811 21-

Revista
de
Ciencias Económicas

PUBLICACION DE LA FACULTAD DE CIENCIAS ECONOMICAS
CENTRO DE ESTUDIANTES Y COLEGIO
DE GRADUADOS

Año XV — Julio 1927 — Serie II N° 72

DIRECCION Y ADMINISTRACION
CALLE CHARCAS, 1835
BUENOS AIRES

Sobre una nueva fórmula de ajustamiento

La tabla de mortalidad de la población de Buenos Aires, calculada en la Facultad de Ciencias Económicas en los años 1922-1923 fué ajustada por medio de una fórmula sobremanera sencilla y que presenta la ventaja de constituir la solución del problema al cual se llega formulando analíticamente el carácter de compromiso entre dos exigencias antagónicas —la de corregir un sistema de datos observados para aumentar su regularidad y la de corregir lo menos posible, porque toda corrección significa una violencia hecha a lo que en las ciencias de observación interesa sobre todo: el resultado de la observación — propio de todo ajustamiento.

Método y fórmula fueron ilustrados por mí en un artículo publicado en la *Zeitschrift für angewandte Mathematik und Mechanik* (1): pero es oportuno que de ellos quede una constancia, por así decir, auténtica, en la Facultad misma.

I

Parece natural admitir esto: que los valores verdaderos $P'_0, P_1', \ldots$ de las probabilidades de muerte correspondientes a las edades 0, 1,... varían con una cierta regularidad, disminuyendo, por ejemplo, hasta alcanzar un cierto valor mínimo aumentado después. Y que si los valores observados $P_0\ P_1, \ldots$ de las mismas probabilidades manifiestan en conjunto la tendencia definida, pero neutralizada en parte por variaciones que no obedecen a ninguna ley aparente, ello es debido a la limitación del campo de observación y a los errores de observación.

Si formamos las diferencias

$$\Delta P_0 = P_1 - P_0,\ \Delta P_1 = P_2 - P_1, \ldots$$
$$\Delta P'_0 = P_1' - P'_0,\ \Delta P'_1 = P'_2 - P'_1, \ldots$$

constatamos que sus sumas $P_n - P_0, P_n' - P_0'$ serán sensiblemente iguales, si es que aproximadamente $P_0 = P'_0, P_n = P'_n$ y n es la edad extrema, mientras que la suma de sus valores numéricos o, lo que prácticamente lo mismo, de sus cuadrados, diferirán una de otra, siendo

(1) Ueber mechanische Ausgleichung, Junio 1925.

$$(\Delta P_0)^2 + (\Delta P_1)^2 + \ldots + (\Delta P_{n-1})^2 > (\Delta P'_0)^2 + (\Delta P'_1)^2 + \ldots + (\Delta P'_{n-1})^2$$

desde que la suma formada es, a paridad de valores extremos, tanto mayor cuanto mayor es la variación o menor la **regularidad** de los datos que se consideran. Precisamente tomaremos

$$R = (\Delta y_0)^2 + (\Delta y_1)^2 + \ldots + (\Delta y_{n-1})^2$$

como expresión de la regularidad de un sistema $y_0, y_1, \ldots, y_n$ de datos ajustados.

Un ajustamiento que obedeciera únicamente a la exigencia de una regularidad máxima nos llevaría a la adopción de datos ajustados que no varían, o que varían linealmente, porque a ellos corresponde un valor mínimo de R.

Pero a la exigencia de regularidad se añade la de fidelidad a los datos observados, que una fidelidad máxima llevaría a adoptar sin corrección de ninguna especie. En este caso la suma de los valores numéricos, o de los cuadrados, de las correcciones

$$y_0 - P_0, \; y_1 - P_1 \ldots$$

sería nula. La fidelidad es tanto mayor cuanto menor es

$$F = (y_0 - P_0)^2 + (y_1 - P_1)^2 + \ldots + (y_n - P_n)^2.$$

Sean $y_0, y_1, \ldots y_n, z_0, z_1, \ldots, z_n$ dos sistemas de valores ajustados. ¿Cuál será mejor? Si R_1 y F_1 son los valores de R y de F que corresponden al sistema de las y, R_2 y F_2 los que corresponden al sistema de las z, y es

$$R_1 < R_2, \; F_1 < F_2.$$

la contestación no es dudosa: el sistema de las y, más regular y más fiel, es mejor. No pasa lo mismo si

$$R_1 > R_2, \; F_1 < F_2$$

La contestación depende entonces de cuál de las dos exigencias de regularidad y de fidelidad nos parece más importante. En el primer caso mejor es el sistema de las y, en el segundo el de las z.

Si para fijar las ideas suponemos sea

$$R_1 = 110, \; F_1 = 70, \; R_2 = 100, \; F_2 = 80$$

ambos sistemas son igualmente buenos si fidelidad y regularidad tienen una misma importancia, desde que

$$R_1 + F_1 = R_2 + F_2 = 180$$

mientras que es mejor el sistema de las y si la fidelidad tiene una importancia (expresada por un peso o coeficiente de importancia) doble que la regularidad, desde que

$$R_1 + 2\,F_1 = 250 < R_2 + 2\,F_2 = 260$$

es mejor el sistema de las z si a la fidelidad corresponde un peso doble que a la regularidad, desde que

$$2\,R_1 + F_1 = 290 > 2R_2 + F_2 = 280.$$

Si en general suponemos de atribuir a la exigencia de regularidad un coeficiente de importancia 1 y a la exigencia de fidelidad un coeficiente de importancia λ ($o \leqslant \lambda \leqslant \infty$) mejor será el sistema de valores ajustados definido por la condición

$$R + \lambda F = \text{mínimo.}$$

Es este el problema de mínimo, que se resolvió, y que lleva a la fórmula

$$y_x = C\,[P_x + \varrho\,(P_{x-1} + P_{x+1}) + \varrho_2\,(P_{x-2} + P_{x+2}) + \ldots] \quad (A)$$

donde $\varrho > 1$ resulta definido en función de λ, por ser

$$\varrho + \frac{1}{\varrho} = 2 + \lambda$$

y C es una constante, igual a $\frac{1 - \varrho}{1 + \varrho}$ si es infinito el número de los términos del segundo miembro de la (A), igual a

$$\frac{1 - \varrho}{1 + \varrho - 2\,\varrho^{n+1}}$$

si

$$y_x = C\,(P_x + \varrho)\,(P_{x-1} + P_{x+1}) + \ldots +$$
$$+ \varrho^n\,(P_{x-n} + P_{x+n})] \quad (B)$$

Es en realidad

$$\frac{1 - \varrho}{1 + \varrho - 2\,\varrho_{n+1}}\,[1 + 2\,\varrho + 2\,\varrho.^2 + \ldots 2\,\varrho^n\,] = 1$$

y la (B) define el valor ajustado correspondiente a la edad x como un promedio ponderal de valores observados correspondientes a la edad x (a la cual corresponde un peso máximo) y a las edades x—1, x—2,... x+1, x+2,... a las cuales corresponden pesos tanto menores cuanto más difieren de x: la suma de los pesos debe ser igual a 1.

II

Pondremos

$$H = R + \lambda F = (y_1 - y_0)^2 + \cdots (y_n - y_{n-1})^2 +$$
$$+ \lambda\,[\,(y_0 - P_0)^2 + \ldots + (y_n - P_n\,)^2\,]$$

Si y_0 e y_n son fijos, H resulta ser función de $y_1, y_2 \cdots y_{n-1}$ lo es de $y_0, y_1, \cdots y_n$

Si y_0 y_n son los que corresponden al problema

$$H = \text{mínimo.}$$

Debe ser en el primer caso

$$\frac{\delta\ H}{\delta\ y_1} = y_1 - y_0 - (y_2 - y_1) + \lambda\,(y_1 - P_1) = 0$$

$$\frac{\delta\ H}{\delta\ y_2} = y_2 - y_1 - (y_3 - y_2) + \lambda\,(y_2 - P_2) = 0$$

$$\frac{\delta H}{\delta y_3} = y_3 - y_2 - (y_4 - y_3) + \lambda (y_3 - P_3) = 0$$

.. ..

$$\frac{\delta H}{\delta y_{n-1}} = y_{n-1} - y_{n-2} - (y_n - y_{n-1}) + \lambda (y_{n-1} - P_{n-1}) = 0$$

en el segundo

$$\frac{\delta H}{\delta y_0} = -(y_1 - y_0) + \lambda (y_0 - P_0) = 0$$

$$\frac{\delta H}{\delta y_1} = y_1 - y_0 - (y_2 - y_1) + \lambda (y_1 \; P_1) = 0$$

.. ..

$$\frac{\delta H}{\delta y_{n-1}} = y_{n-1} - y_{n-2} - (y_n - y_{n-1}) + \lambda (y_{n-1} - P_{n-1}) = 0$$

$$\frac{\delta H}{\delta y_n} = y_n - y_{n-1} + \lambda (y_n - P_n) = 0$$

Todas las ecuaciones del primer sistema, todas las ecuaciones del segundo, menos la primera y la última, son de la forma

$$\lambda (y_r - P_r) - (y_{r+1} - 2\, y_r + y_{r-1}) =$$
$$= \lambda (y_r - P_r) - \Delta^2 y_{r-1} = 0 \quad \text{(C)}$$

Es esta una ecuación en las diferencias lineal de segundo orden no homogénea de coeficientes constantes, cuya integral particular, correspondiente a las condiciones $y(o) = y_0$, $y(n) = y_n$ es la solución del primer problema, siéndolo del segundo la integral particular que corresponde a

$$y_0 = y_{-1}, \quad y_n = y_{n+1}$$

Valen para ecuaciones del tipo considerado proposiciones completamente análogas a las que permiten integrar ecuaciones diferenciales lineales: la integral general de (C) es la suma de una integral particular de la misma y de la integral general de la ecuación homogénea correspondiente

$$y_{r+1} - (2 + \lambda)\, y_r + y_{r-1} - 0 \quad \text{(D)}$$

Si ponemos

$$y_r = C^r$$

la (D) se escribe

$$C^r [C^2 - (2 + \lambda)\, C + 1] = 0$$

y se encuentra satisfecha si C es raíz de la ecuación característica

$$C^2 - (2 + \lambda)\, C + 1 = 0$$

es decir, si es igual a uno cualquiera de los dos valores

$$1 + \frac{\lambda}{2} \pm \sqrt{\lambda + \frac{\lambda^2}{4}}$$

reales por ser $\lambda > 0$ y de los cuales uno es el recíproco del otro. Si

$$\varrho = 1 + \frac{\lambda}{2} - \sqrt{\lambda + \frac{\lambda^2}{4}}$$

es también

$$\frac{1}{\varrho} = 1 + \frac{\lambda}{2} + \sqrt{\lambda + \frac{\lambda^2}{4}}$$

> 1. Tenemos así las dos soluciones particulares ϱ^r, ϱ^{-r} y si a y b son constantes arbitrarias, la solución general

$$a\,\varrho^r + b\,\varrho^{-r}$$

de (D)

Pero resulta, si se pone

$$z_r = \frac{\varrho - 1}{\varrho + 1}\left[P_1\left(\varrho^{-r+1} - \varrho^{r-1}\right) + \right.$$
$$\left. + P_2\left(\varrho^{-r+2} - \varrho^{r-2}\right) + \ldots + P_{r-1}\left(\varrho^{-1} - \varrho\right)\right]$$

y se toma en cuenta que, por la ecuación característica

$$1 - 2\varrho + \varrho^2 = \lambda\varrho,\ 1 - 2\varrho^{-1} + \varrho^{-2} = \lambda\varrho^{-1}$$

$$\frac{1-\varrho}{1+\varrho}\left(\varrho^{-1} - \varrho\right) = \frac{1-\varrho}{1+\varrho}\,\frac{1-\varrho^2}{\varrho} = \frac{1 - 2\varrho + \varrho^2}{\varrho} = \lambda$$

$$z_{r+1} - 2z_r + z_{r-1} = \frac{\varrho - 1}{\varrho + 1}\left[\lambda \left\{ P_1\left(\varrho^{-r+1} - \varrho^{r-1}\right)\right.\right.$$
$$\left.\left. + \ldots + P_{r-1}\left(\varrho^{-1} - \varrho\right)\right\} + P_r\left(\varrho^{-1} - \varrho\right)\right] = \lambda\left(z_r - P_r\right)$$

z_r, que no depende de ninguna constante arbitraria, es una integral particular de (C). La integral general es pues

$$y_r = a\,\varrho_r + b\,\varrho^{-r} + z_r = a\,\varrho^r + b\,\varrho^{-r}$$
$$- \frac{1-\varrho}{1+\varrho}\left[P_1\left(\varrho^{-r+1} - \varrho^{r-1}\right) + \ldots + P_{r-1}\left(\varrho^{-1} - \varrho\right)\right].$$

Si suponemos dados los valores de y_0 e y_n, que corresponde a $r = o, n$, resulta con aproximación suficiente para los fines que se persiguen

$$y_0 = a + b$$

$$y_n = a\,\varrho^n + b\,\varrho^{-n} - \frac{1-\varrho}{1+\varrho}\left[P_1\left(\varrho^{-n+1} - \varrho^{n-1}\right) + \ldots + \right.$$
$$\left. + P_{n-1}\left(\varrho^{-1} - \varrho\right)\right]$$

$$b = \varrho^n \frac{1-\varrho}{1+\varrho}\left[P_1\left(\varrho^{-n+1} - \varrho_{n-1}\right) + \ldots + P^{n-1}\left(\varrho^{-1} - \varrho\right)\right]$$

$$= \frac{1-\varrho}{1+\varrho}\left[P_1\,\varrho + P_2\,\varrho^2 + \ldots + P_{n-1}\,\varrho^{n-1}\right]$$

$$y_r = y_0\,\varrho^r - \varrho^r\,\frac{1-\varrho}{1+\varrho}\left[P_1\,\varrho + \ldots + P_{n-1}\,\varrho^{n-1}\right] +$$
$$+ \varrho^{-r}\,\frac{1-\varrho}{1+\varrho}\left[P_1\,\varrho + \ldots + P_{n-1}\,\varrho^{n-1}\right]$$

$$-\frac{1-\varrho}{1+\varrho}\left[P_1\left(\varrho^{-r+1}-\varrho^{r-1}\right)+\ldots+P^{r-1}\left(\varrho^{-1}-\varrho\right)\right]$$

Queda así finalmente, exclusión hecha de los primeros valores de R.

$$y^r=\frac{1-\varrho}{1+\varrho}\left[P_r+\varrho\left(P_{r-1}+P_{r+1}\right)+\varrho^2\left(P_{r-2}+P_{r+2}\right)+\ldots\right]$$

III

En el caso de la tabla de mortalidad citada no resultó necesario ajustar los valores correspondientes a las 13 primeras edades ni posible ajustar los correspondientes a las edades superiores a 80 y, hechos los cálculos en el supuesto de $\lambda=2$ y de $\lambda=4$, parecieron preferibles los resultados correspondientes a $\lambda=2$.

Julio de 1927

HUGO BROGGI
Profesor de Estadística de la
Facultad de Ciencias Económicas