- $y = ax^n - 1$.
- $y = f(x) + g(x) = (3x - 2) + (c - 4x^3)$.
- $y = f(x) \cdot g(x) = (13x + 2x^3) \cdot (x^5 - 4x + r)$.
- $y = (x - 3)^3$.
- $y = (\frac{x^2+1}{x+1})$.
- $y = 5x^7 + 7x^4 + 3x^2$.
- $y = 5x^8 + 10x^7 - 5x^6 - 5x^5 + 3x^4 + 7x^3 - 2x^2 + x - 1,123$.
- $y = x^3 + x^2 + 1$.
- $y = x^4 - x^3 + x^2 - x + 1$.
- $y = (3x^2 + 4)(2x^3 + 3x + 5)$.
- $y = (5x^3 + 4x^2 + 3x + 2)(7x^5 + 6x^4 + 5x^3 + 4x^2)$.
- $y = (3x^2 + 4) + (2x^3 + 3x + 5)$.
- $y = (5x^5 + 3x^3 + x + 1) - (4x^4 + 2x^2 + 2)$.
- $y = (x + 5)^2$.
- $y = (x^2 + x + 2)^2$.
- $y = \left(\frac{x^2+1}{x+1}\right)^2$.
- $y = \frac{x^3+x^2+x+1}{x^2+x+1}$.

2. Using the rules in this chapter (i.e., you don't need to go back to the definition), differentiate the following:

   a) $f(x) = a_n x^n + a_{n-1}x^{n-1} \ldots + a_0$. Try also expressing the derivative as a series.
   b) $f(x) = (x^3 + 2)\ln(x^4 - 5x + 3)$.
   c) $f(x) = \frac{(x^2-4)}{x^5-x^3+x}$.
   d) $f(x) = e^{x-\ln(x)+5}$.
   e) $f(x) = xg(x) - 7x^2$, where $g(x) = e^x \ln(x)$.
   f) $f(x) = a^x x^2 - b^{x^2}$.
   g) $f(x) = e^{5x}$.
   h) $f(x) = e^{5x^2+x+3}$.
   i) $f(x) = 3e^{2x}$.
   j) $f(x) = \frac{1}{2}e^{\frac{x}{2}}$.
   k) $f(x) = e^{\ln(2x)}$.
   l) $f(x) = e^{g(x)}$, where $g(x) = 7x^3 + 5x^2 - 3x + \ln(x) - 7$.
   m) $f(x) = x^2 g(x) + 6x^2$, where $g(x) = \log_a(x) + x^7$.
   n) $f(x) = ax^2 g(x) + 9x^4$, where $g(x) = e^{\ln(x)+2x^2}$.

3. Show that $\frac{d\log_a(x)}{dx} = \frac{1}{x(\ln(a))}$.

Chapter Seven

The Integral

Recall from Chapter 5 that our primary use of calculus will come in allowing us to deal with continuity usefully. The derivative provides us with the instantaneous change in a continuous function at each point. The derivative, then, permits us to graph the marginal rate of change in any variable that we can represent as a continuous function of another variable. In the next chapter we make extensive use of the derivative to find maxima and minima.

But what if we care less about change than about the *net effect of change*? Say we had some continuous function that represented the marginal change in voter turnout with respect to some aggregate measure of education, and we wanted to know the total level of voter turnout for all levels of education. To get this, we'd need to start at some point, say, where aggregate education is equal to zero, and then add up all the changes in turnout as education increased. This sounds straightforward enough, if time-consuming, save for one little factor: the function we have describes the instantaneous marginal change in turnout as education varies continuously! Somehow we have to add *continuously*, and our summation operator, $\sum$, is not going to cut it. This is where the integral enters the picture.

As we see in more detail below, the integral is like the limit of a sum. In our example, the integral of the marginal change function essentially adds up an infinite number of infinitesimal changes in turnout to produce the required total turnout function. This function specifies the level of turnout at any amount of education.

Though the integral is perhaps most naturally thought of in terms of summing changes, it is not limited to this. For example, other uses of the integral include calculations of areas and volumes. The most common use of the integral in political science relates to its connection to probability density functions, which describe the chance of any particular value of a continuous function occurring. We'll meet these in Part III. For now, we assert that the integral is needed to find cumulative distribution functions (the area under the curve of the probability density function), necessary for statistical inference, and to compute expected values and utilities, necessary for game theory.

We can distinguish between definite and indefinite integrals, and in Section 1 we provide an intuitive description of the definite integral that explicates the concept of a limit of sums. In Section 2 we discuss the antiderivative, or indefinite integral, and offer the fundamental theorem of calculus. In Section 3 we show how integrals can be computed, using the integral's status as the antiderivative in order to derive rules of integration. These rules are summarized in Section 4.

7.1 THE DEFINITE INTEGRAL AS A LIMIT OF SUMS

Take a moment to flip back to the plots of $f(x) = x^2$ in Chapter 5. There we were concerned with discerning the instantaneous rate of change in the function at any point. To get this, we drew secants between neighboring points, and then shrunk the distance between them until the secant became a tangent to the curve.

What if instead we were concerned with the area under the curve? That is, we want to know what the area is between the curve and the x-axis, from some point x_1 to another point x_2. As we alluded to earlier in this chapter, calculating such areas is central to statistical inference, as we'll see in Part III. If we let $x_1 = 1$ and $x_2 = 2$ and use $f(x) = x^2$, then we can visualize this area by dropping lines down from $f(x)$ at the two endpoints of the secant in the figure in Chapter 5. We show this in Figure 7.1, in which the shaded area is the area of interest.

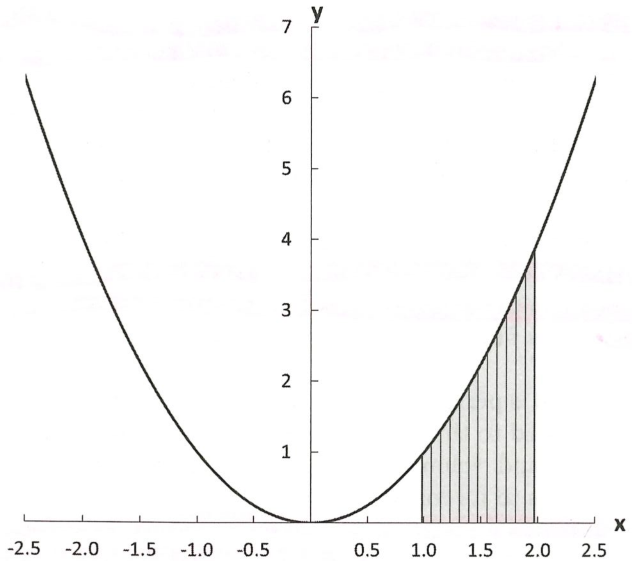

Figure 7.1: Area under $y = x^2$ from $x = 1$ to $x = 2$

How might we go about calculating this area? The intuition behind finding the area under a curve is pretty straightforward and is developed by thinking about finding the area within geometric shapes for which the formula for the area is known. If one were to take a curve and draw rectangles under the curve, then one could approximate the area under the curve by calculating the sum of the areas of the rectangles. That is, we can calculate the area of a rectangle,

and we can sum those areas. Figure 7.2 does this, drawing rectangles with finite width within the area shaded in the previous figure. As we can see, the area contained within these rectangles does get close to matching the shaded area in the previous figure; however, it undercounts by the triangular regions on top of each rectangle.[1]

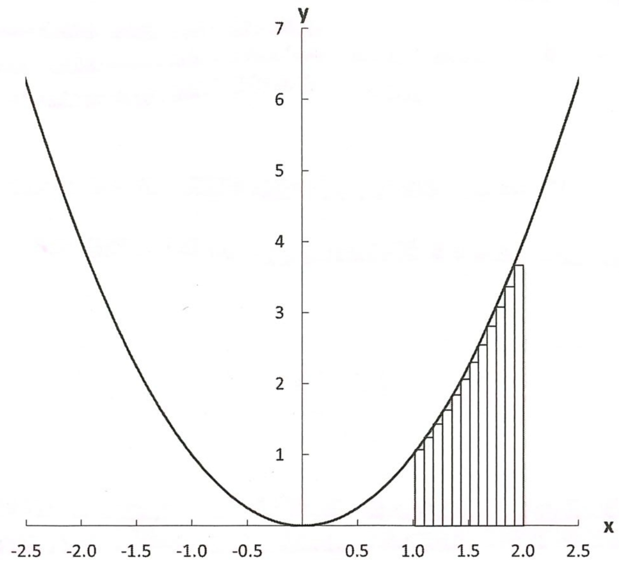

Figure 7.2: Area under $y = x^2$ from $x = 1$ to $x = 2$ with Rectangles

There are ways to do better than rectangles, using a similar method. We could add a triangle to the top of each rectangle, thus minimizing the underestimation. These would form trapezoids. They might overestimate or underestimate the area under the curve, depending on how they are drawn, but they would produce a closer approximation of the area. Or we could use the approximation developed by British mathematician Thomas Simpson, who proposed using tiny parabolas to fill in the gaps left by trapezoids (see Simpson's Rule). This effectively produces a quadratic interpolation, which should produce an exact solution for our case, since we have a quadratic function, $f(x) = x^2$, but will in general still over- or underestimate the area for more complex functions.

There is a way around these problems, however, and it is to draw our shapes very thinly: as they become more and more narrow, the divergence between the sum of their areas and the area under the curve will shrink. The German mathematician Georg Friedrich Bernhard Riemann proposed thinking about the area contained in the rectangles as they approach the limit of zero width. At the limit the area estimate will converge to the true value, and we can use these

[1] We could also have drawn the rectangles to overcount if the top of the rectangles had been drawn from the larger of the two points rather than the smaller.

rectangles not as an approximation of the area under the curve but as the actual area. The rectangle rule, then, can be used to calculate the area under a curve, and you might encounter reference to Riemann integrals or Riemann sums as a result.

How does this work? If we let the width be denoted Δx, then the area of each rectangle is $f(x_i)\Delta x$, where $f(x_i)$ is the value of the function at each (evenly spaced) point we have chosen. The total area might then be written as $\sum_i f(x_i)\Delta x$. We can take the limit of this sum as $\Delta x \to 0$ to produce the true area. We call this limit the **definite integral**, and write it in general as $\int_a^b f(x)dx$.[2]

The integral symbol, $\int$, looks like an S and is meant to, to remind you that it is in essence a sum. The dx in the integral tells you the variable of integration. It is exactly analogous to the dx in the derivative. The $f(x)$, or more generally, the expression multiplying the dx, is known as the **integrand**. The a and b are the bounds (or limits) of integration. They tell you the value of x at which to start and the value at which to stop. Unlike the derivative, which can be represented using a variety of notation, this is the single, common, agreed-upon notation for the integral. For our case, x^2 is the integrand, $a = 1$ and $b = 2$ are the bounds of integration, and the definite integral is $\int_1^2 x^2 dx$.

We have now given a name to our limit of a sum, but what do we do with it? How does one add up an infinite number of infinitesimally small areas? To do this we must first specify the nature of the connection between the integral and the derivative, captured in what is known as the fundamental theorem of calculus. We turn to this now.

7.2 INDEFINITE INTEGRALS AND THE FUNDAMENTAL THEOREM OF CALCULUS

In Chapter 3 we discussed the notion of an inverse of a function. This object is a function $f^{-1}(x)$ such that when composed with $f(x)$ it yields the identity function, x. In symbols, $f^{-1}(f(x)) = x$. In Chapter 1 we discussed additive and multiplicative inverses but did not link them to inverse functions. Though these concepts are not the same, they do have some similarities. The difference operator $(-)$ is in a rough sense the inverse of the sum operator $(+)$: if you add 3 and then subtract 3, you end up where you started. Similarly, if you multiply by some non-zero number and then divide by this same number, you end up where you started. So, in a sense, $\times$ and $\div$ are inverses of each other. We'll see that the derivative and the **antiderivative** (aka indefinite integral) are similarly related.

[2]Strictly speaking this is a Riemann integral, which is very useful but limited in some important ways that we will not discuss in this book. A more flexible integral that is founded on measure theory is the Lebesgue integral. Though that is not the approach this book will take, one can formulate probability theory using measure theory, and statistics or math departments often offer courses that focus on doing just that. Other varieties of integral exist as well.

7.2.1 Antiderivatives and the Indefinite Integral

The derivative is an operator that takes one function and returns another that describes the instantaneous rate of change of the first at each point. This is certainly more complex than addition or subtraction, but it doesn't change the fact that the derivative also has an inverse. We call this inverse the antiderivative. Consider some function $f(x)$. The antiderivative of $f(x)$ is denoted $F(x)$. That is, it's the capital of the letter used to represent the function.[3]

Because they are inverse operations, differentiation and antidifferentiation, applied in sequence, should take you back to where you started. In other words, $\frac{dF(x)}{dx} = f(x)$. Let's see how this works. Start with $f(x) = 1$. What is its antiderivative? Well, we know that the derivative of x is one. So $F(x) = x$ is one antiderivative.

It's not quite that easy, though. Take $F(x) = x + 10$. The derivative of that is also one, as is the derivative of $F(x) = x + 1,000$. In general, because the derivative of a constant is zero, there are lots (an infinite number, actually) of antiderivatives that all produce the same $f(x)$. Luckily, they're all of the same form: $F(x) = x + C$. We call C the constant of integration, and it can be any constant value (i.e., any value that does not depend on x).

We can find other antiderivatives in the same way. So, if $f(x) = x$, then we try to figure out what function $F(x)$, when differentiated, yields x. Well, we know that the derivative of x^2 is $2x$, so that's pretty close. We also know that the derivative is linear, so we can divide x^2 by 2 to get a derivative of x. Thus, the antiderivative of x is $\frac{1}{2}x^2 + C$. (Don't forget the C!)

For a slightly more complicated example, consider $\frac{1}{x}$. There's no polynomial that when differentiated produces this; check for yourself. What about other functions? While e^x certainly doesn't work, its inverse function, $\ln(x)$, does. Since $(\ln(x))' = \frac{1}{x}$, the antiderivative of $\frac{1}{x}$ is $\ln|x| + C$, where the absolute value in the argument of the logarithm arises to account for the fact that the logarithm isn't defined over negative real numbers.

Other antiderivatives can be found in the same fashion. At this point you might be asking why we care about antiderivatives, and it's a fair question. As a way of answering, first note another name for the antiderivative, the *indefinite integral*. We write an indefinite integral as $\int f(x)dx = F(x)$, so it looks the same as a definite integral without the bounds on the integral. The difference is that the **definite integral** returns a value, the area under the curve, while the **indefinite integral** returns a function[4] that, when differentiated, reproduces the integrand. In symbols, $\frac{d\int f(x)dx}{dx} = f(x)$.

[3]You will see this notation a great deal in Part III, as we typically denote probability distribution functions (PDFs) by $f(x)$ or $g(x)$ and their cumulative distribution functions (CDFs) by $F(x)$ or $G(x)$, respectively. The CDF is the antiderivative of the PDF, or, conversely, the PDF is the derivative of the CDF.

[4]Or, more precisely, a set of functions given the "indefiniteness" of having an unknown constant C in the function.

7.2.2 The Fundamental Theorem of Calculus

This still doesn't tell us why we care about taking indefinite integrals, though; it just tells us why the antiderivative is in a chapter on integration. To get at this question, we offer the grand-sounding **fundamental theorem of calculus**:

$$\int_a^b f(x)dx = F(b) - F(a).$$

In other words, the definite integral of a function from a to b is equal to the antiderivative of that function evaluated at b minus the same evaluated at a. The theorem is fundamental because it bridges (or links) differential and integral calculus. We can make the connection even more clear with a little notation $\int_a^b f(x)dx = F(x)|_a^b$, where the vertical line means "evaluate the antiderivative $F(x)$ at b, and subtract the antiderivative evaluated at a."

With this theorem, to calculate the area under the curve—a value that will prove very important for statistical inference—all we need to know is the indefinite integral of the function. We don't even have to worry about the constant C: since it appears in both indefinite integrals on the RHS, it cancels when they are subtracted.[5]

Given this, let's return to the example of the figures above. There $f(x) = x^2$. What's the antiderivative of this? Well, we know that the derivative of x^3 is $3x^2$ from the previous chapter, so an x^3 will be involved. If we further divide by 3 we'll get x^2 without the 3 in front. Thus, $F(x) = \frac{1}{3}x^3$. Consequently, by the fundamental theorem of calculus, $\int_1^2 x^2 dx = F(2) - F(1) = \frac{1}{3}(2)^3 - \frac{1}{3}(1)^3 = \frac{8}{3} - \frac{1}{3} = \frac{7}{3}$, and that is the area under the curve of x^2 between 1 and 2.

Let's try another example. Consider some function $f(x) = 1 + 2x + x^2$. Suppose we want to know the area under the curve over the range from $x = 0$ to $x = 3$. This area is the shaded region in Figure 7.3.

We first need to compute the indefinite integral. Recalling that the derivative is linear, we note that x differentiates to 1, x^2 differentiates to $2x$, and, as we just saw, $\frac{1}{3}x^3$ differentiates to x^2. Putting them together yields $F(x) = x + x^2 + \frac{1}{3}x^3$. Evaluated at $a = 0$, this is 0. At $b = 3$, this is 21. As $21 - 0 = 21$, so $\int_0^3 (1 + 2x + x^2)dx = 21$. Put differently, the sum of the changes in the function $F(x) = x + x^2 + \frac{1}{3}x^3$ (changes given by the derivative of $F(x)$, $f(x) = 1 + 2x + x^2$) from $x = 0$ to $x = 3$ is 21.

7.2.3 Why Should I Care?

One will encounter a number of statistical tables that report integral values. For example, z-scores—which report the area between the mean of a distribution and a point a selected distance from that mean under the standard normal curve—are definite integrals. Recall that definite integrals are values: the sum of the area under the curve between two points. Thus, when you look up z-score

[5] Hence the appellation *definite*: there is no uncertainty here.

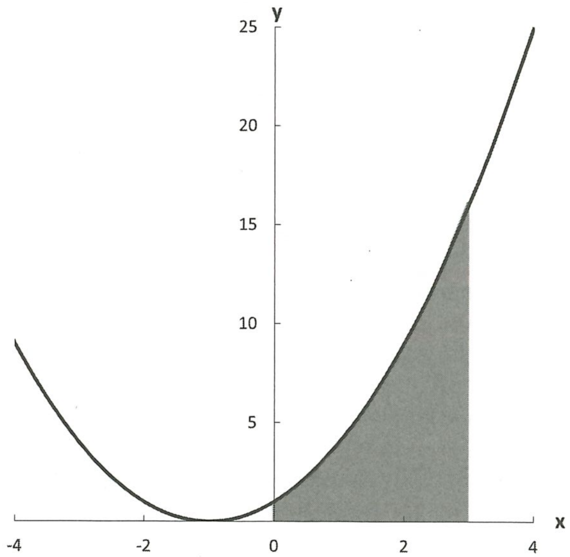

Figure 7.3: Shaded Area under $y = 1 + 2x + x^2$

values in a table in the appendix of a statistics text you are looking up the value of the difference between the antiderivative when $x = 0$ (the mean in a standard normal distribution) and the antiderivative at the point z (i.e., $\frac{X_i - \bar{X}}{s}$, where X_i is a given value of x, $\bar{X}$ is the sample mean, and s is the standard deviation of the sample). That value represents the area under the curve between those points. Conceptually, a z-score is the number of standard deviation units an individual observation is from the sample mean, and the area under the curve between it and zero relates to the chance of its being drawn from a standard normal distribution randomly.[6] Integral values are also important for studying continuous probability distributions in general, as we discuss in Chapter 11 of this book.

Formal theories also make appeals to integrals. Though some outcomes over which individuals may have an interest are certain, it is more often true that there is some level of uncertainty in the outcome, either because one is not sure what others with whom one is interacting will do, or because the payoffs themselves are variable. We can describe the chance that any particular outcome occurs by a probability distribution over outcomes. The expected value of the game (or expected utility) is the definite integral over all possible outcomes (or possible values of one's utility arising from different outcomes), weighted by the chance that each outcome occurs, i.e., by the probability distribution over outcomes.

[6] If you are unfamiliar with z-scores, make a note to return to this discussion when you are introduced to z-scores in your statistics coursework.

We cover continuous probability distributions in Chapter 11, where we provide an example that makes use of the integral of a probability distribution function to compute an expected utility. Here we provide a less intensive example by sticking to a decision theoretic context (i.e., one with a single decision maker). Assume that a bureaucrat must allocate money over two public works projects, one fixing a bridge and one building a new road. The payoff for the road might be known, but the payoff for the bridge is stochastic, since it depends on the likelihood that the bridge will collapse if it is not fixed. In other words, the return on investment in the case of the bridge is uncertain. To make things simpler, let the bureaucrat's decision be dichotomous: either money gets allocated or it doesn't, and money can only be allocated to one project. How would we figure out which project gets the money?

To model the bureaucrat's optimal decision we need to know the expected payout for each option. We show in Chapter 11 that this is computed by taking the expectation of the payout (or utility) function over the probability distribution of possible payouts. Here we simplify. Let the payoff for the road be 2, and the payoff for the bridge be uniformly distributed between -2 (if the bridge wouldn't have collapsed anyway) and 8 (if the bridge were in immediate danger of collapse). A uniform distribution places equal probability on every outcome. A rational actor will put money into the project with the greater payoff, but which is greater?

In Chapter 11 we define the uniform distribution formally. Here we assert that the expected value of a variable distributed according to the uniform distribution provided in the problem is the definite integral $\frac{1}{8-(-2)}\int_{-2}^{8} x dx = \frac{1}{10}\frac{1}{2}(x^2)|_{-2}^{8} = \frac{1}{20}(64-(4)) = 3$.[7] This gives us the payoff we need. Since $3 > 2$, the bureaucrat should fix the bridge.

7.3 COMPUTING INTEGRALS

We now know how to find the area under the curve in theory, but thus far doing so in practice has mostly involved trying to guess integrals from derivatives. Here we present rules for integration akin to those for differentiation. In fact, they are more than merely akin—they follow closely from the rules for differentiation, and we discuss them more or less in the same order, though we discuss integrals of functions first in order to get examples for the more general rules later. Since most of these rules apply for both definite and indefinite integrals, in our rules we refer generically to "integrals" and use their indefinite form, except for rules related to the bounds on the definite integral.

We start by noting that the fundamental theorem of calculus implies a couple of useful properties of the integral. Since $\int_a^b f(x)dx = F(b) - F(a)$, it must also

[7]Note that this is the midpoint of the range of the distribution. This is not an accident; since the uniform distribution places equal weight on all outcomes, the expected outcome is the one in the middle. Once you've read Chapter 11, you can come back here and see that out assertion was correct.

be true that $\int_b^a f(x)dx = F(a) - F(b)$. In words, if you flip the bounds of the integral, you switch the sign of the answer. In short, $\int_a^b f(x)dx = -\int_b^a f(x)dx$. Along these same lines, since $F(a) - F(a) = 0$, $\int_a^a f(x)dx = 0$. We can also split up the bounds of the integral. If $c \in [a, b]$, then $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$. Finally, note that the bounds need not be constants—they may be functions as well. A common way to write a cumulative distribution function, for example, is $F(x) = \int_{-\infty}^{x} f(t)dt$. All that must be true is that the bounds on the integrals should not contain the variable as that over which you are integrating. In the case of the cumulative distribution function just provided, that means we shouldn't have a t in the bounds of the integral because we are integrating over it.

7.3.1 Polynomials and Powers

Recall from the previous chapter that $\frac{dx^n}{dx} = nx^{n-1}$. And recall from the previous sections of this chapter that differentiating the antiderivative must produce the original function. In this case, we'll call the original function $f(x) = x^n$. If $F(x)$ is the antiderivative, we need $\frac{dF(x)}{dx} = x^n$. How do we get this? The easiest way is to use what we know already. If we call $g(x) = nx^{n-1}$, then the definition of the antiderivative implies that $G(x) = x^n$. We can get $G(x)$ from $g(x)$ by first increasing the exponent in $g(x)$ by 1, and then dividing by this new exponent. Doing this transforms the $n-1$ in the exponent of $g(x)$ to n, and then eliminates the n out front. Since this is true for any n, we can use it as our general rule for powers of x. In symbols, $\int x^n dx = \frac{x^{n+1}}{n+1} + C$ if $n \neq -1$. Differentiating the RHS yields x^n, which is the original function in the integral on the LHS, so our rule checks out.

Note a few fine details, though. First, always remember the constant of integration, C! The indefinite integral is specified only up to a constant. Second, this holds only for $n \neq -1$ for the reasons we discussed in the example above: the derivative of x^0 is 0, not x^{-1}, so the antiderivative of x^{-1} can't be a multiple of x^0. As shown in that example, for this case we have $\int x^{-1}dx = \ln|x| + C$.[8]

Without the general rules for integration the examples we offer in this chapter must be kept pretty simple. But we can integrate a wide range of functions with this rule alone. For example, $\int x^3 dx = \frac{x^4}{4} + C$, $\int \frac{1}{x^5}dx = \frac{-1}{4x^4}$, and $\int \sqrt{x}dx = \frac{2}{3}x^{\frac{3}{2}}$.

7.3.2 Exponentials

In the previous chapter we saw that $\frac{de^x}{dx} = e^x$. Since there is no difference between the function and its derivative, it must be the case that the the antideriva-

[8]If you are worried about losing track of these rules, remember that we summarize them all in Section 4. Also note that, while integrals are in general more complicated to compute than derivatives, you will also most likely be doing fewer of them in typical political science applications.

tive is the same as the function as well, since otherwise differentiating it would produce something different from the function. Consequently, $\int e^x dx = e^x + C$. The more general a^x behaves similarly. Since $\frac{a^x}{dx} = (\ln(a))a^x$, the antiderivative must also have an a^x in it for the same reason as for e^x. That leaves only $\ln(a)$. Since we need the antiderivative to differentiate to a^x, and since differentiating multiplies by $\ln(a)$, dividing by this same factor should do the trick. So, $\int a^x dx = \frac{a^x}{\ln(a)} + C$. As there are literally no examples we could do right now that would be interesting here, we move on to the next section.

7.3.3 Logarithms

While integrals of exponentials are common in political science, in large part because the probability distribution function for the normal distribution can be represented by an exponential function, as we'll see in Chapter 11, integrals of logarithms are less so. Consequently we will spend almost no time on them and offer no examples. However, in the interest of completeness and benefit to those in game theory whose utility functions are logarithms, we offer the integral of a logarithm.

To get the antiderivative of $f(x) = \ln(x)$, we must guess what function differentiated produces a natural log. The problem is that we haven't seen any of those, so we're going to have to get creative. We might not be able to use a basic expression, but if we include a log in a product, then at least one of the terms in the derivative of the product will still have a log in it. Then all that will be left is to eliminate the other term. Let's try the easiest product, $x\ln(x)$. The derivative of this, by the product rule, is $\ln(x) + \frac{x}{x} = 1 + \ln(x)$. That's close; all we need to do is get rid of the one. Subtracting x from $x\ln(x)$ will do this since its derivative is one. Thus, $\int \ln(x)dx = x\ln(x) - x + C$. For the more general $\log_a$ we have to deal with the $\ln(a)$ that appears when it is differentiated. Thus we need to divide both terms by $\ln(a)$, so that $\int \log_a(x)dx = \frac{x\ln(x)-x}{\ln(a)} + C$.

7.3.4 Other Functions

Unlike with the derivative, we cannot be so casual with integrating more complex functions or piecewise functions. In general, for our purposes we'll want the functions we're integrating to be continuous (and sometimes differentiable too). Piecewise functions can be handled, however, particularly for definite integrals, by splitting up the integral using the rules on its bounds given above. Consider the example of the previous chapter

$$f(x) = \begin{cases} -(x-2)^2 & : x \leq 2, \\ \ln(x-2) & : x > 2. \end{cases}$$

We could integrate this from $a \leq 2$ to $b > 2$ by splitting it up: $\int_a^b f(x)dx = \int_a^2 (-(x-2)^2)dx + \int_2^b \ln(x-2)dx$.

There are also integrals for trigonometric functions that may sometimes be useful. We list these for completeness: $\int \sin(x)dx = -\cos(x) + C$, $\int \cos(x)dx = \sin(x) + C$, $\int \tan(x)dx = -\ln(|\cos(x)|) + C$.

7.3.5 The Integral Is *Also* a Linear Operator

We saw that the derivative is a linear operator in the previous chapter. This means that $(af + bg)' = af' + bg'$ for functions f and g and constants a and b. The fact that the derivative is linear means that when one differentiates an antiderivative to obtain the original function, each constituent expression in the antiderivative corresponds to one expression in the function. So the integral will also be a linear operator.[9]

More formally, to say that the integral is also a linear operator is to say that $\int(af(x) + bg(x))dx = a\int f(x)dx + b\int g(x)dx$. To show this we need to use the definition of the antiderivative. The proof follows from this definition and the linearity of the derivative, which we showed in the previous chapter.

By the definition of an antiderivative, $y(x) = \int \frac{dy}{dx}dx$ and $z(x) = \int \frac{dz}{dx}dx$. We multiply both sides of the first equation by $a \neq 0$ and both sides of the second by $b \neq 0$, and then add them to get equation (7.1):

$$ay + bz = a\int \frac{dy}{dx}dx + b\int \frac{dz}{dx}dx. \tag{7.1}$$

Now start with the fact that the derivative is a linear operator to get $\frac{d(ay+bz)}{dx} = a\frac{dy}{dx} + b\frac{dz}{dx}$. Next integrate both sides to get

$$\int \frac{d(ay+bz)}{dx}dx = \int \left(a\frac{dy}{dx} + b\frac{dz}{dx}\right)dx.$$

Finally, use the definition of the antiderivative on the LHS to get equation (7.2):

$$ay + bz = \int \left(a\frac{dy}{dx} + b\frac{dz}{dx}\right)dx. \tag{7.2}$$

Compare (7.1) and (7.2). They have the same LHS, so their RHS must also be equal. Therefore, $\int \left(a\frac{dy}{dx} + b\frac{dz}{dx}\right)dx = a\int \frac{dy}{dx}dx + b\int \frac{dz}{dx}dx$. Since the derivatives of y and z are just functions, let $f = \frac{dy}{dx}$ and $g = \frac{dz}{dx}$. Making this substitution yields $\int(af(x) + bg(x))dx = a\int f(x)dx + b\int g(x)dx$, which is what we wanted.

Like each of the three general rules of integration we offer, this rule is used to simplify integrals, turning them into something that looks like the simpler functions above. We saw this when we computed the integral of $f(x) = 1 + 2x + x^2$: we treated each term in the sum separately. Examples utilizing the linearity of the integral are of this form. We offer a few here.

[9] We are ignoring issues with the constant of integration here.

All polynomials can be tackled with this rule. For instance, consider $f(x) = 4x^5 + 2x^2 + 5$. Linearity implies we can treat each term separately. The integral $\int x^5 dx = \frac{x^6}{6} + C$, so $4\int x^5 dx = \frac{2x^6}{3} + C$. The integral $\int x^2 dx = \frac{x^3}{3} + C$, so $2\int x^2 dx = \frac{2x^3}{3} + C$. The integral $\int x^0 dx = x + C$ so $5\int x^0 dx = 5x + C$. Combining all these yields the answer: $\int (4x^5 + 2x^2 + 5)dx = \frac{2x^6}{3} + \frac{x^3}{3} + 5x + C$, where we've combined all the arbitrary constants into one equally arbitrary constant. In a similar fashion, $\int (10x^6 - 4x^4 + \frac{1}{x^2})dx = \frac{10x^7}{7} - \frac{4x^5}{5} - \frac{1}{x} + C$, where all we've done is to consider each term separately and use the rule for powers of x. Nor are we limited to polynomials: $\int (5x^2 + e^x)dx = \frac{5x^3}{3} + e^x + C$.[10]

7.3.6 Integration by Substitution

We start describing the technique of substitution by taking a closer look at the notation in the integral $\int f(x)dx$. What is x here? Could we replace it by, say, u, to get $\int f(u)du$? The answer to this is yes. The variable of integration itself has no meaning. This is no different from saying the sum $\sum_{i=1}^{N} x_i$ is the same as the sum $\sum_{k=1}^{N} x_k$. The variables k and i are just indices in the sums, and x and u are just "infinitesimal indices" in the integral.

We bring this up because the fungibility of the variable of integration signals that we should have the ability to change things within the integral, as long as we are careful. Clearly, we can exchange the *name* of the variable, turning an x into a u. But what if we turn x into an entirely new function? For example, let $x = g(u)$. What then?

Well, if $x = g(u)$, then $f(x) = f(g(u))$, which is a composite function. Integration by substitution says that

$$\int_a^b f(g(u))g'(u)du = \int_{g(a)}^{g(b)} f(x)dx.^{11}$$

In words, we can change variables to an entirely different function if it helps us to compute the integral. We'll see why we'd want to do such a thing in the examples below in this subsection, but first we show why this equality holds.

We use a combination of the definition of an antiderivative, the fundamental theorem of calculus, and the chain rule from differentiation.[12] First, let $x = g(u)$, so that $f(x) = f(g(u))$. Next, define the antiderivative $F(x)$ in the (now) usual fashion. We can compose this with $g(u)$ as well: $F(g(u))$. The chain rule says that $(F(g(u)))' = F'(g(u))g'(u) = f(g(u))g'(u)$, where in the last step we have

[10] We went through these examples pretty quickly, and advise you to work them out on your own with more care.

[11] We could also have written $\int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)du = \int_a^b f(x)dx$, depending on whether we wanted to define the bounds of the integral as measuring differences in u values or x values. The way given in the text makes it slightly simpler to prove.

[12] As you should be used to by now, we skip some steps that don't enhance intuition. In this case, we won't show that the integrals on each side of the substitution rule exist, just that they are equal, as the rule states.

used the fact that the derivative of the antiderivative is the original function $f(x)$.[13]

The fundamental theorem of calculus says that $\int_a^b (F(g(u)))'du = F(g(b)) - F(g(a))$. We can plug in $f(g(u))g'(u)$ to the LHS integral to get $\int_a^b f(g(u))g'(u)du = F(g(b)) - F(g(a))$. We can again use the fundamental theorem of calculus on the RHS to give us $F(g(b)) - F(g(a)) = \int_{g(a)}^{g(b)} f(x)dx$. Plugging this into the RHS of the previous equality gives our desired result, $\int_a^b f(g(u))g'(u)du = \int_{g(a)}^{g(b)} f(x)dx$.

This likely seemed a bit abstract, so let's break substitution down into a useful technique, and then try it out with some examples.[14] Integration by substitution is attempted whenever the integral contains a composite function that one cannot integrate easily. In this sense it is used similarly to the chain rule in differentiation. The big difference is that whereas the chain rule will provide the derivative in most cases, substitution will often fail. Integrals, in general, are less amenable than derivatives, and do not always produce straightforward answers. Sometimes they do, though, so it's worth trying substitution. A couple of examples that are quite important to probability theory will help.

First consider the function $f(x) = \frac{1}{2\pi} x e^{-\frac{x^2}{2}}$. This function is the probability distribution function of a standard normal distribution, multiplied by x. Its integral is $\int \left(\frac{1}{2\pi} x e^{-\frac{x^2}{2}}\right) dx$, which happens to be the expected value of x in a standard normal distribution. We'll see this more in Chapter 11. This integral looks complicated, but it turns out that it reduces quite easily via substitution. The composite function here is $e^{-\frac{x^2}{2}}$, which we don't know how to integrate. But we do know how to integrate e^u, so let's set $u = g(x) = -\frac{x^2}{2}$ and see what happens. We can use the power rule to see that $g'(x) = -x$, which means we can rewrite the integral as $\int \left(\frac{-1}{2\pi} g'(x) e^{g(x)}\right) dx$. Integration by substitution implies that this integral is the same as $\int \left(\frac{-1}{2\pi} e^u\right) du$, which just equals $\frac{-1}{2\pi} e^u + C = \frac{-1}{2\pi} e^{-\frac{x^2}{2}} + C$.[15] So we've completed our integral.

Now instead consider the function $f(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$, which is the probability distribution function of a standard normal distribution. The integral of this is $\int \left(\frac{1}{2\pi} e^{-\frac{x^2}{2}}\right) dx$, and if we took the definite integral from $-\infty$ to ∞ we would get 1, which is true for all probability distribution functions. But how would we take that integral? We can't use our substitution trick even though we still have the same composite function, because there is no longer an x there to satisfy the need for a $g'(x)$ in the integrand. So we're stuck.[16]

[13] If you don't see how the chain rule applies here, set $v = g(u)$. Then $(F(g(u)))' = \frac{F(v)}{dv}\frac{dv}{du} = \frac{F(v)}{dv}\frac{dg(u)}{du} = F'(v)g'(u) = F'(g(u))g'(u)$.

[14] This technique is rough, but it works.

[15] If this is confusing, switch u and x in this example to see that we used integration by substitution here.

[16] It turns out that one can do this integral with tools from complex analysis, but this is considerably beyond the scope of this book.

The moral is that for our purposes, substitution is feasible to try, and as we show, it is not terribly challenging to determine whether or not it will work. First you identify a composite function $f(g(x))$. Then you substitute $u = g(x)$ for the inner function. If there is some multiple of $g'(x)$ in the integrand, you can use substitution. If there is not, you cannot, or at least not so easily.[17] To wit: $\int 3x^2 e^{x^3} dx = \int e^u du = e^{x^3} + C$, since $u = g(x) = x^3$ and $g'(x) = 3x^2$, and $\int x^2 e^{x^3} dx = \frac{1}{3} e^{x^3} + C$ because x^2 is a multiple of $3x^2 = g'(x)$,[18] but neither $\int 3(x^2+1)e^{x^3} dx$ nor $\int 3xe^{x^3} dx$, for example, is integrable in this way.

Another way to see how to perform integration by substitution returns to our discussion at the beginning of this subsection about changing variables of integration. Rather than view the change from x to $u = g(x)$ as merely substituting in a function, one can view it as a change in variables from x to u, where $g(x)$ specifies the relation between them. One can use this relation to link dx to du as well: $du = \frac{dg(x)}{dx} dx$, so $dx = (g'(x))^{-1} du$.[19] Then $\int g'(x) f(g(x)) dx = \int g'(x) f(u) (g'(x))^{-1} du = \int f(u) du$.[20] This way of going about substitution is equivalent to our first method but may be easier procedurally, as it makes clearer that the goal is to eliminate any independent x in the integrand when substituting. For example, to compute $\int x^4 e^{x^5} dx$, we set $u = g(x) = x^5$, so $dx = (g'(x))^{-1} du = \frac{1}{5x^4} dx$, and thus $\int x^4 e^{x^5} dx = \int x^4 e^{x^5} \frac{1}{5x^4} du = \frac{1}{5} \int e^u du$. This is just $\frac{1}{5} e^u + C = \frac{1}{5} e^{x^5} + C$.

Before moving on, let us inject a note on the definite integral. These examples all used indefinite integrals but would have been equally valid as definite integrals (recall that the distinction is between a limited range of x and a function valid over the full range of x). They would have differed only in the presence of the bounds on the integral in the case of definite integrals. With substitution, however, one must take care with the bounds. Specifically, one must convert the bounds as the substitution rule would dictate. For example, let's say we had $\int_1^3 3x^2 e^{x^3} dx$, i.e, our recent example made into a definite integral. With $u = g(x) = x^3$, the integral becomes $\int_{g(1)}^{g(3)} e^u du = \int_1^{27} e^u du = e^{27} - e^1$. If this seems confusing, note that one can always check this by putting the original variable x back in the answer, and then using the original bounds. For our case, $e^u = e^{x^3}$, so $\int_1^3 3x^2 e^{x^3} dx = e^u|_1^{27} = e^{x^3}|_1^3 = e^{3^3} - e^{1^3} = e^{27} - e^1$, as before.

[17] This is of course a simplification, and there are many ways to be clever about this technique. As you become more practiced with integration, you will have plenty of opportunities for such cleverness. Trigonometric identities can figure heavily. We are just providing a guideline that will work in many common situations in political science.

[18] The $\frac{1}{3}$ in $\frac{1}{3}e^{x^3} + C$ arises because you have to multiply by $1 = \frac{3}{3}$ to get the $3x^2$ you need to use substitution in this case.

[19] An easy way to remember this is that $\frac{du}{dx} = \frac{dg(x)}{dx}$; just "multiply" both sides by the dx from the LHS to get the required equation. Don't do this in general, though; this is just a mnemonic!

[20] We could write this more generally: for any function $f(x)$, $\int f(x) dx = \int f(x)(g'(x))^{-1} du = \int f(u) du$ if $u = g(x)$. Of course, if $(g'(x))^{-1} f(x)$ does not simplify into a nice $f(u)$—which it won't if one can't get rid of any dependence of $f(u)$ on x—then substitution won't work. This simplification occurs when the $(g'(x))^{-1}$ cancels with a part of $f(x)$. The more specific way we have written substitution in the main text makes the need for cancellation clearer.

7.3.7 Integration by Parts

The third and final general rule we offer is integration by parts. Its analogue in differentiation is the product rule, and it is used to integrate products of functions. First we'll state the rule, then explain and prove it, and finally offer some examples for how to use it.

Integration by parts states that

$$\int f(x) g'(x) dx = f(x) g(x) - \int f'(x) g(x) dx.$$[21]

Basically, the rule allows you to change around your functions. If you can't integrate $f(x)g'(x)$ but you can integrate $g'(x)$ and $f'(x)g(x)$, it's very useful. Practically, this occurs when $g'(x)$ is something like e^x that does not get more complex when you integrate it, while $f(x)$ is something like x that simplifies when you differentiate it. We'll see this shortly in the examples. First, the proof, which is straightforward.

The product rule states that $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$. Integrating both sides and using the definition of the antiderivative on the LHS and the fact that the derivative is linear on the RHS yields $f(x)g(x) = \int f'(x)g(x)dx + \int f(x)g'(x)dx$. Moving $\int f'(x)g(x)dx$ to the other side of the equation produces the rule.

Now for why we'd use it. Let's start with $\int xe^x dx$. We can't do this directly, nor can we use substitution, because the only compound function available is xe^x and we don't have the derivative of this present so as to allow substitution. So we'll try integration by parts. This is promising, as we can integrate e^x in a way that does not complicate it further, and the derivative of x is just 1. We set $f(x) = x$, $g'(x) = e^x$, and the rule gives us $\int xe^x dx = xe^x - \int e^x dx = xe^x - e^x + C$. We could in theory do this multiple times for more complicated functions. For example, $\int x^2 e^x dx = x^2 e^x - \int 2xe^x dx + C$. We can then use our previous example of integration by parts to complete the problem: $\int x^2 e^x dx = x^2 e^x - 2(xe^x - e^x) + C$.

Most examples you'll see in political science that use integration by parts are of this form. For instance, the example we just did is related to taking the expected value of a function, or finding the mean, given an exponential distribution, which you'll meet in Chapter 11. Other cases you might see, though far less often, would involve a power of x, such as x^2 or x^3, and an integrable root of x, such as $(x+1)^{-\frac{3}{2}}$. For example, $\int x(x+1)^{-\frac{3}{2}} dx$. To solve it, let $f(x) = x$, and $g'(x) = (x+1)^{-\frac{3}{2}}$. We need to know $f'(x)$ and $g(x)$ to use integration by

[21] You will often see this as $\int u dv = uv - \int v du$, where $u = f(x)$, $v = g(x)$, and $du = f'(x)dx$ and $dv = g'(x)dx$ are differentials. These mean the same thing, so we'll stick with the one that is more familiar.

parts. Note that $f'(x) = 1$ and $g(x) = \int (x+1)^{-\frac{3}{2}}dx$. This requires substitution to solve, but a relatively simple one. Set $u = h(x) = x + 1$. Then $h'(x) = 1$, which is automatically in the integral already. So we rewrite the integral as $g(u) = \int u^{-\frac{3}{2}}du$. By the power rule, this is $-2u^{-\frac{1}{2}} + C = -2(x+1)^{-\frac{1}{2}} + C$, after plugging back in for u. Now we plug into the integration by parts rule: $\int x(x+1)^{-\frac{3}{2}}dx = -2x(x+1)^{-\frac{1}{2}} + 2\int (x+1)^{-\frac{1}{2}}dx$. The last integral can be solved by again using the substitution $v = (x+1)$, so that $\int (x+1)^{-\frac{1}{2}}dx = \int v^{-\frac{1}{2}}dv = 2\sqrt{v} + C = 2\sqrt{x+1} + C$. Plugging this back into the integration by parts rule yields the answer: $\int x(x+1)^{-\frac{3}{2}}dx = -2x(x+1)^{-\frac{1}{2}} + 4\sqrt{x+1} + C$.[22]

As with substitution, we have presented integration by parts for the indefinite integral. It is not difficult to switch to the definite integral, though:

$$\int_a^b f(x)g'(x)dx = (f(x)g(x))|_a^b - \int_a^b f'(x)g(x)dx.$$

7.3.8 Why Should I Care?

Though you will largely rely on tables and statistical software to compute integrals of cumulative distribution functions, you will have cause to compute integrals should you choose to delve further into the underpinnings of statistical methodology. This will become clearer in Part III of the book. Further, you will have ample opportunity to take integrals in formal theory when computing either expected values or expected utilities, which are foundational concepts.

7.4 RULES OF INTEGRATION

For convenience, we provide a summary here of the **rules of integration** we have discussed in this chapter. Though not necessary, to make this easier we assume that f and g are both differentiable and integrable functions, and a, b, and C are constants. Procedurally, it usually helps in computing the integrals of more complex functions to first check to see if the linear rule is sufficient; second, attempt substitution; and third, turn to integration by parts if neither of the previous two rules helps. You should also be aware that, unlike with differentiation, there will be times when none of these three techniques yields progress toward an answer. There are integrals you will simply not be able to do, either because they require specialized techniques (as with $e^{-\frac{x}{2}}$) or because they simply are not integrable. (You won't find any of those in our exercises, though.) Finally, though we would expect that in many cases you'd be able to reference a list like that in Table 7.1 should you need to compute an integral by hand, we note that despite the length of the list in Table 7.1, all the rules provided may be derived (more or less) from the corresponding rules for differentiation. Thus, as noted before, learning calculus need not be a matter of memorizing a large number of disconnected rules.

[22]Don't worry if this isn't immediately clear. It's a relatively complicated integral. That said, for this reason it is a good one to practice on.

Table 7.1: List of Rules of Integration

Fundamental theorem of calculus	$\int_a^b f(x)dx = F(b) - F(a)$
Rules for bounds	$\int_a^b f(x)dx = -\int_b^a f(x)dx$ $\int_a^a f(x)dx = 0$ $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$ for $c \in [a, b]$
Linear rule	$\int (af(x) + bg(x))dx = a\int f(x)dx + b\int g(x)dx$
Integration by substitution	$\int_a^b f(g(u))g'(u)du = \int_{g(a)}^{g(b)} f(x)dx$
Integration by parts	$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$
Power rule 1	$\int x^n dx = \frac{x^{n+1}}{n+1} + C$ if $n \neq -1$
Power rule 2	$\int x^{-1}dx = \ln\|x\| + C$
Exponential rule 1	$\int e^x dx = e^x + C$
Exponential rule 2	$\int a^x dx = \frac{a^x}{\ln(a)} + C$
Logarithm rule 1	$\int \ln(x)dx = x\ln(x) - x + C$
Logarithm rule 2	$\int \log_a(x)dx = \frac{x\ln(x)-x}{\ln(a)} + C$
Trigonometric rules	$\int \sin(x)dx = -\cos(x) + C$ $\int \cos(x)dx = \sin(x) + C$ $\int \tan(x)dx = -\ln(\|\cos(x)\|) + C$
Piecewise rules	Split definite integral into corresponding pieces

7.5 SUMMARY

We introduced the integral as a method for calculating the area under a curve. We also defined the indefinite integral as the antiderivative and the definite integral as the antiderivative evaluated over a range of values. The key points are the integral's relationship to the derivative and understanding why it is a tool we can use to measure the area under curves.

We also provided a variety of rules used to compute integrals. We note that, in general, computing derivatives is considerably easier than computing integrals. Although we have provided tools to compute most derivatives you'll encounter, you may quite well find you need to compute integrals that simply are not amenable to computation using the tools you have in your toolkit. For example, you may need to compute the expected utility over a normal distribution, or even simply the integral of the normal distribution's probability distribution function itself. If you should find yourselves in such a scenario, help is readily available. Searching on the phrase "list of integrals" on the Internet should get you far to start, as many of the integrals that you might want to do have been fully or partially computed by others, either via more advanced analytic techniques or by numerical analysis.

7.6 EXERCISES

1. Visit `http://math.furman.edu/~dcs/java/NumericalIntegration.html` and approximate the area under the curve using different rules as described on the page. What does this applet demonstrate?

2. Visit `http://math.furman.edu/~dcs/java/ftc.html`, and drag the red dot (as described on the page). What does this applet demonstrate?

3. Integrate the following derivatives to find y:

   a) $\frac{dy}{dx} = 4x + 3$.
   b) $\frac{dy}{dx} = 3x^2$.
   c) $\frac{dy}{dx} = -2x + 3 - 4x^3$.
   d) $\frac{dy}{dx} = -1$.
   e) $\frac{dy}{dx} = -3 + 4x$.
   f) $\frac{dy}{dx} = 5x^4 - x - 4$.
   g) $\frac{dy}{dx} = 4x^4 + 3x^2$.
   h) $\frac{dy}{dx} = 5x^5$.
   i) $\frac{dy}{dx} = 4x^4 + 3x^3 + 2x^2 + x + 1$.
   j) $\frac{dy}{dx} = 3x^3 - 4x^2 + 5x - 6$.
   k) $\frac{dy}{dx} = x^{-1} + 3x^2$.
   l) $\frac{dy}{dx} = e^{5x}$.
   m) $\frac{dy}{dx} = 2e^{5x}$.
   n) $\frac{dy}{dx} = (20x + 2)e^{5x^2+x}$.
   o) $\frac{dy}{dx} = \ln(3x)$.
   p) $\frac{dy}{dx} = \ln(x^2)$.

4. Take the derivative of the answer to each of the problems in the previous question to check your work.

5. Which of the options below best describes $\int_a^b \frac{dy}{dx} dx$?

   a) It is the indefinite integral of the derivative of y.
   b) It is the area from y to x for the function defined in a and b.

   c) It is the integral of the derivative of y with respect to x over the range a to b.

6. Compute the following integrals:

   a) $\int (a_n x^n + a_{n-1}x^{n-1} \ldots + a_0)dx$. You get a bonus for expressing the integral as a sum.
   b) $\int (3x^{3/2} - 2x^{-5/4} + 4^x)dx$.
   c) $\int_1^{16} (5x^{3/2} - 2x^{-5/4})dx$.
   d) $\int (-\frac{1}{x} \ln(\frac{1}{x}))dx$.
   e) $\int (xe^{3x^2+1})dx$.
   f) $\int (x^2(x^3 + 15)^{3/2})dx$.
   g) $\int (\frac{12x^2-16x+20}{x^3-2x^2+5x})dx$.
   h) $\int (xe^x)dx$.
   i) $\int_2^4 (3x^2 + x + 5)dx$.
   j) $\int_2^2 (3x^2 + x + 5)dx$.
   k) $\int_2^4 (3x^4 + 2x^3 + x^2 + x + 1)dx + \int_4^6 (3x^4 + 2x^3 + x^2 + x + 1)dx$.
   l) $\int_2^6 (3x^4 + 2x^3 + x^2 + x + 1)dx$.