ECONOMETRÍA
APLICADA UTILIZANDO R.

PAPIME PE302513 Libro electrónico y complementos didácticos en medios computacionales, para el fortalecimiento de la enseñanza de la econometría

Anexo.

Repaso básico de estadística, probabilidad y álgebra lineal en R

Luis Quintana Romero

Objetivo

Permitir al alumno recordar una serie de herramientas y conceptos básicos de estadística, probabilidad y álgebra lineal aprendidos en los cursos de matemáticas básicos tomados previamente. Además, se brindará la oportunidad de practicar un poco, antes de entrar de lleno al estudio de la econometría.

Introducción

Dentro de este apartado se brindan las bases y herramientas básicas de estadística, probabilidad y álgebra lineal, este breve repaso se encuentra dentro de tres apartados 1) Revisión de datos, 2) Probabilidad y 3) Álgebra matricial dentro de ellos se encuentran algunos métodos y formulas que posteriormente serán de utilidad en el uso de la econometría.

1) Revisión de datos.

Media aritmetica: es la medida de localización más usual de una muestra.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Donde:

Xi= valores muestrales y n= número de observaciones muestrales

Moda y mediana:

La moda es el valor con la mayor frecuencia en los datos, mientras que la mediana es el valor medio de un conjunto ordenado de datos.

Histogramas:

Presentan los porcentajes de observaciones que quedan dentro de un intervalo en particular.

Varianza muestral (S^2)

Mide la dispersión pero eleva al cuadrado las unidades de medida originales, lo cual resulta difícil de interpretar. Así que para interpretar, en las unidades de medida originales, es mejor considerar la desviación estándar.

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

Desviación estándar (S):

Es la raíz cuadara de la varianza.

Coeficiente de variación (CV)

Se calcula dividiendo la desviación estándar entre la media.

$$CV = \frac{S}{\bar{X}} \times 100$$

Algunas mediciones de la forma de las distribuciones de probabilidad de nuestros datos, aportan información sobre la asimetría de su distribución. Las medidas de forma con las que usualmente vamos a trabajar son el sesgo y la curtosis.

El sesgo (SK):

Mide la simetría de una distribución de frecuencias de nuestros datos con relación a la media

SESGO $$SK = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{X_i-\bar{X}}{S}\right]^3$$

Curtosis (KS):

Mide el achatamiento o agudeza en relación con la forma que presenta una distribución normal de los datos.

CURTOSIS$KS = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{X_i - \bar{X}}{S}\right]^4$

Para suavizar la distribución de los datos se utilizan las funciones de densidad de Kernel, cuyo estimador se define a continuación (Everitt y Horthorn, 2006):

$$\hat{f}(x) = \frac{1}{hn}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - x_i}{h}\right)$$

Donde K es una función kernel y h es la amplitud de banda o parámetro de suavizamiento.

Las funciones kernel más usuales son las siguientes:

a) Rectangular

b) Triangular

c) Gaussiano

2) Probabilidad

La teoría de la probabilidad es fundamental para realizar inferencias de la población a partir de datos muestrales:

$$\Pr(\cup A_i) = \sum \Pr(A_i) \quad si\ A_i \cap A_j = \phi \quad \text{donde } \phi \text{es el evento imposible}$$

Teoremas:

1) Si el evento A es un subconjunto del evento B entonces la $\Pr(A)$ es menor o igual a la Pr(B): $A \subset B \rightarrow \Pr(A) \leq \Pr(B)$.

2) Para cualquier evento A la probabilidad de su complemento, $\Pr(A^c)$, es igual a: $1 - \Pr(A)$.

3) La probabilidad del evento imposible es cero: $\Pr(\phi) = 0$.

4) La probabilidad de la unión de dos eventos es:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

La probabilidad condicional se define de la siguiente manera:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Variables aleatorias:

Son variables de las cuales su resultado nunca es conocido de antemano.

Formalmente una variable aleatoria es una función medible que vincula al conjunto de resultados con el conjunto de los números reales.

Variable aleatoria discreta es aquella que puede tomar sólo un número finito de valores o bien infinito pero que pueden ser contados.

Variable aleatoria continua puede tomar un número infinito de valores en cualquier intervalo.

Las probabilidades asociadas a una variable aleatoria se calculan a través de su función de distribución probabilística.

Para variables aleatorias continuas no es relevante calcular la probabilidad de que X tome un valor particular x debido a que en un intervalo a,b existe un número infinito de puntos. Por ello, la cuestión relevante es calcular la probabilidad de que X tome valores en el intervalo a,b donde a<b;

$$\Pr(a \leq x \leq b) = \int_a^b f(x)dx$$

La función de densidad acumulada de una variable aleatoria continua se define como:

$$\Pr(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Para una variable aleatoria discreta, su media o valor esperado es un promedio ponderado de sus posibles resultados, en donde el ponderador es la probabilidad asociada a cada valor, tal y como se indica a continuación:

$$\mu_x = E(X) = pr_1x_1 + pr_2x_2 + \cdots + pr_nx_n = \sum pr_ix_i$$

El operador de esperanza matemática puede utilizarse para describir las funciones de distribución conjuntas.

La covarianza:

Mide la asociación lineal entre las dos variables

Por ejemplo, la covarianza entre X y Y estará dada por la siguiente expresión:

$Cov(X,Y)=E(X-E(X))(Y-E(Y))$

El problema que tiene la covarianza como medida de asociación lineal es que depende de las unidades de medida de X y Y, para evitar ese problema se normaliza dividiéndola entre las desviaciones estándar de las dos variables. A está covarianza, liberada de unidad de medida, se le conoce como el **coeficiente de correlación de Pearson** y se denota de la siguiente forma:

$\rho(X,Y)=Cov(X,Y)/ \sigma_x\sigma_y$

El coeficiente de correlación tiene un rango que va de -1 a +1.

Distribución Binaria

La variable aleatoria X toma únicamente valores 0 y 1 con probabilidades p y 1-p, a estas variables se les denomina variables aleatorias Bernoulli. Donde P(X=1)=p y la P(X=0)=1-p

Por lo tanto, la función de probabilidad es:

$f(x)=P(X=x)=p^x(1-p)^{1-x}$

Un importante grupo de modelos econométricos utilizan variables dependientes cualitativas que son codificadas como uno y cero, estos modelos se conocen como logit, probit y tobit.

Distribución Binomial

La variable aleatoria mide el número de éxitos en n experimentos de éxito-fracaso, iguales e independientes.

Su función de densidad probabilística es definida con la siguiente expresión:

$$f(x) = P(X = x) = C_x^n p^x (1-p)^{n-x}$$

Distribución Poisson

La variable aleatoria toma una serie de valores en un tiempo o espacio en donde el número de valores es igual a cualquier entero entre cero e infinito. La muestra, n, es grande tal que np=λ y la probabilidad de acierto, p, es pequeña:

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Donde λ es un número dado

Distribución Normal

La distribución Normal es la distribución más usual que se emplea en estadística. La densidad de probabilidad de esa función está dada por la fórmula siguiente:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

donde:

σ es la desviación estándar, μ es la media, π es 3.1416 y e es el número 2.71828

Una variable aleatoria es estandarizada si se le resta su media y se divide entre su desviación estándar, lo que da como resultado una variable con media cero y varianza igual a la unidad.

En el análisis econométrico se utilizan, con gran frecuencia, tres distribuciones derivadas de la normal:

a) Distribución ji cuadrada

b) Distribución t de student

c) La Distribució F de Fisher

3) Álgebra matricial

Una matriz no es más que un arreglo rectangular de elementos o números en renglones o columnas.

$$\mathbf{B} = \begin{bmatrix} 50 & 80 & 100 \\ 30 & 40 & 50 \\ 60 & 50 & 80 \end{bmatrix}$$

De forma general, podemos representar una matriz en un cuadro de entradas y salidas como el siguiente:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

La matriz **A** es de dimensión mxn, dim(**A**)=(mxn) es decir tiene m por n elementos arreglados en m renglones y n columnas.

La matriz **A** transpuesta se representa por **A’** y se obtiene permutando las filas por las columnas de **A**.

Por ejemplo, suponiendo una matriz **A** particular como la siguiente:

$$\mathbf{A} = \begin{bmatrix} 10 & 5 & 8 \\ 20 & 10 & 30 \\ 30 & 40 & 1 \end{bmatrix}$$

La matriz transpuesta para A es:

$$\mathbf{A}' = \begin{bmatrix} 10 & 20 & 30 \\ 5 & 10 & 40 \\ 8 & 30 & 1 \end{bmatrix}$$

Las matrices que tienen las mismas dimensiones se pueden realizar las operaciones básicas de suma y resta. Para que dos matrices sean iguales no solamente deben ser de la misma dimensión sino además deben ser iguales elemento a elemento, en este caso cada elemento a_{ij} en la matriz **A'** debiera ser igual a cada elemento b_{ij} en la matriz **B** para todo i, j.

Suma de matrices: se debe sumar elemento a elemento; si **A'** y **B** tienen el mismo orden, **A'**+**B** es una nueva matriz **C** del mismo orden en donde:

$$\mathbf{A}' + \mathbf{B} = \mathbf{C} = \begin{bmatrix} 10 & 5 & 8 \\ 20 & 10 & 30 \\ 30 & 40 & 1 \end{bmatrix} + \begin{bmatrix} 50 & 80 & 100 \\ 30 & 40 & 50 \\ 60 & 50 & 80 \end{bmatrix} = \begin{bmatrix} 60 & 85 & 108 \\ 50 & 50 & 80 \\ 90 & 90 & 81 \end{bmatrix}$$

La resta o sustracción entre dos matrices **A** y **B** requiere al igual que la suma de la conformabilidad de las matrices, y el resultado es: **C**=aij-bij.

$$\mathbf{A}' - \mathbf{B} = \mathbf{C} = \begin{bmatrix} 10 & 5 & 8 \\ 20 & 10 & 30 \\ 30 & 40 & 1 \end{bmatrix} - \begin{bmatrix} 50 & 80 & 100 \\ 30 & 40 & 50 \\ 60 & 50 & 80 \end{bmatrix} = \begin{bmatrix} -40 & -75 & -92 \\ -10 & -30 & -20 \\ -30 & -10 & -79 \end{bmatrix}$$

Multiplicación escalar, si se multiplica a la matriz A' por el escalar z el resultado es: z**A'**= za_{ij}. Por ejemplo, para duplicar la matriz A' ésta debe multiplicarse por el escalar z=2, tal y como se indica a continuación:

$$\mathbf{z}\mathbf{A}' = 2\mathbf{A}' = \begin{bmatrix} 20 & 10 & 16 \\ 40 & 20 & 60 \\ 60 & 80 & 2 \end{bmatrix}$$

Vectores:

Se constituye de una matriz de dimensión 1xn y se representa como:

$$\mathbf{v} = [v_{11} \quad v_{12} \quad \cdots \quad v_{1n}]$$

Vector columna: vector de dimensión nx1, por ejemplo el siguiente vector **c**:

$$\mathbf{c} = \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{n1} \end{bmatrix}$$

Para cualquier par de vectores **v** y **c** el producto interno de vectores es:

$$\mathbf{vc} = [v_{11} \quad v_{12} \quad \cdots \quad v_{1n}] \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{n1} \end{bmatrix} = v_{11}c_{11} + v_{12}c_{21} + \cdots + v_{1n}c_{n1} = \sum v_{1k}c_{k1}$$

Ejemplo de multiplicación de vectores:

$$\mathbf{vc} = [1 \quad 4 \quad 6] \begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix} = [2 + 20 + 60] = [82]$$

La multiplicación matricial:

Es una generalización de la multiplicación vectorial repetida. Por ejemplo, el producto de las matrices **A'** y **B** está dado por:

$$\mathbf{A}'\mathbf{B} = \begin{bmatrix} 10 & 5 & 8 \\ 20 & 10 & 30 \\ 30 & 40 & 1 \end{bmatrix} \begin{bmatrix} 50 & 80 & 100 \\ 30 & 40 & 50 \\ 60 & 50 & 80 \end{bmatrix} = \begin{bmatrix} 1130 & 1400 & 1890 \\ 3100 & 3500 & 4900 \\ 2760 & 4050 & 5080 \end{bmatrix}$$

El primer elemento de la matriz final anterior se obtiene por producto interno de vectores de la manera siguiente:

$$\mathbf{a_1 b_1} = \begin{bmatrix} 10 & 5 & 8 \end{bmatrix} \begin{bmatrix} 50 \\ 30 \\ 60 \end{bmatrix} = [500 + 150 + 480] = [1130]$$

De la multiplicación matricial se puede deducir:

Si una matriz **A** es de dimensión mxn y una matriz **B** es de orden nxp, el producto **AB** es una matriz de orden mxp.

Para poder realizar la multiplicación debe cumplirse que la dimensión columna de la primer matriz sea igual a la dimensión fila de la segunda matriz.

Existe un tipo de matrices con la peculiar característica que multiplicada por sí misma, cualquier número de veces, es siempre igual a la matriz original, a estas matrices se les llama **idempotentes**:

$\mathbf{A} = \mathbf{A}^2 = \mathbf{A}^3 \ldots \mathbf{A}^n$

Conjunto de propiedades matriciales:

1) La suma de matrices es conmutativa: **A+B=B+A**

2) La multiplicación de matrices en general no es conmutativa. En particular podemos decir que una matriz cuadrada conmuta consigo misma y con la matriz identidad: **AB≠BA AA=AA AI=IA**

La matriz identidad en su diagonal principal tiene únicamente elementos escalares iguales a la unidad y ceros fuera de ella. Dado que es una matriz cuadrada se suele representar como $\mathbf{I_n}$ cuando es de dimensión nxn o simplemente como **I** cuando no hay confusión acerca de su dimensión:

$$\mathbf{I_n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Propiedades matriciales:

La suma y la multiplicación de matrices cumplen con la propiedad asociativa:

Suma: **(A+B)+C=A+(B+C)**

Multiplicación: **(AB)C=A(BC)**

La multiplicación y la multiplicación escalar cumplen con la propiedad distributiva:

Multiplicación: **A(B+C)=AB+AC**

Multiplicación escalar: **z (A+B)= zA+zB**

También las matrices transpuestas cumplen con ciertas propiedades que conviene conocer:

La transpuesta de una matriz transpuesta es igual a la matriz original: **(A')'=A**

La transpuesta de una suma de matrices es igual a la suma de sus transpuestas: **(A+B)'=A'+B'**

La transpuesta del producto de matrices es igual al producto inverso de sus transpuestas: **(AB)'=B'A'** o bien **(ABC)=C'B'A'**

El determinante de **A** es una función de los valores aij de la matriz:

$$\det(\mathbf{A}) = |\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$$

Para cualquier matriz cuadrada, independientemente de su orden, el determinante se obtiene como:

$C_{ij}=(-1)^{i+j} \det(\mathbf{M}_{ij})$

En donde C es llamado el cofactor del elemento (i,j) y **M** es el menor o la submatriz que se obtiene al eliminar la fila i y la columna j de la matriz **A**.

Entonces podemos reescribir el determinante de nuestra matriz **A** de orden 2 como:

$$\det(\mathbf{A}) = |\mathbf{A}| = a_{11}\mathrm{C}_{11} - a_{12}\mathrm{C}_{12}$$

Ahora estamos en condiciones de calcular la inversa $\mathbf{A}^{-1}$ de una matriz **A** cuadrada y no singular, que es la matriz única que cumple con la relación:

$\mathbf{A}\mathbf{A}^{-1}= \mathbf{I} =\mathbf{A}^{-1}\mathbf{A}$

La matriz inversa juega la misma función que el reciproco en el álgebra ordinaria.

La inversa la obtenemos con la siguiente fórmula:

$$\mathbf{A}^{-1} = \frac{1}{|\mathrm{A}|} adj(\mathbf{A})$$

En donde la matriz adjunta de **A**, Adj(**A**), es la matriz transpuesta de cofactores de **A**.

Otras dos operaciones con matrices que nos serán de utilidad son la traza y el producto Kronecker.

La traza de una matriz es la suma de los elementos de su diagonal principal y tiene las siguientes propiedades:

- tr(A+B)=tr(A)+tr(B)
- tr(AB)=tr(BA)
- tr(ABC)=tr(CAB)=tr(BCA)
- tr(k)=k donde k es una constante

El producto Kronecker nos permite multiplicar matrices que por sus dimensiones no son conformables para la multiplicación. Sea **A** una matriz cualquiera con dimensiones (n,k) y **B** una matriz (m,p).

Ecuación característica para λ y **x**.

Esto lo hacemos si tomamos el determinante $|\lambda \mathbf{I}-\mathbf{A}|$, que es un polinomio de grado n y que se le conoce como polinomio característico de la matriz **A**, en donde las raíces del polinomio son las raíces características λ .

Las raíces se llaman eigenvalores y los vectores correspondientes son eigenvectores.

Veamos un ejemplo, consideremos una matriz cuadrada y simétrica A:

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

Construimos la ecuación característica:

$$(\lambda I - A) = \begin{bmatrix} \lambda - 4 & -2 \\ -2 & \lambda - 1 \end{bmatrix}$$

Obtenemos el polinomio característico:

$| \lambda \mathbf{I} - \mathbf{A} | = (\lambda-4)(\lambda-1) - 4 = 0$

$= \lambda^2 - \lambda - 4\lambda + 4 - 4$

$= \lambda^2 - 5\lambda$

Podemos factorizar:

$| \lambda \mathbf{I} - \mathbf{A} | = \lambda(\lambda-5) = 0$

Las raíces que satisfacen la ecuación son:

$\lambda_1 = 0$

$\lambda_2 = 5$

Sustituimos cada una de las raíces en la ecuación característica para resolver en términos de **x.**

Para λ_1 = 0 el vector característico es:

$$A = \begin{matrix} \begin{bmatrix} 0-4 & -2 \\ -2 & 0-1 \end{bmatrix} & [X_1] & \\ & & = \\ \begin{bmatrix} -4 & -2 \\ -2 & -1 \end{bmatrix} & [X_2] & \end{matrix} \begin{matrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{matrix}$$

El sistema de ecuaciones queda así:

$-4x_1 - 2x_2 = 0$

$-2x_1 - x_2 = 0$

Despejando para las x´s:

$x_2 = -4/2\ x_1 = -2x_1$

$x_1 = -1/2x_2$

Por tanto para λ = 0 el eigen vector es:

$$X = \begin{bmatrix} 1/5^{1/2} \\ -2/5^{1/2} \end{bmatrix}$$

Se hace lo mismo para λ = 5

El vector caracteristico para λ =5 es:

$$X = \begin{bmatrix} 2/(5)^{1/2} \\ 1/(5)^{1/2} \end{bmatrix}$$

Se cumple para los eigen vectores de una matriz simétrica que son ortogonales:

Para cada $i \neq j$ $\quad x_i' \, x_j = 0$

De acuerdo con nuestro ejemplo: $X_i' x_j = \left[-2/(5)^{1/2}\right] \begin{bmatrix} 2(5)^{1/2} \\ 1/(5)^{1/2} \end{bmatrix} = \frac{2}{5} - \frac{2}{5} = 0$

Con base en los valores característicos se puede obtener:

1) Rango (**A**) = Rango (Λ)

Es el número de valores diferentes de cero en su diagonal principal; en este caso el rango (**A**) = 1

2) Los valores característicos también nos permiten determinar el signo de una matriz:

Si todas las raíces λ son positivas, **A** es positiva definida.

Si son negativas es negativa definida

Si algunas son cero y las demás negativas es seminegativa definida

Si tiene raíces positivas y negativas es indefinida.