

El Uso del Tiempo

Aplicación de los Datos Composicionales Asimétricos

Lic. Iván Aliaga Casceres

✉ powervan@gmail.com

Resumen. El Uso de tiempo se entiende como el tiempo que dispone cada individuo en sus labores cotidianas que se analizan en un ámbito social, económico y hasta político, en este trabajo se expone la aplicación de la teoría de pruebas de bondad y ajuste de una muestra para la Distribución Normal Asimétrica Multivariante en un espacio muestral restringido y estructurado llamado “Simplex”, que fue desarrollado por Aitchison (1986), el cual fue denominado hoy en día como Análisis de Datos Composicionales. La aplicación se desarrolla sobre datos observados del Uso del Tiempo en tres días de la semana tomadas aleatoriamente, excluyendo el día domingo, en diferentes actividades de 20 personas residentes mayores de 18 años de 2 distritos de la Ciudad de El Alto entre los meses de marzo-mayo de 2011.

1. Introducción

Los datos Composicionales son datos con información numérica real positiva que contienen una restricción subyacente natural y de orden geométrico, matemáticamente pueden ser descritos como: Datos que proceden de una relación de equivalencia del ortante positivo real $\mathbb{R}^{+D}$, cuyo cálculo se basa en un subconjunto de este espacio llamado “Simplex” que contiene una restricción de suma constante y que a su vez mantiene una estructura euclídea propia[1]. El uso de estos datos composicionales hoy en día hacen que muchas de las interpretaciones antes pensadas para $\mathbb{R}^{+D}$ sean incorrectas, sin embargo, durante el último decenio surgieron grandes avances al respecto y ahora con el uso de esta metodología composicional, brinda más luces de investigación sobre un conjunto de problemas cuya aplicación va al “El uso del tiempo de las personas en las diferentes actividades diarias”[6].

De acuerdo con varios autores, el uso del tiempo de las personas en diferentes actividades, tiene un ámbito social, económico y hasta científico laboral, ya que revelan un orden y una estructura de estado de desarrollo de las fuerzas productivas y no productivas de la sociedad civil. La vida cotidiana podría ser estructurada según una rutina que puede variar con el género, la edad, la ocupación, el nivel educacional, etc. El estudio del uso del tiempo evidencia científicamente la situación de los indicadores del nivel de vida de la población y también la disponibilidad de tiempo de cada individuo en sus labores diarias[2].

Se utilizó las encuestas para investigar el uso del tiempo de la población desde el inicio del siglo pasado principalmente en Europa y los Estados Unidos. Existen antecedentes en Londres 1913, URSS 1920, USA 1920 y 1934, posteriormente después de la II Guerra Mundial comenzaron a proliferar este tipo de encuestas, entre las que se pueden resaltar el llamado estudio Szalai auspiciado por la UNESCO entre 1965 y 1966 donde participaron 13 ciudades de 11 países (Bélgica, Checoslovaquia, ex República Federal de Alemania, Francia, Hungría, Bulgaria, Polonia, ex Unión Soviética, USA, Yugoslavia y Perú). Otros países del mundo han sistematizado este tipo de encuestas como Holanda que la realiza cada 5 años y Dinamarca. Gran Bretaña y Francia llevan a cabo cada 10 años. También la llevan a cabo, aunque sin ninguna sistematicidad Bélgica, Alemania, Italia y España[7].

En las Naciones Unidas la medición del tiempo y su utilización constituye una de las prioridades de investigación. En su 31 período de sesiones en marzo del 2000, la Comisión de Estadísticas examinó el Plan de ejecución del proyecto “E/CN.3/2000/13”[8] sobre las cuestiones de género en la medición del trabajo remunerado y no remunerado, reconoció la importancia de esta labor realizada y señaló la labor conexa

adicional que llevaba a cabo diversos países y organismos en particular en relación con los estudios sobre el empleo del tiempo.

El Instituto Nacional de Estadística de Bolivia inicia investigaciones en este campo aproximadamente desde el 2001, sin embargo, es el año 2010-2011 donde se realizó encuestas pilotos por mandato ministerial del 2 de junio enmarcado en el artículo 338 de la Constitución Política del Estado que señala “El Estado reconoce el valor económico del trabajo del hogar como fuente de riqueza y deberá cuantificarse en las Cuentas Públicas”, con el propósito de visualizar esta temática y el rol protagónico para generar políticas públicas, el INE incluye dentro de sus áreas de trabajo el levantamiento de información, procesamiento y publicación de la investigación del Uso del Tiempo, sin embargo, los datos analizados en este paper fueron recolectados por el propio autor tomando para ello una muestra no probabilística en dos distritos de la ciudad de El Alto y cuya metodología de trabajo fue extraída en base a la tesis doctoral de Glória Mateu i Figueras[4].

2. Datos Composicionales

Sea el vector $w = (w_1, w_2, \dots, w_D)^t$ con D elementos o componentes cuyos elementos son positivos y la suma de todos ellos es constante, es decir, que se tiene la definición de un espacio estructurado llamado “Simplex” definido como:

$$\mathcal{S}^D = \left\{ (w_1, w_2, \dots, w_D)^t : w_1 > 0, w_2 > 0, \dots, w_D > 0; \sum_{i=1}^D w_i = K \right\} \quad (1)$$

La restricción de suma constante da una correspondencia específica a las partes, ya que los componentes de la composición $(w_1, w_2, \dots, w_D)^t$ con D partes positivas pueden ser completamente especificadas si se conocen $D - 1$ componentes. Este espacio simplex es un espacio vectorial porque se pueden definir operadores como en el cuerpo de los reales, los operadores de perturbación y potencia son las operaciones interna y externa del espacio vectorial, estos a su vez incluyen a un espacio de medida de probabilidad definiendo claro esta su medida de probabilidad y su respectiva función.

2.1. Transformaciones log-cocientes

Aitchison (1986) define la transformación logcociente Aditiva, denominado alr, del inglés *Additive log ratio*.

Definición 1. Dada una composición con D partes, la transformación log cociente aditiva (alr) de $X \in \mathcal{S}^D$ a $\Upsilon \in \mathbb{R}^{D-1}$ se define como:

$$\Upsilon = alr(X) = \left(\ln \left(\frac{x_1}{x_D} \right), \ln \left(\frac{x_2}{x_D} \right), \dots, \ln \left(\frac{x_{D-1}}{x_D} \right) \right)^t. \quad (2)$$

Aitchison Define una transformación log cociente centrada, denominada clr, del inglés *Centered log ratio*.

Definición 2. Dada una composición con D partes, la transformación log cociente centrada (clr) de $X \in \mathcal{S}^D$ a $Z \in \mathbb{R}^D$ se define como:

$$Z = clr(X) = \left(\ln \left(\frac{x_1}{g(X)} \right), \ln \left(\frac{x_2}{g(X)} \right), \dots, \ln \left(\frac{x_D}{g(X)} \right) \right)^t, \quad (3)$$

donde $g(X)$ es la media geométrica de las D componentes de X .

La transformación *alr* y *clr* son aplicaciones lineales entre los espacios vectoriales $\mathcal{S}^D$ y $\mathbb{R}^{D-1}$ ya que conservan las operaciones interna y externa, pero un inconveniente de la transformación alr es su falta de simetría porque el denominador de cada log cociente adquiere un protagonismo muy especial respecto al resto de los componentes, se podría escoger a priori otra componente como denominador.

Sin embargo la transformación clr es simétrica entre las partes, su imagen es el hiperplano V de $\mathbb{R}^D$ que pasa por el origen y es ortogonal al vector unitario, aquí se encuentra una nueva dificultad ya que la suma de los componentes del vector transformado es igual a cero, a raíz de estas dificultades en el año 2003 Egozcue et al [5]., define una isometría entre los espacios $\mathcal{S}^D$ y $\mathbb{R}^{D-1}$ con objeto de superar los defectos e inconvenientes de las dos transformaciones anteriores, en donde la transformación alr no es una isometría y la transformación clr a pesar de conservar las distancias y el producto escalar, transforma la composición en vectores de un subespacio de $\mathbb{R}^D$ con la restricción adicional que la suma de las componentes es igual a cero. Estos inconvenientes dan ciertas dificultades a la hora de interpretar resultados y provocaron una larga discusión del método propuesto por Aitchison (1986).

La transformación isométrica surge de manera natural, observando la transformación clr cuya condición $\sum z_k = 0$ que satisface las componentes de los vectores del subespacio $V = clr(\mathcal{S}^D)$ indica que el vector $(1, 1, \dots, 1)$ es ortogonal a este hiperplano. Si se escoge una base del espacio $\mathbb{R}^D$ formada por $D - 1$ vectores ortogonales del subespacio V con un vector unitario y normal a V , es decir, $1/\sqrt{D}(1, 1, \dots, 1)$, se obtendrá que su última componente es igual a cero. A continuación aplicando una proyección sobre el hiperplano V da lugar a una isometría entre los espacios $\mathcal{S}^D$ y $\mathbb{R}^{D-1}$.

Definición 3. Dada una base ortonormal del simplex $\mathcal{S}^D$, $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$, se define la transformación log cociente isométrica ilr de una composición $X \in \mathcal{S}^D$ a un vector $v \in \mathbb{R}^{D-1}$ como:

$$v = ilr(X) = (\langle x, \mathbf{e}_1 \rangle_a, \langle x, \mathbf{e}_2 \rangle_a, \dots, \langle x, \mathbf{e}_{D-1} \rangle_a)^t. \quad (4)$$

Con estas transformaciones y definiciones de operaciones en el simplex brindan una mejor comprensión sobre el análisis de datos composicionales cuya definición de espacios de medida y de distribuciones probabilísticas son análogas al espacio euclideo real.

3. Distribución normal asimétrica

Una distribución genérica es capaz de generar casos particulares de distribuciones de probabilidad, ciertamente la distribución normal asimétrica $\mathcal{SN}$ es una de ellas, con relación a la ya conocida distribución normal y sus aplicaciones habituales, la $\mathcal{SN}$ genera a la distribución normal como caso particular. La distribución normal asimétrica univariante, conocida en literatura inglesa como skew-normal, fue introducida y detallada por Azzalini (1985)[3].

Definición 4. Dada una variable aleatoria z , se dice que tiene una distribución normal asimétrica continua, cuya función de densidad es:

$$f(z) = 2\phi(z)\Phi(\lambda z), \quad \lambda, z \in \mathbb{R} \quad (5)$$

donde $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$, $\Phi(\lambda z) = \int_{-\infty}^{\lambda z} \phi(u)du$, $z \in \mathbb{R}$ es la función de densidad y distribución normal estándar, cuya notación es la siguiente $z \sim \mathcal{SN}(\lambda)$.

El parámetro λ es un indicador de asimetría de la distribución, tal que su dominio es toda la recta real, cuando $\lambda = 0$ se obtiene la forma de una $\mathcal{N}(0, 1)$ y cuando $\lambda \rightarrow \pm\infty$ la distribución tiende a una normal truncada en el punto 0. La asimetría crece a medida que el valor absoluto del parámetro λ aumenta, de tal manera que a partir del valor $\lambda = 20$ aquel incremento es ligeramente inapreciable. Para valores λ positivos se obtiene una distribución asimétrica con una cola prolongada a la derecha y para valores λ negativos se tiene una distribución asimétrica con una cola prolongada a la izquierda.

Como el propio autor expresa que esta distribución tiene mejor ajuste en condiciones donde la muestra objeto de análisis se aproxima a una distribución normal. La definición multivariante de esta distribución es la siguiente:

Definición 5. Dado un vector aleatorio Z de dimension $D \times 1$, se dice que tiene una distribución Normal Asimétrica Multivariante continua cuya función de densidad es:

$$f_Z(z) = 2\phi_D(Z; \Omega) \Phi(\alpha^t Z), \quad \alpha, Z \in \mathbb{R}^D \quad (6)$$

donde $\phi_D(\cdot; \Omega)$ representa la función de densidad de un vector normal $(D \times 1)$ -dimensional con marginales estandarizadas y matriz de correlación Ω ; $\Phi(\cdot)$ es la función de distribución multivariante de una $\mathcal{N}(0, I_D)$ en $\alpha^t Z$; y $\alpha \in \mathbb{R}^D$, se utiliza la notación $Z \sim \mathcal{SN}^D(\Omega, \alpha)$ o simplemente $Z \sim \mathcal{SN}^D(\alpha)$, el parámetro de forma es α , cuando $\alpha = (0, 0, \dots, 0)^t$ se obtiene la densidad de una normal multivariante D -dimensional[9].

3.1. Contraste de hipótesis

El modelo normal es una caso particular de la familia Normal Asimétrica, ya que corresponde al caso de $\alpha = 0$, en la práctica solo interesa comparar el modelo normal asimétrico ajustado con el modelo normal. En ese caso tan solo se contrasta la hipótesis nula de $\alpha = 0$ contra la hipótesis alternativa de que $\alpha \neq 0$, aplicando un test de razón de verosimilitud, para esto se necesita el máximo de la función de log verosimilitud bajo la hipótesis de normalidad asimétrica $l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha})$ y el máximo de la función de log verosimilitud bajo la hipótesis de normalidad, $l(\bar{x}, \hat{S}, 0)$, donde $\bar{x}$ y $\hat{S}$ representan los estimadores de máxima verosimilitud de los parámetros μ y Σ bajo la hipótesis de normalidad, el estadístico de prueba es:

$$RV = 2(l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha}) - l(\bar{x}, \hat{S}, 0)), \quad (7)$$

el cual, bajo la hipótesis nula, sigue una distribución χ^2 con D grados de libertad, donde D es el número de componentes del vector aleatorio.

4. Aplicación

Ahora se presentará la aplicación de la teoría de Datos Composicionales y la Distribución Normal Asimétrica con datos del *Uso del tiempo de las personas en diferentes actividades diarias*, realizado en los meses de Marzo-Mayo 2011, con personas que accedieron a ser observadas en su quehacer cotidiano del tiempo en 60 días[2]. La suma de los tiempos usados de todo un conjunto de actividades por cada unidad observacional es constante, suma 24 hrs., se tiene una muestra de 20 personas, 10 hombres y 10 mujeres, tanto profesionales como no profesionales mayores a 18 años residentes en los distritos 4 y 6 de la Ciudad de El Alto, cuya medición del uso del tiempo en horas empleadas en actividades fue durante 3 días para cada persona, estos días fueron escogidos al azar de la semana excluyendo el día domingo, por un lapso de 2 meses y dos días (Marzo-Mayo, 2011), se tuvieron en total 60 días de observación, esto asegura que existan días trabajados alternados en cada semana y no exista tiempo recurrente o posibles arrastres de actividades.

Cuadro 1 ➤ Variables utilizadas

1	TTRAB	Tiempo usado en actividades de trabajo remunerado.
2	TNREM	Tiempo usado en actividades no remuneradas*.
3	TASOC	Tiempo usado en Actividades Sociales.
4	TASEOP	Tiempo empleado en actividades de aseo y cuidado personal.
5	TINV	Tiempo empleado en investigación y/o repasos para el desarrollo profesional.
6	TOAC	Tiempo empleado en otras actividades*.

De esta forma se tiene un conjunto de datos definido composicionalmente como:

$$\mathbf{X} = [x_{k,i}; k = 1, \dots, 60; i = 1, \dots, 6] \quad \mathbf{X} \in \mathcal{S}^6 \quad (8)$$

$$\mathcal{S}^6 = [x_{k,i} > 0; \forall k = 1, \dots, 60, i = 1, \dots, 6; \sum_i x_{k,i} = 24] \quad (9)$$

La interpretación de los datos se observa de forma relativa de sus partes y no pueden justificarse interpretaciones que involucren magnitudes absolutas, por lo tanto cualquier aseveración de una composición debe hacerse en términos de los cocientes entre las partes, los cuales medirán dicha relación relativa, a continuación se observa la matriz de variación composicional.

Cuadro 2 ➤ Matriz de variación composicional de los datos observados

	1	2	3	4	5	6
1		1.215	1.589	0.348	0.632	0.236
2	0.610		0.809	1.394	1.917	1.424
3	2.633	2.023		1.727	2.269	1.735
4	0.662	0.051	-1.972		0.641	0.369
5	1.859	1.249	-0.774	1.198		0.599
6	-0.588	-1.199	-3.222	-1.250	-2.448	

El cuadro (2) muestra la matriz de variación composicional cuya diagonal inferior muestra las esperanzas de log cocientes de las partes, la diagonal superior muestra las varianzas de log cocientes de las mismas partes.

Puede observarse que la variación relativa más alta se encuentra entre los tiempos usados en: Actividades Sociales y Actividades de Investigación para el desarrollo profesional con $\hat{\tau}_{35} = 2.269$. Adicionalmente el valor negativo $\hat{\xi}_{35} = -0.774$ indica que los tiempos en actividades en investigación para el desarrollo profesional tienden a ser ligeramente más grandes que los tiempos dedicados a las actividades sociales.

Otro valor casi extremo es la existencia pequeña de variabilidad relativa entre los tiempos usados en: Actividades de Trabajo Remunerado con actividades relacionadas al Aseo y Cuidado Personal $\hat{\tau}_{14} = 0.348$, el valor positivo de $\hat{\xi}_{14} = 0.662$ muestra que los tiempos dedicados a las actividades de trabajo remunerado tienden a ser grandes que los tiempos dedicados a las actividades de aseo y cuidado personal.

Los estadígrafos centrales y de dispersión se muestran en el cuadro (3).

Cuadro 3 ➤ Resumen Estadístico de las 6 componentes

Componentes	1	2	3	4	5	6
Media	5.87	3.19	0.42	3.03	0.92	10.57
Varianza	3.29	6.21	19.09	5.70	12.47	2.03
Mínimo	2.70	0.41	0.03	0.79	0.23	2.78
1er cuartil	3.90	1.22	0.16	1.98	0.48	7.94
Mediana	5.18	4.38	0.49	2.84	0.70	9.81
3er cuartil	6.61	5.83	0.85	3.67	1.27	11.49
Máximo	8.47	12.12	3.27	6.39	4.79	17.56

Trabajando con la subcomposición TTRAB, TNREM y TASOC que corresponde al tiempo usado en: Actividades de trabajo remunerado, Actividades de trabajo no remunerado y Actividades sociales, se procede a comprobar si esta composición procede de una Distribución Normal Logística Aditiva o una Normal Asimétrica Logística Aditiva, se utilizan todas las componentes como denominadores en cada caso de la transformación alr y se aplica el test de razón de verosimilitud y las pruebas estadísticas de hipótesis de normalidad como ser: A-Darling, CV-Mises y Watson K-Smirnov.

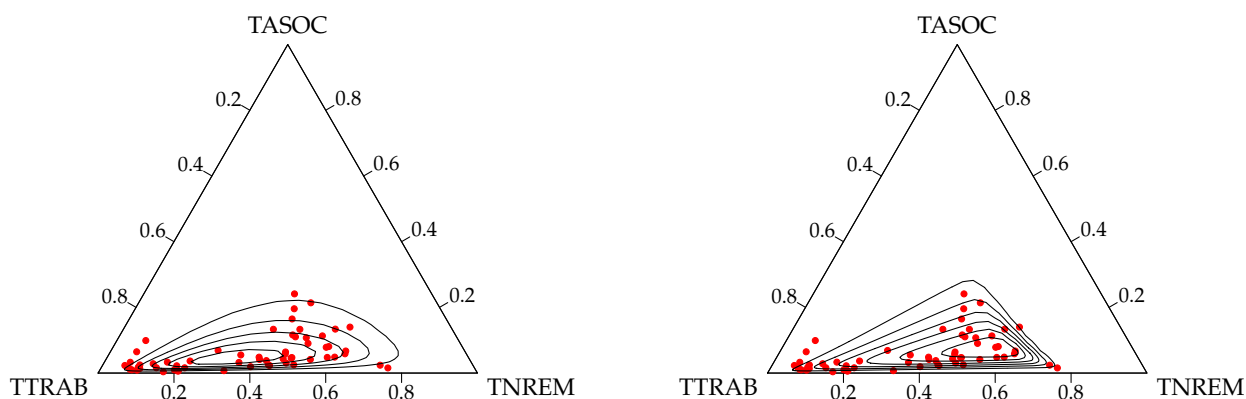


Figura 1 Diagrama Ternario, subcomposición TTRAB, TNREM y TASOC; izquierda: diagrama ternario asumiendo normalidad; derecha: diagrama ternario asumiendo normalidad asimétrica.

Estas dos facetas de representación ternaria composicional asumiendo que la muestra proceda de una Distribución Normal (izquierda) y Distribución Normal Asimétrica (derecha) se ven marcados. Aplicando la transformación logcociente aditiva con la componente TASOC como denominador se tendrán dos marginales $\ln(\text{TTRAB}/\text{TASOC})$ y $\ln(\text{TNREM}/\text{TASOC})$, los estimadores de Máxima Verosimilitud Asimétrica son:

$$\hat{\mu} = (1.22, 2.04)^t \quad \hat{\Sigma} = \begin{pmatrix} 3.56 & 0.56 \\ 0.56 & 0.80 \end{pmatrix} \quad \hat{\alpha} = (14.44, -5.09)^t \quad (10)$$

El valor de la función de logverosimilitud en estos puntos es: $l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha}) = -159.56$, luego los estimadores de Máxima Verosimilitud asumiendo normalidad en los datos son:

$$\bar{x} = (2.63, 2.02)^t \quad \hat{S} = \begin{pmatrix} 1.59 & 0.59 \\ 0.59 & 0.80 \end{pmatrix} \quad (11)$$

El valor de la función de máxima verosimilitud de estos estimadores es $l(\bar{x}, \hat{S}, 0) = -167.29$, comparando numéricamente el test de razón de verosimilitud, se obtiene que la diferencia entre estos dos ajustes es significativa $2(l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha}) - l(\bar{x}, \hat{S}, 0)) = 15.46 = \chi^2_2(1)$. Por lo tanto se presume que la subcomposición tiene un buen ajuste a razón de la Distribución Normal Asimétrica, para corroborar lo mencionado, se procede a calcular los estadísticos de bondad y ajuste bajo la hipótesis nula de normalidad y normalidad asimétrica en las observaciones transformadas.

Cuadro 4 Resumen de las pruebas estadísticas de contraste bajo la hipótesis nula de normalidad; alr_i indica la transformación alr con $i = 1, 2, 3$ componentes como denominador; los números subrayados representan puntos significativos al 95 % de confianza.

			A-Darling	CV-Mises	Watson	K-Smirnov
Marginal	i	1				
		alr_1	<u>2.212</u>	<u>0.383</u>	<u>0.357</u>	<u>1.322</u>
		alr_2	<u>2.213</u>	<u>0.384</u>	<u>0.357</u>	<u>1.322</u>
	2	alr_1	0.334	0.044	0.042	0.574
		alr_2	0.257	0.035	0.033	0.475
		alr_3	0.257	0.034	0.032	0.475
	j	1				
		alr_1	0.959	0.159	0.098	<u>1.193</u>
		alr_2	0.592	0.104	0.099	<u>1.033</u>
Prueba radii	i	2				
		alr_1	0.959	0.159	0.098	<u>1.193</u>
		alr_2	0.592	0.104	0.099	<u>1.033</u>

Se observa que en la marginal 2 no existe evidencia en contra de la hipótesis nula de normalidad en los tres tipos de transformaciones aditivas, en la marginal 1 las pruebas de A-Darling, Watson y K-Smirnov comprueban que en los primeros dos tipos de transformación aditiva se halla la existencia en contra de la hipótesis nula de normalidad, mientras en la prueba del ángulo bivariado, la prueba de Kolmogorov-Smirnov corrobora que existen pruebas suficientes en contra de la hipótesis nula de normalidad conjunta, por lo tanto no puede aseverarse que los datos en tres partes de la subcomposición pueda provenir de una Distribución Normal.

Seguidamente se procede a calcular los mismos estadísticos para las pruebas de bondad y ajuste bajo la hipótesis nula de que los datos provienen de una Distribución Normal Asimétrica.

Cuadro 5 Resumen de las pruebas estadísticas de contraste bajo la hipótesis nula de normalidad asimétrica; alr_i indica la transformación alr con $i = 1, 2, 3$ componentes como denominador; (*) puntos significativos al 95 % de confianza.

		$n = 60$		A-Darling	CV-Mises	Watson	K-Smirnov
Marginal	i						
	1	alr_1	$\lambda_1 = -5.248$	0.929*	0.162*	0.160*	0.869
		alr_2	$\lambda_1 = 5.248$	0.929*	0.163*	0.161*	0.870
		alr_3	$\lambda_1 = 1.896$	0.295	0.041	0.039	0.545
	2	alr_1	$\lambda_2 = -1.896$	0.295	0.040	0.039	0.544
		alr_2	$\lambda_2 = 0.623$	0.245	0.033	0.033	0.470
		alr_3	$\lambda_2 = -0.623$	0.246	0.034	0.034	0.470
	Ángulo Bivariado	$i \quad j$					
	1	2	alr_1	0.456	0.080	0.076	0.837
			alr_2	0.457	0.080	0.077	0.838
			alr_3	0.458	0.081	0.077	0.840

Asumiendo normalidad asimétrica solo en la marginal 1 de los dos primeros tipos de transformaciones son significativas al 95 % de confianza, dando lugar a la no existencia de pruebas en contra de la hipótesis nula de normalidad asimétrica en la marginal 2 ni en la conjunta, puede aseverarse que los datos en composición de tres partes provienen de una población con función de Distribución Normal Asimétrica, sin embargo la influencia del denominador en las transformaciones aditivas 1 y 2 hace que las conclusiones no sean tan validos del todo, porque podría usarse cualquier otra componente como denominador y llegar a las mismas conclusiones, pero al usar estas, se llegan a diferentes puntos de vista, esto es una desventaja propia de la transformación aditiva logística.

A continuación se presenta la tabla de pruebas de bondad y ajuste en todas las variables utilizando para ello la transformación isométrica con la siguiente base ortonormal,

$$U = \begin{bmatrix} 0.91 & & & & \\ -0.18 & 0.89 & & & \\ -0.18 & -0.22 & 0.87 & & \\ -0.18 & -0.22 & -0.29 & 0.82 & \\ -0.18 & -0.22 & -0.29 & -0.41 & 0.71 \\ -0.18 & -0.22 & -0.29 & -0.41 & -0.71 \end{bmatrix} \quad (12)$$

Que consiste en la partición binaria de las coordenadas de las partes según el orden de las mismas[5].

Según los valores de los estadísticos de prueba bajo la hipótesis nula de normalidad asimétrica respecto de la base ortonormal U , estos valores no logran superar los puntos críticos al 95 % de confianza en las marginales, a excepción de la primera marginal para la prueba de Kolmogorov-Smirnov, sin embargo,

todas las otras pruebas multivariantes incluyendo las marginales conjuntas no logran refutar la hipótesis nula de normalidad asimétrica en las observaciones de 6 partes, las observaciones del uso del tiempo en diferentes actividades proceden de una Distribución Normal Asimétrica.

Cuadro 6 ➤ Estadísticos de contraste para la prueba de bondad y ajuste bajo la hipótesis de normalidad asimétrica utilizando coordenadas según la base ortonormal U ; (*) puntos significativos al 95 % de confianza.

		$n = 60$	A-Darling	CV-Mises	Watson	K-Smirnov
Marginal	i					
	1	$\lambda = -7.255$	0.770	0.149	0.110	1.059*
	2	$\lambda = -2.971$	0.529	0.073	0.073	0.622
	3	$\lambda = -2.083$	0.427	0.067	0.067	0.771
	4	$\lambda = -2.609$	0.384	0.061	0.060	0.803
	5	$\lambda = 0.577$	0.369	0.065	0.065	0.640
Test Radii	i	j				
	1	2	1.109	0.212	0.136	1.067
	1	3	0.755	0.126	0.122	0.878
	1	4	1.201	0.224	0.163	1.021
	1	5	1.242	0.229	0.125	1.042
	2	3	0.550	0.088	0.082	0.736
	2	4	0.932	0.142	0.135	1.049
	2	5	0.332	0.045	0.033	0.525
	3	4	1.236	0.213	0.123	0.955
	3	5	0.743	0.078	0.071	0.602
	4	5	0.411	0.072	0.052	0.683
Test Radii	$D = 5$		0.468	0.054	0.050	0.542

5. Conclusiones

El uso de la metodología de datos composicionales brinda un estudio completo y detallado, cuando los datos disponibles son por naturaleza restrictivos y de suma constante, en este paper se propuso contrastar dos distribuciones en un espacio estructurado llamado simplex donde se vio que no existen pruebas en contra de la hipótesis nula de normalidad asimétrica en datos del uso del tiempo de personas en diferentes actividades, sin embargo, esto no quiere decir que en todas las muestras obtenidas del uso del tiempo tengan una misma distribución asimétrica, el hecho de generalizar distribuciones de probabilidad aumentando las operaciones de cálculo, esto debido al ámbito paramétrico, no debe tomarse a la ligera, ya que si bien ayuda a tener una buena bondad de ajuste con la SN , se tiene la desventaja en el cálculo directo de las estimaciones en función de la muestra.

Los datos obtenidos y analizados no muestran observaciones faltantes, razón por la cual los datos no han sido previamente procesados con algún método de imputación, si así fuera, el método escogido tendría que estar modificado bajo la hipótesis nula de que los datos proceden de una población con función de Distribución Normal o Distribución Normal Asimétrica y probar su ajuste bajo las mismas hipótesis mencionadas.

Adicionalmente se recomienda utilizar la distribución asimétrica cuando exista una continua evidencia en contra de la hipótesis nula de normalidad en las observaciones o datos recolectados.

Se debe tomar en cuenta con sumo cuidado en la utilización de la transformación isométrica, cuando se realiza la transformación siempre debe utilizar una base ortonormal, esto con el fin de realizar las interpretaciones adecuadas de los componentes como balances entre las partes.

Referencias

- [1] Aitchison, J. *The Statistical Analysis of Compositional Data*. The Blackburn Press, 1986.
- [2] Aliaga, I. *Aplicación del Modelo Normal Asimétrico Composicional Multivariante al Uso del Tiempo*. Universidad Mayor de San Andrés, 2011.
- [3] Azzalini, A. A note on regions of given probability of the skew normal distribution. *University of Padua, Italy* 1 (2004), 8.
- [4] I Figueras , G. M. *Models De Distribució Sobre El Simplex*. PhD thesis, U.P.C., 2003.
- [5] J. J. Egozc Ue, V. P.-G. Groups of parts and their balances in compositional data analysis. *Mathematical Geology* 37 (2005), 795–828.
- [6] Rosario Aguirre, Cristina García Sainz, C. C. *El Tiempo, Los Tiempos una Vara de Desigualdad*. CEPAL, 2005.
- [7] Teres A Lara Junc O, Neyda González Nápoles, E. M. L. D. J. E. C. M. A. A. L. D. C. C. *Encuesta Sobre el Uso del Tiempo*. Oficina Nacional de Estadísticas, División de Estadísticas de Naciones Unidas, 2001.
- [8] Unid As, N., Ed. *Informe sobre su 31 período de sesiones (29 de febrero a 3 de marzo de 2000), Consejo Económico y Social Suplemento No. 4*. CEPAL, Comisión Estadística, 2000.
- [9] Y A. Dalla Valle, A. A. The multivariate skew-normal distribution. *Biometrika* 83 (1996), 715–726.