

Guía para la ponderación de atributos de la Herramienta de Evaluación de la Calidad de Registros Administrativos – HECRA

INTRODUCCIÓN

La evaluación de la calidad de los Registros Administrativos a ser utilizados con fines estadísticos es un concepto multidimensional.

Los modelos de evaluación multicriterio se caracterizan por estar conformados por atributos con diversos grados de importancia, según el grupo de interés.

La importancia relativa de cada atributo con respecto a los otros estará determinada por información cuantitativa, es decir, por el peso asignado a cada uno de ellos.

BARREDO CANO (1996:125) señala que el objetivo de la ponderación es el de llegar a expresar, en términos cuantitativos, la importancia de los distintos elementos, agrega asimismo que si bien es frecuente asignar pesos a los criterios, la especificación de los mismos es una cuestión en la que no existe un método generalmente aceptado para su determinación, pudiéndose considerar este proceso como un aspecto que puede crear controversias acerca de la asignación de dichos pesos.

Existen diferentes procedimientos para ponderar variables o determinar pesos, dentro de los modelos multicriterio, como son los métodos de Diakoulaki, Entropía, Ordenación Simple, SMART y el Proceso Analítico Jerárquico, entre otros.

Estos métodos, a su vez, pueden ser clasificados de acuerdo a la información que utilizan o necesitan para su aplicación. Los métodos de ponderación a partir de información cuantificada son: Diakoulaki, Entropía y Ordenación Simple. Los métodos de valoración a partir de información cualitativa son: SMART, Proceso Analítico Jerárquico y Modelos de Valoración Multicriterio Individual y Colectivo, utilizados para ponderar atributos con base en las preferencias de los participantes. Claro que la aplicación de estos métodos implica realizar consultas sobre las preferencias o valoraciones que hacen los usuarios sobre cada atributo, dimensión o elemento de la calidad de los registros administrativos contemplados en la HECRA.

En cambio, se pueden utilizar otros métodos basados en el análisis multivariado de los datos ya recabados para los atributos que se quieren ponderar. Las técnicas de análisis multivariado de más comúnmente utilizadas son: Escalamiento Multidimensional o Análisis de Proximidades, Análisis de Componentes Principales, Análisis Factorial, Análisis de Correspondencias, Análisis de Correspondencias Múltiples.

MÉTODO DE PONDERACIÓN DE ATRIBUTOS

Los métodos de análisis multivariado utilizan el criterio de proximidad (similitud o disimilitud) entre los elementos, representadas en un espacio como distancias.

En el caso particular de la HECRA, los atributos que la componen son variables cualitativas, por lo cual se analizó la aplicación del método de análisis multivariado de

datos cualitativos más apropiado de acuerdo a la cantidad de atributos y casos recolectados, optando por el Análisis de Componentes Principales¹.

El método de Análisis de Componentes Principales (ACP) tiene como objetivo reducir la dimensionalidad de un conjunto de datos, quedándose con la mayor variabilidad posible. El ACP sirve para hallar las causas de la variabilidad del conjunto de datos y ordenarlas por importancia. Esta técnica se aplica tomando como insumo la matriz de varianzas y covarianzas o la matriz de correlaciones, como en este caso se cuenta con variables categóricas se optó por aplicar esta técnica sobre la matriz de correlaciones, calculada utilizando correlaciones policóricas.

El método ACP ha sido aplicado sobre los datos (62 variables categóricas) de 56 cuestionarios de la HECRA² recolectados entre octubre de 2011 y diciembre de 2012 por la Secretaría de Planeación y Presupuesto del Estado de Yucatán, en el marco del estudio realizado entre las oficinas que producen y utilizan registros administrativos con fines estadísticos.

El procedimiento que se describe a continuación ha sido elaborado con el apoyo del experto en análisis multivariado Lic. Diego Forteza.

Análisis de Componentes Principales para la ponderación de atributos de la HECRA

1. Análisis del elemento “Fuente de datos administrativa”

El elemento “Fuente” está conformado por 5 dimensiones, cada una formada por distinta cantidad de variables (ver cuadro 1). En una primera instancia se depuraron las variables con más de diez datos faltantes. Por lo tanto para el elemento “Fuente” no se consideraron las siguientes variables (indicadores): I.5.3, I.5.4, I.5.5, I.5.6, I.5.9.

Posteriormente se procedió a calcular las correlaciones policóricas, para ello se fueron considerando las variables de a pares y seleccionando la máxima cantidad de casos posibles para cada par. Para el cálculo de las matrices de correlaciones se utilizó el paquete “polycor”³ del software estadístico “R”⁴. Luego de obtener las matrices de correlaciones se procedió a calcular los pesos mediante ACP utilizando la función “eigen”.

Los pesos de las variables (indicadores) consideradas en la dimensión correspondiente se presentan en la columna Pesos del indicador del cuadro 1.

¹ Peña, D. Análisis de datos multivariantes. McGraw Hill, 2002.

² “Cuestionario de auto-diagnóstico de la calidad de Registros Administrativos a ser usados con fines estadísticos”.

³ Polychoric and Polyserial Correlations, John Fox, 2010. R package version 0.7-8, <http://CRAN.R-project.org/package=polycor>

⁴ A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN:3-900051-07-0, <http://www.R-project.org>

Una vez que se obtuvieron todas las dimensiones se calculó la matriz de correlaciones y se aplicó nuevamente ACP para obtener los pesos de las dimensiones/atributos en el elemento “Fuente” (Peso de la dimensión, cuadro 1).

Los ponderadores de los indicadores y las dimensiones o atributos fueron transformados (reescalados) de modo que su suma sea 1.

Cuadro 1: ponderadores de los indicadores y correspondientes dimensiones o atributos del elemento “Fuente de datos administrativa”.

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
I.1. Relevancia	I.1.3. Demanda de información	0,5	
	I.1.4. Satisfacción del usuario primario	0,5	
			0,175
I.2. Seguridad de la Información y limitaciones en el uso de la información	I.2.1. Marco legal	0,095	
	I.2.2. Protección de datos personales	0,168	
	I.2.3. Limitaciones debido a normativa sobre la confidencialidad	0,101	
	I.2.4. Acuerdos de confidencialidad	0,164	
	I.2.5. Transferencia de datos segura	0,163	
	I.2.6. Integridad, Confidencialidad y Disponibilidad de la información	0,092	
	I.2.7. Protección física de los datos	0,125	
	I.2.8. Políticas de respaldo de los datos	0,092	
			0,148
I.3. Compromiso de entrega de los datos	I.3.1. Costos asociados a la entrega	0,047	
	I.3.2. Acuerdos de la entrega	0,166	
	I.3.3. Frecuencia	0,139	
	I.3.5. Puntualidad	0,126	
	I.3.7. Método alternativo para falta de información	0,003	
	I.3.8. Medio de entrega de los datos	0,174	
	I.3.9. Formato de archivos	0,126	
	I.3.11. Selección de datos	0,219	
			0,220
I.4. Control y mejora	I.4.1. Captura de datos	0,249	

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
continua	I.4.2. Control de consistencia	0,249	
	I.4.3. Control de cambios	0,266	
	I.4.4. Mejora continua	0,236	
			0,276
I.5. Tratamiento de los datos	I.5.1. Controles de la unidad objetivo	0,327	
	I.5.2. Controles de contenido de las variables	0,304	
	I.5.3. Controles de valores extremos	-	
	I.5.4. Modificaciones	-	
	I.5.5. Razón de no modificación	-	
	I.5.6. Modificaciones según procedimiento	-	
	I.5.7. Utiliza manejador de BD	0,133	
	I.5.9. Documentación de la BD	-	
	I.5.10. Integridad de la BD	0,236	
			0,180

2. Análisis del elemento “Metadatos”

El segundo elemento “Metadatos” está conformado por 4 dimensiones o atributos, cada uno formado por distinta cantidad de variables (indicadores). Nuevamente se depuraron las variables con más de diez datos faltantes, los indicadores no considerados son: II.2.3 y II.3.2.

El atributo o dimensión 1, al contar con una única variable se le asignó peso 1. En el atributo 2 se observó que una de las variables toma un único valor de la escala por lo cual no es posible calcular la correlación policórica, en tal caso, se asignó como peso de cada variable la media de la misma. En el atributo 3 se aplicó la técnica correctamente. Mientras que en el atributo 4 las dos variables que lo componen toman un único valor, por lo que se replicó lo hecho en el atributo 2.

Una vez obtenidas las dimensiones del elemento “Metadatos” se calculó la matriz de correlaciones y se aplicó nuevamente ACP para obtener los pesos de las dimensiones/atributos en el elemento (Peso de la dimensión, cuadro 2).

Nuevamente los ponderadores de cada indicador y de los atributos o dimensiones fueron transformados (reescalados) de modo que la suma sea 1.

Cuadro 2: ponderadores de los indicadores y correspondientes dimensiones o atributos del elemento “Metadatos”.

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
II.1. Documentación de metadatos	II.1.1. Documentación de metadatos (evaluación genérica, más allá de su uso previsto)	1	
			0,123
II.2. Completitud y claridad	II.2.1. Definición de las unidades de la población	0,498	
	II.2.2. Descripción de variables	0,502	
	II.2.3. Comunicación de cambios en definiciones/conceptos	-	
			0,386
II.3. Uso de claves únicas	II.3.1. Claves de identificación	0,5	
	II.3.2. Comparabilidad de las variables de identificación	0,5	
	II.3.3. Combinaciones únicas de variables	-	
			0,148
II.4. Comparabilidad	II.4.1. Comparabilidad de la definición de la unidad de estudio	0,5	
	II.4.2. Comparabilidad de la definición del resto de las variables	0,5	
			0,343

3. Análisis del elemento “Datos”

El tercer elemento está conformado por 12 dimensiones o atributos, cada uno formado por distinta cantidad de variables (indicadores). No se consideraron debido a la cantidad de datos faltantes las variables (indicadores): III.1.2, III.1.4, III.1.5, III.3.1, III.7.1, III.7.2, III.9.1, III.9.2, III.9.3, III.9.4, por lo tanto, solo se consideraron como influyentes en el elemento “Datos” las dimensiones o atributos 1, 2, 4, 5, 6, 10, 11 y 12.

Nuevamente, una vez que se obtuvieron las dimensiones se calculó la matriz de correlaciones y se aplicó ACP para obtener los pesos de las dimensiones/atributos en el elemento “Datos” (Peso de la dimensión, cuadro 3).

Cabe destacar que al igual que en los elementos “Fuente” y “Metadatos” los pesos de las dimensiones fueron transformados para que sumen 1. Sin embargo, al calcular los pesos de las dimensiones en el elemento “Datos” se obtuvieron dos pesos negativos, por lo que en este caso la transformación constó de una traslación, además del re-escalamiento aplicado en todos los casos para que la suma de los pesos sea igual a 1.

Cuadro 3: ponderadores de los indicadores y correspondientes dimensiones o atributos del elemento “Datos”.

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
III.1. Controles técnicos	III.1.1. Datos legibles	0,5	
	III.1.2. Redefinir conceptos y metadatos en el caso de tener más de una fuente (si corresponde)	-	
	III.1.3. Correspondencia con Metadatos	0,5	
	III.1.4. Método de unión de registros	-	
	III.1.5. Verificación de eficacia del método de unión de registros	-	
			0,068
III.2. Cobertura	III.2.1. Exceso de cobertura	0,5	
	III.2.2. Errores de clasificación	0,5	
			0,219
III.3. Unión de registros (informáticos) de los archivos de datos	III.3.1. Tasa de registros del archivo unidos con los registros del usuario primario	-	
			-
III.4. Completitud	III.4.1. Tasa de registros sin datos	0,5	
	III.4.2. Tasa de variables sin datos	0,5	
			0,237
III.5. Medición	III.5.1. Control externo al proceso (auditoría interna)	1	
			0
III.6. Claves de identificación	III.6.1. Tasa de registros con clave única	1	
			0,003
III.7. Procesamiento	III.7.1. Ajustes (datos editados)	-	
	III.7.2. Imputación	-	
			-
III.9. Uso de códigos y clasificadores estándar (nacional o internacional)	III.9.1. Uso de clasificadores estándar	-	
	III.9.2. Verificación de la codificación (tasa de errores de codificación, tasa de registros sin código por tipo de variable)	-	
	III.9.3. Tasa de errores de codificación	-	
	III.9.4. Tasa de registros sin código	-	

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
III.10. Actualidad de los datos	III.10.1. El 90% o más de las unidades, que pertenecen a la población objetivo, creadas durante el año t han sido registradas antes del fin del año t+1	1	
			0,183
III.11. Registros múltiples	III.11.1. Tasa de registros múltiples de la misma unidad	1	
			0,133
III.12. Otros controles	III.12.1. Tasa de unidades que contienen valores válidos en las variables clave	0,733	
	III.12.2. Los cuadros de salida de la operación estadística validados a través de métodos automáticos	0,267	
			0,157

4. Análisis del elemento “Producto estadístico final”

Los atributos o dimensiones de la calidad que componen el elemento “Producto estadístico final”, junto con sus ponderadores, han sido establecidos en función de la priorización determinada por los expertos de la región consultados por el Banco Mundial, a estos efectos. En el marco del trabajo realizado por el consultor internacional Alejandro Medina Giopp: “Revisión de marcos de evaluación de calidad estadística - Bases para la elaboración de una guía de auditoría de calidad de los Registros Administrativos del Estado de Yucatán”.

A continuación se transcribe el cuadro de dimensiones o atributos, indicadores y ponderadores del elemento “Producto estadístico final” (cuadro 4).

Cuadro 4: ponderadores de los indicadores y correspondientes dimensiones o atributos del elemento “Datos”.

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
IV.1. Comparabilidad	IV.1.1 - Longitud de las series temporales comparables (principales indicadores)	0,50	
	IV.1.2 - Comparabilidad en el tiempo de los microdatos (principales variables, variables secundarias, variables agregadas, otras variables), si corresponde.	0,50	
			0,20
IV.2. Relevancia	IV.2.1 - Identificación de los usuarios	0,25	
	IV.2.2 - Información sobre los usuarios	0,25	
	IV.2.3 - Tasa de satisfacción del usuario (encuesta)	0,25	
	IV.2.4 - Utilidad (usos previstos del Registro Estadístico)	0,25	

Dimensión o Atributo	Indicador	Peso del indicador	Peso de la dimensión
			0,20
IV.3. Coherencia	IV.3.1 - Coherencia de estadísticas con periodicidades diferentes	0,50	
	IV.3.2 - Coherencia de las estadísticas en el mismo ámbito socioeconómico	0,50	
			0,15
IV.4. Disponibilidad y Transparencia	IV.4.1 - Accesibilidad por Web	0,50	
	IV.4.2 - Tasa de completitud de la información de metadatos del producto estadístico	0,50	
			0,15
IV.5. Precisión y Exactitud	IV.5.1 - Coeficiente de Variación (encuestas por muestreo basadas en el marco del Registro Estadístico)	0,10	
	IV.5.2 - Tasa de falta de respuesta por unidad	0,15	
	IV.5.3 - Tasa de falta de respuesta por pregunta	0,15	
	IV.5.4 - Tasa de imputación de las principales variables	0,15	
	IV.5.5 - Tasa de edición de las principales variables	0,15	
	IV.5.6 - Tasa de exceso de cobertura	0,15	
	IV.5.7 - Tasa de errores de clasificación	0,15	
			0,15
IV.6. Oportunidad y Puntualidad	IV.6.1 - Puntualidad en el calendario de difusión/publicación del producto estadístico	0,33	
	IV.6.2 - Lapso entre el final del período de referencia y la fecha de publicación de los resultados	0,33	
	IV.6.3 - Actualidad de los datos del Registro Estadístico	0,33	
			0,15

Ponderación de los elementos de la calidad

El Análisis de Componentes Principales arrojó resultados prácticamente iguales para los tres elementos analizados: “Fuente”, “Metadatos” y “Datos”, en cuanto a sus pesos en relación al indicador de la calidad global.

Por otra parte, los componentes del elemento “Producto estadístico final” fueron ponderados a través de un método diferente, según se explicó anteriormente.

Por tanto, se tomó el criterio de asignar igual peso a los cuatro elementos. En este punto, con un estudio más detallado con información de los usuarios sobre su opinión en cuanto a la ponderación que les asignarían a cada elemento, se podrían ajustar dichos pesos, aunque creemos que no sería un cambio significativo.

NORMALIZACIÓN DE LOS VALORES

Los modelos de evaluación multicriterio, incluyen atributos que, en muchos casos, utilizan diferentes escalas o unidades de medida, como es el caso de la HECRA. Esto hace necesaria la unificación o normalización de dichos elementos con el propósito de que sean comparables al momento de aplicar los métodos de valoración o ponderación de los atributos.

Los métodos de normalización más comúnmente utilizados son: normalización por la suma, por el mayor elemento o el ideal, y por el rango. A los efectos de esta guía se utilizará el método de normalización por el mayor elemento de la escala ordinal utilizada para cada atributo, pues mantiene la proporcionalidad, lo cual es un punto clave para la valoración.

La normalización por el mayor elemento o el ideal se calcula como el cociente de cada elemento de un atributo i por el mayor elemento de la escala utilizada para dicho atributo X_{ij}

$$X_{ij}^{NORMALIZADO} = \frac{X_{ij}}{Max_Escala(X_i)}$$

El intervalo de los valores normalizados es $0 \leq X_{ij} \leq 1$

En el caso de los atributos que componen los criterios de evaluación de la calidad de los Registros Administrativos a ser usados con fines estadísticos, los valores que éstos toman están determinados, en la mayoría de los casos, por escalas ordinales correspondientes a las opciones de respuesta de las diferentes preguntas del cuestionario de “Evaluación de la calidad de Registros Administrativos a ser usados con fines estadísticos”.

BIBLIOGRAFÍA

Aznar Bellver, J., Guijarro Martínez, F. (2005). *Nuevos métodos de valoración. Modelos multicriterio*. Universidad Politécnica de Valencia. Publicado en: <http://personales.upv.es/~fraguima/Modelos%20multicriterio.pdf>. Último acceso: 10-10-2012.

Barredo Cano, J. (1996). *Sistemas de Información Geográfica y evaluación multicriterio en la ordenación del territorio*. Editorial RA-MA. Madrid, España. 264 páginas.

Diakoulaki, D., Mavrotas, G., y Papayannakis, L. (1992). *Objective weights of criteria for interfirm comparisons*. 36ª Journées du groupe européen Aide Multicritère à la Décision, Luxembourg.

Peña, D. (2002). *Análisis de datos multivariantes*. McGraw Hill.

Saaty, T. (1980). *The Analytic Hierarchy Process*. RWS Publicaciones.

Venables, WN, Ripley, BD. (2002). *Modern Applied Statistics with S*. Springer-Verlag.

Uebersax JS. (2006). The *tetrachoric and polychoric correlation coefficients*. Statistical Methods for Rater Agreement. <http://john-uebersax.com/stat/tetra.htm>