

# Actividad 1: Validación de Supuestos

27 de mayo de 2019

Iniciaremos nuestro trabajo con un modelo lineal simple, de la forma  $Y = \beta_0 + \beta_1 X + e$ , donde además supondremos que se cumplen los cuatro supuestos planteados inicialmente sobre los errores. Para este modelo, establecemos que  $\beta_0 = 15,3$ ,  $\beta_1 = 8,39$  y  $\sigma^2 = 50,3$ .

1. Lo primero que analizaremos es la influencia del tamaño de muestra en las estimaciones del modelo y las propiedades de las mismas. En este caso, iniciaremos con una muestra de tamaño  $n = 4$  y este se debe ir incrementando de forma gradual hasta llegar a una muestra de tamaño  $n = 300$  y evaluaremos los siguientes aspectos:
  - a. En teoría, si el modelo está correctamente especificado, las estimaciones son insesgadas. El proceso de simulación en este caso consiste en evaluar el promedio y la mediana de 100 estimaciones para diferentes tamaños de muestra y construir una curva que permita observar el comportamiento de dichas estimaciones en función del tamaño de muestra.
  - b. Por otra parte, se debe observar la variabilidad de las estimaciones obtenidas. En este caso, se puede adecuar el proceso de simulación para que almacene el cálculo de la varianza de las 100 estimaciones para cada tamaño de muestra y compararla con el valor teórico de las mismas.
  - c. Otra de las características interesantes a evaluar es el Error Cuadrático Medio, calculado como el valor esperado del cuadrado de la diferencia entre el valor estimado y el valor real del parámetro.
  - d. Finalmente, el proceso inferencial y el contraste de hipótesis dependen en gran medida del tamaño de muestra. En este caso, ¿cómo se vería afectada la potencia de las pruebas para estos tamaños de muestra?
2. ¿Qué sucede con el supuesto de correcta especificación cuando no se cumple? Se podría pensar en un modelo poblacional cuadrático y ajustar una línea recta. Por ejemplo, supongamos que el modelo poblacional está dado por:

$$Y = 0,5714 + 1,3X + 2,1071X^2 + e$$

donde  $e \sim N(0, \sigma^2 = 0,01)$ . Sin embargo, asumamos que se plantea el siguiente modelo:

$$Y = \beta_0 + \beta_1 X + e$$

y se toma una muestra aleatoria de tamaño  $n = 50$  y se asume entonces que  $\vec{e} \sim N(\vec{0}, \sigma^2 I)$ .

- a. En teoría,  $\beta_0 = 0,5714$ ,  $\beta_1 = 1,3$  y  $\sigma^2 = 0,01$ . En un proceso de simulación como en el caso anterior, se pueden simular 500 muestras aleatorias de tamaño  $n = 50$  y evaluar la convergencia de las estimaciones.

- b. Se suele pensar que al incrementar el tamaño de la muestra, se pueden solucionar algunos problemas de estimación pues por el Teorema Central del Límite, la convergencia tiene que resultar en algún punto. ¿Qué pasa si incrementamos el tamaño de la muestra? Nuevamente, iniciemos con una muestra de tamaño  $n = 4$  y vamos incrementando de forma gradual hasta llegar a una muestra de tamaño  $n = 300$ . ¿Qué se puede concluir al respecto?
- c. También se puede pensar en que las estimaciones de la variabilidad de las estimaciones pueden verse afectadas y las mismas a su vez, afectar el proceso inferencial. En este caso, adecúe el proceso anterior para calcular el Error Cuadrático Medio de las estimaciones y comente respecto a dicho comportamiento.
- d. Finalmente, evalúe el comportamiento de la potencia del contraste de hipótesis para  $\beta_1$  modificando el tamaño de la muestra.