

VALIDACION DE LA REGRESION MEDIANTE EL ANALISIS DE HOMOCEASTICIDAD

Inq. Roberto Piol, M.Sc. SOITAVE 260 / UPAV 94

Resumen: Continuadamente, se ha minimizado u omitido la importancia del análisis de Homocedasticidad de la data en la aplicación de las técnicas de regresión múltiple. En los libros de econometría, el tratamiento a este importante supuesto o condición; se ha reducido a una muy completa explicación teórica y raramente a sus consecuencias y significancia práctica. Por definición, la regresión solamente existirá para el sector de la data que se comporta "Homocedasticamente"; mientras que los coeficientes del modelo calculado, obviamente no explicarán correctamente el fenómeno estudiado en el sector "Heterocedástico". Se pretende en este Trabajo de Investigación, ofrecer al profesional tasador un nuevo criterio para la validación de los resultados de una regresión; de manera que pueda determinar la sección de la curva donde el modelo estudiado genere resultados valederos.

Palabras Clave: Homocedasticidad, Heterocedasticidad, Homocedástico, Heterocedástico, Regresión

Índice:

I. Fundamentos Teóricos sobre la Homocedasticidad y la Heterocedasticidad, Pag.

II. Determinación de la Heterocedasticidad en un Modelo de Regresión Múltiple, Pag.

III Ejemplo Demostrativo, Pag.

IV Conclusiones y Recomendaciones, Pag.

V Bibliografía

I. Fundamentos Teóricos sobre la Homocedasticidad y la Heterocedasticidad

I.1 Definiciones

Se define que un Modelo de Regresión Múltiple es “Homocedástico”, si la “Varianza del Error” de la regresión ES CONSTANTE para la totalidad de la Data.

Por lo tanto, si la “Varianza del Error” de la regresión, NO ES CONSTANTE a lo largo de la Data; se definirá como “Heterocedástico”.

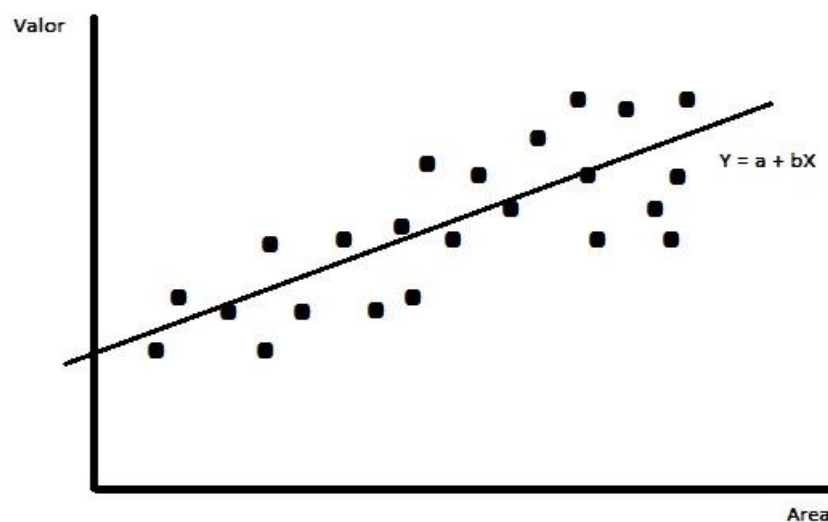
Varianza del Error: Conjunto de irregularidades o variaciones aleatorias que presenta toda data y se define como el cuadrado de la Desviación Estándar de la serie.

En otras palabras, para que un Modelo de Regresión Múltiple sea “Homocedástico”, se deberá cumplir que:

$$Var\left(\frac{y_1 - \hat{y}_1}{x_1}\right) = Var\left(\frac{y_2 - \hat{y}_2}{x_2}\right) = Var\left(\frac{y_n - \hat{y}_n}{x_n}\right) = \sigma^2$$

I.2 Representación Gráfica del Concepto de Homocedasticidad

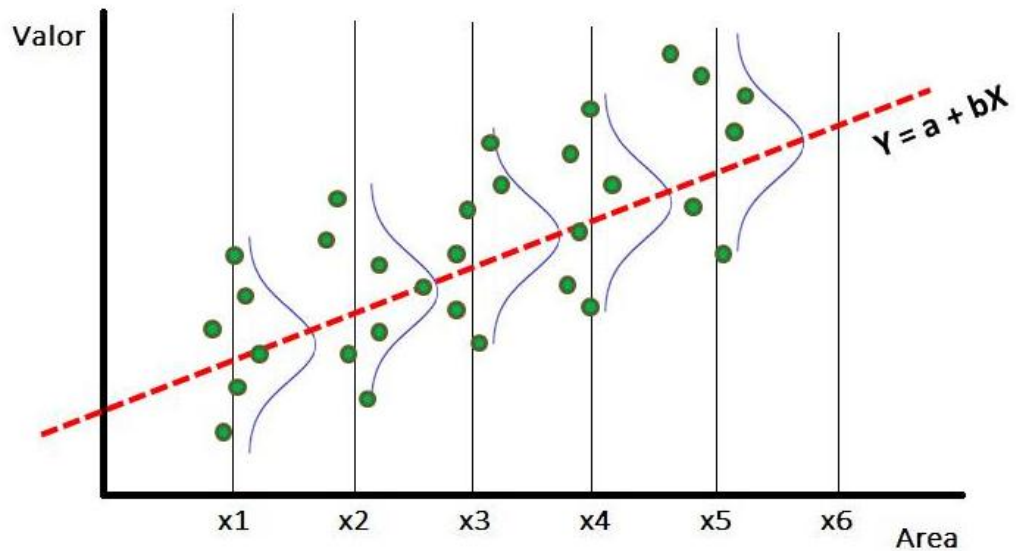
Sea el siguiente Modelo de Regresión (*):



(*) Si bien, para facilitar la explicación; el gráfico se corresponde a un Modelo de Regresión Lineal Simple. Los conceptos teóricos aquí expresados, pueden ser aplicados a cualquier tipo de ecuación con Dos (2) o más variables.

Obsérvese de que la gran mayoría de los pares ordenados (x,y), están aproximadamente equidistantes de la Curva de Regresión $Y = a + bX$.

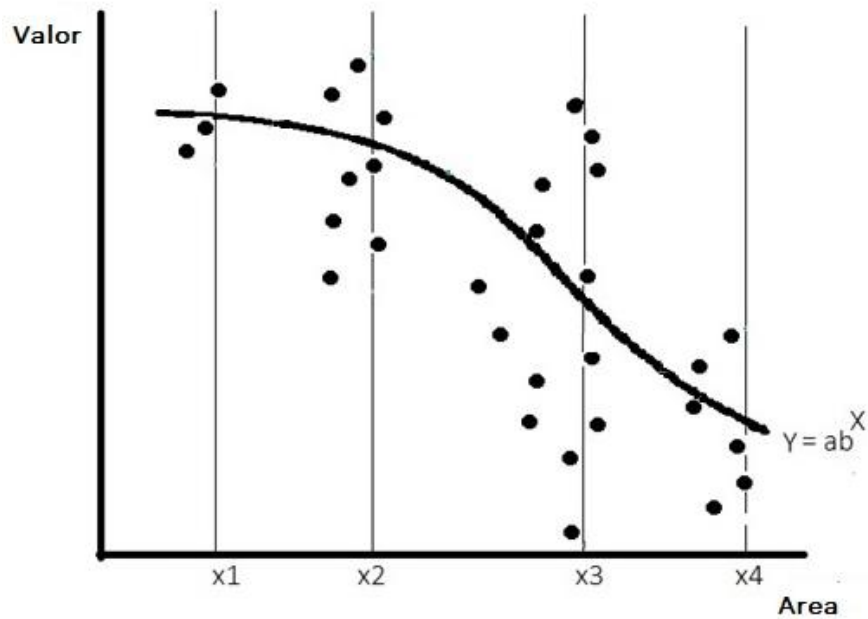
Por lo tanto, al analizar diferentes valores de X a lo largo de la data; podemos observar que los pares ordenados (x,y), pareciera que se distribuyen muy similarmente para cada uno de los valores de X que hemos seleccionado:



Obsérvese en el gráfico que todas las campanas dibujadas, tienen aproximadamente misma forma, es decir: “La misma dispersión”. Concluyendo que el Modelo de Regresión es Homocedástico.

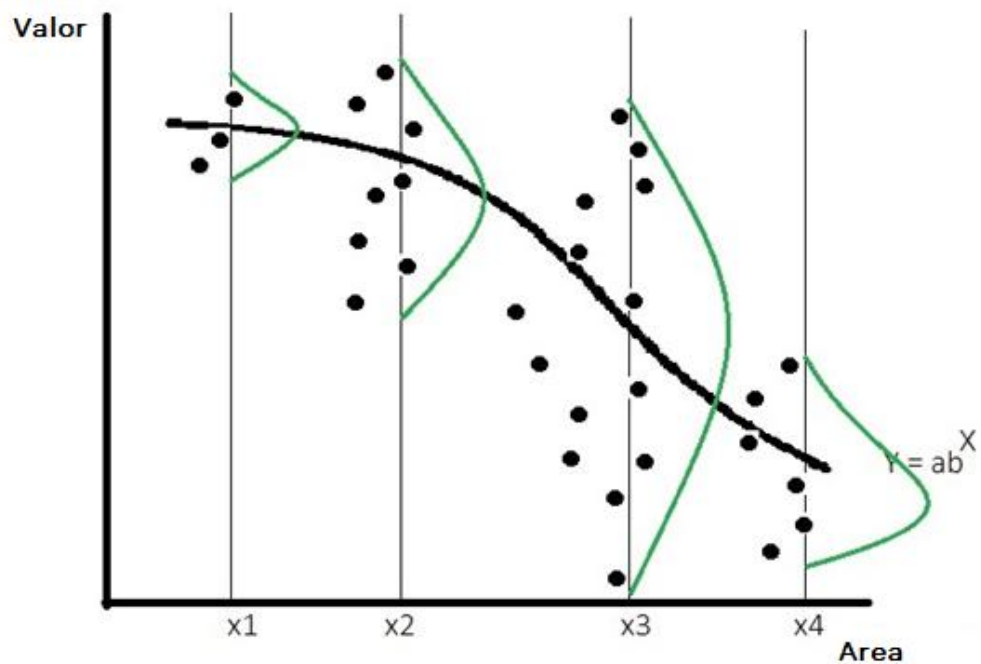
1.3 Representación Gráfica del Concepto de Heterocedasticidad

Sea el siguiente Modelo de Regresión:



Obsérvese que en este caso, además de tratarse de un modelo no lineal, la mayoría de los pares ordenado (x,y) , no se encuentran equidistantes a la Curva de Regresión $Y = a b^X$.

Por lo tanto, al analizar el gráfico podemos observar que los pares ordenado (x,y) , se distribuyen en forma diferente para cada uno de los valores de X que hemos seleccionado:



Obsérvese en el gráfico que todas las campanas dibujadas tienen diferentes formas, es decir: “Diferentes dispersiones”. Concluyendo que este Modelo de Regresión es Heterocedástico.

II. Determinación de la Heterocedasticidad en un Modelo de Regresión Múltiple

II.1 El Análisis Gráfico

II.1.1 El Gráfico de Residuos (Scatter Plot)

El dibujo de los valores observados (x,y) conjuntamente con la curva de regresión en un gráfico, no es suficiente para determinar si es homocedástica o heterocedástica.

Para la determinación gráfica de la Homocedasticidad (o de la Heterocedasticidad) de un modelo de regresión, se utilizan los llamados “Gráficos de Residuos” o “Diagrama de Dispersión”.

El “Gráfico de Residuos” o “Diagrama de Dispersión” (“scatter plot”, en inglés), se construye dibujando los “Residuos Estandarizados” vs. los “Valores Calculados Estandarizados”.

Residuos Estandarizados (ZRes): Valor Absoluto del cociente entre el valor de un residuo ($y - \hat{y}$) y la Desviación Estándar de la Regresión (σ)

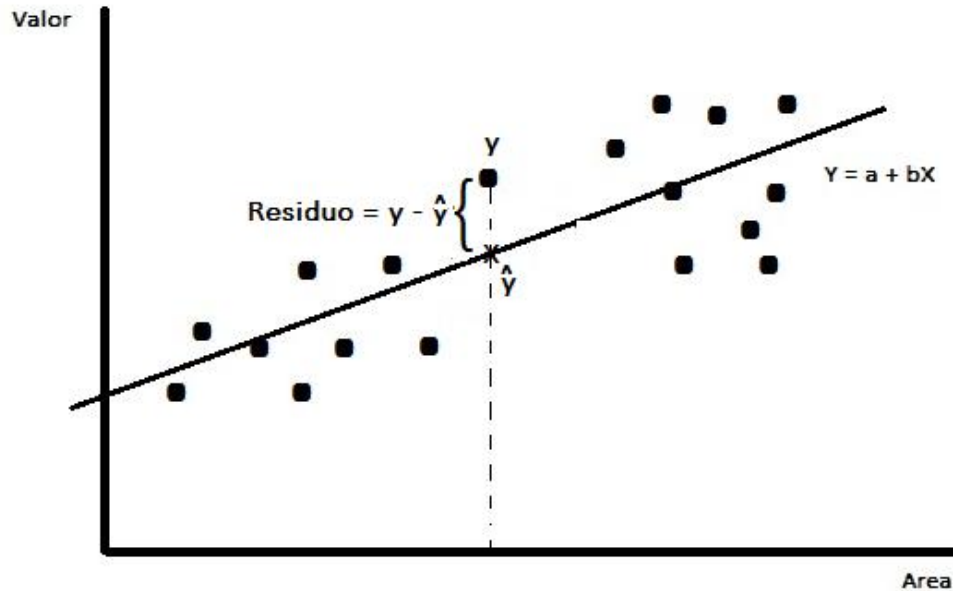
$$ZRes = \left| \frac{y - \hat{y}}{\sigma} \right|$$

Valores Calculados Estandarizados (ZPred): Cociente entre un valor calculado (“fitted value” en inglés) (\hat{y}) y la Desviación Estándar de la regresión (σ)

$$ZPred = \frac{\hat{y}}{\sigma}$$

II.1.1.1 Ejemplo demostrativo Nro. 1

Sea la siguiente representación gráfica de un modelo de regresión:



Procedimiento para construir un “Gráfico de Residuos” o “Diagrama de Dispersión”:

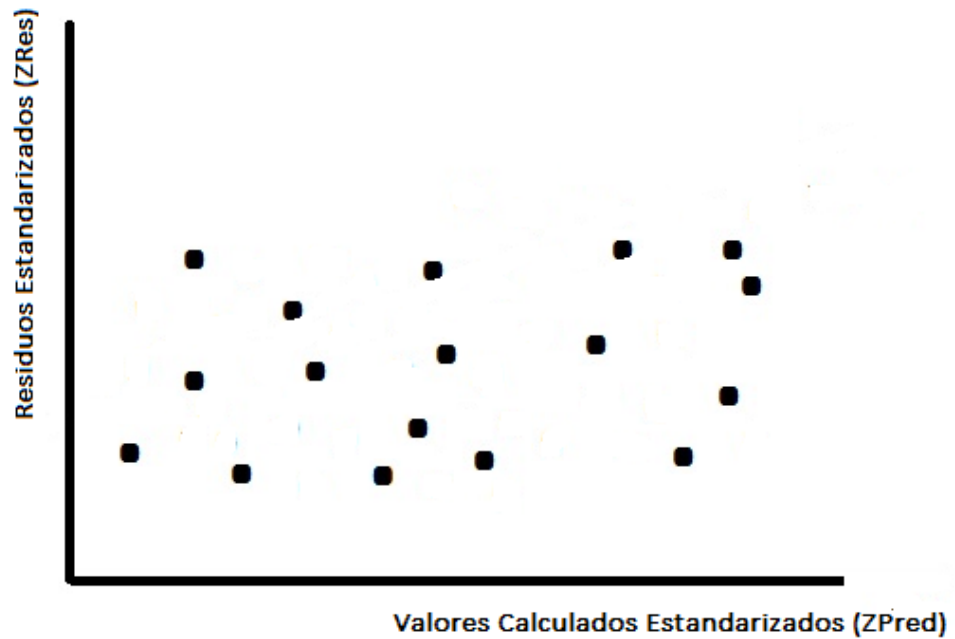
a) Cada uno de los residuos ($y - \hat{y}$) se dividirá entre la desviación estándar de la regresión (σ) y se calculará su valor absoluto (es decir se calcularán los “Residuos Estandarizados”, $ZRes$).

$$ZRes = \left| \frac{y - \hat{y}}{\sigma} \right|$$

b) Cada “Valor Calculado” (\hat{y}) se dividirá entre la desviación estándar de la regresión (σ) (es decir se predecirán los “Valores Calculados Estandarizados”, $ZPred$).

$$ZPred = \frac{\hat{y}}{\sigma}$$

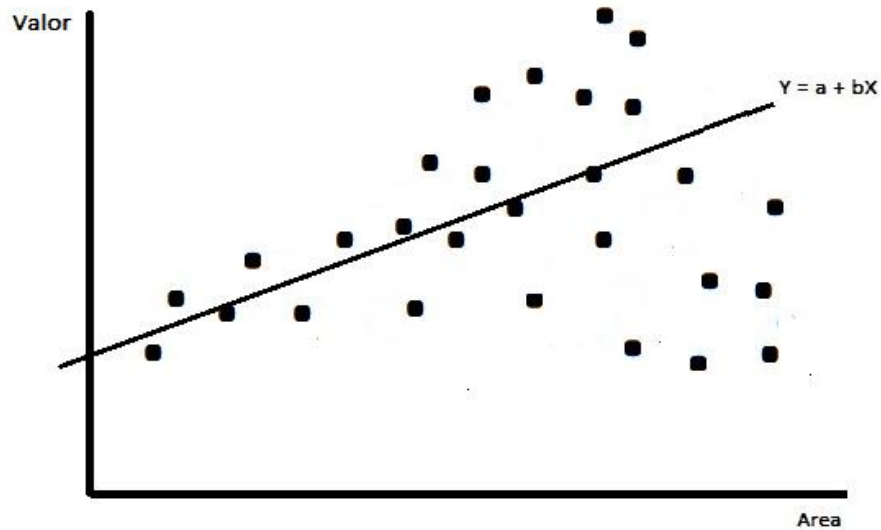
c) Se procederá a dibujar el gráfico: “Residuos Estandarizados” ($ZRes$) vs. “Valores Calculados Estandarizados” ($ZPred$).



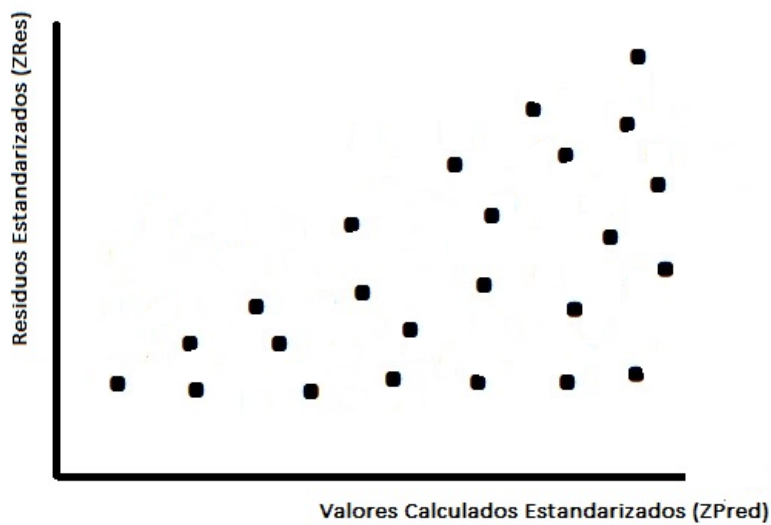
Observamos en el gráfico que la nube de puntos, no revela ningún patrón específico, por lo que se concluye que: La varianza es constante; indicando de esta manera que el modelo de regresión analizado es "Homocedástico".

II.1.1.2 Ejemplo demostrativo Nro. 2

En el siguiente modelo de regresión, al dibujar valores observados (pares ordenados x,y) y su correspondiente curva de regresión; se obtuvo la siguiente gráfica:



Al igual que en el ejemplo anterior, se procederá a dibujar el gráfico: “Residuos Estandarizados” vs. “Valores Calculados Estandarizados”:



Observamos en el gráfico que la nube de puntos, revela un patrón de variación en los residuos. En este caso a medida que el Valor Calculado (\hat{y}) aumenta, la dispersión de la serie también aumenta. Por lo que se concluye que: La varianza NO es constante; indicando de esta manera que el modelo de regresión analizado es “Heterocedástico”.

II.1.2 Homocedasticidad: La gran olvidada

Resulta ser, que no solo entre los profesionales, sino en los propios enseñantes de las Técnicas de Regresión, se ha minimizado u omitido la importancia del análisis de la Homocedasticidad de una serie de datos.

Si bien, el desarrollo de los Test o Pruebas de la Homocedasticidad de una regresión son muy recientes y la mayoría solos pudieron ser desarrollados después de la democratización de las computadoras personales y los software estadísticos.

La realidad es que en los propios libros de econometría, el tratamiento a esta importante condición, se ha reducido a una muy completa explicación teórica y raramente a sus consecuencias y significancia práctica en la regresión.

II.1.3 ¿Homocedasticidad y Heterocedasticidad en el mismo modelo de regresión?

Si bien el supuesto o condición de la Homocedasticidad especifica que:

“...Los errores o residuos $(y - \hat{y})$ de la regresión, deben distribuirse con igual varianza (σ^2) en todo el rango de valores de las variables independientes...”

En el caso de que no se cumpliera este supuesto, estaríamos en presencia de una regresión con comportamiento Heterocedástico.

Sin embargo, es posible que en un mismo modelo de regresión se presentaran supuestos de Homocedasticidad y de Heterocedasticidad a lo largo de la data analizada.

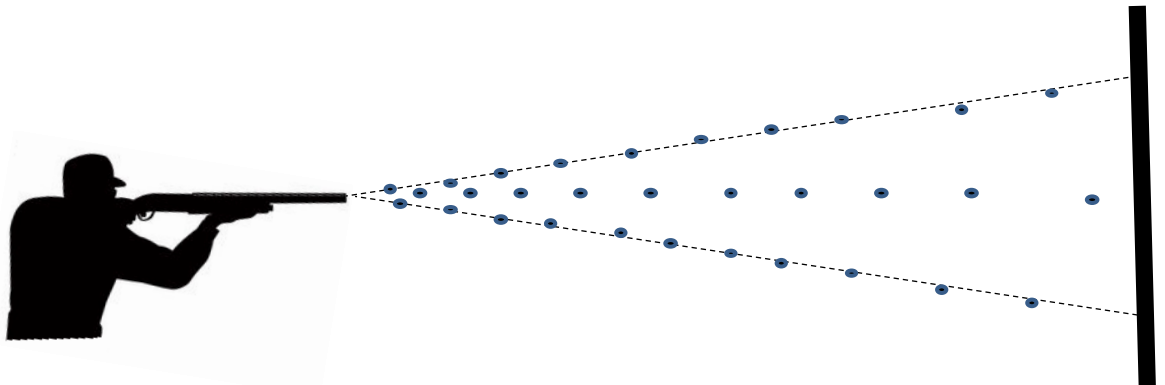
Y por cierto, es muy común la presencia de este contradictorio fenómeno en las técnicas de regresión múltiple; sobre todo cuando la data observada es heterogénea o existe presencia de valores atípicos en la muestra seleccionada.

Cuando esto ocurre, el modelo de regresión solamente existirá para el sector de la data que se comporta homocedasticamente; mientras que los coeficientes del modelo, obviamente no explicarán el fenómeno en el sector Heterocedástico.

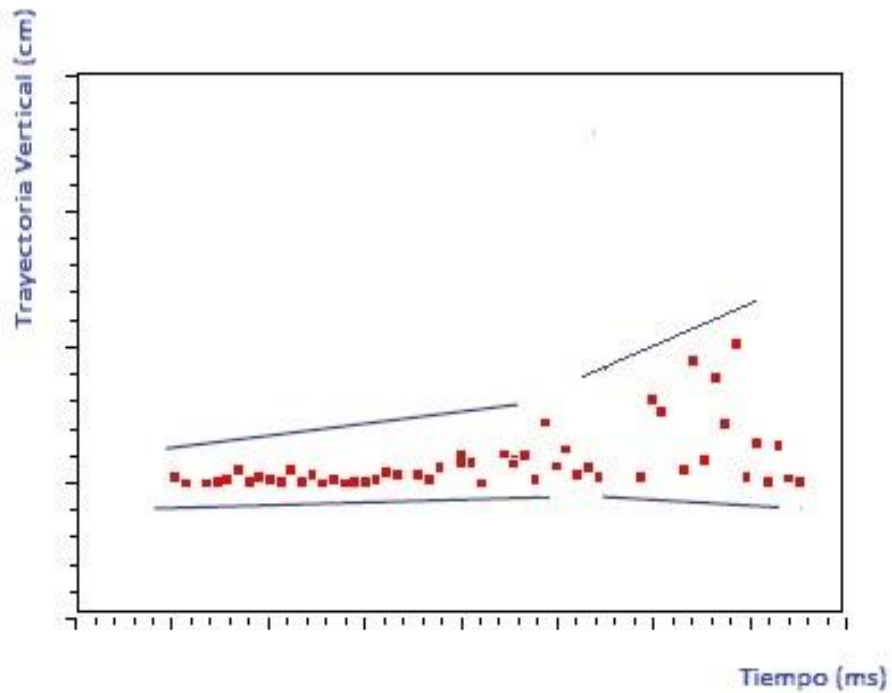
Por lo tanto, es perfectamente posible que la Heterocedasticidad pudiera presentarse solo en una parte del rango de la data, tal como se ilustra en el ejemplo siguiente.

II.1.4 El Caso del cazador

Un cazador utiliza para su deporte una escopeta diseñada para descargar varios proyectiles simultáneamente denominados municiones o perdigones:

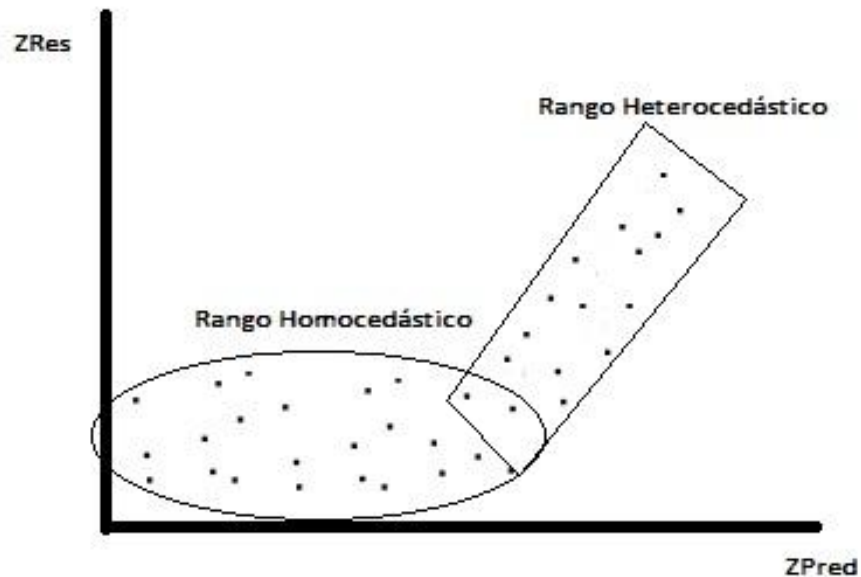


Si observamos la trayectoria de los perdigones a la salida del cañón del arma, comienzan su trayectoria muy compactos y los mismos se empiezan a dispersar a medida que avanzan hacia su objetivo. Podríamos graficarlo de la siguiente manera:



En la gráfica se ha dibujado en el eje de la abscisas la variable independiente “tiempo” (en milisegundos) y en el eje de las ordenadas la variable “trayectoria Vertical” (en centímetros). Observe que los primeros milisegundos, la munición prosigue una trayectoria compacta y los últimos la misma rápidamente se dispersa.

GRAFICO DE RESIDUOS



En el “Gráfico de Residuos” (Scatter Plot), se ha dibujado en el eje de las abscisas los Valores Calculados Estandarizados $ZPred = \frac{\hat{y}}{\sigma}$ y en el eje de las ordenadas los Residuos Estandarizados $ZRes = \left| \frac{y - \hat{y}}{\sigma} \right|$.

Obsérvese que:

a) En el llamado “Rango Homocedástico” la nube de puntos no presenta ningún patrón definido,

b) En el llamado “Rango Heterocedástico” la nube de puntos presenta un patrón definido ascendente.

c) La regresión solo será válida en el denominado “Rango Homocedástico”, ya que en el resto de la data no se cumple con la condición o supuesto de que la distribución de la “Varianza del Error” sea constante. Por lo tanto, en ese rango de la data, los coeficientes del Modelo de Regresión son poco eficientes y resultaría un pronóstico (Valor Calculado) poco adecuado.

II.2 El Análisis Numérico

II.2.1 Preliminares

Si bien todos los paquetes estadísticos especializados tales como StatGraphic, SPSS, E-Views, MiniTAB, SAS, SPlus, etc., poseen un módulo específico para el Análisis de la Homocedasticidad. En este trabajo de investigación, nos limitaremos únicamente a la utilización de la aplicación MS-Excel (que forma parte del paquete Microsoft Office), por tratarse de un software de uso generalizado, que manejan la mayoría de los Profesionales Tasadores y estudiantes.

II.2.2 La Prueba (Test) de Breusch – Pagan – Godfrey (1979)

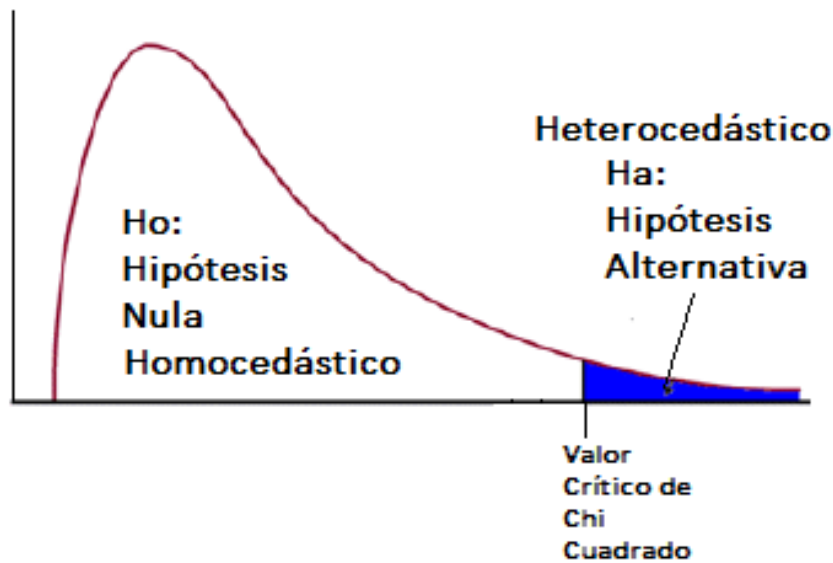
II.2.2.1 Breve planteamiento teórico

El Test de Breusch - Pagan – Godfrey, se utiliza para comprobar la Homocedasticidad en un modelo de regresión.

Esta prueba o test implica las dos siguientes hipótesis:

Ho (hipótesis nula): La regresión es Homocedástica.

Ha (hipótesis alternativa): La regresión es Heterocedástica.



II.2.2.2 Los Autores

Trevor Stanley Breusch, Ph.D. (1953 -) Profesor emérito de Econometría y Ex-Director de la Escuela de Políticas Públicas de la Universidad Nacional de Australia.

Adrian Rodney Pagan, Ph.D. (1947 -) Profesor emérito de Economía en la Escuela de Economía de la Universidad de Sidney. Ex-Director del Banco Central de Australia (Reserve Bank of Australia). Miembro de la Academia de Ciencias Sociales Australiana.

Leslie G. Godfrey, M.Sc. (1955 -) Profesor emérito de Econometría en la Escuela de Economía y Estudios Relacionados de la Universidad de York (Reino Unido).

II.2.2.3 Procedimiento Esquemático de la Prueba o Test de Breusch – Pagan – Godfrey

a) Se calcula el modelo mínimo cuadrático de regresión:

$$y = a + b X_1 + c X_2 + \dots + m X_n \text{ (u otro modelo)}$$

b) Se calcula la “Varianza Estimada” o “Varianza Muestral” de la regresión:

$$\hat{\sigma}^2 = \frac{(y - \hat{y})^2}{n}$$

c) Se declara la Variable Auxiliar:

$$\varepsilon = \frac{(y - \hat{y})^2}{\hat{\sigma}^2}$$

d) Con la Variable Auxiliar, se calcula una nueva regresión de la forma:

$$\varepsilon = a + b X_1 + c X_2 + \dots + m X_n \text{ (u otro modelo)}$$

e) Se Calcula el “Valor Crítico de Chi Cuadrado”:

$$\text{Valor Crítico de } \chi^2 = \frac{1}{2} SCR$$

$$\text{Suma del Cuadrado de la Regresión (SCR)} = \sum (\hat{y} - \bar{y})^2$$

f) Utilizando la Tabla: Distribución Chi Cuadrado, se intercepta:

EJE DE LAS X: Nivel de Significancia Preestablecido (Por lo general 0,05)

EJE DE LAS Y: Grados de Libertad (Número Total de Variables de la regresión – 1)

INTERCEPCION: Chi Cuadrado de la Regresión

g	p										
	0.001	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.999
1	10.827	5.024	3.841	2.706	1.323	0.455	0.102	0.016	0.004	0.001	0
2	13.815	7.378	5.991	4.605	2.773	1.386	0.575	0.211	0.103	0.051	0.002
3	16.266	9.348	7.815	6.251	4.108	2.366	1.213	0.584	0.352	0.216	0.024
4	18.466	11.143	9.488	7.779	5.385	3.357	1.923	1.064	0.711	0.484	0.091
5	20.515	12.832	11.07	9.236	6.626	4.351	2.675	1.61	1.145	0.831	0.21
6	22.457	14.449	12.592	10.645	7.841	5.348	3.455	2.204	1.635	1.237	0.381
7	24.321	16.013	14.067	12.017	9.037	6.346	4.255	2.833	2.167	1.69	0.599
8	26.124	17.535	15.507	13.362	10.219	7.344	5.071	3.49	2.733	2.18	0.857
9	27.877	19.023	16.919	14.684	11.389	8.343	5.899	4.168	3.325	2.7	1.152
10	29.588	20.483	18.307	15.987	12.549	9.342	6.737	4.865	3.94	3.247	1.479
11	31.264	21.92	19.675	17.275	13.701	10.341	7.584	5.578	4.575	3.816	1.834
12	32.909	23.337	21.026	18.549	14.845	11.34	8.438	6.304	5.226	4.404	2.214
13	34.527	24.736	22.362	19.812	15.984	12.34	9.299	7.041	5.892	5.009	2.617
14	36.124	26.119	23.685	21.064	17.117	13.339	10.165	7.79	6.571	5.629	3.041
15	37.698	27.488	24.996	22.307	18.245	14.339	11.037	8.547	7.261	6.262	3.483
16	39.252	28.845	26.296	23.542	19.369	15.338	11.912	9.312	7.962	6.908	3.942
17	40.791	30.191	27.587	24.769	20.489	16.338	12.792	10.085	8.672	7.564	4.416
18	42.312	31.526	28.869	25.989	21.605	17.338	13.675	10.865	9.39	8.231	4.905
19	43.819	32.852	30.144	27.204	22.718	18.338	14.562	11.651	10.117	8.907	5.407
20	45.314	34.17	31.41	28.412	23.828	19.337	15.452	12.443	10.851	9.591	5.921

Tabla de la Distribución Chi Cuadrado χ^2 (fuente: www.bioestadistica.uma.es)

g) Prueba de la Hipótesis Nula:

Si Valor Crítico de $\chi^2 < \chi^2 \rightarrow$ Se cumple la Hipótesis Nula \rightarrow Homocedástica

Si Valor Crítico de $\chi^2 > \chi^2 \rightarrow$ NO se cumple la Hipótesis Nula \rightarrow Heterocedástica

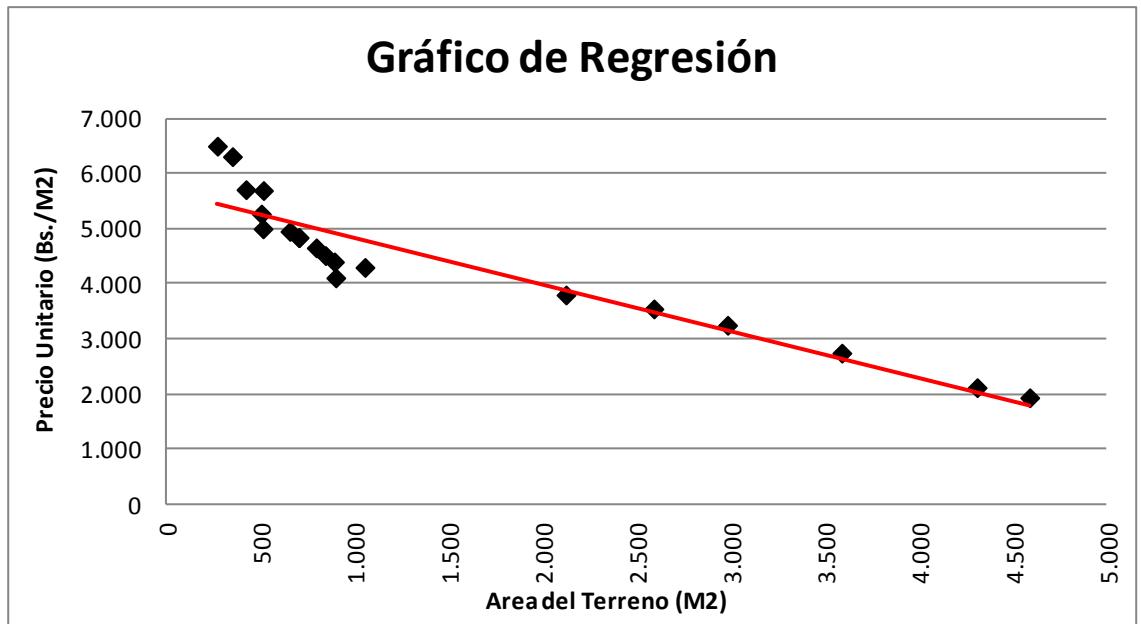
III Ejemplo Demostrativo

Sea la siguiente data correspondiente a comparables de terrenos sin construcción, fuera de la poligonal urbana del área metropolitana de la ciudad de Valencia (Venezuela).

Datos Comparables:

		PRECIO
	AREA	UNITARIO
IDENTIF.	(M2)	(Bs/M2)
1	269,00	6.500
2	349,00	6.310
3	421,00	5.711
4	514,00	5.695
5	503,00	5.269
6	512,00	5.000
7	655,00	4.953
8	704,00	4.842
9	702,00	4.840
10	795,00	4.655
11	844,00	4.515
12	892,00	4.400
13	1.054,00	4.300
14	899,00	4.110
15	2.124,00	3.800
16	2.592,00	3.547
17	2.984,00	3.250
18	3.592,00	2.745
19	4.312,00	2.118
20	4.592,00	1.935

1) Al dibujar los datos y la curva (recta) de regresión en el gráfico, se obtiene:



2) El Modelo de Regresión Lineal es el siguiente:

$$Y = 5.658,4308 - 0,8418 X$$

3) La Salida de la Subrutina de Regresión Múltiple con el software MS-Excel es la siguiente:

a:	5.658,4308	Coficiente del término independiente
b:	-0,8418	Coficiente que acompaña a X (Area)
σ:	475,22	Error estándar de la correlación (Sigma)
R²:	0,8638	Coficiente de Determinación
F:	114,2041	Estadístico F
SCR:	25.791.672	Suma del cuadrado de la regresión
SCE:	4.065.090	Suma del cuadrado del error

4) Se procede al Cálculo de los Residuo, Residuos Estandarizados y Valores Calculados Estandarizados

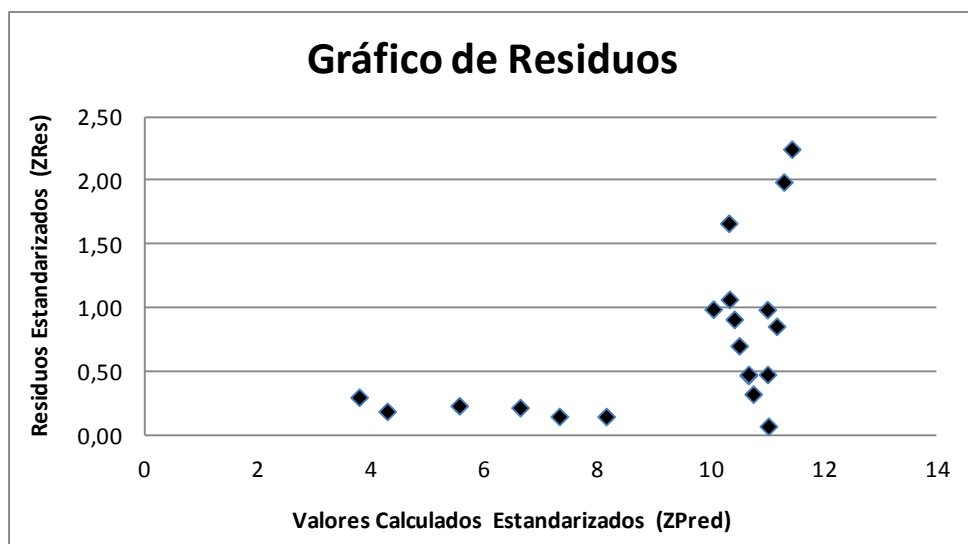
$$Residuos = y - \hat{y}$$

$$Residuos Estandarizados (ZRes) = \left| \frac{(y - \hat{y})}{\sigma} \right|$$

$$Valores Calculados Estandarizados (ZPred) = \frac{\hat{y}}{\sigma}$$

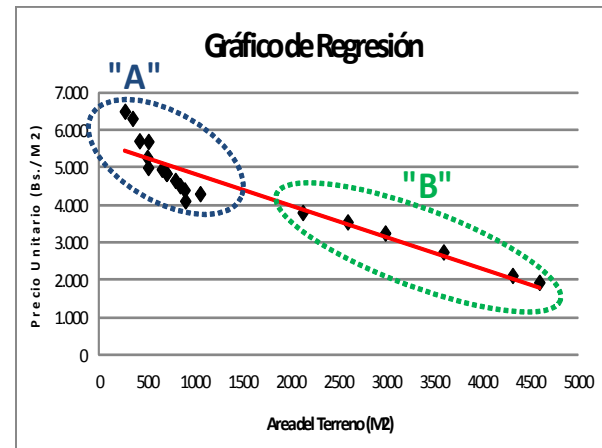
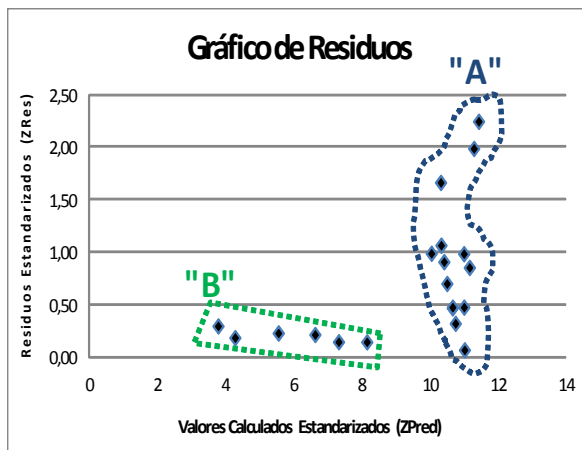
IDENT.	VALORES OBSERV.	VALORES CALCUL.	RESIDUOS	ZPred	ZRes
1	6.500,00	5.431,97	1.068,03	11,43	2,2474
2	6.310,00	5.364,63	945,37	11,29	1,9893
3	5.711,00	5.304,01	406,99	11,16	0,8564
4	5.695,00	5.225,72	469,28	11,00	0,9875
5	5.269,00	5.234,98	34,02	11,02	0,0716
6	5.000,00	5.227,41	-227,41	11,00	0,4785
7	4.953,00	5.107,02	-154,02	10,75	0,3241
8	4.842,00	5.065,77	-223,77	10,66	0,4709
9	4.840,00	5.067,46	-227,46	10,66	0,4786
10	4.655,00	4.989,16	-334,16	10,50	0,7032
11	4.515,00	4.947,91	-432,91	10,41	0,9110
12	4.400,00	4.907,51	-507,51	10,33	1,0679
13	4.300,00	4.771,13	-471,13	10,04	0,9914
14	4.110,00	4.901,61	-791,61	10,31	1,6658
15	3.800,00	3.870,35	-70,35	8,14	0,1480
16	3.547,00	3.476,37	70,63	7,32	0,1486
17	3.250,00	3.146,37	103,63	6,62	0,2181
18	2.745,00	2.634,53	110,47	5,54	0,2325
19	2.118,00	2.028,40	89,60	4,27	0,1885
20	1.935,00	1.792,68	142,32	3,77	0,2995

5) Se dibuja el gráfico Valores Observados Estandarizados (ZPred) vs. Residuos Estandarizados (ZRes):



Obsérvese en el gráfico siguiente, que la "nube de puntos A", no presenta ningún patrón; mientras que la "nube de puntos B", presenta un patrón aproximadamente rectilíneo descendente; indicando que la regresión

muy probablemente sea "Homocedástica" en los datos que conforman la "nube de puntos "A", pero que sea "Heterocedástica en los datos que conforman la "nube de puntos B". Sin embargo, no se puede tener la certeza de esta afirmación, hasta que no se pruebe analíticamente esta condición de "Homocedasticidad".



6) Cálculo de la "Varianza Estimada" de la Regresión y declaración de la Variable Auxiliar

$$\hat{\sigma}^2 = \frac{(y - \hat{y})^2}{n} \quad (\text{Varianza Estimada})$$

$$\varepsilon = \frac{(y - \hat{y})^2}{\hat{\sigma}^2} \quad (\text{Variable Auxiliar})$$

REFER.	RESIDUOS	RESID. AL CUADRADO	VARIABLE AUXILIAR
1	1.068,03	1.140.678,03	5,6121
2	945,37	893.729,82	4,3971
3	406,99	165.637,31	0,8149
4	469,28	220.221,06	1,0835
5	34,02	1.157,15	0,0057
6	-227,41	51.713,73	0,2544
7	-154,02	23.723,02	0,1167
8	-223,77	50.074,09	0,2464
9	-227,46	51.736,27	0,2545
10	-334,16	111.665,95	0,5494
11	-432,91	187.414,69	0,9221
12	-507,51	257.561,99	1,2672
13	-471,13	221.960,53	1,0920
14	-791,61	626.650,74	3,0831
15	-70,35	4.949,60	0,0244
16	70,63	4.988,57	0,0245
17	103,63	10.739,75	0,0528
18	110,47	12.204,54	0,0600
19	89,60	8.028,54	0,0395
20	142,32	20.254,57	0,0997
Sumatoria		4.065.089,94	
Varianza Estimada		203.254,50	

7) Se recalcula la regresión. Sustituyendo a la Variable Dependiente “Y” por la Variable Auxiliar:

REFER.	AREA (M2)	VARIABLE AUXILIAR
1	269,00	5,6121
2	349,00	4,3971
3	421,00	0,8149
4	514,00	1,0835
5	503,00	0,0057
6	512,00	0,2544
7	655,00	0,1167
8	704,00	0,2464
9	702,00	0,2545
10	795,00	0,5494
11	844,00	0,9221
12	892,00	1,2672
13	1054,00	1,0920
14	899,00	3,0831
15	2124,00	0,0244
16	2592,00	0,0245
17	2984,00	0,0528
18	3592,00	0,0600
19	4312,00	0,0395
20	4592,00	0,0997

El Nuevo Modelo de Regresión de la forma $\varepsilon = a + b X$ es el siguiente:

$$\varepsilon = 1,70140 - 0,00048 X$$

La Salida de la Subrutina de la Nueva Regresión Múltiple con el software MS-Excel es la siguiente:

a:	1,70140	Coficiente del término independiente
b:	-0,00048	Coficiente que acompaña a X (Area)
σ:	1,4532	Error estándar de la correlación (Sigma)
R²:	0,1799	Coficiente de Determinación
F:	3,9480	Estadístico F
SCR:	8,3369	Suma del cuadrado de la regresión
SCE:	38,0096	Suma del cuadrado del error

8) Se calcula el Valor Crítico del Chi Cuadrado de la Regresión, que vendrá dado por la siguiente expresión;

$$\text{Valor Crítico de } \chi^2 = \frac{1}{2} SCR$$

$$\text{Suma del Cuadrado de la Regresión (SCR)} = \sum (\hat{y} - \bar{y})^2$$

Valor Crítico de Chi Cuadrado = 4,1684

9) Se calcula el número total de Variables - 1 (Grados de Libertad)

Número de Variables de Regresión (k) = 2

Grados de Libertad = k - 1 = 1

10) Se declara el "Nivel de Significancia" o sea el margen que estamos dispuestos aceptar como error

NOTA: Por lo general, los paquetes estadísticos por defecto asumen un nivel de significancia de 0,05 (95% de confianza), precisión que es más que suficiente para nuestros cálculos.

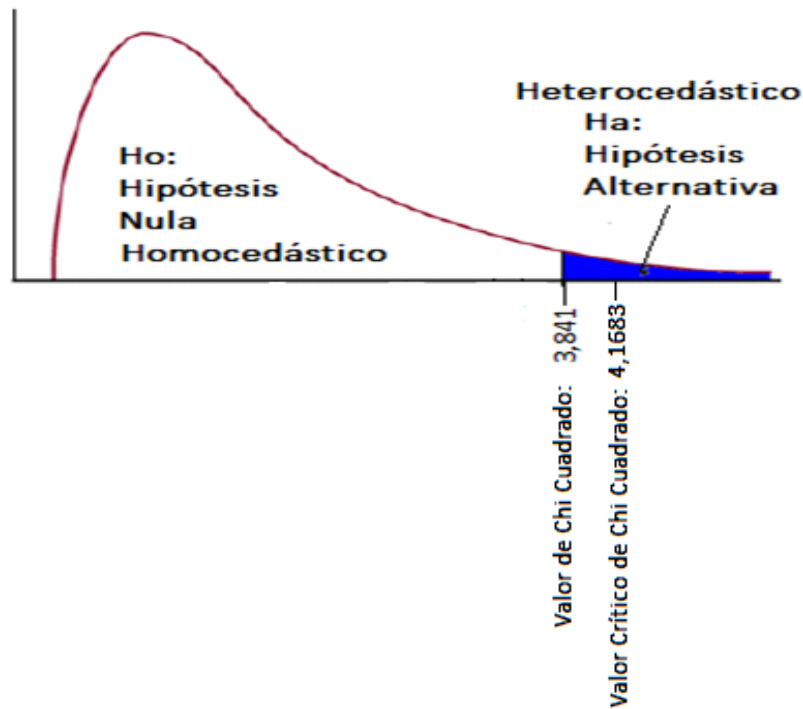
11) Se determina en la tabla siguiente, el valor de Chi Cuadrado, en función de los Grados de Libertad y el Nivel de Significancia:

Tabla de la Distribución Chi Cuadrado χ^2 (fuente: www.bioestadistica.uma.es)

g	SIGNIFICANCIA										
	0.001	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.999
1	10.827	5.024	3.841	2.706	1.323	0.455	0.102	0.016	0.004	0.001	0
2	13.815	7.378	5.991	4.605	2.773	1.386	0.575	0.211	0.103	0.051	0.002
3	16.266	9.348	7.815	6.251	4.108	2.366	1.213	0.584	0.352	0.216	0.024
4	18.466	11.143	9.488	7.779	5.385	3.357	1.923	1.064	0.711	0.484	0.091
5	20.515	12.832	11.07	9.236	6.626	4.351	2.675	1.61	1.145	0.831	0.21
6	22.457	14.449	12.592	10.645	7.841	5.348	3.455	2.204	1.635	1.237	0.381
7	24.321	16.013	14.067	12.017	9.037	6.346	4.255	2.833	2.167	1.69	0.599
8	26.124	17.535	15.507	13.362	10.219	7.344	5.071	3.49	2.733	2.18	0.857
9	27.877	19.023	16.919	14.684	11.389	8.343	5.899	4.168	3.325	2.7	1.152
10	29.588	20.483	18.307	15.987	12.549	9.342	6.737	4.865	3.94	3.247	1.479
11	31.264	21.92	19.675	17.275	13.701	10.341	7.584	5.578	4.575	3.816	1.834
12	32.909	23.337	21.026	18.549	14.845	11.34	8.438	6.304	5.226	4.404	2.214
13	34.527	24.736	22.362	19.812	15.984	12.34	9.299	7.041	5.892	5.009	2.617
14	36.124	26.119	23.685	21.064	17.117	13.339	10.165	7.79	6.571	5.629	3.041
15	37.698	27.488	24.996	22.307	18.245	14.339	11.037	8.547	7.261	6.262	3.483
16	39.252	28.845	26.296	23.542	19.369	15.338	11.912	9.312	7.962	6.908	3.942
17	40.791	30.191	27.587	24.769	20.489	16.338	12.792	10.085	8.672	7.564	4.416

12) Se Verifica o no la Hipótesis Nula en la Prueba (Test) de Breusch – Pagan – Godfrey, en función a los valores de Chi Cuadrado y el valor crítico de Chi Cuadrado:

VALOR DE CHI CUADRADO =	3,8410
VALOR CRITICO DE CHI CUADRADO=	4,1684
EL VALOR CRITICO DE CHI CUADRADO ES MAYOR QUE CHI CUADRADO	



13) Conclusiones de este Ejemplo Demostrativo:

A) SE VERIFICO ANALITICAMENTE QUE LA REGRESION ES HETEROCEDASTICA, TAL COMO SE HABIA PREDICHO EN EL ANALISIS GRAFICO.

B) SI BIEN PARECIERA EN LA REPRESENTACION GRAFICA DE LA REGRESIÓN QUE EL MODELO DEDUCIDO EXPLICARIA LA EL FENOMENO ESTUDIADO EN LA TOTALIDAD DE LA DATA ESTUDIADA. ESTO NO ES ASI.

C) EN EL GRAFICO DE RESIDUOS (SCATTER PLOT) SE DETECTO QUE UN SECTOR DE LA DATA SE COMPORTA EN FORMA

HOMOCEASTICA, MIENTRAS EL OTRO SECTOR SE COMPORTA DEFINITIVAMENTE EN FORMA HETEROCEASTICA.

D) EL MODELO DE REGRESION SOLO EXISTE PARA EL SECTOR DE LA DATA QUE SE COMPORTA HOMOCEASTICAME (NUBE DE PUNTOS "A").

E) POR ENDE, EN EL SECTOR HETEROCEASTICO DE LA DATA (NUBE DE PUNTOS "B"), LOS COEFICIENTES DE REGRESION DEL MODELO OBVIAMENTE NO EXPLICAN EL FENOMENO.

IV Conclusiones y Recomendaciones

IV.1 Condiciones o Supuestos Mínimos que debe cumplir un modelo de regresión

El éxito en el ajuste de un modelo de regresión, la validez de los valores predichos y las conclusiones que se puedan concluir, depende del cumplimiento de los siguientes supuestos o condiciones:

a) El Coeficiente de Determinación (R^2) debe ser alto (por lo menos un 70 – 75%).

b) Se debe verificar el rechazo de la Hipótesis Nula (H_0) de la Prueba o Test de Fisher ($F > F_0$).

c) Las variables independientes no deben ser proporcionales entre sí (No debe presentarse Multicolinealidad entre las Variables Independientes).

d) Los errores o residuos ($y - \hat{y}$) de la regresión, debe distribuirse en forma Normal.

e) Los errores o residuos ($y - \hat{y}$) de la regresión, deben distribuirse con igual varianza (σ^2) en todo el rango de valores de las variables independientes (Comportamiento Homocedástico).

NOTA: También existen otras condiciones o supuestos que debe cumplir una regresión; pero si los Cinco (5) supuestos precitados se cumplen, es muy probable que se cumplan los demás.

IV.2 La Validez de la Regresión frente a la Heterocedasticidad

a) Tal como se demostró en el texto de este Trabajo Técnico, una regresión puede presentar un sector de la data donde presenta un “comportamiento Homocedástico” y otro sector presentar un “comportamiento Heterocedástico”.

b) En este caso (que es muy común) la regresión se considera heterocedástica, ya que no se verifica el Test de Breusch – Pagan – Godfrey:

El Valor Crítico de $\chi^2 > \chi^2$ y por lo tanto NO se cumple la Hipótesis Nula

c) Eliminar la data que presenta Heterocedasticidad y trabajar únicamente con los datos del sector Homocedástico de la regresión no es una opción, ya que esta porción de datos a descartar; la mayoría de las veces resulta en una parte importante de la única información disponible para estudiar un fenómeno.

d) La Gráfica de Residuos (Scatter Plot), permite determinar:

d-1) El “Rango Homocedástico”: Nube de puntos que no presenta ningún patrón definido.

d-2) El “Rango Heterocedástico”: Nube de puntos que presenta un patrón definido.

e) La regresión solo será válida en el denominado “Rango Homocedástico” de la Gráfica de Residuos

f) Fuera ese rango (zona heterocedástica), los coeficientes del Modelo de Regresión son poco eficientes y resultarían en un pronóstico poco adecuado.

BIBLIOGRAFÍA:

Arnold, S.F. (1981). "The Theory of Linear Models and Multivariate Observations". Wiley

Breusch, T. S. y A. R. Pagan., (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation". Revista: Econométrica: Vol. 47. Acceso: <http://onlinelibrary.wiley.com/journal/>

Carmona, F. (2003). "Modelo Lineales". Universitat de Barcelona. Acceso: <http://www.ub.edu/stat/docencia/Diplomatura/ModelsLineals/regre.pdf>

Box, G.E.P, Hunter, W. and J.S. Hunter, J.S. (1988). Estadística para Investigadores. Reverté

Doran, H.E. (1989). "Applied Regression Analysis in Econometrics". Marcel Dekker

Glesjer, H., (1969). «A New Test for Heteroscedasticity», Revista: Journal of the American Statistical Association: vol. 64. Acceso: <http://onlinelibrary.wiley.com/journal/>

Godfrey, L. (1978). "Testing For Multiplicative Heteroscedasticity". Revista: Journal of Econometrics, vol. 8. Acceso: <http://onlinelibrary.wiley.com/journal/>

Johnston, J. (1989). "Métodos de Econometría". Vincens-Vives

Murillo, F. y González B. (2000). "Manual de Econometría". Universidad de las Palmas. Gran Canaria. Acceso: <https://www2.ulpgc.es/>

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach" Revista Econometría. Vol. 75. Acceso: <http://onlinelibrary.wiley.com/journal/>

Peña, D. (1993). "Estadística: Modelos y Métodos". Alianza

Piol, R. (2011). "Estadística Aplicada a la Valuación Inmobiliaria. Parte II. Análisis de Variables Múltiples". Sociedad de Ingeniería de Tasación de Venezuela. 1999.". Acceso: <http://www.rpiol.com/>

Piol, R. (2014). "Análisis Estadísticos de Series Pequeñas y Muy Pequeñas aplicado a la Valuación Inmobiliaria". Sociedad de Ingeniería de Tasación de Venezuela. Acceso: <http://www.rpiol.com/>

Quesada, M. (2011). "Análisis de Series. Modelos Heterocedásticos". Universidad de Granada. Trabajo Especial de Maestría. Acceso: <http://masteres.ugr.es/>

Ramírez, D.C. (S/F). "Heterocedasticidad". Universidad de Los Andes. Acceso: <http://webdelprofesor.ula.ve/economia/dramirez/>

Tsay, R.S. (1987): "Conditional Heteroskedastic Time Series Models". Revista: Journal of American Statistical Association, Vol. 34. Acceso: <http://onlinelibrary.wiley.com/journal/>

White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity". Revista: Econométrica, Vol. 48. Acceso: <http://onlinelibrary.wiley.com/journal/>