

# SUPUESTOS DEL MODELO CLÁSICO DE REGRESIÓN LINEAL Y DE LOS MODELOS LINEALES GENERALIZADOS

Isadore Nabi

<b>I. EL MODELO CLÁSICO DE REGRESIÓN LINEAL</b>	<b>2</b>
<b>I.I. Breve Reseña Histórica</b>	<b>2</b>
<b>I.II. Fundamento Teórico</b>	<b>3</b>
<b>I.III. Supuestos</b>	<b>8</b>
I.III. I. <i>Linealidad en las Variables</i>	8
I.III. II. <i>Linealidad en los Parámetros</i>	8
I.III. III. <i>Valores fijos de X, o valores de X independientes del término de error</i>	9
I.III. IV. <i>El valor medio de la perturbación estocástica es nulo</i>	9
I.III. V. <i>La varianza del término de perturbación estocástica es nula (homocedasticidad)</i>	9
I.III. VI. <i>No existe correlación serial (autocorrelación) entre las perturbaciones</i>	13
I.III. VII. <i>El número de observaciones n-ésimas debe ser mayor que el número de parámetros n-k a estimar</i>	15
I.III. VIII. <i>Independencia lineal entre las variables explicativas</i>	19
<b>II. LOS MODELOS LINEALES GENERALIZADOS (MLG)</b>	<b>23</b>
<b>II.III. Supuestos del MLG</b>	<b>23</b>
<b>Referencias</b>	<b>25</b>

## I. EL MODELO CLÁSICO DE REGRESIÓN LINEAL

### I.I. Breve Reseña Histórica

Como se señala en (Gujarati & Porter, 2010, pág. 15), fue Francis Galton quien acuñó el término “regresión”. En “Family Likeness in Stature”, Proceedings of Royal Society, Londres, vol. 40, 1886, pp. 42-72”. Ahí planteó que, a pesar de la tendencia de los padres de estatura alta a procrear hijos altos y los padres de estatura baja, hijos bajos, la estatura promedio de los niños de padres de una estatura determinada tendía a desplazarse, o “regresar”, a la estatura promedio de la población total. En otras palabras, la estatura de los hijos de padres inusualmente altos o inusualmente bajos tiende a dirigirse a la estatura promedio de la población. Esta ley de regresión universal de Galton fue confirmada por su discípulo y amigo Karl Pearson (junto con A. Lee) en “On the Laws of Inheritance”, Biometrika, vol. 2, noviembre de 1903, pp. 357-462. Ahí, se reúnen más de mil registros de estaturas de miembros de grupos familiares. Pearson descubrió que la estatura promedio de los hijos de un grupo de padres de estatura alta era menor que la estatura de sus padres, y que la estatura promedio de los hijos de un grupo de padres de estatura baja era mayor que la estatura de sus padres; es decir, se trata de un fenómeno mediante el cual los hijos altos e hijos bajos “regresan” por igual a la estatura promedio de todos los demás. En palabras de Galton, se trata de una “regresión a la mediocridad”. La definición moderna de regresión consiste en el “(...) estudio de la dependencia de una variable (variable dependiente) respecto de una o más variables (variables explicativas) con el objetivo de estimar o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las segundas.” (Gujarati & Porter, 2010, pág. 15).

Como metodología empírica, la regresión nace en la obra *Natural Inheritance* (1889) de Francis Galton, mentor de Karl Pearson, quien posteriormente se erigiría como padre de la Estadística Matemática. La obra y el pensamiento de Galton fueron la base sobre la que Pearson (quien cambió “Carl” por “Karl” en honor a Karl Marx) construyó el edificio teórico que completaría su histórico enemigo académico: ni más ni menos que Ronald Fisher, padre de la Bioestadística (santo de los bioestadísticos y psicometristas) y quien se dedicó a criticar en vida todo trabajo de Pearson que fuese posible criticar (casi tanto como criticó toda investigación bioestadística que afirmase que fumar y el cáncer estaban significativamente correlacionados, y de su crítica acuñó la expresión “correlación no implica causalidad”) desde que este se negase a publicar en su revista un artículo de Ronald Fisher aludiendo que no estaba lo suficientemente capacitado para verificar

como gnoseológicamente válido lo planteado en la investigación en cuestión, ganando Fisher todos y cada uno de los enfrentamientos intelectuales que tuvo con Karl Pearson; hay que decir en favor de Fisher que no termina de resultar verosímil (como diría el hijo de Karl más tarde) que Karl Pearson no tuviese la capacidad para fungir de juez. Sería también Ronald Fisher quien fundaría el proceso de prueba de hipótesis (aunque no de la forma en que es conocida hoy, la metodología de Fisher era diferente y no es fácilmente automatizable y reproducible), cuyo marco teórico completaría (para infortunio de Fisher) Jerzey Neyman y Egon Pearson (hijo de Karl Pearson), sobre este último Fisher le “heredaría” el desagrado (académico y personal, factores seguramente linealmente dependientes) que sentía por su padre (quien falleció veintiséis años antes que Fisher). Finalmente, es importante mencionar el hecho que el proceso de prueba de hipótesis actualmente utilizado (conocido como *New Hypothesis Statistical Testing*) es una amalgama incompatible entre las metodologías de Ronald Fisher y de Neyman-Pearson (Egon), como puede verificarse en (Perezgonzalez, 2015, pág. 10).

## **I.II. Fundamento Teórico**

El modelo de regresión lineal clásico como metodología estadística está compuesto fundamentalmente dos técnicas, los cuales son la técnica de regresión lineal y la técnica de mínimos cuadrados ordinarios (que es una forma de realizar ajuste de curvas). De ello se desprende que el método de mínimos cuadrados ordinarios y el modelo de regresión lineal no son sinónimos. Como señala (Bhuptani, 2020), hay que comenzar entonces por resaltar la diferencia existente entre *regresión lineal* y *ajuste de curvas*. Tener un conjunto de puntos y desear dibujar una curva (línea) a través de ellos que se ajuste lo mejor posible a los mismos es un problema puramente geométrico, es decir, los ejes  $x$  e  $y$  no tienen interpretación, puesto que lo que en otro contexto serían *datos*, en este son meramente puntos en el espacio cartesiano. Por su parte, la regresión lineal es una inferencia estadística sobre un problema concreto de la realidad. Los valores de  $y$  (*i.e.*, el conjunto de datos) se interpretan según el contexto analítico en que se encuentre el investigador<sup>1</sup> y con ello se transforman en *información sobre la variable de interés*, ello tras su estudio mediante modelos estadísticos. Por otra parte, los valores de  $x$  se transforman en *información adicional* que se tiene sobre cada elemento de  $y$  que podría ser útil para realizar estimaciones sobre su comportamiento.

Cuando se hace una regresión lineal, se está tratando de construir un modelo probabilístico que describa la variable  $y$  teniendo en cuenta a la variable  $x$ , sin embargo, existen múltiples formas de realizar esto. Un modelo lineal supone que  $y$  tiene una media diferente para cada valor posible de  $x$ . Así, el conjunto de estos

---

<sup>1</sup> Con todas las implicaciones que esto posee.

valores medios sigue una línea recta con una cierta intersección y una cierta pendiente. Como con cualquier problema de inferencia estadística, se estiman los parámetros desconocidos utilizando la estimación de máxima verosimilitud. Sin embargo, como en este caso los parámetros desconocidos son una intersección y una pendiente, el resultado final de la estimación de máxima verosimilitud es básicamente que se está eligiendo una línea recta que se ajuste mejor a los datos observados, por lo que es así como convergen la regresión lineal con el ajuste de curvas.

Una vez planteado lo anterior, como se señala en la fuente citada, podemos pensar en la regresión lineal como una metodología que utiliza la herramienta del ajuste de curvas (en el caso de la regresión lineal simple, específicamente de una línea recta –o de un hiperplano, si es una regresión lineal múltiple–) mediante un conjunto de puntos que llamamos cualitativamente datos. Pero hay muchas estrategias posibles para ajustar una línea a través de un conjunto de puntos. Entre estas se encuentran:

- a) Tomar el punto más a la izquierda y el más a la derecha y dibujar una línea entre ellos.
- b) Calcular las pendientes de las líneas que conectan cada par de puntos y calcular la pendiente promedio, dibujando una línea con esta pendiente que pase por el punto en el promedio de los valores de  $x$  y el promedio de los valores de  $y$ .
- c) Se puede encontrar la línea para la cual hay un número igual de puntos sobre la línea y debajo de la línea.
- d) Es posible dibujar una línea y luego, para cada uno de los puntos de datos, medir la distancia vertical entre el punto y la línea y sumarlos; la línea ajustada sería aquella donde esta suma de distancias es lo más pequeña posible.
- e) También se puede dibujar una línea y luego, para cada uno de los puntos de datos, medir la distancia vertical entre el punto y la línea, elevarlos al cuadrado y sumarlos; la línea ajustada sería aquella donde esta suma de distancias es lo más pequeña posible.

La última estrategia se llama *mínimos cuadrados ordinarios*, de ahora en adelante *MCO*, y su nombre proviene del hecho que se está buscando minimizar la suma de los errores de predicción al cuadrado. Sin embargo, a pesar de que los *MCO* son la técnica más popular que emplea la metodología del Análisis de Regresión, en lo que respecta al ajuste de una línea a través de un conjunto de puntos, cualquiera de las otras estrategias es igualmente válida. Las tres primeras estrategias las inventó el autor de la investigación citada a manera de ejemplo y probablemente no funcionen adecuadamente; sin embargo, la cuarta es una estrategia real llamada

*desviaciones medias absolutas* y es preferida por algunas personas por sobre mínimos cuadrados.

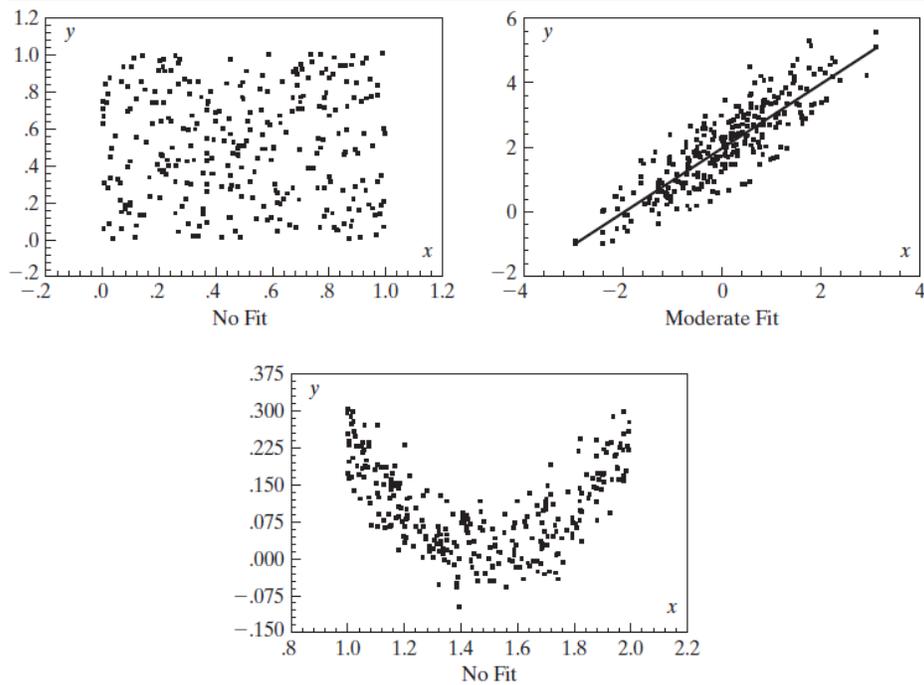
Cabe preguntarse por qué si no es la única técnica entonces es la más utilizada. La razón es que, al resolver el problema de regresión lineal estadística, una suposición de modelado muy común es que por cada valor posible de  $x$ , la cantidad  $y$  se distribuye normalmente con una media que es lineal en  $x$ . Por lo tanto, la función de verosimilitud es esencialmente un producto de funciones de densidad de probabilidad de la distribución normal. Así mismo, el autor señala que se estiman los parámetros desconocidos ( $y$ , por lo tanto, se encuentra la recta de mejor ajuste al conjunto de observaciones) maximizando la función de verosimilitud. Si se observa cómo es el producto de funciones de densidad de probabilidad normales, el lector notará que maximizar esta expresión es equivalente a minimizar la suma de los errores al cuadrado. Es decir, la línea que se obtiene realizando el ajuste de la curva a través de mínimos cuadrados es equivalente a la línea que obtiene realizando una regresión lineal utilizando un modelo distribuido normalmente.

De esta forma, puede observarse que el análisis de regresión es una metodología, mientras que los mínimos cuadrados con una técnica empleada por esta metodología y la regresión como tal (la ecuación de regresión) es también una técnica empleada por el análisis de regresión. Tampoco debe identificarse una regresión con la técnica de mínimos cuadrados lineales (puesto que existen no-lineales), debido a que existen distintos tipos de análisis de regresión y no todos ellos son lineales. Sin embargo, los mínimos cuadrados ordinarios son una de las técnicas posibles en la regresión lineal para encontrar la línea recta de mejor ajuste al conjunto de datos del que se dispone. Véase las figuras presentadas a continuación.

### Figura 1

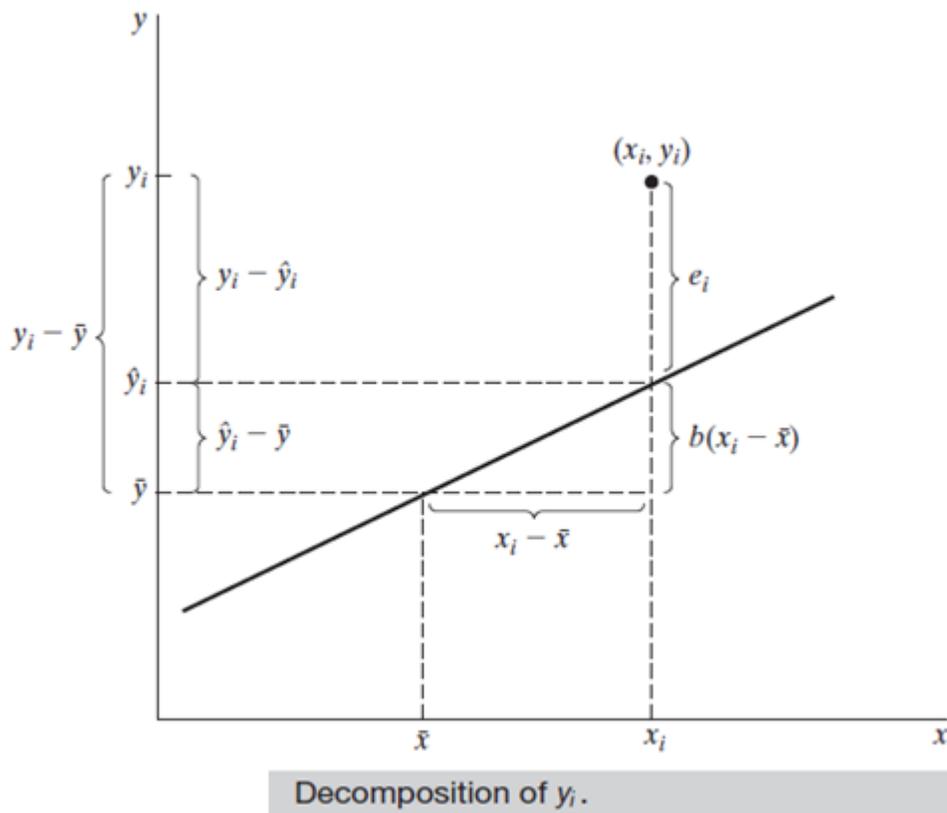
The original fitting criterion, the sum of squared residuals, suggests a measure of the fit of the regression line to the data. However, as can easily be verified, the sum of squared residuals can be scaled arbitrarily just by multiplying all the values of  $y$  by the desired scale factor. Since the fitted values of the regression are based on the values of  $x$ , we might ask instead whether *variation* in  $x$  is a good predictor of *variation* in  $y$ . Figure 3.3 shows three possible cases for a simple linear regression model. The measure of fit described here embodies both the fitting criterion and the covariation of  $y$  and  $x$ .

**FIGURE 3.3** Sample Data.



Fuente: (Greene, 2012, pág. 79).

Figura 2



Fuente: (Greene, 2012, pág. 80).

Además, en el contexto de ciencias de datos existen otras técnicas para entrenar un modelo lineal que no usa mínimos cuadrados lineales, como se señala en (StackExchange Cross Validated, 2017).

En síntesis, los mínimos cuadrados son una técnica en el análisis de regresión para aproximar la solución de sistemas sobre determinados (*i.e.*, conjuntos de ecuaciones en las que hay más ecuaciones que incógnitas) minimizando la suma de los cuadrados de los residuos hechos en los resultados de cada ecuación. Por su parte, los mínimos cuadrados lineales es una técnica de aproximación de funciones lineales a conjuntos de datos por la técnica de mínimos cuadrados. Es un conjunto de formulaciones para resolver problemas estadísticos relacionados con la regresión lineal, incluidas las variantes para residuos ordinarios (no ponderados), ponderados y generalizados (correlacionados). Los métodos numéricos para mínimos cuadrados lineales incluyen la inversión de la matriz de las ecuaciones normales y los métodos

de descomposición ortogonal. Por su parte, los mínimos cuadrados ordinarios son una de las técnicas más comunes (más no las únicas) dentro de la familia de técnicas conocida como “Mínimos Cuadrados Lineales”, que a su vez pertenece a la familia de técnicas conocida como “Mínimos Cuadrados”. Los mínimos cuadrados ordinarios son utilizados en el contexto del modelo clásico de regresión lineal para estimar sus parámetros desconocidos.

En línea con (McCullagh & Nelder, 1989, pág. 4), deben considerarse a las diferentes teorías estadísticas como descripciones de determinados patrones que es posible identificar que siguen los números en la vida real, patrones los cuales en alguna medida pueden sustituir al conjunto de datos en sí mismos (puesto que estos patrones numéricos describen patrones geométricos, es decir, relativo a las formas que adoptan los fenómenos naturales y/o sociales) a través de determinados valores numéricos concretos en los que se cristalizan dichos patrones, conocidos como parámetros o estadísticos de prueba, según se esté en un contexto poblacional o muestral, respectivamente. Es por ello que según las características empíricas del conjunto de datos en concreto que se estudie (obtenido de la medición de fenómenos naturales y/o sociales) los parámetros  $\beta$  generados por tal conjunto tomarán diferentes valores y precisamente de este hecho empírico es que se formulan las teorizaciones estadísticas-matemáticas que actualmente se conocen como *familias de distribuciones de probabilidad*.

El modelo básico de regresión lineal  $y = \alpha + \beta x$  conecta dos variables  $x$  e  $y$  vía el par de parámetros  $(\alpha, \beta)$  y define una relación entre ambas que describe geoméricamente una línea recta.

### **I.III. Supuestos**

#### ***I.III. I. Linealidad en las Variables***

Como señalan (Gujarati & Porter, 2010, pág. 38), consiste en que la esperanza condicional de la variable de respuesta es una función lineal de las variables regresoras. Una función lineal es aquella expresión algebraica en la que la variable constituye la base de la expresión y su potencia es igual a la unidad.

Geoméricamente hablando, la curva de regresión en este caso es una recta.

#### ***I.III. II. Linealidad en los Parámetros***

Como señalan (Gujarati & Porter, 2010, pág. 38), esto consiste en que la esperanza condicional de la variable de respuesta es una función de los parámetros (lineales), calculados estos últimos mediante los mínimos cuadrados ordinarios (o alguna técnica de ajuste de curvas apropiada para tales fines).

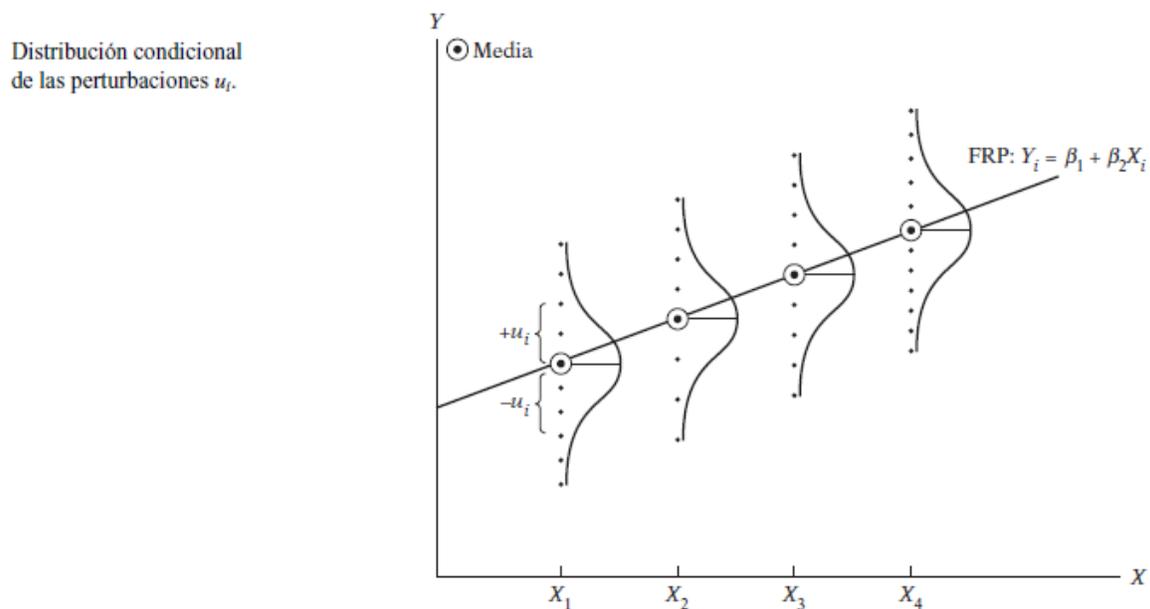
### ***I.III. III. Valores fijos de X, o valores de X independientes del término de error***

Como señalan (Gujarati & Porter, 2010, pág. 62), los valores que toma las variables regresoras pueden considerarse fijos en muestras repetidas (el caso de la regresora fija), o haber sido muestreados junto con la variable dependiente o de respuesta (el caso de la regresora estocástica). En el segundo caso se supone que las variables independientes y el término de error son independientes o, lo que es lo mismo, que su covarianza es nula.

### ***I.III. IV. El valor medio de la perturbación estocástica es nulo***

Como señalan (Gujarati & Porter, 2010, pág. 63), dados los valores de las variables regresoras, la esperanza matemática o valor esperado del término de perturbación estocástica es igual a cero.

Figura 3



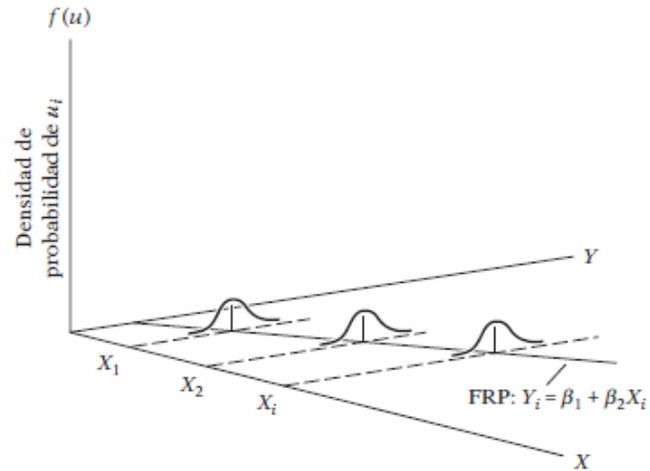
Fuente: (Gujarati & Porter, 2010, pág. 63).

### ***I.III. V. La varianza del término de perturbación estocástica es nula (homocedasticidad)***

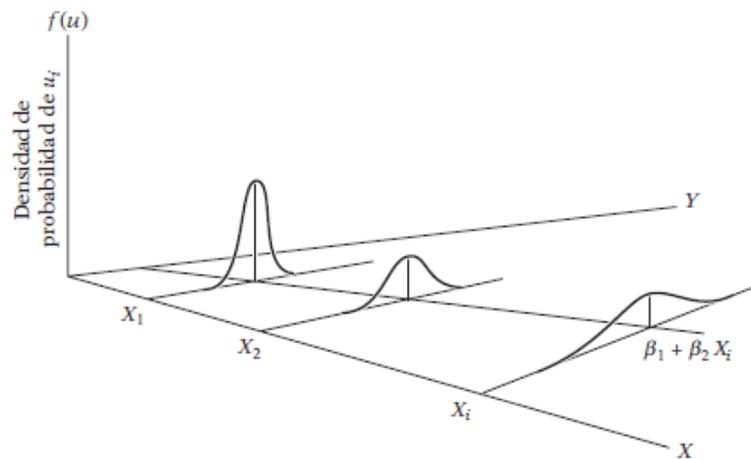
Como señala (Gujarati & Porter, 2010, págs. 365-371), la varianza del término de error es la misma sin importar los valores de las variables regresoras.

Figura 4

Homoscedasticidad.



Heteroscedasticidad.

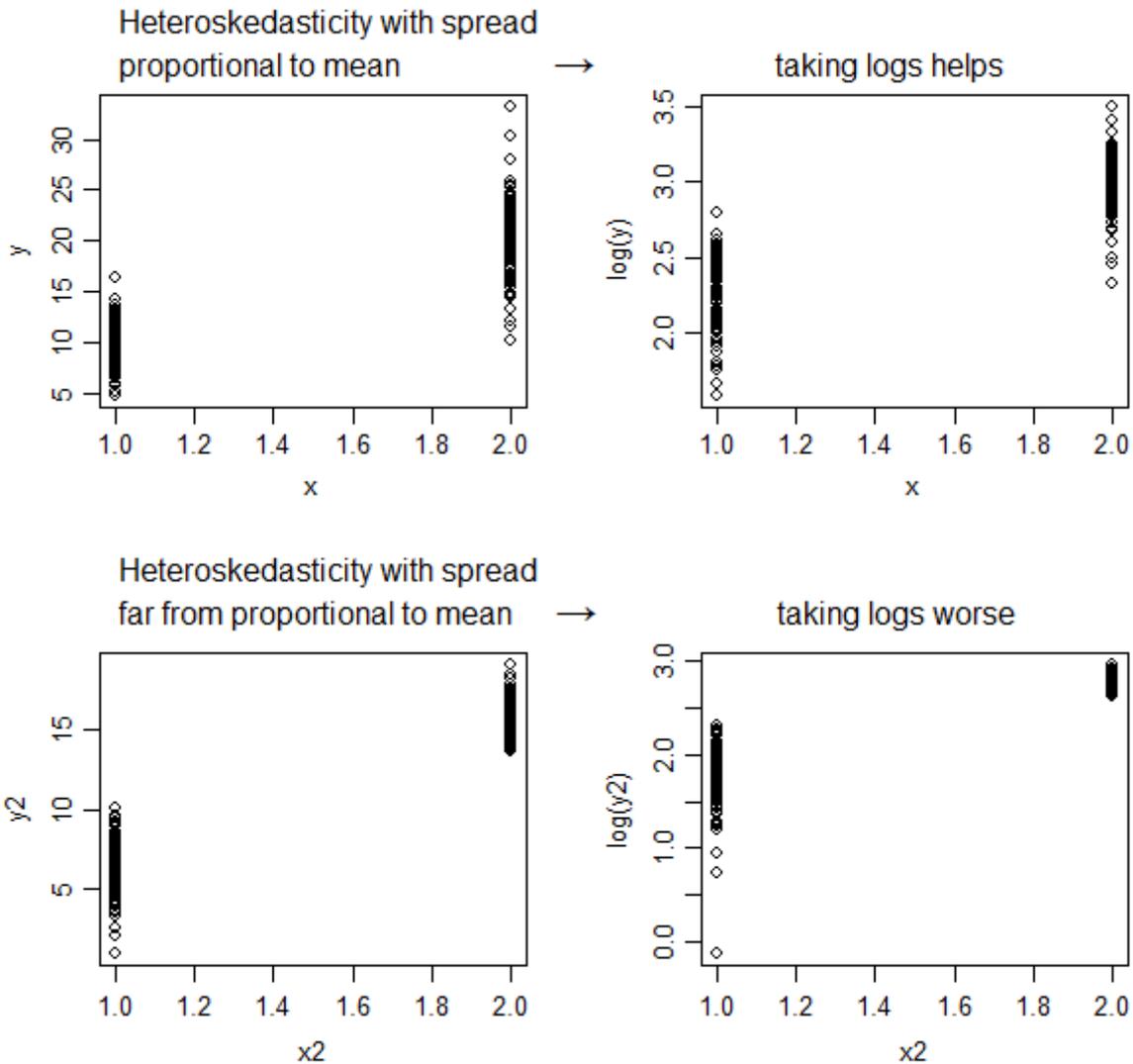


Fuente: (Gujarati & Porter, 2010, pág. 65).

La naturaleza de la heterocedasticidad (ausencia de homocedasticidad) obedece a una o más de las siguientes causas:

1. Con base en los modelos de aprendizaje de los errores, a medida que la gente aprende disminuyen sus errores en el tiempo, por ejemplo, la relación entre horas de práctica de mecanografía y los errores mecanográficos es heterocedástica.

2. Cambios naturales en determinada clase de patrón. Por ejemplo, a medida que aumentan los ingresos, las personas poseen más opciones de compra, por lo que en función de las opciones de compra que les proporcione cada nivel de ingreso (y dado una estructura psíquica de gustos y preferencias, generada por su base genética y su proceso de socialización) pueden variar (o no) su nivel de ahorro.
3. A medida que mejoran las técnicas de recolección de datos, es probable que la varianza se reduzca. Así, es probable que los bancos con equipos complejos de procesamiento de información cometan menos errores en los informes mensuales o trimestrales de sus clientes que bancos que no los posean.
4. La existencia de datos atípicos, los cuales son observaciones que son muy grandes o muy pequeñas en comparación con la abrumadora generalidad de observaciones. Gráficamente esto se expresa como un aislamiento de estas pocas observaciones del resto del conjunto de datos. De forma más precisa, un dato atípico es una observación dentro de la muestra que proviene (estadísticamente hablando) de una población distinta a la que genera las demás observaciones de la muestra.
5. Problemas de especificación en el modelo de regresión.
6. Asimetría en la distribución de una o más regresoras incluidas en el modelo, por ejemplo, al analizar la distribución del ingreso (debido al proceso de acumulación de capital, que lo centralizada progresivamente en cada vez menos individuos).
7. Incorrecta transformación de los datos. Esto ocurre porque la heterocedasticidad es, geoméricamente hablando (que no debe confundirse con su representación gráfica), una dispersión irregular de los datos y, por ello, algunas transformaciones sobre las variables pueden corregirla (al menos parcialmente); sin embargo, también puede empeorarla. Como ejemplo tómese la transformación logarítmica que, como se señala en (Cross Validated, 2018), siempre que la dispersión del error sea aproximadamente proporcional al valor esperado condicional de la variable de respuesta, la heterocedasticidad tenderá a mejorar, sin embargo, si lo anterior no se cumple, entonces la empeorará; geoméricamente hablando, esto ocurre porque la transformación logarítmica redistribuye al conjunto de datos de tal forma que hace que los valores más extremos (los valores más altos) tiendan a situarse más a la derecha del primer cuadrante del plano, mientras que hace que los valores más bajos tiendan a situarse más a la izquierda de dicho cuadrante.



8. Forma funcional incorrecta. Esto ocurre, por ejemplo, cuando se quiere modelar un fenómeno no-lineal con un modelo lineal o viceversa.

El problema de la heterocedasticidad en el contexto de las estimaciones está relacionado con las características deseables de un parámetro o estimador.

Específicamente, como señala (Gujarati & Porter, 2010, págs. 71-73), un estimador debe poseer, en el contexto de la regresión lineal, las siguientes propiedades:

- Es lineal, es decir, es una función lineal de la variable aleatoria.
- Es insesgado, es decir, su valor promedio o esperado es igual al valor verdadero o poblacional.
- Tiene varianza mínima dentro de la clase de todos los estimadores lineales insesgados, es decir, es un estimador eficiente.

Cuando un estimador posee estas tres propiedades se conoce como MELI: Mejor Estimador Lineal Insesgado. Cuando no existe heterocedasticidad, no puede

garantizarse que los estimadores sean de varianza mínima, es decir, no puede garantizarse que sean “los mejores estimadores”, aunque sí seguirán siendo lineales e insesgados. De ello, se desprende naturalmente que las predicciones realizadas con estimadores que no sean de varianza mínima deberán ser tomadas con ciertas reservas, puesto que la variabilidad será elevada en relación al escenario MELI. Lo mismo sucederá, por consiguiente, cuando los modelos no se usen para pronóstico sino para control de políticas de alguna índole (gubernamentales, empresariales, etc.).

### *I.III. VI. No existe correlación serial (autocorrelación) entre las perturbaciones*

Como se señalan en (Gujarati & Porter, 2010, pág. 66), dados dos conjuntos de cualesquiera diferentes valores de las variables independientes, la correlación serial entre dos miembros cualesquiera del término de error es nula o, dicho de otra forma, su covarianza es igual a cero.

Como señalan (Gujarati & Porter, 2010, págs. 414-418), la naturaleza de la autocorrelación es también multidimensional y sus causas por consiguiente son también diversas:

1. Movimiento conjunto de las variables en una determinada dirección, a causa de su misma naturaleza. Esto ocurre, por ejemplo, el caso del movimiento de las variables económicas tras llegar al punto más álgido de la recesión.
2. Sesgo de especificación: caso de variables excluidas. En este escenario, la existencia de autocorrelación es una consecuencia de tener “la pintura incompleta”, puesto que cuando determinadas variables son introducidas en el modelo, la “autocorrelación” se corrige, puesto que en realidad es una falsa autocorrelación. Es conveniente indicar que, si se encuentra que el problema real es de sesgo de especificación y no de autocorrelación, los estimadores MCO pueden ser sesgados e inconsistentes.
3. Sesgo de especificación: forma funcional incorrecta.
4. Fenómeno de la telaraña. Por ejemplo, en el contexto del estudio de los sistemas económicos, la oferta de muchos productos agrícolas reacciona a variaciones en el precio con un rezago de un período debido a que la instrumentación de las decisiones de oferta tarda algún tiempo (el período de gestación de la política de la empresa o agente económico del que se trate).
5. Rezagos. Esto expresa el condicionamiento histórico de las variables, es decir, variables cuyo comportamiento presente es afectado por el comportamiento que tuvo en el pasado, lo que refleja a su vez la continuidad de los fenómenos y la historicidad de los procesos en general.

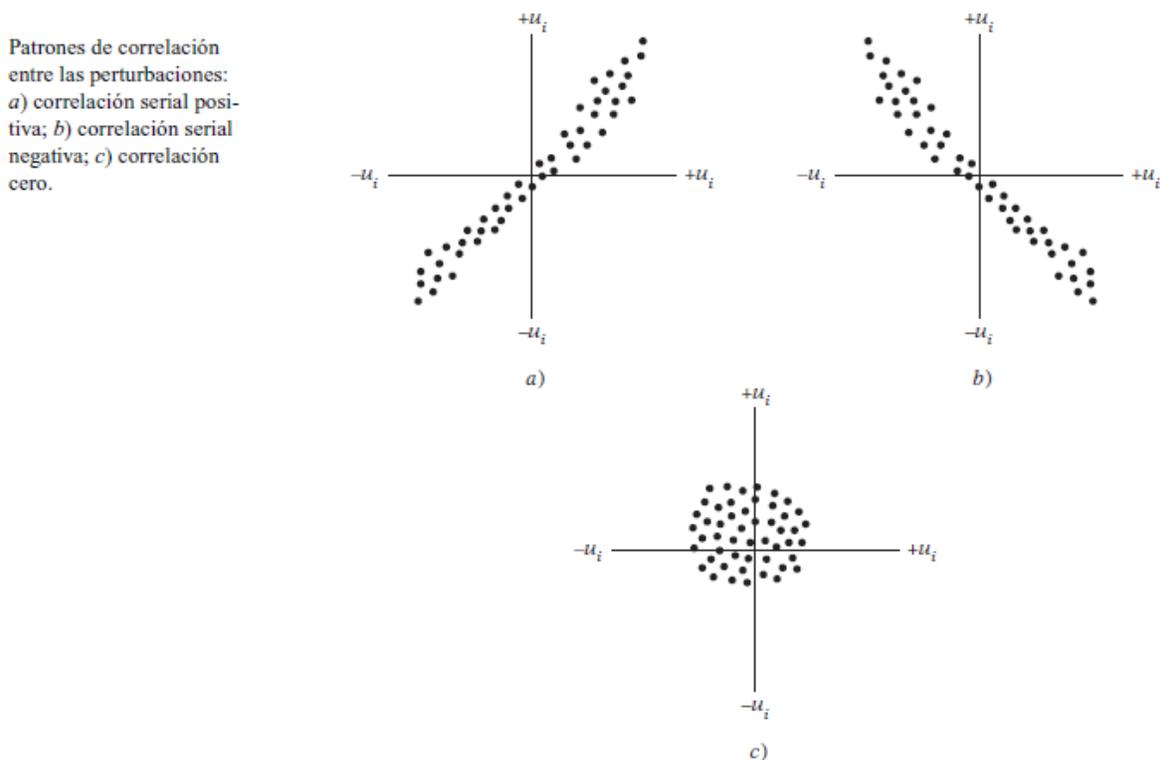
6. Transformaciones simples de datos. En el análisis empírico con frecuencia se hacen este tipo de transformaciones, por ejemplo, en las regresiones de series de tiempo con datos trimestrales. Ahí, por lo general, estos datos provienen de datos mensuales a los que agregan simplemente las observaciones de tres meses y se divide la suma entre 3. Este procedimiento de promediar las cifras suaviza en cierto grado los datos al eliminar las fluctuaciones en los datos mensuales. Por consiguiente, la gráfica referente a datos trimestrales aparece mucho más suave que la que contiene los datos mensuales, y este suavizamiento puede, por sí mismo, inducir un patrón sistemático de perturbaciones, lo que agrega autocorrelación, puesto que el modelo de regresión lineal clásico (MRLC) supone que no existe un patrón distributivo en los miembros del término de error (su distribución es estocástica, producto del azar).
7. Otro tipo de transformaciones sobre los datos. Esto ocurre, por ejemplo, cuando se estudia un conjunto de variables en términos de sus diferencias.
8. No estacionariedad. Informalmente, una serie de tiempo es estacionaria si sus características estadísticas (por ejemplo, la media, varianza y covarianza) son invariantes respecto al tiempo, es decir, no cambian a medida el tiempo transcurre. Cuando ocurre lo contrario entonces la serie de tiempo es no-estacionaria. Cuando las variables regresoras y de respuesta son no-estacionarias, es muy probable que el error también sea no-estacionario y es precisamente en ese escenario en donde el término de perturbación presenta autocorrelación. Por supuesto, existen escenarios en que las variables involucradas son no-estacionarias y que el término de perturbación es estacionario, pero no es la tendencia general de las investigaciones empíricas, según lo reportado en el lugar referido por Gujarati y Porter.

¿Qué sucede entonces con los estimadores MCO y sus varianzas en presencia de autocorrelación? Los estimadores siguen siendo consistentes, lineales e insesgados, sin embargo, dejan de ser eficientes (de varianza mínima). Así, a pesar de que la muestra se incrementa indefinidamente es probable que no cambie el hecho de que los estimadores no son eficientes, es decir, que su ineficiencia en muestras finitas se traslada a su ineficiencia asintótica, aún cuando muchas veces en muestras infinitas se corrigen problemas que aparecen para esas mismas variables en muestras finitas. Estimar MCO ignorando la presencia de autocorrelación puede ocasionar los siguientes problemas:

- a. Es probable que la varianza estimada de los residuos subestime a la verdadera varianza, lo que implica hacer pronósticos asumiendo un nivel de error inferior al error real de predicción.

- b. Como resultado del punto anterior, es probable que se sobreestime el coeficiente de determinación y, con ello, que se establezca una relación estadística entre variables que no existe o no con la misma intensidad que se cree.
- c. Aunque la varianza del modelo no esté subestimada, la varianza de los parámetros estimados puede subestimar la varianza de estos mismos parámetros en presencia de autocorrelación.
- d. Como consecuencia de lo anterior, las pruebas de significancia  $t$  y  $F$  usuales dejan de ser válidas y, de aplicarse, es probable que conduzcan a conclusiones erróneas sobre la significancia estadística de los coeficientes de regresión estimados.

Figura 5



Fuente: (Gujarati & Porter, 2010, pág. 67).

***I.III. VII. El número de observaciones  $n$ -ésimas debe ser mayor que el número de parámetros  $n-k$  a estimar***

Otra forma de expresar el enunciado anterior, como se señala en (Gujarati & Porter, 2010, pág. 67), es decir que sucesivamente el número de observaciones  $n$ -ésimas debe ser mayor que el número de variables explicativas.

Como se señala en (ResearchGate, 2019) y (ResearchGate, 2014), normalmente si el número de parámetros de un modelo es menor que el número de observaciones

disponibles, se puede utilizar algún proceso de optimización (como el método de mínimos cuadrados) para determinar los parámetros óptimos del modelo. Sin embargo, si el número de parámetros es mayor que el tamaño de muestra, entonces no se puede determinar un conjunto único de parámetros. Por supuesto, en ciertos casos, se ha propuesto el enfoque de multiplicadores de Lagrange como solución al problema, como se verifica en (MIT Computer Science & Artificial Intelligence Lab, 2021).

Para fines ilustrativos de lo anterior, considérese algún modelo lineal simple. Estos tienen dos parámetros, el gradiente (el intercepto) y el desplazamiento (la pendiente). Se necesitan dos o más puntos de datos para estimar los valores numéricos de dichos parámetros. Si solo se tuviera una observación, entonces se podría ajustar un número infinito de líneas y serían igualmente viables. Sin embargo, si se conociera la información anterior sobre uno de los parámetros del modelo, entonces el número infinito de soluciones se puede colapsar en una sola solución.

Así, mediante el uso de supuestos razonables, puede ser posible reducir el número de observaciones que serían necesarias, en principio, para determinar completamente (estimar) los parámetros. Por ejemplo, podemos tomar el «diámetro de la partícula de tamaño mínimo permisible» como parámetro de ajuste, obteniendo así su estimación a partir de los datos experimentales disponibles, o podemos aceptar (suposición razonable) que puede tomarse como nulo.

Esta opción reduce el número de grados de libertad en una unidad. Otro supuesto que puede conducir a un grado de libertad menos es aceptar que se puede establecer algún tipo de relación adicional entre dos o más parámetros (por lo tanto, no independientes). Por supuesto, tal relación previamente aceptada también disminuirá el número de grados de libertad en uno, lo que elevaría el conteo total a dos grados de libertad menos. Lo mismo ocurre si uno de los parámetros se fija en algún "valor razonable", por lo que no se ajusta al conjunto considerado de puntos de datos experimentales, es decir, a la muestra original. Evidentemente, tales supuestos restringen la generalidad de la correlación considerada originalmente.

Con un determinado número restante de grados de libertad, todavía requiere al menos el mismo número de observaciones independientes para que se determinen (estimen) los parámetros correspondientes.

En general, como se señala en (Wikipedia, 2021), la terminología se puede describir en términos del concepto de recuento de restricciones. Cada incógnita puede verse

como un grado de libertad disponible. Cada ecuación introducida, que estima cada una de las incógnitas en el sistema, puede verse a su vez como una restricción que resta un grado de libertad a la cantidad de grados totales disponibles. Por tanto, el caso crítico ocurre cuando el número de ecuaciones y el número de variables libres son iguales. Para cada variable que da un grado de libertad, existe una restricción correspondiente. El caso sobredeterminado ocurre cuando el sistema se ha restringido en exceso, es decir, cuando las ecuaciones superan en número a las incógnitas; en estos contextos, el sistema de ecuaciones no tiene solución o, dicho de otra forma, es inconsistente, *i.e.*, es posible derivar de él una contradicción lógica-formal, por ejemplo,  $0 = 1$ . En contraste, el caso indeterminado ocurre cuando el sistema ha sido subrestringido, es decir, cuando el número de ecuaciones es menor que el número de incógnitas. Estos sistemas suelen tener un número infinito de soluciones.

Finalmente, para comprender de forma aplicada el impacto que tiene la violación del supuesto establecido en un modelo estadístico, conviene estudiar este fenómeno desde la perspectiva de la Teoría del Aprendizaje Estadístico. Como señala (Gujarati & Porter, 2010, pág. 67), otra forma de expresar el supuesto analizado en esta sección es decir que “Sucesivamente, el número de observaciones  $n$  debe ser mayor que el número de variables explicativas”. Cuando el tamaño de la muestra es menor que el número de variables explicativas puede ocurrir el fenómeno conocido como *sobreajuste del modelo* o, lo que es lo mismo, se dice que el modelo está sobredeterminado. ¿Qué implicaciones tiene en términos aplicados que un modelo esté sobredeterminado?

Como señala (Banerjee, 2019), para comprender esto es necesario comprender antes las dos cualidades fundamentales que un modelo de aprendizaje estadístico posee:

1. Especificidad: Capacidad para realizar un aprendizaje centrado en los datos de entrenamiento, de modo que el modelo aprenda los detalles minuciosos de los datos.
2. Generalidad: Capacidad para generalizar el modelo para datos de prueba no conocidos durante el entrenamiento (*i.e.*, que no se utilizaron), mediante el entrenamiento realizado con los datos conocidos, es decir, con aquellos que sí se utilizaron.

Así, cuando se dice que un modelo se ajusta a los datos, se hace referencia a un equilibrio entre estas dos cualidades fundamentales, esto es, que es capaz de identificar tendencias o patrones generales sin ser demasiado específico (aunque, evidentemente, es deseable cierto nivel de especificidad).

¿Qué ocasiona entonces el sobreajuste de un modelo en el estudio aplicado de los fenómenos naturales y sociales? Como señala (Zhao, 2017), en el contexto de un

sobreajuste del modelo, la hipótesis de aprendizaje estadístico (que no es más que la métrica que mide las distancias entre los valores observados y estimados, cuya forma funcional variará según el problema estudiado) puede complementarse muy bien, pero fallar al intentar generalizar la tarea para nuevas muestras. Cuando existe un sobreajuste, el modelo carece de generalidad. De forma equivalente, en el contexto de la Teoría Estadística Clásica, esto implica que no se pueden generalizar las conclusiones del modelo (con sobreajuste) construido, lo que, de hecho, suele ser una necesidad imperiosa en las investigaciones científicas. (Zhao, 2017) señala que un sobreajuste implica que la volatilidad del modelo (encarnada en la varianza) es elevada en el contexto de un sobreajuste.

Complementariamente, aunque no es un componente explícito de los supuestos del MCRL, conviene exponer sintéticamente en qué consiste un subajuste del modelo de aprendizaje. Como señala (Banerjee, 2019), un subajuste implica lo inverso de un sobreajuste, es decir, un subajuste es aquel escenario en que al modelo le hacen falta más variables explicativas para realizar inferencias (sean de tipo supervisado o no supervisado) estadísticamente significativas sobre un determinado conjunto de datos, lo que ocasiona que el modelo de aprendizaje no pueda capturar las minucias del conjunto de datos, por lo que tiene alta capacidad de generalización, pero esa capacidad de generalización va de la mano con lo que podría llamarse un “aprendizaje superficial”, es decir, un aprendizaje en el que el modelo no captura ningún aspecto significativo de la generalidad que abarcó. La consecuencia aplicada de ello, como señala (Zhao, 2017) es que el modelo está sesgado. El sesgo de aprendizaje automático, también llamado a veces sesgo de algoritmo o sesgo de inteligencia artificial, es un fenómeno que ocurre cuando un algoritmo produce resultados que tienen prejuicios sistémicos debido a suposiciones erróneas en el proceso de aprendizaje automático. Un elevado sesgo de aprendizaje puede hacer que un algoritmo pierda las relaciones relevantes entre las características y los verdaderos valores de la variable de respuesta.

Finalmente surge entonces la pregunta, ¿cuál es la diferencia entre sesgo de aprendizaje y varianza? Como señala (Guanga, 2018), el sesgo de aprendizaje es entonces la tendencia del algoritmo a aprender constantemente lo incorrecto al no tener en cuenta toda la información de los datos (subajuste), lo mismo que ocurre cuando una persona da su opinión y se dice que “está sesgada” (lo cual no es un problema cualitativo sino cuantitativo, porque siempre existirá sesgo). Por otro lado, la varianza en el contexto del aprendizaje estadístico hace referencia a la sensibilidad de un algoritmo circunscrita a conjuntos específicos del conjunto de entrenamiento de tal forma que esta sensibilidad impide al algoritmo generalizar, es decir, que

tenga la suficiente flexibilidad para aprender la verdadera señal del conjunto de datos. Así, la varianza es un error de sensibilidad a pequeñas fluctuaciones en el conjunto de entrenamiento. Una gran variación puede hacer que un algoritmo modele el ruido aleatorio en los datos de entrenamiento, en lugar de los resultados previstos (sobreajuste).

### ***I.III. VIII. Independencia lineal entre las variables explicativas***

Como señalan (Gujarati & Porter, 2010, págs. 320-342), esto es conocido en el contexto de la estadística aplicada como ausencia de multicolinealidad. La multicolinealidad puede obedecer a los siguientes factores:

1. El método de recolección de la información. Por ejemplo, la obtención de muestras en un intervalo limitado de valores tomados por las regresoras en la población (muestra pequeña).
2. Restricciones en el modelo o en la población objetivo de muestreo. Por ejemplo, en la regresión del consumo de electricidad sobre el ingreso y el tamaño de las viviendas hay una restricción física en la población, pues las familias con ingresos más altos suelen habitar en viviendas más grandes que las familias con ingresos más bajos.
3. Especificación del modelo. Por ejemplo, la adición de términos polinomiales a un modelo de regresión, en especial cuando el rango estadístico de las variables explicativas es pequeño.
4. Un modelo sobredeterminado. Esto sucede cuando el modelo tiene más variables explicativas que el número de observaciones. Esto puede suceder en investigación médica, donde en ocasiones hay un número reducido de pacientes sobre quienes se reúne información respecto de un gran número de variables.
5. Que las regresoras compartan una tendencia común, es decir, que todas aumenten o disminuyan a lo largo del tiempo.

Matemáticamente hablando, las estimaciones en presencia de multicolinealidad perfecta impiden la estimación de los parámetros, puesto que algebraicamente hablando el resultado de su estimación será la operación cero entre cero, cuyo resultado es indeterminado. Más en profundidad, este cociente refleja que en presencia de multicolinealidad perfecta no puede obtenerse una solución única para las ecuaciones por medio de las cuales se encuentran los coeficientes de regresión o parámetros del modelo. Por su parte, la estimación en presencia de multicolinealidad alta, pero imperfecta, es posible de realizar matemáticamente hablando, con independencia de la validez de los resultados de la estimación. Finalmente, las consecuencias prácticas de la multicolinealidad son:

1. Aunque los estimadores de MCO son MELI, presentan varianzas y covarianzas grandes que dificultan la estimación precisa. De lo anterior se

desprende que el hecho de que los estimadores sean eficientes, no significa que sean precisos.

2. Debido a lo anterior, los intervalos de confianza tienden a ser mucho más amplios, lo cual propicia una aceptación más fácil de la “hipótesis nula cero” (es decir, que el verdadero coeficiente poblacional es cero).
3. También debido a la primera consecuencia, ocurre que la razón  $t$  de uno o más coeficientes tiende a ser estadísticamente no significativa.
4. Aunque la razón  $t$  de uno o más coeficientes sea estadísticamente no significativa, el coeficiente de determinación (que es la medida global de bondad de ajuste) puede ser muy alto.
5. Los estimadores MCO y sus errores estándar son sensibles a pequeños cambios en los datos.

Como se señala en (Salmerón Gómez, Blanco Izquierdo, & García García, 2016, págs. 3-4), existen dos tipos de multicolinealidad:

- a. La multicolinealidad sistemática, la cual es debida a un problema estructural, es decir, a la alta correlación lineal de las variables exógenas que en concreto se han especificado.
- b. La multicolinealidad errática, que es debidamente a un problema numérico.

Para este segundo tipo de multicolinealidad es que, según los autores citados en el lugar referido, el famoso economista Arthur Goldberger acuñó el término *micronumerosidad*. Es importante destacar que, según (Gujarati & Porter, 2010, pág. 332), Goldberger acuñó dicho término como una parodia de las consecuencias de la multicolinealidad. Esto se complementa con otras dos referencias presentadas por (Gujarati & Porter, 2010, pág. 320), específicamente una de Edward E. Leamer (reconocido economista, estadístico y catedrático de UCLA), otra de Christopher Achen (reconocido politólogo, estadístico y catedrático de Princeton), otra de Olivier Blanchard (anterior economista jefe del Fondo Monetario Internacional, catedrático de Harvard y MIT, así como uno de los economistas más citados del mundo) y otra de Peter Kennedy (economista, estadístico y catedrático de la Simon Fraser University).

Leamer señala que no hay una expresión más errónea, tanto en los libros de texto de econometría como en la bibliografía aplicada, que la de “problema de multicolinealidad”. Es un hecho que muchas variables explicativas presentan un alto grado de colinealidad; asimismo, resulta muy claro que existen diseños experimentales que serían mucho más convenientes que los diseños que proporciona la experimentación natural (es decir, la muestra disponible). No obstante, no es nada constructivo quejarse de la aparente malevolencia de la naturaleza, y los remedios *ad hoc* para un mal diseño -como una regresión por pasos o una regresión en cadena- pueden ser desastrosamente inapropiados. Es

mejor aceptar de plano que los datos que no se recopilaron mediante experimentos diseñados a veces no proporcionan mucha información sobre los parámetros de interés.

Por su parte, Achen señala que los novatos en el estudio de la metodología en ocasiones se preocupan porque sus variables independientes estén correlacionadas: el llamado problema de multicolinealidad. Sin embargo, la multicolinealidad no viola los supuestos básicos de regresión. Se presentarán estimaciones consistentes e insesgadas y sus errores estándar se estimarán de forma correcta. El único efecto de la multicolinealidad tiene que ver con la dificultad de obtener coeficientes estimados con errores estándar pequeños [aquí “pequeños” no hace referencia a mínimo, que es un concepto matemático, sino que hace referencia a un intervalo de confianza que contenga los valores de los parámetros que para fines de estimación y control de políticas pueda ser considerado pequeño (según el marco científico de referencia, las necesidades objetivas que motivaron la investigación, el criterio experto y, por supuesto, la evidencia objetiva -que la constituyen los hechos, más allá de los datos-)]. Sin embargo, se presenta el mismo problema al contar con un número reducido de observaciones o al tener variables independientes con varianzas pequeñas. De hecho, en el nivel teórico, los conceptos de multicolinealidad, número reducido de observaciones y varianzas pequeñas en las variables independientes forman parte esencial del mismo problema. Por tanto, la pregunta “¿qué debe hacerse entonces con la multicolinealidad?” es similar a “¿qué debe hacerse si no se tienen muchas observaciones?” Al respecto no hay una respuesta estadística.

Blanchard señala que “Cuando los estudiantes efectúan por primera vez la regresión de MCO, el primer problema que suelen afrontar es el de multicolinealidad. Muchos concluyen que hay algo malo con los MCO; otros recurren a nuevas y con frecuencia creativas técnicas a fin de darle la vuelta al problema. Pero eso está mal. La multicolinealidad es la voluntad de Dios, no un problema con los MCO ni con la técnica estadística general.”

Kennedy señala que en el trabajo aplicado que se nutre de información secundaria (la información recopilada por alguna institución, como la información del producto nacional bruto recopilada por el gobierno), es posible que un investigador por sí solo no pueda hacer gran cosa sobre el tamaño de la información muestral, y quizá deba enfrentar la estimación de problemas lo bastante importantes (en términos de las consecuencias prácticas de una mala estimación) para justificar su tratamiento (el de la multicolinealidad) como una violación del modelo clásico de regresión lineal.

Adicionalmente, señalan Gujarati y Porter que:

1. Es cierto que aún en el caso de casi multicolinealidad los estimadores de MCO son insesgados. Pero el insesgamiento es una propiedad multimuestral o de muestreo repetido. Esto significa que, si mantenemos fijos los valores de  $X$ , si obtenemos muestras repetidas y calculamos los estimadores de MCO para cada una de esas muestras, el promedio de los valores muestrales se aproximará a los verdaderos valores poblacionales de los estimadores a medida que aumenta el número de las muestras. Pero esto nada dice sobre las propiedades de los estimadores en una muestra concreta.
2. La multicolinealidad no destruye la propiedad de varianza mínima; en la clase de los estimadores lineales insesgados, los estimadores de MCO tienen varianza mínima; es decir, son eficientes. Pero esto no significa que la varianza de un estimador de MCO necesariamente sea pequeña (en relación con el valor del estimador) en cualquier muestra dada, como demuestran los autores citados en el lugar citado.
3. La multicolinealidad es en esencia un fenómeno de regresión muestral en el sentido en que, aunque las variables  $X$  no estén linealmente relacionadas en la población, pueden estarlo en la muestra particular disponible: cuando se postula la función de regresión teórica o poblacional (FRP), se considera que todas las variables  $X$  incluidas del modelo ejercen una influencia separada o independiente sobre la variable dependiente  $Y$ . Pero puede suceder que en cualquier muestra dada con que se pruebe la FRP, alguna o todas las variables  $X$  sean tan colineales que no sea posible aislar su influencia individual sobre  $Y$ .

Complementariamente, debe decirse que en el contexto de la Bioestadística (de su generalidad, al menos) la multicolinealidad se considera un problema únicamente si es moderada o alta, como se verifica en (Penn State University, Eberly College of Science, 2018) y (Simon Fraser University, 2011). Lo mismo ocurre en el contexto de la Econometría, en el cual el indicador de la magnitud de la multicolinealidad suele ser el *factor de inflación de varianza* (VIF, por sus siglas en inglés) y se considera únicamente un problema cuando este reporta valores mayores o iguales a diez. Como señalan (Gujarati & Porter, 2010, pág. 328), este indicador toma su nombre del hecho que al incrementar la colinealidad incrementa la varianza del estimador también y, en el límite, se vuelve infinito. Ello es así puesto que  $VIF_{ij} = \frac{1}{1-r_{ij}^2}$ , en donde  $r_{ij}^2$  es el coeficiente de determinación entre la variable explicativa  $i$  –ésima y la variable explicativa  $j$  –ésima.

## II. LOS MODELOS LINEALES GENERALIZADOS (MLG)

### II.I. Supuestos del MLG<sup>2</sup>

Como se señala en (Nabi, 2021), el ajuste de una simple relación lineal entre los elementos  $x$  y los elementos  $y$  requiere que sea utilizado un par de valores en particular  $(a, b)$ , seleccionado del conjunto de todos los posibles pares de valores que pueden adoptar los parámetros  $(\alpha, \beta)$ , que genere valores estimados  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$  que se ajusten (en el sentido de generar una figura geométrica -en este caso una línea recta-) mejor a los valores observados  $y_1, y_2, \dots, y_n$ .

Con la finalidad de cuantificar las diferencias en el ajuste entre los valores estimados  $\hat{y}_n$  y los valores observados  $y_n$ , se deben medir las distancias entre tales conjuntos de valores. En general, estas distancias (referidas en textos como el de McCullagh y Nelder como discrepancias) buscan cuantificarse con la finalidad de optimizar el proceso de selección del patrón geométrico teórico que mejor describa el patrón observado descrito por el conjunto de datos. Para cuantificar tales distancias se pueden utilizar diferentes tipos de métricas (criterios de medición), selección que a su vez responderá a las características específicas del conjunto de datos,

Como señalan (McCullagh & Nelder, 1989, pág. 5), existen desde la norma  $L_1$  definida como  $S_1(y, \hat{y}) = \sum |y - \hat{y}|$  hasta la norma  $L_p$  definida como  $S_p(y, \hat{y}) = \max_i |y - \hat{y}|$ . Sin embargo, los espacios que usualmente son de más interés aplicado, así como también aquellos en que se realiza el ajuste clásico por mínimos cuadrados, son los espacios  $L_2$  normados, conocidos también a nivel estadístico como espacios de desviaciones cuadráticas, que son los espacios aritméticos euclidianos  $n - dimensionales$  antes definidos a nivel métrico. A continuación, se define la norma de estos espacios.

$$S_2(y, \hat{y}) = \sum (y_i - \hat{y}_i)^2$$

Así, la validez de las estimaciones realizadas tiene determinados elementos en común con independencia del orden de la norma que se utilice, orden que indica precisamente la potencia a la que se eleva la diferencia entre las localizaciones de los elementos de los conjuntos de datos (del estimado o generado con la regresión y del observado). Estos elementos en común son tres:

- 1) En primer lugar, todas las mediciones (expresadas en las observaciones) han sido realizadas bajo la misma escala física.

---

<sup>2</sup> Si se desea consultar los fundamentos teóricos de este modelo, más allá de sus supuestos, véase (Nabi, 2021).

- 2) Las observaciones son independientes entre sí o, al menos, "(...) ellas son en algún sentido intercambiables, lo que justifica un trato imparcial de los componentes." (McCullagh & Nelder, 1989, pág. 5).
- 3) Cada una de las desviaciones debe ser independiente del valor esperado del conjunto de observaciones.

Con base en (Penn State University, Eberly College of Science, 2021), deben agregarse los siguientes supuestos:

- 4) La variable dependiente no debe estar necesariamente distribuida normalmente, pero debe asumir una distribución perteneciente a las de alguna familia exponencial (binomial, Poisson, multinomial, normal, exponencial, etc.).
- 5) Como resulta evidente, el MLG no asume una relación lineal entre las variables independientes y la variable dependiente, pero sí asume que existe una relación lineal entre la variable de respuesta transformada (mediante la función enlace) y las variables independientes en cuestión.
- 6) No es necesario satisfacer la homogeneidad de la varianza. De hecho, ni siquiera es posible en muchos casos dada la estructura del modelo, y puede haber una sobredispersión (cuando la varianza observada es mayor de lo que supone el modelo).
- 7) Los errores deben ser independientes, pero no necesariamente normalmente distribuidos.
- 8) Utiliza la estimación de máxima verosimilitud (MLE) en lugar de mínimos cuadrados ordinarios (MCO) para estimar los parámetros y, por lo tanto, se basa en aproximaciones de muestras grandes.
- 9) Las medidas de bondad de ajuste se basan en muestras suficientemente grandes, donde una regla heurística es que no más del 20% de los conteos celulares esperados sean inferiores a 5.

Por supuesto, para un lector no familiarizado con la biología molecular, el supuesto 9 no proporciona mayor información. El *conteo celular*, proveniente del inglés "cell counting", consiste en, como su nombre lo indica, realizar un conteo o alguna cuantificación similar sobre las células. Para comprender a qué se refiere al afirmar que "una regla heurística es que no más del 20% de los conteos celulares esperados sean inferiores a 5" hay que comprender primero qué es una regla heurística y qué significa la regla como tal en este contexto.

Intuitivamente, una regla heurística es una regla (o conjunto de reglas) destinada a aumentar la probabilidad de resolver algún problema, en donde esta regla (o conjunto de estas) ha sido formulada con base a la investigación empírica que históricamente se ha realizado en tal o cual disciplina, contrastante con la lógica del planteamiento de soluciones analíticas (en el sentido en que estas son

entendidas en la matemática pura). Complementariamente, si se retoma la terminología de las encuestas por muestreo (cuyo fundamento teórico es más abstracto y, por consiguiente, de mayor alcance aplicativo), se puede verificar que el conteo celular es el equivalente en la teoría de las encuestas por muestreo a las unidades de muestreo (las cuales que pueden tener varios elementos dentro de sí y pertenecen a una estructura de datos más general conocido como marco muestral), es decir, "conteo celular" hace referencia a un conjunto de células encontradas en un conteo dentro de una determinada unidad de análisis (unidad de carácter biológico, molecular o de otra índole, lo cual dependerá del contexto específico de aplicación). Así, puede afirmarse que la regla heurística plantea, en general, lo siguiente: *No más del 20% de las unidades muestrales deben contener dentro de sí menos de 5 observaciones.*

Como se deduce de lo anterior, cada una de las normas  $L_p$  –ésimas se corresponde con un determinado criterio estadístico, específicamente con la potencia a la que se elevan las distancias entre los valores observados y los valores estimados, puesto que así se realiza la medición de los residuos o errores en el contexto de los modelos de mínimos cuadrados.

### III. REFERENCIAS

Banerjee, A. (29 de Octubre de 2019). *Intuition behind model fitting: Overfitting v/s Underfitting*. Obtenido de Towards Data Science:

<https://towardsdatascience.com/intuition-behind-model-fitting-overfitting-v-s-underfitting-d308c21655c7>

Bhuptani, R. (13 de Julio de 2020). *Quora*. Obtenido de What is the difference between linear regression and least squares?:

<https://www.quora.com/What-is-the-difference-between-linear-regression-and-least-squares>

Cross Validated. (23 de Marzo de 2018). *Will log transformation always mitigate heteroskedasticity?* Obtenido de StackExchange:

<https://stats.stackexchange.com/questions/336315/will-log-transformation-always-mitigate-heteroskedasticity>

Greene, W. (2012). *Econometric Analysis* (Séptima ed.). Harlow, Essex, England: Pearson Education Limited.

Guanga, A. (11 de Octubre de 2018). *Machine Learning: Bias VS. Variance*. Obtenido de Becoming Human: Artificial Intelligence Magazine:

<https://becominghuman.ai/machine-learning-bias-vs-variance-641f924e6c57>

Gujarati, D., & Porter, D. (8 de Julio de 2010). *Econometría* (Quinta ed.). México, D.F.: McGrawHill Educación. Obtenido de Homocedasticidad.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (Segunda ed.). London: Chapman and Hall.

MIT Computer Science & Artificial Intelligence Lab. (6 de Mayo de 2021). *Solving over- and under-determined sets of equations*. Obtenido de Articles:

[http://people.csail.mit.edu/bkph/articles/Pseudo\\_Inverse.pdf](http://people.csail.mit.edu/bkph/articles/Pseudo_Inverse.pdf)

Nabi, I. (27 de Agosto de 2021). *MODELOS LINEALES GENERALIZADOS*.

Obtenido de El Blog de Isadore Nabi:

<https://marxianstatistics.files.wordpress.com/2021/08/modelos-lineales-generalizados-isadore-nabi.pdf>

Penn State University, Eberly College of Science. (2018). *10.4 - Multicollinearity*.

Obtenido de Lesson 10: Regression Pitfalls:

<https://online.stat.psu.edu/stat462/node/177/>

Penn State University, Eberly College of Science. (24 de Mayo de 2021). *Introduction to Generalized Linear Models*. Obtenido de Analysis of Discrete Data:

<https://online.stat.psu.edu/stat504/lesson/6/6.1>

Perezgonzalez, J. D. (3 de Marzo de 2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *frontiers in PSYCHOLOGY*, VI(223), 1-11.

ResearchGate. (10 de Noviembre de 2014). *How it can be possible to fit the four-parameter Fedlund model by only 3 PSD points?* Obtenido de [https://www.researchgate.net/post/How\\_it\\_can\\_be\\_possible\\_to\\_fit\\_the\\_four-parameter\\_Fedlund\\_model\\_by\\_only\\_3\\_PSD\\_points](https://www.researchgate.net/post/How_it_can_be_possible_to_fit_the_four-parameter_Fedlund_model_by_only_3_PSD_points)

ResearchGate. (28 de Septiembre de 2019). *s there a rule for how many parameters I can fit to a model, depending on the number of data points I use for the fitting?* Obtenido de <https://www.researchgate.net/post/Is-there-a-rule-for-how-many-parameters-I-can-fit-to-a-model-depending-on-the-number-of-data-points-I-use-for-the-fitting>

Salmerón Gómez, R., Blanco Izquierdo, V., & García García, C. (2016). Micronumerosidad aproximada y regresión lineal múltiple. *Anales de ASEPUMA*(24), 1-17. Obtenido de <https://dialnet.unirioja.es/descarga/articulo/6004585.pdf>

Simon Fraser University. (30 de Septiembre de 2011). *THE CLASSICAL MODEL*. Obtenido de <http://www.sfu.ca/~dsignori/buec333/lecture%2010.pdf>

StackExchange Cross Validated. (2 de Febrero de 2017). *"Least Squares" and "Linear Regression", are they synonyms?* Obtenido de [What is the difference between least squares and linear regression? Is it the same thing?: https://stats.stackexchange.com/questions/259525/least-squares-and-linear-regression-are-they-synonyms](https://stats.stackexchange.com/questions/259525/least-squares-and-linear-regression-are-they-synonyms)

Wikipedia. (18 de Marzo de 2021). *Overdetermined system*. Obtenido de [Partial Differential Equations: https://en.wikipedia.org/wiki/Overdetermined\\_system](https://en.wikipedia.org/wiki/Overdetermined_system)

Zhao, J. (9 de Noviembre de 2017). *More features than data points in linear regression?* Obtenido de Medium: <https://medium.com/@jennifer.zzz/more-features-than-data-points-in-linear-regression-5bcabba6883e>

