

Aplicación de la factorización QR usando transformaciones de Householder en mínimos cuadrados para un análisis de la pobreza en el Perú

Ayrton Fabio Coronado Huamán ¹, Guillermo Joel Borjas Córdova ², Israel Danilo Blas Salas ³, Tomás García Sifuentes ⁴

Facultad de Ciencias 1, Universidad Nacional de Ingeniería 1, e-mail:

Facultad de Ciencias 2, Universidad Nacional de Ingeniería 2, e-mail:

1. Introducción

En este proyecto nos dedicaremos a construir un algoritmo eficiente, en términos de tiempo de ejecución, del método matemático "Transformación de Householder", el cual nos permite obtener un tipo de descomposición de matrices llamado "Descomposición QR". Esto nos servirá en la resolución de un sistema de ecuaciones proveniente del ajuste de una curva utilizando la técnica de mínimos cuadrados donde intervienen una variable dependiente y una independiente, y en la cual la relación entre ellas se aproxima por medio de una línea recta, todo esto con la finalidad de poder aplicarse a una situación particular en el campo de la economía, más concretamente, para hacer un análisis de la pobreza en el Perú y de esta manera hacer predicciones o pronósticos que nos ayudarán a tener un mejor entendimiento de este problema. Con el fin de llevar a cabo este proyecto, se emplearán los datos tomados de una muestra real, extraídos de un informe hecha por el INEI de pobreza, pobreza extrema y el coeficiente de Gini (indicador de la desigualdad económica en una población) en los años 2015 - 2016 de las principales regiones del Perú.

Muchos autores que han hecho estudios sobre modelos de regresión, entre los que se pueden citar a: Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2001), Devore, J. L. (2005), Evans, M., & Rosenthal, J. S. (2005), Freund, J. E., & Simon, G. A. (1994), Levin,

R. I., & Rubin, D. S. (2004) y Miller, I. (2000); coinciden en que siempre que se analizan datos observados o recopilados para llegar a una función o ecuación matemática que describa la relación entre las variables por medio de una regresión, se deben enfrentar tres problemas:

1. Decidir qué clase de curva muestran los puntos y por tanto qué clase de ecuación se debe usar.
2. Encontrar la ecuación particular que mejor se ajuste a los datos.
3. Demostrar que la ecuación particular encontrada cumple con ciertos aspectos referentes a los méritos de ésta para hacer pronósticos.

Para decidir qué clase de función podría ajustarse a la curva, debe hacerse una gráfica de dispersión de los datos observados. Si en dicha gráfica se aprecia que los puntos se distribuyen alrededor de una recta, se procede a realizar un análisis de regresión lineal.

2. Conceptos Previos

2.1. SISTEMAS LINEALES DE ECUACIONES SOBREDETERMINADOS

El problema lineal de mínimos cuadrados se acerca a la optimización si se encuentra un vector \mathbf{x} que en algún sentido da la mejor (aunque no la perfecta) aproximación entre \mathbf{Ax} y \mathbf{b} .

Considérese un sistema de m ecuaciones con n incógnitas escrito en la forma

$$\mathbf{Ax} = \mathbf{b} \quad (\text{II.1})$$

Aquí, \mathbf{A} es de $m \times n$, \mathbf{x} es de $n \times 1$ y \mathbf{b} es de $m \times 1$. Supondremos que el rango de \mathbf{A} es n , de lo que resulta que $m \geq n$. Si $m > n$, se dice que el sistema lineal de ecuaciones es sobredeterminado. Los sistemas sobredeterminados son generalmente no compatibles.

2.2. EL VECTOR RESIDUAL Y EL PROBLEMA DE LOS MÍNIMOS CUADRADOS

Por lo general, el sistema (II.1) no tiene solución debido a que \mathbf{b} no pertenece al subespacio de \mathbf{R}^n de dimensión n , generado por las columnas de \mathbf{A} . Es frecuente en tales casos que se requiera encontrar un \mathbf{x} tal que \mathbf{b} esté cerca de \mathbf{Ax} , es decir que minimice una norma del vector residual $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$. El primer problema será determinar el sentido de cercanía, es decir la norma que mide la diferencia entre \mathbf{b} y \mathbf{Ax} . La interpretación y la naturaleza de la solución de este problema varía dependiendo de la norma que se use. La norma más comúnmente usada es la norma euclidiana, la norma \mathbf{l}_2 . Entonces, la solución en mínimos cuadrados de (II.1) es el vector \mathbf{x} que hace de $\|\mathbf{r}\|_2 = \|\mathbf{b} - \mathbf{Ax}\|_2$ un mínimo

$$\min_{\mathbf{x} \in \mathbf{R}^n} \|\mathbf{b} - \mathbf{Ax}\|_2^2, \quad (\text{II.2})$$

Dado que las normas no pueden ser negativas, minimizar una norma es equivalente a minimizar su cuadrado; hablaremos entonces libremente de normas cuadradas y no. Según lo que se ha supuesto acerca del rango de \mathbf{A} , la solución \mathbf{x} será única.

Veremos que el problema de mínimos cuadrados se puede resolver transformando la matriz \mathbf{A} y resolviendo un sistema compatible a ella relacionado. Al contrario, minimizar la norma \mathbf{l}_1 o la norma \mathbf{l}_∞ es equivalente a un problema de programación lineal no diferenciable, cuya solución necesita una técnica de iteración.

Es útil pensar en el problema de los mínimos cuadrados en términos de los subespacios definidos por la matriz \mathbf{A} . El subespacio rango de \mathbf{A} , $\text{rango}(\mathbf{A})$, consiste en todos los vectores m -dimensionales que son combinaciones lineales de las columnas de \mathbf{A} y el subespacio complementario, el subespacio nulo de \mathbf{A}^t , $\text{nulo}(\mathbf{A})$, contiene todos los vectores m -dimensionales que son z -ortogonales a las columnas de \mathbf{A} , es decir tales que $\mathbf{A}^t \mathbf{z} = \mathbf{0}$. Si r denota el rango de la matriz \mathbf{A} , entonces r es la dimensión del subespacio rango de \mathbf{A} y $(m - r)$ es la dimensión del subespacio nulo de \mathbf{A}^t . Dada una matriz no nula \mathbf{A} , cualquier vector m -

dimensional \mathbf{c} se puede expresar como la suma de un vector \mathbf{c}_R en el rango de \mathbf{A} , $\mathbf{c}_R \in \text{rango}(\mathbf{A})$, y de un vector \mathbf{c}_N en el espacio nulo de \mathbf{A}^t , $\mathbf{c}_N \in \text{nulo}(\mathbf{A}^t)$, es decir

$$\mathbf{c} = \mathbf{c}_R + \mathbf{c}_N. \quad (\text{II.3})$$

Los vectores \mathbf{c}_R y \mathbf{c}_N son únicos y satisfacen

$$\mathbf{c}_R = \mathbf{Ac}_A, \mathbf{A}^t \mathbf{c}_N = \mathbf{0}, \mathbf{c}_R^t \cdot \mathbf{c}_N = 0, \quad (\text{II.4})$$

donde \mathbf{c}_A se refiere a un vector n -dimensional cualquiera tal que $\mathbf{c}_R = \mathbf{Ac}_A$. La unicidad de \mathbf{c}_R y \mathbf{c}_N implica que las componentes en el espacio rango y en el espacio nulo de vectores iguales tienen que ser iguales. En efecto,

$$\mathbf{c} = \mathbf{d} \text{ si y solo si } \mathbf{c}_R = \mathbf{d}_R \text{ y } \mathbf{c}_N = \mathbf{d}_N. \quad (\text{II.5})$$

Aunque \mathbf{c}_R es único, el vector \mathbf{c}_A es único solo si las columnas de \mathbf{A} son linealmente independientes.

Debido al hecho de que la norma euclídea está definida en términos del producto escalar, las relaciones (II.3) y (II.4) tienen una consecuencia muy importante:

$$\|\mathbf{c}\|_2^2 = \mathbf{c}^t \cdot \mathbf{c} = \|\mathbf{c}_R\|_2^2 + \|\mathbf{c}_N\|_2^2, \quad (\text{II.6})$$

de manera que la representación de un vector en términos de sus componentes en el espacio rango y nulo separa el cuadrado de la norma euclídea en dos partes independientes. Esta propiedad no vale para otras normas.

Estas observaciones son relevantes para el problema de los mínimos cuadrados dado que permiten determinar la más pequeña posible norma euclídea del vector residual $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$ para todos vectores \mathbf{x} . Dado que ambos \mathbf{b} y el vector residual \mathbf{r} son vectores m -dimensionales, ambos pueden escribirse en la representación (II.3):

$$\mathbf{b} = \mathbf{b}_R + \mathbf{b}_N \text{ con } \mathbf{b}_R = \mathbf{Ab}_A, \quad (\text{II.7})$$

$$\mathbf{r} = \mathbf{r}_R + \mathbf{r}_N \text{ con } \mathbf{r}_R = \mathbf{Ar}_A$$

Combinando la definición de \mathbf{r} como $\mathbf{b} - \mathbf{Ax}$ con estas relaciones, se obtiene

$$\mathbf{r} = \mathbf{r}_R + \mathbf{r}_N = \mathbf{b} - \mathbf{Ax} = \mathbf{b}_R + \mathbf{b}_N - \mathbf{Ax} = \mathbf{b}_R - \mathbf{Ax} + \mathbf{b}_N \quad (\text{II.8})$$

Un punto obvio pero crucial en esta expresión es que el vector \mathbf{Ax} está enteramente en el rango de \mathbf{A} . Cuando se resta \mathbf{Ax} del vector \mathbf{b} para crear el residual, sigue de (II.5) que la componente en el espacio rango de $\mathbf{b} - \mathbf{Ax}$ tiene que ser $\mathbf{b}_R - \mathbf{Ax}$. En contraste, el restar \mathbf{Ax} no puede eliminar nada a la componente de \mathbf{b} en el espacio nulo, de manera que la componente en el espacio nulo de $\mathbf{b} - \mathbf{Ax}$ es igual a \mathbf{b}_N . Entonces, las componentes en el espacio rango y en el nulo del vector residual satisfacen

$$\mathbf{r}_R = \mathbf{b}_R - \mathbf{Ax} \text{ y } \mathbf{r}_N = \mathbf{b}_N. \quad (\text{II.9})$$

Como ya dicho, el problema de mínimos cuadrados consiste en minimizar la norma l_2 del vector residual. De (II.6) y (II.9) sigue que la norma euclídea del vector residual \mathbf{r} para cualquier vector \mathbf{x} satisface

$$\|\mathbf{r}\|_2^2 = \|\mathbf{b}_R - \mathbf{Ax}\|_2^2 + \|\mathbf{b}_N\|_2^2 \geq \|\mathbf{b}_N\|_2^2. \quad (\text{II.10})$$

Dado que \mathbf{b}_N se mantiene en el residual, se encontrará el mínimo de $\|\mathbf{b} - \mathbf{Ax}\|_2$ cuando la norma de la componente en el subespacio rango, $\|\mathbf{b}_R - \mathbf{Ax}\|_2$, sea la más pequeña posible.

Además, por definición \mathbf{b}_R está en el rango de \mathbf{A} , luego tiene que existir un vector \mathbf{x} tal que $\mathbf{Ax} = \mathbf{b}_R$. Para este vector particular \mathbf{x} , se elimina la componente de \mathbf{b} en el espacio rango cuando se le resta \mathbf{Ax} , que significa que $\mathbf{b} - \mathbf{Ax} = \mathbf{b}_N$ y que la norma euclídea del vector residual es igual a su cota inferior $\|\mathbf{b}_N\|_2$. Escoger \mathbf{x} de esta manera, $\mathbf{Ax} = \mathbf{b}_R$, no solo minimiza la norma del vector residual, sino que además fuerza al residual a estar enteramente en el espacio nulo de \mathbf{A}^t . De esta manera se llega a dos caracterizaciones equivalentes de la solución optimal del problema de mínimos cuadrados:

$$\mathbf{x} \text{ minimiza } \|\mathbf{b} - \mathbf{Ax}\|_2 \iff \mathbf{A}^t(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}, \quad (\text{II.11})$$

$$\mathbf{x} \text{ minimiza } \|\mathbf{b} - \mathbf{Ax}\|_2 \iff \begin{cases} \mathbf{Ax} = \mathbf{b}_R, \\ \mathbf{b} - \mathbf{Ax} = \mathbf{b}_N. \end{cases} \quad (\text{II.12})$$

Un problema de mínimos cuadrados se dice compatible si su vector residual optimal es cero, en este caso $\mathbf{b}_N = \mathbf{0}$ y $\mathbf{b} = \mathbf{b}_R$; si éste no es el caso se dice incompatible. La unicidad del vector \mathbf{b}_N implica que el

vector residual optimal para el problema de mínimos cuadrados es único. El vector \mathbf{b}_R es también único, y el sistema $\mathbf{Ax} = \mathbf{b}_R$ es compatible por definición. Sin embargo, el vector \mathbf{x} que satisface $\mathbf{Ax} = \mathbf{b}_R$ es único si y solo si las columnas de la matriz \mathbf{A} son linealmente independientes. Dado que cualquier vector \mathbf{x} que satisface $\mathbf{Ax} = \mathbf{b}_R$ es una solución de mínimos cuadrados, sigue una importante conclusión: la solución del problema de mínimos cuadrados es única si y solo si la matriz \mathbf{A} es de rango n .

Un aspecto crucial de la relación (II.12) es que revela cómo cambios en el vector \mathbf{b} afectan a la solución \mathbf{x} y al problema de mínimos cuadrados $\min \|\mathbf{b} - \mathbf{Ax}\|_2^2$. Para sistemas de ecuaciones lineales no singulares y compatibles, cambios en \mathbf{b} implicaban encontrar una solución \mathbf{x} diferente. Esta propiedad no se verifica para el problema de los mínimos cuadrados, cuyo análisis es más complicado.

Se supone por simplicidad que \mathbf{A} tiene rango lleno, entonces \mathbf{x} es la única solución del problema de mínimos cuadrados $\min \|\mathbf{b} - \mathbf{Ax}\|_2^2$ y \mathbf{r} es el vector residual optimal, $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$. Si \mathbf{b} viene perturbado por un vector que está enteramente en el subespacio rango de \mathbf{A} , es decir $\tilde{\mathbf{b}} = \mathbf{b} + \delta\mathbf{b}_R$, donde $\delta\mathbf{b}_R = \mathbf{A}\delta\mathbf{x}$ para un único n -vector $\delta\mathbf{x}$, entonces

$$\tilde{\mathbf{b}}_R = \mathbf{b}_R + \delta\mathbf{b}_R = \mathbf{b}_R + \mathbf{A}\delta\mathbf{x} \text{ y } \tilde{\mathbf{b}}_N = \mathbf{b}_N. \quad (\text{II.13})$$

Sigue entonces de (II.12) que la solución de mínimos cuadrados $\tilde{\mathbf{x}}$ y el residual $\tilde{\mathbf{r}}$ correspondientes a $\tilde{\mathbf{b}}$ satisfacen

$$\tilde{\mathbf{x}} = \mathbf{x} + \delta\mathbf{x} \text{ y } \tilde{\mathbf{r}} = \mathbf{r}, \quad (\text{II.14})$$

mostrando que cambios en \mathbf{b} en el espacio rango modifican el vector solución \mathbf{x} pero no el vector residual \mathbf{r} . Por el contrario, si \mathbf{b} viene perturbado por un vector que está enteramente en el subespacio nulo de \mathbf{A}^t , es decir $\mathbf{b} = \mathbf{b} + \mathbf{z}$, donde $\mathbf{A}^t\mathbf{z} = \mathbf{0}$, entonces

$$\mathbf{x} = \mathbf{x} \text{ y } \mathbf{r} = \mathbf{r} + \mathbf{z}. \quad (\text{II.15})$$

mostrando que cambios en \mathbf{b} en el espacio nulo modifican el vector residual \mathbf{r} , mientras que el vector solución \mathbf{x} permanece inalterado.

2.3. LAS ECUACIONES NORMALES

La propiedad más significativa del vector residual optimal del problema de mínimos cuadrados es que está enteramente en el espacio nulo de la matriz A^t . En términos algebraicos, una solución x satisface

$$A^t(b - Ax) = 0, \quad (\text{II.16})$$

que es equivalente a

$$A^t Ax = A^t b. \quad (\text{II.17})$$

Este sistema de ecuaciones se conoce como el de las ecuaciones normales.

La matriz simétrica $A^t A$, conocida como la matriz de la ecuación normal, es semidefinida positiva para una matriz cualquiera A , y es definida positiva si y solo si las columnas de A son linealmente independientes. Las ecuaciones normales son siempre compatibles, es decir existe siempre una solución de (II.17) aunque $A^t A$ sea singular.

Cualquier solución x tiene que satisfacer las ecuaciones normales, y cualquier vector x que satisface las ecuaciones normales tiene que ser una solución de mínimos cuadrados.

Si $A^t A$ es definida positiva, es decir si A tiene rango lleno, las ecuaciones normales (II.17) tienen solución única, que se puede encontrar usando el algoritmo de factorización de Cholesky. Si $A^t A$ es singular, se puede hallar una solución de las ecuaciones normales usando una versión de la factorización de Cholesky que incluya un intercambio simétrico. Entonces, el problema de mínimos cuadrados se puede resolver siguiendo los siguientes pasos:

- (I) Formar la matriz de la ecuación normal $A^t A$ y el vector $A^t b$;
- (II) Calcular la factorización de Cholesky $A^t A = L^t L$, con L triangular superior;
- (III) Resolver los dos sistemas $L^t y = A^t b$ y $Lx = y$.
El vector x es la solución deseada.

Desde el punto de vista computacional, resolver el problema de mínimos cuadrados con las ecuaciones normales es eficiente: para formar $A^t A$ y $A^t b$ se necesitan aproximadamente $m \cdot n^2/2$ operaciones, calcular la factorización de Cholesky usa del orden de $n^3/6$ operaciones y resolver los dos sistemas triangulares implica aproximadamente n^2 operaciones. Desafortunadamente, el uso de las ecuaciones normales presenta el problema de la estabilidad numérica, dado que el número de condición de la matriz $A^t A$ es el cuadrado del número de condición de A . Consecuentemente, la matriz de la ecuación normal está seriamente mal-condicionada si la matriz A misma está ligeramente mal-condicionada.

El mal condicionamiento de la matriz de la ecuación normal asociada puede conducir no solo a una aproximación no precisa de la solución calculada de las ecuaciones normales, sino también a una pérdida de información cuando el rango numérico de A es marginal.

2.4. APLICACIONES

A menudo coleccionamos datos e intentamos encontrar una relación funcional entre las variables. Si los datos son $n + 1$ puntos del plano, es posible encontrar un polinomio de grado n o inferior que pasa por todos los puntos. Este polinomio se llama polinomio de interpolación. Dado que los datos tienen en general error experimental, no hay razón para pedir que las funciones pasen por todos los puntos. De hecho, polinomios de grado inferior que no pasan por los puntos de manera exacta, dan una mejor descripción de la relación entre variables. Por ejemplo, si la relación entre variables es actualmente lineal y los datos tienen un pequeño error, sería desastroso usar un polinomio de interpolación.

Dada una tabla de datos

x	x_1	x_2	\dots	x_m
y	y_1	y_2	\dots	x_m

(II.18)

deseamos encontrar una función lineal

$$y = c_0 + c_1 x \quad (\text{II.19})$$

que mejor aproxima los datos en el sentido de mínimos cuadrados. Si se pide que

$$y_i = c_0 + c_1 x_i \text{ para } i = 1, 2, \dots, m \quad (\text{II.20})$$

obtenemos un sistema de m ecuaciones en dos incógnitas

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad (\text{II.21})$$

La función lineal cuyos coeficientes son la solución de mínimos cuadrados de (II.21) viene a ser la aproximación de mínimos cuadrados a los datos con una función lineal.

Si los datos no aparecen en relación lineal, se podría usar un polinomio de grado mayor. Es decir, para encontrar los coeficientes c_0, c_1, \dots, c_n de la mejor aproximación por mínimos cuadrados a los datos

x	x_1	x_2	\dots	x_m
y	y_1	y_2	\dots	y_m

$$(\text{II.18})$$

con un polinomio de grado n , tenemos que encontrar la solución de mínimos cuadrados al sistema

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^n \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad (\text{II.22})$$

2.5. LAS TRANSFORMACIONES DE HOUSEHOLDER

Debido a las posibles dificultades numéricas del uso de las ecuaciones normales, los modernos métodos de mínimos cuadrados se han desarrollado basándose en las transformaciones ortogonales, que preservan las distancias euclídeas y no empeoran las condiciones de la matriz A . La idea es transformar un problema de mínimos cuadrados de manera que sea fácil de resolver, reduciendo la matriz A a la forma que revela el

rango. El término forma triangular que revela el rango una matriz genérica $m \times n$ de rango r correspondiente a la matriz $\tilde{T} = \begin{pmatrix} T \\ 0 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ 0 & 0 \end{pmatrix}$, con T_{11} matriz $r \times r$ no singular triangular superior y T_{12} matriz $r \times (n - r)$. Cuando \tilde{T} tiene rango lleno de fila entonces no aparecen ambos bloques de ceros; si T' es de rango lleno de columna, la matriz T_{12} y el bloque de ceros derecho no aparecerán. Mas en general, el rango de una matriz $m \times n$, \tilde{F} , es r si \tilde{F} es de la forma $\tilde{F} = \begin{pmatrix} F \\ 0 \end{pmatrix}$, con F matriz $r \times n$ y las filas de F son linealmente independientes.

Dado que ya no estamos resolviendo ecuaciones, el conjunto de transformaciones que se puede aplicar a A sin cambiar la solución está restringido. Las transformaciones ortogonales son obvias en este contexto dado que estas no alteran la norma l_2 . Sea A una matriz no nula $m \times n$, con rango igual a r . Supóngase que se pueda encontrar una matriz $m \times m$ ortogonal Q que produzca la forma que revela el rango

$$QA = \begin{pmatrix} F \\ 0 \end{pmatrix}, \quad (\text{III.1})$$

donde F es una matriz $r \times n$ y tiene las filas linealmente independientes; el bloque de ceros no aparece si $r = m$. La forma (III.1) nos permite resolver el problema de mínimos cuadrados usando las matrices Q y F . Sea d el vector transformado $Q^t b$, descompuesto en

$$d = Q^t b = \begin{pmatrix} d_r \\ d_{r-m} \end{pmatrix}, \quad (\text{III.1})$$

Usando esta definición y la estructura mostrada en (III.1), se puede escribir el vector residual transformado como

$$Q^t(b - Ax) = \begin{pmatrix} d_r \\ d_{r-m} \end{pmatrix} - \begin{pmatrix} Fx \\ 0 \end{pmatrix} = \begin{pmatrix} d_r - Fx \\ d_{m-r} \end{pmatrix}. \quad (\text{III.3})$$

Dado que la matriz Q^t es ortogonal, su aplicación al vector residual no altera la norma euclídea, y

$$\|b - Ax\|_2^2 = \|Q^t(b - Ax)\|_2^2. \quad (\text{III.4})$$

Combinando esta relación con la forma especial del vector residual transformado, se concluye que

$$\|b - Ax\|_2^2 = \|Q^t(b - Ax)\|_2^2 = \|d_r - Fx\|_2^2 + \|d_{m-r}\|_2^2 \geq \|d_{m-r}\|_2^2. \quad (\text{III.5})$$

que significa que $\|b - Ax\|_2$ no puede ser menor que $\|d_{m-r}\|_2$ para cualquier vector x . El valor más pequeño posible para $\|b - Ax\|_2^2$ es la cota inferior. Igualdades con la cota inferior se obtienen si y sólo si el vector x satisface

$$Fx = d_r. \quad (\text{III.6})$$

Dado que el rango de la matriz F es igual a r , el sistema $Fx = d_r$ tiene que ser compatible. Cuando $Fx = d_r$, el vector residual transformado satisface $Q^t(b - Ax) = \begin{pmatrix} 0 \\ d_{r-m} \end{pmatrix}$ y $\|b - Ax\|_2 = \|d_{m-r}\|_2$. Esto demuestra que cualquier vector x que satisfaga $Fx = d_r$ será una solución de mínimos cuadrados. Suponiendo que los sistemas en los cuales aparecen las matrices F son fáciles de resolver, se sigue que el problema de mínimos cuadrados se puede resolver encontrando una matriz ortogonal Q_t que nos dé la forma (III.1).

Antes de presentar la factorización QR para resolver el problema de mínimos cuadrados, es necesario introducir las transformaciones de Householder.

La técnica más popular para construir una matriz ortogonal que reduzca la matriz A en forma triangular usa una clase especial de matrices que son simultáneamente simétricas, elementales y ortogonales. Para cualquier vector no nulo u , la correspondiente transformación de Householder (o matriz de Householder, o reflector de Householder) es una matriz de la forma

$$H = H(u) = I - \frac{2uu^t}{\|u\|_2^2} = I - \frac{uu^t}{\beta}, \text{ con } \beta = \frac{\|u\|_2^2}{2}, \quad (\text{III.7})$$

donde el vector u es el vector de Householder

Teorema Si H es la matriz definida en (III.7), entonces

- 1) $H = H^t$,
- 2) $H = H^{-1}$,

que es lo mismo que decir que la matriz H es simétrica y ortogonal.

Las matrices de Householder son simétricas y ortogonales, y dependen sólo de la dirección del vector u . En el contexto de las reducciones a matrices triangulares, las matrices de Householder poseen dos propiedades cruciales:

- Para cualquier par de vectores distintos de igual norma $\|u\|_2$, existe una transformación de Householder que transforma el uno en el otro,

$$Ha = (I - \frac{uu^t}{\beta})a = b, \quad (\text{III.8})$$

con $\|a\|_2 = \|b\|_2$. Esto implica que el vector u tiene que satisfacer la condición

$$-\frac{u^t a}{\beta} u = b - a, \quad (\text{III.9})$$

es decir, u es un múltiplo de $b - a$; - cualquier vector c transformado por una matriz de Householder posee una forma especial:

$$Hc = (I - \frac{u^t u}{\beta})c = c - \frac{u^t c}{\beta} u, \quad (\text{III.10})$$

de manera que Hc es la diferencia entre el vector original c y un múltiplo especial del vector de Householder u . Claramente el vector c no varía si $u^t c = 0$. Además, calcular el producto Hc no necesita los elementos explícitos de H , sino sólo el vector u y el escalar β .

2.6. LA FACTORIZACIÓN QR

Para una matriz genérica A , $n \times n$, no singular, las propiedades que se acaban de describir permiten construir una sucesión de $n - 1$ matrices de Householder tales que

$$H_{n-1} \dots H_2 H_1 A = R, \quad (\text{III.11})$$

donde R es una matriz $n \times n$ no singular y triangular superior.

El primer paso de este proceso es construir una matriz de Householder H_1 que transforma a_1 (la primera columna de A) en un múltiplo de e_1 , es decir, se desean crear ceros en las componentes 2 hasta n del vector a_1 . La norma euclídea se conserva bajo transformaciones ortogonales, de manera que

$$\mathbf{H}_1 \mathbf{a}_1 = (\mathbf{I} - \frac{\mathbf{u}_1 \mathbf{u}_1^t}{\beta_1}) \mathbf{a}_1 = \pm \|\mathbf{a}_1\|_2 \mathbf{e}_1 = \begin{pmatrix} r_{11} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \quad (\text{III.12})$$

donde $|r_{11}| = \|\mathbf{a}_1\|_2$. De la expresión (III.9) sabemos que el vector \mathbf{u}_1 tiene que ser un múltiplo de $\|\mathbf{a}_1\|_2 \mathbf{e}_1 - \mathbf{a}_1$, y dado que \mathbf{H}_1 depende sólo de la dirección de \mathbf{u}_1 , podemos elegir el vector \mathbf{u}_1 como

$$\mathbf{u}_1 = \begin{pmatrix} \mathbf{a}_{11} - r_{11} \\ \mathbf{a}_{21} \\ \vdots \\ \mathbf{a}_{n1} \end{pmatrix}. \quad (\text{III.13})$$

Al definir \mathbf{u}_1 , el signo de r_{11} se puede elegir positivo o negativo (excepto cuando \mathbf{a}_1 es ya un múltiplo de \mathbf{e}_1), y para evitar el problema de la cancelación de términos parecidos, usualmente se escoge el signo opuesto al signo de \mathbf{a}_{11} , de manera que

$$\text{signo}(r_{11}) = -\text{signo}(\mathbf{a}_{11}). \quad (\text{III.14})$$

Después de la primera aplicación de las transformaciones de Householder, la primera columna de la matriz parcialmente reducida $\mathbf{A}^{(2)} = \mathbf{H}_1 \mathbf{A}$ es un múltiplo de \mathbf{e}_1 , y los demás elementos han sido alterados

$$\mathbf{A}^{(2)} = \mathbf{H}_1 \mathbf{A} = \begin{pmatrix} r_{11} & \mathbf{a}_{12}^{(2)} & \cdots & \mathbf{a}_{1n}^{(2)} \\ \mathbf{0} & \mathbf{a}_{22}^{(2)} & \cdots & \mathbf{a}_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{a}_{n2}^{(2)} & \cdots & \mathbf{a}_{nn}^{(2)} \end{pmatrix}. \quad (\text{III.15})$$

Es muy importante notar que a diferencia de la eliminación Gaussiana, la primera fila de la matriz \mathbf{A} viene modificada por efecto de la transformación de Householder \mathbf{H} .

Al construir la segunda transformación de Householder, el objetivo principal es reducir la segunda columna a la forma correcta, sin alterar la primera fila y la primera columna de la matriz parcialmente reducida. Debido a la propiedad (III.10), se puede obtener este resultado definiendo el segundo vector de Householder \mathbf{u}_2 con la primera componente nula. Habiendo escogido así el vector \mathbf{u}_2 , la aplicación de la matriz de

Householder \mathbf{H}_2 a un vector genérico no cambia la primera componente, y la aplicación a un múltiplo de \mathbf{e}_1 deja el vector entero como está.

Si \mathbf{A} es una matriz no singular, se pueden efectuar $n-1$ pasos de reducción de Householder, para obtener $\mathbf{H}_{n-1} \dots \mathbf{H}_2 \mathbf{H}_1 \mathbf{A} = \mathbf{R}$, con \mathbf{R} una matriz triangular superior no singular. Si se denota con \mathbf{Q}^t la matriz ortogonal $n \times n$

$$\mathbf{Q}^t = \mathbf{H}_{n-1} \dots \mathbf{H}_2 \mathbf{H}_1 \implies \mathbf{Q} = \mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_{n-1}. \quad (\text{III.16})$$

Cualquiera de las dos formas

$$\mathbf{Q}^t \mathbf{A} = \mathbf{R} \quad \text{o} \quad \mathbf{A} = \mathbf{Q} \mathbf{R}. \quad (\text{III.17})$$

se conoce como la factorización QR de la matriz \mathbf{A} . Una vez conocida la factorización QR de \mathbf{A} , ecuación (III.17), la solución al sistema $\mathbf{A} \mathbf{x} = \mathbf{b}$ se obtiene resolviendo el sistema triangular superior $\mathbf{R} \mathbf{x} = \mathbf{Q}^t \mathbf{b}$. En general, es necesario un intercambio de columnas para asegurar que el rango está plenamente revelado. Sin el intercambio de columnas la reducción de Householder terminaría inmediatamente con una matriz no nula cuya primera columna es cero. Si las demás columnas son también ceros, la reducción termina. De otra manera, existe por lo menos una columna no nula, la columna pivote, que se puede elegir como candidata para la siguiente reducción. Como en la eliminación gaussiana, una estrategia de pivoteo pide que se escoja la columna pivote como la primera columna de norma máxima (otra posibilidad es escoger la columna “menos reducida”).

En general, si \mathbf{A} es $m \times n$ de rango r , se necesitarán r permutaciones \mathbf{P}_k y r matrices de Householder \mathbf{H}_k , $k = 1, \dots, r$. Después de estos r pasos, la configuración final será

$$\mathbf{H}_r \dots \mathbf{H}_1 \mathbf{A} \mathbf{P}_1 \dots \mathbf{P}_r = \tilde{\mathbf{R}}, \quad (\text{III.18})$$

donde

$$\tilde{\mathbf{R}} = \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (\text{III.19})$$

es triangular superior que nos revela el rango, y \mathbf{R}_{11} es una matriz $r \times r$ no singular triangular superior. Con una correcta estrategia de pivoteo, dependiente

del problema original, y aritmética exacta, este procedimiento de Householder terminará después de r pasos, cuando la matriz restante se transformará en cero. Combinando los intercambios de columnas en una sola matriz de permutación P y las transformaciones de Householder en una sola matriz ortogonal Q^t , se obtiene

$$Q^t A P = \tilde{R}, \quad (\text{III.20})$$

o equivalentemente,

$$A P = Q \tilde{R}, \quad (\text{III.21})$$

y

$$Q^t A = \tilde{R} P^t = \begin{pmatrix} R \\ 0 \end{pmatrix} P^t = \begin{pmatrix} R P^t \\ 0 \end{pmatrix}. \quad (\text{III.22})$$

Las filas de la matriz $R P^t$ son linealmente independientes, de manera que la matriz transformada $Q^t A$ tiene la forma deseada (III.1), con $F = R P^t$. El problema de mínimos cuadrados se puede entonces resolver con el siguiente algoritmo:

- (I) Calcular la factorización QR de la matriz A , usando la reducción de Householder (que determina el rango r , y las matrices P , Q y R);
- (II) Formar el vector $d = Q^t b$, y denotar con d_r las primeras r componentes de d ;
- (III) Calcular cualquier solución y del sistema $R_y = d_r$;
- (IV) Formar $x = P y$.

El número de operaciones requerido para resolver el problema de mínimos cuadrados de esta manera es del orden de $2 \cdot m \cdot n \cdot r - r^2(m + n) + \frac{2}{3}n^3$.

Si la matriz A posee columnas linealmente independientes ($r = n$), la matriz R es no singular, la solución de $R y = d_r$ es única, y la solución del problema de mínimos cuadrados es única. En este caso resolver el problema con el método QR es aproximadamente dos veces más costoso que resolviendolo con las ecuaciones normales. Cuando las columnas de A son linealmente dependientes, de manera que $r < n$, el sistema

$r \times n \ R y = d_r$ posee un número infinito de soluciones. Dado que R es de la forma $R = (R_{11} R_{12})$, con R_{11} triangular superior no singular, es posible escoger un vector $y = (y_B, 0)^t$ tal que $R y = d_r$ y con la propiedad de que $R_{11} y_B = d_r$. Si y posee esta forma, la solución $x = P y$ es sencillamente una reordenación de y , y es llamada una solución básica del problema de mínimos cuadrados.

3. Análisis

Ahora que hemos explicado como implementaremos el algoritmo para realizar una regresión lineal usando las transformaciones de Householder y la factorización QR veamos los datos que hemos obtenido del INEI

	Pobreza	
	2015	2016
Amazonas	37.2	36.4
Áncash	18.2	15.7
Apurímac	13.9	18.0
Arequipa	11.3	11.1
Ayacucho	24.9	23.3
Cajamarca	25.0	23.9
Callao	8.8	7.0
Cusco	15.9	18.1
Huancavelica	27.5	24.8
Huánuco	26.1	27.4
Ica	12.5	11.8
Junín	26.9	26.8
La Libertad	14.7	13.6
Lambayeque	15.4	11.4
Lima	9.2	9.2
Provincia de Lima 1/	8.4	8.4
Región Lima 2/	17.6	17.6
Loreto	58.7	57.5
Madre de Dios	30.6	29.1
San Martín	41.7	38.3

Figura 1: Datos de pobreza en Perú en 2015 y 2016

Realizamos el diagrama de dispersión para la pobreza en 2015 vs pobreza en 2016 y obtenemos los siguientes puntos:

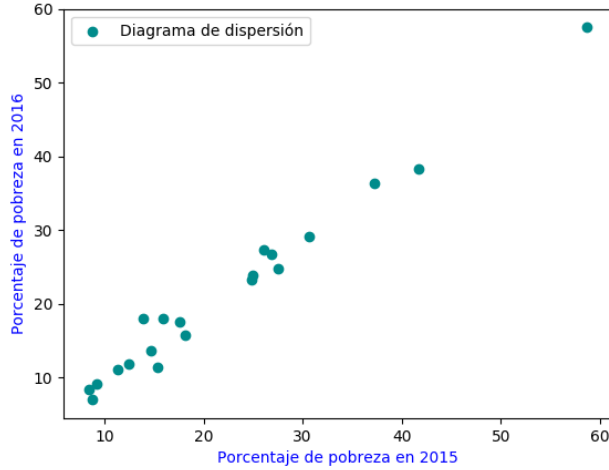


Figura 2: Diagrama de dispersión

Observamos que los datos pueden tener una relación lineal para ello debemos obtener una ecuación de la forma:

$$\hat{y} = c_0 + c_1 x. \text{ (IV.1)}$$

De la ecuación (IV.1) los x_m son los porcentajes de la pobreza en 2015 y los y_m son los porcentajes de la pobreza en 2016.

Utilizando el algoritmo en Python obtenemos la siguiente ecuación:

$$\hat{y} = 0,114103064 + 0,9608952x.$$

Graficando la ecuación obtenida:

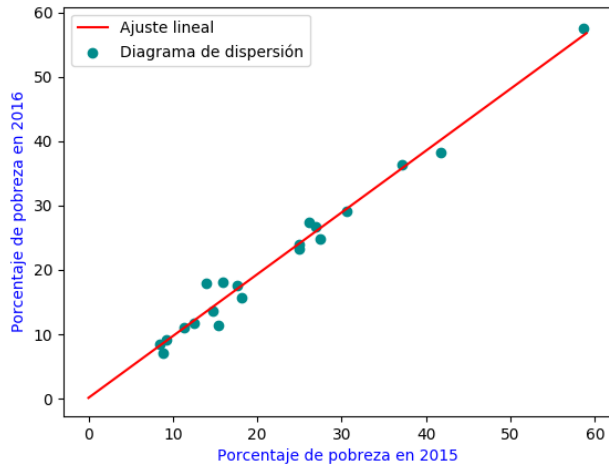


Figura 3: Ajuste lineal

Ahora analizaremos el resultado de nuestro algoritmo:

3.1. Calidad de la solución

Para comparar nuestro resultado usaremos una librería llamada sklearn de python que nos da una ecuación lineal.

$$y = 0,11410127044526774 + x0,96089533$$

Tenemos que el error relativo (E_r) esta en el siguiente intervalo:

$$\frac{\|R\|}{\|b\|} \frac{1}{Cond(A)} \leq E_r \leq Cond(A) \frac{\|R\|}{\|b\|}$$

Donde:

$$R = A\tilde{x} - b$$

Realizando los calculos en python tenemos que:

$$3.44498688e-8 \leq E_r \leq 9.40002724e-5$$

Con este resultado podemos decir que la solución es confiable.

3.2. Análisis de regresión

Con el fin de determinar la pertinencia de la ecuación de regresión hallada, es necesario hacer un análisis de la bondad de ajuste de la recta, demostrar si la relación es estadísticamente significativa. Calcularemos las siguientes variables estadísticas usando Python:

3.2.1. El coeficiente de determinación

Para la i -ésima observación de la muestra, la desviación entre el valor observado de la variable dependiente y i y el valor estimado de la variable dependiente, se llama i -ésimo residual. Representa el error que se comete al usar para estimar y i . También se le conoce como la suma de los cuadrados debidos al error (SSE).

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

$$SSE = 61,406350402$$

El valor de SSE es una medida del error que se comete al usar la ecuación de regresión para calcular los valores de la variable dependiente en la muestra. Otro valor de importancia es la medida del error incurrido al usar para estimar y_i , llamado suma total de cuadrados (SST):

$$SST = \sum (y_i - \tilde{y})^2$$

$$SST = 2913,262$$

Para saber cuánto se desvían los valores de \hat{y}_i medidos en la línea de regresión, de los valores de \tilde{y} , se calcula otra suma de cuadrados. A esa suma se le llama suma de cuadrados debida a la regresión, y se representa por SSR.

$$SSR = \sum (\hat{y}_i - \tilde{y})^2 \quad SSR = 2851,854780525$$

La relación entre las medidas es

$$SST = SSR + SSE$$

La relación SSR/SST , se denomina coeficiente de determinación y se representa por r^2 .

$$r^2 = SSR/SST$$

$$r^2 = 0,978921503$$

Se puede interpretar a r^2 como el porcentaje de la variación de los valores de la variable dependiente que se puede explicar con la ecuación de regresión.

3.2.2. Coeficiente de correlación

Es la segunda medida que se usa para describir qué tan bien explica una variable a la otra. El coeficiente de correlación de la muestra se denota por r y es la raíz cuadrada del coeficiente de determinación:

$$r = (\text{signo de } c_1) \sqrt{r^2}$$

$$r = 0,98940462$$

Muestra un coeficiente de correlación muy alto lo que implica una relación de dependencia lineal fuerte entre los valores de pobreza del 2015 y 2016. Los coeficientes de determinación y correlación no son suficientes para llegar a la conclusión acerca de si la relación es estadísticamente significativa. Necesitamos realizar una prueba de significancia.

3.2.3. Prueba de significancia

La ecuación de regresión lineal simple indica que el valor medio esperado de y es una función lineal de x :

$$E(y) = \beta_0 + \beta_1 x$$

Si $\beta_1 = 0$, entonces $E(y) = \beta_0$. En este caso el valor medio de y no depende del valor de x y se concluye que no existe relación lineal entre las variables. En forma análoga, si el valor de β_1 no es igual a cero, se concluye que las dos variables se relacionan. Así, para probar si hay alguna relación importante de regresión debemos efectuar una prueba de hipótesis para determinar si el valor de β es cero. Usaremos la prueba t de Student pero necesitaremos la varianza del error en el modelo de regresión.

Estimando σ^2 :

La varianza de ϵ , también representa la varianza de los valores de respecto a la línea de regresión. Así, la suma de los residuales al cuadrado, SSE , es una medida de la variabilidad de las observaciones reales respecto a la línea de regresión. Cada suma de cuadrados tiene asociado un número que llamamos grados de libertad. Se ha demostrado que SSE tiene $n-2$ grados de libertad, porque se deben estimar dos parámetros β_0 y β_1 . El error cuadrado medio (s^2) es el estimado de σ^2 . Se calcula mediante la ecuación:

$$s^2 = \frac{SSE}{n-2} \text{ en nuestro caso } n-2 = 18$$

$$3,411463911$$

Desviación estándar de la estimación

El error típico o desviación estándar del estimado se calcula como la raíz cuadrada de la varianza del estimado.

$$s = \sqrt{\frac{SSE}{n-2}}$$

$$1,847014865$$

3.2.4. Realizando la prueba t

Distribución muestral de c_1 :

Valor esperado : $E(c_1) = \beta_1$

Desviación estándar : $\sigma_{c_1} = \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}}$

Forma de la Distribución: *Normal*

Como no conocemos el valor de σ obtenemos una estimación de σ_{c_1} , que denotaremos como s_{c_1} .

$$s_{c_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}, s_{c_1} = 0,033234007$$

Para determinar si la relación es significativa se basa en el hecho estadístico de prueba:

$$\frac{c_1 - \beta_1}{s_{c_1}}$$

sigue una distribución t con $n - 2$ grados de libertad. Si la hipótesis nula es verdadera, entonces β_0 y $t = c_1/s_{c_1}$. Empleando como nivel de significancia $\alpha = 0,05$. $t = \frac{c_1}{s_{c_1}} = 28,913010119$

En las tablas de distribución t se encuentra que para $n - 2 = 18$ grados de libertad, $t = 2,1009$ da un área de **0,025** en la cola superior. Por lo tanto, el área en la cola superior de la distribución t correspondiente al valor estadístico de prueba $t = 28,913010119$ debe ser menor a **0,025**. Como es una prueba de dos colas, este valor sería $2 \cdot (0,025) = 0,01$ y se concluye que el valor- p para $t = 28,913010119$ debe ser menor a **0,01**. Calculando el valor- p con 9 cifras significativas nos da **0**. Se concluye que β_1 no es igual a **0**. Esto es suficiente evidencia para concluir que son existe una relación significativa entre la pobreza del 2015 con la pobreza en 2016.

3.3. Uso de la ecuación de regresión para estimar y predecir

Como comprobamos que existe una relación estadísticamente significativa podemos usar esa ecuación para estimar y predecir. Estimación de intervalo. Con ese fin, se determinan estimaciones de intervalo. El primer tipo de estimado es el de intervalo de confianza, que es un estimado del valor medio de y para determinado valor de x . El segundo tipo es el estimado de intervalo de predicción, que se usa cuando deseamos un estimado de intervalo de valor individual de y que corresponda a determinado valor de x .

3.3.1. Estimado del intervalo de confianza del valor medio de y

Al estimar el porcentaje promedio de pobreza en 2016 de todas las ciudades que en **2015** mostraron un

índice de pobreza de **10,6 %**. El estimado de $E(y_p)$, el valor medio desconocido, es:

$$\hat{y}_p = 0,114103064 + 0,9608952(10,6) = 10,299591979.$$

Donde \hat{y}_p es el estimado del valor particular de y . Dado que no se puede esperar que \hat{y}_p sea exactamente igual a $E(y_p)$. Entonces es necesario considerar la varianza de los estimados basados en la ecuación de regresión. La fórmula para estimar la desviación estándar de \hat{y}_p dado un valor particular de x, x_p es:

$$S_{\hat{y}_p} = s \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

$$S_{\hat{y}_p} = 0,582055397$$

La ecuación general para un estimado del intervalo de confianza de $E(y_p)$ dado un valor particular x de es:

$$\hat{y}_p \pm t_{\alpha/2} \cdot S_{\hat{y}_p}.$$

En donde el coeficiente de confianza es $1 - \alpha$ y $t_{\alpha/2}$ se basa en una distribución t con $n - 2$ grados de libertad.

Para determinar un estimado de intervalo de confianza de **95 %** para el porcentaje promedio de pobreza en 2016 de todas las ciudades que en 2015 mostraron un índice de pobreza de **10,6 %**, necesitamos el valor de t para $\alpha/2 = 0,025$ y $n-2 = 18$ grados de libertad. Así, con $\hat{y}_p = 22,5$ $t_{0,025} = 2,1009$ $s_{\hat{y}_p} = 0,582055397$, tenemos:

$$10,6 \pm 2,1009 \cdot 0,582055397.$$

$$10,6 \pm 1,222840184$$

Entonces, con una confianza del **95 %** se puede decir que el porcentaje promedio de pobreza en 2016 de todas las ciudades que en 2015 mostraron un índice de pobreza de **10,6 %** está entre **9,377159816 %** y **11,822840184 %**.

3.3.2. Estimado del intervalo de predicción para un valor particular de y

Para este análisis, se supone que en vez de estimar el valor medio del porcentaje de pobreza, deseamos

estimar el porcentaje de pobreza en 2016 para la región Tacna con un índice de pobreza de **10,6 %** en 2015.

El estimado para ese valor particular por medio de la ecuación de regresión es:

$$\hat{y}_p = 0,114103064 + 0,9608952(10,6) = 10,299591979.$$

Para determinar un estimado del intervalo de predicción debemos determinar primero la varianza asociada al empleo de \hat{y}_p como estimado de un valor individual de y . Esta varianza está formada por la suma de dos componentes:

- 1) La varianza de los valores individuales de y respecto del promedio cuyo estimado es s^2 .
- 2) La varianza asociada al uso de \hat{y}_p para estimar $E(y_p)$ cuyo estimado es $s_{\hat{y}_p}$.

Así, el estimado de la varianza de un valor individual es:

$$s_{ind}^2 = s^2 + s_{\hat{y}_p}.$$

Por consiguiente, un estimado de la desviación estándar de un valor individual de \hat{y}_p es:

$$S_{ind} = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

$$S_{ind} = 1,93655684$$

La ecuación general para un estimado del intervalo de predicción para un valor individual de y dado un valor particular de x es:

$$\hat{y}_p \pm t_{\alpha/2} \cdot S_{ind}.$$

Se tiene que:

$$10,6 \pm 2,1009 \cdot 1,93655684 .$$

$$10,6 \pm 4,068512265$$

Entonces, con una confianza del **95 %** se puede decir que el porcentaje de pobreza en 2016 de la región Tacna que en 2015 tenía un porcentaje de pobreza de **10,6 %** está entre **6,531487735 %** y **14,668512265 %**.

3.3.3. Estimación de los parámetros del modelo de regresión lineal

Uno de los conceptos fundamentales sobre el que se ha basado este análisis, es que la ecuación de regresión lineal obtenida a partir de los datos de la muestra es un estimado de los parámetros del modelo para la población. Por lo tanto, es posible determinar intervalos de confianza para los coeficientes de la ecuación de regresión:

$$\beta_0 = c_0 \pm t_{\alpha/2} \cdot s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}.$$

$$\beta_1 = c_1 \pm t_{\alpha/2} \frac{s}{\sum (x_i - \bar{x})^2}$$

De estas fórmulas:

$$\beta_0 = 0,114103064 \pm 1,535177749 .$$

$$\beta_1 = 0,960895200 \pm 0,069821325$$

Con esta información encontramos que la tasa de incremento de la pobreza en 2016 está entre **0,891073675 %** y **1,030716325 %** por cada **1 %** de incremento de la pobreza en 2015 entre una región y otra.

4. Observaciones

1. Si incrementamos la cantidad de valores de análisis, es necesario paralelizar las transformaciones de Householder y para una paralelización más sencilla se necesita un nuevo concepto: Rotaciones de Givens.
2. Es importante mencionar que los datos encontrados no son actuales por lo que tal vez no describa perfectamente la tendencia en la actualidad.
3. Se puede cometer un error al utilizar el análisis de regresión, y es suponer que un cambio en una variable es “ocasionado” por un cambio en la otra variable. Los análisis de regresión y correlación no pueden, de ninguna manera, determinar la causa y el efecto.

5. Conclusiones

1. Desde el punto de vista computacional resolver el problema de mínimos cuadrados con las ecuaciones normales es eficiente, pero esafortunadamente el uso de estas presenta un problema de estabilidad numérica.
2. La factorización QR usando transformaciones de Householder es un gran método para realizar ajustes lineales ya que el valor de \mathbf{E}_r es bastante pequeño.
3. El análisis de regresión lineal, como parte de la inferencia estadística, es fundamental para determinar relaciones de dependencia lineal entre variables.
4. Usando la regresión lineal podemos realizar estimaciones y predicciones dentro de un intervalo de confianza deseado.
5. Con esta herramienta que describe la relación entre dos variables podemos hacer ajustes en los procesos, tomar decisiones o establecer políticas.

Referencias

- [1] L. Héctor Juárez V. *Análisis Numérico*, Setiembre 2008
- [2] E. Ward Cheney, David Kincaid *Numerical Mathematics and Computing, Seventh Edition*, Abril 2012
- [3] Robert Johansson *Numerical Python, Second Edition*, Diciembre 2018
- [4] Wes McKinney *Python for Data Analysis, Second Edition*, Noviembre 2017
- [5] David R. Anderson, Dennis J. Sweeney, Thomas A. Williams *Statistics for Business and Economics, Eleventh Edition*, 2010
- [6] Instituto Nacional de Estadística e Informática (INEI) *Encuesta Nacional de Hogares*. 2017