_0)<0$ and a global maximum if $f''(x_0)>0$.
- A function need not have a global max or global min if the domain of that function is an open interval.

DERIVATIVE RULES

Suppose that *k* is an arbitrary constant and that $f(x)$ and $g(x)$ are differentiable at the point $x=x_0$.

- If $z(x)=f(x)+g(x),$ then $z'(x_0)=f'(x_0)+g'(x_0)$
- If $z(x)=kf(x),$ then $z'(x_0)=kf'(x_0)$
- Product Rule: If $z(x)=f(x)g(x),$ then $z'(x_0)=f'(x_0)g(x_0)+f(x_0)g'(x_0)$
- Quotient Rule: If $z(x)=\dfrac{f(x)}{g(x)},$ then $z'(x_0)=\dfrac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}$
- Power Rule: If $z(x)=x^k,$ then $z'(x_0)=kx_0^{k-1}$

- EXP: If $z(x) = e^x$, then $z'(x_0) = e^{x_0}$
- LN: If $z(x) = \ln(x)$, then $z'(x_0) = \frac{1}{x_0}$
- EXPONENTIALS: If $z(x) = b^x$, $b > 0$, then $z'(x_0) = \ln(x_0) b^{x_0}$

COMPOSITE FUNCITONS AND THE CHAIN RULE

- Composite functions just take a function of a function: $h(x) = g\left(f(x) \right)$.
  - Note that $h(x)$ is a function of x, and that $g(\cdot)$ has one argument. Our notation says: "take the output from the function $f(x)$ and use it as the argument, or input, into the function $g(\cdot)$.
  - Specific Example: $f(x) = x^{1/2}$ and $g(x) = x + 1$, so $h(x) = g\left(f(x) \right) = x^{1/2} + 1$.
  - Economic example: The regression $y = \alpha + \beta \ln(x) + \varepsilon$, where $f(x) = \ln(x)$. This makes $g(z) = \alpha + \beta z + \varepsilon$.
- The chain rule for partial differentiation of a composite function is "the derivative of the outside function (evaluated at the inside function) times the derivative of the inside function.
  - Canonical: $h'(x) = g'\left(f(x) \right) \cdot f'(x)$.
  - Specific: $h'(x) = g'\left(f(x) \right) f'(x) = (1)\left(\frac{1}{2} x^{-1/2} \right) = \frac{1}{2} x^{-1/2}$.
  - Economic: $\frac{\partial y}{\partial x} = \beta / x$.

PRACTICE PROBLEMS

1. Simon and Blume 2.11, 3.1, 3.4, 3.9, 3.22, 4.1, 4.3
2. Assume a univariate utility function $u(x)$ that exhibits diminishing marginal utility in the good x.
   a. Graph the total and marginal utility functions on separate graphs.
   b. Confirm that the function $v(x) = x^{1/2}$ is a special case of $u(x)$.
   c. Our utility function is "homogeneous of degree one" if and only if $u(tx) = tu(x)$ for all t>0. Is $v(x) = x^{1/2}$ homogeneous of degree one? Explain.
   d. The economic concept of "strong monotonicity" in this context says that if $x_1 > x_2$, then x_1 is preferred to x_2. Does $u(x)$ represent preferences that are strongly monotone? Explain.
3. Consider a total cost function $c(x)$, which represents the total costs of producing output level x. Assume $c'(x) > 0$ and $c''(x) > 0$.
   a. Formally show that the marginal cost curve intersects the average cost curve $c(x)/x$ at the minimum of average cost.

b. Graph marginal costs and average costs on the same graph.
c. Assuming no fixed costs, the marginal cost curve is the supply curve for a competitive firm. Find an expression for the elasticity of supply for a firm with the cost function $c(x)$.

4. Consider the linear demand curve for $q(p\ ;I)$ which follows the law of demand.
   a. Based on these assumptions alone, what mathematical properties can we ascribe to $q'(p\ ;I)$? $q''(p\ ;I)$? $\frac{\partial q(p\ ;I)}{\partial I}$?
   b. Suppose that demand at price p_0 is elastic (i.e., the absolute value of the own-price elasticity of demand is greater than 1). Formally show that an increase in the price will decrease total expenditures.
   c. Suppose this good is normal and a luxury item. What does that tell us about $q'(p\ ;I)$? $q''(p\ ;I)$? $\frac{\partial q(p\ ;I)}{\partial I}$?
5. Consider the function $y(x;\alpha)=\alpha+\beta\ln x$, where α and β are known, fixed constants.
   a. Graph this function if $\beta>0$. Is this more likely to represent a demand or supply curve? Why?
   b. Graph this function if $\beta<0$. Is this more likely to represent a demand or supply curve? Why?
   c. Give an expression for the elasticity of *y* with respect to *x*.
6. A logistic growth function that describes growth of a stock of a resource as a function of the stock is given by $G(x;r,k)=rx\left[1-x/k\right]$, where r and k are positive parameters. Use the derivative properties of this function to graph the growth curve as a function of the stock, and the evolution of the stock over time starting at a relatively low initial level.
7. Formally show that the marginal revenue for a monopolist (who faces a downward sloping demand curve) is less than the price charged for any quantity greater than zero. Document any assumptions you make.
8. Show that $\frac{\partial \ln y}{\partial \ln x}$ gives the elasticity of *y* with respect to *x*.
9. Given $h(x)=g\left(f(x)\right)$, sign $h'(x)$ and $h''(x)$ (if possible) under the following assumptions:
   a. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign
   b. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of different signs and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign
   c. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of different signs and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign
   d. First derivatives of $f(\cdot)$ and $g(\cdot)$ are of the same sign and second derivatives of $f(\cdot)$ and $g(\cdot)$ are of different signs
10. Review Chapter 5 of Simon and Blume and complete problems 5.6, 5.8, 5.9, and 5.14.

Section II: Unconstrained Optimization
References: S&B Ch 17

Introduction:

- Economic theory is largely based on optimization theory. (An optimization problem is a mathematical problem in which one is either maximizing or minimizing the value of some objective function.)
- The two principle agents in micro theory are consumers and firms. At the most basic level, we assume that consumers maximize their utility while firms are assumed to maximize profit.
- Another key economic concept that evokes optimization is that of economic efficiency (e.g. a given objective can be achieved at a minimum cost).
- The two basic types of economic optimization problems are 1) constrained and 2) unconstrained. In unconstrained optimization problems one is trying to find the values of the choice variables that max or min some objective function, $f(x_1, x_2, ...x_n)$. In the next lecture you will learn about constrained optimization in which one is still attempting to max or min the objective function $f(x_1, x_2, ...x_n)$ but subject to one or more constraints on the choice variables (equalities, inequalities).

Generalized Univariate Example (Framework for Profit Max)

- Firm chooses level of output y to max profits.
- $\pi(y) = r(y) - c(y)$ where $\pi(y)$ = profit, $r(y)$ = total revenue, $c(y)$ = total costs
- One specific case is $\pi(y) = p * y - c(y)$
- To solve, find the first order necessary conditions (*FONC*). The FONC is found by taking a derivative of the profit function with respect to the choice variable – in doing so we are performing *marginal analysis*.
- Why necessary? At optimum, we want *FONC* to equal 0, (e.g., Marginal Revenue-Marginal Cost=0), at this point there is no net benefit from producing an additional unit of output. (Figure a)
- *FONC* $\frac{\partial \pi}{\partial y} = \pi'(y) = p - c'(y) = MR - MC$
- MR=MC means that the extra revenue gained from selling one additional unit of output just equals the extra cost of producing one more unit. An equivalent statement is that marginal profits are zero.
- The *FONC* implicitly define the optimal level of output (*y)* as a function of price. This *y** can be substituted into the profit function to get max profits for a given level of p.

- Next we check the second order sufficient conditions *SONC*, the *SONC* < 0
- Why sufficient? Because the slope of the marginal cost curve must be upward sloping (more generally, marginal profit must be decreasing)
- *SONC* $\frac{\partial^2 \pi}{\partial y^2} = \pi''(y) = -c''(y) < 0$, so the slope of the marginal cost curve, $c''(y)$ >0. See Figure a.

Univariate Numeric Example:

- Consider the case of a competitive firm which sells its output for a constant price *p* and whose optimization problem can be described as below. Using the information below, we will find the firm's supply curve and determine how quantity supplied changes with price and how supply changes with changes in marginal cost.
- $\pi(y) = p * y - \frac{1}{2}ky^2$ where $\pi(y)$ = profit, and total costs, $c(y)$ = $\frac{1}{2}ky^2$
- $FONC \frac{\partial \pi}{\partial y} = \pi'(y) = p - ky = 0$ and $\therefore \; p = ky$
- Solve for optimal level of output, $y^* = \frac{p}{k}$

Related Comparative Statics

- We calculate comparative statistics to examine changes in equilibrium response to a change in underlying economic parameters.
- Comparative: because we are comparing one statistic in equilibrium to the new one that would occur if the parameters were to change.
- Static: because we do not describe the dynamic path of how the equilibrium actually moves from one position to another.
- Mathematically, we are taking the derivatives of our solution function wrt the parameters.
- Recall the previous example where $y^* = \frac{p}{k}$, now see how output would change with an change in price: $\frac{\partial y^*}{\partial p} = \frac{1}{k} > 0$, in this case an increase in price leads to an increase in quantity supplied. See Figure b.
- Now consider and increase in input costs: $\frac{\partial y^*}{\partial k} = -\frac{p}{k^2} < 0$, as costs for the supplier increase, amount supplied goes down. See Figure c.

Multivariate Numeric Example:

- Consider a simple monopoly price discrimination problem where the monopolist sells its product in two separate markets with demand curves given by: $q_1 = 100 - p_1$ and $q_2 = 60 - p_2$. The monopolist will generally want to charge different prices to the two groups and assume that re-selling of the good by consumers is not possible. Let the monopolist's total cost function be defined as: $C = 1000 + 20(q_1 + q_2)$.
  - The monopolist's goal is to max profits (π) and the objective function is: $\pi = p_1 q_1 + p_2 q_2 - 1000 - 20(q_1 + q_2)$. Using the demand functions to solve for q_1 and q_2 in terms of p_1 and p_2. Profit expressed as a function of p_1 and p_2 is: $\pi(p_1, p_2) = 120 p_1 + 80 p_2 - p_1^2 - p_2^2 - 4200$.
  - The first order conditions for maximizing profit are: $\frac{\partial \pi}{\partial p_1}\pi(p_1, p_2) = 120 - 2p_1 = 0$ and $\frac{\partial \pi}{\partial p_2}\pi(p_1, p_2) = 80 - 2p_2 = 0$. Solving for the critical values of p_1 and p_2 results in $p_1^* = 60$ and $p_2^* = 40$.

Econometrics Application: Minimizing the Sum of Squared Errors (MSSE)

- Similar to our unconstrained profit max problems; in the following case we are using a tool (regression analysis) to examine the relationship of a dependent variable to independent or explanatory variables. In the case of a profit maximizing firm we are trying to find output as a function of price.
- In a regression we are trying to determine what is a "good fit" (e.g. which estimated parameters best approximate the true relationship of the variables). To make this determination, we minimize the sum of squared errors. Doing so has several advantages over competing methods 1) avoids problems of sign 2) regression line goes through the middle point 3) squaring emphasizes large errors 4) method is easily manageable 5) has a unique minimum and 5) has a unique, and best, solution.
- Mathematically, MSSE is represented by: $\min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} \varepsilon^2$
- A simple regression model: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$
- Where $\hat{\alpha}$ and $\hat{\beta}$ are estimates of the true, but unknown parameters α and β
- To minimize the sum of squared errors, $\min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - \hat{\alpha} + \hat{\beta} x_i)^2$
- At optimum, $\frac{\partial}{\partial \hat{\alpha}} \sum_{i=1}^{n} \varepsilon_i^{\,2} = 0$
- And, $\frac{\partial}{\partial \hat{\beta}} \sum_{i=1}^{n} \varepsilon_i^{\,2} = 0$
- Where, $\hat{\beta} = \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}$ and $\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$

Practice Problems:

1. A simple monopolist faces a linear inverse demand curve given by $p(y) = a - by$ where $a > 0$ and $b > 0$, and where y is the output of the monopolist. In addition, the monopolist faces the total cost function given by $c(y) = ky$, where $k > 0$. Furthermore, assume that $a > k$.
    a. Set up the profit maximization problem facing the monopolist
    b. Provide both graphical and economic interpretations of the assumption $a > k$. It may help to provide an economic interpretation of parameters a and k individually first.
    c. What does the assumption $a > k$ imply for the profit maxing output level of the monopolist? Use the graph in part b to answer this question.
    d. Find the monopolist's profit maximizing level of output, say $y = y^m(a,b,k)$, as well as its profit maximizing price, say $p = p^m(a,b,k)$. How do you know that you have found the profit maximizing solution?
    e. Derive the comparative statistics, $\frac{\partial y^m}{\partial k}$ and $\frac{\partial p^m}{\partial k}$
    f. Provide an economic interpretation of the above comparative statistics.

2. In this question, we will examine the effect of a *sales tax* on a monopolist's optimal choice of price and output. Assume that the monopolist faces a linear inverse demand

curve given by $p(y) = a - by$ where $a > 0$ and $b > 0$, and a quadratic cost function given by $c(y) = \frac{1}{2}ky^2$, where $k > 0$. The monopolist is also assumed to pay a fraction s, $0 < s < 1$, of its total sales revenue in taxes to the government.

a. Set up the monopolist's profit maximization problem in the presence of the sales tax, and find the profit max level of output, say $y = y^s(a,b,k,s)$.

b. Find the monopolist's optimal price, $p = p^s(a,b,k,s)$.

c. Derive the comparative statistics for an increase in the sales tax, namely $\frac{\partial y^s}{\partial s}$ and $\frac{\partial p^s}{\partial s}$, and provide an economic interpretation of these statistics

d. What is the economic interpretation of an increase in a? You may want to plot a demand function to help with your interpretation.

e. Derive the comparative statics for an increase in the parameter a, namely $\frac{\partial y^s}{\partial a}$ and $\frac{\partial p^s}{\partial a}$, and provide an economic interpretation of these statistics.

f. What is the economic interpretation of an increase in the parameter k?

g. Derive the comparative statics for an increase in the parameter a, namely $\frac{\partial y^s}{\partial k}$ and $\frac{\partial p^s}{\partial k}$, and provide an economic interpretation of these statistics.

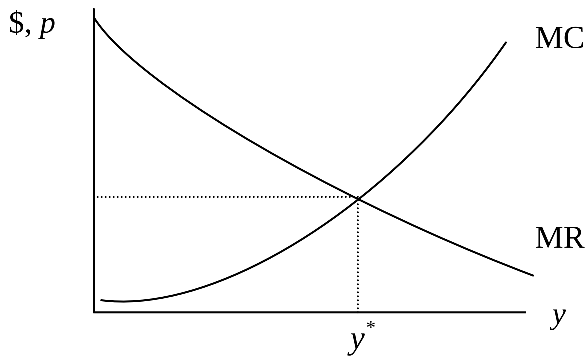


Figure a: Marginal Cost and Marginal Revenue Curves

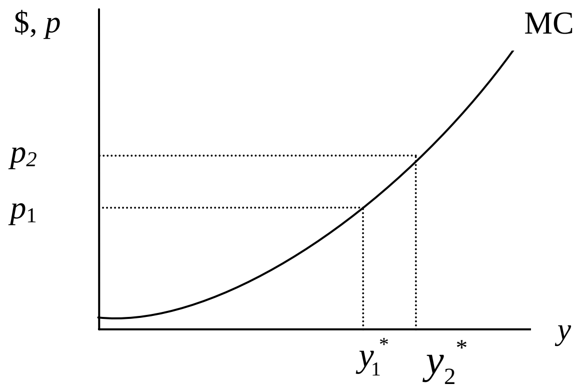


Figure b: Marginal Cost Curve with output price increase

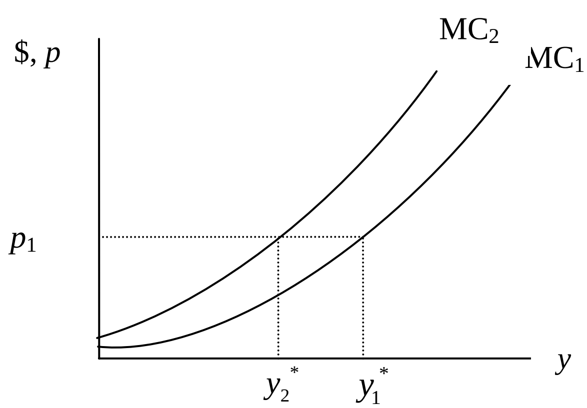


Figure c: Marginal Cost curve with input cost increase

Section III: Basic Statistics Review
Reference: Hill, Griffiths, and Lin, Appendices

Random Variables and PDFs

A **Random Variable** is a variable whose value is unknown until it is observed

- Not perfectly predictable
- A random variable is **discrete** if it can take on a countable number of values
  - E.g., sociodemographic characterics such as male/female, income categories, etc…
- A random variable is **continuous** if it can take on any real value in an interval
  - E.g. Income, prices, etc…

Random variables are characterized by **probability distributions (pdf's)** that describe the likelihood of alternative outcomes from **draws,** or **experiments.**

- The pdf for a discrete random variable gives the probability of each potential outcome
  - **Notation**: $f(x) = \Pr(X = x)$, where $f(x)$ is the probability of outcome x occurring and X is the random variable described by $f(x)$
  - Sometimes, $f(x)$ can be an equation
  - To be a proper pdf, $0 \le f(x) \le 1$ (probabilities must be non-negative) and $\sum_{n=1}^{N} f(x_n) = f(x_1) + f(x_2) + \ldots + f(x_N) = 1.$
- The pdf for a continuous random variable is similar, with a few small differences
  - Since the number of outcomes is not countable:
    a) the probability of any particular outcome is zero
    b) instead of summing, we use integration (the continuous form of summing for non-countable outcomes), so that $\int f(x)dx = 1.$
  - It is still true that $0 \le f(x) \le 1.$

A **cumulative distribution function** gives the same information as the pdf in a different way…it gives the probability that the random variable X is less than or equal to a particular value

- **Notation**: $F(x) = \Pr(X \le x)$, where $F(x)$ is the probability that the variable X takes on a value less than or equal to outcome x

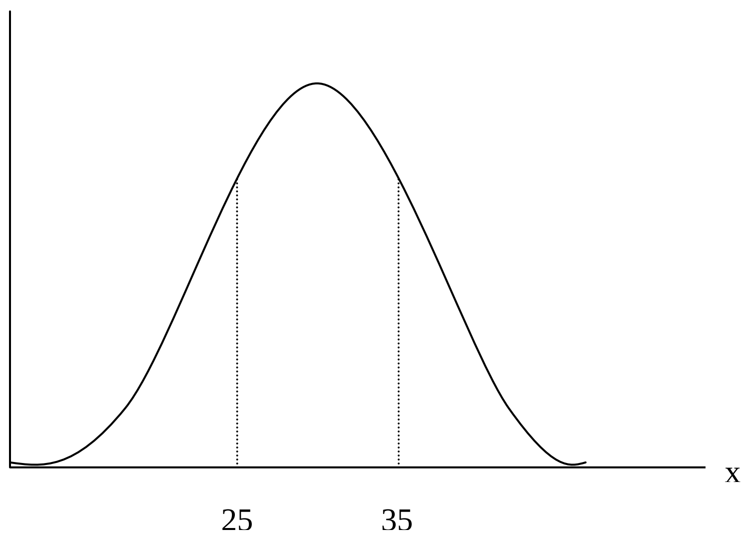
Examples

Consider the following pdf for a **discrete** random variable.

x	f(x)	F(x)
1	0.3	*0.3*
2	0.2	*0.5*
3	0.5	*1*

- *Is it a proper pdf?*A: Yes, f(x) are non-negative for each outcome and the sum of f(x) equals one
- *Can you derive the cdf from the pdf?*
- *What is the probability that X=2?* A: 0.2/1=0.2, or 20%
- *What is the probability that* $X \leq 2$? A: (.3+.2)/1=0.5, or 50%.
  Note $\Pr(X \leq 2) = 1 - \Pr(X > 2)$.

Consider the following pdf for a **continuous** random variable.



- *Is it a proper pdf?*A: Yes, if integral under the curve equals 1
- *Can you plot the cdf?*
- *What is the probability that X=35?* A: 0, due to non-countability
- *What is the probability that* $X \leq 35$? A: The area under the curve to the left of 35.
- *What is the probability that* $25 \leq X \leq 35$? The area under the curve to the left of 35 and to the right of 25.

Joint, Marginal, and Conditional Probabilities

A **joint probability density function** describes the probabilities of the realized values of multiple variables.

Example

	x	
y	0	1
1	0.4	0.15
2	0.25	0.2

- **Notation**: $f(x,y)=\Pr(X=x,\ Y=y\ )$ where $f(x,y)$ is the joint probability of outcome *x* AND *y* occurring with *X* and *Y* as random variables
- *What is the probability that X=1 and Y=2?* A: $f(1,2)=0.2.$
- A joint pdf is proper if $0\le f(x,y)\le 1$ and $\sum_x\sum_y f(x,y)=1.$

A **marginal probability distribution** is the probability distribution of one variable, *regardless of the value of the other(s)*, from a joint distribution. The marginal distribution of *X* is denoted $f_X(x)=\sum_y f(x,y),$ while the marginal distribution of *Y* is denoted $f_Y(y)=\sum_x f(x,y).$ Note that we "sum out" the other variable.

What is the marginal distribution of X? In other words, what are the probabilities that *X*=0 and *X*=1, regardless of the values of *Y*?

Answer:
$$f_X(0)=\sum_y f(0,y)=\Pr(X=0,Y=1)+\Pr(X=0,Y=2)=0.4+0.25=0.65.$$
$$f_X(1)=\sum_y f(1,y)=\Pr(X=1,Y=1)+\Pr(X=1,Y=2)=0.15+0.2=0.35.$$

In table form, this can be represented as

	x		
y	0	1	$f_Y(y)$
1	0.4	0.15	0.55
2	0.25	0.2	0.45
$f_X(x)$	0.65	0.35	1

For continuous variables, we "integrate out" instead of "sum out".

A **conditional probability distribution** gives the distribution of a random variable taking the value of (some of) the others as given.

- **Notation**: $f(y\mid x)=\Pr(Y=y\mid X=x)=\dfrac{\Pr(Y=y,X=x)}{\Pr(X=x)}=\dfrac{f(x,y)}{f_X(x)}.$
- **Read**: "the distribution of y conditional on the value of x"
- Note that the general idea here is that we are changing the total number of outcomes considered with the distribution.

- **Example:** Let's find the conditional distribution of *Y* given *X*=1. Now the number of outcomes we're considering is only 35% of the population, given by $f_X(1).$ Using the formula, $f(1|1) = \frac{.15}{.35} \approx .429,$ and $f(2|1) = \frac{.2}{.35} \approx .571.$ Thus, the probability that *Y*=1 given that *X*=1 is 42.9%. Note that $f(1|1) + f(2|1) = 1,$ as required by a proper pdf.

- Two variables are **statistically independent** if the conditional distribution of a variable is equal to the marginal distribution of the variable; i.e., $f(y|x) = f_Y(y).$ This makes sense…the value of *X* does not influence the chance of seeing *y*.

- Also note that since $f(y|x) = \frac{f(x,y)}{f_Y(y)},$ then statistical independence implies $f(x,y) = f_X(x) \cdot f_Y(y).$

- This provides an easy check of independence if you have the marginal distributions. For example, $f(0,1) = .4,$ but $f_X(0) \cdot f_Y(1) = .65 \cdot .55 = .3575,$ so the variables *X* and *Y* **are not independent**.

Measures of Central Tendency (mean, median, and mode) and Dispersion (variance, standard deviation)

- The **mean** of a random variable *X* (often denoted μ or μ_X) is its **expectation** [often denoted $E(X)$]. It is a weighted sum of potential outcomes, with the probabilities of each outcome as the weights.

- Discrete distribution: $\mu_X = E(X) = \sum_x x f(x).$

- Continuous distribution: $\mu_X = E(X) = \int_x x f(x) dx.$

- **Example:**

x	f(x)
1	0.3
2	0.2
3	0.5

$$\mu_X = E(X) = 1(.3) + 2(.2) + 3(.5) = 2.2.$$

- The **median** (often denoted *m*) is the value of *m* for which $\Pr(X \le m) = .5$ and $\Pr(X \ge m) = .5.$ It tends to be less sensitive to outliers.

- The **mode** is the value of *x* corresponding to the maximum pdf value.

- We will be particularly interested in *functions* of random variables and their expectations. Not surprisingly, $E\left[g(X)\right] = \sum_x g(x) f(x).$

- **Special Case:** If $g(X) = aX$, where *a* is constant, then $E\left[g(X)\right] = E[aX] = \sum_x axf(x) = a\sum_x x f(x) = aE[X] = a\mu.$ Furthermore, by the same logic, $E\left[ag(X)\right] = aE\left[g(x)\right].$

    - However, in general, $E(g(X)) \neq g\left(E(X)\right).$ Thus, $E\left[X^2\right] \neq E\left[X\right]^2.$

- The **variance** of a random variable *X* (often denoted σ^2) is defined as $\text{var}(X) = E\left[X - E(X)\right]^2 = E\left[X - \mu\right]^2 = E(X^2) - \mu^2.$ It measures dispersion around the mean. The larger the variance, the more "spread out" the distribution.

- The **standard deviation** (often denoted σ) is the square root of the variance. It also measures dispersion, but is in the same units as the random variable.

- **Example:** $\text{var}(X) = E\left[X - \mu\right]^2 = .3\left[1 - 2.2\right]^2 + .2\left[2 - 2.2\right]^2 + .5\left[3 - 2.2\right]^2 = 0.76.$

- There are other "moments" of distributions, such as skewness and kurosis.

- **Further Example:** *Find the mean and variance of the function aX+b* where a and b are constants and *X* has mean μ and variance σ^2.

A:
$$\begin{aligned} E\left[g(X)\right] &= E[aX] = \sum_x \left(ax + b\right) f(x) = a\sum_x x f(x) + \sum_x bf(x) \\ &= a\sum_x x f(x) + b\sum_x f(x) = aE\left[X\right] + b(1) = a\mu + b. \end{aligned}$$

$$\begin{aligned} \text{var}(aX + b) &= E\left[(aX + b) - E(aX + b)\right]^2 = E\left[(aX + b) - (a\mu + b)\right]^2 \\ &= E\left[(aX) - (a\mu)\right]^2 = E\left[a(X - \mu)\right]^2 = E\left[a^2(X - \mu)^2\right] \\ &= a^2 E\left[(X - \mu)^2\right] = a^2\,\text{var}(X). \end{aligned}$$

- Our most simple model of economic variables will involve linear functions of random variables!

An Extension to Several Variables

- Not surprisingly, $E\left[g(X,Y)\right] = \sum_x \sum_y g(x,y) f(x,y),$ where $f(x,y)$ is the joint distribution of the (discrete) random variables *X* and *Y*.

- Using algebra (please try this on your own, and see p. 490-491), it can be shown that $E\left[X+Y\right]=E(X)+E(Y)$, and $E\left[aX+bY+c\right]=aE(X)+bE(Y)+c$, where *a*, *b*, and *c* are constants.

- We can also show that if the variables *X* and *Y* are ***independent***, then $E\left[XY\right]=E\left(X\right)E\left(Y\right)$. *How?*

$$E\left[XY\right]=\sum_x\sum_y xyf(x,y)=\sum_x\sum_y xyf_X(x)f_Y(y)=\sum_x xf_X(x)\sum_y yf_Y(y)=E\left(X\right)E\left(Y\right).$$

- We also sometimes like to know how variables move together…in other words, if a draw of *X* is greater than its mean, what can we expect of *Y*?

- The meaningful statistic here is the **covariance** (often denoted σ_{XY}), which is defined by $\text{cov}\left(X,Y\right)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]=E(XY)-E(X)E(Y)$.

- **Example:**

	x		
y	0	1	$f_Y(y)$
1	0.4	0.15	0.55
2	0.25	0.2	0.45
$f_X(x)$	0.65	0.35	1

$$E(XY)=.4(0)+.15(1)+.25(0)+.2(2)=.55.$$
$$E(X)=.65(0)+.35(1)=.35$$
$$E(Y)=.55(1)+.45(2)=1.45$$
$$\text{cov}(X,Y)=E(XY)-E(X)E(Y)=.55-.35(1.45)=.0425>0$$

- The fact that $\text{cov}(X,Y)>0$ suggests that when *X* values are greater than their mean, then *Y* values are greater than their mean.

- The **correlation** between *X* and *Y* (often denoted ρ) standardizes the covariance and lies between -1 (perfect negative linear correlation) and +1 (perfect positive linear correlation). The formula is $\rho=\dfrac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$.

- In our example,

$$\text{var}(X) = .65(0-.35)^2 + .35(1-.35)^2 = .2275$$
$$\text{var}(Y) = .55(1-1.45)^2 + .45(2-1.45)^2 = .2475$$
$$\rho = \frac{.0425}{\sqrt{.2275}\sqrt{.2475}} \approx .1791.$$

- The variance of a linear function of random variables can be shown to be:

$$\text{var}(aX + bY) = a^2\,\text{var}(X) + b^2\,\text{var}(Y) + 2ab\,\text{cov}(X,Y)$$

Review of Statistical Inference

We have now learned about the mean $\mu = E[Y]$ and variance $\sigma^2 = \text{var}(Y)$ of the **population**. As such, these two statistics are **population parameters**.

However, only very rarely in economics do we have data for the entire population (like the over 18 population of the United States, or sales data every day). Instead, we use

- S**ampling** - drawing randomly from the population of interest, to do
- **Statistical Inference –** using a sample to make conclusions about the population parameters, whose estimates are called
- **Estimators** – estimates of population parameters or related measures constructed from the sample data

An Estimator for Central Tendency (Estimating μ)

Assume we have a randomly drawn sample $Y_1, Y_2, \ldots Y_N$ drawn from a distribution $Y \sim (\mu, \sigma^2)$. The notation says that each *Y* is distributed with mean μ and variance σ^2. The two population parameters are ***unknown***, but we'd like to make some conclusions based on our sample.

The **sample mean** is defined by $\bar{Y} = \sum_{i=1}^{N} Y_i / N$, and is an estimator for μ.

- Note that $\bar{Y}$ is a random variable, because it is a (linear) function of random variables.
- As such, it, too, has a **sampling distribution** – a mean and variance that are *not necessarily the same* as the mean and variance of the population

Question: How good of an estimator is $\bar{Y}$?
- Ideally, we'd like it to give us a precise, accurate estimate of μ.
- In other words, we'd like the mean of $\bar{Y} = \mu$, or $E[\bar{Y}] = \mu$, and $\text{var}(\bar{Y})$ to be small.

Why?

Let's check it out – first the mean.

$$E[\bar{Y}] = E[\sum_{i=1}^{N} Y_i / N] = \frac{1}{N} E[\sum_{i=1}^{N} Y_i] = \frac{1}{N}\left[E[Y_1] + E[Y_2] + \ldots + E[Y_N]\right]$$
$$= \frac{1}{N}\left[\mu + \mu + \ldots + \mu\right] = \frac{N}{N}\mu = \mu$$

In other words, the mean of our estimator is the population mean…or said another way, our estimator is centered on the true population mean.

- This estimator is thus termed **unbiased**, which means that our estimator is centered (in repeated samples, the average value is) on the true population value

How about the variance?

- Using the rule $\mathrm{var}(aX) = a^2\,\mathrm{var}(X)$, we obtain

$$\mathrm{var}(\bar{Y}) = \mathrm{var}\left(\frac{1}{N}Y_1 + \frac{1}{N}Y_2 + \ldots + \frac{1}{N}Y_N\right) = \frac{1}{N^2}\mathrm{var}\left(Y_1\right) + \frac{1}{N^2}\mathrm{var}\left(Y_2\right) + \ldots + \frac{1}{N^2}\mathrm{var}\left(Y_N\right)$$
$$= \frac{1}{N^2}\sigma^2 + \frac{1}{N^2}\sigma^2 + \ldots + \frac{1}{N^2}\sigma^2 = \frac{N}{N^2}\sigma^2 = \frac{\sigma^2}{N}.$$

- What does this mean? It says that for any sample size greater than 1, the variance of our estimator is less than the variance of the population.
- Furthermore, it says that the variance tends to zero as the sample size goes up. In other words, the more data we have, the better our estimator!

So we have shown that our estimator is distributed $\bar{Y} \sim (\mu, \frac{\sigma^2}{N})$. Of course, we don't know what these population parameters actually are, but that's OK.

The Central Limit Theorem

However, to be truly useful for statistical inference, we need to know ***what kind*** of distribution it is (e.g., normal, chi-sq, F, etc…). A handy result is the Central Limit Theorem (CLT).

CLT: If $Y_1, \cdots, Y_N$ are independent and identically distributed (iid) random variables with mean μ and variance σ^2, and $\bar{Y} = \sum_{i=1}^{N} Y_i / N$, then

$$Z_N = \frac{\bar{Y} - \mu}{\sigma / N}$$

has a probability distribution that converges to the standard normal N(0,1) as $N \to \infty$.

Key Points about CLT

- Whatever distribution we start with, CLT suggests that the sample mean is approximately normally distributed in large samples (we say asymptotically distributed)
- How large? Rule of thumb is $N \geq 30$.
- Who cares? We can use this information to perform statistical inference (remember, drawing conclusions about the population from the sample), like hypothesis testing and confidence intervals.

The Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right], -\infty < x < \infty$$

A **Standard Normal** random variable is distributed normally with $\mu = 0$ and $\sigma^2 = 1$.

A very handy result is that any normal random variable can be transformed to create a standard normal random variable, which can be used for inference. Specifically, if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

As such the probability that the random variable *X* takes on a value less than *a* is

$$\Pr\left[X \leq a\right] = \Pr\left[\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right] = \Pr\left[Z \leq \frac{a-\mu}{\sigma}\right] = \Phi\left[\frac{a-\mu}{\sigma}\right].$$

We can read this value from a table.

Example: If $X \sim N(3,2)$, what is the probability that any realized value is less than 1?

What would we like to know? $\Pr\left[X < 1\right]$.

How do we find it?

$$\Pr\left[X < 1\right] = \Pr\left[\frac{X-3}{\sqrt{2}} < \frac{1-3}{\sqrt{2}}\right] = \Pr\left[Z < -1.414\right] = 1 - \Pr\left[Z < 1.414\right] \approx 1 - .9207 \approx .079.$$

So the probability that any realized value is less than one is approximately 7.9%.

Bringing it Together: Interval Estimator of μ

We have just shown that $\bar{Y} \sim N(\mu, \frac{\sigma^2}{N})$ asymptotically. But note that this estimator is a random variable itself, and we can't be guaranteed that $\bar{Y}$ and μ are the same.

We can use the information, however, to form a **confidence interval estimator** for the underlying population mean. This interval estimate is a range of values that may contain the true population parameter with a certain level of probability (called the **confidence level**).

Assume that the population variance σ^2 ***is known.*** *Good assumption? Not likely…only to illustrate the concept here!*

1. Transform your point estimate $(here,\ \bar{Y})$ to a known distribution (here, standard normal).
   - $Z = \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{N}} \sim N(0,1).$
2. Find the critical value(s) of the known distribution that correspond to your confidence level in both tails of the distribution. Thus, for a 90% confidence level, we want 5% of the probability in the left tail and 5% in the right.
   - For the standard normal, these critical values are -1.645 and 1.645 from the standard normal table in the book.
3. Set up a probability statement that says "the probability of your transformed variable lying between the critical values equals your confidence level".
   - $\Pr\left[-1.645 \le Z \le 1.645\right] = .90$
   - $\Pr\left[-1.645 \le \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{N}} \le 1.645\right] = .90$
4. Solve for the ***population*** parameter of interest.
   - Start with the left hand inequality: $-1.645 \le \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{N}}$
     - Multiply each side by $\sigma/\sqrt{N}$: $-1.645\left(\sigma/\sqrt{N}\right) \le \bar{Y}-\mu$
     - Subtract $\bar{Y}$ from each side: $-1.645\left(\sigma/\sqrt{N}\right) - \bar{Y} \le -\mu$
     - Multiply through by -1 (remembering to switch the inequality): $1.645\left(\sigma/\sqrt{N}\right) + \bar{Y} \ge \mu$
   - Now do the same thing for the RHS inequality: $\dfrac{\bar{Y}-\mu}{\sigma/\sqrt{N}} \le 1.645$
     - We end up with: $\mu \ge -1.645\left(\sigma/\sqrt{N}\right) + \bar{Y}$
   - Putting it together: $\Pr\left[\bar{Y} + 1.645\left(\sigma/\sqrt{N}\right) \ge \mu \ge \bar{Y} - 1.645\left(\sigma/\sqrt{N}\right)\right] = .90$

5. Note that our **interval estimate of the population mean** at the 90% level of confidence is $\bar{Y} \pm 1.645\left(\sigma/\sqrt{N}\right).$ There is thus a 90% chance that this interval contains the true population mean. NOTE: It is the INTERVAL that is random, NOT the population mean.

Of course, this method generalizes to different confidence levels and distributions.

Bringing it Together: Using $\bar{Y}$ for Inference About the Population Mean

Suppose we have a sample of 50 days of particulate emissions for Ft. Collins, with sample mean 10 ppm and (known population) variance of 4 ppm.

We can use this estimator to ask a great number of questions about the population mean…in other words, use the estimator for statistical inference! We do this formally via **hypothesis testing**.

A **hypothesis test** is a formal means of testing a conjecture about a population parameter using sample data.

ALL hypothesis tests have the following elements:

1. A **null hypothesis**, denoted H_0.
    - H_0 is the belief we maintain until convinced by evidence that we can reject it.
    - In this case, it specifies a value *c* for the population mean.
    - We write it $\mathrm{H}_0: \ \mu = c.$
    - The null is usually stated such that we reject the null hypothesis if our theory is correct

2. An **alternative hypothesis**, denoted H_1 or H_A.
    - The hypothesis we accept if the evidence causes us to reject H_0
    - Three examples are $\mathrm{H}_1: \ \mu > c,$ $\mathrm{H}_1: \ \mu < c,$ and $\mathrm{H}_1: \ \mu \neq c.$ The first two are "one-tailed tests", while the third is a "two-tailed test".

3. A **test statistic**, which is used in conjunction with the test statistic's known distribution under the null hypothesis, to define regions of "Do not reject" and "Reject" the null hypothesis
    - E.g., under the null hypothesis $\mathrm{H}_0: \ \mu = 3$ and assuming a known variance, $z = \dfrac{\bar{Y} - 3}{\sigma / \sqrt{N}} \sim N(0,1)$.

4. The **rejection region** is the region of the test statistic's distribution that most closely corresponds to the alternative hypotheses. Alternatively, it is the region that contains *unlikely* values under $H_{0.}$
    - If the null is true, the test statistic should be close to zero
    - If the null is not true, the absolute value of the test statistic is large
    - For $\mathrm{H}_1: \ \mu > c$, evidence against H_0 is a large (in absolute value), positive test statistic in the right tail of the distribution
    - For $\mathrm{H}_1: \ \mu < c$, evidence against H_0 is a large (in absolute value), negative test statistic in the left tail of the distribution
    - "Large" depends on the **level of significance** of the test, or the probability of an unlikely (under the null) event. The level of significance is often denoted $\alpha.$ Note that the confidence level is $100(1-\alpha)\%.$

5. A **conclusion** either rejects or does not reject the null hypothesis on the basis of the sample evidence.
   - Note that we do not "accept" the null, but rather do not have enough evidence to reject

Let's use our sample to test that the mean daily particulate emissions in Ft Collins are greater than a standard of, say, 12 PPM.

1. $\text{H}_0: \ \mu = 12$
2. $\text{H}_1: \ \mu > 12$
3. Test Statistic: $z = \dfrac{\bar{Y} - \mu_0}{\sigma / \sqrt{N}} = \dfrac{10 - 12}{2 / \sqrt{50}} = -7.071.$
4. Rejection Region: Since our alternative is that the mean is greater than twelve, we get good evidence against the null if the test statistic is large and positive. At the 5% level of significance, the critical z-value (z_{crit}) such that $\Pr\left[z > z_{crit}\right] = (1 - \alpha) = .95$ is 1.645. At the 1% level of significance, it is 2.326.
5. Conclusion: Regardless of the significance level, we cannot reject the null hypothesis. In other words, the sample we have is compatible with the hypothesis that $\mu = 12$.

Note from this sample information, we could also find an interval estimator for the population mean at, say, the 95% level of confidence, namely:
$\bar{Y} \pm 1.96\left(\sigma / \sqrt{N}\right) = 10 \pm 1.96\left(2 / \sqrt{50}\right) = \left(9.446, 10.554\right).$

You should be able to show that for any null hypothesis that posits a mean outside of this range, we will reject the null at the 5% level of significance for a two-tailed test.

An Estimator for Dispersion (Estimating σ^2)

The population variance $\sigma^2 = \text{var}(Y) = E\left[Y - \mu\right]^2$. But since we don't know the true mean, we substitute $\bar{Y}$ instead. Then we essentially take the "average" variance across the sample, but adjust so that the estimator is unbiased. As such, our formula for the **sample variance**, or estimate of the population variance, is

- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{N} (Y_i - \bar{Y})}{(N - 1)}.$

We can use this statistic to estimate the variance of the estimator $\bar{Y}$, namely

- $\hat{\text{var}}(\bar{Y}) = \hat{\sigma}^2 / N,$

and find the standard error of the estimate $\bar{Y}$, or

- $se(\bar{Y}) = \sqrt{\hat{\text{var}}(\bar{Y})} = \hat{\sigma} \ / \sqrt{N}.$

Note the distinction between the estimated variance of the ***population*** and the estimated variance of the ***estimator***.

Interval Estimation and Hypothesis Testing when the Variance is Unknown

When the true population variance is unknown, and instead we use the standard error of the estimate, the transform of the mean estimator has a ***t-distrubution with N-1 degrees of freedom.*** In other words,

- $t = \dfrac{\bar{Y} - \mu}{\hat{\sigma} / \sqrt{N}} \sim t_{(N-1)}$.

We form interval estimators and hypothesis tests, however, in the same manner. The only difference is in using the critical t-values from the proper distribution.

Example:

Suppose a random sample of GPA's from 15 members of this class has a mean of 3.09 and sample variance of 0.06.

1. Find the 95% interval estimator for μ.

   First note that $t = \dfrac{3.09 - \mu}{.245 / \sqrt{15}} \sim t_{14}$, and that the critical t-values with .025% of the probability mass in each tail is 2.145 from the t-table. Then set up the probability statement of interest; namely $\Pr\left[-2.145 \leq \dfrac{3.09 - \mu}{.245 / \sqrt{15}} \leq 2.145\right] = .95$. Solving inside the probability statement for μ yields $\Pr\left[2.95 \leq \mu \leq 3.23\right] = .95$. So our 95% interval estimator for the population mean is (2.95, 3.23), which means there is only a 5% chance that our interval does not contain the true population mean.

2. Test the hypothesis that the GPA of the class equals 2.5 at the 1% level of significance.

   H_0: $\mu = 2.5$
   H_1: $\mu \neq 2.5$
   Test Stat: $t = \dfrac{3.09 - 2.5}{.245 / \sqrt{15}} \sim t_{14}$ under the null hypothesis, and this stat equals -9.33.
   Rejection region: test stat must be greater in absolute value than 2.977.
   Conclusion: Reject the null hypothesis…our sample data is NOT consistent with the theory that the class GPA is 2.5 at the 99% level of confidence.

A Few Other Related Concepts

P-value: The probability value, or **p-value**, associated with a test statistic associated with a hypothesis test is the cumulative probability density in the rejection region of the test. We can use this to get an exact value for the level of significance at which we can reject the null hypothesis. Most software reports both test statistics and p-values.

Type I and Type II errors:

A type I error occurs when the null hypothesis is TRUE and we REJECT it. The probability of a type I error is $\alpha,$ the level of significance.

A type II error occurs when the null hypothesis is FALSE and we DO NOT REJECT it. Although the probability of a type II error is unknown, since it depends on the underlying population parameters, it generally

A) varies inversely with α,

B) declines as the sample size increases, and,

C) is more likely when the null hypothesis value is close to the true value.

Example: Finding Means, Variance, Covariance, and Correlation

	x		
y	0	1	$f_Y(y)$
1	0.4	0.15	0.55
2	0.25	0.2	0.45
$f_X(x)$	0.65	0.35	1

$$E(XY) = .4(0) + .15(1) + .25(0) + .2(2) = .55.$$
$$E(X) = .65(0) + .35(1) = .35$$
$$E(Y) = .55(1) + .45(2) = 1.45$$
$$\text{cov}(X,Y) = E(XY) - E(X)E(Y) = .55 - .35(1.45) = .0425 > 0$$

- The fact that $\text{cov}(X,Y) > 0$ suggests that when *X* values are greater than their mean, then *Y* values are greater than their mean.

- The **correlation** between *X* and *Y* (often denoted ρ) standardizes the covariance and lies between -1 (perfect negative linear correlation) and +1 (perfect positive linear correlation). The formula is $\rho = \dfrac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$.

- In our example,

$$\text{var}(X) = .65(0-.35)^2 + .35(1-.35)^2 = .2275$$
$$\text{var}(Y) = .55(1-1.45)^2 + .45(2-1.45)^2 = .2475$$
$$\rho = \frac{.0425}{\sqrt{.2275}\sqrt{.2475}} \approx .1791.$$

Example: Hypothesis Testing About the Mean: Known Population Variance

We can use this estimator to ask a great number of questions about the population mean…in other words, use the estimator for statistical inference! We do this formally via **hypothesis testing**. A **hypothesis test** is a formal means of testing a conjecture about a population parameter using sample data.

ALL hypothesis tests have the following elements:

1. A **null hypothesis**, denoted H_0.
   - H_0 is the belief we maintain until convinced by evidence that we can reject it.
   - In this case, it specifies a value *c* for the population mean.
   - We write it $\mathrm{H}_0: \ \mu = c$.
   - The null is usually stated such that we reject the null hypothesis if our theory is correct

2. An **alternative hypothesis**, denoted H_1 or H_A.
   - The hypothesis we accept if the evidence causes us to reject H_0
   - Three examples are $\mathrm{H}_1: \ \mu > c,$ $\mathrm{H}_1: \ \mu < c,$ and $\mathrm{H}_1: \ \mu \neq c.$ The first two are "one-tailed tests", while the third is a "two-tailed test".

6. A **test statistic**, which is used in conjunction with the test statistic's known distribution under the null hypothesis, to define regions of "Do not reject" and "Reject" the null hypothesis
   - E.g., under the null hypothesis $\mathrm{H}_0: \ \mu = 3$ and assuming a known variance, $z = \dfrac{\bar{Y} - 3}{\sigma / \sqrt{N}} \sim N(0,1)$.

7. The **rejection region** is the region of the test statistic's distribution that most closely corresponds to the alternative hypotheses. Alternatively, it is the region that contains *unlikely* values under H_0.
   - If the null is true, the test statistic should be close to zero
   - If the null is not true, the absolute value of the test statistic is large
   - For $\mathrm{H}_1: \ \mu > c$, evidence against H_0 is a large (in absolute value), positive test statistic in the right tail of the distribution
   - For $\mathrm{H}_1: \ \mu < c$, evidence against H_0 is a large (in absolute value), negative test statistic in the left tail of the distribution
   - "Large" depends on the **level of significance** of the test, or the probability of an unlikely (under the null) event. The level of significance is often denoted α. Note that the confidence level is $100(1-\alpha)\%$.

8. A **conclusion** either rejects or does not reject the null hypothesis on the basis of the sample evidence.
   - Note that we do not "accept" the null, but rather do not have enough evidence to reject

Suppose we have a sample of 50 days of particulate emissions for Ft. Collins, with sample mean 10 ppm and (known population) variance of 4 ppm.

Let's use our sample to test that the mean daily particulate emissions in Ft Collins are greater than a standard of, say, 12 PPM.

$\mathrm{H}_0: \ \mu = 12$

$\mathrm{H}_1: \ \mu > 12$

Test Statistic: $z = \dfrac{\bar{Y} - \mu_0}{\sigma / \sqrt{N}} = \dfrac{10 - 12}{2 / \sqrt{50}} = -7.071.$

Rejection Region: Since our alternative is that the mean is greater than twelve, we get good evidence against the null if the test statistic is large and positive. At the 5% level of significance, the critical z-value (z_{crit}) such that $\Pr\left[z > z_{crit}\right] = (1 - \alpha) = .95$ is 1.645. At the 1% level of significance, it is 2.326.

Conclusion: Regardless of the significance level, we cannot reject the null hypothesis. In other words, the sample we have is compatible with the hypothesis that $\mu = 12$.

Note from this sample information, we could also find an interval estimator for the population mean at, say, the 95% level of confidence, namely:

$\bar{Y} \pm 1.96\left(\sigma / \sqrt{N}\right) = 10 \pm 1.96\left(2 / \sqrt{50}\right) = \left(9.446, 10.554\right).$

Example: Hypothesis Testing About the Mean and Interval Estimation: Unknown Population Variance

The **sample variance**, or estimate of the population variance, is

- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{(N-1)}.$

We can use this statistic to estimate the variance of the estimator $\bar{Y}$, namely

- $\hat{\mathrm{var}}(\bar{Y}) = \hat{\sigma}^2 / N,$

and find the standard error of the estimate $\bar{Y}$, or

- $se(\bar{Y}) = \sqrt{\hat{\mathrm{var}}(\bar{Y})} = \hat{\sigma} \ / \sqrt{N}.$

When the true population variance is unknown, and instead we use the standard error of the estimate, the transform of the mean estimator has a ***t-distrubution with N-1 degrees of freedom.*** In other words,

- $t = \dfrac{\bar{Y} - \mu}{\hat{\sigma} \ / \sqrt{N}} \sim t_{(N-1)}.$

We form interval estimators and hypothesis tests, however, in the same manner. The only difference is in using the critical t-values from the proper distribution.

Example:

Suppose a random sample of GPA's from 15 members of this class has a mean of 3.09 and sample variance of 0.06.

Find the 95% interval estimator for μ.

First note that $t = \dfrac{3.09 - \mu}{.245/\sqrt{15}} \sim t_{14}$, and that the critical t-values with .025% of the probability mass in each tail is 2.145 from the t-table. Then set up the probability statement of interest; namely $\Pr\left[-2.145 \le \dfrac{3.09 - \mu}{.245/\sqrt{15}} \le 2.145\right] = .95$. Solving inside the probability statement for μ yields $\Pr\left[2.95 \le \mu \le 3.23\right] = .95$. So our 95% interval estimator for the population mean is (2.95, 3.23), which means there is only a 5% chance that our interval does not contain the true population mean.

Test the hypothesis that the GPA of the class equals 2.5 at the 1% level of significance.

H_0: $\mu = 2.5$
H_1: $\mu \ne 2.5$
Test Stat: $t = \dfrac{3.09 - 2.5}{.245/\sqrt{15}} \sim t_{14}$ under the null hypothesis, and this stat equals -9.33.
Rejection region: test stat must be greater in absolute value than 2.977.
Conclusion: Reject the null hypothesis…our sample data is NOT consistent with the theory that the class GPA is 2.5 at the 99% level of confidence.

Example: Creating an Interval Estimator when the True Population Variance is Unknown

Let's assume we have a sample of data, $Y_1 \ldots Y_N$, and have estimated $\bar{Y} = \dfrac{1}{N}\sum_{i=1}^{N} Y_i$ and

$\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{(N-1)}$, where $\hat{\sigma}^2$ is an estimate of the population variance.

1. Transform your point estimate $(here,\ \bar{Y})$ to a known distribution (here, a *t* distribution with *N-1* degrees of freedom).
   - $t = \dfrac{\bar{Y} - \mu}{\hat{\sigma}/\sqrt{N}} \sim t_{(N-1)}$.

2. Find the critical value(s) of the known distribution that correspond to your confidence level in both tails of the distribution. Thus, for a 90% confidence level, we want 5% of the probability in the left tail and 5% in the right.
   - For the t distribution with, say, 27 degrees of freedom (implying *N*=28), these critical values are -1.703 and 1.703 from the *t* tables.

3. Set up a probability statement that says “the probability of your transformed variable lying between the critical values equals your confidence level”.
    - $\Pr\left[-1.703 \le t \le 1.703\right] = .90$
    - $\Pr\left[-1.703 \le \frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{N}} \le 1.703\right] = .90$

4. Solve for the ***population*** parameter of interest.
    - Start with the left hand inequality: $-1.703 \le \frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{N}}$
        - Multiply each side by $\hat{\sigma}/\sqrt{N}$: $-1.703\left(\hat{\sigma}/\sqrt{N}\right) \le \bar{Y} - \mu$
        - Subtract $\bar{Y}$ from each side: $-1.703\left(\hat{\sigma}/\sqrt{N}\right) - \bar{Y} \le -\mu$
        - Multiply through by -1 (remembering to switch the inequality): $1.703\left(\hat{\sigma}/\sqrt{N}\right) + \bar{Y} \ge \mu$
    - Now do the same thing for the RHS inequality: $\frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{N}} \le 1.703$
        - We end up with: $\mu \ge -1.703\left(\hat{\sigma}/\sqrt{N}\right) + \bar{Y}$
    - Putting it together:
      $\Pr\left[\bar{Y} + 1.703\left(\hat{\sigma}/\sqrt{N}\right) \ge \mu \ge \bar{Y} - 1.703\left(\hat{\sigma}/\sqrt{N}\right)\right] = .90$

5. Note that our **interval estimate of the population mean** at the 90% level of confidence is $\bar{Y} \pm 1.703\left(\hat{\sigma}/\sqrt{N}\right)$.

INTERPRETATION: There is thus a 90% chance that this interval contains the true population mean. NOTE: It is the INTERVAL that is random, NOT the population mean.

IT IS NOT CORRECT to say that the probability that the true mean is between $\left[\bar{Y} + 1.703\left(\hat{\sigma}/\sqrt{N}\right), \bar{Y} - 1.703\left(\hat{\sigma}/\sqrt{N}\right)\right]$ is 90%. Since the true mean is non-random, this probability is either one or zero.

Of course, this method generalizes to different confidence levels and distributions.

PRACTICE PROBLEMS

1. Consider the following discrete probability distribution function for the variable X:

x	0	1	2	3	4
f(x)	0.1	0.3	0.3	0.1	0.2

a. Is this a proper probability density function? How do you know?
b. Find the mean, variance, and standard deviation of x, showing your work.
c. What is the probability that any draw from this distribution, say X, is greater than 2 [i.e., find $P(X > 2)$]?

2. Consider the following discrete joint probability function for the variables *Education* and *Income*. *Education* measures each individual's education achievement, with

$$Edu = \begin{cases} 1 & \text{high school or less} \\ 2 & \text{some college} \\ 3 & \text{4 year degree} \\ 4 & \text{advanced degree} \end{cases},$$

and *Income* measures gross income class, with

$$Inc = \begin{cases} 1 & <\$30{,}000/\text{yr} \\ 2 & \$30{,}000\text{-}50{,}000/\text{yr} \\ 3 & >\$50{,}000/\text{yr} \end{cases}.$$

The joint pdf is given as:

	Inc		
Edu	1	2	3
1	0.23	0.20	0.02
2	0.03	0.22	0.02
3	0.03	0.04	0.08
4	0.01	0.04	0.08

a. Is this a proper probability density function? How do you know?
b. What is the probability that an individual drawn from this distribution makes more than \$50,000 per year? What type of probability is this (e.g., joint, marginal, or conditional)?
c. Construct the conditional pdf of *Income* given that an individual has a four year degree but no graduate study; i.e., construct the table that shows $f(Inc \mid Edu = 3)$.
d. What is the probability that an individual earns more than \$30,000 per year given that they have at least some college?
e. Calculate the mean income class for those individuals with a high school degree or less. Is this greater than or less than the mean income class for those with an advanced degree?
f. Find the covariance and correlation of *Edu* and *Inc*, and provide an interpretation of the sign of this statistic.
g. Are Education and Income independent? How do you know?

3. Suppose you collected data on the average number of beers consumed by patrons of the Trailhead on any given Saturday night, and found that from a sample of 23 customers, the sample mean ($\overline{Y} = \sum_{i=1}^{23} Y_i / N$) was 3.14 with an estimated standard error of

$$se(\overline{Y}) = \frac{\hat{\sigma}}{\sqrt{N}} = 1.1.$$

a. Provide an estimate of the mean and the variance of the *population* of Saturday night Trailhead patrons using your sample.
b. Approximately what is the probability that any patron from the population consumes more than 5 beers on any given Sat. night?
c. Approximately what proportion of the population consumes between 2 and 4 beers on any given Sat. night at the Trailhead?
d. At a minimum, how many beers do you estimate the top 5% of drinkers consume on a Sat. night?
e. Provide the 95% confidence interval for the mean number of beers consumed at the Trailhead on Sat. night, and provide an interpretation.
f. Test the null hypothesis that the mean number of beers consumed on Saturday night is 3.5 at the 5% level of significance using a two-tailed test, being sure to write down the null and alternative hypotheses, the test statistic, indicate the rejection region, and provide a conclusion.
g. Repeat part f), but now test the alternative hypothesis that the mean number of beers is greater than 3.5.

4. Suppose $X \sim (3,9)$ and $Y \sim (1,25)$, with $\text{cov}\left[X,Y\right] = -2.$

   a. Find $E\left[X+Y\right].$
   b. Find $E[6X-2Y].$
   c. Find $\text{var}\left[X+Y\right].$
   d. Find $\text{var}\left[6X-2Y\right].$

5. Suppose you collected data on the average weekly snowfall at Breckenridge for 24 weeks over the winters of 2008 and 2009, and found that the sample mean ($\bar{Y} = \sum_{i=1}^{24}(Y_i / N)$) was 4.5 inches with an estimated standard error of $se(\bar{Y}) = \frac{\hat{\sigma}}{\sqrt{N}} = .45.$ Assume that the population data are normally distributed.

   a. Provide an estimate of the mean and the variance of the *population distribution* of weekly wintertime Breckenridge snowfall using your sample.
   b. Using the distribution you calculated in a), compute the probability that any week will see more than 6 inches of snow?
   c. Approximately what proportion of the weeks see between 2 and 4 inches of snowfall?
   d. What is the level of snow associated with the top 5% of all winter weeks?
   e. Provide the 95% confidence interval for the mean number of inches of snow per week at Breckenridge, and provide an interpretation.
   f. Test the null hypothesis that the mean number of weekly inches of snowfall is 3.5 at the 5% level of significance using a two-tailed test, being sure to write down the null and alternative hypotheses, the test statistic, indicate the rejection region, and provide a conclusion.
   g. Repeat part f), but now test the alternative hypothesis that the mean number of inches is greater than 3.5.

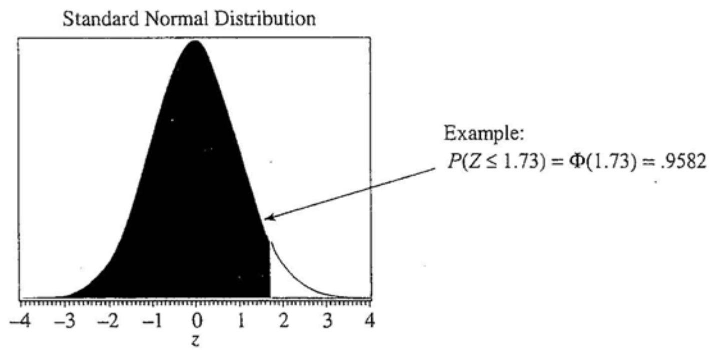


Table 1 **Cumulative Probabilities for the Standard Normal Distribution**
$\Phi(z) = P(Z \le z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Source: This table was generated using the SAS® function PROBNORM ution

08	0.09
319	0.5359
714	0.5753
103	0.6141
480	0.6517
844	0.6879
190	0.7224
517	0.7549
823	0.7852
106	0.8133
365	0.8389
599	0.8621
810	0.8830
997	0.9015
162	0.9177
306	0.9319
429	0.9441
535	0.9545
625	0.9633
699	0.9706
761	0.9767
812	0.9817
854	0.9857
887	0.9890
913	0.9916
934	0.9936
951	0.9952
963	0.9964
973	0.9974
980	0.9981
986	0.9986
990	0.9990

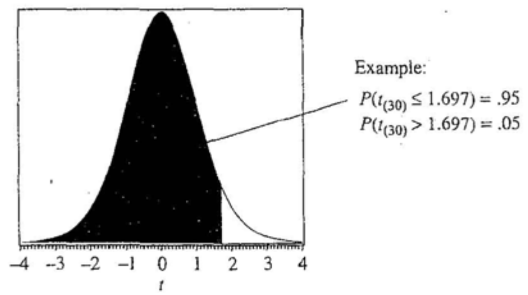


Table 2 **Percentiles of the t-distribution**

df	$t_{(.90,df)}$	$t_{(.95,df)}$	$t_{(.975,df)}$	$t_{(.99,df)}$	$t_{(.995,df)}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715