



CREER...CREAR...CRECER

# USO DE LAS SIMULACIONES EN SAGE Y SUS IMPLICANCIAS EN LA ENSEÑANZA DE LA REGRESIÓN LINEAL

Franco Nicolas Bellomo<sup>1</sup>, Héctor Agnelli<sup>2</sup>

Facultad de Ciencias Exactas, Físico-Químicas y Naturales; Universidad Nacional de Río Cuarto

1. fnbellomo@gmail.com, 2. hagnelli@exa.unrc.edu.ar



## 1. Objetivo

Proponer una alternativa pedagógica utilizando simulaciones con Sage<sup>3</sup>, destinada a la comprensión de supuestos aleatorios que fundamentan la aplicación de la regresión lineal. El trabajo se aborda aplicando el estudio de los residuos, el incumplimiento de los supuestos de normalidad y homocedasticidad

## 2. Introducción

La enseñanza de la regresión lineal, está presente en casi todos los cursos iniciales de estadística, que se desarrollan en los primeros años de diversas carreras universitarias. En particular, el modelo de regresión simple brinda la oportunidad, desde la perspectiva de la enseñanza, de poner en evidencia que es necesario, para la aplicación de una metodología estadística tener en cuenta que se verifique, para los datos bajo análisis, un conjunto de supuestos que le confieran validez a las inferencias que se formulen luego de aplicar el método. Es importante, durante el proceso de enseñanza-aprendizaje enfatizar estos conceptos por sobre la aplicación rutinaria de procedimientos. Como es sabido, el análisis de los residuos permite la verificación de los supuestos del modelo. En el trabajo, se aborda la propuesta aquí presentada, mediante la presentación de simulaciones realizadas en Sage, las distorsiones que se presentan en el comportamiento de los residuos, cuando no se cumple el supuesto distribucional de normalidad o porque se incumple la homogeneidad de varianzas. La simulación permite a los estudiantes cambiar las distribuciones y los valores de la varianza. Pensamos que el uso de estas herramientas puede contribuir a que el estudiante fije ideas para poder establecer el nexo adecuado entre modelo y realidad.

## 3. Modelo

El modelo de ajuste para la regresión lineal es

$$y_i = E(y_i/x_i) + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Donde  $\beta_0, \beta_1, x$  son las componentes sistemáticas (mediciones "exactas"), y  $\varepsilon_i$  son las componentes aleatorias del error. Las hipótesis acerca de los errores son:

- Los errores tienen la misma distribución, con media cero y varianza homogénea

$$E(\varepsilon_i) = 0 \quad \text{VAR}(\varepsilon_i) = \sigma^2$$

- Los errores no están correlacionados entre sí ni con las  $x_i$ , es decir

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{si } i \neq j$$

$$\text{Cov}(\varepsilon_i, x_i) = 0$$

Con estas tres hipótesis (media cero, homogeneidad y no correlación) podemos asegurar que los errores estándar de los estimadores no sean sesgados. Adicionalmente, se asume distribución normal centrada en cero para todos los errores

$$\varepsilon_i \sim N(0, \sigma^2)$$

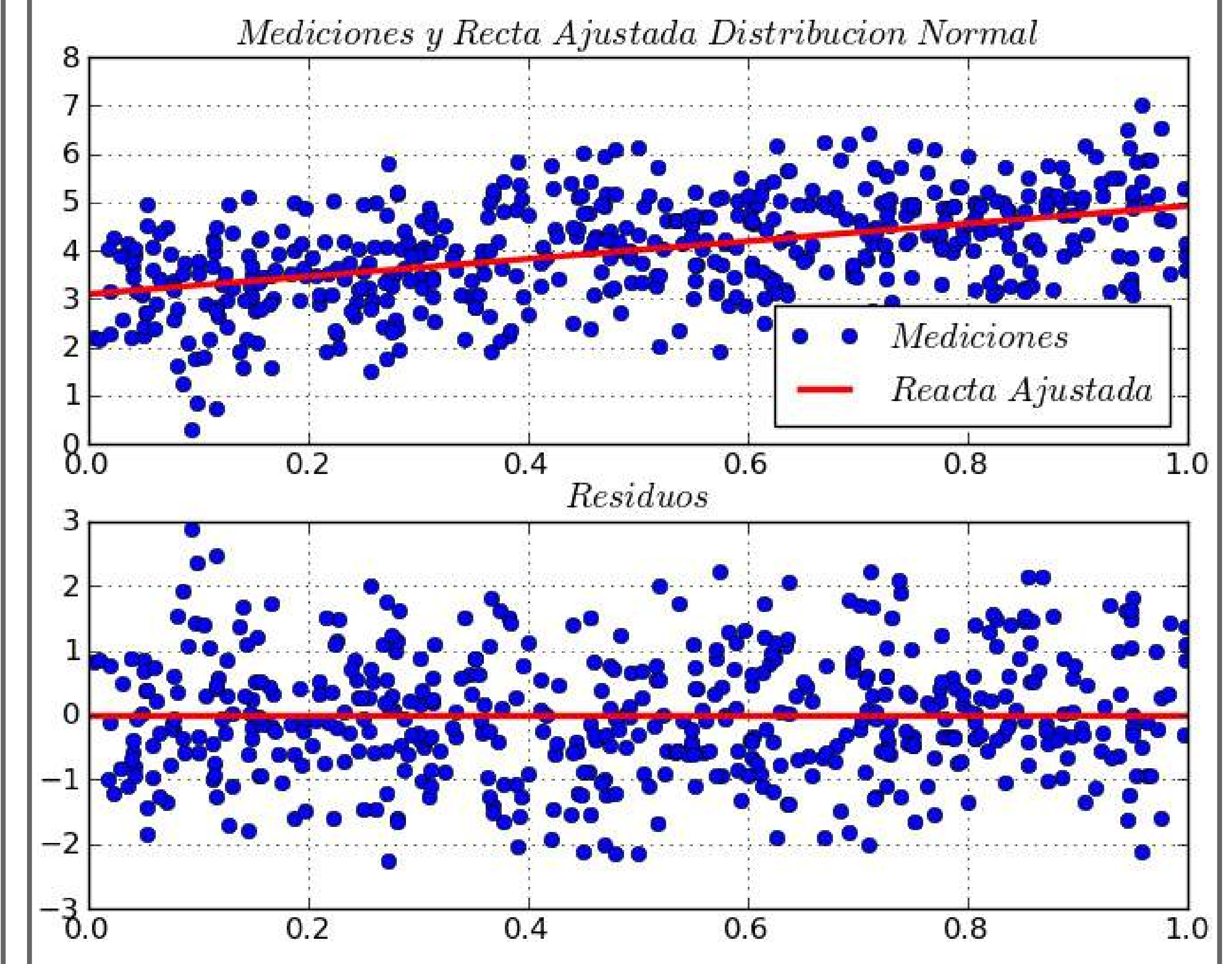
Donde, la no correlación implica independencia entre cada evento (medición). Normalidad: para construir intervalos de confianza y tests de hipótesis. Un fuerte apartamiento afecta los niveles de confianza y niveles de significación.

## 4. Simulaciones

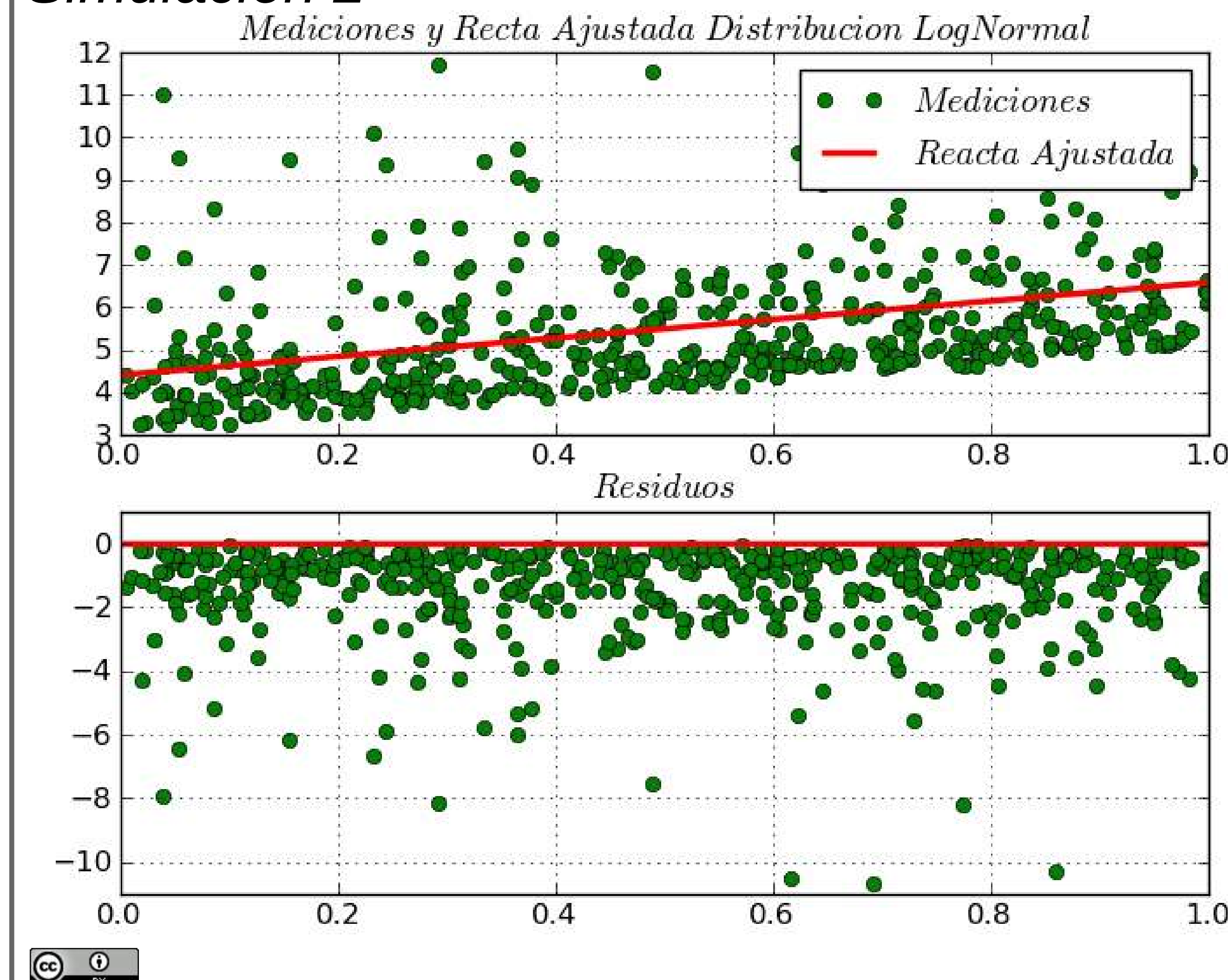
A continuación, se presentan algunas simulaciones, en las que se variaron distintos parámetros, los cuales no cumplieron con las hipótesis.<sup>4-5</sup>

### Simulación 1

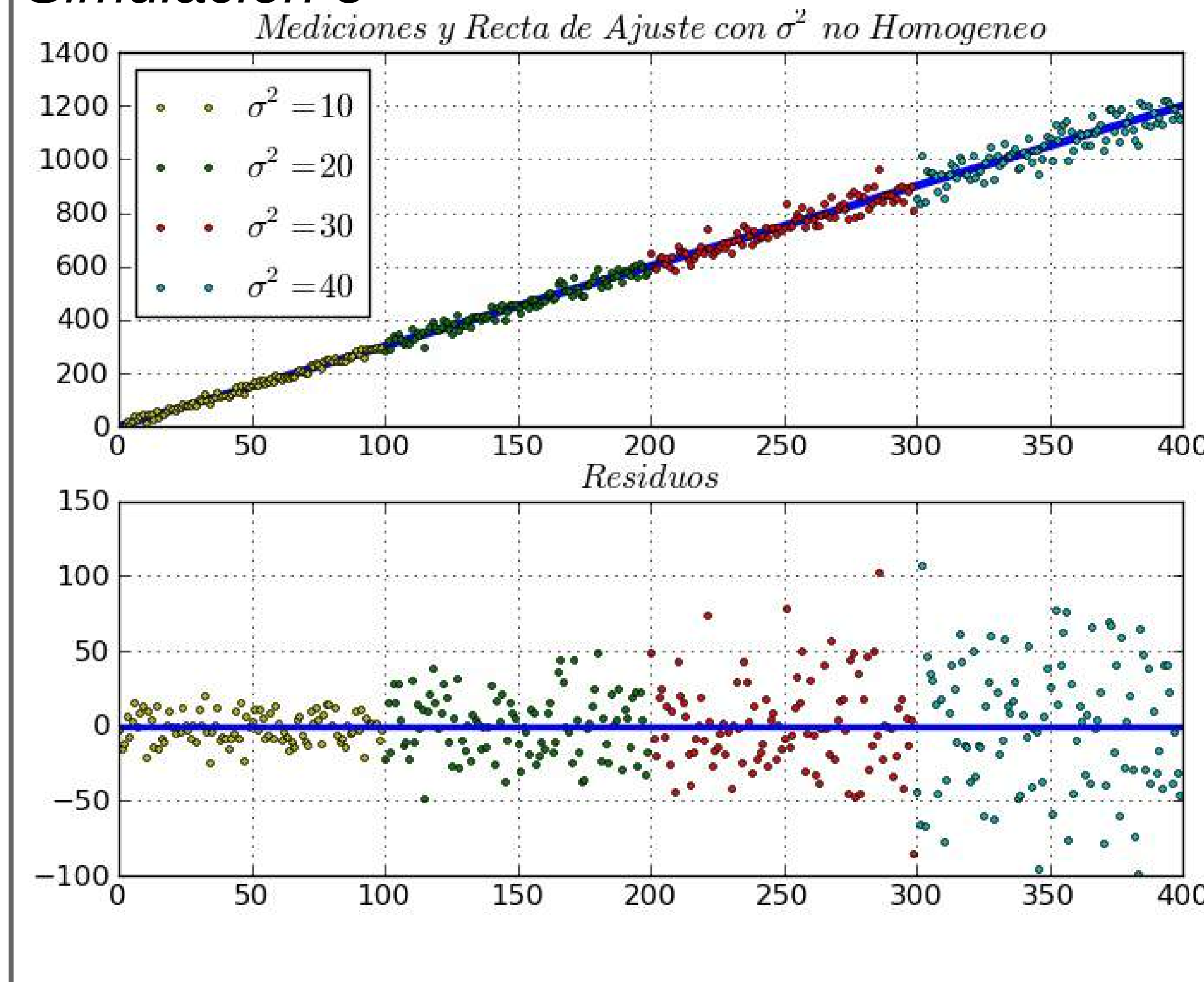
Esta es la simulación patrón cumpliendo todas las hipótesis.



### Simulación 2



### Simulación 3



## 5. Conclusión

Con estos ejemplos, se demuestra la importancia que tiene el observar los residuos luego de haber realizado la regresión. Por medio de ellos, resulta simple verificar si se cumplieron o no las hipótesis. A su vez, al detectar que no cumplimos con alguna, nos puede ayudar en nuestras mediciones a darnos cuenta que estamos incurriendo sistemáticamente en algún error.

El uso de Sage como herramienta de simulación, ofrece la posibilidad de que el usuario pueda interactuar generando modelos con distintos tamaños de muestra, distribuciones y valores de varianza mediante una simple interfaz gráfica. Esto contribuye a que el alumno pueda estar en presencia, sencillamente, de distintas situaciones y alcanzar una valoración adecuada de la importancia de los supuestos al aplicar un procedimiento estadístico.

## 6. Referencias

- 1 - Hamilton L. Regression with Graphics. Duxbury Press. California 1992.
- 2 - Montgomery, D.Peack, E. Introducción al Análisis de Regresión. Grupo Editorial Patria. México 2007.
- 3 - William A. Stein et al. Sage Mathematics Software (Version 5.9), The Sage Development Team, 2013, <http://www.sagemath.org>
- 4 - Código disponible en: <http://sage.ccad.unrc.edu.ar/home/pub/35/>
- 5 - Agradecemos, por el soporte computacional, al Centro de Computación de Alto Desempeño de la Universidad Nacional de Río Cuarto (CCAD UNRC), <http://www.unrc.edu.ar>