

2 VIOLACIONES A LOS SUPUESTOS DEL MODELO LINEAL GENERAL

2.1 Multicolinealidad.

Definición y efectos.

Una de las hipótesis del modelo lineal múltiple establece que la matriz X de orden $n \times k$ es de rango K , ello implica que no existe dependencia lineal entre las variables explicativas. En caso de darse esa dependencia, la matriz $X'X$ no se podría invertir y como consecuencia no sería posible la estimación del vector de coeficientes de regresión. Cuando una o más variables explicativas se expresan como combinación lineal de las restantes se le denomina multicolinealidad extrema o perfecta y su detección es inmediata. La multicolinealidad extrema generalmente no se presenta en la práctica, pero si es muy probable que entre las variables explicativas se presente un alto grado de dependencia lineal, que sin embargo no impide obtener soluciones para los estimadores. En resumen, el problema de la multicolinealidad se refiere a la existencia de relaciones aproximadamente lineales entre las variables explicativas del modelo y de sus efectos.

$$\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_k X_k = 0 \quad \text{Multicolinealidad perfecta}$$

$$\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_k X_k \cong 0 \quad \text{Multicolinealidad aproximada}$$

La multicolinealidad disminuye la precisión de los estimadores y por tanto se experimentará un notable aumento en las varianzas de los estimadores, aunque conservan su propiedad de insesgamiento.

Se complica la posibilidad de aislar adecuadamente la influencia relativa de las diferentes variables explicativas sobre la variable dependiente y se corre el riesgo de eliminar variables del modelo debido a que sus elevados errores estándares pueden inducir a no rechazar la hipótesis de que ciertos coeficientes sean nulos.

Las estimaciones de los coeficientes se tornan muy sensibles a pequeños cambios en los datos muestrales. Los coeficientes experimentan grandes cambios con sólo cambiar o agregar una observación.

Para ilustrar mejor el problema considérese en forma simplificada un modelo de dos variables explicativas expresadas en forma de desviaciones respecto de su media y que por lo tanto el modelo no incluye el término independiente

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \quad i = 1, 2, \dots, n \quad y_i = Y_i - \bar{Y} \quad x_i = X_i - \bar{X}$$

La matriz X adopta la siguiente forma:

$$\begin{bmatrix} x_{21} & x_{31} \\ x_{22} & x_{32} \\ \dots & \dots \\ x_{2n} & x_{3n} \end{bmatrix}$$

y por lo tanto la matriz $X'X$ se expresa:

$$\begin{bmatrix} \sum_{i=1}^n x_{2i}^2 & \sum_{i=1}^n x_{2i} x_{3i} \\ \sum_{i=1}^n x_{2i} x_{3i} & \sum_{i=1}^n x_{3i}^2 \end{bmatrix}$$

La matriz inversa $(X'X)^{-1}$ será entonces de la forma

$$(X'X)^{-1} = \frac{1}{\sum_{i=1}^n x_{3i}^2 \sum_{i=1}^n x_{2i}^2 - (\sum_{i=1}^n x_{2i} x_{3i})^2} \begin{bmatrix} \sum_{i=1}^n x_{3i}^2 & -\sum_{i=1}^n x_{2i} x_{3i} \\ -\sum_{i=1}^n x_{2i} x_{3i} & \sum_{i=1}^n x_{2i}^2 \end{bmatrix}$$

La estimación de la varianza de $\hat{\beta}_2$ se localiza en la celda superior izquierda de la matriz, dividida entre el determinante de la matriz y multiplicada por la varianza del error:

$$\hat{V}(\hat{\beta}_2) = \frac{\sigma^2 \sum_{i=1}^n x_{3i}^2}{\sum_{i=1}^n x_{3i}^2 \sum_{i=1}^n x_{2i}^2 - \left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}$$

Se divide numerador y denominador entre la suma de los cuadrados de x_3

$$\hat{V}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 - \frac{\left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}{\sum_{i=1}^n x_{3i}^2}}$$

A continuación se factoriza en el denominador entre la suma de cuadrados de x_2

$$\hat{V}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 \left(1 - \frac{\left(\sum_{i=1}^n x_{2i} x_{3i} \right)^2}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2} \right)}$$

y entonces la varianza del estimador del coeficiente de x_2 se puede expresar alternativamente en función del cuadrado de la correlación entre x_2 y x_3 :

$$\hat{V}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 (1 - r_{23}^2)} = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2} \frac{1}{(1 - r_{23}^2)}$$

Es obvio que si r_{23} es igual a cero la varianza de $\hat{\beta}_2$ es mínima, pero si r_{23} es diferente de cero, se incrementará la varianza del estimador. En la siguiente tabla se puede apreciar el efecto de diferentes valores de la correlación entre x_2 y x_3 sobre la varianza de los estimadores de los coeficientes de x_2 o x_3 . Una correlación ligeramente mayor a 0.7 duplicará la varianza del estimador y una correlación igual a 0.90 la multiplicará por algo más que 5 veces su varianza.

r	$1-r^2$	$1/(1-r^2)$
0.10	0.990	1.01
0.15	0.978	1.02
0.20	0.960	1.04
0.25	0.938	1.07
0.30	0.910	1.10
0.35	0.878	1.14
0.40	0.840	1.19
0.45	0.798	1.25
0.50	0.750	1.33
0.55	0.698	1.43
0.60	0.640	1.56
0.65	0.578	1.73
0.70	0.510	1.96
0.75	0.438	2.29
0.80	0.360	2.78
0.85	0.278	3.60
0.90	0.190	5.26
0.95	0.098	10.26

2.2 Formas de detectar la Multicolinealidad.

a) Coeficientes de correlación simples.

Existen diversos indicios y estadísticas que pueden sugerir la presencia de la multicolinealidad, el primero es que en la matriz de correlaciones simples entre variables explicativas se observen valores muy altos en valor absoluto. En modelos con dos variables explicativas es un valioso auxiliar, aunque esta condición pudiera no estar presente y sin embargo tenerse un problema de multicolinealidad en modelos con múltiples variables explicativas.

b) Coeficientes de correlación parciales.

Como se sabe el coeficiente de correlación parcial mide el grado de relación lineal entre las variables involucradas en presencia de valores constantes de las restantes variables. La ventaja de los coeficientes de correlación parcial sobre los simples es precisamente la capacidad de aislamiento de la influencia lineal. Entonces coeficientes de correlación parciales elevados entre las variables explicativas se pueden considerar como indicios de multicolinealidad.

c) R^2 elevada y coeficientes no significativos.

Otra forma que no requiere mucho esfuerzo para contar con evidencias iniciales de multicolinealidad es a través de la presencia de un coeficiente de determinación elevado para el modelo ajustado y simultáneamente varios coeficientes con valores de r tan bajos que no dan valores significativos e inducen a no rechazar las hipótesis de nulidad de los coeficientes respectivos.

d) Índice de Tolerancia (IT) y Factor de Incremento de la Varianza (FIV).

En un modelo con más variables, en lugar del cuadrado de la correlación simple entre las dos variables explicativas se calcula R_i^2 , el coeficiente de determinación de la regresión de la variable explicativa de interés, de la que se sospecha sea combinación lineal aproximada, sobre las restantes variables explicativas. Tanto la diferencia $1-R_i^2$ (Índice de Tolerancia) como el recíproco de la diferencia, este último también llamado Factor de Incremento de la Varianza (**FIV**) se suelen adoptar como criterios diagnósticos de la presencia de multicolinealidad. A medida que el **FIV** sea mayor que 1, se podrá evaluar el efecto de la multicolinealidad presente sobre las varianzas del modelo.

$$IT_i = 1 - R_i^2$$

$$FIV_i = \frac{1}{1 - R_i^2}$$

e) Número de Condición (NC).

El número de condición, desarrollado por Rachudel (1971) y Belsley (1980) es igual a la raíz cuadrada de la razón entre la raíz característica más grande λ_{\max} y la raíz característica más pequeña λ_{\min} de la matriz $X'X$ cuyas columnas han sido previamente normalizadas para evitar efectos de las unidades de medición sobre los valores característicos. Es recomendable analizar el **NC** correspondiente a todas las variables explicativas.

Donde λ es el autovalor (eigenvalor)

$$NC = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

Se considera que existe cierta multicolinealidad a partir de 20 y muy severa si el número de condición es superior a 30.

2.3 Medidas Remediales para la Multicolinealidad.

Existen diversas formas de enfrentar la multicolinealidad sin embargo hay que valorar sus ventajas y limitaciones en cada caso.

a) Incremento de Muestra.

La forma más simple es incrementar el tamaño de muestra, aunque en términos prácticos puede resultar imposible agregar más observaciones al modelo. Ello si se tiene al alcance tendrá como efecto la reducción de varianzas, pero no la eliminación de la multicolinealidad.

b) Eliminar variables

Aquellas variables que resultan en una casi combinación lineal de las restantes no están aportando información al modelo pues se da redundancia informativa. Puede ser factible su eliminación, sin embargo esta drástica medida puede tener como desventaja en incurrir en errores de especificación del modelo.

c) Transformación de variables

Transformar variables, por ejemplo, a través de potencias, logaritmos puede suprimir parcial o totalmente la influencia lineal.

d) Crear nuevas variables.

Adoptar variables como diferencia entre períodos de las originales, razones entre diferentes variables o construir combinaciones lineales de las originales mediante componentes principales dará lugar, en este último caso, a nuevas variables no correlacionadas. La desventaja con las transformaciones en general sería la capacidad interpretativa que se daría a las nuevas variables y coeficientes.

e) Regresión Ridge

Otra solución más laboriosa es la de utilizar un estimador "Ridge", (traducido literalmente como espinazo o cresta), para reducir la varianza, esta alternativa propuesta por Hoerl y Kennard en 1970. Consiste en agregar el producto kD donde $k > 0$ es un escalar y D es una matriz que contiene los elementos de la diagonal de la matriz $X'X$

$$\hat{\beta}_R = (X'X + kD)^{-1} X'Y$$

El estimador "Ridge" es sesgado, pero con error cuadrático medio menor. La constante K se selecciona entre valores pequeños pues el sesgo aumenta conforme K aumenta.

$$\begin{aligned}\hat{\beta}_R &= (X'X + kD)^{-1} X'(X\hat{\beta}_R + u) \\ &= (X'X + kD)^{-1} X'X\hat{\beta}_R + (X'X + kD)^{-1} u \\ E(\hat{\beta}_R) &= (X'X + kD)^{-1} X'X\beta_R\end{aligned}$$

2.4 Heteroscedasticidad

Definición y efectos.

El modelo lineal múltiple supone que las perturbaciones u_i tienen la misma varianza σ^2 . Ello las identifica como homoscedásticas, cuando las varianzas de las perturbaciones no son constante, sino que cambia conforme cambian los valores de las variables explicativas, se presenta la heteroscedasticidad. Implica que no se cumpla la siguiente propiedad

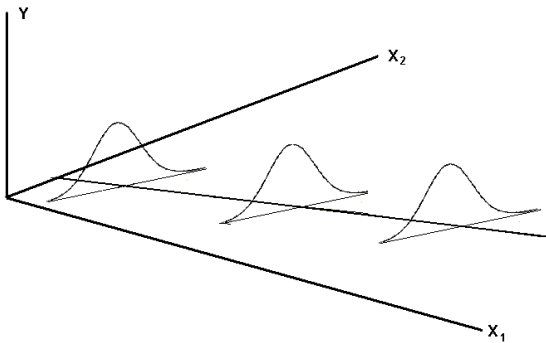
$$E(uu') = \sigma^2 I_n$$

y que se tenga una matriz de la forma:

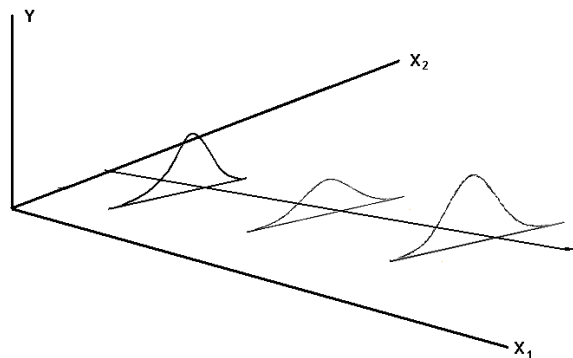
$$E(uu') = \sigma^2 \Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_i^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}$$

En los siguientes diagramas se comparan las condiciones de homoscedasticidad y heteroscedasticidad para un modelo con 2 variables explicativas.

Homoscedasticidad

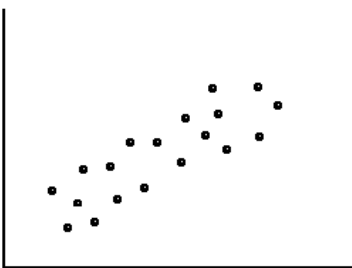


Heteroscedasticidad

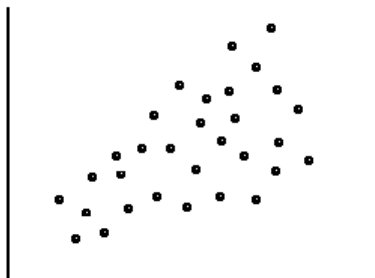


Considerando los diagramas de dispersión para un modelo con una variable explicativa, se pueden observar los siguientes patrones de heteroscedasticidad.

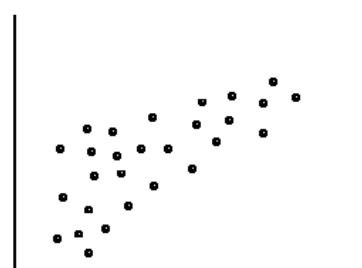
Homoscedasticidad



Heteroscedasticidad



Heteroscedasticidad



El origen de la heteroscedasticidad puede ser diverso, frecuentemente se presenta en modelos con datos transversales, por ejemplo en un modelo que involucre datos sobre consumo en función del ingreso para un grupo de familias. Los hábitos de consumo pueden presentar mayor dispersión en las familias de mayores ingresos. En otras ocasiones en modelos longitudinales, las variables eventualmente pueden experimentar cambios de comportamiento a través del tiempo. También es factible que existan errores de especificación. El efecto inmediato de la heteroscedasticidad en el modelo es que las observaciones con mayor varianza tienen pesos relativos mayores en las estimaciones efectuadas por mínimos cuadrados y se vuelven ineficientes. Los estimadores de mínimos cuadrados de los coeficientes ante la presencia de heteroscedasticidad permanecen insesgados y consistentes pues en la obtención de su esperanza no interviene la varianza del error, pero no son eficientes y por tanto dejan de ser MELI.

$$\begin{aligned} E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'u) \\ &= E(\beta) + (X'X)^{-1}X'E(u) \\ &= E(\beta) \end{aligned}$$

Por otra parte, las varianzas y covarianzas son estimadas en forma sesgada, pues σ^2 no es constante y la siguiente fórmula deja de ser cierta.

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

La matriz de varianzas y covarianzas adopta la forma:

$$V(\hat{\beta}) = ((X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1}) = \sigma^2((X'X)^{-1}X'\Omega X(X'X)^{-1})$$

2.5 Formas de detectar la Heteroscedasticidad.

Existen diversos procedimientos para probar la hipótesis de homoscedasticidad. Algunas ayudan a la identificación de la forma funcional que define el comportamiento de la varianzas y otras se limitan a rechazar o no rechazar la hipótesis.

a) Prueba de Breush-Pagan

Esta prueba propuesta en 1979 por T.S. Breush y A.R. Pagan parte del ajuste del modelo lineal múltiple

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \\ i &= 1, 2, \dots, n \end{aligned}$$

Se propone un modelo lineal múltiple en el cual las varianzas de la serie ordenada se pueden expresar en función de una serie de variables Z_i , que en la práctica se sustituyen por las variables independientes X_i .

$$\begin{aligned} \sigma_i^2 &= \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \dots + \alpha_k X_{ki} + v_i \\ i &= 1, 2, \dots, n \end{aligned}$$

Se plantea como hipótesis nula, la cual implica homoscedasticidad que al ajustar el modelo anterior el vector de los coeficientes α_i no es significativamente diferente de cero.

$$\begin{aligned} H_0 : \alpha_1 &= \alpha_2 = \alpha_3 = \dots = \alpha_k = 0 \\ i &= 1, 2, \dots, n \end{aligned}$$

Procedimiento de prueba

a) Se inicia con el ajuste del modelo propuesto y la obtención de los residuales e_i

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \dots \dots \dots \hat{\beta}_k X_{ki} + e_i$$
$$i = 1, 2, \dots, n$$

b) Se calcula el estimador de máxima verosimilitud de la varianza de los errores:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n}$$

c) Se estandarizan los residuales

$$e_i^* = \frac{e_i}{\hat{\sigma}}$$

d) Se calcula la regresión de los residuales estandarizados al cuadrado sobre las variables independientes:

$$e_i^{*2} = \gamma_1 + \gamma_2 X_{2i} + \gamma_3 X_{3i} + \dots \dots \dots + \gamma_k X_{ki} + d_i$$

Para el modelo ajustado se calcula la suma de cuadrados de la regresión y se divide entre 2. Esta estadística, bajo la hipótesis de normalidad y homoscedasticidad se distribuye Ji cuadrada con k-1 grados de libertad (el número de variables explicativas). Si la estadística resulta significativa, se concluye la existencia de heteroscedasticidad.

$$\frac{SCR}{2} \rightarrow X_{k-1}^2$$

b) Prueba de White

Se plantean las hipótesis nula y alternativa en la siguiente forma:

$$H_0 : \sigma_i^2 = \sigma^2 \text{ para toda } i \text{ Homoscedasticidad}$$

$$H_a : \sigma_i^2 \neq \sigma^2 \text{ para al menos una } i \text{ Heteroscedasticidad}$$

Se parte del ajuste del siguiente modelo por MCO.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \dots \dots \beta_k X_{ki} + u_i$$
$$i = 1, 2, \dots, n$$

Se estima una regresión múltiple considerando como variable dependiente el cuadrado del residual del ajuste del modelo anterior y como variables independientes las variables explicativas originales, con la adición de variables definidas como cuadrados de las originales y productos cruzados 2 a 2. Desde luego hay que vigilar que el número de observaciones supere conservadoramente al número de coeficientes a estimar.

$$e_i^2 = \alpha_1 + \alpha_{21} X_{2i} + \alpha_{31} X_{3i} \dots \alpha_k X_{ki} + \alpha_{22} X_{2i}^2 + \alpha_{32} X_{3i}^2 + \dots + \alpha_{23} X_{2i} X_{3i} + \dots + \alpha_{K-1,K} X_{K-1,i} X_{ki} + v_i$$
$$i = 1, 2, \dots, n$$

La estadística de prueba es la R^2_{e2} asociada a esta regresión, multiplicada por el número de observaciones. La estadística de prueba se distribuye Ji cuadrada con grados de libertad igual al número de variables explicativas involucradas en la regresión de los residuales.

$$\lambda = nR_{e2}^2$$

La prueba de White es sensible a la heteroscedasticidad de cualquier tipo, pero no identifica claramente su origen.

c) Prueba del Coeficiente de Correlación de Spearman

Esta prueba se apoya en el cálculo del coeficiente de correlación por rangos de Spearman. El argumento es que si existe correlación lineal entre los rangos de los valores absolutos de los residuales del modelo original ajustado por MCO y las variables explicativas, entonces habrá una relación lineal significativa.

El coeficiente de correlación de Spearman se calcula de la siguiente forma:

Se toman los valores absolutos de los residuales $|e_i|$ y se ordenan de menor a mayor con la identificación de sus correspondientes rangos de orden. Se procede de manera semejante con las variables explicativas una por una sin considerar su valor absoluto.

Se calculan las diferencias entre rangos correspondientes de los valores de las variables explicativas y los valores absolutos de los residuales.

$$d_i = \text{Rango}(X_{ij}) - \text{Rango}(|e_i|)$$

A partir de las diferencias se calcula el coeficiente de correlación de Spearman y la estadística de prueba que se aproxima a una t de Student.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

La hipótesis nula establece que el coeficiente de correlación de Spearman sea nulo (ausencia de heteroscedasticidad) y la alternativa que sea distinto de cero (evidencia de heteroscedasticidad). La estadística de prueba se distribuye aproximadamente como t de Student con n-2 grados de libertad.

$$T = r_s \sqrt{\frac{n-2}{1-r_s^2}} \rightarrow t_{n-2} \text{ gl}$$

d) Prueba de Goldfeldt Quandt

Esta prueba supone que la varianza σ_i^2 depende de una variable, usualmente una de las variables explicativas y que la varianza crece o disminuye en función de esa variable. Entonces la heteroscedasticidad tendrá como efecto que la varianza calculada para las observaciones menores de la variable causal sea diferente de la varianza calculada para las observaciones mayores.

$$H_0 : \sigma_{Min}^2 = \sigma_{Max}^2$$

$$H_a : \sigma_{Min}^2 \neq \sigma_{Max}^2$$

El procedimiento de prueba consiste en lo siguiente:

1. Se procede a ordenar las observaciones de menor a mayor en base a la variable X_i que se sospecha causante de la heteroscedasticidad.
2. Se omiten c observaciones centrales, donde c es del orden de n/3.

3. Se estima dos veces el modelo original. Una vez con las primeras $(n-c)/2$ observaciones menores y la segunda vez con las $(n-c)/2$ observaciones mayores. Hay que considerar que el cociente $(n-2)/2$ debe ser conservadoramente mayor que el número de coeficientes a estimar.

4. Se obtienen las sumas de cuadrados del error de ambos modelos ajustados y se calcula el siguiente cociente, colocando en el numerador la suma mayor.

$$GQ = \frac{SR_1}{SR_2}$$

La estadística de prueba GQ se distribuye F con $((n-c)/2-k, (n-c)/2-k)$ donde k es el número de variables explicativas.

e) Prueba de Glejser

La prueba de Goldfeldt Quandt intenta identificar la presencia de heteroscedasticidad creciente o decreciente, pero la prueba de Glejser introduce el refinamiento de tratar de identificar la forma funcional de la heteroscedasticidad $\sigma_i^2 = f(X_i)$.

Se plantean los valores absolutos de los residuales como formas funcionales de combinaciones lineales o cuadráticas de las variables explicativas que se estiman por MCO.

$$|e_i| = \gamma_0 + \gamma_1 X_1^k + v_i \quad i=1,2,\dots,n \quad \text{y usualmente } k = -1, -1/2, 1/2, 1$$

y se prueba la hipótesis: $H_0: \gamma_1 = 0$ vs $H_a: \gamma_1 \neq 0$, si se rechaza entonces hay evidencia de heteroscedasticidad y se identifica además la forma funcional. La forma funcional puede ser planteada en forma más amplia por el investigador.

2.6 Medidas Remediales para la Heteroscedasticidad

2.6.1 Mínimos Cuadrados Generalizados.

Se ha mencionado que el modelo lineal múltiple supone que las perturbaciones tienen media cero y varianza constante. La heteroscedasticidad rompe con el segundo supuesto y como consecuencia los estimadores que se calculan por mínimos cuadrados ordinarios si bien mantienen las propiedades de insesgamiento y consistencia, se vuelven ineficientes y la estimación de la matriz de varianzas y covarianzas se torna sesgada, pues otorgan pesos iguales a todas las observaciones, en resumen dejan de ser MELI. Una forma de retornar a estimadores MELI es mediante los mínimos cuadrados generalizados (MCG), los cuales constituyen una forma de mínimos cuadrados ponderados al incorporar pesos diferentes a las observaciones en función de su variabilidad.

Las propiedades de los errores se tornan en las siguientes:

$$E(u) = 0$$

$$E(uu') = \sigma^2 \Omega$$

Considérese inicialmente que se tiene la siguiente forma de la matriz de varianzas y covarianzas.

$$\sigma^2 \Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_i^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}$$

Si se está la afortunada, pero poco probable situación, de que se conozcan las varianzas σ_i^2 . El problema de la heteroscedasticidad se puede corregir mediante la división de los errores entre sus respectivas varianzas $E(u_i^2) = \sigma_i^2$. Las consecuencias serán errores con media cero y varianza 1.

$$E\left(\frac{u_i}{\sigma_i^2}\right) = 0 \quad E\left(\frac{u_i^2}{\sigma_i^2}\right) = \frac{1}{\sigma_i^2} E(u_i^2) = 1$$

El modelo ponderado adopta entonces la siguiente expresión.

$$\frac{Y_i}{\sigma_i^2} = \beta_1^* \frac{1}{\sigma_i^2} + \beta_2^* \frac{X_{2i}}{\sigma_i^2} + \beta_3^* \frac{X_{3i}}{\sigma_i^2} \dots \beta_k^* \frac{X_{ki}}{\sigma_i^2} + \frac{u_i}{\sigma_i^2}$$

$$i = 1, 2, \dots, n$$

Considere los ponderadores $W_i = \frac{1}{\sigma_i^2}$

El modelo transformado que incluye la variable W_i se ajusta por MCO. Observe que el modelo transformado carece del término constante en el origen.

$$Y_i^* = \beta_1^* W_i + \beta_2^* X_{2i} + \beta_3^* X_{3i} \dots \beta_k^* X_{ki} + u_i^*$$

$$i = 1, 2, \dots, n$$

El efecto de la regresión ponderada es restar importancia dentro del modelo a las observaciones con mayor varianza, pues éstas se dividen por un número mayor.

Se ha mencionado que es poco probable que se tenga conocimiento de las varianzas σ_i^2 , pero por ejemplo, si al aplicar la prueba de Gesjler se identifica que la desviación estándar es proporcional a $X_{2i} = \lambda \sigma_i$, entonces se divide el modelo entre X_{2i}

$$\frac{Y_i}{X_{2i}} = \beta_1^* \frac{1}{X_{2i}} + \beta_2^* + \beta_3^* \frac{X_{3i}}{X_{2i}} \dots \beta_k^* \frac{X_{ki}}{X_{2i}} + \frac{u_i}{X_{2i}}$$

$$i = 1, 2, \dots, n$$

Ahora el ponderador adopta la forma $W_i = \frac{1}{X_{2i}}$

El modelo resultante se podrá estimar por MCO y tendrá como término constante el coeficiente de X_2 y varianza constante del error.

$$E\left(\frac{u_i^2}{X_{2i}^2}\right) = E\left(\frac{u_i^2}{\lambda^2 \sigma_i^2}\right) = \frac{1}{\lambda^2}$$

El procedimiento, desde luego puede ser aplicado a otras formas funcionales.

2.6.2 Transformación Logarítmica sobre las variables

Considérese un modelo no lineal, pero linealizable en el que el término del error tiene un efecto multiplicativo y que en términos relativos es constante, por ejemplo de 3% respecto de la variable independiente, pero en términos absolutos el efecto es creciente y provoca heteroscedasticidad.

$$Y = \alpha X^\beta u$$

Al tomar logaritmos el efecto es aditivo, pero puesto que la función logaritmo se incrementa en forma menos que proporcional, el efecto del error correspondiente a los valores altos de X será menor y además reducirá la heteroscedasticidad.

$$\log(Y) = \log(\alpha) + \beta \log(X) + \log(u)$$

El modelo adopta una forma cuyas variables transformadas pueden estimarse por mínimos cuadrados ordinarios.

2.7 Autocorrelación

Definición y efectos.

El modelo lineal múltiple supone que las perturbaciones u_i tienen media cero, varianza constante y son independientes entre sí. Se supone que la perturbación de una observación no es influenciada por la perturbación de otra observación. Se supone entonces la ausencia de correlación serial o autocorrelación. En términos más formales se expresa a través de las siguientes propiedades:

$$E(u) = 0$$

$$E(uu') = \sigma^2 I = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

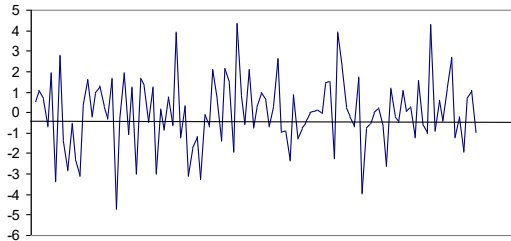
Pero la segunda propiedad puede violarse al no corresponder a una matriz diagonal, sino a una matriz que presenta varianza constante en la diagonal y covarianzas distintas de cero entre los errores., esto es se viola su independencia.

$$E(uu') = \begin{bmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \dots & \sigma_{2n} \\ \dots & 0 & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \frac{\sigma_{12}}{\sigma^2} & \dots & \frac{\sigma_{1n}}{\sigma^2} \\ \frac{\sigma_{21}}{\sigma^2} & 1 & \dots & \frac{\sigma_{2n}}{\sigma^2} \\ \dots & \dots & \dots & \dots \\ \frac{\sigma_{n1}}{\sigma^2} & \frac{\sigma_{n2}}{\sigma^2} & \dots & 1 \end{bmatrix} = \sigma^2 \Omega$$

En forma equivalente sucede $E(u_i u_j) \neq 0$ para $i \neq j$

Una primera evidencia empírica de la presencia de autocorrelación se presenta en los residuales. Se toman los residuales ordenados respecto al tiempo, o al valor de una variable explicativa. En caso de ausencia de autocorrelación la gráfica de la serie presenta un patrón aleatorio similar al siguiente en torno al cero. Con este comportamiento no es factible predecir con alta probabilidad el signo de un residual a partir del anterior.

Residuales no correlacionados

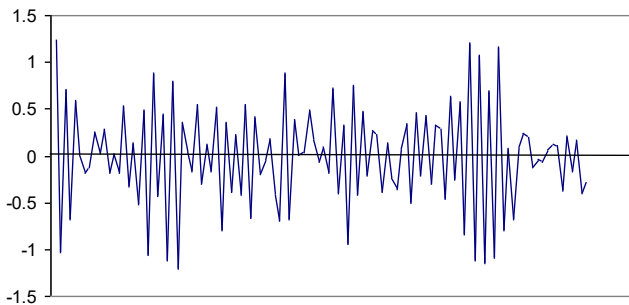
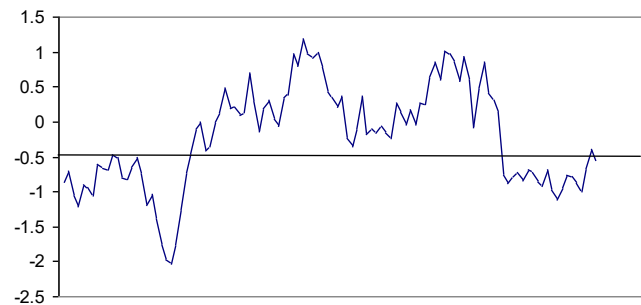


Sin embargo, ante una fuerte autocorrelación positiva, que es la más frecuente, los residuales tendrán la tendencia a permanecer por encima o por abajo del cero, pues a un residual positivo seguirá con alta probabilidad un residual también de signo positivo y uno negativo tendrá un comportamiento similar con otro negativo.

Correlación positiva $r = 0.9$

Si la autocorrelación es negativa la serie presentará cambios de signo muy abruptos y frecuentes, pues un residual positivo tenderá a ser seguido por un residual negativo y viceversa.

Autocorrelación $r = -0.8$



La consecuencia inmediata de la presencia de autocorrelación sobre los estimadores de mínimos cuadrados es que éstos pierden su condición de MELI, si bien se mantienen insesgados, la matriz de varianzas y covarianzas de los estimadores se ve afectada.

$$E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'u) = \beta$$

$$V(\hat{\beta}) = \sigma^2((X'X)^{-1}X'\Omega X(X'X)^{-1})$$

Consideremos como ejemplo el modelo lineal simple:

$$Y_i = \alpha + \beta X_i + u_i$$

El estimador de mínimos cuadrados del coeficiente de regresión se puede expresar de la siguiente forma:

$$V(\hat{\beta}) = E \left(\sum_{i=1}^n \frac{x_i}{\sum_{i=1}^n x_i^2} u_i^2 + 2 \sum_{i<j}^{n-1} \sum_j^n \frac{x_i}{\sum_{i=1}^n x_i^2} \frac{x_j}{\sum_{i=1}^n x_i^2} u_i u_j \right)$$

$$V(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} + 2 \sum_{i<j}^{n-1} \sum_j^n \frac{x_i}{\sum_{i=1}^n x_i^2} \frac{x_j}{\sum_{i=1}^n x_i^2} E(u_i u_j) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \left(1 + 2 \sum_{i<j}^{n-1} \sum_j^n \frac{x_i x_j}{\sum_{i=1}^n x_i^2} E(u_i u_j) \right)$$

En caso de cumplirse $E(u_i u_j) = 0$ la fórmula se reduce a la ya conocida varianza del estimador del coeficiente de regresión, sin embargo, si $E(u_i u_j) \neq 0$, y considerando que muy probablemente $E(u_i u_j) > 0$, la varianza se subestimará notablemente con los obvios efectos sobre las pruebas de hipótesis e intervalos de confianza sobre los coeficientes de regresión. Además la estimación misma de la varianza del error σ^2 será subestimada.

2.7.1 Causas de la Autocorrelación.

- 1) **Ciclos.** Muchas variables económicas presentan comportamientos cíclicos de corto plazo debido a la estacionalidad de los fenómenos.
- 2) **Errores de especificación.** Pueden deberse a formas funcionales incorrectas, por ejemplo el que el comportamiento de una variable responda a un modelo cuadrático y se la especifique lineal. También la falta de inclusión en el modelo de una variable relevante puede tener efectos similares.
- 3) **Rezagos.** El comportamiento de una variable es afectada por valores rezagados y ese tipo de influencia no se haya reflejado en el modelo propuesto.
- 4) **Variables estimadas.** Con frecuencia las series temporales no corresponden a observaciones estrictas en cada período, sino que períodos intermedios son resultado de interpolaciones u otro tipo de combinaciones que disminuyen la variabilidad natural de una variable explicativa y crean dependencia en períodos cortos de unas con otras.

2.7.2 Formas de detectar la Autocorrelación.

1) Estadística de Durbin-Watson

Como en el caso de la heteroscedasticidad, existen diversos procedimientos para detectar y probar la hipótesis de autocorrelación. La prueba más conocida es la de Durbin-Watson (1950), la cual permiten verificar la presencia de autocorrelación de primer orden.

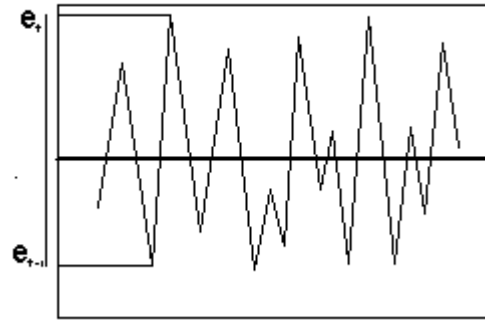
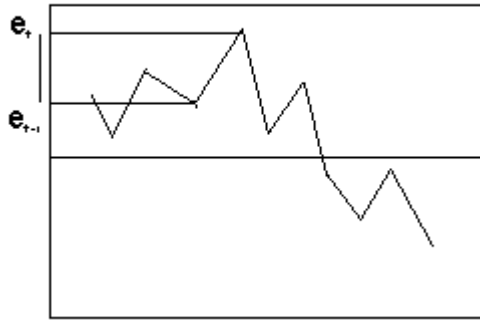
$$\text{Sea } u_t = \rho u_{t-1} + v_t$$

$$H_0: \rho=0 \text{ vs } H_a: \rho \neq 0$$

La estadística de prueba se basa en la suma de productos de los residuales sucesivos dividida entre su suma de cuadrados.

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

Como se ha visto la autocorrelación positiva se caracteriza por residuales del mismo signo, como consecuencia las diferencias y su cuadrado tienden a ser pequeños y su contribución en el numerador de la estadística DW es pequeña. La autocorrelación negativa, por el contrario al ser más frecuentes los signos contrarios, tiende a generar diferencias sustanciales y su contribución al numerador de la DW es importante.



Consideramos que empíricamente el coeficiente de correlación de primer orden entre los residuales se puede calcular como sigue:

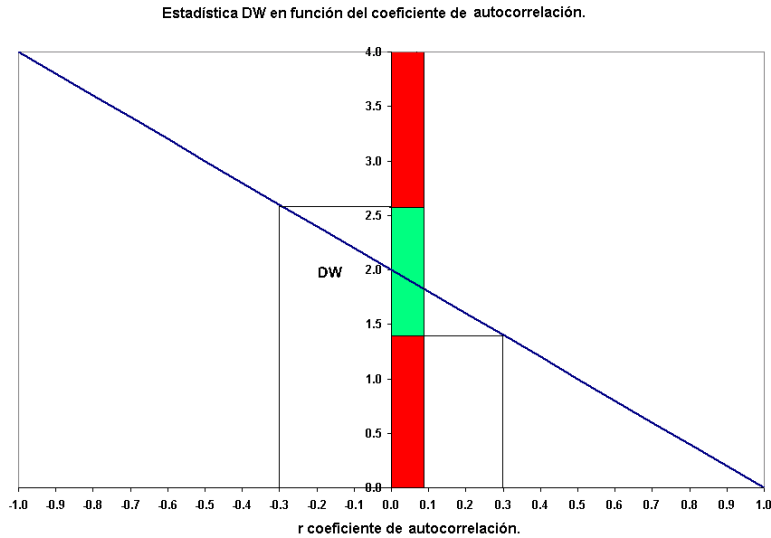
$$r_1 = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$$

Al desarrollar el cuadrado del numerador de la estadística DW y aproximando a uno los dos primeros cocientes se tiene:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = \frac{\sum_{i=2}^n e_i^2}{\sum_{i=1}^n e_i^2} + \frac{\sum_{i=2}^n e_{i-1}^2}{\sum_{i=1}^n e_i^2} + \frac{2 \sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2} \approx 1 + 1 + 2\rho$$

$$DW \approx 2(1 + \rho)$$

Al aplicar esta fórmula de aproximación, se puede observar que en los casos extremos en que $\rho = -1$ y $\rho = 1$ la DW adoptará los valores 4 y 0 respectivamente. Si $\rho = 0$ entonces $DW = 2$. Así a medida que la DW se aleja de 2 en cualquiera de los dos sentidos se puede tener evidencia de autocorrelación. Los límites superior d_U e inferior d_L de la DW se presentan tabulados por los autores en función del nivel de significancia, número de variables en el modelo incluido el término constante y tamaño de la muestra.



2) La h de Durbin

La estadística DW supone que las variables explicativas son determinísticas y que se incluye un término constante. Cuando el modelo se define en función de rezagos de la variable dependiente la DW no detecta la autocorrelación, pues tiende a ubicarse cerca de 2 aun cuando exista autocorrelación serial. Por ejemplo, un modelo de la forma:

$$Y_t = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 Y_{t-1} + \beta_5 Y_{t-2} + u_t$$

$$u_t = \rho u_{t-1} + v_t \quad \text{donde} \quad |\rho| < 1$$

El mismo Durbin, propuso la estadística h, cuya distribución se aproxima a una normal estándar.

$$h = \hat{\rho} \sqrt{\frac{n}{1 - nV(\hat{\beta}_4)}}$$

Donde

$\hat{\rho}$ es la estimación del coeficiente de autocorrelación de primer orden $\hat{\rho} \sim 1 - \frac{dw}{2}$
n es el número de observaciones

$V(\hat{\beta}_4)$ es la varianza estimada del coeficiente del primer retardo de la variable dependiente.

El modelo se ajusta por mínimos cuadrados ordinarios y la estimación del coeficiente de autocorrelación se efectúa a partir de los residuales calculados.

Hay que enfatizar que solamente se toma en cuenta el coeficiente del primer retardo de la variable dependiente, aun cuando el modelo incluya más retardos y que si la varianza del coeficiente del primer retardo es muy grande, el denominador puede ser negativo y la estadística sea no estimable.

3) Prueba de Rachas

Como se mencionó el signo de los residuales consecutivos tiende a ser homogéneo cuando hay autocorrelación positiva y cambiar con mucha frecuencia cuando la autocorrelación es negativa. Ello lleva a la posibilidad de aplicar la prueba no paramétrica de rachas de los signos positivos y negativos de los residuales. La prueba supone que si el patrón es aleatorio no habrá ni pocas ni excesivas rachas.

El número esperado de rachas positivas y negativas está dado por

$$E(R) = 1 + \frac{2n_+ n_-}{n_+ + n_-}$$

y la varianza del número de rachas se expresa:

$$V(R) = \frac{2n_+ n_- (2n_+ n_- - n_+ - n_-)}{(n_+ + n_-)^2 (n_+ + n_- - 1)}$$

Para muestras pequeñas, F.S. Swed y C. Eisenhart (1943) elaboraron tablas para verificar la significancia. Si el tamaño de muestra es suficientemente grande, entonces se puede construir una estadística Z que se aproxima a una normal estándar y probar en la forma usual.

$$Z(R) = \frac{r - E(R)}{\sqrt{V(R)}}$$

2.7.3 Medidas Remediales para la Autocorrelación.

Se conoce la estructura de la autocorrelación.

La forma de la autocorrelación es motivo de un supuesto, pues los errores u_i son variables aleatorias no observables. Frecuentemente se supone el esquema autorregresivo de primer orden para los errores:

$$u_t = \rho u_{t-1} + v_t$$

En el hipotético caso de conocer ρ considere el modelo siguiente:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 Y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + v_t \quad \text{donde} \quad |\rho| < 1$$

Se multiplica el modelo valuado en t-1 por ρ

$$\rho Y_{t-1} = \rho \beta_1 + \rho \beta_2 X_{2,t-1} + \rho \beta_3 X_{3,t-1} + \rho \beta_4 Y_{t-2} + \rho u_{t-1}$$

Se toma la diferencia entre ambas expresiones

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_{2t} - \rho X_{2,t-1}) + \beta_3(X_{3t} - \rho X_{3,t-1}) + \beta_4(Y_{t-1} - \rho Y_{t-2}) + (u_t - \rho u_{t-1})$$

El modelo transformado conocido como de *diferencias generalizadas* que se ajusta por MCO.

$$Y_t^* = \beta_1^* + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + \beta_4 Y_{t-1}^* + u_t^*$$

Al aplicar las diferencias se pierde al menos la primera observación, pues no hay contra que tomar la diferencia, para evitar eso se sugiere la transformación:

$$Y_1 \sqrt{1 - \rho^2} \quad X_1 \sqrt{1 - \rho^2}$$

1) Estimación simple de ρ

Si el valor de ρ es desconocido, lo más probable, entonces una forma práctica de proceder es correr el modelo original, no transformado y a partir de los residuales estimar el valor de ρ

2) Estimación de ρ a partir de la estadística DW.

Hemos verificado que la estadística DW y el coeficiente de autocorrelación de primer orden se relacionan de la siguiente forma:

$$DW = 2(1 - \rho)$$

El valor de la DW es proporcionado habitualmente por la mayoría de los paquetes estadísticos y es fácil despejar ρ y proceder con esta estimación.

$$\rho = 1 - \frac{DW}{2}$$

Como este es un valor aproximado, es posible que sea poco precisa en muestras pequeñas. Theil y Nagar sugirieron una forma corregida que toma en cuenta el tamaño de la muestra y el número de variables explicativas en la ecuación del modelo incluyendo el término independiente.

$$\rho = \frac{n^2 \left(1 - \frac{DW}{2}\right) + K^2}{n^2 - K^2}$$

3) Estimación iterativa de Cochrane y Orcutt

Sugieren tener una estimación inicial de ρ de manera la manera simple, pero una vez ajustado el nuevo modelo corregido calcular una segunda etapa de estimación del coeficiente de correlación a partir de los residuales del modelo transformado. Este coeficiente de segunda etapa resultaría mejor que el anterior. El coeficiente mejorado de segunda etapa se utilizaría para el modelo de diferencias inicial. El procedimiento puede continuar con etapas subsecuentes análogas, hasta que la diferencia entre coeficientes de autocorrelación de etapas consecutivas sea insignificante.

3 EXTENSIONES DEL MODELO. Transformaciones y Variables Ficticias.

3.1 Transformaciones.

El adjetivo en el modelo lineal hace referencia a los coeficientes y no a las variables, por tanto es admisible adoptar diversas funciones de las variables para definir nuevas variables y mejorar la especificación. Ello es de gran utilidad, pues es frecuente que se analicen fenómenos económicos o naturales cuyo comportamiento no es posible representarlo mediante una forma lineal. En tal caso se tienen dos alternativas: ajustar un modelo mediante procedimientos no lineales o aplicar transformaciones que linealicen un modelo y proceder con los procedimientos de ajuste usuales. Las transformaciones utilizadas con mayor frecuencia son la logarítmica, potencias y la recíproca. Hay que tener cuidado al aplicar transformaciones, pues pueden tener efectos secundarios indeseables y dar lugar a violaciones de los supuestos básicos del modelo como falta de normalidad y heteroscedasticidad.

Algunas formas de modelos usuales y sus transformaciones son las siguientes:

Modelo original	Modelo transformado
$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} \dots X_{ki}^{\beta_k} u_i$	$\ln(Y_i) = \ln(\beta_1) + \beta_2 \ln(X_{2i}) + \beta_3 \ln(X_{3i}) + \dots \beta_k \ln(X_{ki}) + \ln(u_i)$
$Y_i = \exp(\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} + u_i)$	$\ln(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} + u_i$
$Y_i = \frac{1}{\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} + u_i}$	$\frac{1}{Y} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} + u_i$
$Y_i = \frac{1}{1 + \beta_1 \exp(\beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} + u_i)}$	$\ln\left(\frac{1}{Y} - 1\right) = \ln(\beta_1) + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} + u_i$

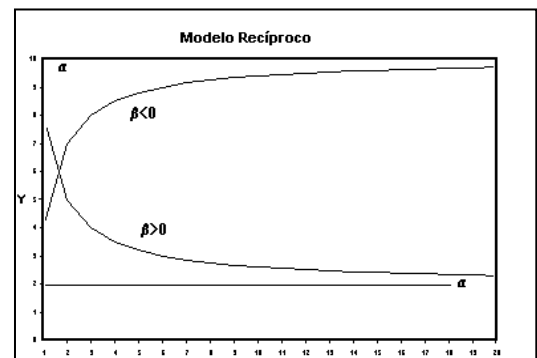
Algunos modelos para dos variables muy populares se mencionan a continuación:

3.2.1 Recíproco

La forma original del modelo es

$$Y = \alpha + \beta/X + u$$

La magnitud y signo de β definen el comportamiento de la curva y el valor de α es la asíntota que limita la función. En las siguientes gráficas se presentan ejemplos de la función para diferentes valores de β



3.2.2 Modelo Semi - log

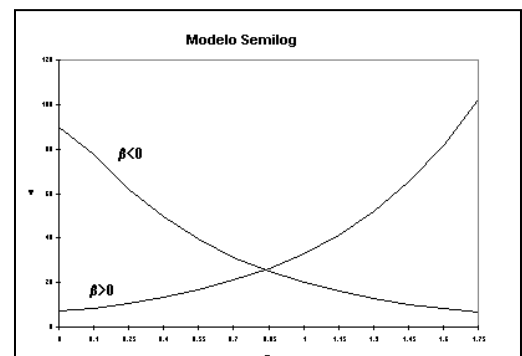
En este caso el modelo original es de la forma:

$$Y = \alpha \beta^X u$$

Si la expresión se transforma mediante logaritmos, que desde luego estarán restringidos a valores positivos se obtiene una expresión en la forma ya conocida del modelo:

$$\text{Log}(Y) = \text{Log}(\alpha) + \text{Log}(\beta)X + \text{Log}(u)$$

$$Y' = \alpha' + \beta'X + u'$$



Este modelo puede ser expresado alternativamente como:

$$Y = e^{(\alpha + \beta X + u)}$$

El ascenso o descenso de la función depende del signo de β y la velocidad de su valor absoluto. La ordenada al origen corresponde a e^α .

3.2.3 Modelo Doble - log

La expresión es similar al caso anterior:

$$Y = \alpha X^\beta u$$

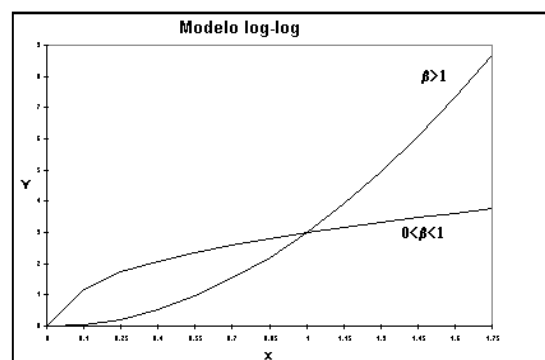
Esta expresión se transforma también tomando logaritmos para linealizar

$$\text{Log}(Y) = \text{Log}(\alpha) + \text{Log}(X)\beta + \text{Log}(u)$$

$$Y' = \alpha' + \beta X' + u'$$

El comportamiento de este modelo depende del signo y magnitud de β . En las siguientes gráficas se pueden observar los efectos de diferentes valores del parámetro. El cual se puede interpretar como la elasticidad.

Tanto en este modelo, como en el anterior se debe observar que originalmente el error es multiplicativo y no aditivo y hay que verificar si al transformar son válidos los supuestos generales del modelo lineal.



3.3 Modelo Logístico

Es un modelo muy popular, que merece atención especial, se caracteriza por estar limitado por dos asíntotas, una superior correspondiente a un valor k y una asíntota inferior al nivel del cero. El modelo logístico ha sido muy utilizado para ajustar tendencias de crecimiento.

$$Y = \frac{K}{1 + \alpha e^{-\beta X}}$$

El modelo de crecimiento geométrico no puede representar de manera continua el comportamiento de una población, ya que la limitación de espacio y de alimentos disminuye la tasa de crecimiento. Esta reducción de la tasa de crecimiento puede provocar que eventualmente la población cese de crecer una vez que se alcanza un punto de saturación del medio ambiente.

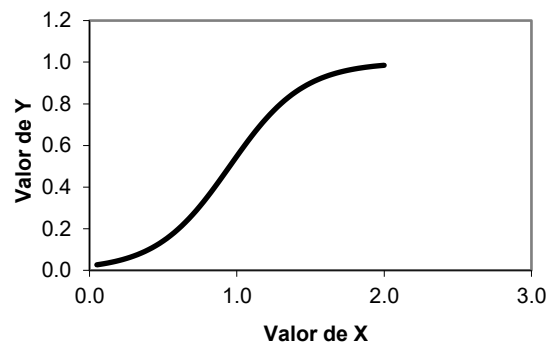
Esta situación puede ser descrita por un tipo de función sigmoide como la de la gráfica

Como se observa en la gráfica la función crece inicialmente a una tasa baja y se eleva gradualmente. Al final la tasa decrece en forma tal que la función es acotada por una asíntota k a la cual se le acerca suavemente.

La función más simple para lograr el comportamiento sigmoide es mediante la introducción de un término que reduzca suavemente la tasa de crecimiento. La relación de crecimiento geométrico o exponencial modificada quedaría como sigue:

$$\frac{\partial N_t}{\partial t} = rN_t \left(\frac{k - N_t}{k} \right)$$

Curva Logística



Donde

Nt	Tamaño de la población al tiempo t
t	Tiempo
r	Tasa de crecimiento
k	Constante asintótica limita el valor máximo de N

Debe observarse que el término $\frac{k - Nt}{k}$ con $k \geq Nt$ se anula cuando $Nt=k$ y la población interrumpe su crecimiento en ese punto.

La ecuación anterior representa la forma diferencial de la curva logística. Esta curva fue sugerida para caracterizar el comportamiento del crecimiento de poblaciones humanas por Verhulst en 1838.

La forma integral de la ecuación logística puede ser descrita como sigue:

$$Nt = \frac{k}{1 + e^{a-rt}}$$

Donde a es una constante de integración que define la posición de la curva relativa al origen y e es la base de logaritmos naturales.

Para linealizar la función y así poder ajustar por mínimos cuadrados para estimar los parámetros a y r se procede inicialmente a despejar la parte exponencial

$$\begin{aligned} 1 + e^{a-rt} &= \frac{k}{Nt} \\ e^{a-rt} &= \frac{k}{Nt} - 1 \\ e^{a-rt} &= \frac{k - Nt}{Nt} \end{aligned}$$

Si a continuación se toma logaritmo natural de ambos miembros, se obtiene

$$a - rt = \log\left(\frac{k - Nt}{Nt}\right)$$

Expresión que tiene analogía con $Y = \alpha + \beta X$ como se presenta a continuación

$$\log\left(\frac{k - Nt}{Nt}\right) = a - rt$$

donde $y = \log\left(\frac{k - Nt}{Nt}\right)$, $\alpha = a$ y $-r$ corresponde a la pendiente β . En este punto es fácil aplicar mínimos cuadrados para estimar a y r .

La curva logística presenta algunos atractivos desde el punto de vista matemático y en su interpretación biológica. Primero, su expresión es relativamente simple. En segundo término, se percibe bastante realista. Tercero, es relativamente simple pues la forma diferencial de la curva logística contiene solamente tres constantes r , a y k . Las constantes r y k se pueden interpretar biológicamente. La r como la tasa de crecimiento poblacional y k como el punto de saturación poblacional.

Actualmente el modelo logístico se utiliza frecuentemente para estimar probabilidades de ocurrencia de cierta característica. La variable dependiente Y es dicotómica (0,1) lo cual indica que se tiene o no la característica de interés y la variable o variables independientes toman medidas en nivel intervalar o de razón. Ello da lugar a requerimientos de estimación mediante máxima verosimilitud.

Como ejemplo de ajuste del modelo logístico se tomarán los datos de un experimento de crecimiento de levadura (en gramos) realizada por Carlson (1913). A partir del análisis de los datos se estima la k de manera un tanto heurística en 665 y posteriormente se efectúan las transformaciones indicadas anteriormente para obtener estimaciones por mínimos cuadrados de los parámetros a y r. Con este procedimiento se tiene

$$a = 4.1635$$

$$r = 0.5306$$

El coeficiente de determinación R^2 para el modelo transformado fue calculado en 0.9996. Los cálculos se resumen en las siguientes tablas

Ajuste de una curva logística
Crecimiento de Levadura (Biomasa)

Horas T	Unidades de levadura Nt	$\log((k-Nt)/Nt)$	Unidades Estimadas $^{\wedge}Nt$	$Nt-^{\wedge}Nt$
0	9.6	4.2235	10.18	-0.5836
1	18.3	3.5650	17.13	1.1706
2	29.0	3.0879	28.61	0.3946
3	47.2	2.5718	47.21	-0.0098
4	71.1	2.1226	76.46	-5.3603
5	119.1	1.5225	120.30	-1.2047
6	174.6	1.0327	181.54	-6.9358
7	257.3	0.4603	259.11	-1.8068
8	350.7	-0.1096	346.10	4.6041
9	441.0	-0.6774	431.26	9.7403
10	513.3	-1.2190	504.24	9.0558
11	559.7	-1.6706	559.99	-0.2887
12	594.8	-2.1369	598.94	-4.1356
13	629.4	-2.8724	624.48	4.9172
14	640.8	-3.2764	640.55	0.2459
15	651.1	-3.8468	650.40	0.7003
16	655.9	-4.2777	656.33	-0.4337
17	659.6	-4.8052	659.87	-0.2749
18	661.8	-5.3318	661.98	-0.1758

Ajuste por Mínimos	Cuadrados
Constante a	4.1635721015
Err.Estd. de Y Est	0.0610788378
R^2	0.9996050634
No. de Observaciones	19
Grados de Libertad	17
Tasa de Crecimiento -r	-
Err.Estd. de Coeficiente	0.5306746627
	0.002558311

Krebs (1978) afirmó que una población sigue un patrón de crecimiento logístico y se hacen implícitamente una serie de supuestos:

- 1) La población tiene una distribución de edad inicial estable. El modelo asume que una población empieza a crecer cuando la razón $(K-N)/K$ es cercana a 1.0 y se incrementa a una tasa aproximada rN
- 2) La densidad se mide en unidades adecuadas. Por ejemplo en el caso de las moscas no basta contar la población adulta, sino además los individuos en estado larvario y huevos. Para evitar problemas es mejor medir la biomasa total en este caso (gramos).
- 3) La relación entre la densidad y la tasa de crecimiento lineal. Esto se aprecia mejor al considerar a la logística en su forma diferencial

$$\frac{\partial Nt}{\partial t} \frac{1}{Nt} = r - \frac{r}{k} Nt$$

Esta relación nos indica que el incremento poblacional por individuo es una función lineal de la densidad $\frac{Nt}{k}$.

- 4) La influencia de la disminución de la densidad sobre la tasa de crecimiento opera instantáneamente, sin retrasos en el tiempo. Sin embargo, es altamente improbable que en organismos con ciclos de vida complejos la tasa de incremento reaccione instantáneamente a cambios en la densidad.

En poblaciones naturales los supuestos del modelo logístico rara vez se podrán cumplir completamente. Los datos recolectados de campo podrán dar indicaciones de la media en que los supuestos se consideren válidos. Debemos recordar que todo modelo es una idealización del comportamiento de la naturaleza.

El modelo logístico ha sido utilizado con profundidad en fenómenos de respuesta discreta binaria y multinomial a los cuales se retornará posteriormente.

3.4 Función de Producción de Cobb Douglas.

En el análisis económico es muy popular la función de producción propuesta en 1929 por Paul Douglas y Charles Cobb en el artículo *A Theory of Production* publicado en el *American Economic Review*. La función explica el monto de producción en función de los factores trabajo y capital.

$$Q = AL^\alpha K^\beta u$$

Donde **Q** es la cantidad producida

L es el insumo de trabajo

K es el insumo de capital

A es un constante

α y **β** son constantes que representan las elasticidades de la producción respecto del trabajo y capital respectivamente.

u es un error aleatorio.

Si los exponentes α y β cumplen la condición $\alpha + \beta = 1$ se tienen rendimientos constantes a escala, esto es que si por ejemplo L y K se incrementan en una proporción λ entonces Q se incrementa en la misma proporción λ . Ello significa que la función de producción es homogénea de grado 1

$$A(\lambda L)^\alpha (\lambda K)^\beta u = A\lambda^\alpha L^\alpha \lambda^\beta K^\beta = A\lambda^{\alpha+\beta} L^\alpha K^\beta = \lambda AL^\alpha K^\beta = \lambda Q$$

El modelo admite sin embargo que si $\alpha + \beta > 1$ se tengan rendimientos crecientes a escala, esto es que a un incremento λ de L y K corresponde un incremento mayor a λ para Q . Si $\alpha + \beta < 1$ entonces los rendimientos serán decrecientes a escala.

Otro atractivo que brinda esta forma funcional es que los exponentes se pueden interpretar como las elasticidades.

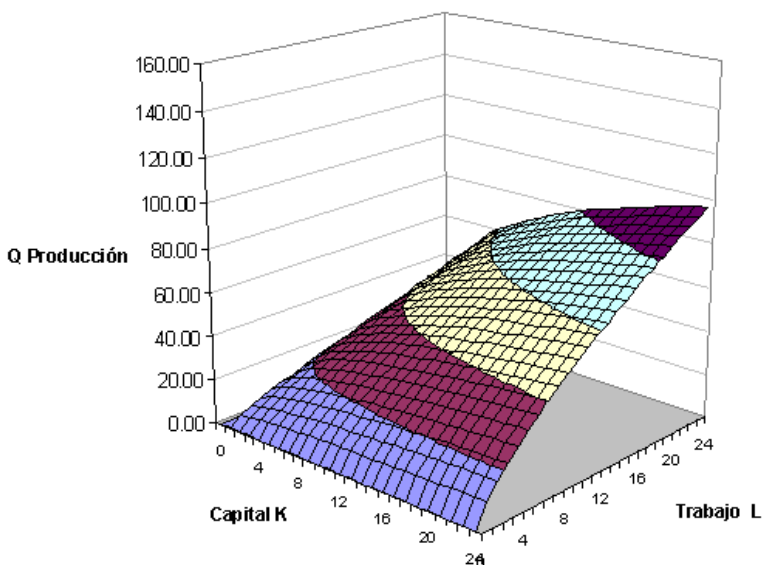
Formalmente la elasticidad de la producción respecto al trabajo se obtiene:

$$\frac{\partial Q}{\partial L} \frac{L}{Q} = \frac{\partial (AL^\alpha K^\beta)}{\partial L} \frac{L}{Q} = \frac{A\alpha L^{\alpha-1} K^\beta L}{AL^\alpha K^\beta} = \alpha$$

Similarmente se obtiene la elasticidad de la producción respecto al capital.

$$\frac{\partial Q}{\partial K} \frac{K}{Q} = \frac{\partial (AL^\alpha K^\beta)}{\partial K} \frac{K}{Q} = \frac{A\beta L^\alpha K^{\beta-1} K}{AL^\alpha K^\beta} = \beta$$

La siguiente gráfica ilustra el comportamiento de la función para $A=4$, $\alpha=0.3$ y $\beta=0.7$. Las curvas que separan las bandas de colores se conocen como isocuantas o curvas de nivel y corresponden a valores constantes de la producción Q que se obtienen con diversas combinaciones de los factores L y K .



El modelo se linealiza fácilmente mediante una transformación logarítmica y así los parámetros se pueden estimar por MCO.

$$Q = AL^\alpha K^\beta u$$

$$\ln(Q) = \ln(A) + \alpha \ln(L) + \beta \ln(K) + \ln(u)$$

$$Q' = A' + \alpha L' + \beta K' + u'$$

Si se consideran rendimientos constantes a escala $\alpha + \beta = 1$ entonces el modelo se reduce a uno lineal simple:

$$\begin{aligned} Q' &= A' + \alpha L' + (1 - \alpha)K' + u' \\ &= K' + A' + \alpha(L' - K') + u' \end{aligned}$$

De donde

$$Q'-K'=A'+\alpha(L'-K')+u'$$

Considere como ejemplo los siguientes datos publicados por D. Salvatore correspondientes a 14 empresas de una industria. Se incluyen los logaritmos de las variables para ajustar por MCO.

Empresa	Horas Hombre	Horas Máquina	Toneladas	Transformaciones logarítmicas		
	L	K	Q	ln(L)	ln(k)	ln(Q)
1	1480	410	240	7.2998	6.0162	5.4806
2	1660	450	400	7.4146	6.1092	5.9915
3	1150	380	110	7.0475	5.9402	4.7005
4	1790	430	530	7.4900	6.0638	6.2729
5	1880	480	590	7.5390	6.1738	6.3801
6	1860	450	470	7.5283	6.1092	6.1527
7	1940	490	450	7.5704	6.1944	6.1092
8	1240	395	160	7.1229	5.9789	5.0752
9	1240	430	290	7.1229	6.0638	5.6699
10	1850	460	490	7.5229	6.1312	6.1944
11	1570	435	350	7.3588	6.0753	5.8579
12	1700	470	550	7.4384	6.1527	6.3099
13	2000	480	560	7.6009	6.1738	6.3279
14	1850	440	430	7.5229	6.0868	6.0638

Los resultados del ajuste por MCO mediante **Casandra** son los siguientes:

Estadísticas de Regresión

Coef. de Correlación Múltiple R ²	0.9370386
Coef. de Determinación R ²	0.8780413
R ² Ajustado	0.8558670
Varianza del Error	0.0365876
Error Estándar de la Estimación	0.1912790
Número de Casos	14

Recuérdese que al utilizar variables transformadas, la relación lineal se expresa en términos de los logaritmos de las variables.

Parámetros estimados	Coeficientes	Error Estándar
A Intercepción	-23.2311819	5.2364620
α lnL	1.4302525	0.5608430
β lnK	3.0453908	1.3672348

Como $\alpha + \beta = 4.4756$ se tienen rendimientos crecientes a escala. Así si L y K se incrementan cada uno en 5%, entonces Q se incrementará en 24.4%

3.5 Variables Ficticias (Dummy).

El modelo lineal general que hasta ahora se ha analizado, supone que las variables independientes corresponden a mediciones en nivel intervalar o de razón. Es frecuente que entre las variables independientes que se desea incluir, presenten niveles de medición en escalas ordinales o categóricas, estas últimas también llamadas cualitativas.

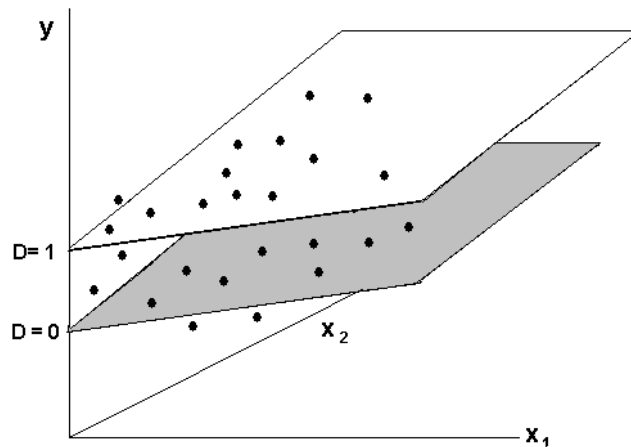
El modelo lineal amplía notablemente sus posibilidades de aplicación al permitir la inclusión de variables categóricas tales como zona geográfica, género de las personas, nivel socioeconómico, estación, período gubernamental, tiempo de paz o guerra, etc.

Un caso de particular interés lo constituyen las variables cualitativas dicotómicas, esto es aquellas que solamente admiten dos categorías, las cuales se pueden asociar a 0 y 1 para identificar ausencia o presencia de la característica de interés. Estas variables se pueden incluir en un modelo en forma de variables ficticias (dummy) identificadas por la letra D.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 D_{4i} + \dots + \beta_k D_i + u_i$$

$$i = 1, 2, \dots, n$$

El efecto que tiene la inclusión de una variable ficticia en el modelo es cambiar el nivel del hiperplano ajustado respecto de la variable dependiente. Así en lugar de ajustar dos diferentes modelos para cada categoría, lo cual produciría sustanciales cambios en los coeficientes de las variables originales, lo único que se altera es la ordenada al origen para cada categoría, pues a la combinación lineal que expresa el modelo se suma cero o el valor del coeficiente de la variable ficticia y los demás coeficientes permanecen sin cambio.



Para cualquier variable con K clasificaciones se requiere la inclusión de k-1 variables ficticias. Por ejemplo, suponga que se tiene un modelo de demanda de un producto que considera tres tipos de consumidores en función de su grupo de edad: Jóvenes, Maduros y Mayores. Las dos variables ficticias a incorporar tomarían los siguientes valores:

Jóvenes

D1= 1 si la observación pertenece a un joven

D1= 0 si la observación no pertenece a un joven

Maduros

D2= 1 si la observación pertenece a un sujeto maduro

D1= 0 si la observación no pertenece a un maduro

Mayores

Los sujetos mayores quedarían definidos implícitamente cuando D1 = 0 y D2=0

La estimación del modelo que incluye una o más variables ficticias se realiza mediante MCO.

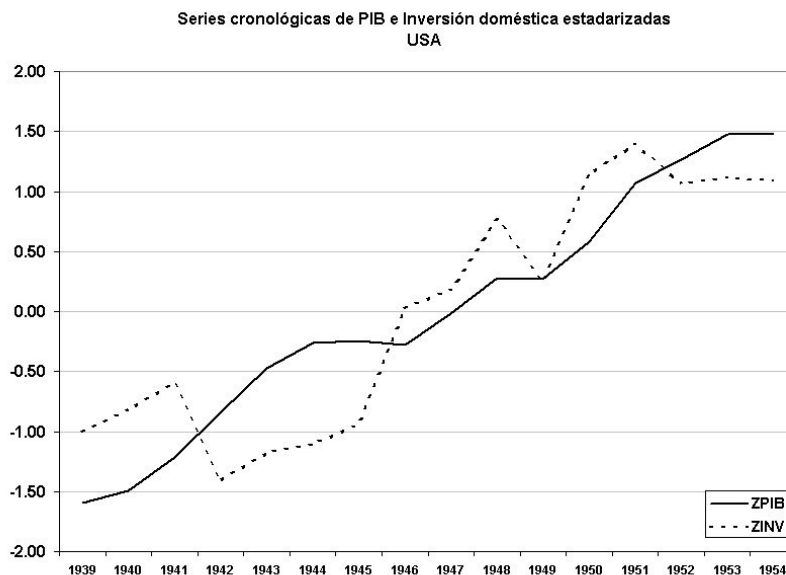
Como ejemplo numérico, considere los siguientes datos correspondientes al PIB y la Inversión interna de los Estados Unidos correspondientes a los años 1939 a 1954. Ambas variables medidas en miles de millones de

dólares. En las dos columnas de la derecha se presentan ambas variables estandarizadas para comparar mejor su comportamiento conjunto en una misma gráfica.

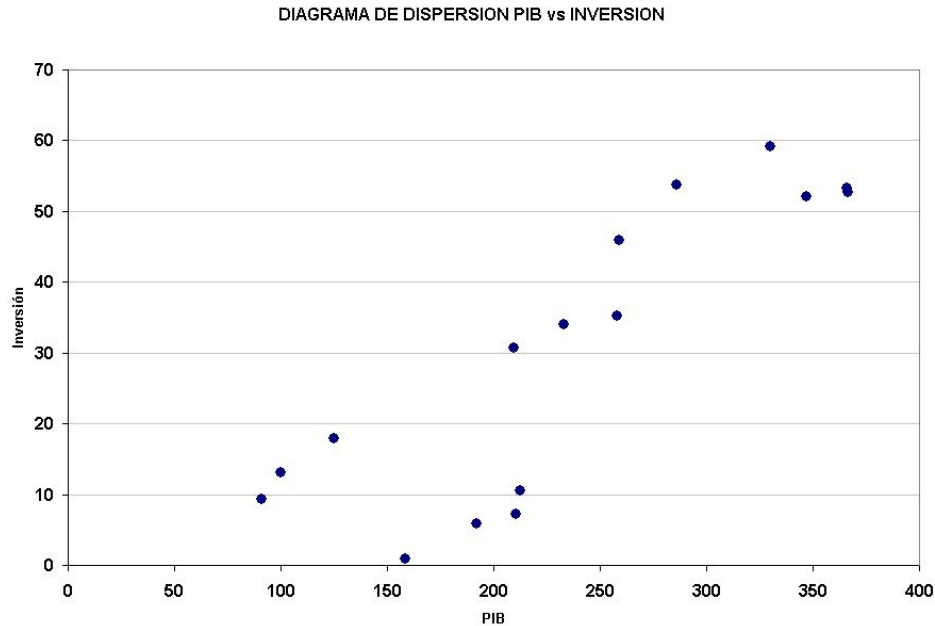
Año	PIB	INVERSION	Variables Estandarizadas	
			ZPIB	ZINV
1939	90.80	9.3	-1.59804815	-1.00351094
1940	100.00	13.1	-1.49538025	-0.82028732
1941	124.90	17.9	-1.21750735	-0.58884696
1942	158.30	0.9	-0.84477825	-1.40853157
1943	192.00	5.8	-0.46870127	-1.17226954
1944	210.50	7.2	-0.26224952	-1.10476610
1945	212.30	10.6	-0.24216232	-0.94082918
1946	209.30	30.7	-0.27564099	0.02832734
1947	232.80	34.0	-0.01339146	0.18744258
1948	259.10	45.9	0.28010481	0.76122181
1949	258.00	35.3	0.26782930	0.25012435
1950	286.20	53.8	0.58252872	1.14213407
1951	330.20	59.2	1.07354911	1.40250448
1952	347.20	52.1	1.26326153	1.06016561
1953	366.10	53.3	1.47417710	1.11802570
1954	366.30	52.7	1.47640901	1.08909566

Miles de millones de dólares

En la siguiente gráfica se presentan las dos series cronológicas de PIB e inversión estandarizadas. Se puede observar la depresión que presenta la inversión y que corresponde a los años de la participación de USA en la Segunda Guerra Mundial 1942-45.



A continuación el diagrama de dispersión del PIB e INVERSION en sus valores originales. Se observan indicios de una relación lineal directa entre PIB e Inversión, pero hay 4 puntos desplazados en la parte inferior de la gráfica y que corresponden al período 1942-45. Así, la guerra tiene una clara influencia.



Se procede a ajustar el modelo lineal simple:

$$Y = \beta_0 + \beta_1 X_1 + u$$

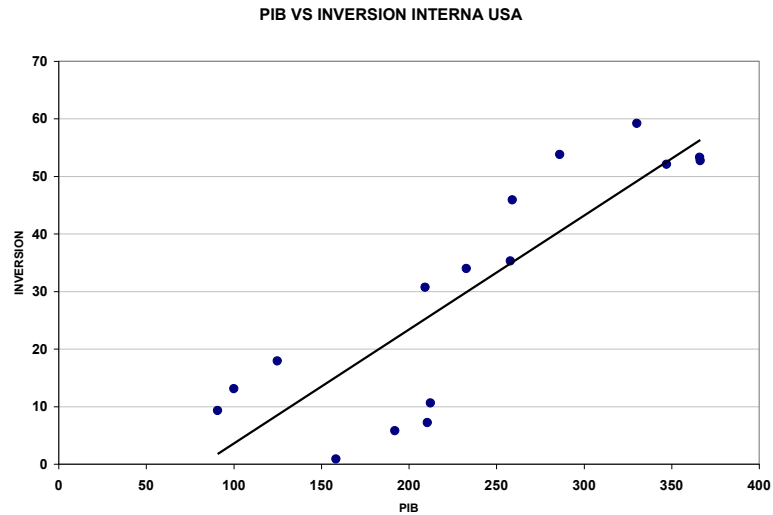
Donde Y Es la inversión interna en millones de USD
 X Es el PIB en miles de millones de USD

Al ajustar el modelo se obtienen los siguientes resultados. Aunque el coeficiente de determinación no corregido por grados de libertad, nos refiere una explicación del 73% de la varianza, y el estadístico T del coeficiente de regresión asociado al PIB lleva a un claro rechazo de la hipótesis de nulidad del coeficiente, la observación de la gráfica de dispersión con el modelo ajustado revela ciertas anomalías.

Coeficiente de correlación múltiple	0.8560258
Coeficiente de determinación R 2	0.7327802
R 2 ajustado	0.7136931
Error estándar de la estimación	11.0973193
Observaciones	16

Variable	Coeficientes	Error Estándar	Estadística T	Probabilidad
Intercepción	-16.248383	7.980083	-2.036117	0.0611167
PIB	0.198123	0.031976	6.196073	0.0000233

La recta ajustada experimenta un efecto de desplazamiento hacia los 4 puntos ya mencionados y un comportamiento anómalo de los residuales, al observar una racha de signo positivo, una de negativo y otra de signo positivo.



La inclusión de una variable indicadora que permita diferenciar los períodos de paz del de guerra se realiza a través de una variable ficticia (dummy) que se incluye en el modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 D_2 + u$$

Donde D_2 adopta el valor 1 para indicar un año dentro del período de guerra y cero para período de paz.

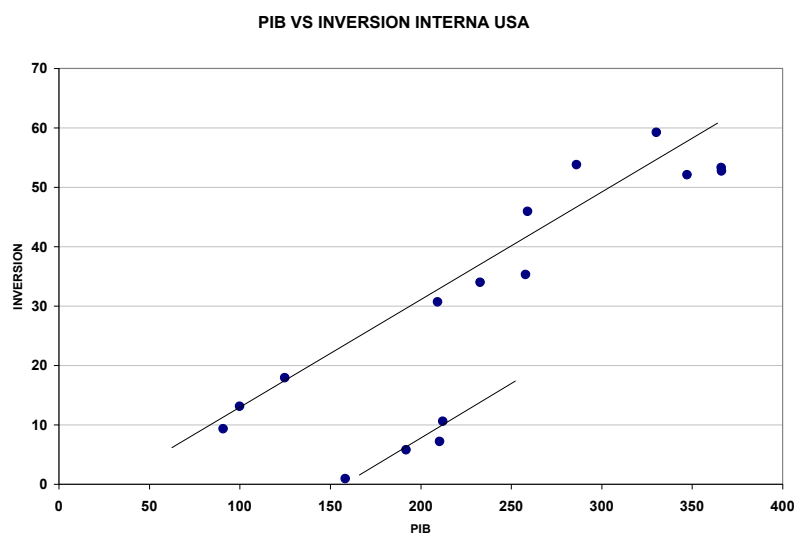
Año	PIB	INVERSION	D2
1939	90.8	9.3	0
1940	100.0	13.1	0
1941	124.9	17.9	0
1942	158.3	0.9	1
1943	192.0	5.8	1
1944	210.5	7.2	1
1945	212.3	10.6	1
1946	209.3	30.7	0
1947	232.8	34.0	0
1948	259.1	45.9	0
1949	258.0	35.3	0
1950	286.2	53.8	0
1951	330.2	59.2	0
1952	347.2	52.1	0
1953	366.1	53.3	0
1954	366.3	52.7	0

Se ajusta el nuevo modelo y se observan notables cambios. El coeficiente de determinación indica que ahora se explica el 95% de la varianza de la inversión. El error estándar de la estimación pasó de 11.09 a 4.5. El error estándar del coeficiente de regresión asociado al PIB pasó de 0.0319 a 0.0135

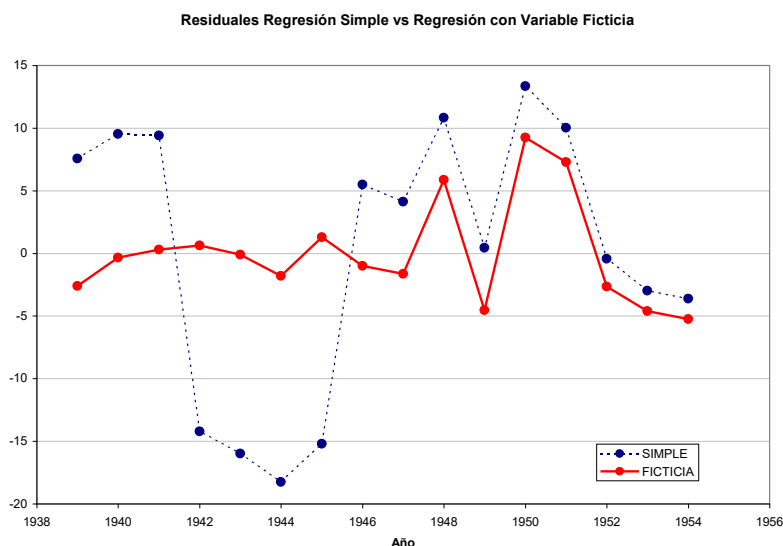
Coeficiente de correlación múltiple	0.97921251
Coeficiente de determinación R2	0.95885713
R2 ajustado	0.95252746
Error estándar de la estimación	4.51880267
Observaciones	16

Variable	Coeficientes	Error típico	Estadística T	Probabilidad
Intercepción	-3.27184406	3.59393271	-0.91037989	0.37919284
PIB	0.16714199	0.01352656	12.35657514	0.00000001
D2	-22.90752345	2.71035143	-8.45186466	0.00000122

El efecto de la variable dicotómica es equivalente a considerar que se tienen dos modelos de regresión simple ajustados. En el primero, tiempos de paz, la ordenada al origen toma el valor -3.27184 y en el segundo, tiempo de guerra, $-3.27184 - 22.90752 = -27.16936$. En ambos el coeficiente del PIB permanece en 0.167141.



En la gráfica siguiente se comparan los comportamientos de los residuales con ambos modelos. Es evidente la disminución de la dispersión y la integración de los residuales del período de guerra a los restantes con el modelo que incluye la variable ficticia.



3.6 Polinomios.

Hasta este punto, se han mencionado transformaciones que involucran a una o más variables independientes, con objeto de linealizar una expresión. Sin embargo también surge la idea de ajustar funciones que involucren una o más potencias de la misma variable independiente o más variables independientes. Puesto que los modelos lineales deben su adjetivo a los parámetros y no a las variables, lo anterior es perfectamente posible. Se sabe que se requieren dos puntos para definir una recta y que es posible ajustar en exactamente un polinomio de grado $n-1$ a través de n puntos. Sin embargo, es prudente ser parsimonioso en lo que respecta al grado del polinomio y adoptar uno de bajo grado.

Considérese un modelo que involucra un polinomio de tercer grado:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 X_{2i}^3 + u_i$$
$$i = 1, 2, \dots, n$$

Se definen las siguientes transformaciones para tener entonces un modelo lineal múltiple en su forma usual.

$$X_3^* = X_2^2$$

$$X_4^* = X_2^3$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}^* + \beta_4 X_{4i}^* + u_i$$
$$i = 1, 2, \dots, n$$

El modelo se ajusta entonces por MCO. El polinomio puede adoptar formas muy diferentes e incluir varias variables definidas con potencias y productos entre ellas que podrían interpretarse como interacciones.

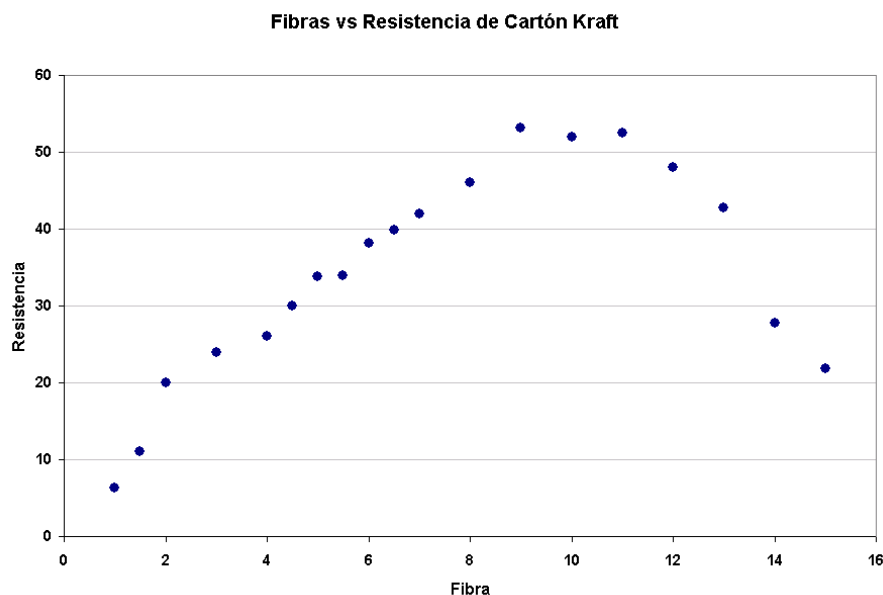
$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 X_{3i}^2 + X_{2i} X_{3i} + u_i$$
$$i = 1, 2, \dots, n$$

Los modelos de polinomios amplían las aplicaciones de los modelos lineales, sin embargo hay que tener cuidado con ellos, sobretodo cuando se trata de hacer predicciones de valores fuera del rango observado debido a la conducta poco previsible que pueden adoptar los polinomios de alto grado.

En la elección del grado del polinomio hay que considerar la significancia de los coeficientes de las potencias de mayor grado y la ganancia marginal en la proporción de la varianza explicada por el modelo que se tiene al incrementar el grado de un polinomio. Esto último se puede medir con la comparación de los coeficientes de determinación R^2 de los modelos ajustados en forma consecutiva con los grados $K-1$ y K .

Ejemplo, considere los siguientes datos relativos a la concentración de fibra en la pulpa para elaborar cartón kraft medida en porcentaje y la resistencia a la tensión en el cartón fabricado. Los datos presentan un comportamiento que se aleja sensiblemente de un modelo lineal simple.

Fibra X	Resistencia Y
1.0	6.3
1.5	11.1
2.0	20.0
3.0	24.0
4.0	26.1
4.5	30.0
5.0	33.8
5.5	34.0
6.0	38.1
6.5	39.9
7.0	42.0
8.0	46.1
9.0	53.1
10.0	52.0
11.0	52.5
12.0	48.0
13.0	42.8
14.0	27.8
15.0	21.9



Conscientes de que no es la mejor opción, se procede a ajustar un modelo lineal simple, con objeto de comparar la ganancia que se tiene al ajustar polinomios de grado superior. El modelo ajustado toma la siguiente forma:

$$\hat{Y} = 21.321262 + 1.770986X$$

Las estadísticas básicas del ajuste revelan un modelo cuya $R^2 = 0.3053$ indica que solamente explica el 30.5% de la varianza total de la resistencia, la R^2 ajustada por grados de libertad cuyo valor es 0.26451 resulta sensiblemente menor. El error estándar de la estimación 11.81 es también muy grande si se lo compara con las observaciones de la variable dependiente:

Estadísticas	
Coefficiente de correlación múltiple	0.5526064
Coefficiente de determinación R^2	0.3053739
R^2 ajustado	0.2645135
Error estándar de la estimación	11.8158862
Casos	19

A continuación se procede a ajustar un polinomio de segundo grado. Hay evidente mejoría del modelo que se nota con la $R^2 = 0.9085020$ y la R^2 ajustada 0.897064. El error estándar de la estimación también mejora, pues descendió de 11.81 a 4.420.

Estadísticas	
Coefficiente de correlación múltiple	0.9531537
Coefficiente de determinación R^2	0.9085020
R^2 ajustado	0.8970647
Error estándar de la estimación	4.4203951
Casos	19

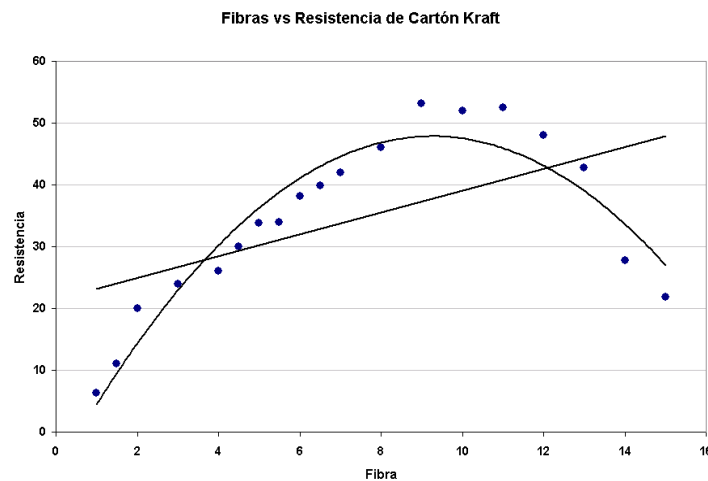
El polinomio ajustado adquiere los siguientes coeficientes:

$$\hat{Y} = -6.674192 + 11.764006X - 0.634549X^2$$

En la siguiente tabla se presentan los coeficientes acompañados de sus errores estándares, estadísticas T y probabilidades asociadas. Las pequeñas probabilidades asociadas a las estadísticas T de los coeficientes de X y de X^2 indican que éstos son claramente distintos de cero.

Variables	Coeficientes	Error típico	Estadístico T	Probabilidad
Intercepción	-6.674192	3.399708	-1.963166	0.067252028
X	11.764006	1.002782	11.731366	0.000000003
X2	-0.634549	0.061788	-10.269727	0.000000019

En la siguiente gráfica se pueden apreciar los dos modelos ajustados hasta este punto. Evidentemente polinomio de segundo grado responde mejor al comportamiento de los datos observados.



Al observar la gráfica de las sucesiones residuales ordenadas según orden de las observaciones para el modelo basado en la recta, como para el modelo del polinomio de segundo grado, se nota el comportamiento anómalo de los correspondientes a la recta (línea punteada) con una racha larga de residuales positivos, en tanto que la secuencia de residuales del polinomio de segundo grado tiene un comportamiento más errático.

