

CENTRO LATINOAMERICANO DE
DEMOGRAFIA - CELADE SAN JOSE
X) CURSO REGIONAL INTENSIVO
DE ANALISIS DEMOGRAFICO
PARA EL DESARROLLO
CRIAQ - 1988

ANALISIS MULTIVARIADO DE INTERCORRELACIONES
ENTRE FECUNDIDAD Y MORTALIDAD CON ALGUNAS
VARIABLES SOCIOECONOMICAS

JOSE VENCES RIVERA
DOMINGO TAPIA MATEO
MARTHA X. CARBALLO

San José, Costa Rica

Noviembre de 1988

CELADE - SISTEMA DOCPAL
DOCUMENTACION
SOBRE POBLACION EN
AMERICA LATINA

CONTENIDO

	Página
1. INTRODUCCION.....	1
2. DEFINICIONES Y CONCEPTOS.....	3
3. METODOLOGIA UTILIZADA.....	5
3.1. Correlación Canónica.....	5
3.2. Procedimientos de Cálculo.....	6
4. ANALISIS DE RESULTADOS.....	8
4.1. Tomando en cuenta todas las Variables.....	8
4.2. Excluyendo la Variable Edad.....	10
4.3. Comparación de los resultados obtenidos por los dos casos.....	11
5. BIBLIOGRAFIA.....	15
6. CONCLUSIONES Y RECOMENDACIONES.....	16
APENDICE. El Analisis de Correlación Canónica.....	18

1. INTRODUCCION

A través del tiempo se ha venido demostrando que existe una intercorrelación entre variables demográficas y variables socioeconómicas; que de alguna manera, las primeras tienen importantes repercusiones en la determinación de los niveles y tendencias de las segundas, y viceversa.

En la actualidad muchos de los trabajos realizados se han limitado al estudio de la relación entre pares de variables, consideradas por separado, por ejemplo, la relación entre Fecundidad y Educación, entre Fecundidad y Área de Residencia, etc.; sin embargo, esta relación puede estar influenciada por otras variables intervinientes no consideradas en el estudio, y por otro lado no se puede decidir cuál o cuáles son las variables (independientes) que más influyen en el comportamiento de otras (dependientes).

Ante esta situación es evidente que en muchos problemas prácticos resulta de interés estudiar las intercorrelaciones de variables de manera simultánea.

El propósito de este trabajo es determinar la incidencia que tienen algunas variables socioeconómicas sobre las variables demográficas fecundidad y mortalidad, y al mismo tiempo ilustrar la técnica estadística utilizada.

La fuente de datos fue una muestra del 1% obtenida de la
cuenta magnética del Censo de Panamá 1980, de todas las mujeres
entre 15 y 44 años de edad que declararon haber tenido un número
de hijos nacidos vivos. Además, de esta muestra se tomaron las
mujeres que declararon estar económicamente activas.

2. DEFINICIONES Y CONCEPTOS

A continuación se presentan los principales conceptos utilizados en el estudio:

FECUNDIDAD:

Es la frecuencia de los nacimientos que ocurren en el seno de conjuntos o subconjuntos humanos en edad de procrear.

MORTALIDAD:

Bajo la denominación de mortalidad se estudia la acción de muerte sobre la población, es la frecuencia de las defunciones o fallecimientos que ocurren en una población o en una subpoblación.

EDUCACION:

Es el nivel de escolaridad alcanzado por la población de acuerdo al sistema de enseñanza ó sistema educativo que se les ofrece: Primaria, Secundaria, Universitario.

HORAS TRABAJADAS:

Tiempo laborado por las personas que participan en determinada actividad económica.

4

INGRESO:

Remuneración que perciben las personas por el empleo de su fuerza de trabajo en un bien o servicio.

EDAD:

Es el tiempo vivido por las personas desde su nacimiento, expresado, en este contexto, en años cumplidos.

PROVINCIA:

Constituye la primera división político-administrativa en que se encuentra dividido el país.

ESTADO CONYUGAL:

Es el tipo de unión que sostiene una persona con otras de sexo diferente.

3. METODOLOGIA UTILIZADA

3.1. Correlación Canónica

El procedimiento para estudiar las intercorrelaciones entre Fecundidad y Mortalidad con algunas variables socioeconómicas, fue a través de la técnica estadística conocida como "Análisis de Correlación Canónica" (ver Apéndice).

Para esto las variables consideradas como independientes (Predictoras) son:

- X1 = Educación (EDUC)
- X2 = Horas Trabajadas (HORAstra) a la semana
- X3 = Ingreso (INGRESO)
- X4 = Provincia (PROVIN)
- X5 = Estado Conyugal (ESTCONYU)
- X6 = Edad (EDAD)

Las variables consideradas como dependientes (Criterio) son:

- Y1 = Fecundidad (FECUNDI)
- Y2 = Mortalidad (MORTALI)

Los indicadores que se utilizarán en el estudio, derivados de dicho análisis serán:

1. La correlación canónica entre los dos grupos de variables
2. Las Fargas canónicas
3. La redundancia

3.2. Procedimientos de cálculo

El paquete computacional utilizado para llevar a cabo la investigación fue el SPSS PLUS, versión para microcomputadores.

Con el propósito de utilizar las variables involucradas de manera consistente, se realizó un previo tratamiento a la información original, el cual consistió en la eliminación de valores no aceptados y en una recodificación.

Los códigos que finalmente tomaron las variables se describen a continuación:

EDUC:	0	=	Sin educación
	1	=	Primaria
	2	=	Secundaria
	3	=	Universitario
HORASTRA:	0	=	0 a 2 horas trabajadas
	1	=	3 a 9
	2	=	10 a 19
	3	=	20 a 29
	4	=	30 y más
INGRESO:	0	=	0 a 60 unidades monetarias
	1	=	61 a 150
	2	=	151 a 239
	3	=	240 a 360
	4	=	361 a 8660

EDAD: 0 = 15 a 19 años
 1 = 20 a 24
 2 = 25 a 29
 3 = 30 a 34
 4 = 35 a 40
 5 = 40 a 44

PROVIN: 0 = Darien
 1 = Bocas del Toro
 2 = Los Santos
 3 = Herrera
 4 = Veraguas
 5 = Cocle
 6 = Colón
 7 = Chiriquí
 8 = Panamá

ESTCONYU: 0 = Soltero
 1 = Casado y los demás estados conyugales

FECUNDI: 0 = 0 hijos tenidos nacidos vivos
 1 = 1
 2 = 2
 3 = 3
 4 = 4 a 5
 5 = 6 y más

MORTALI: 0 = Sin hijos muertos
 1 = Con hijos muertos

Los valores de las variables HORAstra, INGRESO y MORTALI se tomaron en base a un análisis de percentiles, una vez obtenida la distribución de frecuencias respectiva.

Los valores de la variable PROVIN fueron ordenados de menor a mayor de acuerdo al tamaño de la población de cada provincia obtenido por la muestra.

4. ANALISIS DE RESULTADOS

En los Cuadros 1 y 2 se presenta los resultados obtenidos del análisis canónico de fecundidad y mortalidad con las variables socioeconómicas.

En el análisis de hará énfasis en la incidencia que tienen las variables socioeconómicas sobre la mortalidad.

4.1. Tomando en cuenta todas las variables consideradas.

Al considerar las variables fecundidad y mortalidad con las seis variables socioeconómicas, se encontró que la correlación ($r_1 = 0.66$, $r_2 = 0.3$) entre estos grupos es significativamente diferente de cero ($\hat{a} \approx 0$) en las dos funciones canónicas; por lo tanto las dos serán interpretadas en el análisis.

Función F1:

Al observar, pues, las cargas canónicas (en los cuadros, marcados con el símbolo *) ocurre que las mujeres con mayor edad (0.37)¹, menor educación (-0.63), menor ingreso (-0.45) y mayor tiempo dedicado al trabajo (-0.43) tienden a producir

¹ Es la carga canónica correspondiente.

un incremento en fecundidad (0.97) y en la mortalidad (0.58), siendo aproximadamente dos veces más importante la primera que la segunda. Las variables más relevantes en la explicación de la fecundidad y mortalidad son la edad y la educación; el ingreso y las horas trabajadas presentan una contribución muy similar aunque de menor grado que las anteriores; provincia y estado conyugal no son de importancia por su poca influencia (-0.15 y 0.18 respectivamente) sobre las variables dependientes.

Función F2:

Cuando las mujeres presentan una disminución considerable en el tiempo de trabajo (-0.70), con poco ingreso (-0.67), con poca educación (-0.64), que están más jóvenes (-0.59) y viven en una provincia pequeña (-0.58), entonces la mortalidad de sus hijos tiende a ser grande (0.81). Estas variables independientes contribuyen de manera similar a la explicación del comportamiento de la mortalidad. En este caso, la diferencia de F1, la provincia tiene influencia importante en las variables dependientes, pues su carga pasa de -0.13 a -0.58, y en consecuencia junto con las demás variable independientes (consideradas significativas) provoca un aumento en la mortalidad, ya que su carga pasa de 0.58 a 0.81. También puede observarse que esta variables

independientes consideradas conjuntamente tiene poca influencia en la explicación de la fecundidad (-0.23). Por otro lado, el estado conyugal sigue sin presentar contribución aparente.

4.2. Excluyendo la variable Edad

Al considerar la fecundidad y mortalidad con solamente cinco variables socioeconómicas (excluyendo edad, aunque realmente no es una variable socioeconómica sino más bien una variable de control) se encontró que la correlación ($r_1 = 0.48$, $r_2 = 0.17$) entre estos grupos es significativamente diferente de cero ($\hat{\alpha} \approx 0$) en las dos funciones canónicas, por lo tanto ambas serán interpretadas en el análisis.

Función F1:

Las mujeres que tienen un nivel de educación muy bajo (-0.94), con poco ingreso (-0.73) y que dedican poco tiempo al trabajo (-0.71) presentan un aumento de importancia similar en la fecundidad y en la mortalidad de sus hijos tenidos (0.79, 0.87 respectivamente). La variable más relevante en la explicación de la fecundidad y mortalidad es la educación, seguida de ingreso y horas trabajadas, las cuales contribuyen de manera parecida; provincia y estado conyugal al igual que en el caso anterior no son de importancia por su poca influencia (-0.33, 0.20 respectivamente) sobre las variables dependientes.

Función F2:

Las mujeres, madres solteras (-0.48) que viven en provincias de menor población presentan una disminución relativamente importante en su fecundidad (-0.61), acompañadas de un incremento en la mortalidad de sus hijos nacidos vivos (0.49). En este caso las variables educación, horas trabajadas e ingreso no juegan un papel importante en la explicación de las variables dependientes. Por otro lado, a diferencia de F1, la provincia tiene una contribución importante, pues su carga pasa de -0.33 a -0.69 y en consecuencia junto básicamente con el estado conyugal están provocando estos cambios en la fecundidad y mortalidad.

4.3. Comparación de los resultados obtenidos por los dos casos

En el primer caso considerado (Cuadro 1) se obtuvo una correlación entre los dos grupos de variables de 0.66 y 0.48 para las dos funciones 1 y 2 respectivamente, y una redundancia global de 31%; sin embargo, en el segundo caso cuadro 2 al excluir la edad, tanto la correlación como la redundancia bajaron a 0.30, 0.17 y 17% respectivamente. Esto indica que la edad es la variable que está influyendo fuertemente en la explicación del comportamiento de la

fecundidad y mortalidad. Por otro lado, se tiene que, para la función canónica F_2 , la carga de la variable estado conyugal aumentó en valor absoluto de 0.11 a 0.48 al pasar del primer caso al segundo; indicando con esto que al considerar la edad en el conjunto de variables socioeconómicas le resta importancia al estado conyugal, por lo tanto es conveniente excluirla o utilizarla como una variable de control para un estudio más sustancial.

Cuadro 1. Análisis canónico entre fecundidad y mortalidad con las seis variables socioeconómicas consideradas.

VARIABLES	PESOS		CARGAS	
	F1	F2	F1	F2
Conjunto Indepte.				
X1 EDUC	-.41	-.31	-.63*	-.64*
X2 HORA STRA	-.03	-.32	-.43*	-.70*
X3 INGRESO	-.24	-.07	-.45*	-.67*
X4 PROVIN	-.04	-.37	-.15	-.58*
X5 ESTCONYU	.20	-.01	.18	-.11
X6 EDAD	.76	-.54	.77*	-.59*
Conjunto Deppte.				
Y1 FECUNDI	.88	-.63	.97*	-.23
Y2 MORTALI	.24	1.06	.58*	.81*
Correlación Canónica	.66	.30		
Raíz característica	.43	.09		
Valor Ji-cuadrada	3669.00	52.30		
Grados de libertad	12.00	5.00		
Significancia	0.00	0.00		
Redundancia	0.28	0.03		

NOTA: F1, F2 son las funciones canónicas obtenidas.

* son las cargas canónicas que más influyen en el respectivo conjunto de variables.

El tamaño de la muestra fué de $n = 567$.

Cuadro 2. Análisis canónico entre fecundidad y mortalidad con cinco variables socioeconómicas, excluyendo la edad.

VARIABLES	PESOS		CARGAS	
	F1	F2	F1	F2
Conjunto Indepte.				
X1 EDUC	-.79	.70	-.94*	-.001
X2 HORAstra	-.12	-.46	-.71*	-.42
X3 INGRESO	-.11	-.26	-.73*	-.35
X4 PROVIN	-.07	-.72	-.33	-.69*
X5 ESTCONYU	.30	-.47	.20	-.48
X6 EDAD				
Conjunto Depte.				
Y1 FECUNDI	.53	-.94	.79*	-.61*
Y2 MORTALI	.67	.86	.87*	.49*
Correlación Canónica	.48	.17		
Raiz característica	.23	.03		
Valor Ji-cuadrada	164.00	17.00		
Grados de libertad	10.00	4.00		
Significancia	0.00	<0.01		
Redundancia	0.16	0.01		

NOTA: F1, F2 son las funciones canónicas obtenidas.

- * son las cargas canónicas que más influyen en el respectivo conjunto de variables.

El tamaño de la muestra fué de $n = 567$.

5. CONCLUSIONES Y RECOMENDACIONES

En este estudio se encontró que las variable socioeconómicas consideradas influyen sobre la fecundidad y la mortalidad de la siguiente manera:

Las mujeres que tienen un nivel de educación muy bajo, con poco tiempo de ocupación y con bajos ingresos presentan un aumento en la fecundidad y en la mortalidad de los hijos tenidos vivos con efecto muy similar.

Por otro lado, se tiene que la mujeres que viven en provincias menos pobladas y que son consideradas madres solteras, presentan disminución en la fecundidad mientras que la mortalidad de sus hijos tenidos nacidos vivos tiende a aumentar.

En estudios posteriores con caracter similar al presente se sugiere tomar a la edad como una variables de control, además se recomienda ampliamente utilizar el método de Correlación Canónica como una alterantiva de investigación.

6. BIBLIOGRAFIA

Canales, J. y Krotki, C. (1988). Aspectos socioeconómicos de la población, apuntes de clase. CELADE, San José, Costa Rica.

Elizaga, J. C. (1974). Dinámica y economía de la población. CELADE, Santiago, Chile.

. Fecundidad (1977). Artículo seleccionados, serie D. Nº 1033, volumen 1. San José, Costa Rica.

García, V. y Primante D. (1988). Mortalidad, apuntes de clase. CELADE, San José, Costa Rica.

Gnadessikan (1977). Methods for statistical data analysis of multivariate observations. John Wiley.

Hair, J. F., Anderson R. E. y otros (1979). Multivariate data analysis with readings. Division of Petroleum publishing company.

Harris, R. S. (1975). A primer of multivariate statistics. Academic Press.

Nacción, G. A. Diccionario demográfico multilingüe. CELADE.

. Notas de población, Nº 34. Población y desarrollo de América Latina, pag. 11-77. CELADE.

Timm, N. H. (1975). Multivariate analysis with applications in education and psychology. Brooks / cole publishing company.

A P E N D I C E

EL ANALISIS DE CORRELACION CANONICA

El ACC es una técnica estadística multivariada, la cual facilita el estudio de intercorrelaciones entre un grupo de variables dependientes (criterio) y otro grupo de variables independientes (predictoras). Este análisis fue desarrollado por Hotelling en 1936.

La meta del ACC es encontrar para cada grupo una combinación lineal de variables, llamada variable canónica tal que la correlación entre las dos variables canónicas es maximizada. Esta correlación entre las dos variables es la correlación canónica.

Matemáticamente, estas combinaciones se representan como sigue:

$$U1 = a1 X1 + a2 X2 + \dots + ap Xp \quad (\text{para var. indepentes})$$

$$V1 = b1 Y1 + b2 Y2 + \dots + bq Yq \quad (\text{para var. depts}).$$

Esto se continua encontrando un segundo conjunto de variables canónicas, el cual produce el segundo más alto coeficiente de correlación. El proceso de construcción de variables canónicas continua hasta que el número de pares de variables canónicas (funciones canónicas) es igual al número de variables en el grupo más pequeño.

OBJETIVO DEL ACC

Los objetivos del ACC son los siguientes:

1. Determinar si dos conjuntos de variables (medidas sobre los mismos objetos) son independientes uno de otro, o bien determinar la magnitud de la relación que pueda existir entre los dos conjuntos.
2. Determinar un conjunto de pesos para cada conjunto de variables dependientes e independientes, tal que las combinaciones lineales de ellas obtenidas sean lo más correlacionadas posibles.
3. Derivar funciones lineales adicionales que maximicen la correlación remanente, éstas deben ser independientes con las anteriores.
4. Explicar la naturaleza de cualquier relación entre los conjuntos de variables dependientes e independientes, generalmente por la medición de las contribuciones relativas de cada variable original a las funciones canónicas que son extraídas.

INFORMACION OBTENIDA DEL ACC

La información obtenida a través del ACC es la siguiente:

1. Las variables canónicas
2. La correlación canónica entre las variables
3. La significancia de la CC
4. Los pesos canónicos
5. Las cargas canónicas
6. La medida de la redundancia de la varianza explicada por las funciones canónicas.

Las variables canónicas son interpretadas en base a un conjunto de coeficientes de correlación, usualmente las cargas canónicas (C) o correlaciones de estructura (1). Estos coeficientes reflejan la importancia de las variables originales en la obtención de las variables C. Así, el más grande es el más importante. El signo de los pesos indica la dirección en la cual varía la variable respectiva.

(1) Es la correlación entre cada variable original y la variable canónica correspondiente.

Para determinar la significancia de estos coeficientes, se tiene que para tamaños de muestra mayores que 300, cargas de $\pm .11$ y $\pm .15$ son significativas al 5 y 1% respectivamente.

Los resultado del ACC se acostumbra exhibirlos en la siguiente forma:

VARIABLES	PESOS	CARGAS
	F1... Fm(2)	F1...Fm
X1		
.		
.		
Xn		
Y1		
.		
Ym		
Correlación canónica		
Raíz característica		
Valor Ji - Cuadrada		
Grados de libertad		
Significancia		
Redundancia		

(2) Se considera que $m < n$

Explicación de la información contenida en el cuadro.

Las funciones (funciones 1,....., función m) que únicamente serán interpretadas en el análisis son aquellas que tienen correlación significativa.

Correlación canónica: es la correlación entre los dos conjuntos de variables

Raíz característica:

es la correlación canónica elevada al cuadrado e indica la variación común explicada por las variables canónicas.

Este indicador no tiene mucho significado práctico, sino más bien es utilizada en el cálculo de la redundancia.

Valor Ji-cuadrada:

Es la estadística que junto con sus grados de libertad determinan la significancia para probar la hipótesis nula:

H₀: No existe correlación entre dos conjuntos de variables.

C O N T R A

H_a: Existe correlación entre los dos conjuntos de variables.

Significancia:

El nivel mínimo de probabilidades de rechazar la hipótesis nula aún siendo verdadera.

Cargas canónicas (C):

Son las correlaciones entre las variables originales y las variables canónicas correspondientes.

Redundancia:

Es la proporción de la varianza explicada en el conjunto de variables dependientes por la combinación lineal de las variables independientes. Este indicador da idea de que tan apropiado es el conjunto de variables independientes para predecir el conjunto de variables independientes.

RESUMEN

El ACC es una técnica útil en la exploración de relaciones entre grupos de variables.

El ACC responde a preguntas relativas a la intensidad de la relación entre grupos de variables y a explicar la naturaleza de las relaciones así definidas.