



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ECONOMÍA

CURSO DE ECONOMETRÍA 1 - SEMESTRE 2021.1

FORMATO DE REPORTE INDIVIDUAL



**Fecha:** 28/Sept/2020

**Título:** PRUEBAS DE NORMALIDAD

**Autor/Alumno:** Guillermo Droppelmann

**Correo:** guillermo.droppelmann@meds.cl

**Resumen:**

Es frecuente en la literatura biomédica no comunicar el tipo de distribución. Existen pruebas que permiten conocer la distribución de los datos. Cuando las variables no poseen una distribución normal estas se pueden transformar. El siguiente artículo tiene como propósito identificar las principales pruebas de normalidad en STATA.

**Palabras Claves:** normalidad, pruebas, stata.

**Introducción:**

Un aspecto fundamental en las investigaciones y en gran medida subvalorado por los clínicos es el correcto uso de las pruebas que permiten el análisis de los datos, los errores estadísticos constituyen un aspecto común en las publicaciones biomédicas y se han reportado hasta en un 50%. A su vez, es bastante frecuente encontrar un sin número de pruebas que se utilizan arbitrariamente con el propósito de poder contrastar la docimasia de hipótesis de la mejor manera posible para obtener la

ansiada significancia estadística que tanto busca el grupo de investigadores. No obstante, antes de hacer cualquier tipo intervención en los datos se debe realizar algunas pruebas de normalidad que nos permitan saber que pruebas estadísticas se utilizaran posteriormente para obtener los resultados.

El presente artículo tiene como objetivo identificar las principales pruebas de normalidad en el programa estadístico STATA.

## **Contenido:** DISTIRBUCCIÓN NORMAL

Antes de analizar cualquier tipo de dato, el investigador o el profesional clínico deberá considerar la distribución que poseen estos, es decir, el comportamiento de probabilidades que presentan las variables según su naturaleza continua, la cual presenta la conocida gráfica en función de su densidad denominada campana, campana de Gauss o en su aplicación práctica que se asemeje lo más posible a un histograma.

La importancia de su forma radica principalmente en que constituye un acercamiento inicial para determinar el tipo de test estadístico que se debe utilizar para obtener las diferencias en el contraste de las hipótesis de estudio.

Para poder saber si nuestros datos presentan o no una distribución normal, se podrán utilizar algunos test que permitan reconocer esta característica. No obstante, 5 son los pasos previos que se deben considerar antes de hacer cualquier análisis.

### **III. Evaluación De La Normalidad: Detección y comandos**

#### **3.1. Métodos Numéricos**

Para evaluar la normalidad que presentan nuestras bases de datos se puede realizar a

Tabla 1. Métodos numéricos para evaluar normalidad.

Métodos numéricos	Sintaxis
Skewness / Kurtosis	sktest variable
Shapiro Wilk	swilk variable
Shapiro – Francia	sfrancia variable
Kolmogorov - Smirnov	ksmirnov variable

Lo primero es conocer la naturaleza de las variables de estudio, es decir si estas son cualitativas (nominales, ordinales) o cuantitativas (discretas, continuas).

Lo segundo, es conocer cuántos valores presenta la variable de interés, ya que permitirá decidir que prueba seleccionar con mayor precisión.

El tercer punto, es que los test entregaran un valor y que luego se debe confirmar con métodos gráficos. Cuarto, es que el valor entregado permitirá saber la posterior utilización de pruebas para el análisis de los datos clasificándolas en paramétricas (con distribución normal) o pruebas no paramétricas (sin distribución normal).

El quinto punto es que el valor de normalidad pudiera ser susceptible al tamaño de la muestra teniendo siempre presente que grandes tamaños muestrales permiten mayores probabilidades de poseer una distribución normal de los datos.

—través de los métodos numéricos dentro de los que se encuentran los test que aparecen en la Tabla 1.

El problema que presenta la prueba de Skewness / Kurtosis es que es muy susceptible al tamaño de la muestra, por lo que con muestras grandes la probabilidad entregada del valor p siempre será pequeña y con muestras

pequeñas tiende a arrojar una probabilidad grande.

Por lo que este test acepta su uso con tamaños entre 30 a 60 valores por variable. La interpretación de los valores de p entregados por el programa estadístico para

las pruebas Skewness / Kurtosis, Shapiro Wilk, Shapiro – Francia y Kolmogorov – Smirnov es que si el valor es mayor o igual a 0,05 sí existe normalidad y si es menor la distribución es no normal. La prueba de Kolmogorov Smirnov no es recomendada por el programa estadístico STATA debido a que no es lo suficientemente poderosa para determinar normalidad, sin embargo, por convención se ha asumido que se use cuando se tienen más de 50 valores.

Y a su vez, cuando se tienen menos de 50 valores en la variable se recomienda utilizar la prueba de normalidad Shapiro Wilk debido a que estas ajustan mejor con este tipo de tamaños.

A su vez, se debe siempre de contrastar con los métodos gráficos ya que el valor p analizado únicamente a través de los métodos numéricos son susceptibles a los tamaños de la muestra.

### 3.2. Métodos gráficos

Los métodos gráficos también sirven para contrastar los valores entregados por los métodos numéricos, dentro de los que se recomiendan se encuentran los mencionados en la Tabla 2.

Tabla 2. Métodos gráficos para evaluar normalidad.	
Métodos gráficos	Sintaxis
Tallos y hojas	stem variable
Histograma	histogram variable, normal
Gráfico P-P	pnorm variable
Gráfico Q-Q	qnorm variable
Densidad de Kernel	kdensity variable, normal
Gráfico de dispersión	dotplot variable
Gráfico de cajas	graph box variable

La forma para aceptar la normalidad a través de este método es que la gráfica entregada se asemeje lo más posible a la campana de Gauss.

### 3.3. Métodos de transformaciones

Cuando los datos no presentan una distribución normal se pueden transformar para obtener la distribución. Para poder acceder se debe incorporar el comando ladder variable, la cual nos entregará la información de la Tabla 3

Transformaciones
Cúbica
Cuadrática
Identidad
Raíz cuadrada
Logarítmica
1 / (raíz cuadrada)
Inversa
1 / cuadrática
1 / cúbica

El valor de p presenta la misma interpretación que el método numérico. Por otro lado, el método de transformaciones también posee una representación gráfica a través del **comando** *gladder* variable, el cual arrojará una serie de gráficos para cada una de las transformaciones mencionadas. Se seleccionará aquel gráfico que represente de mejor forma la campana gaussiana para aceptar la normalidad de los datos.

## IV. Consideraciones finales

Se sugiere considerar los siguientes aspectos:

1. Los parámetros de la distribución normal son el promedio y la desviación estándar.
2. La distribución es una función de la distribución estándar.
3. El tamaño de muestra juega un rol en la distribución normal.
4. La distribución normal de los datos puede ser corroborada con el uso de pruebas.

Es importante resaltar que las variables cuya naturaleza es cuantitativa (discretas, continuas) son las que pueden ser sometidas a análisis de promedios y desviaciones estándar y para las variables cualitativas (nominales, ordinales) se pueden utilizar otro tipo de análisis como las medianas

#### ***V. Conclusiones***

Los datos presentan una distribución la cual debe ser confirmada con test estadísticos a través de métodos numéricos y gráficos. Cuando los datos no presentan una distribución normal se pueden utilizar métodos que permitan su transformación. Conocer la distribución normal de nuestros datos permite el uso correcto de pruebas paramétricas y no paramétricas para el análisis de los datos obteniendo mejores resultados.

#### ***VI. Referencias***

1. Droppelmann G. Importancia de la Distribución Normal. Revista Archivos de la Sociedad Chilena de Medicina del Deporte. 2017;62(2):5-6.
2. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. Am J Physiol Endocrinol Metab. 2004;287(2):E189–91.
3. Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for NonStatisticians. International Journal of Endocrinology and Metabolism. 2012;10(2):486-489.
4. Altman D, Bland M. The normal distribution. BMJ 1995. 1995:298. 5. Indiana University. Knowledge Base. Consultado el 3 de mayo del 2018. Disponible en: <https://kb.iu.edu/d/alug> 6.

#### **Guía de requerimientos**

- Título
- Fecha
- Correo:
- Contenido
- Resumen
- Autor/Alumno
- Palabras Clave
- Introducción
- .Desarrollo del tema
- Métodos Numéricos
- Métodos gráficos
- Comandos en SOFTWARE.
- Conclusiones
- Referencias