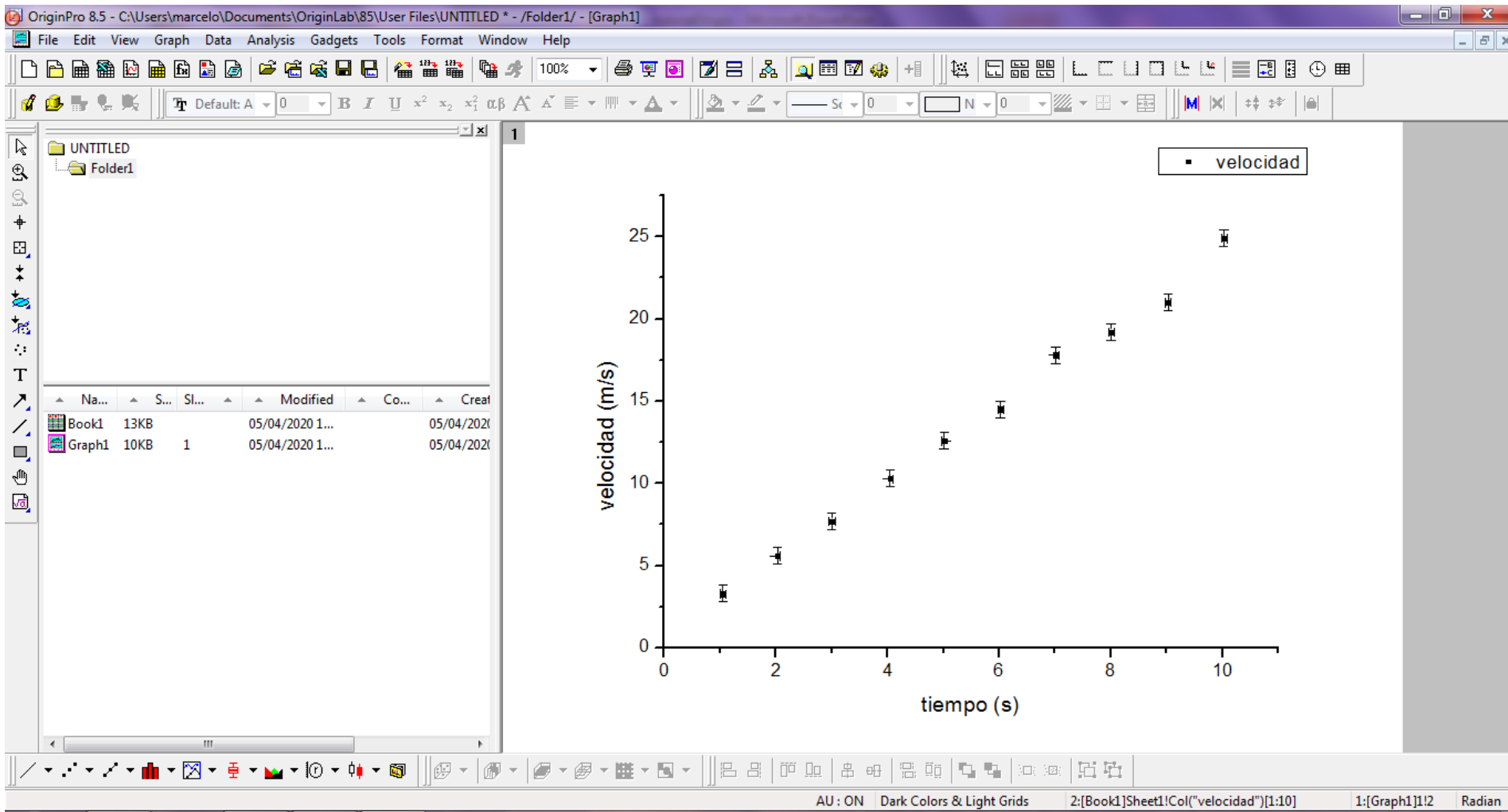


¿Cómo realizar con Origin una
regresión lineal por cuadrados
mínimos?

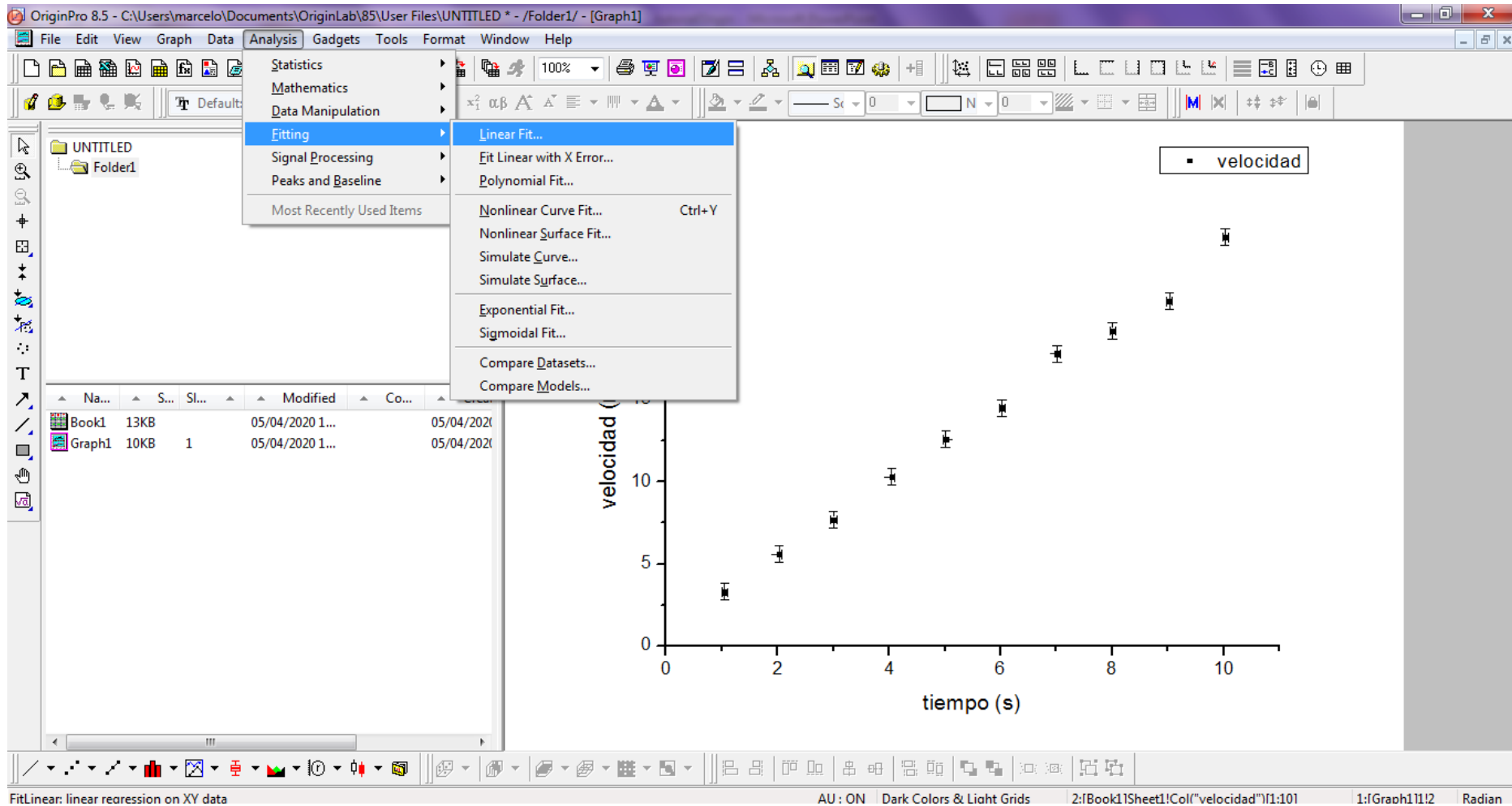
Marcelo Otero

Partimos de un scatter plot de ambas variables.

Para hacer cuadrados mínimos lineales teniendo en cuenta las barras de error (cuadrados mínimos ponderados) recordar poner en el eje Y aquella variable con “mayor error relativo”.



Habiendo leído previamente en que consiste la técnica de regresión lineal por cuadrados mínimos, procederemos a ejecutarla en Origin. Seleccione en el menú: Analysis -> Fitting -> Linear Fit



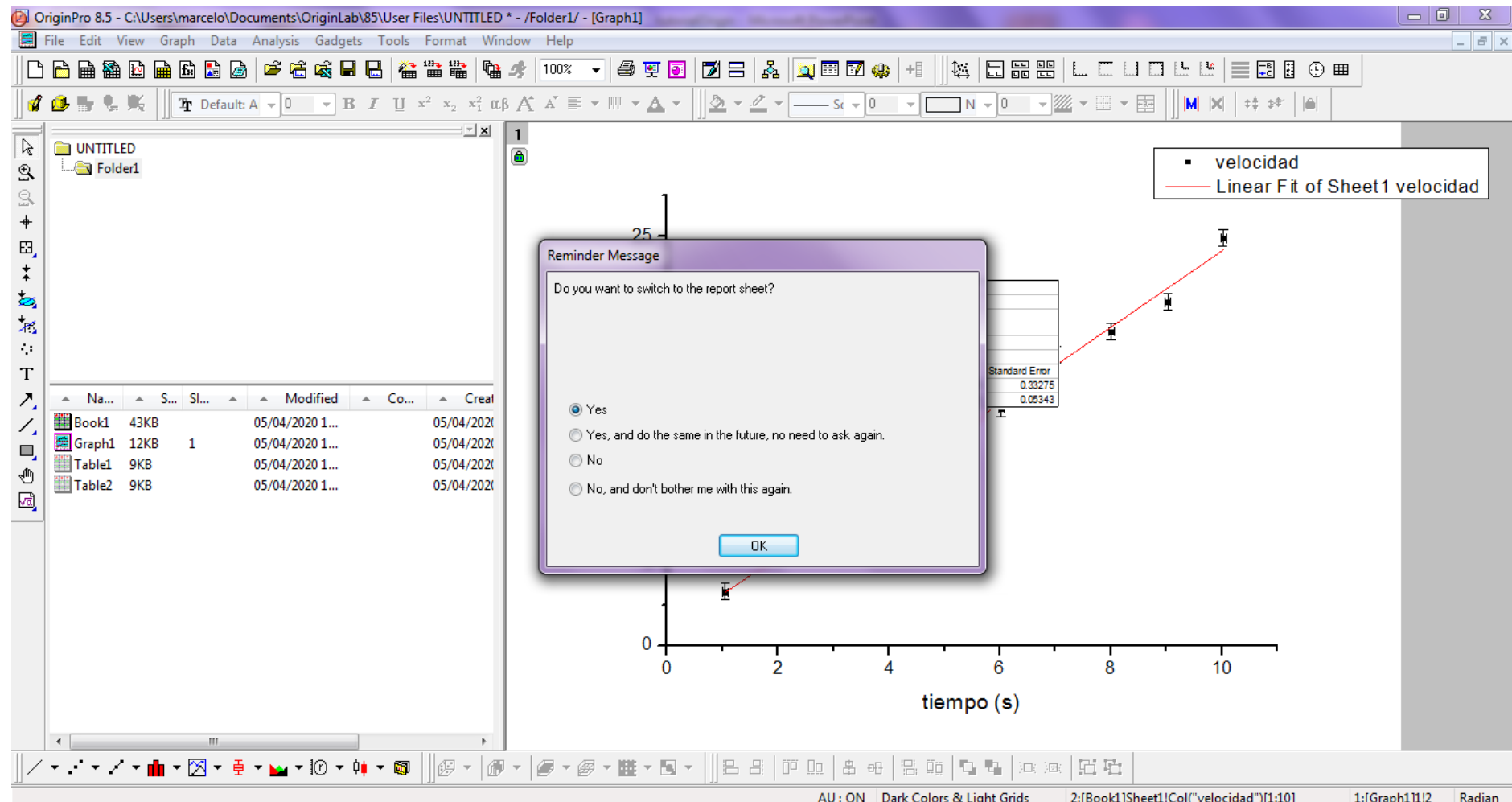
Como partimos de un gráfico donde las variables poseen barras de incertezas, Origin elige en forma automática Cuadrados Mínimos Ponderados (Error as Weight: Instrumental) como herramienta para realizar la regresión lineal. Basta solamente presionar OK

The image shows the OriginPro 8.5 software interface. A 'Linear Fit' dialog box is open, displaying the following settings:

- Dialog Theme: [Default]
- Description: Perform Linear Fitting
- Recalculate: Manual
- Multi-Data Fit Mode: Independent - Consolidated Report
- Input Data: [[Graph1]!12"velocidad"]
- Range 1: [[Graph1]!12"velocidad"]
- Fit Options:
 - Errors as Weight: Instrumental
 - Fix Intercept:
 - Fix Intercept at: 0
 - Fix Slope:
 - Fix Slope at: 1
 - Use Reduced Chi-Sqr:
 - Apparent Fit:
- Quantities to Compute:
 - Residual Analysis:
 - Output Settings:
 - Fitted Curves Plot:
 - Find Specific X/Y:
 - Residual Plots:

The background shows a graph with a data series named 'velocidad' plotted against an unlabeled x-axis. The data points have vertical error bars. The x-axis has major ticks at 8 and 10. The status bar at the bottom indicates 'AU : ON', 'Dark Colors & Light Grids', and the active cell address '2:[Book1]Sheet1!Col("velocidad")!1:101'.

Aca damos OK a la opción por default (Yes) para que Origin nos guarde una hoja con toda la información de la regresión lineal.



En esta hoja, Origin nos brinda información sobre la regresión, la cual resume también en el gráfico, al que pueden acceder seleccionándolo en la ventana inferior izquierda (Graph 1)

The screenshot displays the OriginPro 8.5 software interface. The main workspace shows a linear fit analysis for the variable 'velocidad'. The statistical summary is as follows:

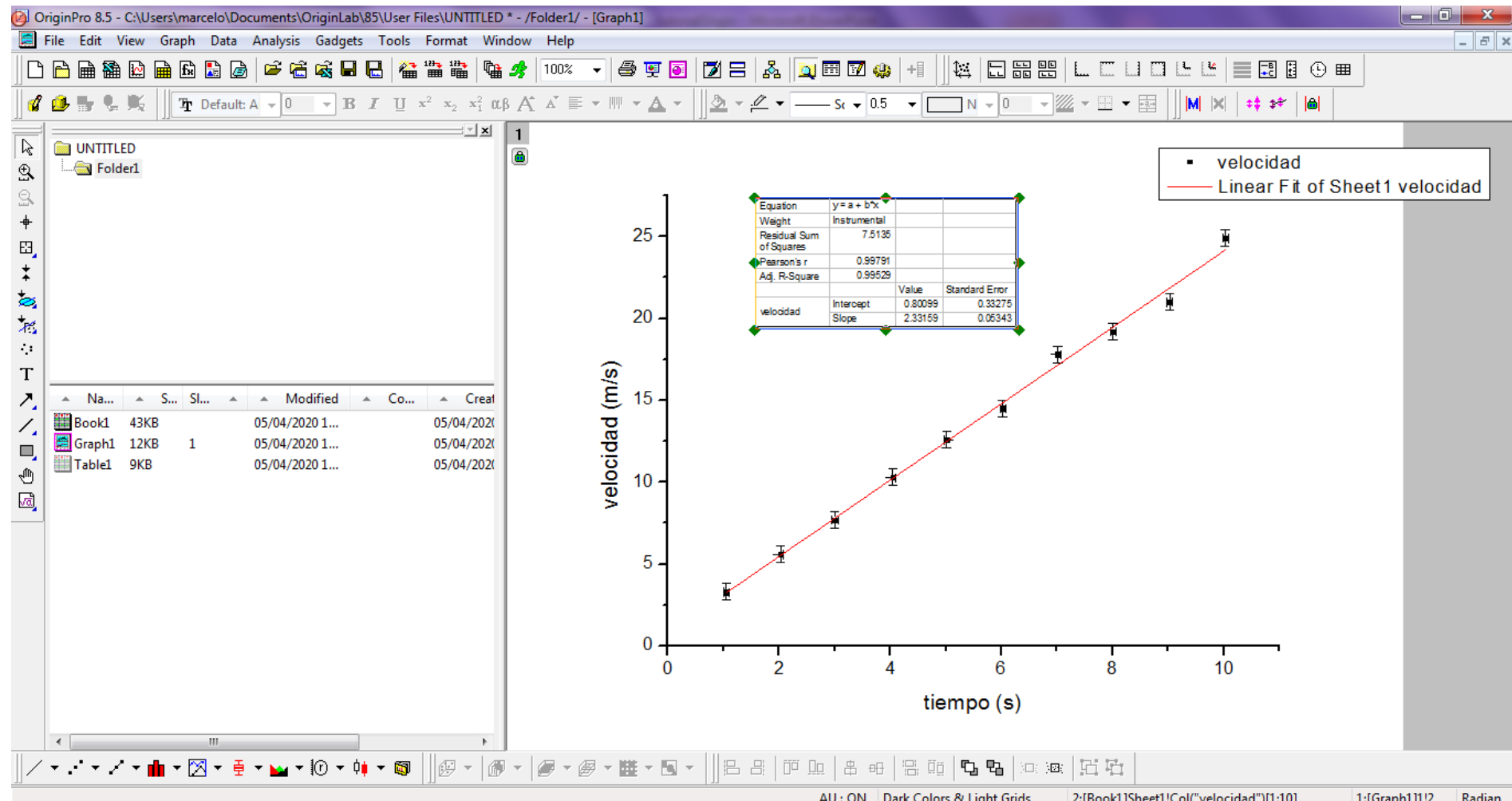
		Intercept		Slope		Statistics
		Value	Standard Error	Value	Standard Error	Adj. R-Square
velocidad	Intercept	0.80099	0.33275			
	Slope			2.33159	0.05343	

		velocidad	
	Number of Points		10
	Degrees of Freedom		8
	Residual Sum of Squares		7.5135
	Pearson's r		0.99791
	Adj. R-Square		0.99529

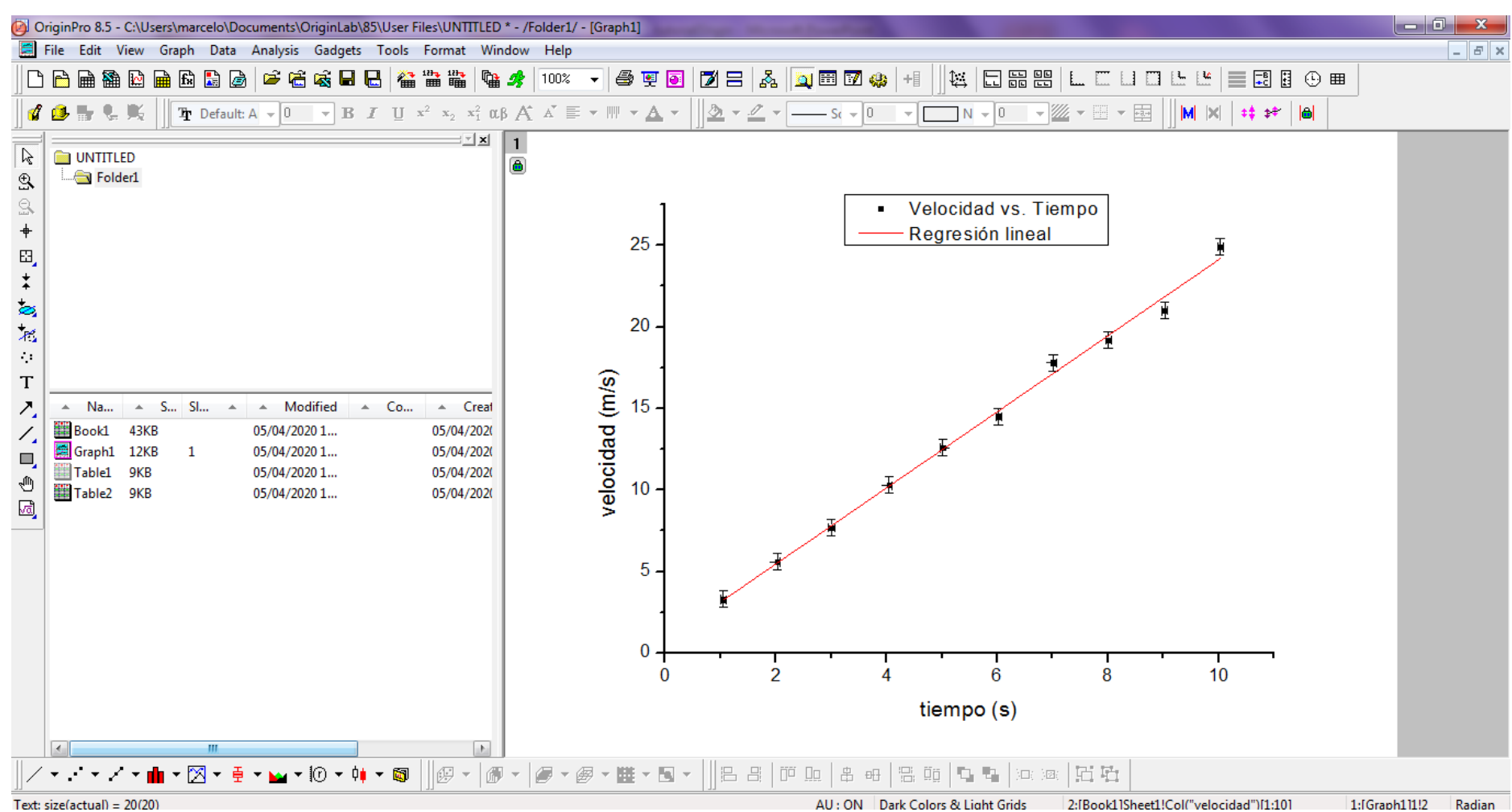
		DF	Sum of Squares	Mean Square	F Value	Prob>F
velocidad	Model	1	1788.3625	1788.3625	1904.15837	8.39171E-11
	Error	8	7.5135	0.93919		
	Total	9	1795.876			

The interface also shows a 'Fitted Curves Plot' for 'velocidad' in the bottom left corner. The status bar at the bottom indicates 'AU: ON' and '2: IBook1\FitLinear1! Radian'.

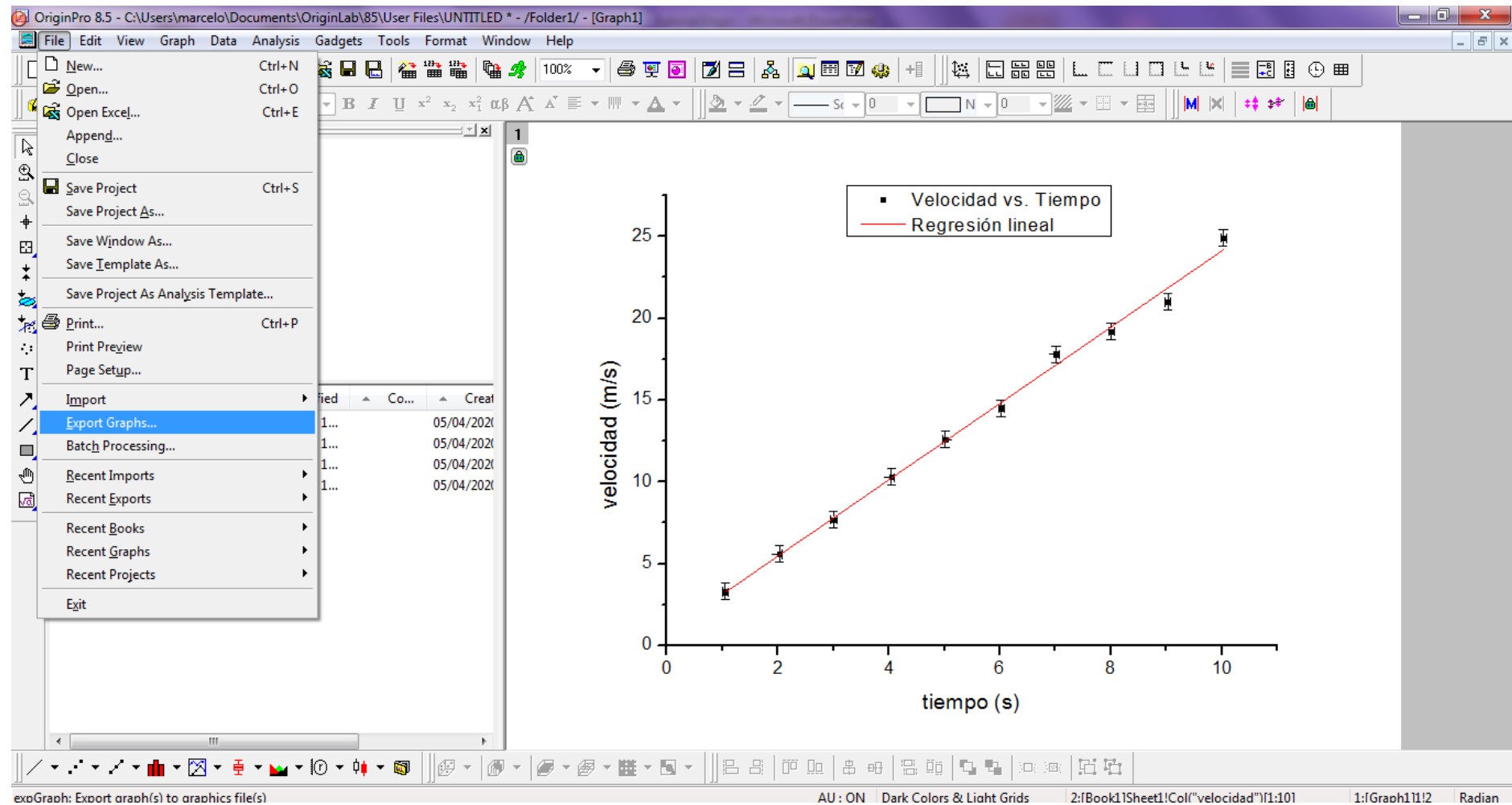
Ya tenemos listo el gráfico de dispersión con la regresión lineal por cuadrados mínimos. Antes de exportar el gráfico para el informe vamos a copiar la tabla con la información de la regresión lineal y vamos a mejorar la leyenda.



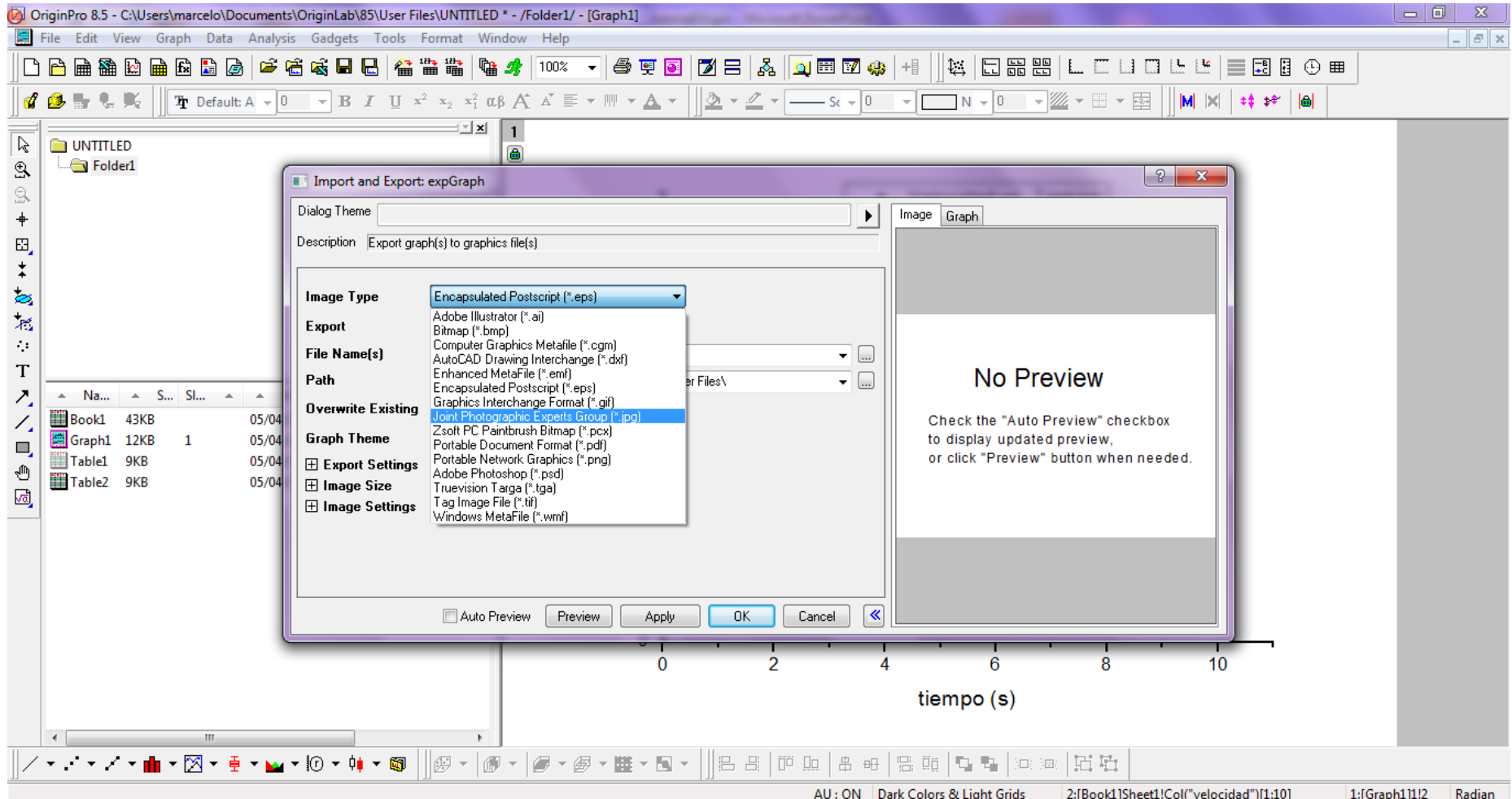
La leyenda se modifica haciendo doble click sobre la misma. Indicamos que los puntos corresponden al gráfico de una variable vs. la otra (en este caso velocidad vs. tiempo) y que la recta roja corresponde a la regresión lineal.



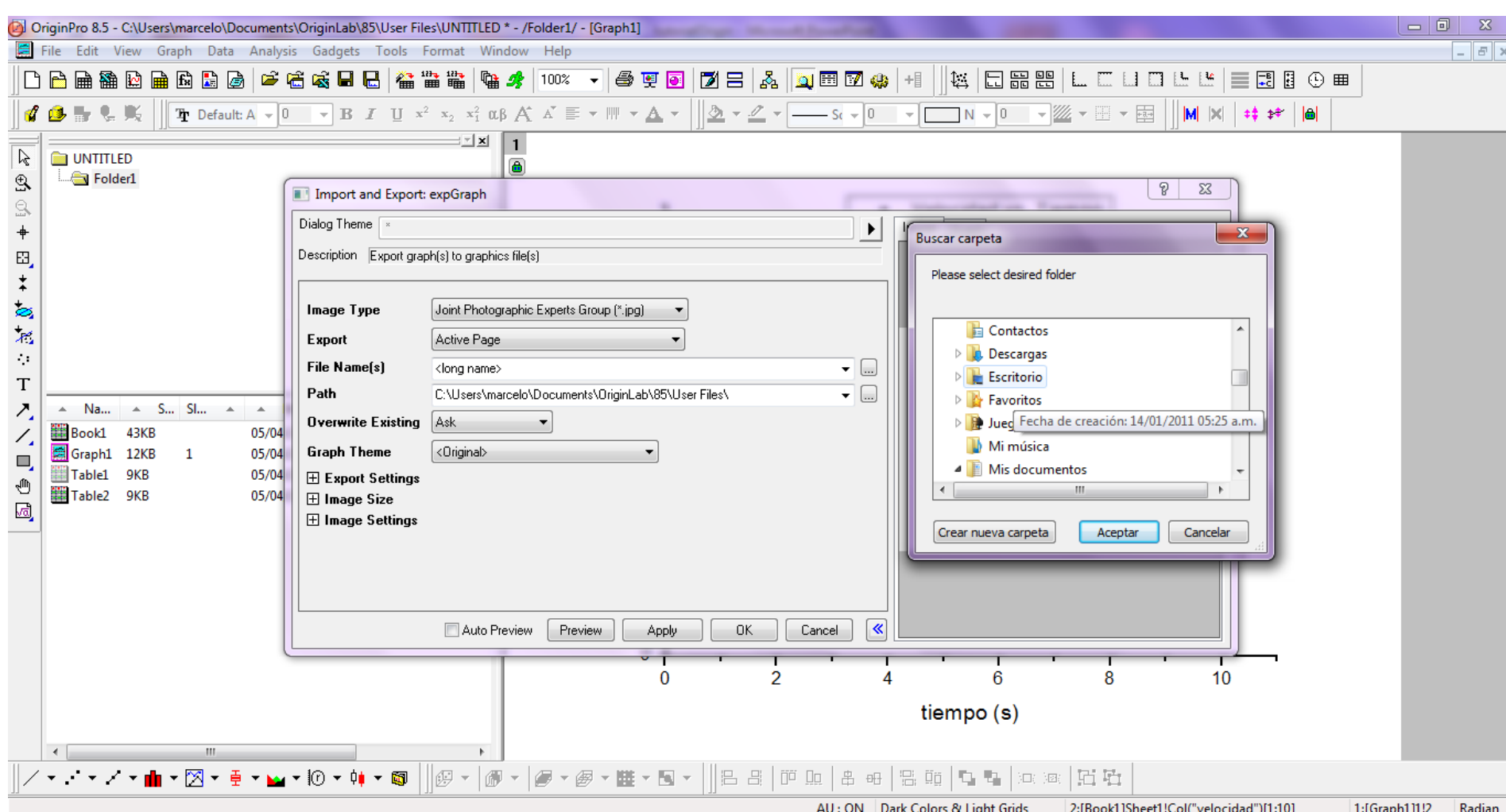
Para exportar el gráfico para el informe vamos al menú principal:
File -> Export Graphs...



En el menú que se despliega, seleccionamos que exporte el gráfico en formato JPG.



Y luego seleccionamos que exporte el gráfico en el escritorio.
Luego seleccionamos Apply -> OK y listo!!!



¿Qué información de la regresión
vamos a utilizar?

La tabla que genera Origin en el gráfico nos brinda cierta información que vamos a utilizar:

Equation $y = a + b*x$

Nos indica cual es la ecuación utilizada en la regresión, de manera tal que “a” corresponde a la ordenada al origen y “b” a la pendiente de la recta

Weight Instrumental

Significa que realizó regresión lineal por cuadrados mínimos ponderados (ver apunte)

Pearson's r 0.99791

Adj. R-Square 0.99529

Ambos son índices que miden la calidad del ajuste. El coeficiente de R de Pearson nos indica que tan buena es la correlación lineal entre las variables y toma valores en el intervalo [-1,1]. Si $R = 0$ no hay correlación lineal, si $|R| = 1$ los datos son colineales. Si $R > 0$ la pendiente es positiva y si $R < 0$ la pendiente es negativa. Cuanto más cercano a 1 es R, mayor es la correlación lineal entre las variables.

El R cuadrado ajustado adj R-Square, mide la correlación lineal entre las variables pero toma valores en el intervalo [0,1]. Cuanto más cercano a 1 es el valor, mayor es la correlación lineal entre las variables.

Parámetros del ajuste (ordenada al origen y pendiente con sus incertezas):

		Value	Standard Error
velocidad	Intercept	0.80099	0.33275
velocidad	Slope	2.33159	0.05343

Intercept corresponde a la ordenada al origen de la recta, **Slope** corresponde a la pendiente de la recta y **Standard Error** corresponde a la incerteza de cada uno de los parámetros de la regresión.

Origin lo expresa con muchos dígitos. Los nombres de los parámetros deben estar siempre en castellano y deben ser expresados en el informe con las cifras significativas adecuadas y con unidades.

Por ejemplo, considerando una sola cifra significativa en la incerteza:

La ordenada al origen es: $a = (0.8 \pm 0.3) \text{ m/s}$

La pendiente es: $b = (2.33 \pm 0.05) \text{ m/s}^2$

En este ejemplo si las variables corresponden al tiempo y a la velocidad en un MRUV, entonces la regresión lineal se corresponde con la ecuación:

$$V = V_0 + a_c * t$$

donde V_0 es la velocidad inicial y a_c la aceleración. Por lo cual, en este ejemplo se podría estimar que

$$V_0 = a = (0.8 \pm 0.3) \text{ m/s}$$

$$a_c = b = (2.33 \pm 0.05) \text{ m/s}^2$$

Si nuestro objetivo era estimar la velocidad inicial V_0 y la aceleración a_c , en este caso los mismos coinciden con los parámetros de la regresión “a” y “b”.

Sin embargo, las magnitudes que uno quiere estimar no siempre coinciden con los parámetros de la regresión lineal “a” y “b”, sino que a veces son una función de los mismos.

Ejemplo: Supongamos un experimento de termodinámica de gases ideales a temperatura constante (T) en el cual medimos la presión (P_i) de un gas para distintos volúmenes (V_i) siempre para un mismo número de moles (n) de un gas ideal. P_i , V_i , T y n son determinados experimentalmente y tienen asociados una incerteza (ΔP_i , ΔV_i , ΔT y Δn). Nuestro objetivo es calcular la constante de los gases ideales R .

Hacemos 10 mediciones en el laboratorio y tenemos 10 pares de datos ($P_i \pm \Delta P_i$, $V_i \pm \Delta V_i$).

Conociendo la ecuación de estado de un gas ideal, sabemos que la relación entre la presión y el volumen está dada por la ecuación: $PV = nRT$, la cual podemos reescribir como: $P = nRT(1/V)$

Ésta no es ni más ni menos que una relación lineal entre la presión P y la inversa del volumen ($1/V$), donde la pendiente corresponde a nRT .

Si realizamos una regresión lineal de los valores $(P_i, 1/V_i)$ incluyendo sus incertezas: ΔP_i y $\Delta(1/V_i)$, esperamos obtener una recta $Y = a + bX$, donde Y corresponde a la variable P y X corresponde a la variable $1/V$.
Aclaración: $\Delta(1/V_i)$ se determina por propagación a partir de la incerteza $\Delta(V_i)$.

La regresión lineal nos arrojará una ordenada a $\pm \Delta a$ y una pendiente $b \pm \Delta b$, donde Δa y Δb corresponden a las incertezas de la ordenada al origen y la pendiente respectivamente.

En este caso esperamos:

- que la ordenada al origen valga cero, o lo que es equivalente, que el valor cero esté contenido en el intervalo: $[a - \Delta a, a + \Delta a]$
- que la pendiente sea $b = nRT$, por lo cual podemos despejar R de la expresión obteniendo: $R = b/nT$

Como vemos, la constante de los gases a determinar R , no coincide con la pendiente de la recta de regresión, sino que es función de la pendiente b y las variables n y T medidas experimentalmente.

En este caso $R = b/nT$ se calculará a partir de los valores de b , n y T observados y la incerteza de R se calculará por propagación de los errores de ΔT , Δn y Δb (ver propagación de errores, guía 2).