

Tutorial

Introducción al análisis y visualización de datos utilizando una planilla de cálculo

Autor: Pablo Poggi (pmpoggi@gmail.com)

1. Introducción

Las planillas de cálculo son documentos compuestos de datos numéricos y alfanuméricos que se hayan dispuestos en tablas. Estos documentos pueden ser creados, editados y visualizados con distintos programas, los cuales permiten realizar operaciones matemáticas, crear tablas dinámicas, dibujar gráficos y muchas cosas más. Entre los programas más conocidos se encuentran el Microsoft Excel y el Open Calc (o LibreCalc, en su versión más reciente). Google también permite trabajar con planillas de cálculo online a través de Google Docs. En este tutorial se trabajará con Open Calc, pero la totalidad de los conceptos y funcionalidades que se describen son compartidas con el Excel.

La unidad básica que constituye una planilla de cálculo es la celda, que es cada uno de los lugares donde se pueden ingresar datos. En la Fig. 1.1 se observa el entorno de trabajo del programa Open Calc, en el cual se distingue el arreglo tabular de las celdas. Cada celda tiene una identificación, que está dada por la columna y fila donde se encuentra (como en una matriz). Es común en las planillas de cálculo que las columna estén nombradas con letras y las filas, con números. De esta manera, al referirnos a una celda lo haremos de la siguiente manera: A3, B5, C7, etc. (ver Fig. 1.1). En cada celda podremos ingresar *letras*, por ejemplo para el encabezado de una columna; *números*, por ejemplo datos o mediciones de un experimento, o también *fórmulas* que involucren otras celdas, que el programa interpretará mostrando resultados numéricos en dicha celda.

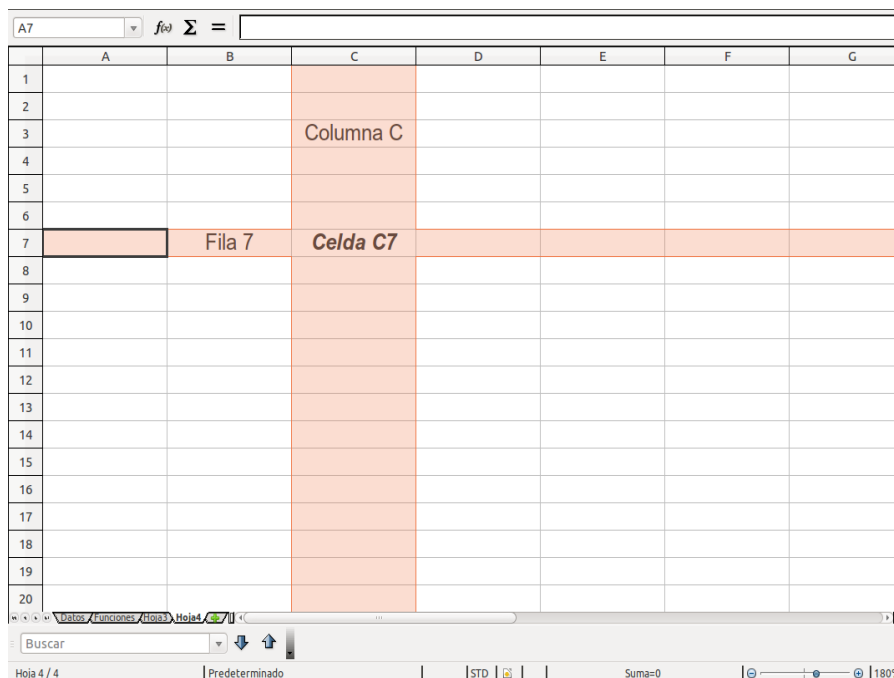


Figura 1.1

2. Operaciones básicas con celdas

Para ilustrar algunas de las funcionalidades básicas de la planilla de cálculo, vamos a usar un ejemplo. Vamos a confeccionar una tabla en la cual queremos volcar el dinero que gastamos en cada

uno de los días de una semana. Para ello vamos a usar en principio tres columnas: en la primera pondremos el día de la semana, en la segunda el número del día (de 1 a 7) y en la tercera, el dinero correspondiente a cada día. Para el llenado de las primeras dos columnas utilizaremos una de las funciones más útiles de la planilla de cálculo: el llenado automático (o autocompletado). Si ingresamos “Lunes” en la primera fila de la primera columna, podemos hacer que el programa rellene las filas siguientes con el resto de los días. Para ello, hay que seleccionar la celda, clickear y mantener sobre el pequeño cuadrado que se encuentra abajo y a la derecha de la misma y arrastrar hacia abajo (ver Fig. 2.1). El programa nos va mostrando con qué valores se rellenarán las celdas de las filas siguientes. Podemos hacer lo mismo para rellenar con los números del 1 al 7 en la columna siguiente. En la sección 4, se da otro ejemplo de llenado automático.

	A	B	C	D
1				
2		Día	Número	Gastos
3		Lunes		
4				
5				
6				
7				
8				
9				
10				
11				

	A	B	C	D
1				
2		Día	Número	Gastos
3		Lunes		
4		martes		
5		miércoles		
6		jueves		
7		viernes		
8		sábado		
9		domingo		
10				
11				

Figura 2.1

Luego de seguir los pasos descriptos anteriormente, podemos llenar la columna “Gastos” con valores a fin de completar la tabla, que ahora tendrá un aspecto similar al de la Fig. 2.2. Notar que los números de la última columna aparecen con cifras decimales y con el símbolo de peso; para lograr esto hay que establecer el Formato de la columna. El mismo se modifica a través del menú Formato → Celdas → Números. Allí es posible elegir entre diversas opciones de formato para los números ingresados en una o varias celdas.

En la Fig 2.2 se observa también que se ha agregado una fila en la cual se consigna el total de los gastos. Para ello, en la celda D10 se introdujo una fórmula, que es la suma de los valores de las siete celdas que se encuentran por encima. Para ello, se puede introducir la suma término a término:

$$=D3+D4+D5+D6+D7+D8+D9$$

o, como se hizo en el ejemplo de la Fig. 2.2, utilizar una de las *funciones* con las que viene OpenCalc (y también Excel):

$$=SUMA(D3:D9)$$

	A	B	C	D
1				
2		Día	Número	Gastos
3		Lunes	1	\$47,00
4		martes	2	\$68,00
5		miércoles	3	\$59,00
6		jueves	4	\$72,00
7		viernes	5	\$70,00
8		sábado	6	\$61,00
9		domingo	7	\$53,00
10			Total	\$430,00
11				

Figura 2.2

En primera instancia, notemos que ambas fórmulas comienzan con un signo de igualdad “=”. Cada vez que ingresemos ese símbolo en una celda, el programa interpretará que lo que ingresemos a continuación es una fórmula. Si logra interpretar dicha fórmula, mostrará en la celda el resultado, caso contrario nos devolverá un error. Como se observa en la Fig. 2.2, si bien el contenido de la celda es un número (el resultado), la fórmula original todavía se observa en la barra de fórmulas (rectángulo verde). Las fórmulas pueden ser cálculos simples entre celdas, como por ejemplo

$$=2*D4 + D5/D6$$

$$=D4-D6+D7/2$$

o funciones

$$=RAIZ(D4)$$

$$=SENO(D6)$$

Haciendo click en el ícono $f(x)$ (ver recuadro rojo en Fig. 2.2), se accede al listado de la enorme cantidad de funciones que provee el programa OpenCalc, junto con su descripción.

En segundo lugar, observemos que si queremos ingresar en una celda la función SUMA, al escribir

$$=SUMA($$

podemos elegir entre ingresar el rango a mano (p. ej. escribir D3:D10) o directamente *pintar* las celdas que nos interesa sumar con el mouse (luego deberemos cerrar el paréntesis). Ésta funcionalidad es típica de las planillas de cálculo: al introducir una fórmula, podemos seleccionar las celdas (en vez de ingresar su nombre) con las que queremos operar. Pruebe de realizar algunos

de los cálculos descriptos en el párrafo anterior, seleccionando las celdas a medida que deben ser ingresadas en la fórmula (esta funcionalidad se vuelve a tratar en el ejemplo de la Sección 4.1).


Por último, observemos que una vez ingresada una fórmula en una celda, el valor que muestre la misma se actualizará automáticamente en el caso de que alguna de las celdas que intervienen en la fórmula sea modificada.

3. Gráficos de datos

Al trabajar con tablas de datos, como la que se observa en la Fig. 2.2, resulta muchas veces provechoso incluir un gráfico que permita visualizarlos de una manera mas directa. La implementación de gráficos en las planillas de cálculo es muy sencilla, y cuenta además con una gran variedad de opciones de formato y presentación. Siguiendo con el ejemplo anterior, vamos a construir un gráfico de “Gasto” en función del “número de día”. El primer paso es pintar las celdas correspondientes a los datos que querramos visualizar en el gráfico; dichas celdas deben abarcar dos columnas y el programa interpretará que la primera corresponde a las abcisas y la segunda, a las ordenadas. En nuestro caso, seleccionaremos las celdas tal como se muestra en la Fig. 3.1 (encabezado incluido).

	A	B	C	D
1				
2		Día	Número	Gastos
3		Lunes	1	\$47,00
4		martes	2	\$68,00
5		miércoles	3	\$59,00
6		jueves	4	\$72,00
7		viernes	5	\$70,00
8		sábado	6	\$61,00
9		domingo	7	\$53,00
10			Total	\$430,00
11				

Figura 3.1

Luego hay que dirigirse al menú Insertar → Gráfico (o seleccionar el botón correspondiente en la barra de herramientas: ), lo cual abre una ventana que nos guía a través de varios pasos para elegir el formato deseado para el gráfico. En primaria instancia, se escoge el tipo de gráfico a realizar, siendo en este caso la opción “XY” (o “Dispersión XY”) la indicada. El resto de las opciones (por ejemplo, el gráfico tipo “torta”) son útiles para representar una sola serie (columna) de datos (haga la prueba). Para cada opción, el programa ofrece una serie de formatos para elegir. En el caso de los gráficos XY, se puede elegir graficar sólo puntos, o puntos y líneas, etc. En las ventanas subsiguientes, se puede modificar el rango de datos utilizado, y escoger los títulos de los ejes, entre otras opciones. Al finalizar, se inserta el gráfico como un objeto en el documento, tal como se observa en la Fig. 3.2.

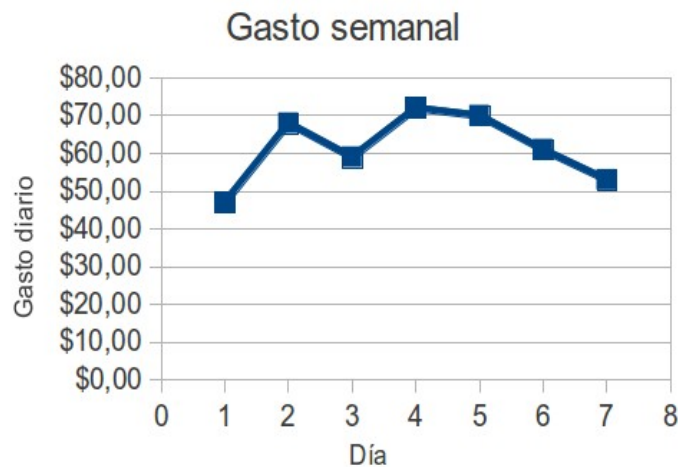


Figura 3.2

El gráfico puede seguir siendo editado una vez creado. Para ello hay que hacer doble click en cualquier lugar del gráfico y luego seleccionar que parte queremos editar (las líneas si queremos cambiar el color o ancho, los ejes si queremos cambiar la fuente o el rango, etc.). Notemos que el gráfico se actualiza automáticamente si modificamos los valores de las celdas que están representadas en él, tal como pasaba con las fórmulas.

4. Algunos ejemplos de aplicación

4.1. Gráficos de funciones

Muchas veces resulta útil tener la posibilidad de graficar no solo una serie de datos discretos, sino también una función dada. En las planillas de cálculo, el procedimiento es, formalmente, idéntico al caso anterior, ya que lo que le vamos a pedir al programa es que grafique una serie de puntos, pero con la salvedad de que en este caso nosotros vamos a fijar los valores de las abcisas y el programa se encargará de calcular las ordenadas.

Dada una función $f(x)$, vamos a ingresar al programa una serie de valores

$$x_1, x_2, x_3, \dots, x_N,$$

para luego calcular los correspondientes

$$f(x_1), f(x_2), f(x_3), \dots, f(x_N).$$

Finalmente, le vamos a pedir al programa que haga un gráfico de los puntos $(x_i, f(x_i))$. Veamos ahora como elegir e introducir los valores de las abcisas. Para ello debemos definir tres cosas: el rango del gráfico y el denominado *paso*. El rango representa los valores de x entre los cuales me interesa graficar la función (x_1 y x_N), y el paso es la diferencia entre dos valores consecutivos de x

$$\Delta x = x_{i+1} - x_i$$

Tanto el rango como el paso determinaran la cantidad de puntos (N) que tendrá nuestro gráfico; más aún, es fácil ver que

$$N = (x_N - x_1) / \Delta x,$$

con lo cual será necesario ajustar ambos parámetros para determinar la cantidad total de puntos del gráfico. Como queremos generar una curva continua, resulta natural que mientras más grande sea N , mejor será el gráfico. Ésto es verdad, pero no resulta cómodo trabajar con una cantidad excesivamente grande de datos, y en algunos casos, por ejemplo si queremos graficar una recta, no tendrá sentido generar 100 puntos para construir su gráfico. En definitiva, la conclusión es que el paso deberá ser lo suficientemente pequeño como para poder resolver las variaciones de la función, pero no lo suficientemente grande como para generarnos inconvenientes al tener que manejar un conjunto muy grande de datos.

Para mostrar el procedimiento a seguir, utilicemos un ejemplo. Supongamos que se quiere graficar la función cuadrática $f(x)=-2x^2+4x+1$, en el rango $(-1,3)$. En una hoja de cálculo, elijo una columna y la titulo "Abcisa". Debajo, ingreso el primer valor del rango, $x_1=-1$. Elegimos un paso de $\Delta x=0,2$, y entonces debajo del primer valor, ingreso $x_1+\Delta x$, o sea $-0,8$. Luego, selecciono ambas celdas y arrastro hacia abajo el cuadrado que aparece abajo y a la derecha de la selección (ver Fig. 4.1). Al hacer esto, el programa entiende que queremos rellenar las filas inferiores a la selección, y mas aún, entiende que queremos hacer eso de manera tal que el valor en cada celda sea el valor de la de arriba más $\Delta x=0,2$. Dejamos de arrastrar cuando vemos que llegamos al límite superior de nuestro rango, en este caso, 3.

	A	B	C
1			
2		Abcisa	Ordenada
3		-1,0	
4		-0,8	
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			

	A	B	C
1			
2		Abcisa	Ordenada
3		-1,0	
4		-0,8	
5		-0,6	
6		-0,4	
7		-0,2	
8		0,0	
9		0,2	
10		0,4	
11		0,6	
12		0,8	
13		1,0	
14		1,2	
15		1,4	
16		1,6	
17		1,8	
18		2,0	
19		2,2	
20		2,4	
21		2,6	
22		2,8	
23		3,0	

Figura 4.1

Para completar la serie de puntos, debemos especificar que ordenada le corresponde a cada valor de x . Para ello, primero calculemos la ordenada $f(x_1)$. En la celda que se encuentra a la derecha de x_1 (en el caso del gráfico, C3), queremos introducir $f(-1)$, por lo que ingresamos

$$=-2*B3^2+4*B3+1$$

Como se había tratado anteriormente, al querer operar con un número que se encuentra en una celda, introducimos el nombre de la celda en la fórmula. Recordemos que no es necesario escribirlo, sino que el programa lo hará automáticamente si hacemos click en la celda; o sea que la manera más cómoda de ingresar la fórmula sería

$$=-2*(\text{click en celda B3})^2+4*(\text{click en celda B3})+1$$

Una vez hecho esto, el programa calculó $f(-1)=-5$ y lo introdujo en la celda C3. Notemos que al seleccionar dicha celda, en la barra de fórmula (ver Figura 4.2), no aparece el valor numérico, sino precisamente la fórmula que introducimos.

El próximo paso sería calcular el resto de las ordenadas. Para esto, no es necesario repetir el procedimiento: simplemente debemos seleccionar la primera celda en la que hicimos el cálculo y arrastrar hacia abajo como se explicó antes. El programa interpreta automáticamente que, al llenar la celda C3 con la fórmula ingresada evaluada en B3, nosotros pretendemos llenar la C4 con el valor de B4 y así sucesivamente. Arrastrando la selección hacia abajo y hasta el final de los datos, obtenemos una tabla de datos que generamos de una manera automatizada (ver Fig. 4.2). Observemos que tenemos $N=(3-(-1))/2=20$ puntos

A partir de la tabla de datos, podemos generar un gráfico siguiendo el procedimiento de la Sección 3 (ver Fig. 4.2). Nuevamente elegiremos la opción “Dispersión XY”, y ésta vez, podemos elegir la opción “Sólo línea” del gráfico, ya que nos interesa una curva sólida y no ver graficada la sucesión de puntos.

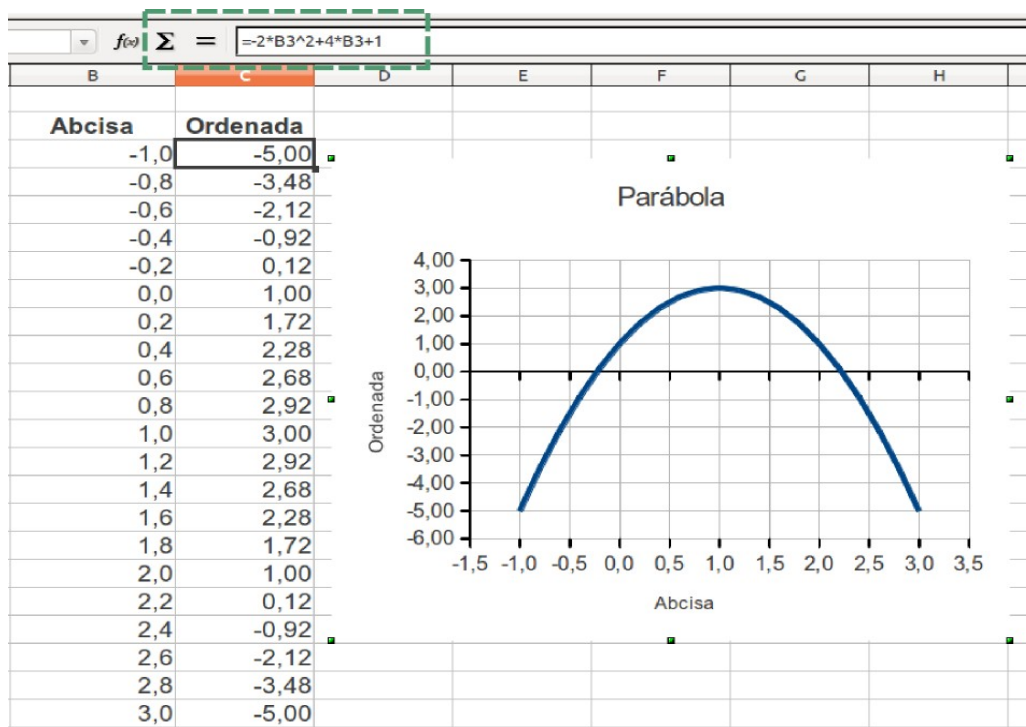


Figura 4.2

Como ya fue remarcado en la Sección 2, una de las principales ventajas que tiene trabajar con planillas de cálculo construidas de ésta forma, es su carácter netamente dinámico. La columna de las ordenadas fue calculada a partir de una fórmula que depende de los valores que se encuentran en la columna que está a su izquierda, y por lo tanto si queremos cambiar, por alguna razón, los valores de dicha columna (la de las abcisas), el programa recalculará automáticamente la columna de las ordenadas. Por ejemplo, borre el contenido de la columna de abcisas y pruebe qué pasa si, repitiendo el mismo rango, usamos un paso de $\Delta x=0,5$, o peor aún, de $\Delta x=1$.

4.2. Cálculo numérico de derivadas

Al trabajar con una serie de datos, por ejemplo de la posición en función del tiempo de un cuerpo, puede ser necesario obtener una estimación de cual es la velocidad del cuerpo en cada instante. Como la definición de derivada implica el cálculo de un límite, y la información que tenemos está restringida a un conjunto finito de datos, solamente obtener una estimación del valor de la derivada en cada punto. Supongamos entonces que tenemos una serie de registros $\{x_1, x_2, \dots, x_N\}$ de la posición en los correspondientes instantes $\{t_1, t_2, \dots, t_N\}$. Una buena estimación de la velocidad v_i en el instante t_i estará dada por el cociente incremental:

$$v_i = \frac{x_{i+1} - x_i}{t_{i+1} - t_i}$$

que es, por definición, la velocidad media entre los instantes t_i y t_{i+1} .

Veamos como implementar esto en una planilla de cálculo. En la Fig. 4.3 se observan los datos de la coordenada vertical (y) de una partícula en un tiro oblicuo (columna C), como función del tiempo (columna B).

	A	B	C	D
2				
3		Tiempo (s)	Coord. y (m)	Velocidad y (m/s)
4		0	0,00	
5		0,03	0,13	
6		0,06	0,24	
7		0,09	0,36	
8		0,12	0,47	
9		0,15	0,57	
10		0,18	0,66	
11		0,21	0,74	
12		0,24	0,80	
13		0,27	0,86	
14		0,3	0,92	
15		0,33	0,95	
16		0,36	0,99	
17		0,39	1,02	
18		0,42	1,03	
19		0,45	1,04	
20		0,48	1,03	
21		0,51	1,01	

Figura 4.3

En la columna D, calcularemos las velocidades en cada instante. Para ello, primero calculamos el primer valor, que corresponde al instante $t=0$. En la celda D4 debemos introducir una fórmula que calcule la velocidad media entre $t=0$ y $t=0,03$, a partir de la expresión que se dió en el párrafo anterior. Dicha fórmula es

$$=(C5-C4)/(B5-B4)$$

Recordemos que para ingresarla, lo más comodo es tipear el signo de igualdad, y luego ir clickeando con el mouse en las celdas que intervienen el cálculo, intercalando obviamente las operaciones correspondientes (y los paréntesis). Una vez ingresada la fórmula, en la celda D4 se lee 4,47 (es una buena idea formatear las celdas previo a los cálculos para que muestren los números solo con dos cifras decimales, como explicó en la Sección 2). Para el resto de los valores, volvemos a aprovechar el autocompletado de celdas. Basta con seleccionar la celda D4 y luego arrastrar

(haciendo click sobre el cuadrado de la esquina inferior derecha) hacia abajo la selección, para que las celdas de las filas inferiores se rellenen con las velocidades medias correspondientes. Este representa un ejemplo más de la característica de autollenado, además de los vistos en las secciones 1 y 4.1, y en este caso vemos que en la fórmula intervienen varias celdas distintas, sin embargo el programa entiende cuál es el cálculo que queremos hacer cuando extendemos la fórmula a las celdas subsiguientes. El resultado del cálculo se aprecia en la Fig. 4.4, junto con el gráfico de los datos, que fue realizado tal como se describió en la Sección 3. Notemos que las columnas involucradas en el gráfico (B y D) no son adyacentes; la manera de seleccionarlas de manera conjunta (sin incluir a la columna C) es seleccionando los datos de la primera (correspondiente a las abcisas), y luego presionando la tecla Control y seleccionando la segunda (la de las ordenadas).

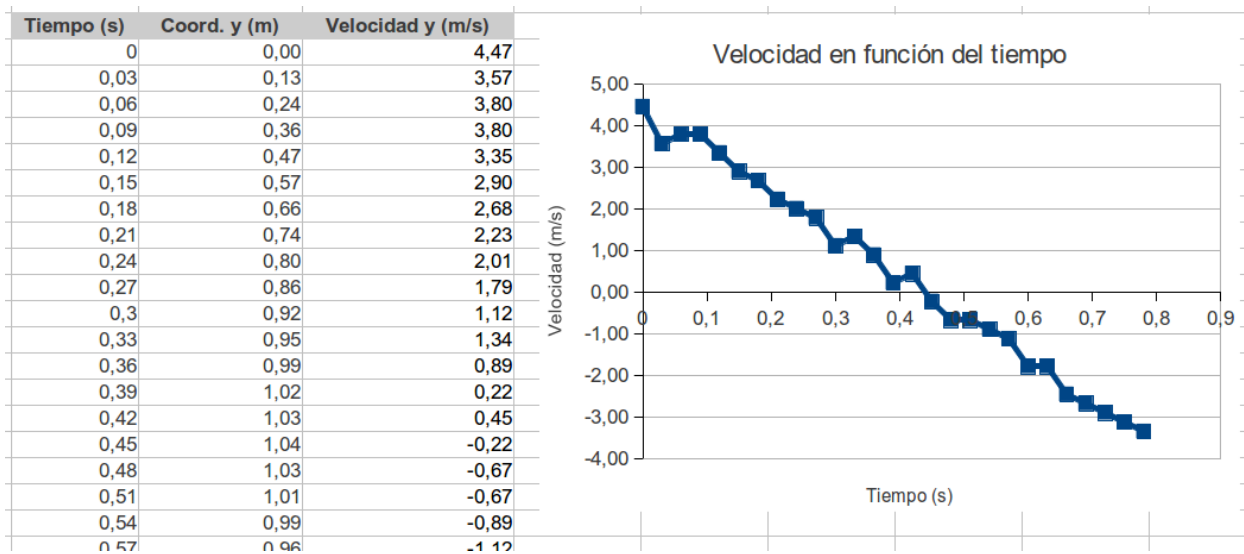


Figura 4.4