CEU
Universidad
Cardenal Herrera

Matemáticas y Estadística

Mónica Alacreu García

monica.alacreu@uchceu.es

Matemáticas y Estadística

1º de Grado en Farmacia

Recordatorio

0.1. Potencias.

$x^0 = 1$
$x^a y^a = (xy)^a$
$x^a x^b = (x)^{a+b}$
$\frac{x^a}{x^b} = x^{a-b}$
$(x^a)^b = x^{ab}$
$x^{-a} = \frac{1}{x^a}$
$x^{\frac{a}{b}} = \sqrt[b]{x^a}$

0.2. Identidades notables.

$(a+b)^2 = a^2 + 2ab + b^2$
$(a-b)^2 = a^2 - 2ab + b^2$
$(a+b)(a-b) = a^2 - b^2$

0.3. Raíces de polinomios.

$ax^2 + bx + c = 0 \Rightarrow r_i = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ *con* $i = 1, 2$. *Entonces* $ax^2 + bx + c = a(x - r_1)(x - r_2)$.

$a_n x^n + a_{n-1} x^{n-1} + ... + a_0 = 0$

Regla de Ruffini:

$$\begin{array}{c|ccccc} & a_n & a_{n-1} & ... & a_1 & a_0 \\ r & & b_{n-1}r & ... & b_1 r & b_0 r \\ \hline & a_n & a_{n-1} + b_{n-1}r & ... & a_1 + b_1 r & a_0 + b_0 r \\ & = b_{n-1} & = b_{n-2} & ... & = b_0 & =0 \end{array}$$

Entonces $a_n x^n + a_{n-1} x^{n-1} + ... + a_0 = (x - r)(b_{n-1} x^{n-1} + b_{n-2} x^{n-2} + ... + b_0)$

0.4. Trigonometría.

x	30º	45º	60º	90º
sen(x)	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
cos(x)	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
tg(x)	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	∞

$tg(x) = \frac{sen(x)}{cos(x)}$; $sec(x) = \frac{1}{cos(x)}$; $cosec(x) = \frac{1}{sen(x)}$
$cos^2(x) + sen^2(x) = 1$
$cos^2(x) - sen^2(x) = cos(2x)$
$sen(x \pm y) = sen(x)cos(y) \pm cos(x)sen(y)$
$cos(x \pm y) = cos(x)cos(y) \mp sen(x)sen(y)$

0.5. Logaritmos.

$log_a(x) = b \Longleftrightarrow a^b = x$
$log_e(x) = ln(x)$
$log_a(xy) = log_a(x) + log_a(y)$
$log_a(\frac{x}{y}) = log_a(x) - log_a(y)$
$nlog_a(x) = log_a(x^n)$

0.6. Función módulo.

$$|x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

0.7. Vectores.

$\mathbf{u} = (u_1, u_2, u_3)$; $\mathbf{v} = (v_1, v_2, v_3)$

Módulo
$|\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + u_3^3}$

Suma y resta
$\mathbf{u} \pm \mathbf{v} = (u_1 \pm v_1, u_2 \pm v_2, u_3 \pm v_3)$

Producto por un escalar
$k\mathbf{u} = (ku_1, ku_2, ku_3)$

Producto escalar
$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3 = |\mathbf{u}||\mathbf{v}| \cos\theta$

Producto vectorial
$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)$

Módulo del producto vectorial
$|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}| \sin\theta$

Índice general

Capítulo 1

Cálculo Diferencial

1.1. La derivada y el problema de la recta tangente.

El **problema de la recta tangente** consiste en determinar la ecuación de la recta tangente a la gráfica de una función, f, en un punto, P, de la misma. La primera solución general a este problema se atribuye a Isaac Newton (1642-1727) y a Gottfried Leibniz (1646-1716). La solución que dio Newton al problema procedía de su interés por la refracción de la luz y la óptica.

En esencia, el problema de calcular la recta tangente en un punto P se reduce al de calcular su pendiente en ese punto.

Recordemos, que la ecuación de la recta punto-pendiente se escribe como aparece junto a la Figura 1.1:

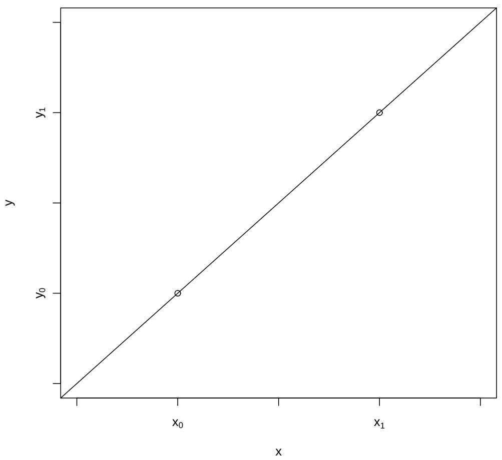


Figura 1.1: Recta que pasa por los puntos (x_0, y_0) y (x_1, y_1).

$$y - y_0 = m(x - x_0)$$

, donde m es la pendiente de la recta

$$y - y_0 = \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$$

Para aproximar la pendiente de la recta tangente a la gráfica de una función, f, en un punto, P, usaremos la **recta secante** que pasa por P y otro punto cercano de la curva, Q, como se muestra en la Figura 1.2:

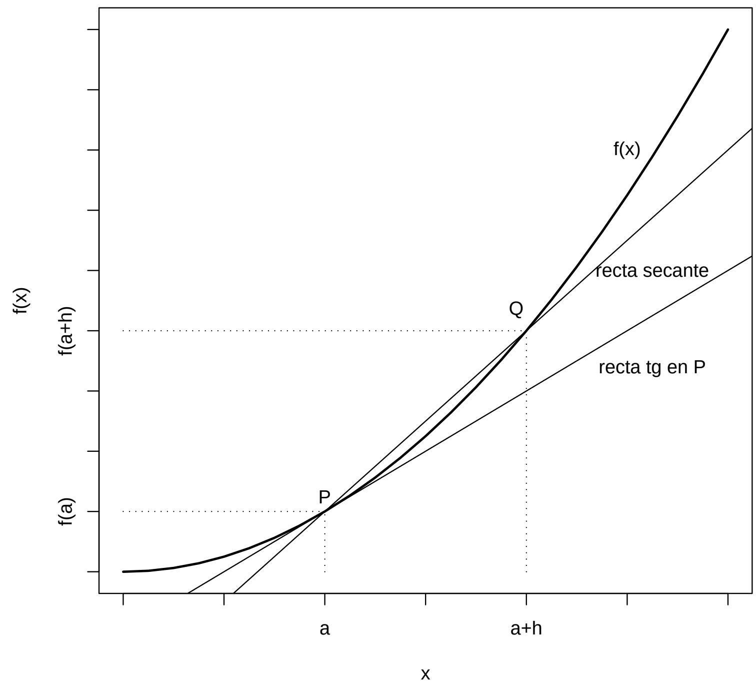


Figura 1.2: Recta tangente a la gráfica de una función, f, en un punto, P, y recta secante entre los puntos P y Q.

La **ecuación de la recta secante** se expresa como:

$$y - f(a) = \frac{f(a+h) - f(a)}{(a+h) - a}(x - a)$$

donde la pendiente de la recta secante es $m_{sec} = \frac{f(a+h)-f(a)}{(a+h)-a} = \frac{f(a+h)-f(a)}{h}$.

Cuando se consideran puntos, Q, de la gráfica cada vez más próximos a P, es decir, cuando la distancia h es cada vez más pequeña (h tiende a 0, $h \to 0$), la ecuación de la recta secante tiende a la **ecuación de la recta tangente** a f en el punto P:

$$y - f(a) = m_{tg}(x - a)$$

donde la pendiente de la recta tangente es $m_{tg} = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$.

Ejemplo 1.1.

Calcula la ecuación de la recta tangente a y=x^2 en el punto (-1,1).

Solución:

La ecuación de la recta tangente es $y-1 = m_{tg}(x-(-1))$, donde
$m_{tg} = \lim_{h\to 0} \frac{f(-1+h)-f(-1)}{h} = \lim_{h\to 0} \frac{(-1+h)^2-(-1)^2}{h} = \lim_{h\to 0} \frac{1-2h+h^2-1}{h} = \lim_{h\to 0} \frac{h(-2+h)}{h} = -2$
$y-1 = -2(x+1)$

El límite utilizado para definir la pendiente de la recta tangente también se utiliza para definir una de las operaciones fundamentales del cálculo: la derivada.

La **derivada** de f en x viene dada por

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

siempre que exista ese límite. Luego $m_{tg} = f'(x_0)$ donde x_0 es la abscisa del punto tangente.

En definitiva, la ecuación de la recta tangente a una curva f(x) en el punto (x_0, y_0), es:

$$y - y_0 = f'(x_0)(x - x_0)$$

Definición

Una **función** es **derivable** en un punto x de su dominio, si en ese punto existe su derivada. Es decir, existe el límite, $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$.

Ejemplo 1.2.

En la próxima sección se utilizarán fórmulas que permitirán generalizar el cálculo de la derivada de una función. Mientras tanto, calcula la derivada mediante el proceso del límite de:
a) f(x)=x^2
b) f(x)=x^3
c) f(x)=x^n
d) f(x)=$\frac{1}{x-1}$

Solución:
a) $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{(x+h)^2-(x)^2}{h} = \lim_{h\to 0} \frac{(x^2+2xh+h^2)-x^2}{h} = \lim_{h\to 0} \frac{2xh+h^2}{h} = \lim_{h\to 0} \frac{h(2x+h)}{h} = 2x$

b) $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{(x+h)^3-(x)^3}{h} = \lim_{h\to 0} \frac{(x^3+3x^2h+3xh^2+h^3)-x^3}{h} = \lim_{h\to 0} \frac{3x^2h+3xh^2+h^3}{h} = \lim_{h\to 0} \frac{h(3x^2+3xh+h^2)}{h} = 3x^2$

b) $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{(x+h)^n-(x)^n}{h} = \lim_{h\to 0} \frac{(x^n+nx^{n-1}h+...+nxh^{n-1}+h^n)-x^n}{h} = \lim_{h\to 0} \frac{nx^{n-1}h+...+nxh^{n-1}+h^n}{h} = \lim_{h\to 0} \frac{h(nx^{n-1}+...+h^{n-1})}{h} = nx^{n-1}$

d) $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{\frac{1}{(x+h)-1}-\frac{1}{x-1}}{h} = \lim_{h\to 0} \frac{\frac{(x-1)-((x+h)-1)}{((x+h)-1)(x-1)}}{h} = \lim_{h\to 0} \frac{\frac{x-1-x-h+1}{((x+h)-1)(x-1)}}{h} = \lim_{h\to 0} \frac{-h}{h((x+h)-1)(x-1)} = \frac{-1}{(x-1)^2}$

Una forma alternativa de expresar la derivada de f en un punto c y que es útil para investigar la relación entre derivabilidad y continuidad es

$$f'(c) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c}$$

Obsérvese que la existencia del límite en esta forma alternativa requiere que los límites unilaterales

$$\lim_{x\to c^+} \frac{f(x)-f(c)}{x-c} \quad y \quad \lim_{x\to c^-} \frac{f(x)-f(c)}{x-c}$$

existan y sean iguales.

Ejemplo 1.3.

La función f(x)=$|x-2|$ es continua en x=2. Sin embargo, los límites unilaterales no son iguales

$\lim_{x\to 2^-} \frac{f(x)-f(2)}{x-2} = \lim_{x\to 2^-} \frac{|x-2|-0}{x-2} = -1$
$\lim_{x\to 2^+} \frac{f(x)-f(2)}{x-2} = \lim_{x\to 2^+} \frac{|x-2|-0}{x-2} = 1$
Por consiguiente, f no es derivable en x=2 y la gráfica de f no tiene una recta tangente en el punto (2,0).

Autoevaluación

1. Halla la ecuación de la recta tangente a la gráfica $f(x) = \frac{1}{x+1}$ en el punto (0,1).
   Sol: y-1= -1 (x-0)

2. Encontrar la derivada, mediante el proceso al límite, para las siguientes funciones:

   a) $f(x) = 2x^2 + x - 1$. Sol: $4x + 1$

   b) $g(x) = \sqrt{x-4}$. Sol: $\frac{1}{2\sqrt{x-4}}$

1.2. Reglas de derivación

En la sección anterior se ha usado la definición por medio del límite para calcular la derivada de una función. En esta sección se presentan varias "reglas de derivación" que permiten calcula las derivadas sin el uso directo de la definición por límites.

La regla de la constante

La derivada de una función constante es 0. Es decir, si f(x)=c es un número real, entonces f'(x)=0

La regla de las potencias

Si n es un número racional, entonces la función f(x)=x^n es derivable y su derivada es f'(x)=nx^{n-1}. Para que f sea derivable en x=0, n debe ser un número tal que x^{n-1} se encuentre definido en un intervalo que contenga al 0.

Las derivadas de las funciones seno y coseno

(sen)'(x)=cos(x)
(cos)'(x)=-sen(x)

Derivadas de las funciones trigonométricas

(tg)'(x)=sec^2(x)
(sec)'(x)=sec(x)tg(x)
(ctg)'(x)=-$csec^2$(x)
(csec)'(x)=-csec(x)ctg(x)

La regla de múltiplo constante

Si f es una función derivable y c un número real, entonces cf también es derivable y su derivada es (cf)'(x)=cf'(x).

Las reglas de la suma y diferencia

La derivada de la suma (o de la diferencia) de dos funciones derivables f y g es derivable en sí. Además, la derivada de f+g (o f-g) es igual a la suma (o diferencia) de las derivadas de f y g.

(f+g)'(x)=f'(x)+g'(x)

(f-g)'(x)=f'(x)-g'(x)

La regla del producto

El producto de dos funciones derivables f y g es derivable en sí. Además, su derivada es

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x)$$

La regla del cociente

El cociente f/g de dos funciones derivables f y g es derivable en sí para todos los valores de x para los que g(x)≠0. Además, la derivada de f/g es

$$(f/g)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$$

La regla de la cadena

Si y=f(u) es una función derivable de u y además u=g(x) es una función derivable de x, entonces y=(f∘g)(x)=f(g(x)) es una función derivable de x y su derivada es

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

Ejemplo 1.4.

Calcula la derivada de las funciones:
a)f(x)=$\sqrt[3]{(x^2-1)^2}$***.***
b)g(t)=$sen^3(4t)$

Solución:
a)Reescribimos la función de manera que sea fácil aplicar la regla de la cadena y la regla de las potencias, es decir, f(x)=$(x^2-1)^{\frac{2}{3}}$. Entonces su derivada es:

$$f'(x) = \frac{2}{3}(x^2-1)^{-\frac{1}{3}}(2x) = \frac{4x}{3\sqrt[3]{(x^2-1)}}$$

b)Aplicamos reiteradamente la regla de la cadena:

$$g'(t) = 3sen^2(4t)cos(4t)4 = 12sen^2(4t)cos(4t)$$

Autoevaluación

1. Halla la derivada en las siguientes funciones, utilizando las reglas de derivación:

   a) $f(x) = x^2 - \frac{4}{x}$. Sol: $f'(x) = 2x + \frac{4}{x^2}$

   b) $g(x) = \frac{x^3-3x^2+4}{x^2}$. Sol: $g'(x) = \frac{x^3-8}{x^3}$

   c) $f(x) = 2sen(x) + 3cos(x)$. Sol: $f'(x) = 2cos(x) - 3sen(x)$

   d) $h(t) = 4\sqrt{t} + 3cos(t)$. Sol:$h'(t) = \frac{2}{\sqrt{t}} - 3sen(t)$

   e) $f(x) = ln(x)(x^3 + 6x)$. Sol:$f'(x) = (x^2 + 6) + ln(x)(3x^2 + 6)$

1.3. Tabla de derivadas

Función	Derivada	Función ejemplo	Solución derivada
y= k	y'= 0	y= 8	y'= 0
y= x	y'= 1	y= x	y'= 1
y= u^m	y'= $m\ u^{m-1}\ u'$	y= $(2x^2+1)^3$	y'= $3(2x^2+1)^2(4x)$
		y= $\frac{1}{(2x+1)^3}$	y'= $\frac{-6}{(2x+1)^4}$
		y=$\sqrt{5x}$	y'= $\frac{5}{2\sqrt{5x}}$
		y= $\sqrt[5]{3x^2}$	y'= $\frac{6x}{5\sqrt[5]{(3x^2)^4}}$
y= e^u	y'= $u'e^u$	y= e^{3x^2+1}	y'= $6xe^{3x^2+1}$
y= a^u	y'= $u'a^u ln(a)$	y= 5^{3x-4}	y'= $3\cdot 5^{3x-4}ln(5)$
y= $ln(u)$	y'= $\frac{u'}{u}$	y= $ln(x^2+7x)$	y'= $\frac{2x+7}{x^2+7x}$
y= $log_a(u)$	y'= $\frac{u'}{u}log_a(e)$	y= $log_2(5x+7)$	y'= $\frac{5}{5x+7}log_2(e)$
y= $sen(u)$	y'= $u'cos(u)$	y= $sen(5x)$	y'= $5cos(5x)$
y= $cos(u)$	y'= $-u'sen(u)$	y= $cos(3x^2)$	y'= $-6xsen(3x^2)$
y= $tg(u)$	y'= $u'sec^2(u)$	y= $tg(7x)$	y'= $7sec^2(7x)$
y= $cotg(u)$	y'= $-u'cosec^2(u)$	y= $cotg(4x+5)$	y'= $-4cosec^2(4x+5)$
y= $sec(u)$	y'= $u'sec(u)tg(u)$	y= $sec(x^3)$	y'= $3x^2sec(x^3)tg(x^3)$
y= $cosec(u)$	y'= $-u'cosec(u)cotg(u)$	y= $cosec(x^2)$	y'= $-2xcosec(x^2)cotg(x^2)$
y= $arcsen(u)$	y'= $\frac{u'}{\sqrt{1-u^2}}$	y= $arcsen(x^2)$	y'= $\frac{2x}{\sqrt{1-x^4}}$
y= $arccos(u)$	y'= $\frac{-u'}{\sqrt{1-u^2}}$	y= $arccos(3x)$	y'= $\frac{-3}{\sqrt{1-9x^2}}$
y= $arctg(u)$	y'= $\frac{u'}{1+u^2}$	y= $arctg(3x)$	y'= $\frac{3}{1+9x^2}$
y= ku	y'= ku'	y= $3x^5$	y'= $3\cdot 5x^4$
y= $u+v-w$	y'= $u'+v'-w'$	y= $3x^2-2x+5$	y'= $6x-2$
y= uv	y'= $u'v+uv'$	y= $x^2cos(x)$	y'= $2xcos(x)+x^2(-sen(x))$
y= $\frac{u}{v}$	y'= $\frac{u'v-uv'}{v^2}$	y= $\frac{2x^2}{x^3-1}$	y'= $\frac{4x(x^3-1)-2x^2(3x^2)}{(x^3-1)^2}$

Autoevaluación

1. Comprueba que obtienes las mismas soluciones en los ejemplos de la tabla de derivadas.

1.4. Derivación implícita

Hasta ahora, la mayor parte de las funciones estudiadas se han enunciado de **forma explícita**, es decir, la variable dependiente está expresada explícitamente como función de la variable independiente, x. Este es el caso, por ejemplo, de la función $y = 3x^2 - 5$. Sin embargo, algunas funciones sólo se enuncian de **forma implícita**, como por ejemplo la función que representa la circunferencia, $x^2 + y^2 = 1$.

Supongamos que se pide calcular la derivada dy/dx=y' de esta función. Podemos escribir "y" como función explícita de "x" y luego derivar.

Forma implícita	Forma explícita	Derivada
$x^2 + y^2 = 1$	$y = (1 - x^2)^{\frac{1}{2}}$	$y' = \frac{1}{2}(1 - x^2)^{-\frac{1}{2}}(-2x) = \frac{-2x}{2\sqrt{1-x^2}}$

Esta estrategia funciona siempre que se pueda despejar "y" en función de "x", de lo contrario, este método no es viable. Por ejemplo, ¿cómo se puede calcular la derivada dy/dx=y' de la función $x^2 - 2y^3 + 4y = 2$?. Nótese que en este caso no es posible despejar la variable y. En esta situación se debe usar la llamada **derivación implícita**.

Para comprender esta técnica, es preciso tener en cuenta que la derivación se efectúa con respecto a "x". Esto quiere decir que cuando haya que derivar un término donde aparezca "x", la derivación será la habitual. Sin embargo, cuando haya que derivar un término donde aparezca "y", será necesario aplicar la regla de la cadena, ya que se está suponiendo que "y" está definida implícitamente como función derivable de x.

Pasos a seguir para la derivación implícita

1. Derivar ambos lados de la ecuación respecto de x.
2. Agrupar todos los términos en que aparezca y' en el lado izquierdo de la ecuación y pasar todos los demás a la derecha.
3. Factorizar y' del lado izquierdo de la ecuación.
4. Despejar y'.

Ejemplo 1.5.

a)Calcula la derivada de la función $x^2 - 2y^3 + 4y = 2$**.**
b)Calcula la derivada de la función $x^2 - 2xy^3 + 4y = 2$**.**

Solución:
a)$2x - 2 \cdot 3y^2 \cdot y' + 4y' = 0 \quad \rightarrow \quad -2 \cdot 3y^2 \cdot y' + 4y' = -2x \quad \rightarrow \quad y'(-6y^2 + 4) = -2x \quad \rightarrow \quad y' = \frac{-2x}{(4-6y^2)}$
b)$2x - (2 \cdot y^3 + 2x \cdot 3y^2 \cdot y') + 4y' = 0 \quad \rightarrow \quad 2x - 2y^3 - 6xy^2 \cdot y' + 4y' = 0 \quad \rightarrow \quad y'(4 - 6xy^2) = 2y^3 - 2x \quad \rightarrow \quad y' = \frac{2y^3-2x}{4-6xy^2}$

Ejemplo 1.6.

Calcula la ecuación de la recta tangente a la gráfica de $x^2 + 4y^2 = 4$ ***en el punto*** ***(***$\sqrt{2}$***,*** $\frac{-1}{\sqrt{2}}$***).***

Solución:
La ecuación de la recta tangente a una función f en el punto (x_0, y_0) viene expresada por $y - y_0 = m_{tg}(x - x_0)$, donde la pendiente de la recta es $m_{tg} = f'(x_0, y_0)$. En particular para nuestros datos $m_{tg} = f'(\sqrt{2}, \frac{-1}{\sqrt{2}})$. Para averiguar su valor, calculamos la derivada de la función mediante derivación implícita:

$$2x + 4 \cdot 2y \cdot y' = 0 \quad \rightarrow \quad y' = \frac{-2x}{8y} \quad \rightarrow \quad m_{tg} = f'(\sqrt{2}, \frac{-1}{\sqrt{2}}) = \frac{-2\sqrt{2}}{8\frac{-1}{\sqrt{2}}} = \frac{-2(\sqrt{2})^2}{-8} = \frac{1}{2}$$

En definitiva, la ecuación de la recta tangente es $y - \frac{-1}{\sqrt{2}} = \frac{1}{2}(x - \sqrt{2})$.

Autoevaluación

1. Calcula la derivada en las siguientes funciones:

   a) $x^2y + y^2x = -2$. Sol: $y' = \frac{-2xy - y^2}{x^2 + 2xy}$

   b) $x^3 - 2x^2y + 3xy^2 = 38$. Sol: $y' = \frac{-3x^2 + 4xy - 3y^2}{-2x^2 + 6xy}$

   c) $sen(x) + 2cos(2y) = 1$. Sol: $y' = \frac{cos(x)}{4sen(2y)}$

2. Calcula la ecuación de la recta tangente a la gráfica $(x + y)^3 = x^3 + y^3$ en el punto (-1,1).
   Sol: y-1=-1(x+1).

Cálculo implícito de la segunda derivada

Se trata de repetir el procedimiento de derivación implícita, pero esta vez, sobre la función de la primera derivada de forma que los términos y' se sustituyen por la expresión obtenida de la primera derivada.

Ejemplo 1.7.

Calcula la segunda derivada de la función $x^2 - 2y^3 + 4y = 2$.

Solución:

La primera derivada se obtiene del siguiente modo:
$2x - 2 \cdot 3y^2 \cdot y' + 4y' = 0 \quad \rightarrow \quad -2 \cdot 3y^2 \cdot y' + 4y' = -2x \quad \rightarrow \quad y'(-6y^2+4) = -2x \rightarrow \quad y' = \frac{-2x}{(4-6y^2)}$
Ahora repetimos el procedimiento sobre la expresión de y':

$$y'' = \frac{-2(4-6y^2) - (-2x)(-6 \cdot 2y \cdot y')}{(4-6y^2)^2} = \frac{-8 + 12y^2 - 24xyy'}{(4-6y^2)^2}$$

Por último, sustituimos y' por su valor y simplificamos:

$$y'' = \frac{-8 + 12y^2 - 24xy\frac{-2x}{(4-6y^2)}}{(4-6y^2)^2} =$$

$$\frac{\frac{-8(4-6y^2)+12y^2(4-6y^2)-24xy(-2x)}{(4-6y^2)}}{(4-6y^2)^2} =$$

$$\frac{-8(4-6y^2) + 12y^2(4-6y^2) - 24xy(-2x)}{(4-6y^2)^3} =$$

$$\frac{-32 + 48y^2 + 48y^2 - 72y^4 + 48x^2y}{(4-6y^2)^3} = \frac{-32 + 96y^2 - 72y^4 + 48x^2y}{(4-6y^2)^3}$$

Autoevaluación

1. Halla la segunda derivada en las siguientes funciones:

   a) $xy + y = 4$. Sol: $y'' = \frac{2y}{(x+1)^2}$

   b) $xy^2 + y = 1$. Sol: $y'' = \frac{4y^3}{(2xy+1)^2} - \frac{2xy^4}{(2xy+1)^3}$

Recordatorio

Un intervalo real, I, es un espacio métrico comprendido entre dos valores que está incluido en la recta real $\Re$. La notación utilizada de ahora en adelante será:
I es un **intervalo cerrado** si I=[a, b]= $\{x \in \Re / \ a \leq x \leq b\}$
I es un **intervalo abierto** si I=]a, b[= (a, b)= $\{x \in \Re / \ a < x < b\}$
I es un **intervalo semiabierto por la izquierda** si I=]a, b]= (a, b]= $\{x \in \Re / \ a < x \leq b\}$
I es un **intervalo semiabierto por la derecha** si I=[a, b[= [a, b)= $\{x \in \Re / \ a \leq x < b\}$

1.5. Extremos absolutos, extremos relativos y monotonía de una función

Una función, f, tiene un **máximo absoluto** en x= c si f(c)$\geq$ f(x) para todo x perteneciente al dominio de f.

Una función, f, tiene un **mínimo absoluto** en x= c si f(c)$\leq$ f(x) para todo x perteneciente al dominio de f.

Una función, f, tiene un **máximo relativo** en x= c si existe un intervalo I tal que c $\in$ I y f(c) $\geq$ f(x) para todo x perteneciente a I.

Una función, f, tiene un **mínimo relativo** en x= c si existe un intervalo I tal que c $\in$ I y f(c) $\leq$ f(x) para todo x perteneciente a I.

Los mínimos y máximos de una función en un intervalo son los **puntos extremos** de la función.

Teorema de valor extremo

Si f es continua en el intervalo cerrado [a,b], entonces f tiene tanto un mínimo relativo como un máximo relativo en el intervalo.

Sea f definida en c. Si f'(c)=0 o si f no es derivable en c, entonces c es un **punto crítico** de f.

Teorema

Si f tiene un mínimo relativo o un máximo relativo en x=c, entonces c es un punto crítico de f.

Ejemplo 1.8.

a) f(x)= $\frac{1}{x^2-1}$ tiene puntos críticos en x= -1 y x= 1 porque no pertenecen al dominio de la función y por lo tanto, no es derivable en esos puntos y también en x= 0, ya que f'(0)= 0.

b) f(x)= $|x|$ tiene un mínimo relativo en x= 0, por lo que x= 0 es un punto crítico de la función.

El teorema del valor extremo establece que una función continua en un intervalo cerrado [a,b] debe tener tanto un mínimo como un máximo en el intervalo, aunque sea en los extremos del intervalo. El siguiente teorema proporciona las condiciones que garantizan la existencia de un valor extremo en el interior de un intervalo cerrado.

Teorema de Rolle

Sea f continua en el intervalo cerrado [a,b] y derivable en el intervalo abierto (a,b). Si f(a)=f(b) entonces existe al menos un número c en (a,b) tal que f'(c)=0.

Teorema del valor medio

Si f es continua en el intervalo cerrado [a,b] y derivable en el intervalo abierto (a,b), entonces existe un número c en (a,b) tal que f'(c)=$\frac{f(b)-f(a)}{b-a}$.

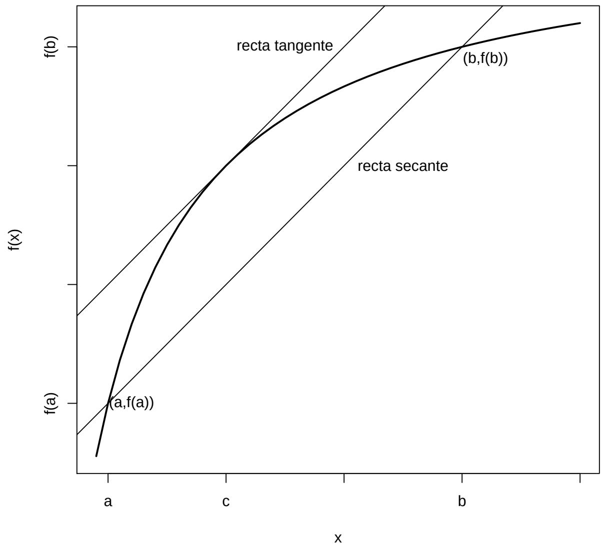


Figura 1.3: La recta tangente en c tiene la misma pendiente que la recta secante entre los puntos (a,f(a)) y (b,f(b)).

Una función f es **creciente** en un intervalo si para cualesquiera dos números x_1 y x_2, en el intervalo, $x_1 < x_2$ implica que $f(x_1) < f(x_2)$.

Una función f es **decreciente** en un intervalo si para cualesquiera dos números x_1 y x_2, en el intervalo, $x_1 < x_2$ implica que $f(x_1) > f(x_2)$.

Teorema Criterio para determinar la monotonía (crecimiento y decrecimiento)

Sea f una función que es continua en el intervalo cerrado [a,b] y derivable en el intervalo abierto (a,b).

1. Si f'(x)>0 para todo x en (a,b), entonces f es creciente en [a,b].
2. Si f'(x)<0 para todo x en (a,b), entonces f es decreciente en [a,b].
3. Si f'(x)=0 para todo x en (a,b), entonces f es constante en [a,b].

Estrategia para estudiar los puntos extremos y monotonía de una función

1. Calcular los puntos críticos de f en (a,b).

2. Determinar el signo de f' en los intervalos que definen los puntos críticos. En cada intervalo la función será creciente, decreciente o contante según el criterio del teorema anterior.

3. Decidir si los puntos críticos son máximos o mínimos dependiendo de si la función pasa de creciente a decreciente o de decreciente a creciente, respectivamente.

Ejemplo 1.9.

Determina los puntos extremos y la monotonía de la función $f(x) = \frac{x^2+1}{x^2-4}$

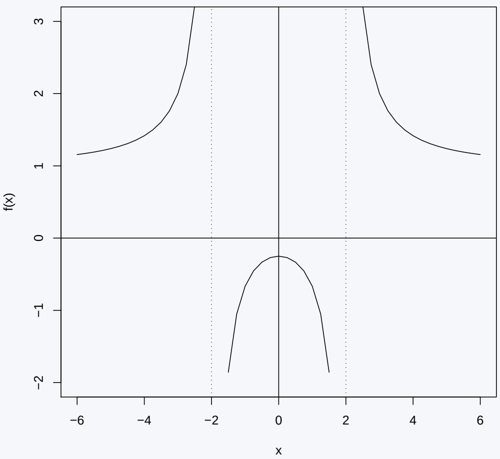
Solución:

En primer lugar calculamos los puntos críticos, es decir los puntos que hacen 0 la primera derivada y los puntos donde f no es derivable.
En x=-2 y x=2 la función no existe, con lo cual f no es derivable tampoco.

$$f'(x) = \frac{(2x)(x^2-4)-(x^2+1)(2x)}{(x^2-4)^2} = \frac{-10x}{(x^2-4)^2} = 0 \qquad \rightarrow \qquad x = 0$$

Los puntos críticos son x=-2, x=0 y x=2 y por lo tanto la recta real queda dividida en 4 intervalos. Sólo queda estudiar el signo de f' en estos 4 intervalos para determinar la monotonía de la función:

Intervalo	]-∞,-2[	x=-2	]-2,0[	x=0	]0,2[	x=2	]2,∞[
Valor de prueba	x=-3		x=-1		x=1		x=3
Signo de f'	f'(-3)>0		f'(-1)>0		f'(1)<0		f'(3)<0
Conclusión	creciente	$\nexists$	creciente	máximo	decreciente	$\nexists$	decreciente



1.6. Puntos de inflexión y curvatura de una función

Sea f derivable en un intervalo abierto I. La gráfica de f es **convexa** sobre I si f' es creciente en el intervalo y es **concava** en I si f' es decreciente en I.

El siguiente teorema muestra cómo utilizar la segunda derivada de una función f para determinar su curvatura.

Teorema

Sea f una función cuya segunda derivada existe en un intervalo abierto I.

a) Si f"(x)>0 para todo x en I, entonces la gráfica de f es convexa.

b) Si f"(x)<0 para todo x en I, entonces la gráfica de f es cóncava.

Sea f una función que es continua en un intervalo abierto y sea c un punto en ese intervalo. Si la gráfica de f tiene una recta tangente en ese punto (c,f(c)) y en él cambia la función de cóncava a convexa entonces (c,f(c)) es un **punto de inflexión**.

Teorema

Si (c,f(c)) es un punto de inflexión de la gráfica de f, entonces f"(x)=0 o f" no existe en x=c.

Ejemplo 1.10.

Determina los puntos de inflexión y analiza la curvatura de la gráfica de f(x)= $x^4 - 4x^3$**.**

Solución:

En primer lugar, se determinan los posibles puntos de inflexión.

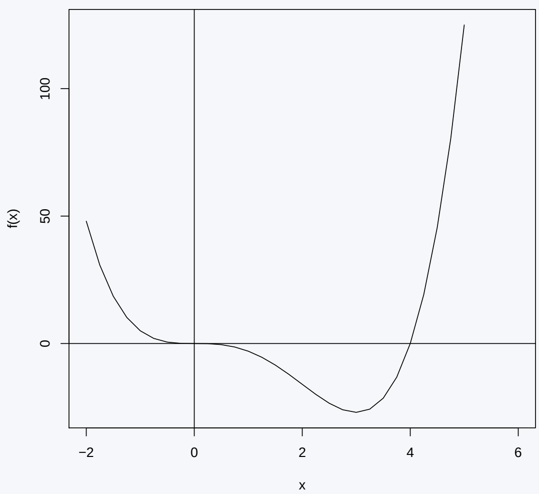
f(x)=$x^4 - 4x^3$

f'(x)=$4x^3 - 12x^2$

f"(x)=$12x^2 - 24x = 12x(x-2) = 0 \qquad \rightarrow \qquad x = 0 \; y \; x = 2$

En este caso, la recta real queda dividida en 3 intervalos. A continuación, se estudia el signo de f" en los 3 intervalos para determinar la curvatura de la función.

Intervalo	]-∞,0[	x=0	]0,2[	x=2	]2,∞[
Valor de prueba	x=-1		x=1		x=3
Signo de f"	f"(-1)>0		f"(1)<0		f"(3)>0
Conclusión	convexa	pto inflexión	concava	pto inflexión	convexa



Autoevaluación

1. Estudia la monotonía y la curvatura de las siguientes funciones, determinando los puntos extremos y los puntos de inflexión:

a) $f(x) = \frac{-(x^2+1)}{x^2-1}$.

Intervalo	]-∞,-1[	x=-1	]-1,0[	x=0	]0,1[	x=1	]1,∞[
Valor de prueba	x=-2		x=-1/2		x=1/2		x=2
Signo de f'	f'(-2)<0		f'(-1/2)<0		f'(1/2)>0		f'(2)>0
Conclusión	decreciente	∄	decreciente	mínimo	creciente	∄	creciente

Intervalo	]-∞,-1[	x=-1	]-1,1[	x=1	]1,∞[
Valor de prueba	x=-2		x=0		x=2
Signo de f''	f''(-2)<0		f''(0)>0		f''(2)<0
Conclusión	cóncava	∄	convexa	∄	cóncava

b) $g(x) = \sqrt{x^2-4}$:

Intervalo	]-∞,-2]	]-2,2[	[2,∞[
Valor de prueba	x=-3	∄	x=3
Signo de f'	f'(-3)<0	∄	f'(3)>0
Conclusión	decreciente	∄	creciente

Intervalo	]-∞,-2[	]-2,2[	]2,∞[
Valor de prueba	x=-3	∄	x=3
Signo de f''	f''(-3)<0	∄	f''(3)<0
Conclusión	cóncava	∄	cóncava

c) $f(x) = \frac{x^2-1}{x^2-4}$:

Intervalo	]-∞,-2[	x=-2	]-2,0[	x=0	]0,2[	x=2	]2,∞[
Valor de prueba	x=-3		x=-1		x=1		x=3
Signo de f'	f'(-3)>0		f'(-1)>0		f'(1)<0		f'(3)<0
Conclusión	creciente	∄	creciente	máximo	decreciente	∄	decreciente

Intervalo	]-∞,-2[	x=-2	]-2,2[	x=2	]2,∞[
Valor de prueba	x=-3		x=0		x=3
Signo de f''	f''(-3)>0		f''(0)<0		f''(3)>0
Conclusión	convexa	∄	cóncava	∄	convexa

1.7. Seminario de CÁLCULO DIFERENCIAL

(Algunos de estos ejercicios han sido obtenidos del libro **CÁLCULO**, de Larson, Hostetler y Edwards (McGrawHill) *Octava edición*. En este libro se pueden encontrar multitud de ejercicios de cada una de las sesiones de este tema.)

(Cálculo de la ecuación de la recta tangente utilizando el límite)

1. Calcula la ecuación de la recta tangente a las siguientes funciones en los puntos indicados:

   a) $f(x) = x^3 + 2$ en el punto (1,3).
   b) $f(x) = x^2 + 2x - 1$ en el punto (0,-1).

(Reglas básicas de derivación)

2. Calcular la derivada de las siguientes funciones:

   a) $f(x) = 2x^3 - x^2 + 3x$
   b) $g(x) = \sqrt[4]{x}$
   c) $h(t) = \pi\cos(t)$
   d) $f(y) = 5 + \sin(y)$
   e) $f(x) = \dfrac{5}{(2x)^3} + 2\cos(x)$
   f) $f(x) = x + \dfrac{1}{x^2}$
   g) $h(x) = \dfrac{2x^2 - 3x + 1}{x}$
   h) $f(x) = \dfrac{2}{\sqrt[3]{x}} + 3\cos(x)$
   i) $f(x) = \dfrac{\sin(x)}{x}$
   j) $f(x) = x \cdot \sin(x) + \cos(x)$
   k) $f(x) = \sin(x) \cdot \cos(x)$
   l) $f(x) = \dfrac{3(1 - \sin(x))}{2\cos(x)}$

(Derivadas de orden superior)

3. En los siguientes ejercicios encontrar la/s derivada/s de orden superior que se indica:

   a) $f(x) = x^2 - 2x + 3$. Hallar $f'(x)$, $f''(x)$ y $f'''(x)$.
   b) $f(x) = 4\sqrt{x}$. Hallar $f'(x)$ y $f''(x)$.
   c) $f(x) = 2 - \dfrac{2}{x}$. Hallar $f'(x)$, $f''(x)$ y $f'''(x)$.

(La regla de la cadena)

4. Encontrar la derivada de las siguientes funciones:

   a) $f(x) = 4(5 - x^2)^5$
   b) $g(x) = \sqrt{x^3 - 3x + 2}$
   c) $g(t) = \sqrt{\frac{1}{t^2+4}}$
   d) $f(x) = \frac{x}{\sqrt{x^4+4}}$
   e) $f(x) = x^2 tan(\frac{1}{x})$
   f) $h(t) = 5t - 4(cos(\pi t))^2$
   g) $g(x) = sen(\sqrt[3]{x^2}) + \sqrt[3]{sen(x)}$
   h) $f(x) = (\frac{4x^2-3}{3x+4})^3$
   i) $f(x) = \frac{e^{x^2+5}}{\ln(x^4-2)}$
   j) $g(x) = \ln(x^4 \cdot e^{2x^3+8})$

(Derivación implícita)

5. Hallar $\dfrac{\partial y}{\partial x}$ por medio de la derivación implícita y calcular la derivada en el punto indicado:

   a) $xy = 4$, en $(-4, -1)$
   b) $x^2 - y^3 = 0$, en $(1, 1)$
   c) $(x + y)^3 = x^3 + y^3$, en $(-1, 1)$
   d) $x^3 + y^3 = 4xy + 1$, en $(2, 1)$
   e) $x \cos(y) = 1$, en $(2, \frac{\pi}{3})$
   f) $e^{xy} = x^2 + y^2$ en $(0, 1)$

(Extremos en un intervalo. Monotonía de una función)

6. Hallar los extremos absolutos de la función en el intervalo cerrado y estudia la monotonía y los extremos en todo su dominio.

   a) $f(x) = \dfrac{2x + 5}{3}$, en $[0, 5]$
   b) $f(x) = x^3 - 12x$, en $[0, 4]$
   c) $f(x) = \dfrac{2x}{x^2 + 1}$, en $[-2, 2]$

7. En cada una de las siguientes funciones se pide: encontrar los puntos críticos (si los hay), determinar el(los) intervalo(s) abierto(s) sobre los cuales la función es creciente o decreciente e indicar cuáles son los extremos relativos. En definitiva, estudiar la monotonía de la función y determinar sus extremos relativos (si los hay).

   a) $f(x) = x^2 + 8x + 10$
   b) $f(x) = (x + 2)^2(x - 1)$
   c) $f(x) = \dfrac{x}{x + 1}$
   d) $f(x) = \dfrac{x^2 - 3x - 4}{x - 2}$

(Concavidad y el criterio de la segunda derivada)

8. En las siguientes funciones, analiza la concavidad de la gráfica de la función, es decir, indica el(los) intervalo(s) abierto(s) sobre los cuales la función es cóncava o convexa y cuáles son los puntos de inflexión (si los hay).

   a) $f(x) = 2x^3 - 3x^2 - 12x + 5$
   b) $f(x) = x^3(x - 4)$
   c) $g(x) = \dfrac{x + 1}{\sqrt{x}}$

9. En las siguientes funciones, estudia la curvatura y los puntos de inflexión .

   a) $f(x) = x^2 + 3x - 8$
   b) $g(x) = \sqrt{x^2 + 1}$
   c) $f(x) = \dfrac{x}{x - 1}$

(Algunos ejercicios que han aparecido en exámenes)

10. (2010) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{x}{x^2 - 4}$.

11. (2011) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{1}{x-1}$.

12. (2011) Calcula la ecuación de la recta tangente a la curva $x^2 + 2x^2y^2 + y = 11$ en el punto (1,2).

13. (2012) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{x^3}{(x-1)^2}$.

14. (2012) Calcula la ecuación de la recta tangente a la curva $x^2y - 3xy^2 + y^2 = -1$ en el punto (1,1).

15. (2013) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{x}{x^2-1}$.

16. (2013) Calcula la ecuación de la recta tangente a la curva $x^3 + 3x^3y^2 + y = 15x + 3$ en el punto (2,1).

17. (2014) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función f(x)= xe^x.

18. (2014) Calcula la ecuación de la recta tangente a la curva $x^2 + 2x^2y^3 + y = 2x + 23$ en el punto (1,2).

19. (2014) Estudia la monotonía y la curvatura de la función $f(x) = \frac{x^3}{x^2-1}$.

20. (2015) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = x^3 - 3x + 2$.

21. (2015) Calcula la ecuación de la recta tangente a la curva $x^2 + xy + 2xy^2 = 10x$ en el punto (1,2).

22. (2015) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{x^2}{x^2-9}$.

23. (2016) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{x+1}{x-1}$.

24. (2016) Calcula la ecuación de la recta tangente a la curva $3x^2y + xy^2 = 3x + y$ en el punto (1,-3).

25. (2016) Estudia los intervalos de crecimiento, decrecimiento, concavidad y convexidad de la función $f(x) = \frac{2(x+3)}{x-1}$.

1.8. Soluciones del Seminario de Cálculo Diferencial

1a) y-3=3(x-1)
1b) y+1=2x
2a) $f'(x) = 6x^2 - 2x + 3$
2b) $g'(x) = \frac{1}{4\sqrt[4]{x^3}}$
2c) $h'(t) = -\pi sen(t)$
2d) $f'(y) = cos(y)$
2e) $f'(x) = \frac{-15}{8x^4} - 2sen(x)$
2f) $f'(x) = 1 - \frac{2}{x^3}$
2g) $h'(x) = \frac{2x^2-1}{x^2}$
2h) $f'(x) = \frac{-2}{3\sqrt[3]{x^4}} - 3sen(x)$
2i) $f'(x) = \frac{cos(x)\ x - sen(x)}{x^2}$
2j) $f'(x) = x\ cos(x)$
2k) $f'(x) = cos^2(x) - sen^2(x)$
2l) $f'(x) = \frac{-3cos^2(x)+3sen(x)-3sen^2(x)}{2cos^2(x)}$
3a) $f'(x) = 2x - 2$, $f''(x) = 2$, $f'''(x) = 0$
3b) $f'(x) = \frac{2}{\sqrt{x}}$, $f''(x) = \frac{-1}{\sqrt{x^3}}$
3c) $f'(x) = \frac{2}{x^2}$, $f''(x) = \frac{-4}{x^3}$, $f'''(x) = \frac{12}{x^4}$
4a) $f'(x) = -40x(5 - x^2)^4$
4b) $f'(x) = \frac{3x^2-3}{2\sqrt{x^3-3x+2}}$
4c) $g'(t) = \frac{-t}{\sqrt{(t^2+4)^3}}$
4d) $f'(x) = \frac{-x^4+4}{\sqrt{(x^4+4)^3}}$
4e) $f'(x) = 2xtg(\frac{1}{x}) - \frac{1}{cos^2(\frac{1}{x})}$
4f) $h'(t) = 5 + 8\pi cos(\pi t)sen(\pi t)$
4g) $g'(x) = \frac{2cos(\sqrt[3]{x^2})}{3\sqrt[3]{x}} + \frac{cos(x)}{3\sqrt[3]{sen^2(x)}}$
4h) $f'(x) = \frac{3(4x^2-3)^2(12x^2+32x+9)}{(3x+4)^4}$
4i) $f'(x) = \frac{e^{x^2+5}2xln(x^4-2)-e^{x^2+5}\frac{4x^3}{x^4-2}}{(ln(x^4-2))^2}$
4j) $g'(x) = \frac{4x^3e^{2x^3+8}+x^4e^{2x^3+8}6x^2}{x^4e^{2x^3+8}}$
5a) $y' = \frac{-y}{x}; y'(-4,-1) = \frac{-1}{4}$
5b) $y' = \frac{2x}{3y^2};\ y'(1,1) = \frac{2}{3}$
5c) $y' = \frac{x^2-(x+y)^2}{(x+y)^2-y^2};\ y'(-1,1) = -1$
5d) $y' = \frac{4y-3x^2}{3y^2-4x};\ y'(2,1) = \frac{8}{5}$
5e) $y' = \frac{-cos(y)}{-xsen(y)};\ y'(2,\frac{\pi}{3}) = \frac{\sqrt{3}}{6}$
5f) $y' = \frac{2x-ye^{xy}}{xe^{xy}-2y};\ y'(0,1) = \frac{1}{2}$
6a) Mínimo= $(0,\frac{5}{3})$ Máximo= $(5,5)$
6b) Mínimo= $(2,-16)$ Máximo= $(4,16)$
6c) Mínimo= $(-1,-1)$ Máximo= $(1,1)$
7a) f es creciente en $]-4,+\infty[$ y decreciente en $]-\infty,-4[$. Mínimo: x=-4.
7b) f es creciente en $]-\infty,-2[\cup]0,+\infty[$ y decreciente en $]-2,0[$. Mínimo: x=0 y Máximo: x=-2.
7c) f es creciente en $]-\infty,-1[\cup]-1,+\infty[$.

7d) f es creciente en $]-\infty, 2[\cup]2, +\infty[$.
8a) f es cóncava en $]-\infty, \frac{1}{2}[$ y convexa en $]\frac{1}{2}, +\infty[$. Punto de inflexión: x=$\frac{1}{2}$
8b) f es cóncava en $]0, 2[$ y convexa en $]-\infty, 0[\cup]2, +\infty[$. Puntos de inflexión: x=0 y x=2.
8c) f es cóncava en $]3, +\infty[$ y convexa en $]0, 3[$. Punto de inflexión: x= 3.
9a) f es convexa en $\Re$
9b) g es convexa en $\Re$
9c) f es cóncava en $]-\infty, 1[$ y convexa en $]1, +\infty[$.
10) f es decreciente en $]-\infty, -2[\,\cup\,]-2, 2[\,\cup\,]2, +\infty[$. No tiene extremos.
f es cóncava en $]-\infty, -2[\,\cup\,]0, 2[$ y convexa en $]-2, 0[\,\cup\,]2, +\infty[$. Punto de inflexión x=0.
11) f es decreciente en $]-\infty, 1[\,\cup\,]1, +\infty[$. No tiene extremos.
f es cóncava en $]-\infty, 1[$ y convexa en $]1, +\infty[$. No tiene puntos de inflexión.
12) $y - 2 = -2 * (x - 1)$
13) f es creciente en $]-\infty, 1[\,\cup\,]3, +\infty[$ y decreciente en $]1, 3[$. Mínimo: x=3.
f es cóncava en $]-\infty, 0[$ y convexa en $]0, 1[\,\cup\,]1, \infty[$. Punto de inflexión: x=0.
14) $y - 1 = \frac{-1}{3}(x - 1)$
15) f es decreciente en $]-\infty, -1[\,\cup\,]-1, 1[\,\cup\,]1, +\infty[$. No tiene extremos.
f es cóncava en $]-\infty, -1[\,\cup\,]0, 1[$ y convexa en $]-1, 0[\,\cup\,]1, +\infty[$. Punto de inflexión x=0.
16) $y - 1 = \frac{-33}{49}(x - 2)$
17) f es decreciente en $]-\infty, -1[$. f es creciente en $]-1, +\infty[$. Mínimo: x= -1.
f es cóncava en $]-\infty, -2[$ y convexa en $]-2, +\infty[$. Punto de inflexión x= -2.
18) $y - 2 = \frac{-32}{25}(x - 1)$
19) f es decreciente en $]-\sqrt{3}, -1[\,\cup\,]-1, 1[\,\cup\,]1, \sqrt{3}[$. f es creciente en $]-\infty, -\sqrt{3}[\,\cup\,]\sqrt{3}, +\infty[$.
Máximo= x= $-\sqrt{3}$. Mínimo: x= $\sqrt{3}$.
f es cóncava en $]-\infty, -1[\,\cup\,]0, 1[$ y convexa en $]-1, 0[\,\cup\,]1, +\infty[$. Punto de inflexión x= 0.
20) f es creciente en $]-\infty, -1]\ \cup\ [1, \infty[$. f es decreciente en $]-1, 1[$.
Máximo= x= -1. Mínimo: x= 1.
f es cóncava en $]-\infty, 0[$ y convexa en $]0, +\infty[$. Punto de inflexión x= 0.
21) $y - 2 = \frac{-2}{9}(x - 1)$
22) f es creciente en $]-\infty, -3[\,\cup\,]-3, 0]$. f es decreciente en $]0, 3[\,\cup\,]3, \infty[$.
Máximo= x= 0.
f es cóncava en $]-3, 3[$ y convexa en $]-\infty, -3[\,\cup\,]3, \infty[$. No hay puntos de inflexión.
23) f es decreciente en $]-\infty, 1[\,\cup\,]1, \infty[$. No hay extremos.
f es cóncava en $]-\infty, 1[$. f es convexa en $]1, \infty[$. No hay puntos de inflexión.
24) $y + 3 = -3(x - 1)$
25) f es decreciente en $]-\infty, 1[\,\cup\,]1, \infty[$. No hay extremos.
f es cóncava en $]-\infty, 1[$. f es convexa en $]1, \infty[$. No hay puntos de inflexión.

Capítulo 2

Cálculo Integral

2.1. Primitivas o antiderivadas

Se dice que una función F es una **primitiva** o **antiderivada** de f, en un intervalo I si F'(x)=f(x) para todo x en I.

Ejemplo 2.1.

Encontrar una función F que sea una primitiva de f(x)=$3x^2$.

Solución:
Por ejemplo F(x)=x^3, ya que F'(x)=$3x^2$=f(x). Nótese que F(x)=x^3+5 también sería una primitiva de f(x), ya que F'(x)=$3x^2$=f(x).

Esto motiva el siguiente teorema.

Teorema
Si F es una primitiva de f en un intervalo I, entonces G es una primitiva de f en el intervalo I si y sólo si G es de la forma G(x)=F(x)+C, para todo x en I, donde C es una constante.

El teorema anterior sirve para representar la **familia completa de primitivas** de una función agregando una constante a una primitiva conocida, F. La familia de funciones, F, es la **primitiva general de f** y la constante C recibe el nombre de **constante de integración**.

La operación que determina la familia completa de primitivas de una función se denomina **integral indefinida** o **antiderivación** y se denota mediante un signo integral $\int$. De manera que la primitiva de f con respecto a x se calcula como:

$$\int f(x)dx = F(x) + C$$

2.1.1. Propiedades

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

$$\int f(x) * g(x) dx \neq \int f(x) dx * \int g(x) dx$$

$$\int f(x)/g(x) dx \neq \int f(x) dx / \int g(x) dx$$

2.2. Reglas de integración inmediata

Funciones simples	Funciones compuestas
$\int dx = x + C$	
$\int k dx = kx + C$	
$\int x^n dx = \frac{x^{n+1}}{n+1} + C$	$\int u^n \cdot u' dx = \frac{u^{n+1}}{n+1} + C$
$\int \frac{1}{x} dx = ln\vert x\vert + C$ (*)	$\int \frac{u'}{u} dx = ln\vert u\vert + C$
$\int e^x dx = e^x + C$	$\int e^u \cdot u' dx = e^u + C$
$\int a^x dx = \frac{a^x}{ln(a)} + C$	$\int a^u \cdot u' \cdot dx = \frac{a^u}{ln(a)} + C$
$\int cos(x) dx = sen(x) + C$	$\int cos(u) \cdot u' \cdot dx = sen(u) + C$
$\int sen(x) dx = -cos(x) + C$	$\int sen(u) \cdot u' \cdot dx = -cos(u) + C$
$\int \frac{1}{cos^2(x)} dx = tg(x) + C$	$\int \frac{u'}{cos^2(u)} dx = tg(u) + C$
$\int (1 + tg^2(x)) dx = tg(x) + C$	$\int (1 + tg^2(u)) \cdot u' \cdot dx = tg(u) + C$
$\int \frac{-1}{sen^2(x)} dx = cotg(x) + C$	$\int \frac{-u'}{sen^2(u)} dx = cotg(u) + C$
$\int \frac{1}{1+x^2} dx = arctg(x) + C$	$\int \frac{u'}{1+u^2} dx = arctg(u) + C$
$\int \frac{-1}{1+x^2} dx = arccotg(x) + C$	$\int \frac{-u'}{1+u^2} dx = arccotg(u) + C$
$\int \frac{1}{\sqrt{1-x^2}} dx = arcsen(x) + C$	$\int \frac{u'}{\sqrt{1-u^2}} dx = arcsen(u) + C$
$\int \frac{-1}{\sqrt{1-x^2}} dx = arccos(x) + C$	$\int \frac{-u'}{\sqrt{1-u^2}} dx = arccos(u) + C$

(*)
Para x>0: $\int \frac{1}{x} dx = lnx + C$.
Para x<0: $\int \frac{1}{x} dx = \int \frac{-dx}{-x} = ln(-x) + C$
Luego $\forall$ x $\in \Re$, x$\neq$0: $\int \frac{1}{x} dx = ln|x| + C$

Ejemplo 2.2.

Calcula la familia de primitivas de:
a)f(x)= $(x^2+5)^3x$
b)f(x)= $\frac{2x}{x^2}$
c)f(x)= $e^{x^2}2x$
d)f(x)= $5^{x^2}2x$
e)f(x)= $cos(x^3+2)x^2$
f)f(x)= $\frac{3x^2}{1+(x^3)^2}$

Solución:

a) $\int(x^2+5)^3xdx = \frac{(x^2+5)^4}{2\cdot 4} + C$
b) $\int \frac{2x}{x^2}dx = ln(x^2) + C$
c) $\int e^{x^2}2xdx = e^{x^2} + C$
d) $\int 5^{x^2}2xdx = \frac{5^{x^2}}{ln(5)} + C$
e) $\int cos(x^3+2)x^2dx = \frac{sen(x^3+2)}{3} + C$
f) $\int \frac{3x^2}{1+(x^3)^2}dx = arctg(x^3) + C$

Uno de los pasos más importantes antes de realizar una integral es reescribir el integrando en una forma que corresponda con las reglas básicas de integración.

Ejemplo 2.3.

Calcula la familia de primitivas de la función f(x)= $\frac{x+1}{\sqrt{x}}$***.***

Solución:

$\int \frac{x+1}{\sqrt{x}}dx = \int(\frac{x}{\sqrt{x}} + \frac{1}{\sqrt{x}})dx = \int(x^{1/2} + x^{-1/2})dx =$

$\frac{x^{3/2}}{3/2} + \frac{x^{1/2}}{1/2} + C = \frac{2}{3}x^{3/2} + 2x^{1/2} + C = \frac{2}{3}\sqrt{x}(x+3) + C$

Como ya se ha señalado antes, cuando se integran cocientes, no se debe integrar por separado el numerado y el denominador. Esto es incorrecto tanto en la integración como en la derivación.

Autoevaluación

1. Calcula la familia de primitivas de las siguientes funciones:

   a) $\int 3x^2 - 1dx$ Sol: $x^3 - x + C$

   b) $\int(t^2+1)^2dt$ Sol: $\frac{t^5}{5} + \frac{2t^3}{3} + t + C$

   c) $\int \frac{x^2-1}{x^{3/2}}dx$ Sol: $\frac{2(x^2+3)}{3\sqrt{x}} + C$

d) $\int 2sen(t)dt$ Sol: $-2cos(t) + C$

e) $\int tcos(t^2)dt$ Sol: $\frac{1}{2}sen(t^2) + C$

f) $\int 3x^4 + 2x^2 + \frac{3}{x^4}dx$ Sol.: $\frac{3x^5}{5} + \frac{2x^3}{3} - \frac{1}{x^3} + C$

g) $\int (x^3+4)^5 x^2 dx$ Sol.: $\frac{(x^3+4)^6}{18} + C$

h) $\int (x^2+5)^3 x dx$ Sol.: $\frac{(x^2+5)^4}{8} + C$

i) $\int \frac{x^2}{x^3+5}dx$ Sol.: $\frac{ln|x^3+5|}{3} + C$

j) $\int \frac{x}{3+x^2}dx$ Sol.: $\frac{ln|3+x^2|}{2} + C$

k) $\int e^{4x^3}x^2dx$ Sol.: $\frac{e^{4x^3}}{12} + C$

l) $\int 7^{4x^3}x^2dx$ Sol.: $\frac{7^{4x^3}}{12ln(7)} + C$

m) $\int cos(x^3)x^2dx$ Sol.: $\frac{sen(x^3)}{3} + C$

n) $\int sen(4x^2+1)2xdx$ Sol.: $\frac{-cos(4x^2+1)}{4} + C$

ñ) $\int \frac{1}{cos^2(3x+2)}dx$ Sol.: $\frac{tg(3x+2)}{3} + C$

o) $\int \frac{1}{sen^2(3x+2)}dx$ Sol.: $\frac{-cotg(3x+2)}{3} + C$

p) $\int \frac{x^3}{1+(2x^4+1)^2}dx$ Sol.: $\frac{arctg(2x^4+1)}{8} + C$

q) $\int \frac{-x}{1+(x^2+2)^2}dx$ Sol.: $\frac{arccotg(x^2+2)}{2} + C$

r) $\int \frac{1}{\sqrt{1-(3x)^2}}dx$ Sol.: $\frac{arcsen(3x)}{3} + C$

s) $\int \frac{-1}{\sqrt{1-9x^2}}dx$ Sol.: $\frac{arccos(3x)}{3} + C$

2.3. Técnicas de Integración

A continuación, introducimos algunas de las técnicas básicas de integración, apropiadas en el caso en que no es posible calcular la primitiva de una función mediante las reglas de integración inmediata.

2.3.1. Método de integración por partes

Esta técnica puede aplicarse a una amplia variedad de funciones y es particularmente útil para integrandos que contengan productos de funciones algebraicas y trascendentes.

Está basada en la fórmula de la derivada de un producto (uv)'=u'v+uv', donde u y v son derivables respecto de x.

Si u' y v' son continuas, se pueden integrar ambos lados de esta ecuación para obtener
$uv = \int u'vdx + \int uv'dx = \int vdu + \int udv$.
Volviendo a escribir la ecuación se obtiene que:
Si u y v son funciones de x y tienen derivadas continuas, entonces

$$\int udv = uv - \int vdu$$

Regla memotécnica para recordar la fórmula:

$\int$ **un día vi** = **una vaca** $- \int$ **v**estida **de** **u**niforme

Ejemplo 2.4.

Calcula $\int x^2 lnx dx$

Solución:
Como en este caso x^2 se puede integrar directamente y ln(x) no, identificamos los términos de la fórmula de integración por partes como sigue:
$u = lnx \rightarrow du = \frac{1}{x}dx$
$dv = x^2dx \rightarrow v = \int x^2dx = \frac{x^3}{3}$

Aplicando la fórmula se obtiene:
$\int x^2lnxdx = \int udv = uv - \int vdu = lnx\ \frac{x^3}{3} - \int \frac{x^3}{3}\frac{1}{x}dx = lnx\ \frac{x^3}{3} - \frac{1}{3}\int x^2dx = lnx\frac{x^3}{3} - \frac{x^3}{9} + C$

Autoevaluación

1. En cada apartado, calcula la primitiva de la función utilizando integración por partes:
   a) $\int xe^x dx$. Sol: $xe^x - e^x + C$
   b) $\int x^2 sen(x)dx$. Sol: $-x^2cos(x) + 2xsen(x) + 2cos(x) + C$
   c) $\int ln(x)dx$. Sol: $-x + xln(x) + C$

2.3.2. Método de integración para funciones racionales

Método para resolver integrales de la forma $\int \frac{P(x)}{Q(x)}dx$, donde $P(x)$ y $Q(x)$ son polinomios.
Si $gradP \geq gradQ$, se efectúa la división (euclídea) de polinomios:

$$\int \frac{P(x)}{Q(x)}dx = \int C(x)dx + \int \frac{R(x)}{Q(x)}dx.$$

Si $gradP < gradQ$, se efectúa la descomposición factorial de $Q(x)$:

$$Q(x) = (x - a_1)(x - a_2)\cdots(x - a_n)$$

donde a_i son las raíces (puede que, a pares, aparezcan raíces complejas, se analiza como un caso independiente).

<u>**Caso 1:**</u>
Si las raíces de $Q(x)$ son reales y distintas, se aplica la siguiente identificación:

$$\frac{P(x)}{Q(x)} = \frac{A_1}{x - a_1} + \frac{A_2}{x - a_2} + \cdots + \frac{A_n}{x - a_n}$$

A continuación, se determinan los coeficientes indeterminados A_i, efectuando la suma de fracciones simples:

$$\frac{P(x)}{Q(x)} = \frac{A_1(x-a_2)\cdots(x-a_n)+A_2(x-a_1)\cdots(x-a_n)}{Q(x)} + \cdots + \frac{A_n(x-a_1)\cdots(x-a_{n-1})}{Q(x)}$$

La integral produce como resultado:

$$\int \frac{P(x)}{Q(x)}dx = A_1 \ln|x-a_1| + A_2 \ln|x-a_2| + \cdots + A_n \ln|x-a_n|$$

Caso 2:

Si el denominador tiene también raíces múltiples del tipo $(x-b)^k$, por cada una de ellas añadimos a la suma de fracciones simples del caso anterior las siguientes:

$$\frac{B_1}{(x-b)} + \frac{B_2}{(x-b)^2} + \cdots + \frac{B_k}{(x-b)^k}$$

obteniendo la integral como suma de logaritmos y potencias de exponente negativo.

Caso 3:

Si en el denominador aparece un factor cuadrático irreducible, ax^2+bx+c, con raíces imaginarias, añadimos a la suma de fracciones de los casos anteriores una fracción del tipo $\frac{Mx+N}{ax^2+bx+c}$. Una vez identificados los coeficientes M y N, a dicha fracción le corresponderán integrales del tipo logarítmico y arco tangente.

Ejemplo 2.5.

Calcula $\int \frac{x^4-4x^2+x+1}{x^3+x^2-4x-4}dx$.

Solución:
Dado que $gradP \geq gradQ$, efectuamos la división euclídea hasta que el resto sea de menor grado que el denominador:

$$\begin{array}{lll|l}
x^4 & & -4x^2+\ x+1 & x^3+x^2-4x-4 \\ \cline{4-4}
x^4 & +x^3 & -4x^2-4x & x-1 \\
 & -x^3 & \quad\ +5x+1 & \\
 & -x^3 & -\ x^2+4x+4 & \\
 & & \quad x^2+\ x-3 &
\end{array}$$

Es decir:

$$\frac{x^4-4x^2+x+1}{x^3+x^2-4x-4} = x-1+\frac{x^2+x-3}{x^3+x^2-4x-4}.$$

Por tanto:

$$\int \frac{x^4-4x^2+x+1}{x^3+x^2-4x-4}dx = \frac{x^2}{2} - x + \int \frac{x^2+x-3}{x^3+x^2-4x-4}dx$$

Nótese que ahora en la segunda integral, el numerador es de grado menor que el denominador.

Se descompone $x^3+x^2-4x-4 = (x-2)(x+2)(x+1)$ y, por lo tanto:

$$\frac{x^2+x-3}{x^3+x^2-4x-4} = \frac{A_1}{x-2} + \frac{A_2}{x+2} + \frac{A_3}{x+1} =$$

$$= \frac{A_1(x+2)(x+1)+A_2(x-2)(x+1)+A_3(x-2)(x+2)}{(x-2)(x+2)(x+1)}$$

Identificando numeradores se tendrá:

$$x^2+x-3 = A_1(x+2)(x+1)+A_2(x-2)(x+1)+A_3(x-2)(x+2)$$

Para $x=2$, será $3=12A_1$, para $x=-2$, $\ -1=4A_2$, y para $x=-1$, $\ -3=-3A_3$.

Del sistema $\left\{\begin{array}{rcl} 3 & = & 12A_1 \\ -1 & = & 4A_2 \\ -3 & = & -3A_3 \end{array}\right.$, obtenemos: $A_1 = \frac{1}{4}, A_2 = -\frac{1}{4}, A_3 = 1$.

De ese modo, la integral descompuesta en fracciones simples queda:

$$\int \frac{x^2+x-3}{x^3+x^2-4x-4}dx = \frac{1}{4}\int \frac{1}{x-2}dx - \frac{1}{4}\int \frac{1}{x+2}dx + \int \frac{1}{x+1}dx =$$

$$= \frac{1}{4}\ln|x-2| - \frac{1}{4}\ln|x+2| + \ln|x+1| + C$$

Ejemplo 2.6.

Calcula $\int \frac{x^2+x+3}{x^3+3x^2-4}dx$.

Solución:
Descomponemos el denominador $x^3+3x^2-4=(x+2)^2(x-1)$
Las fracciones simples serán:

$$\frac{x^2+x+3}{x^3+3x^2-4}=\frac{A}{x-1}+\frac{B_1}{(x+2)}+\frac{B_2}{(x+2)^2}=\frac{A(x+2)^2+B_1(x-1)(x+2)+B_2(x-1)}{(x-1)(x+2)^2}$$

Identificando numeradores, $x^2+x+3=A(x+2)^2+B_1(x-1)(x+2)+B_2(x-1)$,
para $x=1,\ 5=9A$ para $x=-2,\ 5=-3B_2$, y (p. ej.) para $x=0,\ 3=4A-2B_1-B_2$.

El sistema será:

$$\left\{\begin{array}{lcl} 5 & = & 9A \\ 5 & = & -3B_2 \\ 3 & = & 4A-2B_1-B_2 \end{array}\right.$$

de donde $A=\frac{5}{9}, B_1=\frac{4}{9}$, y $B_2=-\frac{5}{3}$.
Finalmente,

$$I=\frac{5}{9}\int\frac{dx}{x-1}+\frac{4}{9}\int\frac{dx}{x+2}-\frac{5}{3}\int\frac{dx}{(x+2)^2}=\frac{5}{9}\ln|x-1|+\frac{4}{9}\ln|x+2|-\frac{5}{3}\frac{(x+2)^{-2+1}}{-2+1}+C$$

Ejemplo 2.7.

Calcular $\int \frac{6x^2+13x+19}{x^3+3x^2+5x+3}dx$.

Solución:
La descomposición del denominador es $x^3 + 3x^2 + 5x + 3 = (x + 1)(x^2 + 2x + 3)$. Por lo tanto, al factor cuadrático le corresponden raíces complejas. Identificamos $\frac{6x^2+13x+19}{x^3+2x^2+2x+1} = \frac{A}{x+1} + \frac{Mx+N}{x^2+2x+3}$ y de aquí obtenemos la igualdad para los numeradores:

$$6x^2 + 13x + 19 = A(x^2 + 2x + 3) + (Mx + N)(x + 1)$$

Tomando $x = -1$, determinamos A: $12 = 2A$. Podemos construir un sistema con resolución escalonada si, a continuación, tomamos $x = 0$. En efecto, ahora tendremos la ecuación $19 = 3A + N$ y, en consecuencia $N = 1$. Por último, podríamos tomar $x = 1$, por ejemplo, para completar el sistema:

$$\begin{cases} 12 = 2A \\ 19 = 3A + N \\ 38 = 6A + 2N + 2M \end{cases}$$

y obtener $M = 0$. De ese modo:

$$\int \frac{6x^2 + 13x + 19}{x^3 + 3x^2 + 5x + 3}dx = \underbrace{6\int \frac{dx}{x+1}}_{I_1} + \underbrace{\int \frac{1}{x^2 + 2x + 3}dx}_{I_2}.$$

La primera integral, $I_1 = 6\ln(x + 1)$.
La segunda integral, $I_2 = \int \frac{1}{x^2+2x+3}dx = \int \frac{1}{(x+1)^2+2}dx = \int \frac{1}{2+(x+1)^2}dx = \int \frac{1}{2}\frac{1}{\frac{2}{2}+\frac{(x+1)^2}{2}}dx = \frac{1}{2}\int \frac{1}{1+(\frac{x+1}{\sqrt{2}})^2}dx = \frac{\sqrt{2}}{2}\int \frac{\frac{1}{\sqrt{2}}}{1+(\frac{x+1}{\sqrt{2}})^2}dx = \frac{\sqrt{2}}{2}arctg(\frac{x+1}{\sqrt{2}}) + C$

Autoevaluación

1. En cada apartado, calcula la primitiva de la función:

   a) $\int \frac{2x^3-4x^2-15x+5}{x^2-2x-8}dx$. Sol: $x^2 + \frac{3}{2}ln|x - 4| - \frac{1}{2}ln|x + 2| + C$

   b) $\int \frac{1}{x^2-5x+6}dx$. Sol: $ln|x - 3| - ln|x - 2| + C$

   c) $\int \frac{5x^2+20x+6}{x^3+2x^2+x}dx$. Sol: $6ln|x| - ln|x + 1| - \frac{9}{x+1} + C$

   d) $\int \frac{2x^3-4x-8}{(x^2-x)(x^2+4)}dx$. Sol: $2ln|x| - 2ln|x - 1| + ln(x^2 + 4) + 2arctg(\frac{x}{2}) + C$

2.3.3. Método de integración por sustitución

Este método está basado en el siguiente resultado teórico.

Teorema (Antiderivación de una función compuesta)
Sea g una función cuyo recorrido o rango es un intervalo I, y sea f una función continua en I. Si g es derivable en su dominio y F es una antiderivada o primitiva de f en I, entonces

$\int f(g(x))g'(x)dx = F(g(x)) + C$. Si t=g(x), entonces dt=g'(x)dx y $\int f(t)dt = F(t) + C$

Esta metodología se utiliza especialmente en la resolución de integrales con los siguientes patrones. En cada caso, se especifica el cambio de variable más aconsejable. Notación: R(f(x), g(x)) significa función radical entre las funciones f(x) y g(x). Recuérdese que una función, f(x) es par si f(-x)= f(x) y es impar si f(-x)= -f(x).

1. $\int R(x, \sqrt{a^2 - x^2})dx$ cambio: $x = a \cdot sen(t)$
2. $\int R(x, \sqrt{a^2 + x^2})dx$ cambio: $x = a \cdot tg(t)$
3. $\int R(x, \sqrt{x^2 - a^2})dx$ cambio: $x = a \cdot sec(t)$
4. $\int R(x, \sqrt[n]{\frac{ax+b}{cx+d}})dx$ cambio: $t = \frac{ax+b}{cx+d}$
5. $\int R(sen(x), cos(x))dx$ con R(sen(x), cos(x)) par, cambio: $t = tg(x)$.
Esto implica que: $sen(x) = \frac{t}{\sqrt{1+t^2}}$, $cos(x) = \frac{1}{\sqrt{1+t^2}}$ y $dx = \frac{dt}{1+t^2}$
6. $\int R(sen(x), cos(x))dx$ con R(sen(x), cos(x)) no par, cambio: $t = tg(\frac{x}{2})$.
Esto implica que: $sen(x) = \frac{2t}{1+t^2}$, $cos(x) = \frac{1-t^2}{1+t^2}$, $tg(x) = \frac{2t}{1-t^2}$ y $dx = \frac{2dt}{1+t^2}$

Ejemplo 2.8.

Calcula $\int \frac{e^{4x}+3}{e^{3x}}dx$

Solución:
Considerando el cambio $t = e^x \to dt = e^x dx \to dx = \frac{dt}{t}$, la integral se reescribe como:
$\int \frac{t^4+3}{t^3}\frac{1}{t}dt = \int \frac{t^4+3}{t^4}dt = \int dt + \int \frac{3}{t^4}dt = \int dt + \int 3t^{-4}dt = t + 3\frac{t^{-3}}{-3} + C = t - \frac{1}{t^3} + C$
Deshaciendo el cambio: $e^x - \frac{1}{e^{3x}} + C$

Autoevaluación

1. En cada apartado, calcula la primitiva de la función utilizando el cambio de variable adecuado:

   a) $\int \sqrt{2x-1}dx$. Sol: $\frac{1}{3}(2x-1)^{3/2} + C$
   b) $\int x\sqrt{2x-1}dx$. Sol: $\frac{1}{10}(2x-1)^{5/2} + \frac{1}{6}(2x-1)^{3/2} + C$
   c) $\int sen^2(3x)cos(3x)dx$. Sol: $\frac{1}{9}sen^3(3x) + C$

2.4. Áreas

En lo que llevamos de tema, se ha introducido la antiderivación o cálculo de primitivas. A continuación, vamos a considerar el problema de calcular el área de una región en un plano. A simple vista, estas dos ideas parecen no relacionarse, aunque más adelante veremos que se relacionan de manera estrecha por medio de un teorema muy importante conocido como Teorema Fundamental del Cálculo.

En primer lugar, vamos a familiarizarnos con la notación de la suma finita de términos para utilizarla en razonamientos posteriores.

La **suma de n términos**, $a_1, a_2, ..., a_n$, se escribe como

$$a_1 + a_2 + ... + a_n = \sum_{i=1}^{n} a_i$$

donde i es el **índice de la suma**, a_i es el **i-ésimo término** de la suma y los **límites superior** e **inferior** de la suma son n y 1, respectivamente. Los límites superior e inferior de la suma han de ser constantes respecto al índice de la suma. Sin embargo, el límite inferior no tiene por qué ser 1. Cualquier entero menor o igual al límite superior es legítimo. De las propiedades asociativa y conmutativa de la suma y de la propiedad distributiva de la adición sobre la multiplicación, se deduce:

1. $\sum_{i=1}^{n} ka_i = k\sum_{i=1}^{n} a_i$, donde k es constante.
2. $\sum_{i=1}^{n}(a_i \pm b_i) = \sum_{i=1}^{n} a_i \pm \sum_{i=1}^{n} b_i$
3. $\sum_{i=1}^{n} c = cn$
4. $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$
5. $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
6. $\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$

En la geometría euclídea, el tipo más simple de región plana es el rectángulo y la definición de su área es **A=bh**, donde b representa la base del rectángulo y h su altura.

Veamos un procedimiento con el que se puede aproximar el área de una región plana.

Consideremos una región plana limitada en su parte superior por la gráfica de una función continua no negativa y=f(x), en su parte inferior por el eje x y en los laterales por las rectas verticales x=a y x=b, como se muestra en la Figura 2.1.

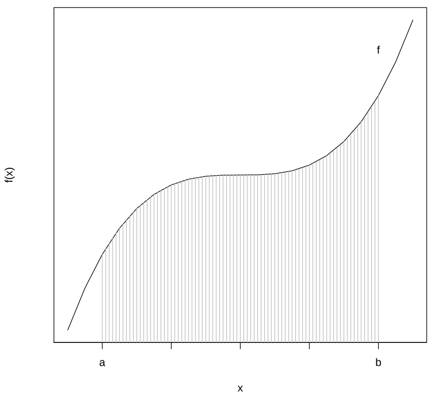


Figura 2.1: Área de una región plana bajo una curva.

Para aproximar el área de la región, se empieza subdividiendo el intervalo [a,b] en n subintervalos, cada uno de longitud $\Delta x = \frac{b-a}{n}$. Los puntos terminales de los intervalos son los siguientes:

$$\underbrace{a + 0(\Delta x)}_{a=x_0} < \underbrace{a + 1(\Delta x)}_{x_1} < \underbrace{a + 2(\Delta x)}_{x_2} < ... < \underbrace{a + n(\Delta x)}_{x_n}$$

Como f es continua, el **teorema del valor extremo**, enunciado en el capítulo anterior, garantiza la existencia de un valor mínimo y un valor máximo de f(x) en cada subintervalo:

f(m_i)= valor mínimo de f(x) en el i-ésimo subintervalo.
f(M_i)= valor máximo de f(x) en el i-ésimo subintervalo.

A continuación, se definen los **rectángulos inscritos** dentro de cada i-ésima región con una altura f(m_i) y los **rectángulos circunscritos** por encima de cada i-ésima región con una altura f(M_i).

Para cada i, el área del rectángulo inscrito es menor o igual que el área del rectángulo circunscrito.

$$\textit{Área del rectángulo inscrito} \;=\; \Delta x f(m_i) \;\leq\; \Delta x f(M_i) = \;\textit{Área del rectángulo circunscrito}$$

La suma de las áreas de los rectángulos inscritos recibe el nombre de **suma inferior** y la suma de las áreas de los rectángulos circunscritos se conoce como **suma superior**.

suma inferior = s(n)= $\sum_{i=1}^{n} \Delta x f(m_i)$
suma superior = S(n)= $\sum_{i=1}^{n} \Delta x f(M_i)$

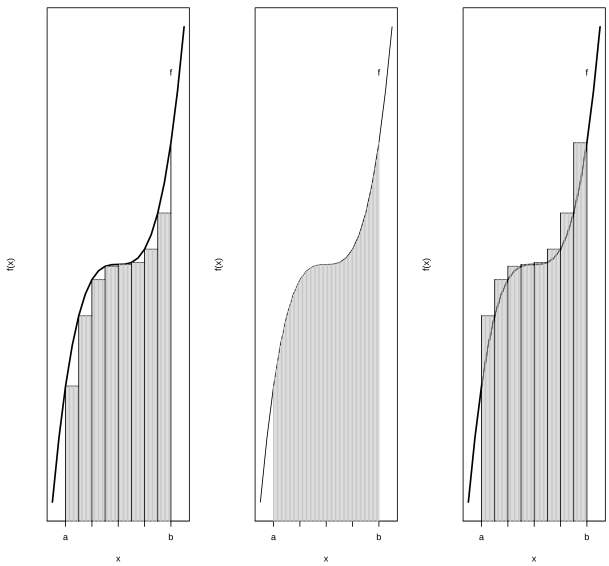


Figura 2.2: Suma inferior, área de la región y suma superior.

En la Figura 2.2 se puede observar que la suma inferior s(n) es menor o igual que la suma superior S(n). Además, el área de la región limitada por la función, f, se encuentra entre estas dos sumas.

$$s(n) \leq \text{Área de la región} \leq S(n)$$

Teorema

Sea f continua y no negativa en el intervalo [a,b]. Los límites cuando n$\rightarrow \infty$ de las sumas inferior y superior existen y son iguales entre sí. Esto es

$$lim_{n\rightarrow\infty} s(n) = lim_{n\rightarrow\infty} \sum_{i=1}^{n} \Delta x f(m_i) = lim_{n\rightarrow\infty} \sum_{i=1}^{n} \Delta x f(M_i) = lim_{n\rightarrow\infty} S(n)$$

donde $\Delta x = \frac{(b-a)}{n}$ y f(m_i) y f(M_i) son los valores mínimo y máximo de f en el subintervalo i-ésimo.

Debido a que se alcanza el mismo límite tanto con el valor mínimo f(m_i) como con el valor máximo f(M_i), se deduce que la elección de x en el i-ésimo intervalo no afecta al límite. Esto significa que se puede elegir cualquier valor arbitrario x en el i-ésimo subintervalo, como apunta la siguiente definición.

Sea f continua y no negativa en el intervalo $[a,b]$. El **área de la región limitada por la gráfica de f**, el eje x y las rectas verticales x=a y x=b es

$$\acute{A}rea = lim_{n\rightarrow\infty} \sum_{i=1}^{n} \Delta x f(c_i), \qquad x_{i-1} \leq c_i \leq x_i$$

donde $\Delta x = \frac{(b-a)}{n}$

Hasta ahora hemos supuesto que las particiones de los subintervalos tienen igual ancho, Δx. En realidad, se ha especificado así por conveniencia de cálculo pero no es necesario que los subintervalos tengan el mismo ancho.

2.5. Sumas de Riemann e integrales definidas

Sea f definida en el intervalo cerrado [a,b], y sea Δ una partición de [a,b] dada por

$$a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b$$

donde Δx_i es el ancho del i-ésimo subintervalo. Si c_i es cualquier punto en el i-ésimo subintervalo entonces la suma

$$\sum_{i=1}^{n} \Delta x_i f(c_i), \qquad x_{i-1} \leq c_i \leq x_i$$

se denomina una **suma de Riemann de f para la partición**, Δ.

El ancho del subintervalo más grande de la partición Δ es la norma de la partición y se denota por $||\Delta||$.

Cuando la **norma** de la partición tiende a cero ($||\Delta|| \rightarrow 0$) implica que el número de subintervalos tiende a infinito ($n \rightarrow \infty$).

Si f se define en un intervalo cerrado $[a,b]$ y el límite $lim_{||\Delta||\to 0}\sum_{i=1}^{n}\Delta x_i f(c_i)$ existe, entonces f es **integrable** en $[a,\ b]$ y el límite se denota por

$$lim_{||\Delta||\to 0}\sum_{i=1}^{n}\Delta x_i f(c_i)=\int_a^b f(x)dx$$

El límite recibe el nombre de **integral definida** de f en a y b. El número a es el **límite inferior de integración**, y el número de b es el **límite superior de integración**.

No es coincidencia que la notación para las integrales definidas sea similar a la que se utilizó para las integrales indefinidas. La razón se verá en la siguiente sección cuando se introduzca el Teorema Fundamental del Cálculo. Es importante observar que las integrales definidas y las indefinidas son identidades diferentes. Una integral definida es un número. Sin embargo, una integral indefinida es una familia de funciones.

Teorema
Si una función f es continua en un intervalo cerrado $[a,b]$, entonces f es integrable en $[a,b]$.
Las integrales definidas pueden ser positivas, negativas o cero. Para que una integral definida sea interpretada como un área, la función f debe ser continua y no negativa en $[a,b]$, como establece el siguiente teorema.

Teorema
Si f es continua y no negativa en un intervalo cerrado $[a,b]$, entonces el área de la región acotada por la gráfica de f, el eje x y las rectas verticales x=a y x=b viene dada por

$$Área=\int_a^b f(x)dx$$

Propiedades de las integrales definidas

1. Si f está definida en x=a, entonces se define $\int_a^a f(x)dx=0$.
2. Si f es integrable en [a,b], entonces se define $\int_b^a f(x)dx=-\int_a^b f(x)dx$.
3. Si f es integrable en los tres intervalos cerrados determinados por a, b y c, entonces $\int_a^b f(x)dx=\int_a^c f(x)dx+\int_c^b f(x)dx$.
4. Si f es integrable en [a,b] y k es una constante, entonces kf es integrable en [a,b] y $\int_a^b kf(x)dx=k\int_a^b f(x)dx$.
5. Si f y g son integrables en [a,b], entonces f±g son integrables en [a,b] y $\int_a^b[f(x)\pm g(x)]dx=\int_a^b f(x)dx\pm\int_a^b g(x)dx$.
   Esta propiedad se puede extender a cualquier número finito de funciones.
6. Si f es integrable y no negativa en el intervalo cerrado [a,b], entonces 0≤ $\int_a^b f(x)dx$.
7. Si f y g son integrables en el intervalo cerrado [a,b] y f(x)≤g(x) para x en [a,b], entonces $\int_a^b f(x)dx\le\int_a^b g(x)dx$.

2.6. Teorema Fundamental del Cálculo

Se han visto ya dos de las principales ramas del cálculo: el cálculo diferencial (presentado con el problema de la recta tangente) y el cálculo integral (presentado con el problema del cálculo del área). En este punto podría parecer que estos dos problemas no se relacionan, aunque tienen una conexión muy estrecha. La conexión fue descubierta independientemente por Newton y Leibniz y está enunciada en el siguiente teorema. De manera informal el teorema establece que la derivación y la integración definida son operaciones inversas, en el mismo sentido que lo son la división y la multiplicación.

Teorema Fundamental del Cálculo

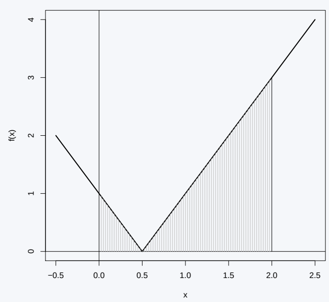
Si una función f es continua en el intervalo cerrado [a,b] y F es una primitiva de f en el intervalo [a,b], entonces

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

Ejemplo 2.9.

Calcula $\int_0^2 |2x-1|dx$

Solución:



Utilizando la Figura y la definición de valor absoluto,

$$f(x) = \begin{cases} -(2x-1), & x < \frac{1}{2} \\ 2x-1, & x \geq \frac{1}{2} \end{cases}$$

se puede reescribir el integrando como se indica:

$\int_0^2 |2x-1|dx = \int_0^{\frac{1}{2}} -(2x-1)dx + \int_{\frac{1}{2}}^2 2x - 1dx = [-x^2+x]_0^{\frac{1}{2}} + [x^2-x]_{\frac{1}{2}}^2 =$

$(-\frac{1}{4}+\frac{1}{2}) - (0+0) + (4-2) - (\frac{1}{4}-\frac{1}{2}) = \frac{5}{2}$

Ejemplo 2.10.

Calcula $\int_0^1 x(x^2+1)dx$

Solución: Considerando el cambio $u = x^2 + 1 \rightarrow du = 2xdx \rightarrow xdx = \frac{1}{2}du$.
El cambio se utiliza para modificar, también, los límites de integración del siguiente modo:
si $x = 0 \rightarrow u = 0^2 + 1 = 1$ y si $x = 1 \rightarrow u = 1^2 + 1 = 2$, de modo que la integral
se reescribe como:
$\int_0^1 x(x^2+1)dx = \frac{1}{2}\int_1^2 udu = \frac{1}{2}[\frac{u^2}{2}]_1^2 = \frac{1}{2}(2 - \frac{1}{2}) = \frac{3}{4}$

Autoevaluación

1. En cada apartado, calcula el área de la región delimitada por la función el eje x y las rectas especificadas:

   a) $f(x) = x + 2$, x=0 y x=3. Sol: $\frac{21}{2}$

   b) $g(t) = -t^2 + 4t - 3, t = 1yt = 3$. Sol: $\frac{4}{3}$

   c) $h(x) = 4 - x^2$, x=-2 y x=2. Sol: $\frac{32}{3}$

   d) $f(x) = \frac{1}{x^2+1}$, x=-1 y x=1 . Sol: $\frac{\pi}{2}$

   e) $f(x) = 2x^2 - 3x + 2$, x=0 y x=2. Sol: $\frac{10}{3}$

2.7. Seminario de CÁLCULO INTEGRAL

(Algunos de estos ejercicios han sido obtenidos del libro **CÁLCULO**, de Larson, Hostetler y Edwards (McGrawHill) *Octava edición*. En este libro se pueden encontrar multitud de ejercicios de cada una de las sesiones de este tema.)

(Reglas de integración: integrales inmediatas)

1. En los siguientes ejercicios encuentra la integral indefinida (y verifica el resultado mediante derivación):

   a) $\int (4-x)\ dx$

   b) $\int (2x^3+4x^2-1)\ dx$

   c) $\int (2x^3-2x+4)\ dx$

   d) $\int \frac{1}{x^2}\ dx$

   e) $\int \frac{3x+2}{\sqrt{x}}\ dx$

   f) $\int (\sqrt{x}+\frac{1}{2\sqrt{x}})\ dx$

   g) $\int (\sqrt[4]{x^3}+3)\ dx$

   h) $\int \frac{1}{x^5}\ dx$

   i) $\int \frac{2x^2+x-4}{x^4}\ dx$

   j) $\int (2t^2-2)^2\ dt$

   k) $\int (1+2t)t^2\ dt$

   l) $\int 2\ dt$

   m) $\int \frac{x^3-2x^2+4x}{x}\ dx$

   n) $\int \frac{2}{\sqrt{x}}\ dx$

   ñ) $\int (4x+2)(x-1)\ dx$

   o) $\int (4x+3)^2\ dx$

   p) $\int (2\sqrt{x}-\sqrt[3]{x}-x^4)\ dx$

   q) $\int (\frac{3}{x}-\frac{x}{3})\ dx$

2. En los siguientes ejercicios encuentra la integral indefinida (y verifica el resultado mediante derivación):

   a) $\int (3\sin(x)+2\cos(x))\ dx$

   b) $\int (x^3+\sin(x))\ dx$

   c) $\int (t^2+\sec^2(t))\ dt$

   d) $\int \frac{\sin^2(x)+\cos^2(x)}{\cos^2(x)}\ dx$

   e) $\int \frac{1}{1+x^2}\ dx$

   f) $\int \frac{1}{\sqrt{1-x^2}}\ dx$

   g) $\int (\sin(x)+7\cos(x)-1)\ dx$

(Integración por partes)

3. En los siguientes ejercicios encuentra la integral indefinida (y verifica el resultado mediante derivación):

a) $\displaystyle\int \frac{2x}{e^x}\ dx$

b) $\displaystyle\int x^4 \ln(x)\ dx$

c) $\displaystyle\int \frac{\ln(x)}{x^2}\ dx$

d) $\displaystyle\int x \sin(x)\ dx$

e) $\displaystyle\int x^2 \cos(x)\ dx$

f) $\displaystyle\int e^x \cos(2x)\ dx$

g) $\displaystyle\int e^{2x} \sin(x)\ dx$

h) $\displaystyle\int (x^2-1)e^x\ dx$

(Fracciones simples)

4. En los siguientes ejercicios encuentra la integral indefinida (y verifica el resultado mediante derivación):

a) $\displaystyle\int \frac{1}{x^2-1}\ dx$

b) $\displaystyle\int \frac{x+1}{x^2+4x+3}\ dx$

c) $\displaystyle\int \frac{x+2}{x^2-4x}\ dx$

d) $\displaystyle\int \frac{5-x}{2x^2+x-1}\ dx$

e) $\displaystyle\int \frac{x+2}{x^2-4x+4}\ dx$

f) $\displaystyle\int \frac{x+1}{x^3+4x^2+5x+2}\ dx$

g) $\displaystyle\int \frac{x^2}{x^4-2x^2-8}\ dx$

h) $\displaystyle\int \frac{6x}{x^3-8}\ dx$

i) $\displaystyle\int \frac{1}{3+x^2}\ dx$

j) $\displaystyle\int \frac{2}{x^2+4}\ dx$

k) $\displaystyle\int \frac{1}{x^2+4x+9}\ dx$

l) $\displaystyle\int \frac{1}{x^2+2x+10}\ dx$

m) $\displaystyle\int \frac{3x^3+5x}{x^2-x-2}\ dx$

n) $\displaystyle\int \frac{x^3+2}{x^2}\ dx$

(Cambio de variable)

5. En los siguientes ejercicios encuentra la integral indefinida (y verifica el resultado mediante derivación):

a) $\displaystyle\int \sqrt[3]{(1-x^2)}(-4x)\ dx$

b) $\displaystyle\int \frac{x^3}{(1+x^4)^2}\ dx$

c) $\displaystyle\int \frac{x^3}{\sqrt{1+x^4}}\ dx$

d) $\displaystyle\int x^3 \sin(x^4)\ dx$

e) $\displaystyle\int \frac{1}{x-5}\ dx$

f) $\displaystyle\int \frac{x^2}{3-x^3}\ dx$

g) $\displaystyle\int \frac{e^{\frac{1}{x^2}}}{x^3}\ dx$

h) $\displaystyle\int \frac{e^{2x}}{1+e^{2x}}\ dx$

(Cálculo de Áreas)

6. En cada uno de los siguientes ejercicios, hallar el área de la región delimitada por la gráfica de la función $f(x)$, el eje x y las rectas $x=a$ y $x=b$ que se indican:

a) $f(x) = 3x^2 + 1$, $x = 0$ y $x = 2$.

b) $f(x) = 1 + \sqrt[3]{x}$, $x = 0$ y $x = 8$.

c) $f(x) = \sin(x)$, $x = 0$ y $x = \pi$.

d) $f(x) = \cos(x)$, $x = 0$ y $x = \frac{\pi}{2}$.

(Algunos ejercicios que han aparecido en exámenes)

7. (2010) Resuelve la integral $\int \frac{x}{x^2-4x+4}dx$

8. (2010) Resuelve la integral $\int xln(2x)dx$

9. (2011) Resuelve la integral $\int \frac{3x}{(x-3)^2}dx$

10. (2011) Halla el área de la región delimitada por la gráfica de la función f(x)=cos(x), el eje x y las rectas $x = \frac{\pi}{2}$ y $x = \frac{3\pi}{2}$

11. (2011) Calcula el área comprendido entre la función f(x)=ln(x), el eje de abscisas y las rectas x= 1 y x= e.

12. (2012) Resuelve la integral $\int \frac{x+1}{x^2-4x}dx$

13. (2013) Calcula la integral $\int \frac{x^2-1}{e^x}dx$

14. (2013) Halla el área de la región delimitada por la gráfica de la función f(x)=$\sqrt{2x+1}$, el eje x y las rectas $x = 0$ y $x = 4$.

15. (2013) Calcula el área comprendida entre la función f(x)=$\frac{x+1}{9-x^2}$, el eje de abscisas y las rectas x=0 y x=2.

16. (2014) Calcula la integral $\int \frac{x}{(x-2)^2}dx$.

17. (2014) Hala el área de la región delimitada por la gráfica de la función f(x)= xln(x), en el eje x y las rectas $x = 1$ y $x = e$.

18. (2014) Resuelve la siguiente integral $\int x^2e^xdx$.

19. (2015) Calcula el área comprendida entre la función f(x)= xe^x y las rectas x=0 y x=1.

20. (2015) Calcula $\int \frac{3x-5}{x^2-5x}dx$.

21. (2015) Calcula $\int xsen(x)dx$.

22. (2016) Calcula $\int \frac{(2x^2+x)}{(x-4)(2+x^2)}dx$.

23. (2016) Calcula $\int xcos(x)dx$.

2.8. Soluciones del Seminario de Cálculo Integral

1a) $4x - \frac{x^2}{2} + C$
1b) $\frac{x^4}{2} + \frac{4x^3}{3} - x + C$
1c) $\frac{x^4}{2} - x^2 + 4x + C$
1d) $\frac{-1}{x} + C$
1e) $2\sqrt{x^3} + 4\sqrt{x} + C$
1f) $\frac{2}{3}\sqrt{x^3} + \sqrt{x} + C$
1g) $\frac{4}{7}\sqrt[4]{x^7} + 3x + C$
1h) $\frac{-1}{4x^4} + C$
1i) $\frac{-2}{x} - \frac{1}{2x^2} + \frac{4}{3x^3} + C$
1j) $\frac{4t^5}{5} - \frac{8t^3}{3} + 4t + C$
1k) $\frac{t^3}{3} + \frac{t^4}{2} + C$
1l) $2t + C$
1m) $\frac{x^3}{3} - x^2 + 4x + C$
1n) $4\sqrt{x} + C$
1ñ) $\frac{4x^3}{3} - x^2 - 2x + C$
1o) $\frac{(4x+3)^3}{12} + C$
1p) $\frac{4}{3}\sqrt{x^3} - \frac{3}{4}\sqrt[3]{x^4} - \frac{x^5}{5} + C$
1q) $3ln|x| - \frac{x^2}{6} + C$
2a) $-3cos(x) + 2sen(x) + C$
2b) $\frac{x^4}{4} - cos(x) + C$
2c) $\frac{t^3}{3} + tg(t) + C$
2d) $tg(x) + C$
2e) $arctg(x) + C$
2f) $arcsen(x) + C$
2g) $-cos(x) + 7sen(x) - x + C$
3a) $-2xe^{-x} - 2e^{-x} + C$
3b) $ln(x)\frac{x^5}{5} - \frac{x^5}{25} + C$
3c) $ln(x)\frac{-1}{x} - \frac{1}{x} + C$
3d) $-xcos(x) + sen(x) + C$
3e) $x^2sen(x) + 2xcos(x) - 2sen(x) + C$
3f) $\frac{e^x(cos(2x)+2sen(2x))}{5} + C$
3g) $\frac{e^{2x}(2sen(x)-cos(x))}{5} + C$
3h) $(x^2-1)e^x - 2xe^x + 2e^x + C$
4a) $\frac{-1}{2}ln|x+1| + \frac{1}{2}ln|x-1| + C$
4b) $ln|x+3| + C$
4c) $\frac{-1}{2}ln|x| + \frac{3}{2}ln|x-4| + C$
4d) $3ln|x-\frac{1}{2}| - 4ln|x+1| + C$
4e) $ln|x-2| - \frac{4}{(x-2)} + C$
4f) $ln|x+1| - ln|x+2| + C$
4g) $\frac{1}{6}ln|x-2| - \frac{1}{6}ln|x+2| + \frac{\sqrt{2}}{6}arctg(\frac{x}{\sqrt{2}}) + C$
4h) $ln|x-2| - \frac{1}{2}ln|x^2+2x+4| + \sqrt{3}arctg(\frac{x+1}{\sqrt{3}}) + C$
4i) $\frac{\sqrt{3}}{3}arctg(\frac{x}{\sqrt{3}}) + C$
4j) $arctg(\frac{x}{2}) + C$

4k) $\frac{\sqrt{5}}{5}arctg(\frac{x+2}{\sqrt{5}})+C$
4l) $\frac{1}{3}arctg(\frac{x+1}{3})+C$
4m) $\frac{3}{2}x^2+3x+\frac{34}{3}ln(x-2)+\frac{8}{3}ln(x+1)+C$
4n) $\frac{x^2}{2}-\frac{2}{x}+C$
5a) $\frac{3(1-x^2)\sqrt{(1-x^2)}}{2}+C$
5b) $\frac{-1}{4(1+x^4)}+C$
5c) $\frac{\sqrt{(1+x^4)}}{2}+C$
5d) $\frac{-1}{4}cos(x^4)+C$
5e) $ln|x-5|+C$
5f) $\frac{-1}{3}ln|3-x^3|+C$
5g) $\frac{1}{2}e^{\frac{1}{x^2}}+C$
5h) $\frac{1}{2}ln|1+e^{2x}|+C$
6a) $[x^3+x]_0^2=10$
6b) $[x+\frac{x^{\frac{4}{3}}}{\frac{4}{3}}]_0^8=20$
6c) $[-cos(x)]_0^{\pi}=2$
6d) $[sen(x)]_0^{\frac{\pi}{2}}=1$
7) $ln|x-2|-\frac{2}{(x-2)}+C$
8) $\frac{x^2}{2}ln|2x|-\frac{x^2}{4}+C$
9) $3ln|x-3|-\frac{9}{x-3}+C$
10) $[sen(x)]_{\frac{\pi}{2}}^{\frac{3\pi}{2}}=2$
11) $[xln(x)-x]_1^e=1$
12) $\frac{-1}{4}ln|x|+\frac{5}{4}ln|x-4|+C$
13) $(1-x^2)e^{-x}-2xe^{-x}-2e^{-x}+C$
14) $[\frac{\sqrt{(2x+1)^3}}{3}]_0^4=\frac{26}{3}$
15) $[\frac{2}{3}ln|3-x|-\frac{1}{3}ln|3+x|]_0^2=ln(3)-\frac{1}{3}ln(5)$
16) $ln|x-2|-\frac{2}{(x-2)}+C$
17) $\frac{e^2+1}{4}$
18) $x^2e^x-2xe^x+2e^x+C$
19) $[xe^x-e^x]_0^1=1$
20) $ln|x|+2ln|x-5|+C$
21) $-xcos(x)+sen(x)+C$
22) $2ln|x-4|+\frac{\sqrt{2}}{2}arctg(\frac{x}{\sqrt{2}})+C$
23) $xsen(x)+cos(x)+C$

Capítulo 3

Ecuaciones Diferenciales

3.1. Introducción

Una **ecuación diferencial** es una ecuación que relaciona de manera no trivial a una función desconocida y una o más derivadas de esta función desconocida con respecto a una o más variables independientes. Por ejemplo, la expresión $(y' - y)^2 = (y')^2 - 2y'y + y^2$, no sería una ecuación diferencial porque cualquier función derivable, "y", verificaría esta expresión.

Si la función desconocida depende de una sola variable la ecuación diferencial se llama **ordinaria**, por el contrario, si depende de más de una variable, se llama **parcial**.

Una función y=f(x) es **solución** de una ecuación diferencial si satisface la expresión de la ecuación diferencial. Por ejemplo, $y = e^{-2x}$ es solución de la ecuación diferencial $y' + 2y = 0$. En realidad, cada solución de esta ecuación diferencial es de la forma $y = Ce^{-2x}$, donde C es cualquier número real. A esta solución se le llama **solución general**. Algunas ecuaciones diferenciales tienen **soluciones singulares** que no se pueden escribir como casos especiales de la solución general.

El **orden** de una ecuación diferencial se determina por la derivada de mayor orden en la ecuación. Por ejemplo, y'=4y es una ecuación diferencial de primer orden y su solución general es $y = Ce^{4x}$. La ecuación s"(t)=-32 es una ecuación diferencial de segundo orden y su solución general es $s(t) = -16t^2 + C_1t + C_2$. Se puede demostrar que una ecuación diferencial de orden n tiene una solución general con n constantes arbitrarias.

Geométricamente, la solución general de una ecuación diferencial de primer orden representa una familia de curvas conocidas como **curvas solución**, una para cada valor asignado a la constante arbitraria. Las **soluciones particulares** de la ecuación diferencial se obtienen de las **condiciones iniciales**, que para un valor particular de la variable independiente proporciona valores para la variable dependiente y para sus derivadas.

Ejemplo 3.1.

Dada la ecuación diferencial xy' - 3y=0, verifica que y=Cx^3 es una solución, y encuentra la solución particular determinada por la condición inicial y=2 cuando x=-3.

Solución:

y=Cx^3 es una solución de la ecuación diferencial, ya que y'=3Cx^2 y xy'-3y=x(3Cx^2)-3(Cx^3)=0. De la condición inicial (-3,2) se deduce que 2=C$(-3)^3$ $\rightarrow$ C=$\frac{-2}{27}$. Esta constante determina la solución particular $y = \frac{-2x^3}{27}$.

NOTA: Para determinar una solución particular, el número de condiciones iniciales debe corresponder al número de constantes en la solución general.

Autoevaluación

1. Determina si las siguientes funciones son solución de la ecuación diferencial y"-y=0:

   a) y=senx. Sol: No es solución.

   b) $y = 4e^{-x}$. Sol: Sí es solución.

   c) $y = Ce^x$. Sol: Sí es solución.

2. En los siguientes apartados verifica que la función dada es solución de la ecuación diferencial:

	Ecuación diferencial	Solución
a)	$y' = \frac{xy}{y^2-1}$	$y^2 - 2lny = x^2$
b)	$y'' + y = 0$	$y = C_1 cosx + C_2 senx$
c)	$y'' + 2y' = 2e^x$	$y = \frac{2}{3}(e^{-2x} + e^x)$

3.2. Separación de variables

Consideremos una ecuación diferencial que pueda escribirse de la forma

$$M(x) + N(y)y' = 0$$

con $y' = \frac{dy}{dx}$, donde M y N son funciones de x y de y, respectivamente, continuas. Este tipo de **ecuaciones** se denominan **separables**, ya que todos los términos que dependen únicamente de x se pueden agrupar con dx en una parte de la igualdad y todos los términos que dependen únicamente de y se pueden agrupar con dy en la otra parte de la igualdad. La solución de este tipo de ecuaciones se obtiene integrando a ambos lados de la igualdad.

Ejemplo 3.2.

Encontrar la solución general de $(x^2+4)\ y' = xy$***.***

Solución:

$(x^2+4)\frac{dy}{dx} = xy$
$(x^2+4)\ dy = xy\ dx\ \rightarrow\ \frac{dy}{y} = \frac{x}{x^2+4}dx$
Integramos a ambos lados de la igualdad:
$\int \frac{dy}{y} = \int \frac{x}{x^2+4}dx = \frac{1}{2}\int \frac{2x}{x^2+4}dx$
$ln|y| = \frac{1}{2}ln|x^2+4| + C_1$
$ln|y| = ln|x^2+4|^{\frac{1}{2}} + C_1$
$ln|y| = ln\sqrt{x^2+4} + C_1$
$e^{ln|y|} = e^{ln\sqrt{x^2+4}+C_1} = e^{ln\sqrt{x^2+4}}e^{C_1}$
$|y| = e^{C_1}\sqrt{x^2+4}$
$y = \pm e^{C_1}\sqrt{x^2+4}$
$y = K\sqrt{x^2+4}$

Ejemplo 3.3.

Dada la condición inicial y(0)=1, encontrar la solución particular de la ecuación

$$xy + e^{-x^2}(y^2-1)y' = 0$$

Solución:

$xy + e^{-x^2}(y^2-1)\frac{dy}{dx} = 0$
$dx(xy + e^{-x^2}(y^2-1)\frac{dy}{dx}) = dx0$
$xy\ dx + e^{-x^2}(y^2-1)\ dy = 0$
$e^{-x^2}(y^2-1)\ dy = -xy\ dx$
$\frac{y^2-1}{y}dy = \frac{-x}{e^{-x^2}}dx$
$\int(y - \frac{1}{y})\ dy = \int -xe^{x^2}\ dx$
$\int(y - \frac{1}{y})\ dy = \frac{-1}{2}\int 2xe^{x^2}\ dx$
$\frac{y^2}{2} - ln|y| = -\frac{1}{2}e^{x^2} + C$

De la condición inicial y(0)=1, se tiene que $\frac{1}{2} - 0 = -\frac{1}{2} + C$, lo cual implica que C=1. Así la solución particular tiene la forma:
$\frac{y^2}{2} - ln|y| = -\frac{1}{2}e^{x^2} + 1$
$y^2 - lny^2 + e^{x^2} = 2$

Autoevaluación

1. Calcula la solución general de las siguientes ecuaciones diferenciales:

   a) $y' = \frac{x}{y}$. Sol: $y = \sqrt{x^2 + C}$.

   b) $y' = \frac{x^2+2}{3y^2}$. Sol: $y^3 = \frac{x^3}{3} + 2x + C$.

   c) $xy' = y$. Sol: $y = Kx$.

   d) $yy' = 6cos(\pi x)$. Sol: $y^2 = \frac{12}{\pi} sen(\pi x) + C$.

3.3. Ecuaciones diferenciales lineales de primer orden

Una ecuación diferencial **lineal** de primer orden es una ecuación de la forma

$$y' + P(x)y = Q(x)$$

donde P y Q son funciones continuas de x. Esta ecuación diferencial lineal de primer orden se dice que está expresada en **forma normal**.

Para resolver una ecuación diferencial lineal, hay que escribirla en forma normal para identificar las funciones P(x) y Q(x). La solución general de la ecuación es

$$y = \frac{1}{u(x)}[\int Q(x)u(x)dx + C]$$

donde u(x)=$e^{\int P(x)dx}$ se denomina **factor integrante**.

Ejemplo 3.4.

Encontrar la solución de la ecuación $xy' - 2y = x^2$***:***

Solución:

La forma normal,$y' + P(x)y = Q(x)$, de la ecuación dada se obtiene dividiéndola entre x:

$$y' - (\frac{2}{x})y = x$$

Así, P(x)=$\frac{-2}{x}$, Q(x)=x y el factor integrante es u(x)=$e^{\int P(x)dx} = e^{\int \frac{-2}{x}dx} = e^{-2ln|x|} = \frac{1}{e^{lnx^2}} = \frac{1}{x^2}$

Con estas funciones, la solución general de la ecuación es:

$$y = \frac{1}{u(x)}[\int Q(x)u(x)dx + C] = \frac{1}{1/x^2}[\int x\frac{1}{x^2}dx + C] = x^2[ln|x| + C]$$

Autoevaluación

1. Calcula la solución general de las siguientes ecuaciones diferenciales:

   a) $y' - 3x^2y = e^{x^3}$. Sol: $y = e^{x^3}[x + C]$.

   b) $y' + \frac{2}{x}y = 3x + 2$. Sol: $y = \frac{3x^2}{4} + \frac{2x}{3} + \frac{C}{x^2}$.

   c) $y' + 2xy = 4x$. Sol: $y = 2 + \frac{C}{e^{x^2}}$.

   d) $(y - 1)senxdx - dy = 0$. Sol: $y = 1 + \frac{C}{e^{cosx}}$.

3.4. Seminario de ECUACIONES DIFERENCIALES

(Algunos de estos ejercicios han sido obtenidos del libro **CÁLCULO**, de Larson, Hostetler y Edwards (McGrawHill) *Octava edición*. En este libro se pueden encontrar multitud de ejercicios de cada una de las sesiones de este tema.)

(Verificación de soluciones de ecuaciones diferenciales)

1. En los siguientes ejercicios verifica que la función dada es la solución de la ecuación diferencial:

	Ecuación diferencial	Solución
a)	$y' = 4y$	$y = C \cdot e^{4x}$
b)	$3y' + 4y = e^{-x}$	$y = e^{-x}$
c)	$y' = \frac{2xy}{(x^2-y^2)}$	$x^2 + y^2 = C \cdot y$
d)	$y'' + 2y' + 2y = 0$	$y = C_1 \cdot e^{-x}\cos(x) + C_2 \cdot e^{-x}\sin(x)$

(Comprobar soluciones particulares de una ecuación diferencial)

2. En los siguientes ejercicios verifica que la función dada es solución de la ecuación diferencial y comprueba que se trata de una solución particular que cumple las condiciones dadas:

	Ecuación diferencial	Solución	Condición inicial
a)	$2y + y' = 2\sin(2x) - 1$	$y = \sin(x)\cos(x) - \cos^2(x)$	$y(\frac{\pi}{4}) = 0$
b)	$y' = x + 4\sin(x)$	$y = \frac{1}{2}x^2 - 4\cos(x) + 2$	$y(0) = -2$
c)	$y' = -4xy$	$y = 6e^{-2x^2}$	$y(0) = 6$
d)	$y' = y\sin(x)$	$y = e^{-cos(x)}$	$y(\frac{\pi}{2}) = 1$

(Calcular soluciones particulares de una ecuación diferencial dada la solución general)

3. En los siguientes ejercicios verificar que la solución general satisface la ecuación diferencial y encontrar la solución particular que satisface las condiciones dadas:

	Ecuación diferencial	Solución	Condición inicial
a)	$y' + 2y = 0$	$y = C \cdot e^{-2x}$	$y = 3$ cuando $x = 0$
b)	$3x + 2yy' = 0$	$3x^2 + 2y^2 = C$	$y = 3$ cuando $x = 1$
c)	$xy'' + y' = 0$	$y = C_1 + C_2\ln(x)$	$y = 0$ cuando $x = 2$; $y' = \frac{1}{2}$ cuando $x = 2$
d)	$y'' + 9y = 0$	$y = C_1\sin(3x) + C_2\cos(3x)$	$y = 2$ cuando $x = \frac{\pi}{6}$; $y' = 1$ cuando $x = \frac{\pi}{6}$

(Variables separables)

4. En los siguientes ejercicios resolver la ecuación diferencial:

   a) $y' = 4 - x$
   b) $y' = (0{,}05)y$
   c) $yy' = \sin(x)$
   d) $y' = \frac{\sqrt{x}}{3y}$
   e) $y' = x(1 + y)$
   f) $y' = \frac{2x}{y}$
   g) $y\ln(x) - xy' = 0$
   h) $\sqrt{1-4x^2}y' = x$

5. En los siguientes ejercicios resolver la ecuación diferencial y encontrar la solución particular que satisface la condición inicial:

   a) $yy' - e^x = 0$ (Condición inicial: $y(0) = 4$)
   b) $y(x + 1) + y' = 0$ (Condición inicial: $y(-2) = 1$)
   c) $y(1 + x^2)y' - x(1 + y^2) = 0$ (Condición inicial: $y(0) = \sqrt{3}$)

(Ecuaciones lineales de primer orden)

6. En los siguientes ejercicios resolver la ecuación diferencial:

   a) $\frac{dy}{dx} + (\frac{1}{x})y = 3x + 4$
   b) $y' - y = 10$
   c) $(y + 1)\cos(x)dx - dy = 0$
   d) $(x - 1)y' + y = x^2 - 1$
   e) $y' + 3y = e^{3x}$
   f) $y' - y = cosx$

7. En los siguientes ejercicios resolver la ecuación diferencial y encontrar la solución particular que satisface la condición inicial:

   a) $y'cos^2x + y - 1 = 0$ (Condición inicial: $y(0) = 5$)
   b) $x^3y' + 2y = e^{1/x^2}$ (Condición inicial: $y(1) = e$)
   c) $y' + \frac{1}{x}y = 0$ (Condición inicial: $y(2) = 2$)
   d) $y' + (2x - 1)y = 0$ (Condición inicial: $y(1) = 2$)

(Algunos ejercicios que han aparecido en exámenes)

8. (2010) Obtén la solución particular de la ecuación diferencial $2x^2 + xy' = x$, que verifica la condición inicial (1,2).

9. (2010) Obtén la solución particular de la ecuación diferencial $x^2y' - 2xy = x^3$, que verifica la condición inicial (1,2).

10. (2011) Obtén la solución particular de la ecuación diferencial $y' - \frac{1}{x}y = x$, que verifica la condición inicial (1,5).

11. (2011) Obtén la solución particular de la ecuación diferencial $xy' - 2x^2y = x^2$ que verifica la condición inicial (1,5).

12. (2012) Resuelve la ecuación diferencial $y' = x(2 + y)$.

13. (2012) Resuelve la siguiente ecuación diferencial $(x-2)y' + y = x^2 - 4$.

14. (2013) Obtén la solución particular de la ecuación diferencial $xy' + 2y = 3x^2 + 2x$ que verifica la condición inicial (1,0).

15. (2013) Obtén la solución particular de la ecuación diferencial $xy' - y = x^3 + x$ que verifica la condición inicial $(1, \frac{3}{2})$.

16. (2014) Obtén la solución particular de la ecuación diferencial $(x^2+2)y' - xy = 0$ que verifica la condición inicial ($\sqrt{2}$,6).

17. (2014) Obtén la solución particular de la ecuación diferencial $(x+2)y' - xy = 0$ que verifica la condición inicial (-1,e).

18. (2015) Obtén la solución particular de la ecuación particular de la ecuación diferencial $(x-1)y' - y = (x^2-1)$, que verifica la condición inicial (2,5).

19. (2016) Obtén la solución particular de la ecuación particular de la ecuación diferencial $(x-2)y' - y = (x^2-4)$, que verifica la condición inicial (3,6).

3.5. Soluciones del Seminario de Ecuaciones Diferenciales

3. a) C=3 b) C=21 c) $C_1 = -ln|2|$ C_2=1 d) $C_1 = 2, C_2 = \frac{-1}{3}$
4a) $y = 4x - \frac{x^2}{2} + C$
4b) $y = Ke^{0,05x}$
4c) $y = \sqrt{C - 2cos(x)}$
4d) $y = \sqrt{\frac{4}{9}x^{\frac{3}{2}} + C}$
4e) $y = Ke^{\frac{x^2}{2}} - 1$
4f) $y = \sqrt{2x^2 + C}$
4g) $ln|y| = \frac{1}{2}(ln(x))^2 + C$
4h) $y = \frac{-1}{4}\sqrt{1 - 4x^2} + C$
5a) $y = \sqrt{2e^x + 14}$
5b) $y = e^{-\frac{x^2}{2} - x}$
5c) $y = \sqrt{4(1 + x^2) - 1}$
6a) $y = x^2 + 2x + \frac{C}{x}$
6b) $y = -10 + Ce^x$
6c) $y = Ke^{sen(x)} - 1$
6d) $y = \frac{1}{x-1}[\frac{x^3}{3} - x + C]$
6e) $y = \frac{e^{3x}}{6} + Ce^{-3x}$
6f) $y = \frac{-cos(x) - sen(x)}{2} + Ce^x$
7a) $y = 1 + 4e^{-tg(x)}$
7b) $y = e^{\frac{1}{x^2}}[\frac{-1}{2x^2} + \frac{3}{2}]$
7c) $y = \frac{4}{x}$
7d) $y = \frac{2}{e^{x^2 - x}}$
8) $y = x - x^2 + 2$
9) $y = x^2[ln|x| + 2]$
10) $y = x^2 + 4x$
11) $y = \frac{-1 + 11e^{x^2 - 1}}{2}$
12) $y = Ke^{\frac{x^2}{2}} - 2s$
13) $y = \frac{1}{x-2}[\frac{x^3}{3} - 4x + C]$
14) $y = \frac{3}{4}x^2 + \frac{2}{3}x - \frac{17}{12}x^{-2}$
15) $y = \frac{x^2}{2} + xln|x| + x$
16) $y = 3\sqrt{x^2 + 2}$
17) $ln|y| = x - 2ln|x + 2|$
18) $y = (x - 1)[x + 2ln|x - 1| + 3]$
19) $y = (x - 2)[x + 4ln(x - 2) + 3]$

Capítulo 4

Análisis Descriptivo de Datos

4.1. ¿Por qué estudiar estadística?

Cada vez es más frecuente que, para hacer una valoración inteligente de la información de que disponemos, necesitemos asimilar e interpretar una cantidad considerable de datos. Afirmaciones como "el índice de precios de consumo bajó un 0,8 por ciento el mes pasado" o "el 98 por ciento de los pacientes de un estudio clínico no experimentó ningún efecto secundario significativo con un nuevo medicamento contra el cáncer de mama" necesitan procesar previamente un gran volumen de datos. Las técnicas estadísticas permiten resumir, describir y valorar la información recogida para ayudar posteriormente en la toma de decisiones.

Una **población** es el conjunto completo de todos los objetos que interesan a un investigador. El tamaño de la población, N, puede ser muy grande o incluso infinito. Una **muestra** es un subconjunto observado de valores poblacionales que tiene un **tamaño muestral** que viene dado por n.

Ejemplo 4.1.

Ejemplos de poblaciones son:

* Todas las familias que viven en una ciudad.
* Todas las reclamaciones que recibe en un año dado una compañía de seguros médicos.
* Todos los peces que habitan en un río.

El **muestreo aleatorio simple** es un método que se emplea para seleccionar una muestra de n objetos de una población en el que cada miembro de la población se elige estrictamente al azar, cada miembro de la población se elige con la misma probabilidad y todas las muestras posibles de un tamaño dado, n, tienen la misma probabilidad de ser seleccionadas. La muestra resultante se denomina muestra aleatoria.

Se necesita estudiar estadística no para hacer afirmaciones sobre la muestra sino, para extraer conclusiones sobre la población en general. La estadística estudia cómo se toman decisiones sobre la población cuando la información procede de una muestra.

La **estadística descriptiva** está formada por los métodos gráficos y numéricos que se utilizan para resumir y procesar los datos y transformarlos en información. La **estadística inferencial** constituye la base para hacer predicciones, previsiones y estimaciones que se utilizan para transformar la información en conocimiento.

4.2. Clasificación de variables

Una **variable** es una característica que puede ser medida y que puede adoptar valores diferentes para cada uno de los elementos que constituyen la población de estudio. Los **datos** son los valores de la variable para cada elemento de la muestra.

Ejemplo 4.2.

Ejemplos de variables son:

Grado de satisfacción sobre un servicio recibido: nada satisfecho, poco satisfecho, satisfecho, muy satisfecho.
Color de una pieza: rojo, verde, blanco, etc.
Número de piezas defectuosas: 0, 1, 2, 3, etc.
Peso al nacer de un niño: 3.3 Kg., 2.7 Kg., etc.

Las variables se pueden clasificar en función de su nivel de medición en **cualitativas** y **cuantitativas**. Los datos que se obtienen en una variable cualitativa no tienen ningún significado mensurable. Por ejemplo, si a un jugador de baloncesto se le asigna el número "20" y a otro el número "10", no se extrae la conclusión que uno sea el doble mejor que el otro. Sin embargo los datos que se obtienen de una variable cuantitativa tiene un significado mensurable. Por ejemplo, cuando un estudiante obtiene una puntuación de 8 en un examen y otro obtiene una puntuación de 4, la diferencia entre ellos es mensurable y se puede valorar cual de ellos es mejor.

A su vez las variables cualitativas se pueden clasificar en **nominales** y **ordinales**. Variables como el sexo, nacionalidad y estado civil son nominales porque los datos que se generan de ellas describen categorías. Si se asignan números para codificar estas respuestas, estos no determinan ningún orden. Los datos ordinales indican el orden que ocupan los objetos. Por ejemplo, la valoración de la calidad de un producto: (1: malo, 2: medio y 3: bueno) o la valoración de satisfacción con el servicio del comedor de la universidad: (1: muy insatisfecho, 2: moderadamente insatisfecho, 3: ninguna opinión, 4: moderadamente satisfecho, 5: muy satisfecho). Las respuestas de una variable ordinal siguen un orden pero la "diferencia" entre ellas no tienen ningún significado mensurable. Es decir, la diferencia entre la primera opción y la segunda puede no ser igual que la diferencia entre la segunda y la tercera.

Por último, las variables cuantitativas se pueden clasificar en **discretas** y **continuas**. Una variable discreta puede tener (pero no necesariamente) un número finito de valores. Sin embargo, el tipo más frecuente es el que produce una respuesta que proviene de un proceso de recuento. Por ejemplo, el número de alumnos matriculados en una clase o el número de hermanos que tiene cada individuo. Una variable continua puede tomar cualquier valor de un intervalo dado de números reales y normalmente proviene de un proceso de medición (no de recuento). Por ejemplo, la altura, el peso, el tiempo, la distancia o la temperatura.

Esquemáticamente la clasificación de variables sería la siguiente:

$$variables \begin{cases} cualitativas \begin{cases} nominales & sexo: 1(masculino),\ 2(femenino) \\ ordinales & calidad: 1(malo),\ 2(medio),\ 3(bueno) \end{cases} \\ cuantitativas \begin{cases} discretas & n\ de\ hijos:\ 0,\ 1,\ 2,\ ... \\ continuas & peso:\ 51'3,\ 60'2,\ 72,\ ... \end{cases} \end{cases}$$

Autoevaluación

1. Clasifica con los dos adjetivos adecuados las siguientes variables:

    a) Talla de una camiseta (S, M, L, XL, XXL). Sol: Cualitativa ordinal.

    b) Número de calzado. Sol: Cuantitativa discreta.

    c) Temperatura corporal de un paciente. Sol: Cuantitativa continua.

    d) Día de la semana. Sol: Cualitativa ordinal.

    e) Día del mes. Sol: Cuantitativa discreta.

    f) Mes del año. Sol: Cualitativa ordinal.

4.3. Tablas de frecuencias

Las **tablas de frecuencias** son una herramienta útil para organizar y resumir los datos.

Si la variable de estudio es una **variable cualitativa** o **cuantitativa discreta** y los datos se recogen a partir de una muestra de tamaño n, la tabla tiene una primera columna (llamada **clases** o grupos) que contiene todas las posibles respuestas de dicha variable. A continuación se le añade la columna de **frecuencias absolutas** (f_i) que contiene el número de observaciones correspondientes a cada clase. La siguiente columna recoge las **frecuencias absolutas acumuladas** (F_i) y se calcula como suma de las frecuencias absolutas de las clases anteriores. La siguiente columna contiene las **frecuencias relativas** (h_i) que se calculan como cociente entre las frecuencias absolutas y el tamaño de muestra. A continuación se añaden las **frecuencias relativas acumuladas** (H_i) calculando la suma de las frecuencias relativas de las clases anteriores. Por último, la tabla concluye con las frecuencias relativas y relativas acumuladas en tantos por cien ($h_i\,\%$ y $H_i\,\%$).

clases	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
		n		1		100
Totales	n		1		100	

Ejemplo 4.3.

En un colegio se decide hacer un estudio sobre la salud bucal de los alumnos, para ello se toma una muestra representativa de 50 alumnos y se anota el número de caries que padecen cada uno de ellos.

alumno	***1***	***2***	***3***	***4***	***........***	***48***	***49***	***50***
nº de caries	***3***	***0***	***1***	***0***	***........***	***0***	***2***	***1***

Resume los datos utilizando una tabla de frecuencias.

Solución:

nº de caries	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
0	21	21	0.42	0.42	42	42
1	8	29	0.16	0.58	16	58
2	10	39	0.20	0.78	20	78
3	4	43	0.08	0.86	8	86
4	6	49	0.12	0.98	12	98
5	1	50	0.02	1	2	100
Totales	50		1		100	

Si la variable de estudio es **cuantitativa continua** y los datos se recogen a partir de una muestra de tamaño n, para construir la tabla de frecuencias es necesario decidir:

- En cuántos intervalos distribuiremos los datos (clases o grupos).
- De qué amplitud deben ser los intervalos.

En definitiva, la tabla de frecuencias tendría una estructura similar a la de una variable cualitativa sólo que con una columna más, que recogería el punto medio de los intervalos. Esta columna recibe el nombre de **marca de clase** (x_i).

clases	x_i	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
			n		1		100
Totales		n		1		100	

Ejemplo 4.4.

Se desea estudiar el índice de obesidad de los jóvenes de entre 18 y 25 años. Para ello se extrae una muestra representativa de la población de tamaño 500 y se anota el peso de cada individuo.

individuo	***1***	***2***	***3***	***4***	***........***	***498***	***499***	***500***
peso (Kg.)	***50.1***	***62.3***	***58.1***	***49.5***	***........***	***43.7***	***72.4***	***103.3***

Construye una tabla de frecuencias con los datos anteriores.

Solución:

La variable de estudio es una variable cuantitativa continua. En este caso el tamaño muestral es n=500 y vamos a distribuir los datos, por ejemplo, en 10 intervalos. Imaginemos que la persona con menor peso ha sido un individuo con 42.7 Kg. y la persona con mayor peso pesa 112.5 Kg. Redondeando, consideramos que el rango de peso oscila entre 40 y 120 Kg. Así, la amplitud de los intervalos sería:

$$w = \frac{120 - 40}{10} = 8$$

La tabla de frecuencias quedaría del siguiente modo

peso (Kg.)	x_i	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
[40, 48[	44	26	26	0,052	0,052	5,2	5,2
[48, 56[	52	200	226	0,400	0,452	40	45,2
[56, 64[	60	150	376	0,300	0,752	30	75,2
[64, 72[	68	50	426	0,100	0,852	10	85,2
[72, 80[	76	20	446	0,040	0,892	4	89,2
[80, 88[	84	21	467	0,042	0,934	4,2	93,4
[88, 96[	92	20	487	0,040	0,974	4	97,4
[96, 104[	100	10	497	0,020	0,994	2	99,4
[104, 112[	108	0	497	0	0,994	0	99,4
[112, 120]	116	3	500	0,006	1	0,6	100
Totales		500		1		100	

4.4. Medidas de localización

Las medidas de localización resumen entre qué cantidades cuantitativas oscilan los valores, numéricos, del banco de datos.

Mínimo: Es el valor más pequeño del banco de datos.

Percentil 25 o **primer cuartil** (P_{25} *o* Q_1) : Ordenando el banco de datos de forma ascendente, es el valor que deja por debajo de él el 25 % de los datos y ocupa la posición pos(P_{25})=0.25(n+1).

Percentil 50 o **segundo cuartil** (P_{50} *o* Q_2) : Ordenando el banco de datos de forma ascendente, es el valor que deja por debajo de él el 50 % de los datos y ocupa la posición pos(P_{50})=0.50(n+1).

Percentil 75 o **tercer cuartil** (P_{75} *o* Q_3) : Ordenando el banco de datos de forma ascendente, es el valor que deja por debajo de él el 75 % de los datos y ocupa la posición pos(P_{75})=0.75(n+1).

Percentil x con $1 < x < 100$ (P_x) : Ordenando el banco de datos de forma ascendente, $Z_1, Z_2, ..., Z_n$, es el valor que deja por debajo de él el x % de los datos y ocupa la posición y=pos(P_x)=$\frac{x}{100}$(n+1).

$P_x = (parte\ decimal\ de\ y) * Z_{(parte\ entera\ de\ y)+1} + (1 - parte\ decimal\ de\ y) * Z_{parte\ entera\ de\ y}$

Máximo: Es el valor más grande del banco de datos.

El **resumen de los cinco números** se refiere a las cinco medidas descriptivas: mínimo, primer cuartil, mediana, tercer cuartil y máximo:

$$\text{Mínimo} < Q_1 < \text{Mediana} < Q_3 < \text{Máximo}$$

Ejemplo 4.5.

Calcula los percentiles 33 y 40 del siguiente conjunto de datos:

12.1, 15.7, 13.5, 14.2, 11.6

.

Solución:
En primer lugar se ordenan los datos de menor a mayor:

11.6, 12.1, 13.5, 14.2, 15.7

$Pos(P_{33}) = 0{,}33(5 + 1) = 1{,}98 \rightarrow P_{33} = 0{,}98 * 12{,}1 + 0{,}02 * 11{,}6 = 12{,}09$
$Pos(P_{40}) = 0{,}4(5 + 1) = 2{,}4 \rightarrow P_{40} = 0{,}6 * 12{,}1 + 0{,}4 * 13{,}5 = 12{,}66$

Autoevaluación

1. Para determinar los posibles efectos adversos de una severa dieta, un nutricionista estudia el nivel de hemoglobina en sangre de 10 pacientes y obtiene los siguientes resultados (expresados en g/dl):

$$10'9, 14'7, 10'0, 11'3, 15'1, 18'2, 12'3, 11'5, 14'7, 14'8$$

Se conoce que los niveles normales oscilan ente [13’8,17’2] en hombres y [12’1,15’1] en mujeres.

Calcula Mínimo, P_{25}, P_{50}, P_{75} y el Máximo.

Sol: Mínimo= 10’0, P_{25}=11.2, P_{50}=13.5, P_{75}=14.88 y el Máximo=18’2.

4.5. Medidas de centralización

Las medidas de centralización suministran información numérica sobre una observación "típica" de los datos. En este apartado analizamos la media, la moda y la simetría de los datos. Consideremos que los datos proceden de una muestra de tamaño n.

La **media muestral**, $\overline{x}$, es un estadístico que viene definido como:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}, \tag{4.1}$$

La **mediana** es la observación que ocupa el lugar central de un conjunto de observaciones ordenadas en sentido ascendente. Si el tamaño muestral, n, es un número impar, la mediana es la observación que se encuentra en el centro. Si el tamaño de la muestra, n es un número par, la mediana es la media de las dos observaciones que se encuentran en el centro. La mediana coincide con el Percentil 50.

La **moda**, si existe, es el valor que aparece con más frecuencia.

Los datos cualitativos se describen, generalmente, mejor mediante la mediana y la moda.

Ejemplo 4.6.

La dirección de un hospital quiere valorar las reclamaciones que recibe por la atención a sus pacientes. Los registros de un periodo de 5 semanas muestran el siguiente número de reclamaciones:

13, 15, 8, 16 y 8

a) Calcula el número medio de reclamaciones semanales.
b) Calcula el número mediano de reclamaciones semanales.
c) Calcula la moda.

Solución:

a) $\overline{x} = \frac{13+15+8+16+8}{5} = 12$.
b) Para calcular la mediana, se ordenan los datos de menor a mayor (8, 8, 13, 15 y 16). Como hay 5 datos, la mediana será el dato que está en la posción 3ª. Me = 13.
c) Mo = 8.

Autoevaluación

1. Para determinar los posibles efectos adversos de una severa dieta, un nutricionista estudia el nivel de hemoglobina en sangre de 10 pacientes y obtiene los siguientes resultados (expresados en g/dl):

$$10'9, 14'7, 10'0, 11'3, 15'1, 18'2, 12'3, 11'5, 14'7, 14'8$$

Se conoce que los niveles normales oscilan ente [13'8,17'2] en hombres y [12'1,15'1] en mujeres.

Calcula el nivel medio de hemoglobina en sangre, el nivel mediano de hemoglobina en sangre y la moda de estos pacientes,.

Sol: $\overline{x} = 13{,}35$, Mediana=P_{50}=13.5, Mo=14.7.

4.6. Medidas de dispersión

La media no es por sí sola una descripción completa o suficente de los datos. En este apartado veremos descriptivos que miden la variabilidad o dispersión de las observaciones con respecto a la media, como por ejemplo el rango, el rango intercuartílico, la varianza, la desviación típica y el coeficiente de variación.

El **rango** es la diferencia entre la observación mayor y la menor.

El **rango intercuartílico** (RIC) mide la dispersión que hay en el 50 por ciento central de los datos; es la diferencia entre la observación Q_3, **tercer cuartil** o **percentil 75** y la observación Q_1, **primer cuartil** o **percentil 25**.

$$RIC = Q_3 - Q_1$$

donde Q_3 y Q_1 se encuentran situados en la posición 0'75(n+1) y 0'25(n+1), respectivamente, cuando los datos se encuentra ordenados en sentido ascendente.

Aunque el rango y el rango intercuartílico miden la dispersión de los datos, ambas medidas sólo tiene en cuenta dos de los valores de los datos. Sería deseable disponer de una medida que promedie la distancia de cada uno de los valores y la media.

La **varianza muestral**, s^2, es la suma de los cuadrados de las diferencias entre cada observación y la media muestral dividida por el tamaño de la muestra, n, menos 1:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

La **desviación típica muestral**, s, es la raíz cuadrada (positiva) de la varianza muestral:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

La desviación típica se mide en las mismas unidades que la variable de donde proceden los datos y describe la dispersión media entorno a la media.

Los **valores atípicos** son aquellos que quedan fuera del intervalo $[Q_1 - 1{,}5 * RIC;\ Q_3 + 1{,}5 * RIC]$

Ejemplo 4.7.

Se ha analizado una muestra de 20 lotes de un producto químico para hallar la concentración de impurezas. Los resultados obtenidos son:

1, 1, 3, 3, 3, 5, 5, 5, 5, 5, 5, 7, 7, 7, 7, 7, 9, 9, 9, 9

Calcula todas las medidas de centralización y dispersión que conozcas.

Solución:

EN CASO DE NO DISPONER DE CALCULADORA CON FUNCIONES DE ESTADÍSTICA, para llevar a cabo los cálculos, construiremos una tabla que nos ayude con los cálculos de las medidas de centralización y dispersión:

x_i	f_i	$x_i f_i$	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$f_i(x_i - \overline{x})^2$
1	2	2	-4'6	21'16	42'32
3	3	9	-2'6	6'76	20'28
5	6	30	-0'6	0'36	2'16
7	5	35	1'4	1'96	9'8
9	4	36	3'4	11'56	46'24
Suma	20	112			120'8

Medidas de centralización:

$n = \sum_{i=1}^{k} f_i = 20$

$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{112}{20} = 5'6$

Para calcular la mediana se calcula la posición del dato central, pos(Q_2)=0.5(n+1)=10'5. Como n es par se toman los datos que están en la posición 10 y 11 y se calcula su media. $Me \approx 5$

Mo=5

Medidas de localización:

Para calcular el primer cuartil o Percentil 25, se calcula la posición, pos(P_{25})=0.25(n+1)=5'25. Los datos involucrados en el calculo se encuentran en la quita y sexta posición:
$P_{25} = (0{,}25 * 5) + (0{,}75 * 3) = 3{,}5$.

Para calcular el tercer cuartil o Percentil 75, se calcula la posición, pos(P_{75})=0.75(n+1)=15'75. Los datos involucrados se encuentran en la posición 15 y 16:
$P_{75} = (0{,}75 * 7) + (0{,}25 * 7) = 7$.

Medidas de dispersión:

rango= 9 - 1= 8.

rango intercuartílico RIC= 7 - 3.5= 3.5.

$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \overline{x})^2}{n-1} = \frac{120'8}{19} = 6'3579$

$s = \sqrt{s^2} = \sqrt{6'3579} = 2'52$

Autoevaluación

1. Para determinar los posibles efectos adversos de una severa dieta, un nutricionista estudia el nivel de hemoglobina en sangre de 10 pacientes y obtiene los siguientes resultados (expresados en g/dl):

$$10'9, 14'7, 10'0, 11'3, 15'1, 18'2, 12'3, 11'5, 14'7, 14'8$$

Se conoce que los niveles normales oscilan ente [13'8,17'2] en hombres y [12'1,15'1] en mujeres.

Calcula el rango, el RIC, la varianza muestral y la desviación típica muestral. Determina si existen valores atípicos.

Sol: rango=8.2, RIC=3.68, $s^2 = 2{,}54^2$=6.45, s=2.54. No existen valores atípicos.

4.7. Gráficos para describir variables cualitativas y cuantitativas discretas

Las variables cualitativas y cuantitativas discretas pueden describirse utilizando gráficos de barras y gráficos de tarta.

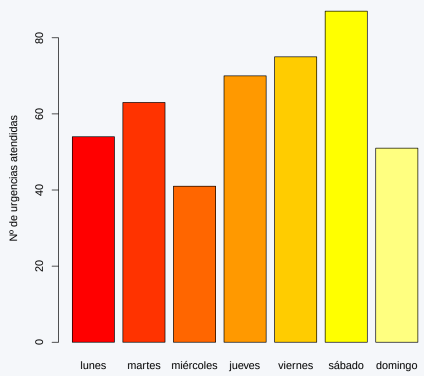
Si el objetivo es llamar la atención sobre la frecuencia de cada categoría lo más adecuado es utilizar un **gráfico de barras**, en el que la altura de un rectángulo representa esta frecuencia. Lo habitual es que cada rectángulo esté separado por un espacio en blanco para distinguir cada categoría.

Ejemplo 4.8.

En el siguiente gráfico de barras se resume el número de pacientes atendidos en urgencias, en cierto hospital de Valencia, durante una semana de febrero:

Día de la semana	*Nº de urgencias* (f_i)	(h_i %)
Lunes	***54***	***12.24***
Martes	***63***	***14.29***
Miércoles	***41***	***9.30***
Jueves	***70***	***15.87***
Viernes	***75***	***17.01***
Sábado	***87***	***19.73***
Domingo	***51***	***11.56***

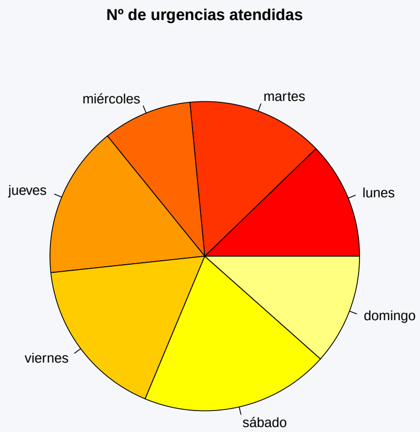
.



Si el objetivo es hacer hincapié en la proporción de las frecuencias de cada categoría, el gráfico adecuado es el **diagrama de sectores**. El círculo (o "tarta") representa la proporción total y los sectores (o "trozos de la tarta") que lo parten tienen un área proporcional a la frecuencia de cada categoría.

Ejemplo 4.9.

El siguiente gráfico de tarta representa la información del ejemplo anterior:



Para calcular los grados que les corresponde a cada porción del gráfico de tarta hay que plantear reglas tres:

100 % ——— 360º
12.24 % —— lunes $\rightarrow lunes = 44{,}064$

100 % ——— 360º
14.29 % —— martes $\rightarrow martes = 51{,}44$

.
.
.

etc.

4.8. Gráficos para describir variables cuantitativas continuas

En este apartado presentamos brevemente diagramas de cajas y histogramas que resumen y describen datos numéricos.

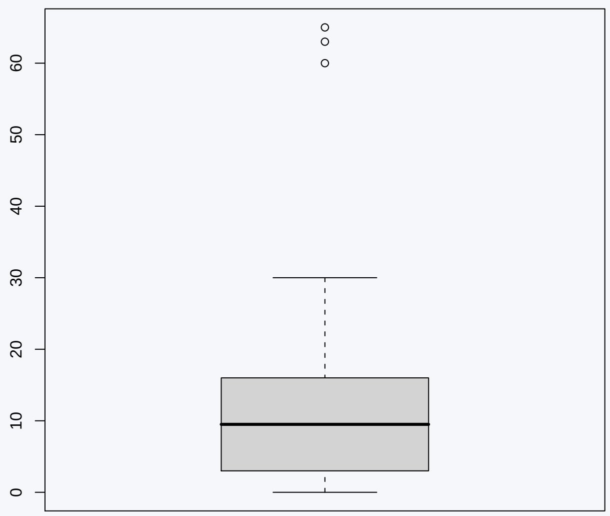
El **diagrama de cajas** de una variable consiste en la representación gráfica del resumen de los cinco números (Mínimo, P_{25}, P_{50}, P_{75} y Máximo). Se compone de un rectángulo o caja determinado por el P_{25} y P_{75}, que representa el rango intercuartílico y dos segmentos o bigotes en cada extremo de la caja. El bigote inferior tiene una longitud igual al Máx{mínimo, $P_{25} - 1{,}5RIC$} y el bigote superior tiene una longitud igual al Mín{máximo, $P_{75} + 1{,}5RIC$}. Si hay datos más allá de estos segmentos se señalan en el gráfico con puntos y se llaman valores atípicos.

Ejemplo 4.10.

Realiza un diagrama de caja para el siguiente conjunto de datos:

$$0, 1, 4, 5, 6, 7, 7, 9, 11, 12, 12, 15, 58, 65, 68$$

Solución:
$pos(P_{25}) = 0{,}25 * 16 = 4$, $pos(P_{50}) = 0{,}5 * 16 = 8$, $pos(P_{75}) = 0{,}75 * 16 = 12$
Mínimo=0, $P_{25} = 5$, $P_{50} = 9$, $P_{75} = 15$, Máximo= 68, $P_{25} - 1{,}5RIC = -10$ y $P_{75} + 1{,}5RIC = 30$



Autoevaluación

1. Para determinar los posibles efectos adversos de una severa dieta, un nutricionista estudia el nivel de hemoglobina en sangre de 10 pacientes y obtiene los siguientes resultados (expresados en g/dl):

$$10'9, 14'7, 10'0, 11'3, 15'1, 18'2, 12'3, 11'5, 14'7, 14'8$$

Se conoce que los niveles normales oscilan ente [13'8,17'2] en hombres y [12'1,15'1] en mujeres.

Dibuja un diagrama de caja de los datos anteriores:.

Sol:

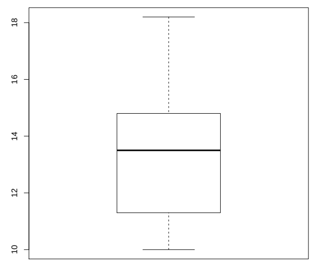


Figura 4.1: *Autoevaluación.*

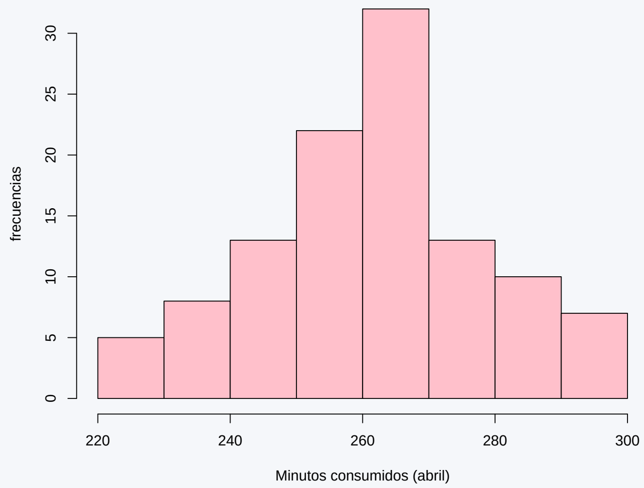
Un **histograma** es un gráfico formado por barras verticales construidas sobre una linea recta horizontal delimitada por los intervalos de la variable estudiada. Los intervalos corresponden a los construidos en una tabla de frecuencias. La altura de cada barra es proporcional al número de observaciones que hay en ese intervalo. El número de observaciones puede indicarse encima de las barras.

Ejemplo 4.11.

Se considera la siguiente tabla de frecuencias sobre el uso del teléfono móvil (en minutos) de una muestra de 110 personas.

Uso del tl. móvil (en minutos)	f_i	H_i %
$[220, 230[$	*5*	*4,5*
$[230, 240[$	*8*	*11,8*
$[240, 250[$	*13*	*23,6*
$[250, 260[$	*22*	*43,6*
$[260, 270[$	*32*	*72,7*
$[270, 280[$	*13*	*84,5*
$[280, 290[$	*10*	*93,6*
$[290, 300]$	*7*	*100*

.



4.9. Seminario de ANÁLISIS DESCRIPTIVO DE DATOS

1. En una farmacia se está recogiendo información sobre el grado de satisfacción de los clientes respecto a su servicio nocturno, concretamente se está preguntando cuál es la opinión de los clientes en cuanto la relación calidad-precio de este servicio nocturno. Las respuestas dadas por los clientes encuestados han sido codificadas según los códigos:

   - 0: Muy desfavorable
   - 1: Desfavorable
   - 2: Favorable
   - 3: Muy favorable

   Se ha preguntado a un total de 50 clientes, y sus respuestas codificadas numéricamente han sido las siguientes:

   0 1 3 0 1 1 2 3 0 0 3 3 3 2 1 2 0 3 0 2 1 0 0 2 3
   2 2 2 1 1 2 2 0 3 0 2 2 0 3 3 0 3 0 1 2 2 2 0 2 1

   a) Determina la población de estudio, la muestra, la variable de interés y el tipo de la misma.

   b) Resume los datos en una tabla de frecuencias.

   c) Realiza (a mano alzada) un gráfico de barras.

   d) Realiza (a mano alzada) un gráfico de tarta.

2. En una farmacia se realiza seguimiento de la Hipertensión Arterial de algunos pacientes. Se dispone de 20 mediciones de la tensión arterial sistólica (TAS) realizadas en el día de hoy, las cuales se muestran a continuación:

   173,03 165,54 141,59 158,66 158,81 156,49 150,29 154,53 162,50 158,49
   151,11 166,13 147,47 152,83 166,99 135,62 138,77 168,11 162,04 176,77

   a) Determina la población de estudio, la muestra, la variable de interés y el tipo de la misma.

   b) Resume los datos de esta variable en una tabla de frecuencias

   c) Calcula los siguientes estadísticos:

   - Mínimo
   - Máximo
   - P_{10}
   - $P_{25}(= Q_1)$
   - $P_{50}(= Q_2)$
   - $P_{75}(= Q_3)$
   - P_{90}
   - Media
   - Mediana
   - Moda
   - Rango
   - Rango intercuartílico
   - Varianza (s^2)
   - Desviación típica (s)

   d) Determina si existen valores atípicos.

   e) Realiza (a mano alzada) un diagrama de cajas.

   f) Realiza (a mano alzada) un histograma.

3. Se dispone del peso (en gramos) de 16 niños de un mes de edad. Los datos se muestran a continuación:

$$\begin{array}{cccccccc} 4123 & 4336 & 4160 & 4165 & 4422 & 3853 & 3281 & 3990 \\ 4096 & 4166 & 3596 & 4127 & 4017 & 3769 & 4240 & 4194 \end{array}$$

   a) Determina la población de estudio, la muestra, la variable de interés y el tipo de la misma.

   b) Calcula los siguientes estadísticos:

   - Mínimo
   - Máximo
   - P_{10}
   - $P_{25}(= Q_1)$
   - $P_{50}(= Q_2)$
   - $P_{75}(= Q_3)$
   - P_{90}
   - Media
   - Mediana
   - Moda
   - Rango
   - Rango intercuartílico
   - Varianza (s^2)
   - Desviación típica (s)

   c) Determina si existen valores atípicos.

   d) Realiza (a mano alzada) un diagrama de cajas.

   e) Realiza (a mano alzada) un histograma.

4. En una encuesta a personas con hipertensión arterial, se les ha preguntado el número de veces que han recibido control de su presión arterial en los últimos 6 meses. Las respuestas se muestran a continuación:

3 5 2 0 2 1 6 2 0 6 2 0 4 3 3 5 2 0 0 1
5 3 6 6 4 6 0 3 1 1 0 5 6 4 4 6 2 3 3 6

   a) Determina la población de estudio, la muestra, la variable de interés y el tipo de la misma.

   b) Resume los datos de esta variable en una tabla de frecuencias.

5. Se han tomado muestras a 40 niños de entre 1 y 5 años del nivel de cobre en orina, obteniéndose los siguientes valores:

0.10 0.30 0.34 0.36 0.42 0.42 0.45 0.48 0.50 0.52
0.55 0.58 0.62 0.63 0.64 0.65 0.65 0.66 0.69 0.70
0.72 0.73 0.74 0.74 0.75 0.76 0.77 0.78 0.81 0.83
0.85 0.86 0.88 0.90 0.94 0.98 1.04 1.12 1.16 1.24

   a) Determina la población de estudio, la muestra, la variable de interés y el tipo de la misma.

   b) Calcula la mediana, el máximo, el mínimo y el rango de estos datos.

   c) Calcula la media y la desviación típica.

   d) Calcula el primer y tercer cuartil, rango intercuartílico, percentil 10 y percentil 95.

   e) Determina si existen valores atípicos.

   f) Realiza (a mano alzada) un histograma y diagrama de cajas.

(Algunos ejercicios que han aparecido en exámenes)

6. (2012) Los hemogramas de 7 hombres adultos, sanos, en ayuna, proporcionaron los siguientes valores de hemoglobina glucosilada

$$5'7, 6'8, 7'4, 6'9, 6'7, 7'8, 5'9$$

Indica cuál es la población y la variable de estudio. Calcula el P_{33} e interpreta el resultado.

7. (2013) La analítica de 8 niños menores de un año diagnosticados de diabetes indica los siguientes valores de glucosa en sangre en ayunas:

$$85, 75, 68, 98, 72, 58, 65, 90$$

a) Calcula la media y la desviación típica.

b) Realiza los cálculos necesarios para indicar razonadamente si existen o no entre estos datos valores atípicos.

4.10. Soluciones del Seminario de Análisis Descriptivo de Datos

1a) Población: clientes de la farmacia. Muestra: los 50 clientes que participan en el estudio. Variable de interés: el grado de satisfacción respecto al servicio nocturno. Tipo de variable: cualitativa ordinal.

1b)

	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
0	14	14	0.28	0.28	28	28
1	9	23	0.18	0.46	18	46
2	16	39	0.32	0.78	32	78
3	11	50	0.22	1.00	22	100

1c y 1d)

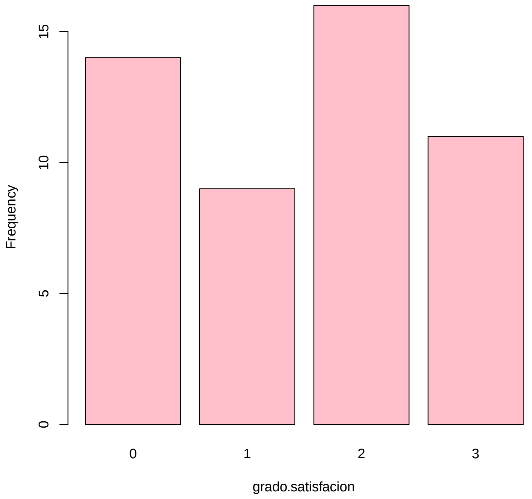


Figura 4.2: *Gráfico de barras.*

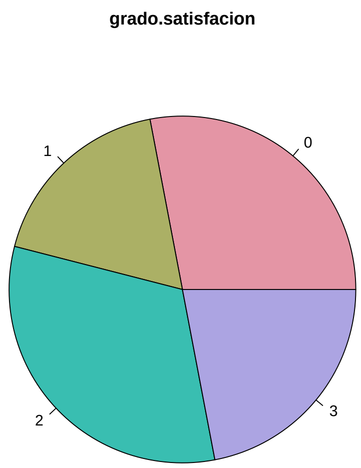


Figura 4.3: *Gráfico de sectores.*

2a)Población: pacientes hipertensos de la farmacia. Muestra: los 20 pacientes hipertensos. Variable de interés: tensión arterial sistólica (TAS). Tipo de variable: cuantitativa continua.

2b)

clases	x_i	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
$[130,140[$	135	2	2	0.1	0.1	10	10
$[140,150[$	145	2	4	0.1	0.2	10	20
$[150,160[$	155	8	12	0.4	0.6	40	60
$[160,170[$	165	6	18	0.3	0.9	30	90
$[170,180[$	175	2	20	0.1	1.0	10	100

2c) Mínimo= 135.62, Máximo= 176.77, P10=139.052, P25=150.4950, P50=158.5750, P75=165.9825, P90=172.5380, Media= 157.2885, Mediana=158.5750, Moda=No existe, Rango=41.15, RIC=15.4875 , Varianza=121.6192 , Desviación típica= 11.02812

2c) [127.26, 189.21]

2e y 2f)

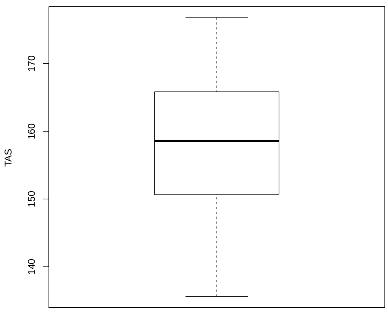


Figura 4.4: *Diagrama de cajas.*

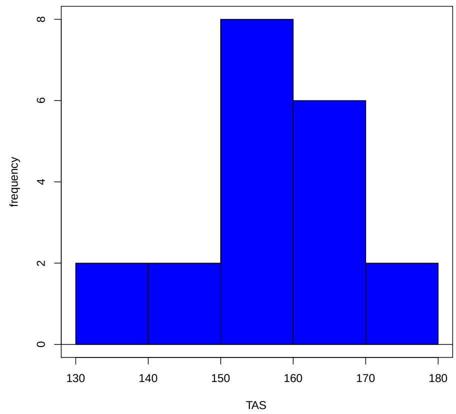


Figura 4.5: *Histograma.*

3a)Población: niños de un mes de edad. Muestra: 16 niños de un mes de edad. Variable de interés: Peso. Tipo de variable: cuantitativa continua.

3b) Mínimo= 3281, Máximo= 4422, P10=3501.50, P25=3887.25, P50=4125.00, P75=4187.00, P90=4361.80, Media= 4033.438, Mediana=4125.00, Moda=No existe, Rango=1141, RIC=299.75 , Varianza=82981.21 , Desviación típica= 288.0646

3c) [3437.625, 4636.625]

3d y 3f)

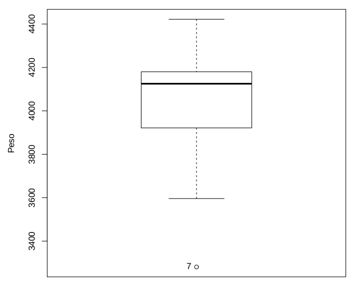


Figura 4.6: *Diagrama de cajas.*

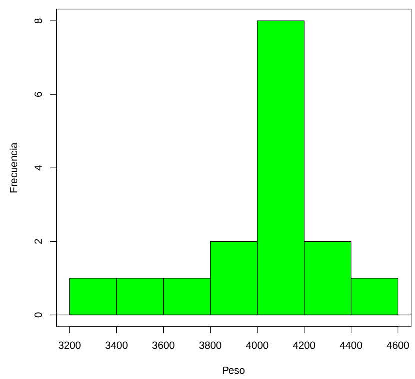


Figura 4.7: *Histograma.*

4a)Población: hipertensos. Muestra: 40 hipertensos. Variable de interés: número de veces que han recibido control. Tipo de variable: cuantitativa discreta.

4b)

	f_i	F_i	h_i	H_i	$h_i\,\%$	$H_i\,\%$
0	7	7	0.175	0.175	17.5	17.5
1	4	11	0.100	0.275	10.0	27.5
2	6	17	0.150	0.425	15.0	42.5
3	7	24	0.175	0.600	17.5	60.0
4	4	28	0.100	0.700	10.0	70.0
5	4	32	0.100	0.800	10.0	80.0
6	8	40	0.200	1.000	20.0	100.0

5a)Población: niños entre 1 y 5 años. Muestra: 40 niños entre 1 y 5 años. Variable de interés: nivel de cobre en la orina. Tipo de variable: cuantitativa continua.

5b) Mediana=0.7100, Máximo=1.2400, Mínimo=0.1000, Rango=1.14.

5c) Media=0.6965, Desviación típica=0.2410399.

5d) P25=0.5275, P75=0.8450, RIC=0.3175, P10=0.3660, P95=1.1580.

5e) [0.0513, 1.3213]

5f)

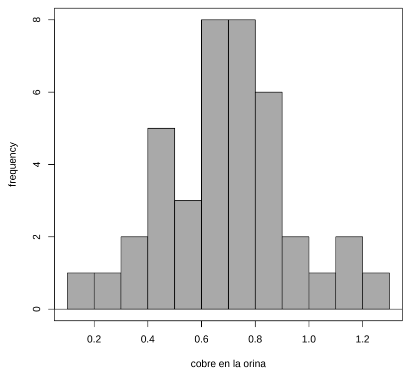


Figura 4.8: *Histograma.*

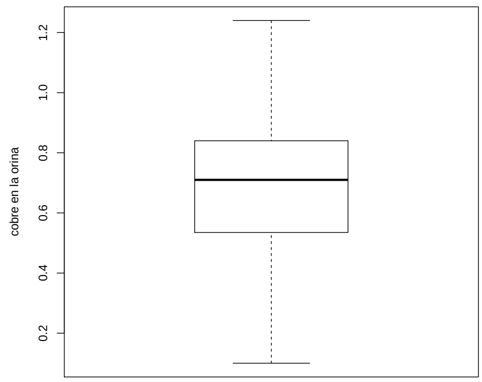


Figura 4.9: *Diagrama de cajas.*

6) Población: hombres adultos. Variable de interés: hemoglobina glucosilada. P33=6.412.

7a) $\overline{x}$= 76.375 , s= 13.55347

7b) Todos los datos se encuentran dentro del intervalo [31.25,123.25]. No existen valores atípicos.

Capítulo 5

Probabilidad. Estimación de una población

5.1. Función de densidad de probabilidad y función de distribución acumulada

Supongamos que tenemos un conjunto de mediciones de una variable aleatoria continua y que representamos sus frecuencias relativas mediante un histograma para describir su distribución. Cuanto mayor sea el número de mediciones, menos irregular se verá el contorno del histograma. Cuando el número de mediciones es muy grande y la amplitud de clases o intervalos se reduce, el histograma se parece cada vez más a una curva como la que aparece en la última viñeta de la Figura. Esta curva describe la distribución de probabilidad de la variable aleatoria continua y la llamaremos **función de densidad de probabilidad** f(x).

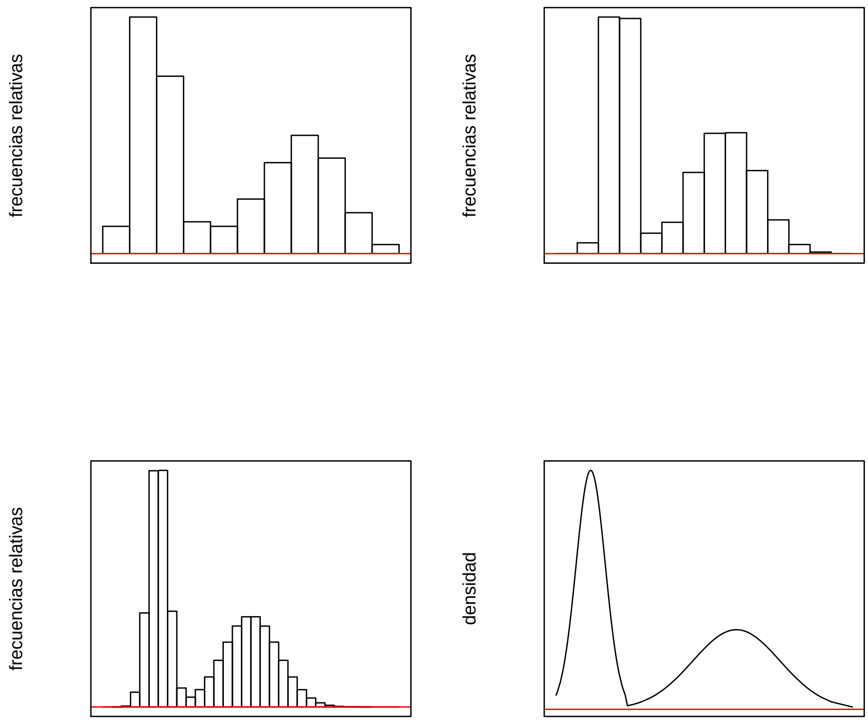


Figura 5.1: *Histogramas de frecuencias relativas para tamaños de muestra cada vez más grandes.*

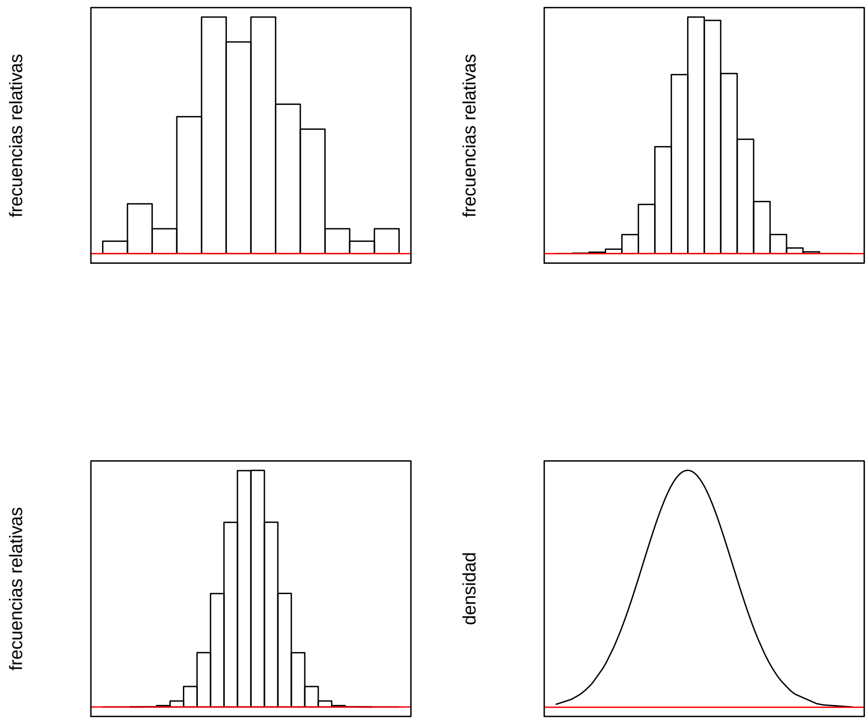


Figura 5.2: *Histogramas de frecuencias relativas para tamaños de muestra cada vez más grandes.*

5.1.1. Propiedades de la función de densidad.

Las propiedades que cumple la función de densidad de una variable continua son las siguientes:

1. El área bajo la función de densidad es 1, es decir $\int_{-\infty}^{\infty} f(x)\,dx{=}1$.

2. La probabilidad de que la variable continua X tome valores en un intervalo [a,b], con $-\infty \leq a \leq b \leq \infty$, es igual al área bajo la función de densidad entre los puntos a y b, es decir

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx \qquad \text{y} \qquad 0 \leq P(a \leq X \leq b) \leq 1.$$

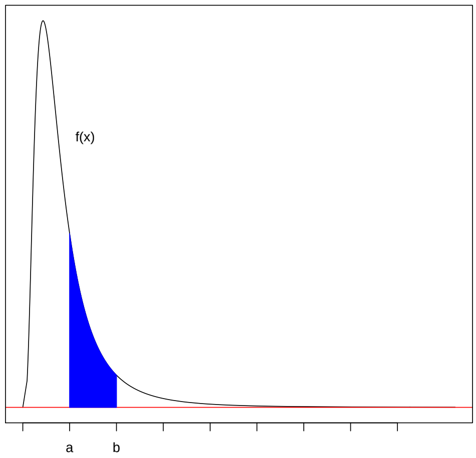


Figura 5.3: *$P(a \leq X \leq b)$ es el área sombreada por debajo de la función de densidad f(x).*

La **función de distribución acumulada**, F(x), de una variable aleatoria continua X expresa la probabilidad de que X tome valores menores o iguales que x, es decir $F(x) = P(X \leq x)$

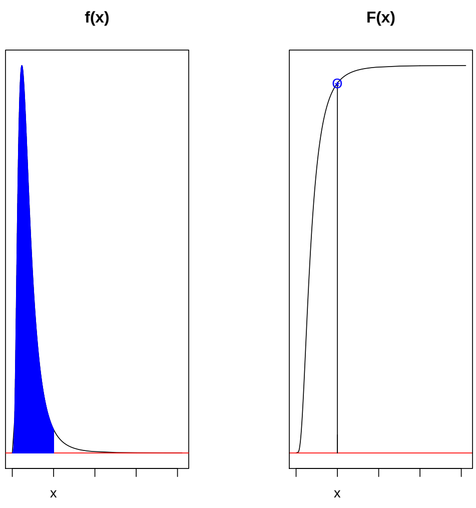


Figura 5.4: *Función de densidad y función de distribución acumulada.*

Como consecuencia de esta definición, la probabilidad de que la variable continua X tome valores entre a y b se puede expresar como $P(a \leq X \leq b)$=F(b)-F(a).

5.2. La distribución normal

La distribución normal es la distribución de probabilidad que más se utiliza para modelizar variables aleatorias continuas.

La **función de densidad de** una variable aleatoria continua X que sigue **una distribución normal** N(μ,σ) es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

donde μ y σ son números tales que $-\infty < \mu < \infty$ y $0 < \sigma < \infty$. Además e y π son las constantes e=2,71828... y π=3,14159...

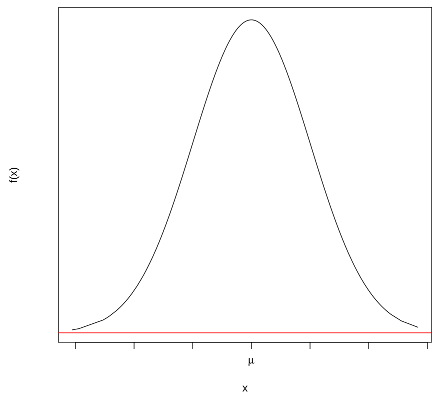


Figura 5.5: Función de densidad de una distribución normal.

Si X$\sim$ $N(\mu,\sigma)$, se cumplen las siguientes propiedades:
1. μ es el valor esperado de X, E(X), o media poblacional.
2. σ es la desviación típica poblacional de X.
3. La forma de la función de densidad es una curva simétrica en forma de campana centrada en la media, μ,como se ve en la Figura 5.5.
4. Dos funciones de densidad con la misma desviación típica y distintas medias producen gráficas idénticas pero desplazadas (primera gráfica de la Figura 5.6).
5. Dos funciones de densidad con la misma media y distinta desviación típica producen gráficas centradas en torno a la media común,pero la que tiene mayor desviación típica se corresponde con una campana más baja y con mayor dispersión (segunda gráfica de la Figura 5.6).
6. Se cumple que los intervalos $[\mu - \sigma,\ \mu + \sigma]$ y $[\mu - 2\sigma,\ \mu + 2\sigma]$ contienen aproximadamente el 62 % y el 95 % de los datos de la población $N(\mu,\sigma)$, respectivamente.

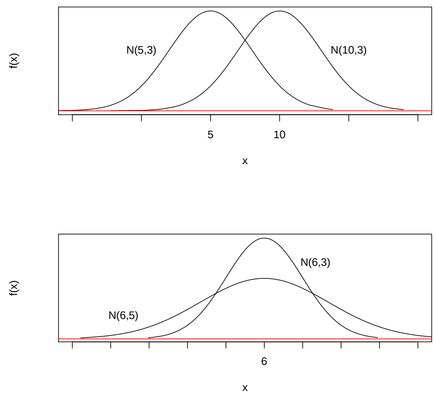


Figura 5.6: Efectos de μ y σ en la función de densidad de una variable aleatoria normal.

Al igual que hemos visto en el apartado anterior, la probabilidad de que una variable aleatoria X$\sim$N(μ,σ) tome valores en un intervalo [a,b], se corresponde con el área situada bajo la función de densidad entre a y b, como muestra la Figura 5.7. En general, el cálculo de esta probabilidad se obtiene utilizando la función de distribución acumulada F(x), de tal manera que:

$$P(a < X < b) = P(X < b) - P(X < a) = F(b) - F(a)$$

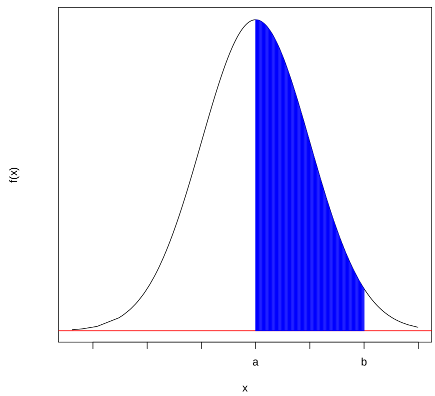


Figura 5.7: Función de densidad normal. El área sombreada indica $P(a<X<b)$.

Sin embargo, en el caso de la distribución normal, no disponemos de una expresión algebraica para la función de distribución acumulada F(x). La metodología para resolver probabilidades en este caso se basa en dos pasos:

1. Convertir X$\sim$N(μ,σ) en Z$\sim$N(0,1).
2. Utilizar la tabla de la función de distribución acumulada de la normal N(0,1).

La distribución normal estándar

Una variable aleatoria normal Z que tiene media 0 y varianza 1, N(0,1), se dice que sigue una distribución **normal estándar**.

Cualquier variable aleatoria normal X$\sim$N(μ,σ) se puede convertir en una variable normal estándar, Z, utilizando la siguiente transformación, a la que se denomina **tipificar**:

$$Z = \frac{X - \mu}{\sigma}$$

Este importante resultado nos permite utilizar la tabla de la función de distribución acumulada de una normal estándar para calcular las probabilidades de cualquier variable normalmente distribuida.

Tabla de probabilidades de la distribución N(0,1)
$[P(Z < z)]$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5039	0.5079	0.5119	0.5159	0.5199	0.5239	0.5279	0.5318	0.5358
0.1	0.5398	0.5437	0.5477	0.5517	0.5556	0.5596	0.5635	0.5674	0.5714	0.5753
0.2	0.5792	0.5831	0.5870	0.5909	0.5948	0.5987	0.6025	0.6064	0.6102	0.6140
0.3	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6405	0.6443	0.6480	0.6517
0.4	0.6554	0.6590	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6843	0.6879
0.5	0.6914	0.6949	0.6984	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7290	0.7323	0.7356	0.7389	0.7421	0.7453	0.7485	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7733	0.7763	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7938	0.7967	0.7995	0.8023	0.8051	0.8078	0.8105	0.8132
0.9	0.8159	0.8185	0.8212	0.8238	0.8263	0.8289	0.8314	0.8339	0.8364	0.8389
1.0	0.8413	0.8437	0.8461	0.8484	0.8508	0.8531	0.8554	0.8576	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8707	0.8728	0.8749	0.8769	0.8790	0.8810	0.8829
1.2	0.8849	0.8868	0.8887	0.8906	0.8925	0.8943	0.8961	0.8979	0.8997	0.9014
1.3	0.9032	0.9049	0.9065	0.9082	0.9098	0.9114	0.9130	0.9146	0.9162	0.9177
1.4	0.9192	0.9207	0.9221	0.9236	0.9250	0.9264	0.9278	0.9292	0.9305	0.9318
1.5	0.9331	0.9344	0.9357	0.9369	0.9382	0.9394	0.9406	0.9417	0.9429	0.9440
1.6	0.9452	0.9463	0.9473	0.9484	0.9494	0.9505	0.9515	0.9525	0.9535	0.9544
1.7	0.9554	0.9563	0.9572	0.9581	0.9590	0.9599	0.9607	0.9616	0.9624	0.9632
1.8	0.9640	0.9648	0.9656	0.9663	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.9	0.9712	0.9719	0.9725	0.9731	0.9738	0.9744	0.9750	0.9755	0.9761	0.9767
2.0	0.9772	0.9777	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9816
2.1	0.9821	0.9825	0.9829	0.9834	0.9838	0.9842	0.9846	0.9849	0.9853	0.9857
2.2	0.9860	0.9864	0.9867	0.9871	0.9874	0.9877	0.9880	0.9883	0.9886	0.9889
2.3	0.9892	0.9895	0.9898	0.9900	0.9903	0.9906	0.9908	0.9911	0.9913	0.9915
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9937	0.9939	0.9941	0.9942	0.9944	0.9946	0.9947	0.9949	0.9950	0.9952
2.6	0.9953	0.9954	0.9956	0.9957	0.9958	0.9959	0.9960	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9975	0.9976	0.9977	0.9978	0.9978	0.9979	0.9980	0.9980
2.9	0.9981	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.0	0.9986	0.9986	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9989	0.9989

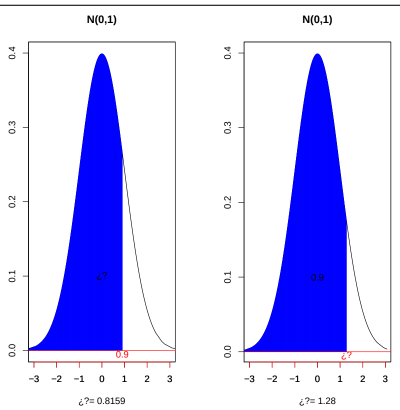


Figura 5.8: Ejemplo de manejo de la tabla de la distribución N(0,1).

Veamos como utilizar esta tabla para calcular las probabilidades de una normal estándar. La tabla proporciona las probabilidades F(z)=P(Z≤z) correspondientes a los valores no negativos de z. Por ejemplo, P(Z≤1.25)=F(1.25)=0.8943. Esto significa que en una distribución N(0,1), el 89.43 % de los valores son menores que 1.25, por lo que 1.25 es el percentil 89.43, es decir, $P_{89,43} = 1{,}25$. Los valores de la función de distribución acumulada correspondiente a los valores negativos de Z pueden deducirse utilizando la simetría de la función de densidad. Por ejemplo, P(Z≤-1.25)=P(Z≥1.25) 1-P(Z≤1.25)= 1-F(1.25)= 1-0.8944= 0.1057.

Ejemplo 5.1.

Considera la población N(0,1). Calcula el percentil P_{85} y un intervalo centrado en el 0 que contenga el 70 % de los datos de la población.

Solución:
El P_{85} es un valor de la población que verifica $P(Z < P_{85}) = 0{,}85$. Para calcularlo, se busca en el centro de la tabla el valor más próximo a 0.85. Se puede comprobar que este valor aparece en la fila 1.0 cruzada con la columna 0.04. Por lo tanto, $P_{85} = 1{,}04$
Por otra parte, el intervalo que contiene un 70 % de los datos de la población se puede denotar $[z_1, z_2]$ donde $z_2 > 0$, $z_1 < 0$ y $z_1 = -z_2$. Como $P(z_1 < Z < z_2) = 0{,}7$, se cumple que $P(Z < z_2) = 0{,}85$. Luego $z_2 = P_{85} = 1{,}04$ y $[z_1, z_2]$= [-1.04, 1.04]

Ejemplo 5.2.

Las batería de los oxímetros hospitalarios tienen una duración según una distribución normal que tiene una media de 1.200 horas y una desviación típica de 250 horas.
a) ¿Cuál es la probabilidad de que un determinado oxímetro dure entre 900 y 1.300 horas?
b) Calcula un intervalo centrado que contenga el 95 % de la duración de las baterías de los oxímetros hospitalarios.

Solución:
a) P(900<X<1.300)= P($\frac{900-1{,}200}{250} < \frac{X-1{,}200}{250} < \frac{1{,}300-1{,}200}{250}$))= P(-1.2<Z<0.4)=

= F(0.4)-F(-1.2)= F(0.4)-(1-F(1.2))= 0.6554-(1-0.8849)= 0.5403.

Por lo tanto, la probabilidad de que un oxímetro dure entre 900 y 1.300 horas es aproximadamente de 0.54.

b) 0.95= P($x_1 < X < x_2$)= P($\frac{x_1-1{,}200}{250} < Z < \frac{x_2-1{,}200}{250}$)

$\frac{x_1-1{,}200}{250} = -1{,}96 \rightarrow x_1 = 710$

$\frac{x_2-1{,}200}{250} = 1{,}96 \rightarrow x_2 = 1690$

El 95 % de las baterías de los oxímetros hospitalarios tienen una duración de entre 710 y 1690 horas.

Ejemplo 5.3.

Un grupo muy numeroso de estudiantes obtiene unas calificaciones de 0 a 100 puntos, que siguen una distribución normal con media 60 puntos y una desviación típica de 15 puntos. ¿Qué proporción de los estudiantes obtiene una calificación de entre 85 y 95 puntos?

Solución:

P(85<X<95)= P($\frac{85-60}{15} < \frac{X-60}{15} < \frac{95-60}{15}$))= P(1.67<Z<2.33)=

= F(2.33)-F(1.67)= 0.9900-0.9525= 0.0376.

Es decir, el 3.76 por ciento de los estudiantes obtuvo una calificación comprendida entre 85 y 95.

Autoevaluación

1. Si Z∼N(0,1), Calcula:
   a) P(Z<2.13). Sol: 0.9834
   b) P(Z>2.13). Sol: 0.0166
   c) P(Z<-2.13). Sol: 0.0166
   d) P_{50}. Sol: 0
   e) P_{87}. Sol: 1.13
   f) P_{90}. Sol: 1.28
2. Se sabe que la hemoglobina en sangre de la población española, sigue una distribución normal de media $\mu = 14{,}1$ g/dl y desviación típica $\sigma = 5$ g/dl. Se considera que una persona tiene anemia si tiene valores de hemoglobina por debajo de 10 g/dl. Por otra parte, se sabe que los pacientes con cardiopatía congénita tienen valores de hemoglobina por encima de 19 g/dl.
   a) Calcula la probabilidad de una persona tenga una nivel de hemoglobina entre 15.1 y 16.9. Sol: 0.133
   b) Calcula el porcentaje esperado de anémicos en la población española. Sol: 20.61 %
   c) Calcula el porcentaje esperado de pacientes con cardiopatía congénita. Sol: 16.35 %

A continuación veremos dos métodos para estimar parámetros poblacionales basados en la información que contiene una muestra aleatoria. En primer lugar, se estimará un parámetro poblacional desconocido mediante un único número llamado **estimación puntual**. Después analizaremos un segundo método que tiene en cuenta una medida de variabilidad, estableciendo un **intervalo de confianza** de valores en el que es probable que se encuentre la cantidad que queremos estimar.

5.3. Estimadores puntuales

Un **estimador** de un parámetro poblacional es una variable aleatoria que depende de la información de la muestra: su valor proporciona aproximaciones a este parámetro desconocido. Un valor específico de esa variable aleatoria se llama **estimación**.

Consideremos un parámetro poblacional como la media poblacional, μ. Un **estimador puntual** de un parámetro poblacional es una función de la información de la muestra que genera un único número llamado **estimación puntal**. Por ejemplo, la media muestral $\overline{X}$ es un estimador puntual de la media poblacional, μ, y el valor que toma $\overline{X}$ para un conjunto dado de datos se llama estimación puntual, $\overline{x}$. La mediana muestral es, también, un estimador puntual, de la media poblacional, alternativo a la media muestral, sin embargo, no es el mejor estimador de la media de algunas distribuciones.

La varianza poblacional, σ^2, es otro parámetro poblacional y su estimación puntual es la varianza muestral s^2.

La desviación poblacional, σ, es otro parámetro poblacional y su estimación puntual es la desviación típica muestral s.

Por último, si la variable aleatoria de interés es cualitativa, el porcentaje de una característica poblacional, p, es otro parámetro poblacional y su estimador puntual es el porcentaje muestral, $\widehat{p}$.

5.4. Intervalos de confianza de la media: desviación típica poblacional conocida

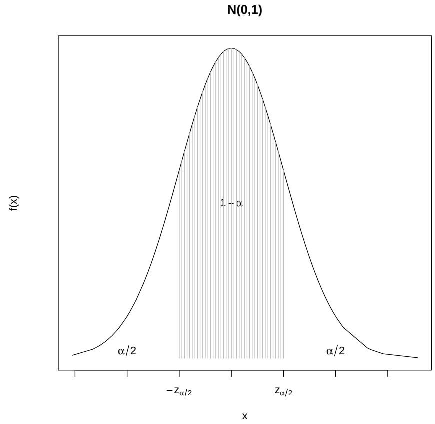
En este apartado supondremos que las muestras se extraen de una población con media, μ, desconocida y **desviación típica, σ, conocida**. El objetivo será hallar un intervalo de valores en el que posiblemente se encuentre μ, en lugar de proporcionar un único valor que estime este parámetro (estimación puntual).

Intervalo de confianza y nivel de confianza.

Sea θ un parámetro desconocido de una población de la que se obtiene una muestra aleatoria. Obtener un intervalo [a, b] para estimar θ con un nivel de confianza del (1 - α)100 %, consiste en averiguar a y b, con información relativa a la muestra, de tal manera que P(a $< \theta <$ b)=1 - α, siendo α un valor entre 0 y 1. Esto significa que, una vez averiguado [a, b], no tenemos total seguridad de que θ esté en el intervalo, sino que si se extrajeran muchas muestras de la población y para cada una de ellas se obtuviera su correspondiente intervalo [a, b] que estima θ, se tendría la confianza de que el parámetro estaría en el (1 - α)100 % de estos intervalos.

Intervalos basados en la distribución normal.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de n observaciones extraídas de una población que sigue una distribución normal de media, μ, desconocida y desviación típica, σ, conocida. Supongamos que queremos averiguar un **intervalo de confianza de la media poblacional al (1 - α)100 %**. Se puede demostrar teóricamente que la variable aleatoria Z=$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sigue una distribución normal N(0,1) y que $z_{1-\alpha/2}$ es el valor de la distribución normal estándar tal que la probabilidad de la cola superior es $\alpha/2$, como se puede ver en la siguiente figura.



$$1-\alpha = P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = P(-z_{1-\alpha/2} < \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}) =$$

$$P(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{x}-\mu < z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}) =$$

$$= P(\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Por lo tanto, el intervalo de confianza al (1 - α)100 % de la media poblacional, cuando la desviación típica es conocida es:

$$[\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}] \ \equiv\ \overline{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},$$

donde $\overline{x}$ es la media muestral y $z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ es el **margen de error**.

Ejemplo 5.4.

Supongamos que el tiempo de recuperación de una intervención sobre trasplante de riñón sigue una distribución normal. Una muestra aleatoria de 16 pacientes trasplantados tuvo un tiempo medio de recuperación de 25 días. Supongamos que σ=6 días. Halla un intervalo de confianza que contenga el verdadero tiempo medio de recuperación al 95 %.

Solución:

Como fijamos el nivel de confianza al 95 %, entonces α=0.05.

$[\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}] = [25 - z_{0,975}\frac{6}{\sqrt{16}},\ 25 + z_{0,975}\frac{6}{\sqrt{16}}] = [22'06, 27'94]$

Con una confianza del 95 %, el verdadero tiempo medio de recuperación para los pacientes trasplantados de riñón en la población, oscila entre 22.06 y 27.94 días.

Si se extrajeran muchas muestras de 16 pacientes trasplantados y para cada muestra obtuviéramos su intervalo de confianza al 95 %, podríamos afirmar que el 95 % de los intervalos contendrían al verdadero tiempo medio de recuperación pero esto no significa que el intervalo que acabamos de calcular contenga, sin lugar a dudas, dicho valor.

Ejemplo 5.5.

Consideremos el mismo contexto del ejemplo anterior y los mismos datos muestrales. Si tras el estudio estadístico se estima que el verdadero tiempo medio de recuperación de los pacientes trasplantados oscila en [23'08, 26'92] días, calcula la confianza a la que se ha calculado dicho intervalo.

Solución:
$[23'08, 26'92] = [25 - z_{1-\alpha/2}\frac{6}{\sqrt{16}},\ 25 + z_{1-\alpha/2}\frac{6}{\sqrt{16}}] \Rightarrow 26'92 = 25 + z_{1-\alpha/2}\frac{6}{\sqrt{16}} \Rightarrow z_{1-\alpha/2} = 1{,}3 \Rightarrow 1 - \alpha/2 = 0{,}9032 \Rightarrow \alpha = 0{,}1936 \Rightarrow 1 - \alpha = 0{,}8064$
En definitiva, la confianza del intervalo es, aproximadamente, del 80 %

Autoevaluación

1. Los siguientes valores corresponden con los niveles de albúmina en sangre de 8 personas sanas (expresados en g/dl):

   3'5, 4'1, 3'7, 5'2, 5'1, 3'8, 4'2, 4'7

   Supongamos que el nivel de albúmina en sangre sigue una distribución normal con una desviación típica poblacional de σ=1'3 g/dl.
   Calcula un intervalo de confianza, al 80 %, para la media e **interpreta el resultado**. Sol:[3'70,4'88]

5.5. Intervalos de confianza de la media: desviación típica poblacional desconocida

En el apartado anterior hemos visto intervalos de confianza de la media de una población normal cuando se conoce la desviación típica poblacional. En el caso en que no exista ninguna información histórica sobre la media poblacional, ni sobre la desviación típica poblacional, para poder construir intervalos de confianza para la media poblacional es necesario utilizar una nueva distribución de probabilidad.

Distribución t de Student

Los intervalos de confianza para la media vistos en el apartado anterior se basan en el hecho de que la variable aleatoria, Z=$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ sigue una distribución normal estándar. En el caso en que la desviación típica poblacional es desconocida, este resultado no puede utilizarse directamente. En esas circunstancias, es lógico considerar la variable aleatoria obtenida sustituyendo la σ desconocida por la desviación típica muestral, s, lo que nos da

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

Esta variable aleatoria no sigue una distribución normal estándar. Su distribución pertenece a una familia de distribuciones llamadas t de Student.

En concreto, dada una muestra aleatoria de n observaciones, de media $\overline{x}$ y desviación típica s, extraída de una población que sigue una distribución normal de media μ, la variable aleatoria $t = \frac{\overline{x}-\mu}{s/\sqrt{n}}$ sigue la **distribución t de Student con (n-1) grados de libertad**, $\mathbf{t_{n-1}}$.

La forma de la distribución t de Student es bastante parecida a la de la distribución normal estándar. Ambas distribuciones tienen una media de 0 y las dos funciones de densidad son simétricas en torno a sus medias. Sin embargo, la función densidad de la distribución t de Student tiene mayor dispersión que la distribución normal estándar.

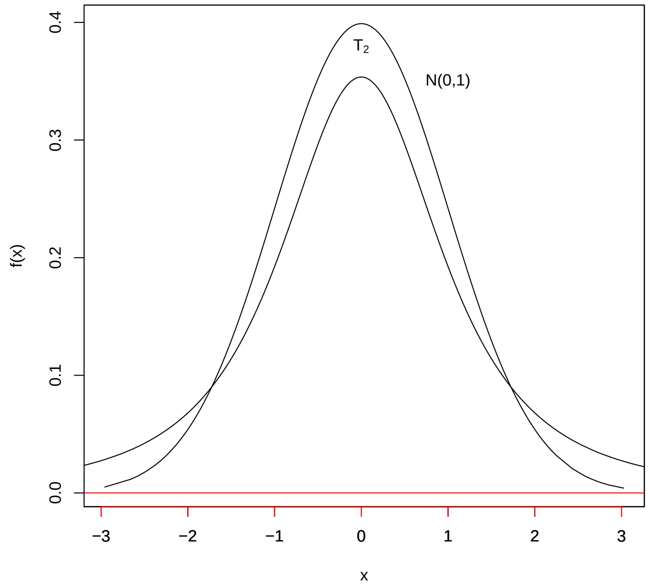


Figura 5.9: Funciones de densidad de la distribución normal estándar y la distribución t de Student con 2 grados de libertad.

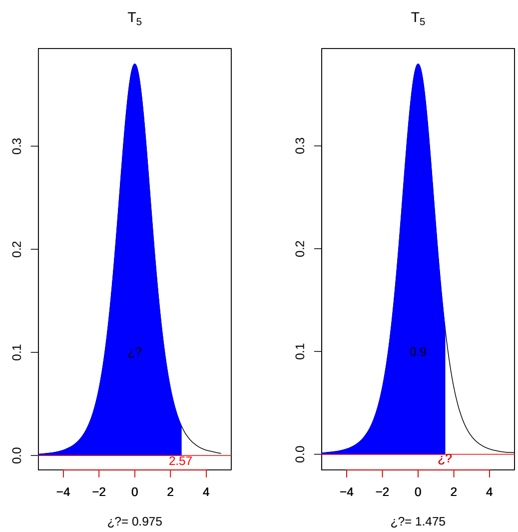


Figura 5.10: Ejemplo de manejo de la tabla de la distribución T Student.

Tabla de probabilidades de la distribución *t de Student*

$[P(t < T)]$

$g \diagdown T$	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.9750	0.990	0.995
1	0.509	0.726	1.000	1.376	1.962	3.077	6.313	12.706	31.820	63.656
2	0.444	0.617	0.816	1.060	1.386	1.885	2.919	4.302	6.964	9.924
3	0.424	0.584	0.764	0.978	1.249	1.637	2.353	3.182	4.540	5.840
4	0.414	0.568	0.740	0.940	1.189	1.533	2.131	2.776	3.746	4.604
5	0.408	0.559	0.726	0.919	1.155	1.475	2.015	2.570	3.364	4.032
6	0.404	0.553	0.717	0.905	1.134	1.439	1.943	2.446	3.142	3.707
7	0.401	0.549	0.711	0.896	1.119	1.414	1.894	2.364	2.997	3.499
8	0.399	0.545	0.706	0.888	1.108	1.396	1.859	2.306	2.896	3.355
9	0.397	0.543	0.702	0.883	1.099	1.383	1.833	2.262	2.821	3.249
10	0.396	0.541	0.699	0.879	1.093	1.372	1.812	2.228	2.763	3.169
11	0.395	0.539	0.697	0.875	1.087	1.363	1.795	2.200	2.718	3.105
12	0.394	0.538	0.695	0.872	1.083	1.356	1.782	2.178	2.680	3.054
13	0.393	0.537	0.693	0.870	1.079	1.350	1.770	2.160	2.650	3.012
14	0.393	0.536	0.692	0.868	1.076	1.345	1.761	2.144	2.624	2.976
15	0.392	0.535	0.691	0.866	1.073	1.340	1.753	2.131	2.602	2.946
16	0.392	0.535	0.690	0.864	1.071	1.336	1.745	2.119	2.583	2.920
17	0.391	0.534	0.689	0.863	1.069	1.333	1.739	2.109	2.566	2.898
18	0.391	0.533	0.688	0.862	1.067	1.330	1.734	2.100	2.552	2.878
19	0.391	0.533	0.687	0.860	1.065	1.327	1.729	2.093	2.539	2.860
20	0.390	0.532	0.686	0.859	1.064	1.325	1.724	2.085	2.527	2.845
21	0.390	0.532	0.686	0.859	1.062	1.323	1.720	2.079	2.517	2.831
22	0.390	0.532	0.685	0.858	1.061	1.321	1.717	2.073	2.508	2.818
23	0.390	0.531	0.685	0.857	1.060	1.319	1.713	2.068	2.499	2.807
24	0.389	0.531	0.684	0.856	1.059	1.317	1.710	2.063	2.492	2.796
25	0.389	0.531	0.684	0.856	1.058	1.316	1.708	2.059	2.485	2.787
26	0.389	0.530	0.684	0.855	1.057	1.314	1.705	2.055	2.478	2.778
27	0.389	0.530	0.683	0.855	1.056	1.313	1.703	2.051	2.472	2.770
28	0.389	0.530	0.683	0.854	1.055	1.312	1.701	2.048	2.467	2.763
29	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.389	0.530	0.682	0.853	1.054	1.310	1.697	2.042	2.457	2.749
40	0.388	0.528	0.680	0.850	1.050	1.303	1.683	2.021	2.423	2.704
60	0.387	0.527	0.678	0.847	1.045	1.295	1.670	2.000	2.390	2.660
120	0.386	0.525	0.676	0.844	1.040	1.288	1.657	1.979	2.357	2.617
∞	0.385	0.524	0.674	0.841	1.036	1.281	1.644	1.959	2.326	2.575

A medida que aumenta el número de grados de libertad, es decir, cuando aumenta la muestra, la distribución t de Student es cada vez más parecida a la distribución normal estándar. Este resultado es intuitivamente razonable y se deduce del hecho de que cuando la muestra es grande, la desviación típica muestral es un estimador muy preciso de la desviación típica poblacional.

Para basar las inferencias sobre una media poblacional en la distribución t de Student se necesitan valores críticos análogos a los $z_{1-\alpha/2}$ de la normal estándar. De la misma forma que $z_{1-\alpha/2}$ es el valor de la distribución normal estándar tal que la probabilidad de la cola superior es $\alpha/2$, $t_{n-1,1-\alpha/2}$ es el valor de la distribución t de Student con n-1 grados de libertad tal que la probabilidad de la cola superior es $\alpha/2$.

Intervalos basados en la distribución t de Student.

Supongamos que tenemos una muestra aleatoria de n observaciones extraídas de una distribución normal de media μ y desviación típica desconocida. Si la media y la desviación típica muestrales son $\overline{x}$ y s, respectivamente, entonces el **intervalo de confianza al** $(1-\alpha)100\,\%$ **de la media poblacional, cuando la varianza es desconocida**, viene dado por

$$[\overline{x} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}, \quad \overline{x} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}] \equiv \overline{x} \pm t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$$

donde $t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$ es el **margen de error**.

Ejemplo 5.6.

Estima el nivel medio de hemoglobina para pacientes con una cardiopatía congénita, con una confianza del 90 por ciento, suponiendo que los niveles de hemoglobina de 24 pacientes con esta enfermedad han sido:

$$15'5,\ 21'0,\ 18'5,\ 19'3,\ 19'7,\ 16'9,\ 20'2,\ 14'5,\ 16'5,\ 19'2,\ 18'7,\ 18'2,$$
$$18'0,\ 17'5,\ 18'5,\ 20'5,\ 18'6,\ 19'1,\ 19'8,\ 18'0,\ 19'8,\ 18'2,\ 20'3,\ 21'8$$

Solución:
Calculamos la media y la desviación típica muestral:
$\overline{x} = 18'68$ $s = 1'69526$, $t_{n-1,1-\alpha/2} = t_{23,0'95} = 1'713$
El intervalo de confianza al 90 por ciento es:
$\overline{x} \pm t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}} = 18'68 \pm t_{23,0'95}\frac{1'69526}{\sqrt{24}} = 18'68 \pm 1'713 \times 0'3460 = 18'68 \pm 0'5930 = [18'09,\ 19'27]$

Ejemplo 5.7.

Consideremos el mismo contexto del ejemplo anterior y los mismos datos muestrales. Si tras el estudio estadístico se estima que el verdadero nivel medio de hemoglobina para pacientes con una cardiopatía congénita oscila en [17'71, 19'65] , calcula la confianza a la que se ha calculado dicho intervalo.

Solución:
$[17'71, 19'65] = [18,68 - t_{23,1-\alpha/2}\frac{1'69526}{\sqrt{24}},\ 18'68 + t_{23,1-\alpha/2}\frac{1'69526}{\sqrt{24}}] \Rightarrow 19'65 = 18'68 + t_{23,1-\alpha/2}\frac{1'69526}{\sqrt{24}}$
$\Rightarrow t_{23,1-\alpha/2} = 2{,}80 \Rightarrow 1-\alpha/2 = 0{,}995 \Rightarrow \alpha = 0{,}01 \Rightarrow 1-\alpha = 0{,}99$
En definitiva, la confianza del intervalo es, aproximadamente, del 99 %

Autoevaluación

1. Los siguientes valores corresponden con los niveles de albúmina en sangre de 8 personas sanas (expresados en g/dl):

   3’5, 4’1, 3’7, 5’2, 5’1, 3’8, 4’2, 4’7

   Supongamos que el nivel de albúmina en sangre sigue una distribución normal.

   Calcula un intervalo de confianza, al 80 %, para la media e **interpreta el resultado**. Sol:[3’95,4’60]

5.6. Intervalos de confianza de porcentajes de la población (grandes muestras)

Imaginemos que estamos interesados en estudiar el porcentaje de miembros de la población que posee una característica específica. Si se toma una muestra aleatoria de la población, el porcentaje muestral constituye un estimador puntual natural del porcentaje de la población. Veamos como se desarrollan intervalos de confianza para el porcentaje de una población.

Teóricamente, para muestras suficientemente grandes, la variable aleatoria $\widehat{P} \sim N(p, \frac{p(100-p)}{n})$ y por lo tanto, $Z = \frac{\widehat{P}-p}{\sqrt{\frac{p(100-p)}{n}}} \sim N(0,1)$

Cuando la muestra es suficientemente grande, se obtiene una buena aproximación si se sustituye p por el estimador puntual $\widehat{P}$ en el denominador, es decir,

$$\sqrt{\frac{p(100-p)}{n}} \approx \sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}$$

Por lo tanto, cuando el tamaño de la muestra es grande, la distribución de la variable aleatoria $Z = \frac{\widehat{P}-p}{\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}} \sim N(0,1)$

Razonando del mismo modo que en los apartados anteriores:

$$1-\alpha = P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = P(-z_{1-\alpha/2} < \frac{\widehat{P}-p}{\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}} < z_{1-\alpha/2}) =$$

$$P(-z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}} < \widehat{P}-p < z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}) =$$

$$P(\widehat{P}-z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}} < p < \widehat{P}+z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}})$$

Así, se obtiene un **intervalo de confianza al** $(1-\alpha)100\,\%$ **de la proporción de la población**:

$$[\widehat{P}-z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}},\ \widehat{P}+z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}]$$

donde $z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}$ es el **margen de error**.

Ejemplo 5.8.

Se desea estimar el porcentaje de niños que han tenido al menos una revisión bucal tras la dentición definitiva. Para ello, se extrae una muestra aleatoria de 344 niños, de entre 6 y 7 años, de los cuales 261 han acudido a una revisión odontológica. Halla un intervalo de confianza al 90 por ciento del verdadero porcentaje de niños que revisan su boca tras la dentición definitiva.

Solución:

Si p representa el verdadero porcentaje de la población y $\widehat{P}$ el porcentaje muestral, los intervalos de confianza del porcentaje de la población se obtienen por medio de:

$$[\widehat{P} - z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}},\ \widehat{P} + z_{1-\alpha/2}\sqrt{\frac{\widehat{P}(100-\widehat{P})}{n}}]$$

donde en el caso de un intervalo de confianza al 90 por ciento, $\alpha = 0{,}10$, por lo que a partir de la distribución normal estándar, $\alpha/2 = 0'05$ y $z_{1-\alpha/2} = z_{0'95} = 1'645$
Se deduce que n= 344, $\widehat{P} = \frac{261}{344}100$=75'9, por lo tanto el intervalo de confianza del porcentaje de la población al 90 por ciento es

$$[75'9 - 1'645\sqrt{\frac{75'9 \cdot 24'1}{344}},\ 75'9 + 1'645\sqrt{\frac{75'9 \cdot 24'1}{344}}] = [72'1, 79'7]$$

Podríamos decir que alrededor del 76 por ciento de los niños con dentición definitiva se revisan los dientes, con un margen de error del 4 por ciento, bajo un nivel de confianza del 90 por ciento.

Autoevaluación

1. El último estudio sanitario en la Comunidad Valenciana, alerta que de 27 farmacias sondeadas 14 registran problemas de suministro de medicamentos básicos.

   Estima el porcentaje de farmacias con problemas se suministro con un intervalo de confianza del 80 %. Sol:[39'54, 64'16]

5.7. Seminario de PROBABILIDAD Y ESTIMACIÓN DE UNA POBLACIÓN

(Distribución Normal(0,1))

1. Calcula las siguientes probabilidades, considerando que Z sigue una distribución $Normal(0, 1)$
   a) $P(Z < 1{,}56)$
   b) $P(Z < 2{,}78)$
   c) $P(Z > 3{,}00)$
   d) $P(Z > 1{,}01)$
   e) $P(Z < -1{,}5)$
   f) $P(Z > -2{,}61)$
   g) $P(Z < -0{,}32)$
   h) $P(Z > -1{,}63)$
   i) $P(0{,}83 < Z < 1{,}64)$
   j) $P(-1{,}25 < Z < 2{,}37)$
   k) $P(-2{,}36 < Z < -1{,}33)$
   l) Valor z_1 tal que $P(X > z_1) = 0{,}9978$
2. Calcula los siguientes intervalos:
   a) Intervalo centrado en 0 que contenga entre sus valores una probabilidad de 0,90
   b) Intervalo centrado en 0 que contenga entre sus valores una probabilidad de 0,95
   c) Intervalo centrado en 0 que contenga entre sus valores una probabilidad de 0,99
   d) Intervalo de la $N(0, 1)$ que contiene el 95 % de los valores mayores, es decir, que deja fuera un 5 % de los valores menores.
   e) Intervalo de la $N(0, 1)$ que contiene el 95 % de los valores menores, es decir, que deja fuera un 5 % de los valores mayores.
3. Calcula los siguientes percentiles:
   a) Percentil 50.
   b) Percentil 75.
   c) Percentil 90.
   d) Percentil 10.
   e) Percentil 25.

(Distribuciones Normales)

4. Se sabe que el peso de los niños de 1 año de edad sigue (aproximadamente) una distribución $N(7, 2)$ (en kg).
   a) Calcula un intervalo centrado que cubra el peso del 95 % de los niños de 1 año.
   b) Acude a la clínica de un pediatra una madre con un niño de 1 año que pesa 10,5 kg. ¿En qué percentil se encuentra el niño en cuanto a peso?, es decir, ¿qué porcentaje de niños de esa edad pesa menos que él?

c) Calcula un intervalo centrado que contenga el 99 % de los valores del peso de niños de un año.

d) Calcula un intervalo que contenga el 90 % de los pesos más altos (dejando fuera el 10 % de los pesos más bajos)

e) ¿Qué porcentaje de niños de 1 año pesa menos de 3,5 kg?

f) ¿Qué porcentaje de niños de 1 año pesa más de 4,5 kg?

g) ¿Qué porcentaje de niños de 1 año pesa entre 6 y 8 kg? ¿y entre 8 y 9?, ¿y entre 4 y 5?

5. Se sabe que la estatura de los alumnos varones matriculados en primero de la universidad CEU-Cardenal Herrera tiene una distribución $N(175, 8)$:

   a) Calcula un intervalo para la estatura de los alumnos, centrado en la media y que incluya al 95 % de éstos.

   b) Calcula un intervalo para la estatura de los alumnos, que contenga el 95 % de los alumnos de menor estatura.

   c) Calcula un intervalo para la estatura de los alumnos, que contenga el 95 % de los alumnos de mayor estatura.

   d) Que porcentaje de alumnos mide más de 190.

   e) Que porcentaje de alumnos mide menos de 182.

   f) Que porcentaje de alumnos mide entre 170 y 185.

6. El diámetro máximo de los hematíes de una persona con malaria por Plamodium vivax presenta las siguientes características: Si la célula esta infectada dicha variable se distribuye de forma normal con media 7,6 micras y desviación típica 0,81 micras, y si la célula no está infectada dicha variable se distribuye de forma normal con media 9,6 micras y desviación típica 1,0 micras. Calcular :

   a) Proporción de células no infectadas con un diámetro máximo mayor que 9,4 micras.

   b) Proporción de células no infectadas con un diámetro máximo inferior a 7 micras.

   c) Proporción de células infectadas con un diámetro máximo inferior a 9,4 micras.

   d) Da un intervalo centrado que contenga el 95 % de las células infectadas y repite el proceso para las células no infectadas.

(Intervalos de confianza)

Para todos los problemas que se proponen a continuación determina:

- Población en estudio.
- Variable en estudio y tipo de la misma.
- Parámetro de interés para el que se calcula el intervalo de confianza.
- Interpretación del intervalo de confianza obtenido.

(*Intervalo de confianza para la media poblacional: varianza poblacional conocida*)

7. En un experimento diseñado para estimar el número medio de latidos por minuto del corazón en los niños de 5 años se encontró que el número medio de latidos por minuto en una muestra aleatoria de 49 niños de 5 años era de 90 pulsaciones. Si la varianza de la variable en la población es $\sigma^2 = 100$, calcula un intervalo de confianza al 95 % para el número medio de latidos por minuto de esta población e interpreta el resultado.

8. En un estudio de la duración de la hospitalización realizado por varios hospitales en cooperación, se extrajo una muestra aleatoria de 64 pacientes con úlcera péptica de una lista de todos los pacientes con esta enfermedad admitidos alguna vez en los hospitales y se determinó, para cada uno, su duración de hospitalización. Se encontró que la duración media de la hospitalización fue de 8,25 días. Si se sabe que la desviación típica de la población es de 3 días, halla el intervalo al 90 % para la duración de hospitalización media poblacional e interpreta el resultado.
9. Una muestra de 100 hombres adultos aparentemente normales, de 25 años de edad, mostró una presión sistólica sanguínea media de 125. Si se sabe que la desviación típica de esta medida en la población es de 15 unidades, calcula el intervalo de confianza al 99 % para la presión sistólica sanguínea media de la población e interpreta el resultado.

(***Intervalo de confianza para la media poblacional: varianza poblacional desconocida***)

10. Una muestra de 25 niños de diez años de edad proporcionó un peso medio y una desviación típica de 36,5 y 5 kg respectivamente. Halla un intervalo de confianza al 90 % para el peso medio de niños de diez años de la población e interpreta el resultado.
11. Los análisis de los gases de la sangre arterial de una muestra de 15 hombres adultos físicamente activos dieron los siguientes valores de PaQ_2 en reposo:
$75, 80, 84, 74, 84, 78, 89, 72, 83, 76, 75, 87, 78, 79, 88$
Calcula un intervalo de confianza al 99 % para el nivel medio de PaQ_2 de la población en estudio e interpreta el resultado.
12. Los siguientes valores son las concentraciones de bilirrubina en suero de una muestra de 10 pacientes admitidos en un hospital para el tratamiento de la hepatitis:
$20'5, 14'8, 21'3, 12'7, 15'2, 26'6, 23'4, 22'9, 15'7, 19'2$ Construye un intervalo de confianza al 95 % para la concentración media de bilirrubina en suero de este tipo de pacientes e interpreta el resultado.
13. Como parte de un estudio de tiempos y movimientos conducido en un centro de salud, una muestra de 100 pacientes pasó en promedio 23 minutos en la sala de espera entre sus registros y su atención por un miembro del grupo médico. La desviación típica obtenida de los datos fue de 10. Calcula un intervalo de confianza al 90 % para el tiempo medio de espera en el centro de salud e interpreta el resultado.

(***Intervalo de confianza para el porcentaje poblacional***)

14. Un encargado del archivo de expedientes médicos extrajo al azar una muestra de 100 expedientes de pacientes y encontró que en el 8 por ciento de ellos la carátula tenía, al menos, un detalle de información que contradecía el resto de la información que aparecía en el expediente. Construye un intervalo de confianza al 99 % por ciento para el porcentaje de los expedientes que contienen dichas discrepancias.
15. Una encuesta a una muestra aleatoria de 150 familias en cierta comunidad urbana, reveló que, en el 87 por ciento de los casos, por lo menos uno de los miembros de la familia tenía alguna forma de seguro relacionado con la salud. Construye un intervalo de confianza al 90 % por ciento para el porcentaje de familias en la comunidad con al menos un seguro relacionado con la salud.
16. Estudios epidemiológicos revelan que en Italia, alrededor del 10 % de los mayores de 65 años tienen diabetes. Para averiguar si en España la prevalencia (porcentaje) de la enfermedad es significativamente diferente a la de Italia, se extrae una muestra de 500 mayores de 65 años y se determina que 38 de ellos tienen diabetes.
a) Indica cuál es la población de estudio, cuál es la variable de interés y clasifica de qué tipo de variable se trata.
b) Calcula un intervalo de confianza al 85 % para el porcentaje de diabéticos, mayores de 65 años, en España y determina, razonadamente, si la prevalencia de la diabetes de los dos países es diferente.

(Algunos ejercicios que han aparecido en exámenes)

17. (2010) Se sabe que la glucemia basal se distribuye como una normal con media poblacional de 106 mg y una desviación típica poblacional de 8 mg.
a) Calcula la probabilidad de que un paciente tenga una glucemia basal menor de 100 mg.
b) Calcula la probabilidad de que un paciente tenga una glucemia basal de al menos 100 mg.
c) Calcula la probabilidad de que un paciente tenga una glucemia basal de más de 116 mg.
d) Calcula la probabilidad de que un paciente tenga una glucemia basal de entre 86 y 106 mg.

18. (2010) Los historiales de una clínica de adelgazamiento revelan que el último grupo de 16 pacientes adelgazó una media de $\overline{x} = 15$ kilos con una desviación de s=2 kilos.
a) Calcula un intervalo de confianza al 80 % en el que se encuentra el verdadero peso medio de adelgazamiento, μ.
b) Averigua el nivel de confianza del intervalo [14.57, 15.43] para el peso medio de adelgazamiento, μ.

19. (2010) Se sabe que el nivel de plomo en la orina, en la población infantil, se distribuye según una normal con una media poblacional de 0.7 $\mu mol/24h$ y una desviación típica poblacional de 0.14 $\mu mol/24h$.
a) Calcula la probabilidad de que un niño tenga un nivel de plomo en la orina de menos de 0.5 $\mu mol/24h$.
b) Calcula la probabilidad de que un niño tenga un nivel de plomo en la orina de al menos 0.5 $\mu mol/24h$.
c) Calcula la probabilidad de que un niño tenga un nivel de plomo en la orina de más de 1 $\mu mol/24h$.
d) Calcula la probabilidad de que un niño tenga un nivel de plomo en la orina de entre 0.6 y 0.9 $\mu mol/24h$.

20. (2010) A 9 pacientes que sufren la misma incapacidad física se les pidió que llevaran a cabo cierta tarea como parte de un experimento. El tiempo medio requerido para realizar la tarea fue de $\overline{x} = 7$ minutos con una desviación típica de s= 2 minutos.
a) Calcula un intervalo de confianza al 95 % en el que se encuentre el verdadero tiempo medio, μ, para que este tipo de pacientes efectúen dicha tarea.
b) Averigua el porcentaje de confianza del intervalo [6.07, 7.93] para el verdadero tiempo medio de la realización de dicha tarea, μ.

21. (2011) Los análisis de los gases de la sangre arterial de una muestra de 5 hombres adultos, físicamente activos, dieron los siguientes valores de PaQ_2 en reposo:

75, 80, 84, 74, 84

a) Calcula un intervalo de confianza al 90 % para el verdadero nivel medio de PaQ_2.
b) Suponiendo que los niveles de PaQ_2 se distribuyen según una normal con media $\mu = 82$ y desviación típica $\sigma = 10$, calcula la probabilidad de que un hombre escogido al azar tenga un nivel de PaQ_2 mayor de 66.

22. (2011) Se han tomado muestras a 5 niños de entre 1 y 5 años del nivel de cobre en orina, obteniéndose los siguientes valores:

0'75, 0'80, 0'84, 0'74, 0'84

a) Calcula un intervalo de confianza al 90 % para el verdadero nivel medio de cobre en la orina e **interpreta su significado**.
b) Suponiendo que el nivel cobre en orina sigue una distribución normal con media $\mu = 0'82$ y desviación típica $\sigma = 0'1$, calcula la probabilidad de que un niño escogido al azar tenga un nivel de cobre en orina menor de 0'66.

23. (2011) Tras realizar una analítica a 345 pacientes valencianos, se comprueba que 29 de ellos han resultado tener diabetes. En función de estos resultados, [5'48, 11'34] es un intervalo de confianza para la proporción de diabéticos de la Comunidad Valenciana. Determina la confianza del intervalo.

24. (2011) . Se sabe que la estatura de los alumnos varones matriculados en primero de la universidad CEU-Cardenal Herrera tiene una distribución normal con media μ= 175 cm y varianza $\sigma^2 = 64\ cm^2$.

    a) Calcula el percentil 33 e **interpreta su significado**.

    b) Calcula el porcentaje de alumnos que mide más de 185.

25. (2012) Los hemogramas de 7 hombres adultos, sanos, en ayunas, proporcionaron los siguientes valores de hemoblogina glucosilada:

    5'7, 6'8, 7'4, 6'9 , 6'7, 7'8, 5'9

    a) Indica cuál es la población y cuál es la variable de estudio.

    b) Calcula un intervalo de confianza al 99 % para el nivel medio de la variable en la población e **interpreta su significado**.

    c) Suponiendo que los niveles de hemoglobina glucosilada se distribuyen según una normal con media μ = 6'5 y desviación típica σ= 1, calcula la probabilidad de que un hombre adulto sano escogido al azar tenga un nivel de esta variable mayor de 5.5.

26. (2012) A partir de una muestra de 197 personas tratadas habitualmente con un fármaco anti-hipertensivo se comprobó que 13 sufrían efectos adversos.

    a) Calcula un intervalo de confianza al 90 % para el porcentaje de efectos adversos de este fármaco hipertensivo e **interpreta su significado**.

    b) Sin calcular nada más, comenta cómo esperarías que fuera un intervalo de confianza para el mismo parámetro pero con un 99 % de confianza, más amplio o más estrecho.

27. (2013) Tras varias quejas realizadas sobre las recetas emitidas en un departamento de salud se está realizando un estudio sobre la calidad de las mismas. Se considera que si el porcentaje de fallos es superior al 7 % se debe abrir una investigación. Se han analizado 546 recetas seleccionadas al azar, y en 24 de ellas se ha detectado algún tipo de error en cuanto a la dosis o fármaco recetado. Calcula un intervalo de confianza al 95 % para el porcentaje de recetas erróneas de ese departamento de salud e interpreta su resultado. A la vista del intervalo obtenido, ¿crees que se debería abrir una investigación?. Razona tu respuesta

28. (2013) Es conocido por estudios recientes que el tiempo en el que una nueva anestesia infantil es efectiva sigue una distribución Normal con media 25 minutos y desviación típica 3.5 minutos (es decir, una distribución N(25,3.5)).

    a) Calcula el percentil 65 de esta distribución e interpreta su significado.

    b) Calcula la probabilidad de que el tiempo de efectividad de esta nueva anestesia sea superior a 30 minutos.

29. (2013) La analítica de 8 niños menores de un año diagnosticados de diabetes indica los siguientes valores de glucosa en sangre en ayunas:

    $$85, 75, 68, 98, 72, 58, 65, 90$$

    Calcula un intervalo de confianza al 99 % para el nivel medio de glucosa de los pacientes de la población en estudio e interpreta el resultado (en el contexto del problema).

30. (2013) La diosmina es el principal componente de ciertos comprimidos que se utilizan en el tratamiento del edema y los síntomas relacionados con la insuficiencia venosa crónica. Una de las empresas farmacéuticas que los fabrica, asegura en su prospecto que el contenido medio de diosmina por comprimido es de 450 mg. Se toma una muestra de 8 comprimidos y se analiza la cantidad de diosmina que contienen:

    $$403, 445, 408, 403, 444, 432, 407, 422$$

a) Calcula el percentil 35 e interpreta el resultado en el contexto del ejercicio.

b) Calcula un intervalo de confianza al 98 % e interpreta el resultado en el contexto del ejercicio.

c) En función del resultado obtenido en el apartado anterior, ¿crees que la empresa farmacéutica miente en su prospecto? Razona tu respuesta.

31. (2013) Se desea estimar qué porcentaje de compañías farmacéuticas han reducido su presupuesto en la investigación de enfermedades infecciosas, como la malaria, propias de países subdesarrollados. De demostrarse que el porcentaje es superior al 30 %, los altos mandatarios emprenderían acciones legales para rectificar esta grave tendencia. El estudio se lleva a cabo con un sondeo de 50 compañías farmacéuticas de las que 16 han reducido su presupuesto en la investigación de enfermedades infecciosas.

    a) Indica cuál la población de estudio, cuál es la variable de interés, su clasificación y el parámetro.

    b) Calcula un intervalo de confianza al 95 % e interpreta su resultado en el contexto del ejercicio.

    c) A la vista del resultado del apartado anterior, ¿crees que los altos mandatarios emprenderán acciones legales? Razona tu respuesta

32. (2014) En campañas anteriores, el Ministerio de Sanidad tasó la eficacia de la vacuna contra la gripe estacional, entre las personas pertenecientes a los grupos de riesgo, en torno al 85 %. Para llevar a cabo el mismo estudio respecto a la campaña anterior, se ha considerado el seguimiento sobre una muestra de 500 pacientes vacunados, de los cuales 400 no han enfermado de gripe.

    a) Determina cuál es la población de estudio, la variable de interés, su clasificación y el parámetro de estudio.

    b) Calcula un intervalo de confianza al 95 % e interpreta su resultado en el contexto del ejercicio.

    c) A la vista del resultado, razona si se puede considerar que la vacuna ha sido igual de eficaz que en campañas anteriores.

33. (2014) La longitud de un feto, en la semana 20 de gestación, sigue una distribución normal con media $\mu = 22.5$ cm y desviación típica, $\sigma = 2.85$ cm.

    a) Calcula la probabilidad de que, monitorizado un feto al azar, tenga una longitud inferior a 15 cm.

    b) Calcula un intervalo centrado que contenga el 80 % de las longitudes de fetos con 20 semanas de gestación.

    c) Calcula el percentil 25 e interpreta el resultado en el contexto del ejercicio.

34. (2014) La epinefrina es un componente habitual en medicamentos descongestivos antigripales. Se desea valorar si pacientes tratados con este componente tiene un nivel medio de glucemia basal (en ayunas) significativamente superior al normal, estimado entre 70 y 110 mg/dl. Para ello, se obtiene la glucemia a 8 adultos que han tomado epinefrina:

$$95, 135, 102, 120, 140, 100, 165, 122$$

    a) Calcula el P_{42} e interpreta su resultado en el contexto del ejercicio.

    b) Calcula un intervalo de confianza al 99 % e interpreta el resultado en el contexto del problema.

    c) A la vista del resultado, los pacientes tratados con epinefrina, ¿tienen un nivel medio de glucemia basal significativamente superior al normal?

35. (2015) En una investigación llevada a cabo en 2011, se concluyó que el 20 % de los pacientes enfermos del corazón, padecían hipertensión. Se desea estimar si, este porcentaje se ha modificado significativamente. Para ello, se seleccionan 152 pacientes enfermos de corazón y se observa que 38 de ellos padecen hipertensión.

    a) Determina en este problema la población de estudio, la variable de estudio, el tipo de la variable y el parámetro de interés.

    b) Calcula un intervalo de confianza al 95 % para estimar el parámetro de interés e **interpreta su resultado** en el contexto del ejercicio.

    c) A la vista del resultado del apartado anterior, razona si el porcentaje de hipertensos que están enfermos de corazón ha disminuido o aumentado significativamente respecto a la investigación de 2011.

36. (2015) Considera que el nivel de colesterol, LDL, en pacientes con más de 5 años de medicación, crónica con Atorvastatina calcio trihidrato, sigue una distribución Normal con media $\mu = 134{,}1$ mg/dL y una desviación típica $\sigma = 12{,}3$ mg/dL.

    a) Calcula la probabilidad de que un paciente de este tipo presente un nivel de colesterol inferior a 153.5 mg/dL.

    b) Calcula un intervalo que contenga el 99 % de los valores mayores de colesterol, LDL, en pacientes medicados más de 5 años con Atorvastatina calcio trihidrato.

37. (2015) Se desea estimar el porcentaje de nuevos casos de epilepsia en España. Para ello, se selecciona una muestra de 100.000 ciudadanos y como resultado, se diagnostican 30 nuevos casos de epilepsia.

    a) Determina en este problema la población de estudio, la variable de estudio, el tipo de la variable y el parámetro de interés.

    b) Calcula un intervalo de confianza al 95 % para estimar el parámetro de interés e **interpreta su resultado** en el contexto del ejercicio.

38. (2016) a) Los siguientes datos corresponden al nivel de colesterol, LDL, de 8 pacientes que llevan más de 5 años de medicación crónica con Atorvastatina calcio trihidrato:

    $$130{,}9; 158{,}2; 125{,}7; 155{,}5; 115{,}4; 123{,}3; 113{,}0; 137{,}1$$

    Calcula el percentil P_{42} e interpreta el resultado en el contexto del ejercicio.

    b) Considera ahora que el nivel de colesterol, LDL, en pacientes con más de 5 años de medicación crónica con Atorvastatina calcio trihidrato, sigue una distribución Normal con media $\mu = 134{,}1$ mg/dL y una desviación típica $\sigma = 12{,}3$ mg/dL. Calcula la probabilidad de que un paciente de este tipo presente un nivel de colesterol inferior a 153.5 mg/dL.

39. (2016) Se desea estimar el porcentaje de nuevos casos de Alzheimer en España, en personas de más de 60 años. Para ello, se selecciona una muestra de 1000 españoles de más de 60 años y son diagnosticados de Alzheimer 30 de ellos.

    a) Determina en este problema la población de estudio, la variable de estudio, el tipo de la variable y el parámetro de interés.

    b) Calcula un intervalo de confianza al 99 % para estimar el parámetro de interés e **interpreta su resultado** en el contexto del ejercicio.

5.8. Soluciones del Seminario de probabilidad y estimación de una población

1a) 0.9406
1b) 0.9972
1c) 0.0014
1d) 0.1563
1e) 0.0669
1f) 0.9954
1g) 0.3745
1h) 0.9484
1i) 0.1527
1j) 0.8854
1k) 0.0826
1l) -2.85
2a) [-1.64, 1.64]
2b) [-1.96, 1.96]
2c) [-2.58, 2.58]
2d) [-1.64, $+\infty$)
2e) ($-\infty$, 1.64]
3a) P50=0
3b) P75=0.67
3c) P90=1.28
3d) P10=-1.28
3e) P25=-0.67
4a) [3.08, 10.92]
4b) P95.99=10.5
4c) [1.84, 12.16]
4d) [4.44, $+\infty$)
4e) 4.01 %
4f) 89.43 %
4g) 38.28 %, 14.99 %, 9.18 %
5a) [159.32, 190.68]
5b) ($-\infty$, 188.12]
5c) [161.88, $+\infty$)
5d) 3.01 %
5e) 81.05 %
5f) 62.99 %
6a) 0.5792
6b) 0.0047
6c) 0.9864
6d) [6.01, 9.19], [7.64, 11.56]

7) Población: niños de 5 años. Variable: número de latidos por minuto del corazón. Tipo: cuantitativa discreta. Parámetro: número medio de latidos por minuto del corazón.

[87.2, 92.8]

8) Población: pacientes con úlcera péptica. Variable: duración de hospitalización. Tipo: cuantitativa continua. Parámetro: duración media de hospitalización.

[7.64, 8.87]

9) Población: hombres. Variable: presión sistólica sanguínea. Tipo: cuantitativa continua. Parámetro: presión sistólica media.

[121.13, 128.87]

10) Población: niños de 10 años. Variable: peso. Tipo: cuantitativa continua. Parámetro: peso medio.

[34.79, 38.21]

11) Población: hombres adultos, físicamente activos. Variable: nivel de los gases de la sangre arterial (P_aQ_2). Tipo: cuantitativa continua. Parámetro: nivel medio de P_aQ_2.

[75.98, 84.29]

12) Población: pacientes con hepatitis. Variable: concentraciones de bilirrubina en suero. Tipo: cuantitativa continua. Parámetro: concentración media de bilirrubina en suero.

[16.01, 22.45]

13) Población: pacientes de un centro de salud. Variable: tiempo en la sala de espera. Tipo: cuantitativa continua. Parámetro: tiempo medio de espera.

[21.35, 24.65]

14) Población: expedientes médicos. Variable: expediente erróneo (SI/NO). Tipo: cualitativa nominal. Parámetro: porcentaje de expedientes erróneos.

[1, 15]

15) Población: familias de una comunidad urbana. Variable: seguro de salud (SI/NO). Tipo: cualitativa nominal. Parámetro: porcentaje de ciudadanos con seguro de salud.

[82.5, 91.5]

16) Población: españoles mayores de 65 años. Variable: ¿padece diabetes? (SI/NO). Tipo: cualitativa nominal. Parámetro: porcentaje de diabéticos en España.

[5.89, 9.31]

17a) 0.2267

17b) 0.7733

17c) 0.1057

17d) 0.4937

18a) [14.33, 15.67]

18b) 60 %

19a) 0.0764

19b) 0.9236

19c) 0.0162

19d) 0.6847

20a) [5.46, 8.54]

20b) 80 %

21a) [74.85, 83.95]

21b) 0.9452

22a) [0.748, 0.84]

22b) 0.0548

23) 95 %

24a) P_{33}= 171.48

24b) 10.57 %

25a) Población: hombres adultos sanos. Variable: hemoglobina glucosilada.

25b) [5.69, 7.79]

25c) 0.8413

26a) [3.7, 9.5]

26b) IC(99 %) es más amplio que el IC(90 %)

27) [2.68, 6.12]

28a) P_{65}= 26.33

28b) 0.0764

29) [59.62, 93.14]

30a) P_{35}=407.15

30b) [401.55, 439.45]

31a) Población: compañías farmacéuticas. Variable: si han reducido o no su presupuesto en investigación sobre enfermedades infecciosas. Tipo: cualitativa nominal. Parámetro: porcentaje de compañías farmacéuticas que han reducido su presupuesto en investigación sobre enfermedades infecciosas.

31b) [19.07, 44.93]

32a) Población: personas vacunadas, pertenecientes a los grupos de riesgo. Variable: enfermar o no de gripe. Tipo: cualitativa nominal. Parámetro: porcentaje de eficacia de la vacuna.

32b) [76.49 %, 83.51 %]
33a) 0.0043
33b) [18.852, 26.148]
33c) $P_{25} = 20.59$
34a) $P_{42} = 116.04$
34b) [92.96, 151.8]
35a) Población: pacientes enfermos de corazón. Variable: si tienen o no hipertensión. Tipo: cualitativa nominal. Parámetro: porcentaje de hipertensos enfermos de corazón.
35b) [18.12, 31.88]
35c) Como $20\,\% \in [18.12, 31.88]$, no hay diferencias significativas con respecto a la investigación de 2011.
36a) 0.9429
36b) $[105.44, \infty)$
37a) Población: ciudadanos españoles. Variable: si es un nuevo caso de epilepsia o no. Tipo: cualitativa nominal. Parámetro: porcentaje de casos de epilepsia en España.
37b) [0.0192, 0.0407]
38a) $P_{42} = 125{,}172$
38b) 0.9429
39a) Población: españoles de más de 60 años. Variable: si tienen o no Alzheimer. Tipo: cualitativa nominal. Parámetro: porcentaje de nuevos casos de Alzheimer en España, de más de 60 años.
39b) [1.61, 4.39]

Capítulo 6

Regresión Lineal

La regresión lineal es una técnica estadística que modela la relación entre una variable dependiente, Y, y una o varias variables independientes, X_i.

6.1. Introducción

Tipos de relaciones entre variables cuantitativas

Relaciones determinísticas (relación perfecta, exacta)
Ejemplos:
V=K/P [relación entre la presión y el volumen de un gas]
E=VxT [espacio recorrido por un móvil con velocidad cte.]
Dosis de un medicamento= 0.02 * Peso
Relaciones aleatorias (relación sólo aproximada)
Ejemplos:
Talla $\approx$ K x EDAD [relación aproximada entre la talla (cm) y la edad (meses)]

Paciente	Edad (meses)	Talla (cm)	Peso (Kg)	IMC=$\frac{Peso}{Talla^2}$	Sexo	Tipo parto
1	3	55	4.7	15.54	niña	natural
2	6	68	7.3	15.79	niño	natural
3	5	64	6.2	15.14	niña	cesárea
4	5	66	6.7	15.38	niña	natural
5	3	62	5.5	14.31	niño	natural
6	4	65	5.1	9.47	niña	natural
7	9	74	6.5	16.44	niño	natural
8	8	75	6.9	14.22	niño	natural
9	9	73	6.7	16.89	niña	cesárea
10	7	69	5.3	14.70	niño	natural
11	6	73	4.9	11.26	niña	natural
12	5	68	5.3	10.81	niña	cesárea
13	8	73	7.3	15.01	niña	natural
14	6	71	7.5	11.90	niño	natural

Autoevaluación

1. En las siguientes parejas determina cuál es la variable dependiente y cuál es la variable independiente.
   a) Grados de alcohol de una bebida y número de calorías que proporcionan.

b) Número de trabajadores realizando un proyecto y número de días que dura el proyecto.

c) Cantidad de sal ingerida y tensión arterial

Sol: a) alcohol variable independiente y calorías variable dependiente. b) trabajadores variable independiente y días variable dependiente. c) sal variable independiente y tensión arterial variable dependiente.

Idea intuitiva de regresión

Dadas n parejas de valores (x_i, y_i) obtenidos de una muestra aleatoria de dos variables cuantitativas, su representación gráfica en el plano cartesiano da lugar a una nube de puntos o diagrama de dispersión. Si a dicha nube se le puede ajustar alguna curva entonces se dice que existe regresión. A tal curva se le llama línea de regresión. La variable ubicada en el eje horizontal (abcisas) es la llama variable independiente y a la ubicada en el eje vertical (ordenadas) es la variable dependiente.

En la siguiente figura se presentan varios ejemplos de diagramas de dispersión. En el primer ejemplo se aprecia que cuando aumenta el valor de la variable X, también aumenta el valor de la variable Y, además los puntos tienden a colocarse siguiendo un linea por lo que se dice que ambas variables tienen una **relación lineal positiva**. En el ejemplo 2, los puntos también se disponen de forma lineal pero esta vez cuando X aumenta, Y disminuye, por lo que se dice que las variables tienen una **relación lineal negativa**. La nueve de puntos del tercer ejemplo no muestra ninguna tendencia entre las dos variables. Decimos entonces que las **variables son independientes** y que el conocimiento de una de ellas no proporciona información sobre el valor de la otra. Por último, en el cuarto ejemplo se observa que las variables están relacionadas pero esta **relación no es lineal**.

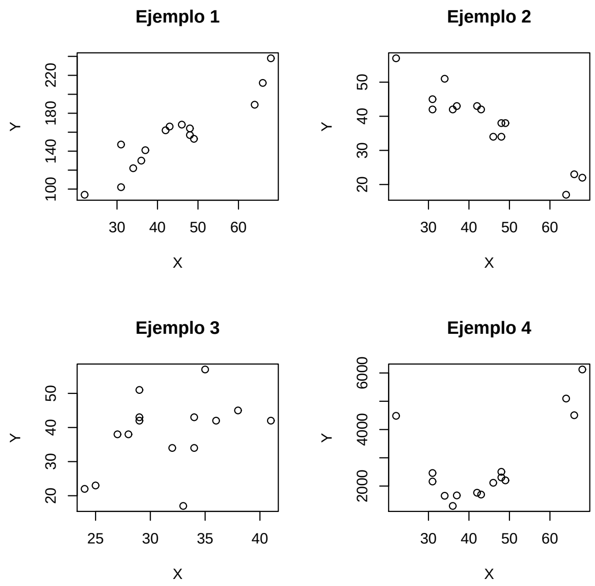


Figura 6.1: *Diferentes tipos de diagramas de dispersión. Ejm 1: Asociación lineal positiva; Ejm 2: Asociación lineal negativa; Ejm 3: Ausencia de asociación; Ejm 4: Asociación no lineal.*

Un diagrama de dispersión es capaz de caracterizar varios aspectos de las variables que lo componen, como por ejemplo el rango en que varían cada una de las variables, la posible asociación de los datos y una indicación de los casos atípicos.

6.2. Covarianza

La covarianza es una medida de asociación lineal entre dos variables que resume la información existente en un gráfico de dispersión. Mide el sentido de la relación lineal. Un valor positivo de la covarianza indica una relación lineal directa o creciente, es decir cuando los valores de una variable crecen los valores de la otra variable también crecen y viceversa. Un valor negativo indica una relación lineal inversa o decreciente, es decir si una variable crece la otra decrece y viceversa.

La **covarianza muestral** tiene la siguiente expresión:

$$cov(X,Y) = s_{XY} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1},$$

donde x_i e y_i son los valores observados de las variables X e Y, $\overline{x}$ e $\overline{y}$ son las medias muestrales y n es el tamaño de la muestra.

6.3. El coeficiente de correlación

El coeficiente de correlación muestral nos da una medida estandarizada de la relación lineal entre dos variables. Generalmente es una medida más útil, ya que indica tanto el sentido como el grado de relación. La covarianza y el coeficiente de correlación correspondiente tienen el mismo signo (ambos son positivos o ambos son negativos).

El **coeficiente de correlación muestral**, r, se calcula como:

$$r = \frac{s_{XY}}{s_X s_Y},$$

donde s_{XY} es la covarianza entre las dos variables y s_X y s_Y son las desviaciones típicas muestrales de X e Y, respectivamente.

Una regla práctica útil es considerar que existe una relación lineal si $|r| \geq \frac{2}{\sqrt{n}}$

El coeficiente de correlación varía entre -1 y +1. Cuanto más cerca se encuentra r de +1 más cerca se encuentran los datos de alinearse sobre una recta ascendente que indica una relación lineal positiva. Cuanto más cerca se encuentra r de -1 más cerca se encuentran los datos de alinearse sobre una recta descendente que indica una relación lineal negativa. Cuando r=0, no existe ninguna relación lineal entre X e Y, pero eso no quiere decir necesariamente que no exista ningún tipo de relación entre las variables. r sólo detecta asociaciones lineales.

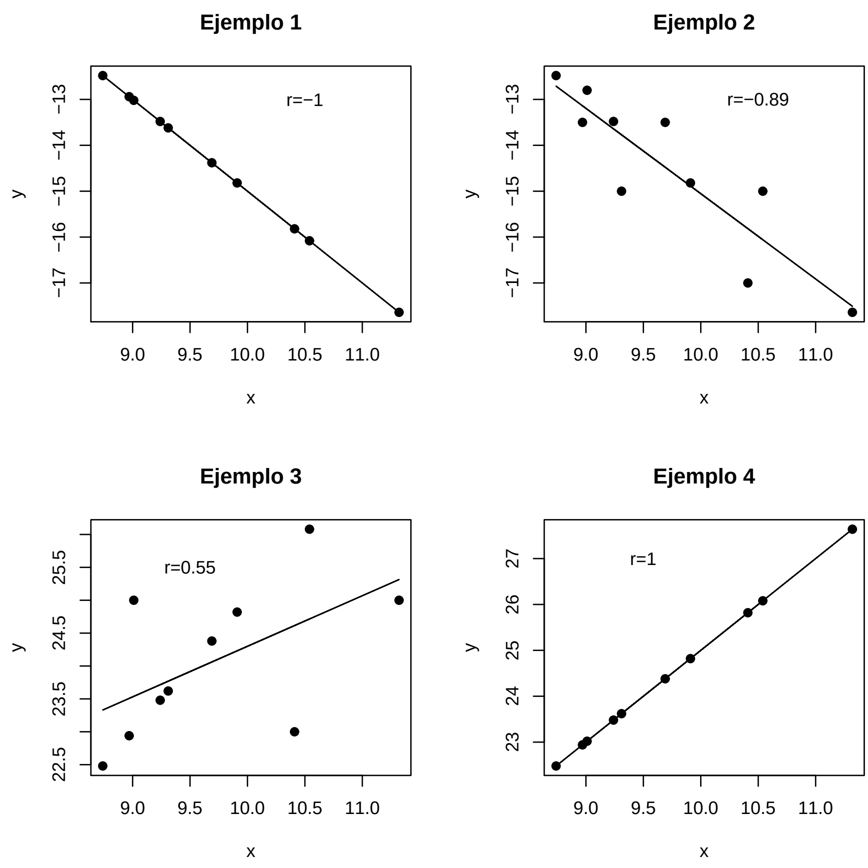


Figura 6.2: *Diferentes relaciones entre dos variables junto con sus coeficientes de correlación.*

6.4. El coeficiente de determinación

El **coeficiente de determinación** es una medida de la bondad de ajuste de un modelo lineal. Se calcula como el cuadrado del coeficiente de correlación, r^2. Por lo tanto, se puede interpretar como una proporción, ya que varía entre 0 y 1 (o equivalentemente en tantos por cien $r^2 * 100\,\%$). Cuantifica qué proporción o porcentaje de variabilidad es capaz de explicar la variable X de la variable Y.

Ejemplo 6.1.

Se desea investigar si existe relación entre la Edad (X) y la Talla (Y) de los bebés de entre 3 y 9 meses. Para ello, se toma una muestra aleatoria de 14 bebés. En la tabla se muestran los resultados obtenidos. Analiza brevemente la relación entre las dos variables.

x_i	y_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
3	55	-3	9	-13.3	176.5	39.9
6	68	0	0	-0.3	0.1	0.0
5	64	-1	1	-4.3	18.4	4.3
5	66	-1	1	-2.3	5.2	2.3
3	62	-3	9	-6.3	39.5	18.9
4	65	-2	4	-3.3	10.8	6.6
9	74	3	9	5.7	32.7	17.1
8	75	2	4	6.7	45.1	13.4
9	73	3	9	4.7	22.2	14.1
7	69	1	1	0.7	0.5	0.7
6	73	0	0	4.7	22.2	0.0
5	68	-1	1	-0.3	0.1	0.3
8	73	2	4	4.7	22.2	9.4
6	71	0	0	2.7	7.4	0.0
84	956		52		402.88	127

SI NO SE DISPONE DE CALCULADORA CON FUNCIONES ESTADÍSTICAS:

$s_{XY} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1} = \frac{127}{13} = 9{,}77$ (asociación positiva)

EN CASO DE DISPONER DE CALCULADORA CON FUNCIONES ESTADÍSTICAS:

$s_{XY} = r * s_X * s_Y$
$r = \frac{s_{XY}}{s_X s_Y} = \frac{9{,}77}{2*5{,}57} = 0{,}88$

$|0,88| \geq \frac{2}{\sqrt{14}} \cong 0,53$

Llegamos a la conclusión de que existe una estrecha relación lineal positiva entre la Edad y la Talla de los bebés.
$r^2 * 100 = (0,88)^2 * 100 = 77{,}44\,\%$ Es decir, la Edad de los bebés es capaz de explicar el 77.44 % de la variabilidad de la Talla de los bebés. Existe casi un 23 % de la variabilidad de la Talla que estaría explicada por otros factores que no han sido tenidos en cuenta, como por ejemplo: factores genéticos, la nutrición, etc.

6.5. La recta de regresión

En este apartado estudiaremos el ajuste de un modelo de regresión adecuado a la relación lineal entre dos variables continuas: la recta de regresión. Parece natural medir la dependencia entre dos variables describiendo cómo varía la media de la variable dependiente en función de la variable independiente.

Cálculo de la recta de regresión.

Llamaremos y a la variable dependiente y x a la variable independiente. La recta de regresión indica el valor medio de y para cada valor observado de x. En su cálculo supondremos que, aunque las medias de las distribuciones condicionadas no se localicen exactamente en una recta, lo estarían si tomásemos una población de datos más amplia. La recta de regresión va a darnos unos valores estimados o aproximados de estas medias condicionadas a partir de los datos. Llamaremos $\widehat{y}(x)$ a las estimaciones de las medias condicionadas que vamos a obtener con la recta, para indicar que dependen del valor particular de x.

El procedimiento de cálculo de la recta de regresión a partir de un conjunto de datos es el siguiente. Representaremos la ecuación de la recta por

$$\widehat{y} = a + bx$$

Esta ecuación depende de dos constantes, a y b, que deben estimarse a partir de los datos observados. **El parámetro** b es el más importante. Se denomina pendiente de la recta y nos dice cuánto aumenta de media la variable dependiente cuando la independiente aumenta en una unidad.

El **parámetro** a se denomina ordenada en el origen y representa el valor medio de la variable dependiente cuando la independiente es cero. En muchos problemas no tiene sentido considerar un valor cero para x: por ejemplo, es absurdo preguntarse cuál es el valor de la estatura cuando el peso es cero. Entonces a debe tomarse como un valor de referencia necesario para describir la recta, pero sin pretende atribuirle una interpretación lógica en el problema. Sólo tiene interpretación lógica si la variable X puede tomar el valor 0. Es decir, la nube de puntos se representa entorno al origen de coordenadas.

Para hallar a y b a partir de los datos se define el error de predicción o residuo de la recta, que es el error al predecir cada uno de los valores observados de la variable dependiente y.

residuo = error de predicción = valor observado - valor de la recta.

La recta se calcula imponiendo la condición de que el error promedio, definido como la raíz cuadrada de la suma de los cuadrados de los errores al prever cada punto con la recta, sea mínimo. Este criterio se denomina **mínimos cuadrados**.

$$Min\left(\sqrt{\frac{\sum_{i=1}^{n}[y_i - (bx_i + a)]^2}{n}}\right)$$

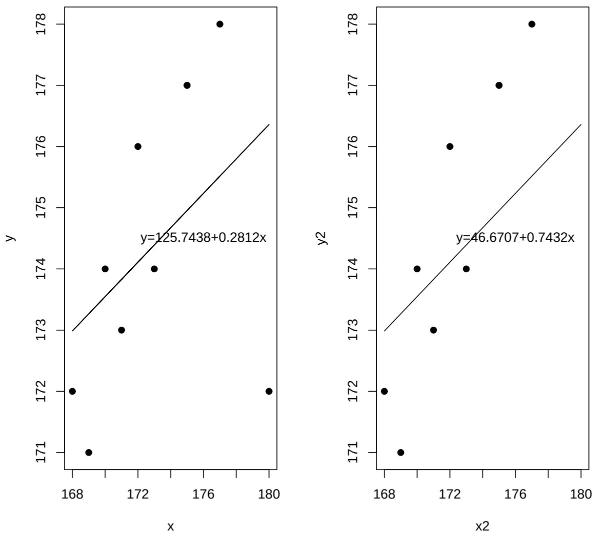


Figura 6.3: *Dos nubes de puntos que difieren en un punto .*

Como resultado de aplicar este criterio se obtiene que la pendiente b viene dada por la ecuación

$$b = \frac{s_{XY}}{s_x^2}$$

es decir, la pendiente de la recta resulta de estandarizar la covarianza de manera que tenga las unidades apropiadas de aumento en y por unidad de aumento x. Observemos que si la covarianza es cero también lo será la pendiente de la recta, que es la medida básica de la asociación, ya que si $b = 0$, no hay relación entre las variables. Es importante resaltar que la pendiente está muy relacionada con el coeficiente de correlación ya que se verifica que

$$b = r\frac{s_y}{s_x}$$

El segundo parámetro, a, ordenada en el origen de la recta, indica su posición concreta y se calcula mediante

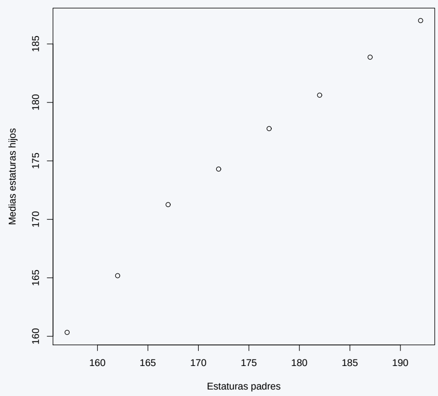
$$a = \overline{y} - b\overline{x}$$

ecuación que indica que la recta debe pasar por el punto $(\overline{x},\overline{y})$, es decir, por el centro de los datos.

Ejemplo 6.2.

Calculemos la recta de regresión a partir de las estaturas de 10 padres y la de sus hijos respectivamente, expresadas en centímetros.

Padre (X)	170	177	168	175	180	175	169	172	171	173
Hijo (Y)	174	178	172	177	178	177	171	176	173	174



$\overline{x}$=173 cm. $\overline{y}$=175 cm.
Si restamos a cada variable su media, obtenemos la siguiente tabla:

$(x_i - \overline{x})$	-3	4	-5	2	7	2	-4	-1	-2	0
$(y_i - \overline{y})$	-1	3	-3	2	3	2	-4	-1	-2	-1

s_{xy}=cov(x,y)=$\frac{(-3)(-1)+...+(0)(-1)}{9}$=8.67 cm^2
s_x=$\sqrt{\frac{(-3)^2+(-4)^2+...+(-2)^2+(0)^2}{9}}$=3.771 cm
s_y=$\sqrt{\frac{(-1)^2+(3)^2+...+(-2)^2+(-1)^2}{9}}$=2.539 cm
$r = \frac{cov(x,y)}{s_x \cdot s_y} = \frac{8{,}67}{3{,}771 \cdot 2{,}539}$=0.9. Concluimos que hay una fuerte relación lineal positiva entre las variables. Es decir, a mayor estatura del padre se espera mayor estatura del hijo. Por lo tanto, tiene sentido construir una recta de regresión para predecir valores de las estaturas de los hijos en función de las estaturas de los padres.
$r^2 100\,\% = 81\,\%$. Es decir, la estatura de los padres es capaz de explicar el 81 % de la variabilidad de la estatura de los hijos.
b=$\frac{cov(x,y)}{s_x^2} = \frac{8{,}67}{3{,}771^2}$= 0.6094
$a = \overline{y} - b\overline{x}$ = 175-0.6094·173= 69.5781
La recta de regresión es: $\widehat{y} = 69{,}5781 + 0{,}6094x$

Importante: La recta es una buena predictora únicamente en el rango de estaturas de padres [168,180]. Fuera de este rango no se pueden obtener conclusiones.

La predicción obtenida con estos datos de la estatura de un descendiente de un hombre de 175 cm es:
$\widehat{y} = 69{,}5781 + 0{,}6094 \cdot 175$=176.22 cm,
mientras que para un padre de 172 centímetros, esta estatura es
$\widehat{y} = 69{,}5781 + 0{,}6094 \cdot 172$=174.39 cm.

La desviación típica residual

La desviación típica residual es una medida de variabilidad de los puntos con respecto a la recta de regresión. Su expresión se calcula como

$$s_r = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n}}$$

Su interpretación es similar a la desviación típica: representa la variabilidad promedio de los datos observados con relación a la recta de regresión.

Ejemplo 6.3.

(continuación)

La desviación típica residual se calculará a partir de la siguiente tabla:

X	Y	$\widehat{Y}$	$(Y - \widehat{Y})^2$
170	174	173.1719	0.6858
177	178	177.4375	0.3164
168	172	171.9531	0.0022
175	177	176.2188	0.6104
180	178	179.2656	1.6018
175	177	176.2188	0.6104
169	171	172.5625	2.4414
172	176	174.3906	2.5901
171	173	173.7813	0.6104
173	174	175.0000	1.0000

$$s_r = \sqrt{\frac{0{,}6858 + 0{,}3164 + 0{,}0022 + 0{,}6104 + 1{,}6018 + 0{,}6104 + 2{,}4414 + 2{,}5901 + 0{,}6104 + 1{,}0000}{10}}$$

$= 1{,}02 cm.$

Así, el error promedio al estimar la estatura de un hijo en función de la de su padre, con la recta de regresión, es de aproximadamente 1 cm.

6.6. Reflexiones

Correlación y heterogeneidad.

Cuando se estudia la relación entre dos variables es importante asegurarse de que los elementos estudiados son homogéneos respecto a dichas variables. Por ejemplo, la Figura 6.4 presenta dos casos frecuentes de heterogeneidad. En el **Ejemplo 1** hay un **dato atípico** o discordante con el resto, que modifica el signo de la correlación. Puede comprobarse que si el punto con coordenadas (6,1), del ejemplo 1, no existiese, el coeficiente de correlación sería positivo, mientras que su presencia hace la correlación negativa. Ante una situación de este estilo conviene asegurarse de que:
1. No se ha cometido un error de medida o de transcripción del dato.
2. El elemento de la población al que le corresponde el dato atípico es homogéneo con respecto a los demás.

Si existe un error de medida conviene eliminar el dato y si el elemento es distinto de los demás por alguna razón objetiva conviene también suprimirlo del cálculo del coeficiente de correlación.

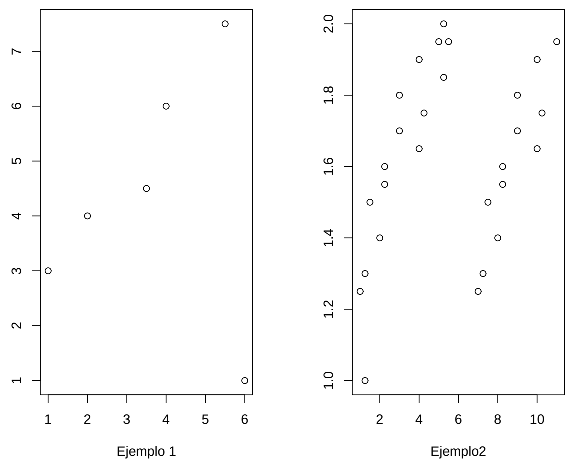


Figura 6.4: *Dos casos frecuentes de heterogeneidad.*

En el **Ejemplo 2** de la Figura 6.4 se muestra otro caso de heterogeneidad. En este caso el gráfico indica que la relación entre las variables es distinta para los elementos de las dos zonas de puntos y si calculamos un coeficiente de correlación para todos los datos obtendremos un valor muy pequeño. Sin embargo, si obtenemos los coeficientes para los dos grupos de puntos separadamente, encontraremos que dentro de cada grupo hay una relación fuerte.

La conclusión fundamental de este análisis es que un coeficiente de correlación es el resumen de la relación presente en el gráfico de dispersión. Conviene, pues, asegurarse mirando este gráfico que el coeficiente es un buen resumen del mismo. Tratar de interpretar un coeficiente de correlación sin haber visto previamente el gráfico de las variables puede ser peligroso.

Correlación y causalidad

Hemos visto que un coeficiente de correlación alto entre dos variables índica que los elementos observados toman valores relacionados entre sí, pero no permite concluir la existencia de ninguna relación de causalidad de una variable respecto de otra. Por ejemplo, si consideramos el número de matrimonios mensuales en una ciudad y la temperatura media mensual registrada, el coeficiente de correlación entre ambas variables es muy alto. Sin embargo, es obvio que no existe una relación causal entre ambas variables: ni es probable que un aumento de los matrimonios eleve la temperatura media del mes, ni es esperable que una ola de calor cause una avalancha de matrimonios. La razón del alto coeficiente de correlación es que los matrimonios tienden a producirse en verano, ya que de esta manera las parejas pueden aprovechar sus vacaciones para tener más tiempo en el inicio de su vida en común.

Este tipo de correlaciones se denomina **correlaciones espurias** y se deben al efecto de otra variable (las vacaciones veraniegas) que al tener una relación de dependencia con las variables que observamos (matrimonios y temperatura) crea la relación entre ellas.

Por otra parte, si no encontramos correlación entre dos variables tampoco podemos deducir que no exista relación lineal entre ellas. En primer lugar, si las observaciones tienen un rango de variación pequeño y existen otros factores que producen variabilidad es posible que no veamos causalidad aunque exista. Por ejemplo, consideremos la relación entre el tamaño de un apartamento y su precio de alquiler en un conjunto de pisos de una gran ciudad. Supongamos que tomamos una muestra de pisos entre 80 y 100 m^2. Como el tamaño de estos pisos es bastante homogéneo, su precio dependerá de su localización y de otros factores. En consecuencia, el coeficiente de correlación ente precio y tamaño será pequeño. Sin embargo, no podemos concluir que el tamaño del piso no influye en su precio. Si tomásemos una muestra que incluyese pisos pequeños y grandes la relación entre estas dos variables aparecería claramente.

Resumiendo, la dependencia que existe entre dos variables puede deducirse de un gráfico de dispersión, que es la herramienta más completa para juzgar la relación entre ellas. Una medida resumen de este gráfico, que es útil para medir la relación lineal entre dos variables, es la covarianza, que se expresa en unidades que son el producto de la variables. Al convertirla en una medida adimensional de la relación lineal, que no depende de las unidades de medida, se obtiene el coeficiente de correlación, que puede interpretarse como la varianza entre variables estandarizadas.

6.7. Seminario de REGRESIÓN LINEAL

1. Indicar cuál de las dos variables en los pares siguiente es la dependiente y si la relación esperada es positiva o negativa.
   a)potencia de un coche y precio.
   b)calorías que consume una persona y su peso.
   c)consumo de tabaco y esperanza de vida.

2. Un estudiante que busca piso ha tomado los siguientes datos de los precios de alquiler semanal y de la superficie de los pisos en metros cuadrados.

superficie	60	60	80	90	100	110	120
precio	70	85	80	90	85	110	115

   a) Estudia si existe relación lineal entre la superficie de los pisos y su precio de alquiler, explicando el grado y el sentido de asociación, de manera justificada mediante el coeficiente de correlación.

   b) Calcula e interpreta el coeficiente de determinación.

   c) Obtén la recta de regresión que modeliza el precio de alquiler en función de la superficie e interpreta, en el contexto del ejercicio, los coeficientes de la recta.

   d) Estima el error que se espera cometer cuando se predice el precio de alquiler con la recta de regresión obtenida en el apartado anterior.

3. Una encuesta de salarios entre graduados proporciona los datos siguiente:

Edad	28	28	32	35	38	44	49	52	58	62	66	70
Salario	2.2	2.2	3.8	4.2	4.2	5.3	7.3	6.4	6.7	5.3	6.0	5.1

   a) Estudia si existe relación lineal entre el salario percibido y la edad, explicando el grado y el sentido de asociación, de manera justificada mediante el coeficiente de correlación.

   b) Calcula e interpreta el coeficiente de determinación.

   c) Obtén la recta de regresión que modeliza el salario en función de la edad e interpreta, en el contexto del ejercicio, los coeficientes de la recta.

   d) Estima el error que se espera cometer cuando se predice el salario con la recta de regresión obtenida en el apartado anterior.

4. El estradiol es una hormona esteroide sexual femenina. Se desea estudiar si el nivel en sangre de estradiol tiene relación lineal con la edad de las mujeres, con el objetivo de predecir y modificar su nivel farmacológicamente en edades que lo necesiten. Para ello, se considera una muestra de 10 mujeres de las que se ha tomado su edad (en años) y su nivel de estradiol (en pg/ml):

Edad	14.3	21.2	25.7	35.2	38.2	41.8	47.2	51.3	54.5	62.7
Estradiol	193.7	195.2	185.3	152.7	120.7	88.3	75.2	47.5	25.1	24.2

   a) Estudia razonadamente si el nivel de estradiol femenino se puede explicar linealmente en función de la edad indicando el grado y sentido de asociación.

   b) Calcula e interpreta el coeficiente de determinación.

   c) Calcula la recta regresión que mejor ajusta los datos e interpreta sus coeficientes en el contexto del ejercicio.

   d) ¿Qué nivel medio de estradiol se espera en una mujer con 32 años? ¿Y en una mujer con 72 años? ¿Son fiables los niveles medios obtenidos?

e) Está recomendado tratar farmacológicamente a mujeres con niveles de estradiol menores o iguales a 35 pg/ml. A la vista del estudio, ¿a partir de qué edad es recomendable iniciar el tratamiento?

f) Estima el error de predicción que se comete cuando se utiliza la recta de regresión del apartado b).

(Algunos ejercicios que han aparecido en exámenes)

5. (2011) La siguiente tabla contiene la tensión arterial sistólica (TAS), medida en mm de Hg y la hemoglobina glicosilada (HBA_1), expresada en %, de una muestra de 5 pacientes.

TAS	150	120	145	140	130
HBA_1	8.2	6.1	7.2	7	6

a) Estudia si existe relación lineal entre la tensión arterial sistólica y la hemoglobina glicosilada, explicando el grado y el sentido de asociación, de manera justificada mediante el coeficiente de correlación.

b) Calcula e interpreta el coeficiente de determinación.

c) Obtén la recta de regresión que modeliza la tensión arterial sistólica en función de la hemoglobina glicosilada e interpreta, en el contexto del ejercicio, los coeficientes de la recta.

d) Estima el error que se espera cometer cuando se predice la tensión arterial sistólica con la recta de regresión obtenida en el apartado anterior.

6. (2012) Se desea investigar si el peso de las personas tiene influencia lineal sobre el colesterol LDL.

LDL (mg/dl)	131	143	178	189	121	99
Peso (Kg)	67	71.24	89.56	92.50	81.70	65.80

a) Estudia si existe relación lineal entre el colesterol y el peso, explicando el grado y el sentido de asociación, de manera justificada mediante el coeficiente de correlación.

b) Calcula e interpreta el coeficiente de determinación.

c) Obtén la recta de regresión que modeliza el colesterol en función del peso e interpreta, en el contexto del ejercicio, los coeficientes de la recta.

d) Estima el error que se espera cometer cuando se predice el colesterol con la recta de regresión obtenida en el apartado anterior.

7. **Problema de Técnicas Analíticas (2º Grado en Farmacia)**

La determinación de metales pesados puede llevarse a cabo polarográficamente por medida de la intensidad de corriente originada en su reducción con un electrodo de gotas de mercurio. Para determinar cadmio en una muestra de aguas residuales, se obtuvo una corriente de 41 mA, operando en determinadas condiciones. El calibrado se obtuvo operando en las mismas condiciones sobre patrones de cadmio de concentración conocida, obteniéndose los siguientes valores:

$[Cd^2](M)$	$1\ 10^{-4}$	$2\ 10^{-4}$	$3\ 10^{-4}$	$4\ 10^{-4}$	$5\ 10^{-4}$	$6\ 10^{-4}$	$8\ 10^{-4}$	$10\ 10^{-4}$
I (mA)	9	20	29	39	41	56	73	87

Obtén la concentración de cadmio en la muestra problema.

8. **Problema de Sistemas de Calidad de Laboratorio (2º Grado en Farmacia)**

CONVERSIÓN DE UNA DISTRIBUCIÓN CURVA (CINÉTICA EXPONENCIAL O CINÉTICA DE ORDEN UNO) EN UNA RECTA Y DETERMINACIÓN DE LOS PARÁMETROS DE LA ECUACIÓN MEDIANTE REGRESIÓN LINEAL POR MINIMOS CUADRADOS

Etamsilato es un fármaco antihemorrágico que se utiliza como tratamiento preventivo previo a algunas intervenciones quirúrgicas tanto en cirugía humana como veterinaria. La tabla que se adjunta corresponde a los valores de cantidad de fármaco en plasma a distintos tiempos tras ser administrado por vía intravenosa (i.v.) a un ternero de 70 kg de peso a la dosis de 7.5 mg/kg de peso. ¿Cuánto tiempo tardará en desaparecer el 95 % de la dosis de fármaco del plasma?

NOTA: La ecuación que relaciona la concentración con el tiempo es una curva exponencial que se puede transformar en una recta semilogarítmica (ln Q vs t). $Q = Q_0 e^{-kt} \rightarrow ln(Q) = ln(Q_0) - kt$

Tiempo (h)	0.5	1	2	3	4	5	6	7	8
Concentración (mg)	562.32	420.75	145.86	70.95	35.31	20.46	12.54	11.55	5.94

6.8. Soluciones del Seminario de regresión lineal

1.

	V. Dependiente	V. Independiente	Relación
a)	Precio	Potencia del coche	positiva
b)	Peso	Calorías	positiva
c)	Duración de la vida	Consumo de tabaco	negativa

2a) r= 0.8613
2b) $r^2 100\,\% = 74{,}18\,\%$
2c)$\widehat{y} = 37{,}957 + 0{,}596X$
2d) s_r= 7.61
3a) r= 0.7293
3b) $r^2 100\,\% = 53{,}19\,\%$
3c) $\widehat{y} = 1{,}1349 + 0{,}0802X$
3d) s_r= 1.073
4a) r= -0.97
4b) $r^2 100\,\% = 94{,}09\,\%$
4c) $\widehat{y} = 279{,}23 - 4{,}296X$
4d) 141.76; -30.071
4e) 56.852
4f) 15.51
5a) r= 0.9112
5b) $r^2 100\,\% = 83{,}03\,\%$
5c)$\widehat{y} = 52{,}88 + 12{,}19X$
5d) s_r= 4.44
6a) r= 0.832
6b) $r^2 100\,\% = 69{,}22\,\%$
6c) $\widehat{y} = -48{,}68 + 2{,}46X$
6d) s_r= 17.39
7) $4{,}5\ 10^{-4}$
8) 4.96 h

Apéndice A

Tablas de derivadas, integrales. Ecuaciones diferenciales.

Tabla de derivadas

Función	Derivada
y= k	y'= 0
y= x	y'= 1
y= u^m	y'= $m\ u^{m-1}\ u'$
y= e^u	y'= $u'e^u$
y= a^u	y'= $u'a^uL(a)$
y= $L(u)$	y'= $\frac{u'}{u}$
y= $log_a(u)$	y'= $\frac{u'}{u}log_a(e)$
y= $sen(u)$	y'= $u'cos(u)$
y= $cos(u)$	y'= $-u'sen(u)$
y= $tg(u)$	y'= $u'sec^2(u)$
y= $cotg(u)$	y'= $-u'cosec^2(u)$
y= $sec(u)$	y'= $u'sec(u)tg(u)$
y= $cosec(u)$	y'= $-u'cosec(u)cotg(u)$
y= $arcsen(u)$	y'= $\frac{u'}{\sqrt{1-u^2}}$
y= $arccos(u)$	y'= $\frac{-u'}{\sqrt{1-u^2}}$
y= $arctg(u)$	y'= $\frac{u'}{1+u^2}$
y= ku	y'= ku'
y= $u+v-w$	y'= $u'+v'-w'$
y= uv	y'= $u'v+uv'$
y= $\frac{u}{v}$	y'= $\frac{u'v-uv'}{v^2}$

Tabla de integrales

Funciones simples	Funciones compuestas
$\int dx = x + C$	
$\int kdx = kx + C$	
$\int x^n dx = \frac{x^{n+1}}{n+1} + C$	$\int u^n \cdot u' dx = \frac{u^{n+1}}{n+1} + C$
$\int \frac{1}{x} dx = Ln\|x\| + C$ (*)	$\int \frac{u'}{u} dx = Ln\|u\| + C$
$\int e^x dx = e^x + C$	$\int e^u \cdot u' dx = e^u + C$
$\int a^x dx = \frac{a^x}{Ln(a)} + C$	$\int a^u \cdot u' \cdot dx = \frac{a^u}{Ln(a)} + C$
$\int cos(x) dx = sen(x) + C$	$\int cos(u) \cdot u' \cdot dx = sen(u) + C$
$\int sen(x) dx = -cos(x) + C$	$\int sen(u) \cdot u' \cdot dx = -cos(u) + C$
$\int \frac{1}{cos^2(x)} dx = tg(x) + C$	$\int \frac{u'}{cos^2(u)} dx = tg(u) + C$
$\int (1 + tg^2(x)) dx = tg(x) + C$	$\int (1 + tg^2(u)) \cdot u' \cdot dx = tg(u) + C$
$\int \frac{-1}{sen^2(x)} dx = cotg(x) + C$	$\int \frac{-u'}{sen^2(u)} dx = cotg(u) + C$
$\int \frac{1}{1+x^2} dx = arctg(x) + C$	$\int \frac{u'}{1+u^2} dx = arctg(u) + C$
$\int \frac{-1}{1+x^2} dx = arccotg(x) + C$	$\int \frac{-u'}{1+u^2} dx = arccotg(u) + C$
$\int \frac{1}{\sqrt{1-x^2}} dx = arcsen(x) + C$	$\int \frac{u'}{\sqrt{1-u^2}} dx = arcsen(u) + C$
$\int \frac{-1}{\sqrt{1-x^2}} dx = arccos(x) + C$	$\int \frac{-u'}{\sqrt{1-u^2}} dx = arccos(u) + C$

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

$$\int f(x) * g(x) dx \neq \int f(x) dx * \int g(x) dx$$

$$\int \frac{f(x)}{g(x)} dx \neq \frac{\int f(x) dx}{\int g(x) dx}$$

Integración por partes

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

Ecuaciones diferenciales

Ec. diferencial de variables separables:

$M(x) + N(y)y' = 0$

Ec. diferencial lineal de primer orden:

$y' + P(x)y = Q(x)$

$y = \frac{1}{u(x)}[\int Q(x)u(x)dx + C]$ donde u(x)=$e^{\int P(x)dx}$

Apéndice B

Tablas y formulario de Estadística

Tabla de probabilidades de la distribución N(0,1)
$[P(Z < z)]$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5039	0.5079	0.5119	0.5159	0.5199	0.5239	0.5279	0.5318	0.5358
0.1	0.5398	0.5437	0.5477	0.5517	0.5556	0.5596	0.5635	0.5674	0.5714	0.5753
0.2	0.5792	0.5831	0.5870	0.5909	0.5948	0.5987	0.6025	0.6064	0.6102	0.6140
0.3	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6405	0.6443	0.6480	0.6517
0.4	0.6554	0.6590	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6843	0.6879
0.5	0.6914	0.6949	0.6984	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7290	0.7323	0.7356	0.7389	0.7421	0.7453	0.7485	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7733	0.7763	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7938	0.7967	0.7995	0.8023	0.8051	0.8078	0.8105	0.8132
0.9	0.8159	0.8185	0.8212	0.8238	0.8263	0.8289	0.8314	0.8339	0.8364	0.8389
1.0	0.8413	0.8437	0.8461	0.8484	0.8508	0.8531	0.8554	0.8576	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8707	0.8728	0.8749	0.8769	0.8790	0.8810	0.8829
1.2	0.8849	0.8868	0.8887	0.8906	0.8925	0.8943	0.8961	0.8979	0.8997	0.9014
1.3	0.9032	0.9049	0.9065	0.9082	0.9098	0.9114	0.9130	0.9146	0.9162	0.9177
1.4	0.9192	0.9207	0.9221	0.9236	0.9250	0.9264	0.9278	0.9292	0.9305	0.9318
1.5	0.9331	0.9344	0.9357	0.9369	0.9382	0.9394	0.9406	0.9417	0.9429	0.9440
1.6	0.9452	0.9463	0.9473	0.9484	0.9494	0.9505	0.9515	0.9525	0.9535	0.9544
1.7	0.9554	0.9563	0.9572	0.9581	0.9590	0.9599	0.9607	0.9616	0.9624	0.9632
1.8	0.9640	0.9648	0.9656	0.9663	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.9	0.9712	0.9719	0.9725	0.9731	0.9738	0.9744	0.9750	0.9755	0.9761	0.9767
2.0	0.9772	0.9777	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9816
2.1	0.9821	0.9825	0.9829	0.9834	0.9838	0.9842	0.9846	0.9849	0.9853	0.9857
2.2	0.9860	0.9864	0.9867	0.9871	0.9874	0.9877	0.9880	0.9883	0.9886	0.9889
2.3	0.9892	0.9895	0.9898	0.9900	0.9903	0.9906	0.9908	0.9911	0.9913	0.9915
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9937	0.9939	0.9941	0.9942	0.9944	0.9946	0.9947	0.9949	0.9950	0.9952
2.6	0.9953	0.9954	0.9956	0.9957	0.9958	0.9959	0.9960	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9975	0.9976	0.9977	0.9978	0.9978	0.9979	0.9980	0.9980
2.9	0.9981	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.0	0.9986	0.9986	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9989	0.9989

Tabla de probabilidades de la distribución *t de Student*
$[P(t < T)]$

$g \diagdown T$	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.9750	0.990	0.995
1	0.509	0.726	1.000	1.376	1.962	3.077	6.313	12.706	31.820	63.656
2	0.444	0.617	0.816	1.060	1.386	1.885	2.919	4.302	6.964	9.924
3	0.424	0.584	0.764	0.978	1.249	1.637	2.353	3.182	4.540	5.840
4	0.414	0.568	0.740	0.940	1.189	1.533	2.131	2.776	3.746	4.604
5	0.408	0.559	0.726	0.919	1.155	1.475	2.015	2.570	3.364	4.032
6	0.404	0.553	0.717	0.905	1.134	1.439	1.943	2.446	3.142	3.707
7	0.401	0.549	0.711	0.896	1.119	1.414	1.894	2.364	2.997	3.499
8	0.399	0.545	0.706	0.888	1.108	1.396	1.859	2.306	2.896	3.355
9	0.397	0.543	0.702	0.883	1.099	1.383	1.833	2.262	2.821	3.249
10	0.396	0.541	0.699	0.879	1.093	1.372	1.812	2.228	2.763	3.169
11	0.395	0.539	0.697	0.875	1.087	1.363	1.795	2.200	2.718	3.105
12	0.394	0.538	0.695	0.872	1.083	1.356	1.782	2.178	2.680	3.054
13	0.393	0.537	0.693	0.870	1.079	1.350	1.770	2.160	2.650	3.012
14	0.393	0.536	0.692	0.868	1.076	1.345	1.761	2.144	2.624	2.976
15	0.392	0.535	0.691	0.866	1.073	1.340	1.753	2.131	2.602	2.946
16	0.392	0.535	0.690	0.864	1.071	1.336	1.745	2.119	2.583	2.920
17	0.391	0.534	0.689	0.863	1.069	1.333	1.739	2.109	2.566	2.898
18	0.391	0.533	0.688	0.862	1.067	1.330	1.734	2.100	2.552	2.878
19	0.391	0.533	0.687	0.860	1.065	1.327	1.729	2.093	2.539	2.860
20	0.390	0.532	0.686	0.859	1.064	1.325	1.724	2.085	2.527	2.845
21	0.390	0.532	0.686	0.859	1.062	1.323	1.720	2.079	2.517	2.831
22	0.390	0.532	0.685	0.858	1.061	1.321	1.717	2.073	2.508	2.818
23	0.390	0.531	0.685	0.857	1.060	1.319	1.713	2.068	2.499	2.807
24	0.389	0.531	0.684	0.856	1.059	1.317	1.710	2.063	2.492	2.796
25	0.389	0.531	0.684	0.856	1.058	1.316	1.708	2.059	2.485	2.787
26	0.389	0.530	0.684	0.855	1.057	1.314	1.705	2.055	2.478	2.778
27	0.389	0.530	0.683	0.855	1.056	1.313	1.703	2.051	2.472	2.770
28	0.389	0.530	0.683	0.854	1.055	1.312	1.701	2.048	2.467	2.763
29	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.389	0.530	0.682	0.853	1.054	1.310	1.697	2.042	2.457	2.749
40	0.388	0.528	0.680	0.850	1.050	1.303	1.683	2.021	2.423	2.704
60	0.387	0.527	0.678	0.847	1.045	1.295	1.670	2.000	2.390	2.660
120	0.386	0.525	0.676	0.844	1.040	1.288	1.657	1.979	2.357	2.617
∞	0.385	0.524	0.674	0.841	1.036	1.281	1.644	1.959	2.326	2.575

Estadística descriptiva

Percentil P_x $Pos = (n+1)\cdot\frac{x}{100}$ $P_x = deci(Pos)\cdot X_{[Pos]+1} + (1-deci(Pos))\cdot X_{[Pos]}$	**Media** $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}$
Rango $Rango = Máximo - Mínimo$	**Rango intercuartílico** $R.I.C. = Q_3 - Q_1 = P_{75} - P_{25}$
Varianza $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$	**Desviación típica** $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$
Valores no atípicos $[Q_1 - 1{,}5\cdot R.I.C.,\quad Q_3 + 1{,}5\cdot R.I.C.]$	

Aritmética de variables normales

$X \sim N(\mu,\sigma) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
$Z \sim N(0,1) \Rightarrow X = \sigma\cdot Z + \mu \sim N(\mu,\sigma)$

Intervalos de confianza

Una media   a) $\left[\overline{x} - Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x} + Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right]$; b) $\left[\overline{x} - t_{(n-1,1-\frac{\alpha}{2})}\cdot\frac{S}{\sqrt{n}}\ ,\ \overline{x} + t_{(n-1,1-\frac{\alpha}{2})}\cdot\frac{S}{\sqrt{n}}\right]$
Un porcentaje $\left[\widehat{P} - Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\widehat{P}\cdot(100-\widehat{P})}{n}}\ ,\ \widehat{P} + Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\widehat{P}\cdot(100-\widehat{P})}{n}}\right]$

Regresión lineal

Covarianza $cov(X,Y) = s_{XY} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$
Coeficiente de correlación $r = \frac{s_{XY}}{s_X s_Y}$; hay asociación si $\lvert r\rvert \geq \frac{2}{\sqrt{n}}$
Coeficiente de determinación $r^2\cdot 100\,\%$
Ecuación de la recta de regresión $\widehat{y} = a + bx$; $b = \frac{s_{XY}}{s_X^2} = r\frac{s_Y}{s_X}$; $a = \overline{y} - b\overline{x}$;
Desviación típica residual $s_r = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}}$