

Universidad Panamericana
Maestría en Ciencia de Datos
Econometría

Manual de Supuestos de Regresión:
Prueba de Jarque-Bera

Enrique Ulises Báez Gómez Tagle

17 de septiembre de 2025

Índice

1	Introducción	2
2	Preparación de Datos	2
3	Normalidad	2
4	Interpretación de los Resultados	3
5	Solución a las Violaciones de Supuestos	3
6	Link al repositorio con código fuente	4

1. Introducción

En regresión lineal, la normalidad de los residuos es un supuesto que asegura la validez de inferencias como pruebas t y F . Si los residuos no son normales, las conclusiones sobre significancia pueden ser incorrectas, sobre todo en muestras pequeñas. Aunque el teorema central del límite mitiga este efecto en grandes muestras, en la práctica es necesario verificarlo. La prueba de Jarque-Bera, que analiza asimetría y curtosis, es una de las herramientas más usadas para evaluar la normalidad tras ajustar un modelo de regresión.

2. Preparación de Datos

Para aplicar la prueba de Jarque-Bera en Python, es necesario seguir una serie de pasos de preparación de datos:

1. **Cargar los datos:** Utilizar la librería `pandas` para importar el conjunto de datos, por ejemplo desde un archivo CSV.
2. **Ajustar un modelo de regresión:** Emplear `statsmodels` para ajustar un modelo de regresión lineal simple o múltiple. Es importante asegurarse de agregar una constante al conjunto de predictores si se desea incluir el intercepto.
3. **Obtener los residuos:** Extraer los residuos del modelo ajustado, que serán utilizados en la prueba de normalidad.

A continuación se muestra un ejemplo básico de estos pasos en Python:

```
import pandas as pd
import statsmodels.api as sm

# Paso 1: Cargar datos
datos = pd.read_csv('ruta/archivo.csv')
X = datos[['variable_independiente']]
y = datos['variable_dependiente']

# Paso 2: Ajustar modelo de regresion
X = sm.add_constant(X) # Agregar intercepto
modelo = sm.OLS(y, X).fit()

# Paso 3: Extraer residuos
residuos = modelo.resid
```

Estos residuos serán la base para realizar la prueba de normalidad de Jarque-Bera.

3. Normalidad

En muestras pequeñas, la falta de normalidad puede invalidar las pruebas estadísticas. Una de las pruebas más utilizadas para verificarlo es la **prueba de Jarque-Bera**, que evalúa la asimetría (skewness) y la curtosis (kurtosis) de los residuos.

Hipótesis de la prueba de Jarque-Bera:

- **Hipótesis nula (H_0):** Los residuos siguen una distribución normal.
- **Hipótesis alternativa (H_1):** Los residuos no siguen una distribución normal.

La estadística de Jarque-Bera se calcula como:

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right)$$

donde n es el tamaño de la muestra, S es la asimetría muestral y K es la curtosis muestral *no de exceso* (bajo normalidad $K = 3$). Asintóticamente, $JB \sim \chi^2_{(2)}$, por lo que la decisión puede tomarse vía el valor p o comparando con el cuantil crítico de una χ^2 con 2 g.l.

Implementación en Python:

```
import pandas as pd
import statsmodels.api as sm
from scipy.stats import jarque_bera
from statsmodels.stats.stattools import jarque_bera as jb_sm

# Carga de datos
datos = pd.read_csv('ruta/archivo.csv')
X = datos[['variable_independiente']]
y = datos['variable_dependiente']
X = sm.add_constant(X)          # Agrega el intercepto.
modelo = sm.OLS(y, X).fit()
residuos = modelo.resid

# JarqueBera con SciPy
jb_stat, jb_pvalue = jarque_bera(residuos)
print(f'JB: {jb_stat:.4f}')
print(f'p-valor: {jb_pvalue:.6f}')
alpha = 0.05

if jb_pvalue < alpha:
    print('Rechazamos H0: los residuos no son normales.')
else:
    print('No rechazamos H0: no hay evidencia contra la normalidad.')
jb_stat_sm, jb_pvalue_sm, skew, kurt = jb_sm(residuos)

print(f'(statsmodels) JB: {jb_stat_sm:.4f}, p-valor: {jb_pvalue_sm:.6f},
      skew: {skew:.4f}, kurtosis: {kurt:.4f}')
```

4. Interpretación de los Resultados

La interpretación de la prueba de Jarque-Bera se centra en el valor p .

- Si $p < 0.05$, se rechaza la hipótesis nula y se concluye que los residuos no son normales.
- Si $p \geq 0.05$, no se rechaza la hipótesis nula y los residuos pueden considerarse normales.

En muestras grandes, la prueba puede detectar desviaciones pequeñas sin importancia práctica, mientras que en muestras pequeñas podría no detectarlas. Por ello, se recomienda complementar la prueba con gráficos (histograma, QQ plot) y con el conocimiento del contexto.

Además del valor p , es recomendable fijarse en el propio estadístico JB y en los valores de asimetría (skewness) y curtosis. Un estadístico JB pequeño junto con un skew cercano a 0 y una curtosis cercana a 3 refuerzan la evidencia de normalidad. En cambio, valores alejados de estos puntos de referencia sugieren desviaciones que deben ser evaluadas en conjunto con el contexto y los gráficos.

5. Solución a las Violaciones de Supuestos

Si la prueba de Jarque-Bera indica que los residuos no son normales (es decir, se rechaza la hipótesis nula), existen varias estrategias para abordar esta violación del supuesto:

- **Transformaciones de variables:** Aplicar transformaciones a la variable dependiente o a las independientes (por ejemplo, logaritmo, raíz cuadrada o Box-Cox) puede ayudar a normalizar los residuos.

- **Modelos robustos:** Utilizar métodos de estimación robustos, como la regresión robusta o mínimos cuadrados ponderados, que son menos sensibles a la no normalidad.
- **Métodos no paramétricos:** Considerar técnicas no paramétricas que no requieren el supuesto de normalidad, como los modelos de regresión cuantílica.
- **Revisión de datos atípicos:** Identificar y tratar valores atípicos o leverage points que puedan estar afectando la distribución de los residuos.

La elección de la solución dependerá del contexto del análisis, los objetivos del estudio y la naturaleza de los datos.

6. Link al repositorio con código fuente

En el siguiente link se puede encontrar el código fuente en Python, Julia, R, Matlab <https://github.com/enriquegomeztagle/MCD-Econometria/tree/main/HWs/RegressionAssumptionsManual>