**CAPÍTULO 2**

# 2. EL MÉTODO JACKNIFE

**2.1 Introducción**

En este capítulo revisamos la parte teórica del método Jacknife y se hacen ilustraciones del mismo; en la sección 2.2 presentamos una breve historia del método Jacknife, en la sección 2.3 se muestra la forma general o metodología del método Jacknife, en la sección 2.4 encontraremos el algoritmo para la obtención del estimador Jacknife, en la sección 2.5 el algoritmo para la obtención del sesgo y la desviación estándar del estimador Jacknife, en la sección 2.6 se ha realizado el diagrama de flujo para la obtención del estimador Jacknife y en la sección 2.7 tenemos el diagrama de flujo para la obtención del sesgo y la desviación estándar del estimador Jacknife.

**2.2 Historia**

En 1948, Maurice Quenouille, presentó una investigación en la cual lograba reducir el sesgo de los estimadores para los coeficientes de correlación parcial y autocorrelación en las series temporales aplicando el método que denominó Jacknife, como lo podemos constatar en (12). Este método se basa en la obtención de datos ficticios a partir de los datos originales, se estima la variabilidad del estimador a través de la variabilidad sobre los conjuntos de datos ficticios; es usado cuando no se puede determinar el sesgo y error estándar de los estimadores.

Se lo suele clasificar como un método de remuestreo o método intensivo por computador, ya que las medidas de precisión y dicho estimador, se estiman tomando muestras repetidas de los datos obtenidos.

Solemos obtener estimadores para los parámetros poblacionales, a través de los métodos conocidos de máxima verosimilitud, de los momentos o de suficiencia mínima y mínima varianza los cuales por lo general nos dan estimadores con las características deseadas de insesgadez y mínima varianza, sin embargo cuando nos enfrentamos a estimadores que no cumplen con dichas características, desearíamos lograr reducir su sesgo, esto lo podemos conseguir al aplicar el método de estimación Jacknife.

Muchas inferencias estadísticas son posibles realizarlas suponiendo que los datos cumplen, con la hipótesis de normalidad. Pero si no se cumpliese este supuesto, las inferencias que se realicen en base a dicha hipótesis, no serán confiables; este método ayuda a solucionar este inconveniente sin suponer que los datos siguen determinada distribución.

En 1956, Quenouille generalizó la idea de la siguiente manera:

Se puede expresar el sesgo de cualquier estimador , de un parámetro poblacional θ basado en una muestra aleatoria de tamaño n, como una serie de potencias de , esto se puede verificar al desarrollar la serie de Taylor de la expresión (-θ) y obteniendo el valor esperado de la misma, es así que se obtiene, que para muchos estadísticos se cumple que el sesgo tiene la forma:



Observamos que el valor esperado del estimador depende del parámetro poblacional que está siendo estimado y el sesgo el cual lo constituye una serie de potencias , el orden del sesgo viene dado por el término mayor de esa serie, que sería el que tiene más peso en el sesgo del estimador.

En la serie original del sesgo, la serie de potencias, tenía un orden O() ó O(n-1), al aplicar el método Jacknife observamos que el sesgo se reduce a un orden O(n-2) ya que se elimina el término de .

Quenouille, propuso también un estimador que reduce el sesgo de orden 2, es decir al estimador Jacknife que reduce el sesgo de O(n-1), o algún otro estimador en el cual el sesgo sea de O(n-2), como sigue:



Podemos observar que el sesgo del estimador de O(n-2), se reduce al orden O(n-3).

**2.3 Forma General**

 El método Jacknife consiste en particionar la muestra aleatoria en g grupos iguales de tamaño h cada uno. Si denotamos a  como el estimador de un parámetro desconocido θ y  el estimador de un parámetro desconocido pero sobre la muestra de tamaño (g-1)\*h, donde el i-ésimo grupo de tamaño h en la muestra original ha sido eliminado.

Definamos los pseudovalores:



Tukey, en 1958 sugirió que la distribución del estimador Jacknife sigue una distribución t-student, ya que en muchos casos los g grupos pueden ser tratados como variables aleatorias independientes e idénticamente distribuidas, teniendo el estadístico siguiente:



Que tendrá aproximadamente una distribución t con (g-1) grados de libertad, si la varianza es desconocida, para g lo suficientemente grande el estadístico converge en distribución a una Normal. Este es un resultado deseable, ya que podemos obtener intervalos de confianza para el parámetro estimado, se suele escoger g=n y h=1.

Efron y Tibshirani en 1993 obtuvieron, una expresión para el sesgo estimado, y la varianza estimada del método Jacknife.

La estimación del sesgo de estimador Jacknife que se obtiene del método Jacknife está dado por:



 La estimación del error estándar usando Jacknife es:

 Usando los pseudovalores  tenemos:

 **Ejemplo 1:**

 

 

**Ejemplo 2:**

 

##

##  Ilustraciones:

 Este ejemplo es una ilustración de la aplicación del método Jacknife, para obtener el estimador de la media muestral, su sesgo y su varianza. Se ha tomado una muestra aleatoria de tamaño 7 de una población exponencial con media 36, como se observa en la Tabla III.

**Tabla III**

*Estimación por el Método Jacknife*

**Muestra Aleatoria de un población Exponencial con media 36**

|  |  |
| --- | --- |
| **Nº.de unidad observada** | **Muestra** |
| 1 | 35.6741 |
| 2 | 45.4633 |
| 3 | 26.9745 |
| 4 | 38.9127 |
| 5 | 36.7966 |
| 6 | 39.8383 |
| 7 | 28.2845 |

**Elaboración**: *R. Plúa*

El estimador convencional de la media con la muestra anterior toma el valor de 35.992. Se obtienen las siete submuestras a partir de la anterior, el estimador de la media de cada una de ellas y los pseudovalores respectivos, como se observa en la Tabla IV.

**Tabla IV**

*Estimación por el Método Jacknife*

**Ilustración para la obtención de los Pseudovalores en la estimación Jacknife para la media poblacional**

|  |  |
| --- | --- |
|  | **Submuestras** |
| **Nº.de unidad observada** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| 1 | 45.463 | 35.674 | 35.674 | 35.674 | 35.674 | 35.674 | 35.674 |
| 2 | 26.975 | 26.975 | 45.463 | 45.463 | 45.463 | 45.463 | 45.463 |
| 3 | 38.913 | 38.913 | 38.913 | 26.975 | 26.975 | 26.975 | 26.975 |
| 4 | 36.797 | 36.797 | 36.797 | 36.797 | 38.913 | 38.913 | 38.913 |
| 5 | 39.838 | 39.838 | 39.838 | 39.838 | 39.838 | 36.797 | 36.797 |
| 6 | 28.285 | 28.285 | 28.285 | 28.285 | 28.285 | 28.285 | 39.838 |
| **Media Muestral** | ***36.045*** | ***34.413*** | ***37.495*** | ***35.505*** | ***35.858*** | ***35.351*** | ***37.277*** |
| **Pseudovalores** | ***35.674*** | ***45.463*** | ***26.975*** | ***38.913*** | ***36.797*** | ***39.838*** | ***28.285*** |

**Elaboración**: *R. Plúa*

El estimador Jacknife para la media poblacional, es decir el promedio de los pseudovalores es 35.992, cuyo valor coincide con el estimador convencional por lo que, el sesgo es por la estimación Jacknife el mismo que obtendríamos la estimación normalmente utilizada, es decir  y su varianza es =218.14.

Utilizaremos la misma población y muestra aleatoria estipulada anteriormente para la estimación de la media poblacional, para ilustrar el método Jacknife pero ahora estimando la mediana poblacional cuyo valor es de 24.95.

El estimador de la mediana con la muestra dada en la Tabla III toma el valor de 38.913.

Se obtienen las siete submuestras a partir de la anterior, el estimador de la mediana de cada una de ellas y los pseudovalores respectivos, como se muestra en la Tabla V. El estimador Jacknife para la mediana poblacional es 63.008, por tanto el sesgo de estimación es  y su varianza es =1442.85.

**Tabla V**

*Estimación por el Método Jacknife*

**Ilustración para la obtención de los Pseudovalores en la estimación Jacknife para la mediana poblacional**

|  |  |
| --- | --- |
|  | **Submuestras** |
| **Nº.de unidad observada** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| 1 | 45.463 | 35.674 | 35.674 | 35.674 | 35.674 | 35.674 | 35.674 |
| 2 | 26.975 | 26.975 | 45.463 | 45.463 | 45.463 | 45.463 | 45.463 |
| 3 | 38.913 | 38.913 | 38.913 | 26.975 | 26.975 | 26.975 | 26.975 |
| 4 | 36.797 | 36.797 | 36.797 | 36.797 | 38.913 | 38.913 | 38.913 |
| 5 | 39.838 | 39.838 | 39.838 | 39.838 | 39.838 | 36.797 | 36.797 |
| 6 | 28.285 | 28.285 | 28.285 | 28.285 | 28.285 | 28.285 | 39.838 |
| **Mediana** | ***37.855*** | ***37.855*** | ***37.855*** | ***31.886*** | ***32.944*** | ***32.944*** | ***32.944*** |
| **Pseudovalores** | ***45.263*** | ***45.263*** | ***45.263*** | ***81.078*** | ***74.729*** | ***74.729*** | ***74.729*** |

**Elaboración**: *R. Plúa*

En algunas ocasiones, el estimador Jacknife no funciona, y es cuando se utilizan estimadores que no son suaves, un estimador suave es aquel que para ser obtenido utiliza toda la información de la muestra aleatoria, como ejemplo de estimadores suaves tenemos la media muestral, varianza muestral, coeficiente de correlación, etc. Y como ejemplo de estimadores que no son suaves tenemos la mediana muestral. Citemos un caso, para estimar la mediana de una población exponencial podemos obtener valores negativos para este estimador, lo cual no tiene sentido ya que el dominio de la función de densidad exponencial es mayor igual a cero, y por tanto el método Jacknife no es apropiado.

Algunos estudios empíricos revelan que para el caso de los estimadores que se encuentran limitados dentro de un rango como es el caso del coeficiente de correlación, , y la varianza, , el método Jacknife puede trabajar mejor si el estadístico es transformado, para el estimador del coeficiente de correlación la transformación de Fisher de los coeficientes de correlación y el logaritmo de la varianza para los estimadores de la varianza poblacional, esto es sugerencia de muchos estudios empíricos, ya que se suelen obtener valores negativos, aún siendo estimadores suaves.

## 2.4 Algoritmo para obtener el estimador Jacknife

1. Definir un vector aleatorio X ∈ ℜn, donde se encuentra la muestra aleatoria.
2. Definir el estimador  theta sobrero, de un parámetro θ de la población, basado en una muestra aleatoria.
3. Definimos el tamaño de cada grupo como , donde g es el # de grupos en que será particionada la muestra.
4. Se define
5. Para todo i=1,..g

 

6. 

1. Presentar 

**2.5 Algoritmo para obtener el sesgo y la desviación estándar del estimador jacknife**

* + - 1. Definir un vector X ∈ ℜn, donde se encuentra la muestra aleatoria.
			2. Definir el estimador  theta, de un parámetro θ de la población, el cual está basado en una muestra aleatoria.

3. Se define

4. Para todo i=1,..g

 

5.

6.

7. Presentar Sesgo y σjack.

## 2.6 DIÁGRAMA DE FLUJO PARA OBTENER EL ESTIMADOR JACKNIFE

INICIO

n, g

i=1,...n

X[i]

i=1,...g

z=1

T=Tetha(X,n)

h=n/g

j=1,...n

1

X1[z]=X[j]

z=z+1

h\*(i-1)<=j

Y

j<=h\*i

TAN=Tetha(X1,(g-1)\*h

T1[i]=g\*T-(g-1)\*TAN

i=1,..,g

R=R+T1[I]

R=R/g

FIN

SI

NO

## Significado de variables usadas:

n= tamaño de la muestra aleatoria.

g= # de grupos en la muestra

X = vector que contiene la muestra aleatoria.

T = Estimador evaluado sobre los datos de la muestra aleatoria de tamaño n.

X1= vector que contiene la muestra de tamaño (n-1) tomada de la muestra de tamaño n.

TAN= valor del estimador evaluado en la muestra i-ésima de tamaño (n-1)

T1= vector que contiene los pseudovalores con los que trabaja el método jacknife.

R= Valor del estimador jacknife.

i,j,z = Contadores

## 2.7 Diágrama de flujo para obtener el sesgo y desviación estándar del estimador jacknife

INICIO

n

i=1,...n

X[i]

i=1,...g

z=1

T=Tetha(X,n)

g=n; h=g/n

j=1,...n

1

X1[z]=X[j]

z=z+1

h\*(i-1)<=j

Y

j<=h\*i

F[i]=Tetha(X1,(g-1)\*h

R=R+F[i]

i=1,..,g

D=D+(F[I]-R)2

D=(n-1)/n\*D

FIN

SI

NO

S=(n-1)\*(R/n-T)

## Significado de variables usadas:

n= tamaño de la muestra aleatoria.

g= # de grupos en la muestra

X = vector que contiene la muestra aleatoria.

T = Estimador evaluado sobre los datos de la muestra aleatoria de tamaño n.

X1= vector que contiene la muestra de tamaño (n-1) tomada de la muestra de tamaño n.

F= vector que contiene los valores del estimador evaluado en la muestra i-ésima de tamaño (n-1)

R= Promedio de las muestras de tamaño (n-1)

S= Sesgo del estimador

D=Varianza del estimador

i,j,z = Contadores