Course 1 Key Concepts

Calculus

Lesson 4

Integration and the Definite Integral

Overview

In this lesson, we introduce the concept of the definite integral. From a simple geometric point of view, the definite integral is equal to the area under the graph of the function $y = f(x)$. As we'll see, definite integrals are calculated using anti-derivatives. The function $F(x)$ is an anti-derivative of $f(x)$ if $F'(x) = f(x)$.

We'll start by looking at the interpretation and calculation of definite integrals, and then we'll study several important rules that will help us to integrate (*ie* calculate the anti-derivative of) common functions. Then we'll see how these techniques can be used to solve problems, including basic differential equations and motion problems.

BPP Learning Objectives

This lesson covers the following BPP learning objectives:

(C10) *Calculate the anti-derivative of common functions using basic anti-derivative rules, the method of integration by parts, and the method of substitution.*

(C11) *Calculate the area under a function over the interval [a,b] using definite integrals.*

(C12) *Understand numerical methods used to calculate the area under a function, including Riemann sums.*

(C13) *Solve basic differential equations.*

(C14) *Using differentiation and integration to solve basic motion problems.*

Theory

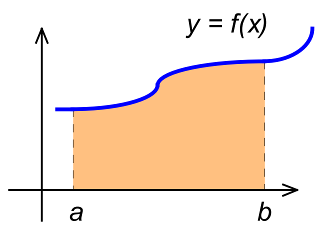
Area under the graph of a function

Integration is most often used to measure the area under a graph, so we'll start with a geometric definition:

Definition

Suppose that $y = f(x)$ is a non-negative continuous function on the interval $[a,b]$.

The **definite integral** of this function over the interval $[a,b]$ is denoted $\int_a^b f(x)dx$ and is equal to the area of the region that is above the interval and beneath the graph of $y = f(x)$.

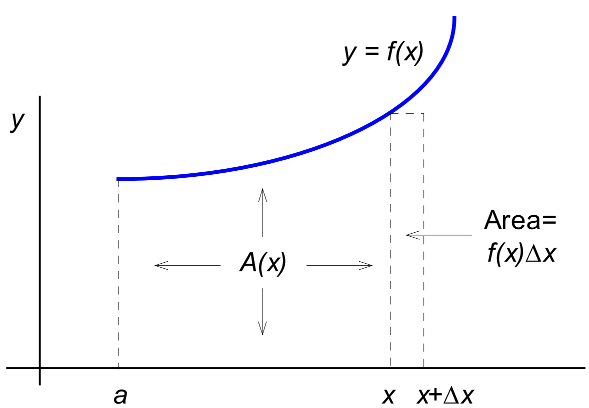


But how do we calculate the definite integral? The Fundamental Theorem of Calculus is of great importance here.

The Fundamental Theorem of Calculus (FTC)

Suppose that $y = f(x)$ is a non-negative continuous function on the interval $[a,b]$. Let $A(x)$ be the area beneath the graph of $y = f(x)$ between the fixed point a and a variable x.

Let's calculate the derivative of $A(x)$, *ie* the rate at which the area increases as *x* increases.



The area of the rectangle, $f(x)\Delta x$, is approximately the area under the graph above the interval $[x, x+\Delta x]$, which can be thought of as $A(x+\Delta x) - A(x)$. The approximation improves as Δx gets smaller.

Theory

Using the definition of the derivative (see Calculus Lesson 2):

$$A'(x) = \lim_{\Delta x \to 0} \frac{A(x+\Delta x) - A(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x)\Delta x}{\Delta x} = f(x)$$

This relation states that $f(x)$ is the derivative of $A(x)$. Alternatively, $A(x)$ is an **anti-derivative** of $y = f(x)$.

Why is $A(x)$ an anti-derivative of $f(x)$ rather than the anti-derivative? For the simple reason that any function of the type $A(x)+c$ is also an anti-derivative of $f(x)$ since:

$$(A(x)+c)' = A'(x) + 0 = A'(x) = f(x)$$

You may also see the notation $F(x) = \int f(x)dx$ used to represent the fact that $F'(x) = f(x)$.

Let's summarize the discussion by stating the Fundamental Theorem of Calculus (FTC):

Fundamental Theorem of Calculus (FTC)

Suppose that $y = f(x)$ is a non-negative continuous function on $[a,b]$.

If $F(x)$ is any anti-derivative of $f(x)$, then the area under the graph $y = f(x)$ and above the interval $[a,b]$, $\int_a^b f(x)dx$, can be calculated as $F(b) - F(a)$.

Note: Any anti-derivative $F(x)$ of $f(x)$ is of the form $A(x)+c$, so:

$$F(b) - F(a) = (A(b)+c) - (A(a)+c) = A(b) - A(a)$$

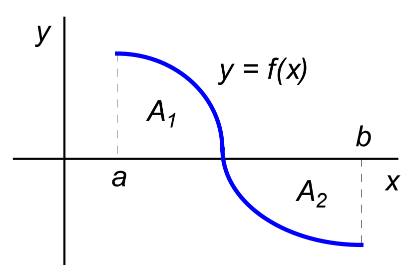
The difference $F(b) - F(a)$ is usually denoted by $F(x)\Big|_a^b$.

So the conclusion of the above theorem is often written in the form:

$$\int_a^b f(x)dx = F(x)\Big|_a^b \quad \text{where } F'(x) = f(x).$$

Suppose that we drop the assumption that the function is non-negative.

In this case, the calculation of the definite integral treats areas below the x-axis as being negative valued, since if $(x, f(x))$ is below the x-axis, then $f(x)$ is negative



In the figure on the right, we would have:

$$\int_a^b f(x)dx = F(b) - F(a) = A_1 - A_2.$$

Theory

Finding anti-derivatives

So, how do we find the anti-derivative of a function? We can identify a set of rules from the work we did on differentiation (but in reverse):

Anti-derivative rules – Summary

General rules

(i) $\int c\,f(x)\,dx = c\int f(x)\,dx$

(ii) $\int f(x) \pm g(x)\,dx = \int f(x)\,dx \pm \int g(x)\,dx$

Anti-derivatives of common functions

(i) $\int x^p\,dx = \dfrac{x^{p+1}}{p+1} + c$ if $p \neq -1$ $\int x^{-1}\,dx = \ln(x) + c$

(ii) $\int \sin(x)\,dx = -\cos(x) + c$ $\int \cos(x)\,dx = \sin(x) + c$

(iii) $\int e^x\,dx = e^x + c$

The anti-derivative counterpart to the product rule for differentiation is known as the method of integration by parts.

The method of integration by parts

If u and v are functions of x, then:

$$\int u\,dv = uv - \int v\,du$$

The idea of this method is to split the function being integrated into 2 parts in such a way that a difficult anti-derivative is traded for two easier ones.

For example, if $f(x) = xe^x$, then we can find the anti-derivative by setting $u = x$ and $v = e^x$. We have:

$$\frac{du}{dx} = 1 \Rightarrow du = dx \quad \text{and} \quad \frac{dv}{dx} = e^x \Rightarrow dv = e^x dx$$

$$\int u\,dv = uv - \int v\,du \Rightarrow \int \underbrace{x}_{u}\,\underbrace{e^x\,dx}_{dv} = \underbrace{x}_{u}\,\underbrace{e^x}_{v} - \int \underbrace{e^x}_{v}\,\underbrace{dx}_{du} = x\,e^x - e^x + c$$

Checking our work by differentiating gives:

$$\left(x\,e^x - e^x + c\right)' = \left(x\,e^x\right)' - \left(e^x\right)' + (c)' = \left(xe^x + e^x\right) - \left(e^x\right) + (0) = x\,e^x \quad \text{(as required)}$$

Theory

The anti-derivative counterpart to the chain rule for differentiation is known as the method of substitution.

The method of substitution

Suppose that $u = f(x)$ and $G'(x) = g(x)$. Then:

$$\int g(f(x))f'(x)dx = \int g(u)du = G(u) + c = G(f(x)) + c$$

This follows directly from the chain rule. If $G'(x) = g(x)$ then the chain rule can be written in the form:

$$\left(G(f(x))\right)' = G'(f(x))f'(x) = g(f(x))f'(x)$$

The anti-derivative counterpart of this formula is (as stated above):

$$G(f(x)) + c = \int g(f(x))\, f'(x)dx\,.$$

If we substitute $u = f(x)$ and $du = f'(x)dx$ into this relation we have the method of substitution.

For example, let's calculate $\int e^{ax}\, dx$

If we set $g(x) = e^x$ and $u = ax$, then:

$$du = a\; dx \Rightarrow dx = \frac{1}{a}du$$

The anti-derivative becomes:

$$\int e^{ax}\, dx = \int g(u)\frac{1}{a}du = \int e^u\,\frac{1}{a}\, du = \frac{1}{a}e^u + c = \frac{1}{a}e^{ax} + c$$

Checking our work by differentiating gives:

$$\left(\frac{1}{a}e^{ax} + c\right)' = \frac{1}{a}\left(e^{ax}\right)(a) + 0 = e^{ax} \qquad \text{(as required)}$$

You need to be able to use the method of integration by parts and the method of substitution efficiently in the Course 1 exam. As ever, practice makes perfect. The more problems that you do, the easier it will be to identify which method to use, and which substitutions to make in order to simplify the calculations.

Theory

When the anti-derivative cannot be calculated

The Fundamental Theorem of Calculus offers a way to calculate the area beneath a curve precisely, but to implement it we need to be able to calculate an anti-derivative, which can be difficult at times and impossible at others.

The alternative is to calculate the area using a **Riemann sum**.

This can be described as a 3-step process.

1. We divide the interval $[a, b]$ into n equal subintervals, each of length $(b-a)/n$.

2. For each subinterval, we construct a rectangle with a base equal to the length of the subinterval and the height equal to the value of the function at some point within the subinterval.

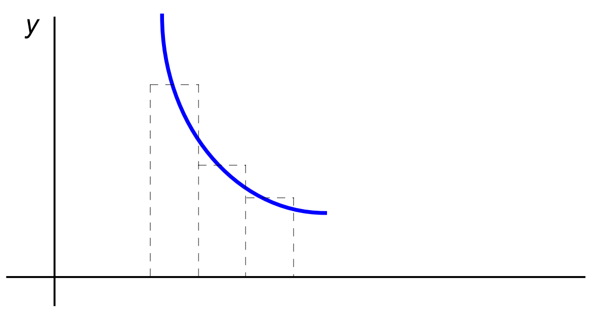
3. The area beneath the curve can be approximated as the total of the areas of the n rectangles. As the value of n increases, the better the approximation becomes.

In mathematical terms:

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} \left(\underbrace{x_i - x_{i-1}}_{base} \right) \underbrace{f(c_i)}_{height} \quad \text{where } x_i = a + i \times \frac{b-a}{n} \text{ and } x_{i-1} \le c_i \le x_i$$

The values c_i are often chosen as the left endpoints, midpoints, or right endpoints of the subintervals $[x_{i-1}, x_i]$.

Graphically (using midpoints):



If the limit as $n \to \infty$ is difficult to calculate, we can approximate the area with a Riemann sum by using a large n.

Numerical methods like these are used to calculate the value of definite integrals for which no closed form solution exists, including the probability distribution function for a standard normal distribution:

$$\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Theory

The Second Fundamental Theorem of Calculus

The discussion of the fact that $F'(x) = f(x)$ leads to the following formula that is sometimes referred to as the Second Fundamental Theorem of Calculus:

$$\frac{d}{dt}\left(\int_{g(t)}^{h(t)} f(x)dx\right) = \frac{d}{dt}\left(F\left(h(t)\right) - F\left(g(t)\right)\right) = f\left(h(t)\right)h'(t) - f\left(g(t)\right)g'(t)$$

Basic differential equations

An equation involving a function y and its derivative $\frac{dy}{dx}$ is called a **first order differential equation**.

The simplest type of first order differential equation is:

$$\frac{dy}{dx} = f(x)$$

Any solution to this equation is an anti-derivative of $f(x)$.

A slightly more complicated first order differential equation is:

$$\frac{dy}{dx} = k\,y$$

This equation says that y changes at a rate proportional to y. This model is commonly used for compound interest in finance, for radioactive decay in physics, and is sometimes used to model population growth over a short period of time.

If you recall that $\left(e^{kx}\right)' = k\,e^{kx}$, you will see that any function of the form $y = a\,e^{kx}$ is a solution of this differential equation. The constant a is sometimes called an **initial value** since:

$$y(0) = a\,e^{0} = a$$

The most complicated first order differential equation to appear on a Course 1 exam has been the limited compound growth model:

$$\frac{dy}{dx} = k\,y\,(L - y)$$

We'll look at this in more detail in Question 4.6 below.

Theory

The average value of a function

If $y = f(x)$ is continuous on the interval $[a,b]$, then the **average value** of the function over this interval is defined by:

$$\bar{y} = \frac{1}{b-a}\int_a^b f(x)dx$$

Note: This is the same idea as the expected value of the random variable $f(X)$ if X is uniformly distributed on the interval $[a,b]$.

For example, if $y = c$, the definition also results in $\bar{y} = c$. This is intuitive.

Now suppose that we have a linear function $y = cx + d$:

$$\bar{y} = \frac{1}{b-a}\int_a^b f(x)dx = \frac{1}{b-a}\int_a^b (cx+d)dx = \frac{0.5cx^2 + dx}{b-a}\Bigg|_a^b$$

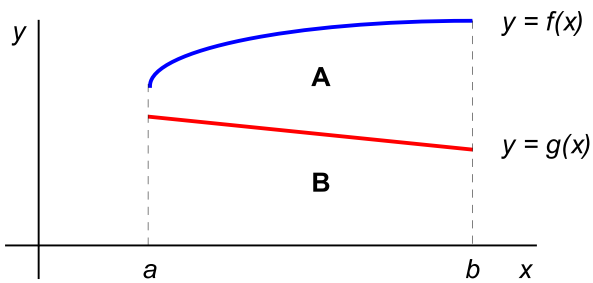
$$= \frac{0.5c\left(b^2 - a^2\right) + d(b-a)}{b-a} = c \times \left(\frac{b+a}{2}\right) + d$$

This should also be intuitive, since $f(x)$ increases linearly from $f(a) = ca + d$ to $f(b) = cb + d$, so the average function value should be mid-way between these two numbers. That is exactly the formula you see above for $\bar{y}$.

The area between two curves

Suppose that we have two continuous functions $y = f(x)$ and $y = g(x)$ defined on the interval $[a,b]$.

We want to be able to compute the area above the interval $[a,b]$ that is between the two curves. The simplest case (see the figure below) occurs when one of the functions is "always on top." The area between the curves is A.

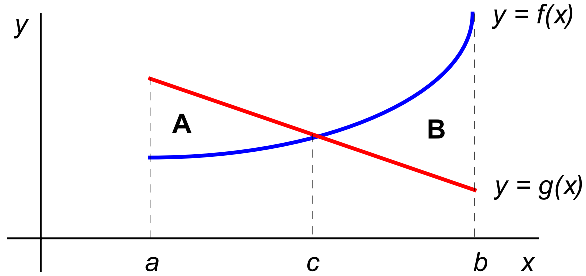


In this case, it is easy to see that we have:

$$A = (A+B) - B = \int_a^b f(x)dx - \int_a^b g(x)dx = \int_a^b \left(f(x) - g(x)\right)dx$$

Theory

Now consider the case where the two graphs cross at $x = c$ between $x = a$ and $x = b$ (figure below).



The area between the two curves in this picture is $A + B$. Each of the two pieces of area can be computed as earlier (when one of the two curves is always on top).

This idea leads to a general formula for area between curves that covers all cases: the definite integral of the absolute difference of the two functions over the interval $[a,b]$.

$$A + B = \int_a^c (g(x) - f(x))dx + \int_c^b (f(x) - g(x))dx = \int_a^c |g(x) - f(x)|dx + \int_c^b |f(x) - g(x)|dx$$

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

Motion problems

Suppose that an object starts in motion at time $t = a$. Suppose that $d(t)$ represents the distance traveled by the object by time $t > a$.

- The rate of change of distance with respect to time, $s(t) = d'(t)$, is called **speed**.
- The rate of change of speed with respect to time, $a(t) = s'(t) = d''(t)$, is called **acceleration**.

Due to the FTC, we have the following relations:

$$s(b) - s(a) = \int_a^b s'(t)dt = \int_a^b a(t)dt$$

$$d(b) - d(a) = \int_a^b d'(t)dt = \int_a^b s(t)dt$$

In a more general linear motion model the **direction** of the motion is incorporated. Let $x(t)$ be the position on the number line at time t. The **displacement** by time t is defined by $x(t) - x(a)$, the **velocity** at time t is $v(t) = x'(t)$, and the **acceleration** is $a(t) = v'(t) = x''(t)$. A positive velocity indicates that the object is moving to the right, while a negative velocity indicates motion to the left on the number line. The distance traveled by time $t = b$ is calculated as $\int_a^b |v(t)|dt$ if the motion begins at time $t = a$.

Worked examples

Example 4.1

Reference: May 2000, Question 15

In a certain town, the rate of deaths at time t due to a particular disease is modeled by:

$$\frac{10{,}000}{(t+3)^3} \quad \text{for } t \geq 0$$

What is the total number of deaths from this disease predicted by the model?

(A) 243
(B) 370
(C) 556
(D) 1,111
(E) 10,000

Solution

Let $D(t)$ be the total number of deaths by time t predicted by this model.

The *rate of deaths* is given by the derivative:

$$D'(t) = \frac{10{,}000}{(t+3)^3}$$

To find the total number of deaths at time t_1, we have:

$$D(t_1) = D(t_1) - D(0) = \int_0^{t_1} D'(t)\,dt$$

$$= \int_0^{t_1} 10{,}000(3+t)^{-3}\,dt = \left.\frac{10{,}000(3+t)^{-2}}{-2}\right|_0^{t_1}$$

$$= -\frac{5{,}000}{(3+t_1)^2} + \frac{5{,}000}{3^2}$$

The total number of deaths predicted by the model is the limiting value of $D(t_1)$ as $t_1 \to \infty$, *ie*:

$$\lim_{t_1 \to \infty} D(t_1) = \lim_{t_1 \to \infty} \left(-\frac{5{,}000}{(3+t_1)^2} + \frac{5{,}000}{3^2} \right)$$

$$= \frac{5{,}000}{3^2} = 556$$

So, the correct answer is **C**.

Worked examples

Example 4.2

Reference: May 2000, Question 32

A study indicates that t years from now the proportion of a population that will be infected with a disease can be modeled by:

$$I(t) = \frac{(t+1)^2}{100} \quad \text{for } t \le 5$$

Determine the time when the actual proportion infected equals the average proportion infected over the time interval from $t = 0$ to $t = 3$.

(A) 1.38
(B) 1.50
(C) 1.58
(D) 1.65
(E) 1.68

Solution

We must solve the equation:

$$I(t_1) = \overline{I} = \frac{1}{3}\int_0^3 I(t)\,dt$$

for the unknown time t_1.

We have:

$$\overline{I} = \frac{1}{3}\int_0^3 I(t)\,dt = \frac{1}{300}\int_0^3 (t+1)^2\,dt = \left.\frac{(t+1)^3}{900}\right|_0^3 = \frac{64-1}{900} = 0.07$$

Now:

$$I(t_1) = 0.07$$
$$\Rightarrow (t_1+1)^2 = 7$$
$$\Rightarrow t_1 = \sqrt{7} - 1 = 1.646$$

So, the correct answer is **D**.

Worked examples

Example 4.3

Reference: November 2000, Question 31

Let $f(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 4 - x & 1 \le x \le 4 \end{cases}$

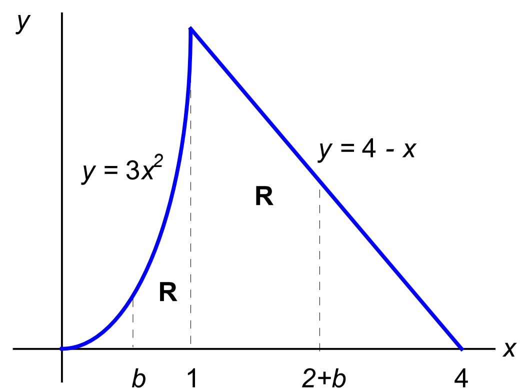
Let R be the region bounded by the graph of f, the x-axis, and the lines $x = b$ and $x = b + 2$ where $0 \le b \le 1$. Determine the value of b that maximizes the area of R.

(A) 0
(B) 1/2
(C) 2/3
(D) 3/4
(E) 1

Solution

There are "two pieces" making up the region R, so let's start by drawing a picture of the region.

We can compute $A(b)$, the area of R as a function of b:



$$A(b) = \int_b^1 3x^2\,dx + \int_1^{2+b} 4 - x\,dx = x^3\Big|_b^1 + \left(4x - \frac{x^2}{2}\right)\Bigg|_1^{2+b}$$

$$= \left(1 - b^3\right) + \left(8 + 4b - \frac{1}{2}\left(4 + 4b + b^2\right)\right) - \left(4 - \frac{1}{2}\right)$$

$$= -b^3 - \frac{1}{2}b^2 + 2b + 3.5$$

The critical points of $A(b)$ can be found as follows:

$$0 = A'(b) = -3b^2 - b + 2 = -(3b - 2)(b + 1) \quad \Rightarrow \quad b = \frac{2}{3} \text{ and } b = -1$$

The maximum value must occur at either the endpoints of the interval (*ie* $b = 0$ or $b = 1$) or at the interior critical point $b = 2/3$:

$$A(0) = 3.5 \qquad A(1) = 4.0 \qquad A(2/3) = 4.315$$

So, the maximum occurs when $b = 2/3$.

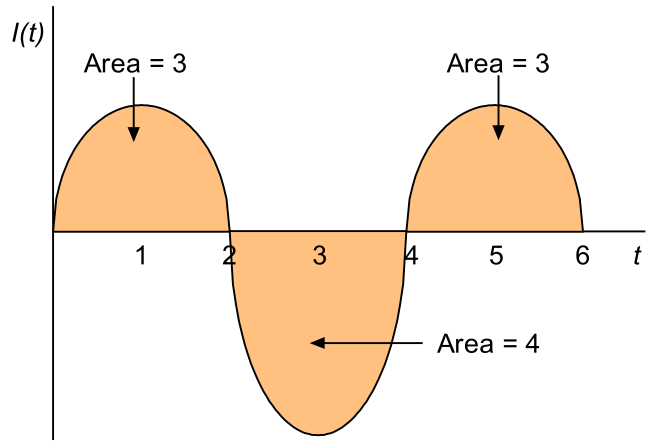
So, the correct answer is **C**.

Practice questions

Question 4.1

Reference: May 2000, Question 28

Inflation is defined as the rate of change in price as a function of time. The figure below is a graph of inflation, $I(t)$, versus time, t.



Price at time $t = 0$ is 100.

What is the next time at which the price is 100?

(A) At some time t, $t \in (0,2)$.

(B) 2

(C) At some time t, $t \in (2,4)$.

(D) 4

(E) At some time t, $t \in (4,6)$.

Question 4.2

Reference: November 2000, Question 35

A company's value at time t is growing at a rate proportional to the difference between 20 and its value at time t.

At time $t = 0$, its value is 2. At time $t = 1$, the value is 3.

Calculate the value at time $t = 3$.

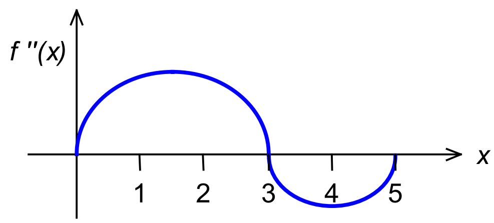
(A) 4.84
(B) 5.00
(C) 5.87
(D) 6.39
(E) 6.75

Practice questions

Question 4.3

Reference: November 2001, Question 12

Let f be a function such that $f'(0)=0$. The graph of the second derivative f'' is shown below.



Determine the *x*-value on the interval $[0,5]$ at which the maximum value of f occurs.

(A) $x=0$
(B) At some x between 1 and 2.
(C) $x=3$
(D) At some x between 3 and 4
(E) $x=5$

Question 4.4

Reference: November 2001, Question 26

An insurance company introduces a new annuity at time $t=0$, where t is in years.

The company has found that, using its current marketing strategies, the instantaneous rate of change of sales of an annuity can be modeled by $s'(t)=t+5/2$.

At time $t=2$, a new advertising campaign is introduced. The instantaneous rate of sales increase changes to $t^2+1/2$.

Calculate the difference in total sales from time $t=2$ to time $t=4$ over what total sales would have been without the new advertising campaign.

(A) $\frac{16}{3}$

(B) 7

(C) $\frac{26}{3}$

(D) 10

(E) $\frac{59}{3}$

Practice questions

Question 4.5 *Reference: November 2001, Question 31*

A town's annual birth rate and annual death rate are each proportional to its population, y, with constants of proportionality k_1 and k_2, respectively. As a result, the new growth of the town can be modeled by the equation

$$\frac{dy}{dt} = (k_1 - k_2)y$$

where t is measured in years.

The town's population doubles every 24 years, but it would be halved in 8 years if there were no births.

Determine k_2.

(A) $-\frac{\ln 2}{6}$

(B) $-\frac{\ln 2}{8}$

(C) $\frac{\ln 2}{24}$

(D) $\frac{\ln 2}{12}$

(E) $\frac{\ln 2}{8}$

Practice questions

Question 4.6

Reference: May 2001, Question 21

The rate at which a disease spreads through a town can be modeled by the differential equation:

$$\frac{dQ}{dt} = Q(N - Q)$$

where $Q(t)$ is the number of residents infected at time t and N is the total number of residents.

Which of the following is a solution for $Q(t)$?

(A) ae^t where a is a constant

(B) $\dfrac{aNe^t - 1}{ae^t}$ where a is a constant

(C) $\dfrac{aNe^t + 1}{ae^t}$ where a is a constant

(D) $\dfrac{aNe^{Nt}}{1 - ae^{Nt}}$ where a is a constant

(E) $\dfrac{aNe^{Nt}}{1 + ae^{Nt}}$ where a is a constant

