Modelos Log – log / Lin – log / Log – lin.

Se le presenta una Base que busca dar cuenta de diversas variables referidas a la calidad de vida y al nivel de desarrollo de los ciudadanos estadounidenses. Si bien las variables se encuentran en Inglés, a continuación se le presenta el significado de cada una de ellas.

Por su parte, se le informa que los tipos de análisis que se esperan de Usted como experto, dicen relación con formas funcionales de modelos de regresión.

Presentación de las variables contenidas en la base ”BigMac2003.sav”:

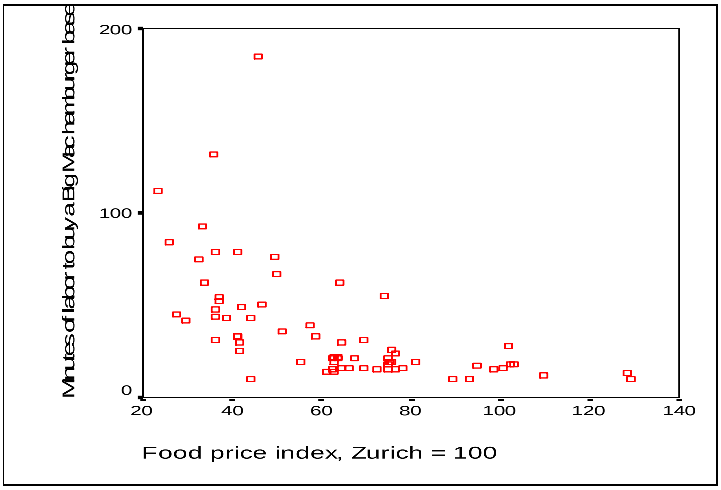
VARIABLE	DESCRIPCIÓN
City	Nombre de la cuidad
BigMac	Minutos de trabajo para comprar una Big Mac basado en un típico salario, promediado en 13 ocupaciones
Bread	Minutos de trabajo para comprar 1kg de pan
Rice	Minutos de trabajo para comprar 1kg de arroz
FoodIndex	Índice de precio de comida, Zurich = 100
Bus	Costo más bajo de 10 km de transporte público
Apt	Renta mensual en Dólares por un típico apartamento de 3 piezas
TeachGI	Sueldo anual bruto en miles de USD de una profesora de primaria
TeachNI	Sueldo anual neto en miles de USD de una profesora de primaria
TaxRate	100*(TeachGI-TeachNI)/TeachGI. En algunas ciudades este indicador es negativo, sugiriendo subsidio, en vez de impuestos
TeachHours	Horas de trabajo semanales de una profesora de primaria

Se le solicita realizar un análisis con los datos, incorporando en cada uno de ellos los planteamientos de supuestos y la sintaxis utilizadas. Entienda en sus análisis, que se le pide analizar en términos de calidad de vida y nivel de desarrollo, lo cual no supondrá un problema para Usted como Magíster en Gestión y Políticas Públicas.

Parte I: Modelo Log-Lin.

1. Realice un gráfico de dispersión de la variable BigMac en el eje vertical y FoodIndex en el eje horizontal. Describa el gráfico.

```
GRAPH
  /SCATTERPLOT(BIVAR)=foodinde WITH bigmac
  /MISSING=LISTWISE .
```



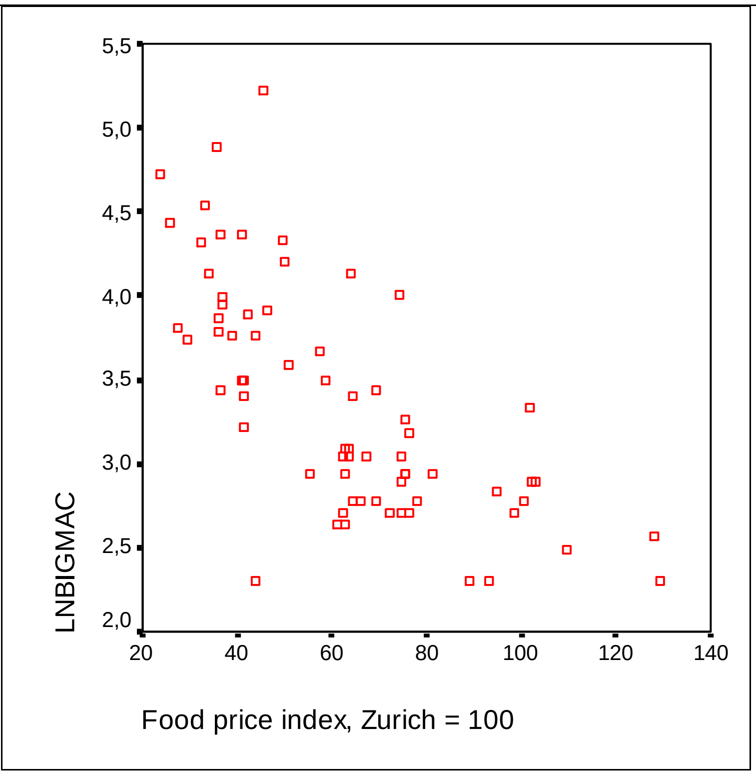
En el gráfico anterior se observa una relación inversa entre el indicador BigMac y el índice de precio de comida. Esto quiere decir, que a mayor valor de BigMac, menor es el valor de FoodIndex.

Esta relación refleja que entre menor es el tiempo de trabajo para comprar una BigMac, mayor es el precio de la comida (tomando como referencia Zurich). Esto claramente refleja que a mayor desarrollo (en los países desarrollados se debe trabajar menos tiempo para comprar un BigMac), la vida es más cara. La relación que se observa parece ser no lineal.

2. Transforme la variable *BigMac* tomando el logaritmo. ¿Le parece que el gráfico de dispersión ahora tiene una función media lineal?

```
COMPUTE LNBIGMAC = LN(bigmac) .
EXECUTE .

GRAPH
  /SCATTERPLOT(BIVAR)=foodinde WITH lnbigmac
  /MISSING=LISTWISE .
```



El gráfico de dispersión deja ver una relación lineal entre la variable transformada BigMac y el FoodIndex. Con la transformación la relación inversa se mantiene. Se podría decir que en el gráfico de dispersión la función media es lineal.

3.- Ajuste un modelo simple log-lin y determine la matriz de correlaciones entre las variables seleccionadas para su modelo, ¿es consistente el coeficiente de correlación con el modelo de regresión?

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT lnbigmac
  /METHOD=ENTER foodinde
  /SCATTERPLOT=(*ZRESID ,*ZPRED )
  /RESIDUALS HIST(ZRESID) NORM(ZRESID)
  /SAVE PRED RESID ZRESID .


CORRELATIONS
  /VARIABLES=lnbigmac foodinde
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .
```

1.

Modelo simple de Regresión simple log – lin.

Resumen del modelo(b)

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,713(a)	,508	,501	,486

a Variables predictoras: (Constante), FOODINDE Food price index, Zurich = 100
b Variable dependiente: LNBIGMAC

ANOVA(b)

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	16,340	1	16,340	69,146	,000(a)
	Residual	15,833	67	,236		
	Total	32,173	68			

a Variables predictoras: (Constante), FOODINDE Food price index, Zurich = 100
b Variable dependiente: LNBIGMAC

Coeficientes(a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	4,598	,160		28,804	,000
	FOODINDE Food price index, Zurich = 100	-,020	,002	-,713	-8,315	,000

a Variable dependiente: LNBIGMAC

<u>El modelo de regresión simple log lin de regresión queda como sigue:</u>

$$\boxed{LNBIGMAC = 4{,}598 + (-0{,}020)*FOODINDEX}$$

Matriz de Correlaciones.

Correlaciones

		LNBIGMAC	FOODINDE Food price index, Zurich = 100
LNBIGMAC	Correlación de Pearson	1	-,713(**)
	Sig. (bilateral)	.	,000
	N	69	69
FOODINDE Food price index, Zurich = 100	Correlación de Pearson	-,713(**)	1
	Sig. (bilateral)	,000	.
	N	69	69

** La correlación es significativa al nivel 0,01 (bilateral).

Como sabemos, la hipótesis nula de una prueba de correlación, plantea que no existe relación entre las variables. En nuestro caso, se ve que existe evidencia estadística para rechazar dicha hipótesis, en tanto la significancia es de 0,000 (menor a 0,05). Luego, sí existe correlación entre las variables, la cual es negativa y media alta con un valor de - 0,713, el que coincide con el valor del estadístico R de un modelo de regresión simple (en valor absoluto). Por tanto, el coeficiente de correlación es consistente con lo que plantea nuestro modelo de regresión.

4.- Interprete el R2 e interprete la pendiente.

R2.

El R2 del modelo, nos indica que el 50,8% de la variabilidad del ln(BIGMAC) está siendo explicada por el índice de precios.

Pendiente.

Por su parte, la pendiente del modelo, nos expresa que ***"los minutos de trabajo para comprar una Big Mac basado en un típico salario, promediado en 13 ocupaciones disminuye en un 2%, por cada punto de aumento en el Índice de precio de comida".***

5.- ¿Son los coeficientes del modelo significativos? Argumente su respuesta.

Recordemos que estamos testeando la hipótesis nula de que los betas son iguales a cero, lo cual implicaría que no son significativos.

En otros términos:

$$H0 : b1 = 0,$$
$$H1 : b1 \text{ es distinto de cero.}$$

ANOVA(b)

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	16,340	1	16,340	69,146	,000(a)
	Residual	15,833	67	,236		
	Total	32,173	68			

a Variables predictoras: (Constante), FOODINDE Food price index, Zurich = 100
b Variable dependiente: LNBIGMAC

Coeficientes(a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	4,598	,160		28,804	,000
	FOODINDE Food price index, Zurich = 100	-,020	,002	-,713	-8,315	,000

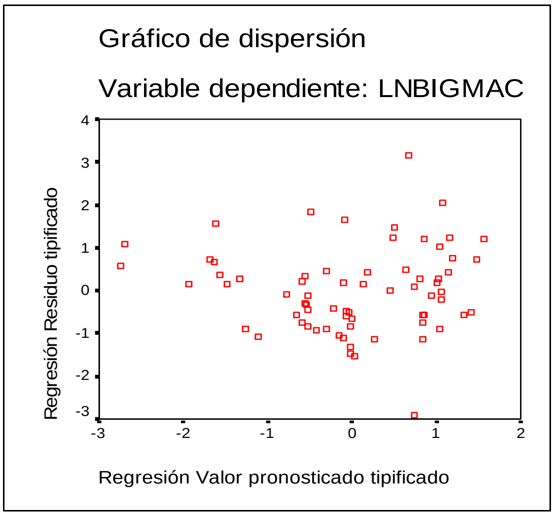
a Variable dependiente: LNBIGMAC

Como vemos tanto en ANOVA, como en la tabla de coeficientes, existe evidencia estadística para rechazar la hipótesis nula, en tanto la significancia de ANOVA como la significancia de los coeficientes es menor a 0,05, presentando un correlato entre los estadísticos. Recordemos que podemos usar ANOVA, solo por tratarse de un modelo simple. De lo contrario, veríamos directamente la tabla de coeficientes.

6.- Evalúe si se cumplen los supuestos de homocedasticidad y normalidad.

Supuesto de homocedasticidad.

H0: La varianza de los residuos es homogénea.
H1: La varianza de los residuos no es homogénea.



Vemos que, en general, no se presentan patrones en el comportamiento de los residuos, por lo que podríamos decir que, en términos generales, se cumple la hipótesis de homocedasticidad de los residuos.

Supuesto de Normalidad.

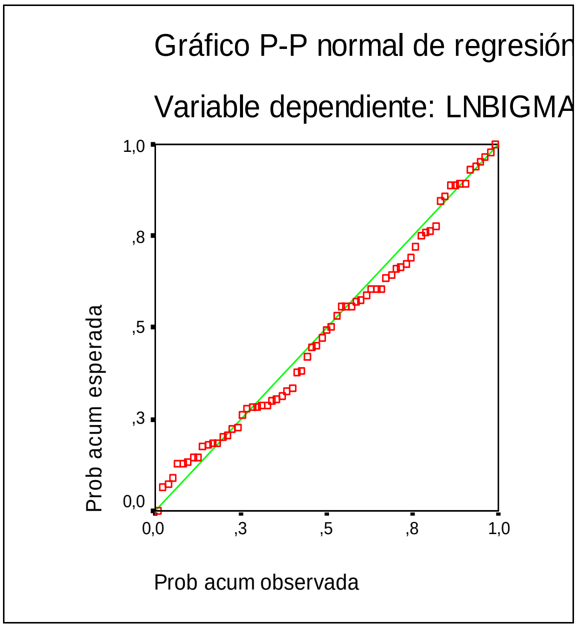
H0: Los residuos distribuyen normal
H1: Los residuos no distribuyen normal.

```
EXAMINE
  VARIABLES=zre_1
  /PLOT BOXPLOT NPPLOT
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Pruebas de normalidad

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Standardized Residual	,073	69	,200(*)	,976	69	,218

* Este es un límite inferior de la significación verdadera.
a Corrección de la significación de Lilliefors

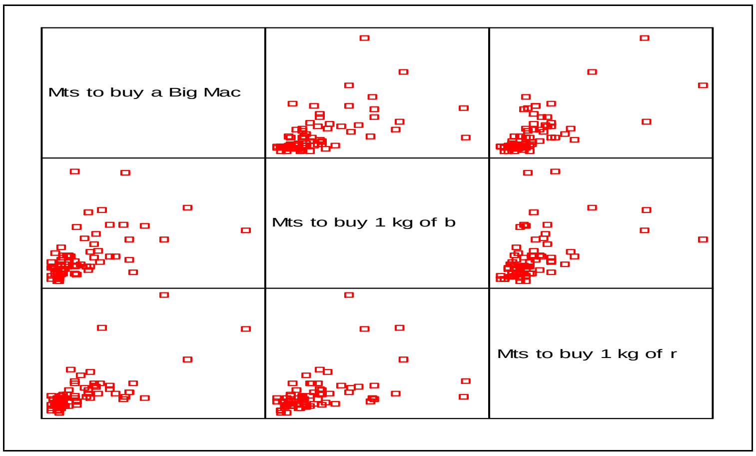


Se observa que los residuos presentan un comportamiento normal. Lo anterior, al corroborar el valor que arroja la prueba de K-S que presenta una significancia mayor a 0,05, pudiendo aceptar estadísticamente la hipótesis de normalidad de los residuos.

Parte II: Modelo Log-Log.

1.- Realice gráficos de dispersión de las variables BigMac, Bread y Rice. Describa los gráficos.

```
GRAPH
  /SCATTERPLOT(MATRIX)=bigmac bread rice
  /MISSING=LISTWISE .
```



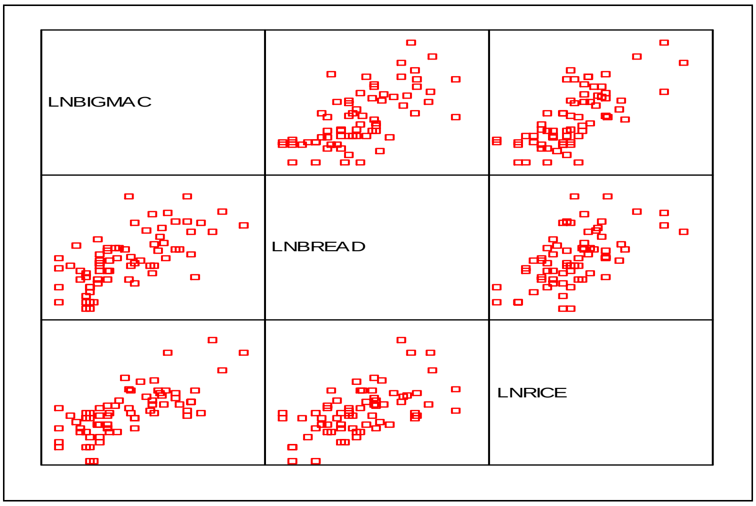
Se observa que, en términos generales, los tres pares de variables, presentan una tendencia lineal positiva. De los gráficos de dispersión se puede concluir que existe una relación lineal entre el indicador BigMac y el indicador Rice así como también entre el indicador BigMac y el indicador Bread. Es decir, los minutos de trabajo necesarios para comprar una BigMac están relacionados en forma lineal directa con los minutos de trabajo necesarios para comprar 1 kg de pan y de arroz. Sin embargo, se observa en la relación lineal una variabilidad. A mayor valor de BigMac y de Bread y Rice, mayor es la variabilidad de la recta media de ajuste. Esto claramente da indicios de la existencia de heterocedasticidad en los datos.

2.- Compare las variables en su forma logarítmica a través de un gráfico que muestre la dispersión entre ellas. ¿Le parece que los gráficos ahora tienen una función media lineal?

```
COMPUTE LNBREAD = LN(bread) .
EXECUTE .

COMPUTE LNRICE = LN(rice) .
EXECUTE .

GRAPH
  /SCATTERPLOT(MATRIX)=lnbigmac lnbread lnrice
  /MISSING=LISTWISE .
```



Se observa que la relación entre ln(BigMac) y ln(Bread) y ln(Rice) toma una forma más lineal. Además, se aprecia que la heterocedasticidad es menos pronunciada que en el gráfico de dispersión anterior.

3.- Ajuste un modelo log-log de regresión múltiple. Interprete la bondad de ajuste del modelo y los coeficientes.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT lnbigmac
  /METHOD=ENTER lnbread lnrice
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*ZRESID ,*ZPRED )
  /RESIDUALS HIST(ZRESID) NORM(ZRESID)
  /SAVE PRED RESID ZRESID .
```

Resumen del modelo(b)

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,774(a)	,599	,586	,442

a Variables predictoras: (Constante), LNRICE, LNBREAD
b Variable dependiente: LNBIGMAC

Coeficientes(a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	T	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	,510	,292		1,744	,086		
	LNBREAD	,391	,104	,366	3,762	,000	,642	1,558
	LNRICE	,599	,117	,497	5,104	,000	,642	1,558

a Variable dependiente: LNBIGMAC

De este modo, el modelo de regresión múltiple queda de la forma:

LNBIGMAC = 0,510 + (0,391 * LN BREAD) + (0,599 * LN RICE)

Interpretación del R2 corregido.

Al tratarse de un modelo múltiple, nos fijamos en el valor que toma el R2 corregido (aunque entre este estadístico, y el R cuadrado tiende a observarse consistencia). En este caso, observamos que éste toma un valor de 0,586, lo que refleja que, tomadas juntas (tanto el log de BREAD como el log de RICE) las dos variables incluidas en el análisis explican el 58,6% de la variabilidad del logaritmo del tiempo de trabajo que toma en minutos comprar una Big Mac.

Interpretación los coeficientes del modelo.

Recordemos que un modelo log log, nos explica lo que en economía se conoce como "elasticidad punto", la cual ve el cambio porcentual en un punto determinado de una función.

Debemos tener cuidado en la interpretación, por tratarse de un modelo múltiple y del tipo log – log. Así, por ejemplo, para el coeficiente del logaritmo de BREAD, que equivale a 0,391, tenemos que éste indica que, "manteniendo ***el resto de términos de la ecuación constantes, si el tiempo en minutos que toma comprar un kilo de pan aumenta en un 1%, en promedio, el tiempo en minutos requerido para comprar una BigMac aumenta en un 0,391%".***

Para el caso de la segunda regresora (LN RICE Con beta 0,599), tenemos que ***"si el resto de términos de la ecuación se mantienen constantes, si el tiempo en minutos de trabajo que toma comprar un kilo de arroz, aumenta en un 1%, en promedio, el tiempo en minutos requerido para comprar un BigMac aumenta en un 0,599%".***

4.- ¿Son los coeficientes del modelo significativos? Argumente su respuesta.

H0 : b1 = b2 =0,
H1 : algún beta es distinto de cero.

Coeficientes(a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	,510	,292		1,744	,086		
	LNBREAD	,391	,104	,366	3,762	,000	,642	1,558
	LNRICE	,599	,117	,497	5,104	,000	,642	1,558

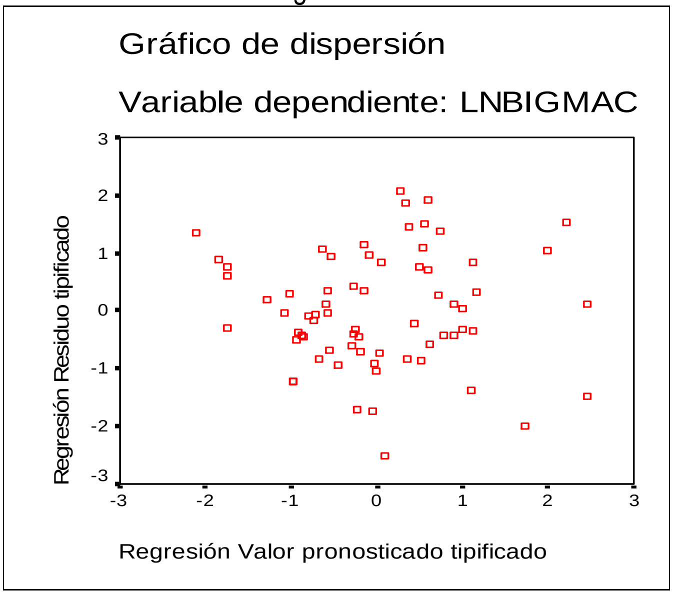
a Variable dependiente: LNBIGMAC

Observamos que solo en uno de los casos (en el caso de la constante) se acepta la hipótesis nula. Sin embargo, apreciamos que la significancia de las dos regresoras es estadísticamente significativa.

5.- Determine si se cumplen los supuestos de homocedasticidad y normalidad.

Supuesto de homocedasticidad.

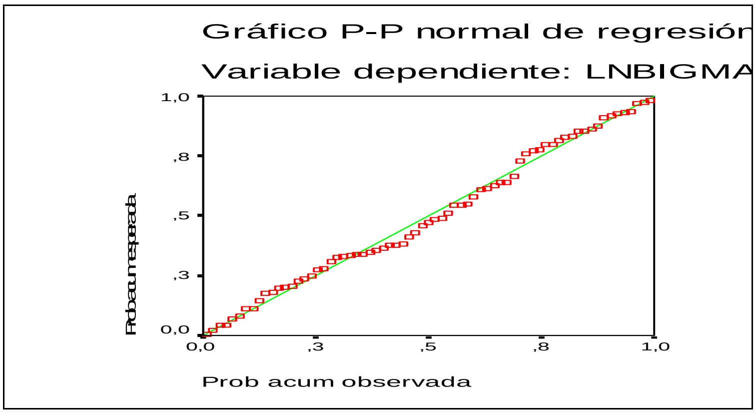
H0: La varianza de los residuos es homogénea.
H1: La varianza de los residuos no es homogénea.



Del gráfico se observa que los residuos en general tienden a presentar un comportamiento de tipo homocedástico, en tanto la nube de puntos se distribuye más bien de forma homogénea en torno al valor de cero, y se encuentran entre -3 y +3. En base a lo anterior, habría evidencia estadística para decir que se cumple la hipótesis nula de homocedasticidad de los residuos.

Supuesto de Normalidad.

H0: Los residuos distribuyen normal
H1: Los residuos no distribuyen normal.




```
EXAMINE
  VARIABLES=zre_2
  /PLOT BOXPLOT NPPLOT
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Pruebas de normalidad

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Standardized Residual	,070	69	,200(*)	,990	69	,870

* Este es un límite inferior de la significación verdadera.
a Corrección de la significación de Lilliefors

En base, tanto al gráfico p –p, como a la prueba de Kolmogorov – Smirnov, podemos decir que existe evidencia estadística para aceptar la hipótesis nula de normalidad de los residuos.

6.- Determine si se satisface la no colinealidad de las variables independientes.

Coeficientes(a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	,510	,292		1,744	,086		
	LNBREAD	,391	,104	,366	3,762	,000	,642	1,558
	LNRICE	,599	,117	,497	5,104	,000	,642	1,558

a Variable dependiente: LNBIGMAC

Diagnósticos de colinealidad(a)

Modelo	Dimensión	Autovalor	Indice de condición	Proporciones de la varianza		
				(Constante)	LNBREAD	LNRICE
1	1	2,961	1,000	,00	,00	,00
	2	,023	11,441	,93	,38	,06
	3	,016	13,459	,07	,62	,93

a Variable dependiente: LNBIGMAC

Observamos en base a los estadísticos de colinealidad que la tolerancia alcanza un valor del 64,2%, lo que estaría explicando que el 64,2% de la variabilidad de una regresora No es explicada por la otra regresora incorporada en el modelo. En otros términos, más de la mitad de la variabilidad de ln(Rice) no es explicada por ln(Bread) y viceversa. En principio esto indica que no hay multicolinealidad.

En la tabla de diagnósticos de colinealidad se observa que ningún índice de condición es mayor a 15 y, por lo tanto, se concluye que no existe colinealidad entre las variables independientes.

<u>**Formalidades:**</u>

1. Plazo de entrega: viernes 16 de abril de 2010, durante la clase auxiliar. Entregar en formato impreso. **No enviar por e-mail.**
2. Se debe realizar de a dos personas.
3. Consultas pueden realizarse por correo electrónico al profesor auxiliar, hasta el día martes 13 a las 12:00 horas.