Modelos de Regresión y Series de Tiempo (MRST)
2024 - 02

Clase 5 – MRLS
Verificación de supuestos

Docente: Natalia Jaramillo Quiceno
Escuela de Ingenierías
natalia.jaramilloq@upb.edu.co

Regresión lineal simple

Validación de supuestos (adecuación del modelo)

Las principales premisas que se han hecho hasta ahora al estudiar el análisis de regresión son las siguientes:

- La relación entre la respuesta y y la variable regresora es lineal, al menos en una forma aproximada.
- El término de error ϵ tiene media cero
- El término de error ϵ tiene varianza σ^2 constante
- Los errores no están correlacionados
- Los errores tienen una distribución normal – **Se requiere de esta premisa para probar hipótesis y para estimar intervalos.**

Comprobación de la adecuación del modelo $\equiv$ **Verificación de supuestos**

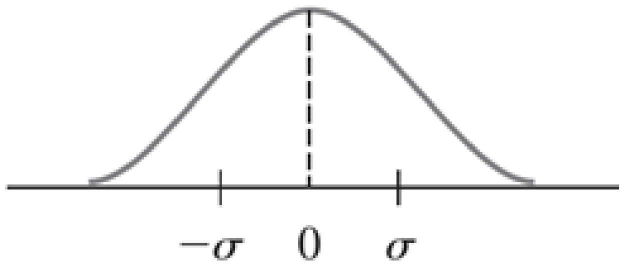
Normalidad y varianza constante de los errores (ε o e)

Regresión lineal simple

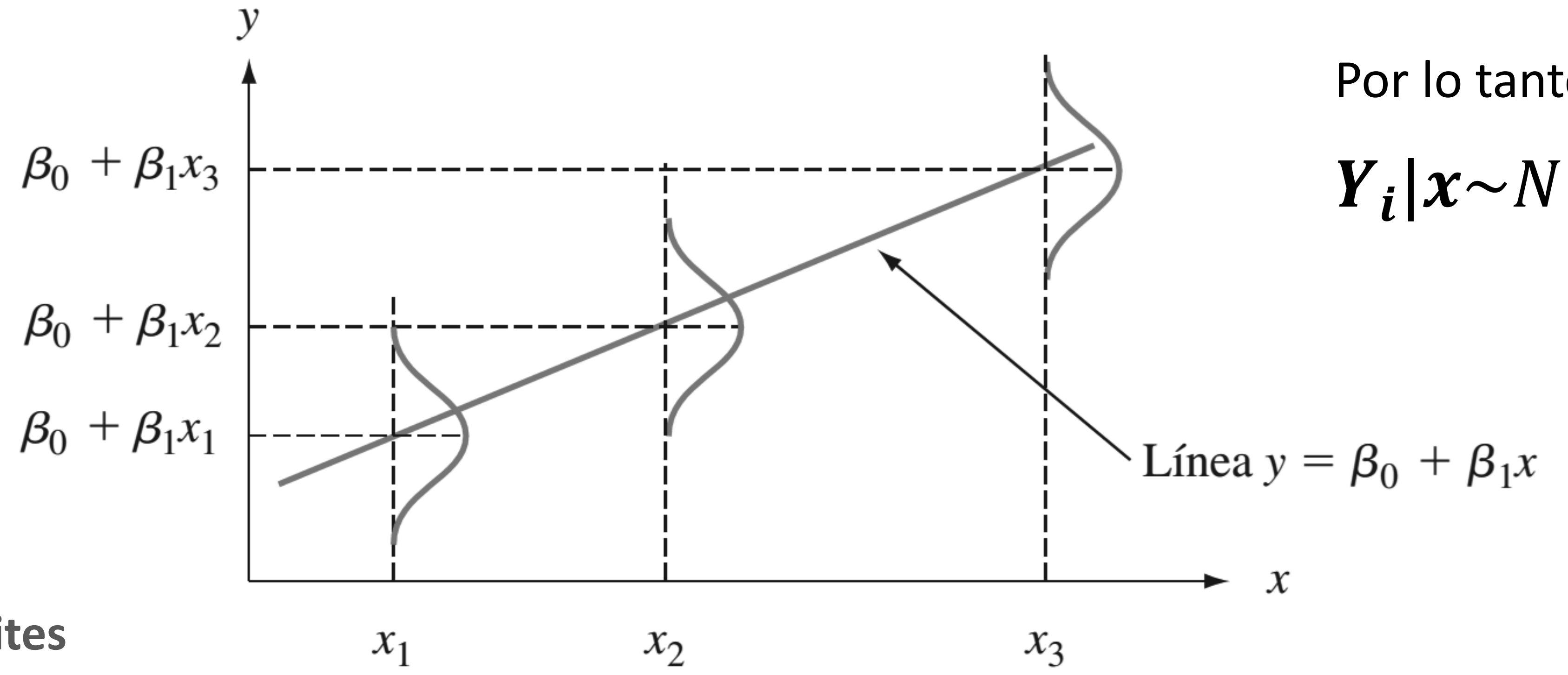
Supuestos del modelo de RLS

$\boldsymbol{\epsilon_i \sim N(0, \sigma^2)}$, independiente e idénticamente distribuidos*

$*Corr(\epsilon_i, \epsilon_j) = 0$ y $Cov(\epsilon_i, \epsilon_j) = 0$ para $i \neq j$



Según el modelo de RLS, Y_i es una función de los ϵ_i, por lo tanto, también es una v.a. cuya media y varianza están dadas por:



Por lo tanto,

$$\boldsymbol{Y_i | x} \sim N(\beta_0 + \beta_1 \text{x}, \sigma^2)$$

Regresión lineal simple
Validación de supuestos (adecuación del modelo)

Los métodos para evaluar el cumplimiento de estas premisas se basan principalmente en el estudio de los **residuales** del modelo.

Recordemos que los residuales de un modelo de regresión son,

$$e_i = y_i - \hat{y}_i, i = 1, 2, \ldots, n$$

En general, los estadísticos de resumen t, F o R^2 son **propiedades "globales"** del modelo, y como tal no aseguran la adecuación del mismo.

Regresión lineal simple

Validación de supuestos – Normalidad de los ε

Prueba de hipótesis:

$$\boldsymbol{H_0}: Los\ errores\ se\ distribuyen\ normal$$

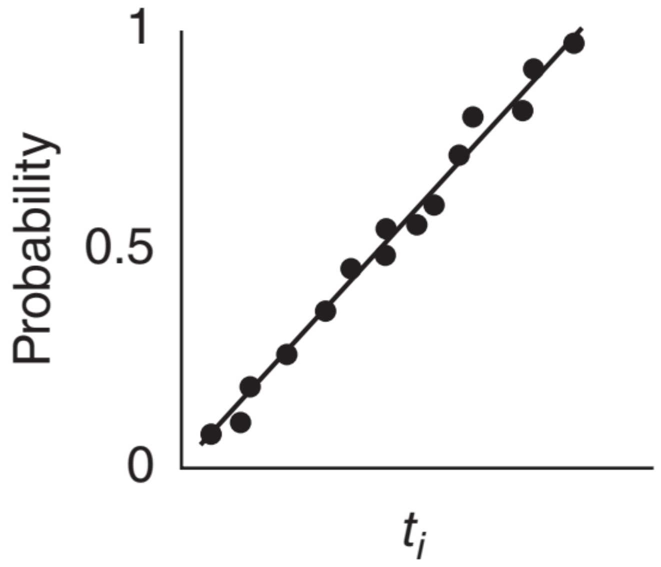
$$\boldsymbol{H_a}: Los\ errores\ no\ se\ distribuyen\ normal$$

Existen dos mecanismos para verificar el supuesto de distribución normal:

a – A través del gráfico "Q-Q" o de probabilidad normal:

Con una inspección visual se identifica si los datos se ajustan a una recta:

CASO IDEAL



b – Mediante test

Pruebas analíticas disponibles:

- Shapiro-Wilks
- Kolmogorov-Smirnov
- Anderson Darling
- Cramér-con Mises

Criterio de decisión

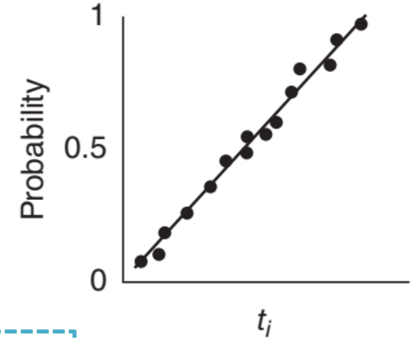
Rechazar $\boldsymbol{H_0}$ si:

$$Valor\ p < \alpha$$

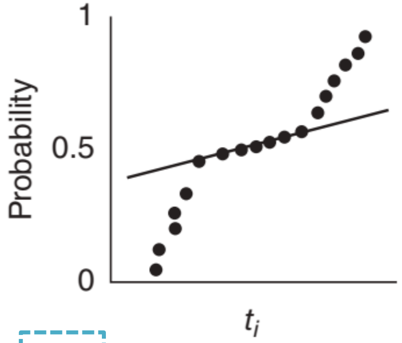
Regresión lineal simple

Validación de supuestos – Normalidad de los ε

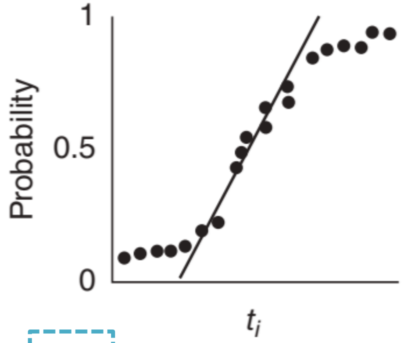
Posibles resultados del gráfico de probabilidad normal



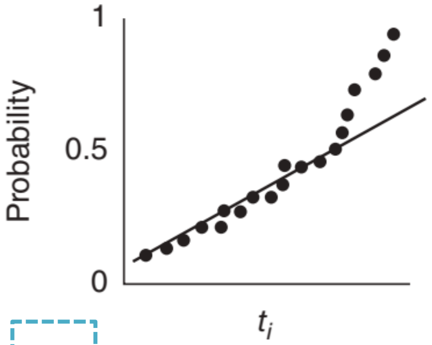
A



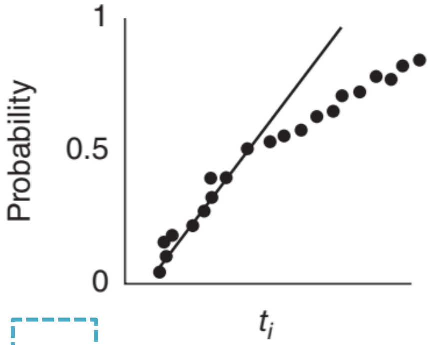
B



C



D



E

A - Caso ideal, se cumple el supuesto

No se cumple el supuesto en:

B - Colas de distribución demasiado gruesas para considerarse como normal.

C – Colas de distribución más delgadas que la normal.

D y E – Asimetría positiva y negativa.

Regresión lineal simple

Validación de supuestos – Normalidad de los ε

Para la verificación aproximada de la **normalidad** se puede construir:

- Histograma de frecuencias de los residuales

También se pueden **estandarizar los errores** como $d_i = e_i / \sqrt{\hat{\sigma}^2}, \; i = 1, 2, \ldots, n.$

- Si los errores siguen una distribución normal, entonces aproximadamente 95% de los residuales estandarizados deberán estar incluidos en el intervalo (-2, +2).

- Los residuales que se aparten de este intervalo pueden indicar la presencia de un punto atípico.

Regresión lineal simple

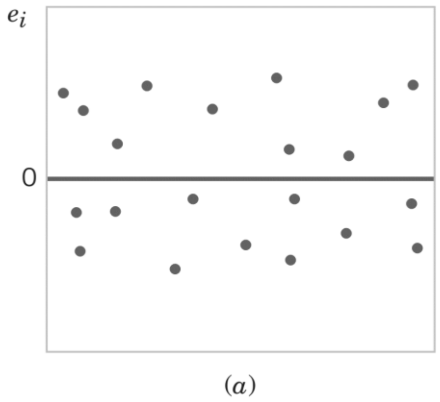
Validación de supuestos – Varianza constante de los ε

Prueba de hipótesis :

$$\boldsymbol{H_0}: Los\ errores\ tiene\ varianza\ constante$$
$$\boldsymbol{H_a}: Los\ errores\ NO\ tienen\ varianza\ constante$$

Principal mecanismo para verificar el supuesto de varianza constante…

CASO IDEAL



(a)

a - A través de alguno de los siguientes gráficos:

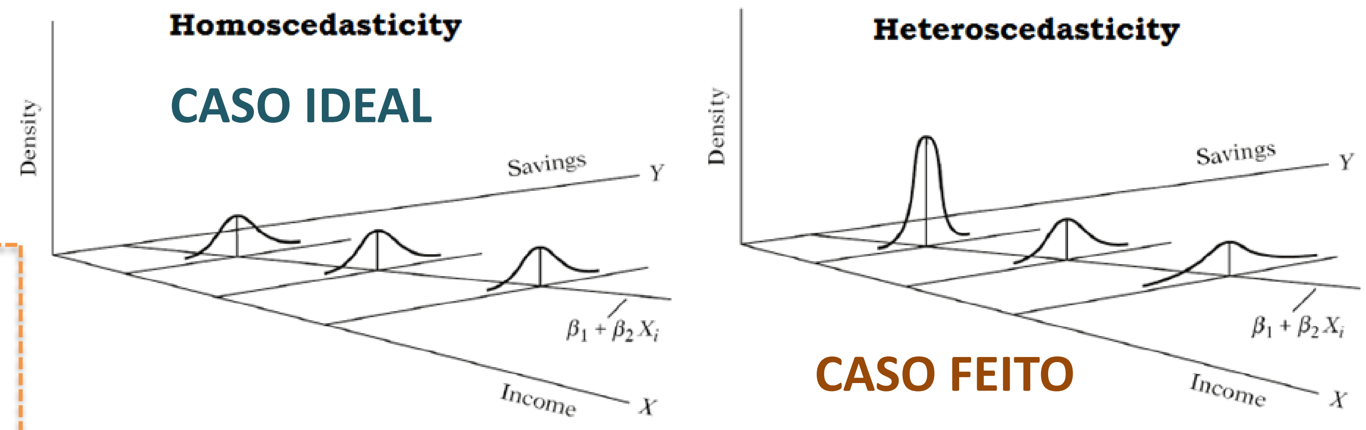
- Residuales en función de x_i.
- Residuales en función de $\hat{y}_i$.

b – Mediante test

- Breusch-Pagan

Criterio de decisión
Rechazar $\boldsymbol{H_0}$ si:

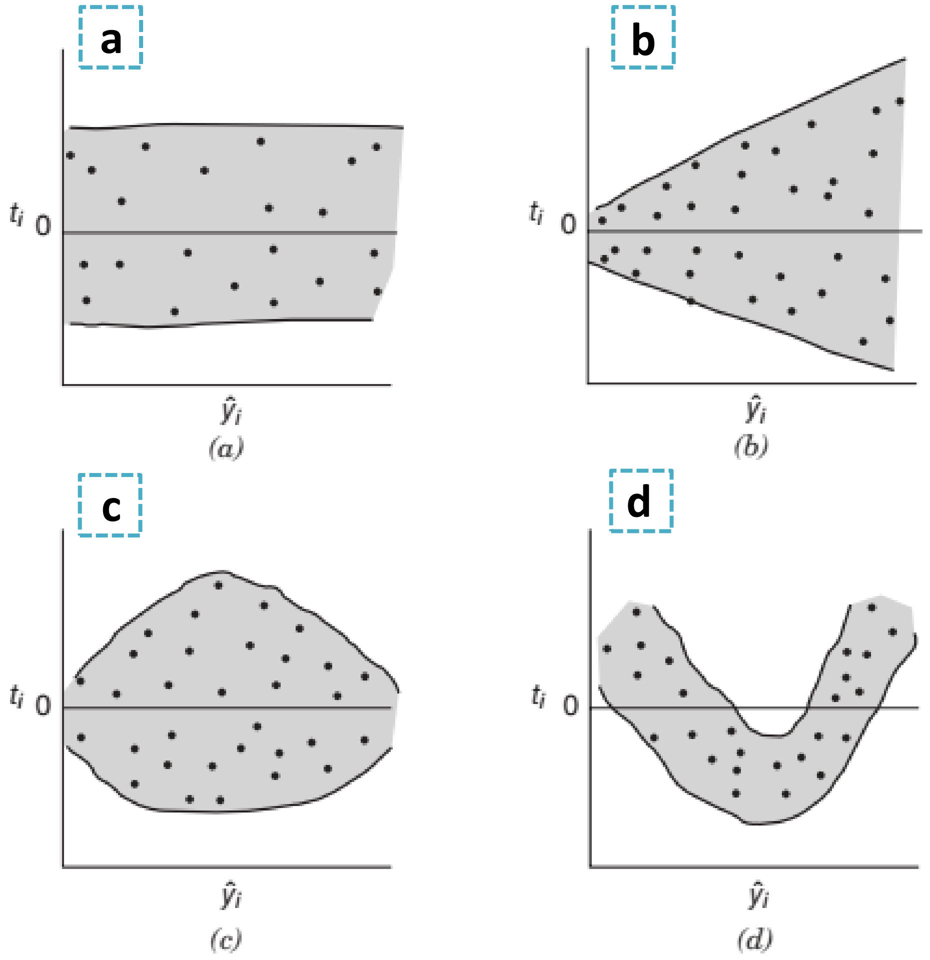
$$Valor\ p < \alpha$$



Regresión lineal simple

Validación de supuestos – Varianza constante de los ε

Se valida utilizando gráficos de residuales vs. predichos o residuales vs. cada variable regresora. Los patrones más comunes son:



a - Caso ideal, se cumple el supuesto

No se cumple el supuesto en :

b, c, y d – La varianza de los $\boldsymbol{e_i}$ parece ser función de los valores de $\boldsymbol{y}$.

Regresión lineal simple

Validación de supuestos en R – Pruebas numéricas

NORMALIDAD

```
        Shapiro-Wilk normality test

data:  data$residual
W = 0.98843, p-value = 0.985
```

Valor p de la prueba
Comparar con α

VARIANZA CONSTANTE

```
Breusch Pagan Test for Heteroskedasticity
-----------------------------------------
Ho: the variance is constant
Ha: the variance is not constant

              Data
-------------------------------
Response : gan
Variables: fitted values of gan

         Test Summary
----------------------------
DF             =   1
Chi2           =   0.1836735
Prob > Chi2    =   0.6682351
```

Valor p de la prueba
Comparar con α

MUCHAS GRACIAS

Natalia Jaramillo Quiceno

e-mail: natalia.jaramilloq@upb.edu.co