

# Análisis estadístico de tablas de contingencia y chi-cuadrado para medir el flujo migratorio de origen y destino en el Ecuador año 2018

## Statistical analysis of contingency and chi-square tables to measure the migratory flow of origin and destination in Ecuador year 2018

David Lastre Baquerizo<sup>1</sup>, Melanie Páez Santana<sup>2</sup>, y Olga López Tumbaco<sup>3</sup>

### RESUMEN

En este estudio se propone la mejora de un sistema de inventarios, con el objetivo de reducir el costo y maximizar las ventas en la farmacia "La Voluntad de Dios", debido a que existe manejo empírico del inventario por parte de los bodegueros, a consecuencia de esto, no se logra satisfacer correctamente a los clientes y en ocasiones se generan pérdidas por falta de espacio en la bodega o expiración de productos. Por este motivo se decidió optimizar la gestión inventario aplicando el algoritmo de búsqueda tabú obteniendo una función objetivo basada en distribuciones de probabilidad normal de 3 productos, para esto se toma muestra del histórico de inventario del año 2017 al 2019, se realizó una proyección de ventas para el año 2020; el dando como resultado un 14% más de ganancia con respecto al año 2019 y un sistema de inventario que proyecta el uso aproximado del 95% de los productos en stock, obteniendo valores óptimos para satisfacer la alta demanda y mantener control del inventario.

**Palabras clave:** Tablas de contingencia, Chi-cuadrado, Flujo migratorio, Ecuador.

### ABSTRACT

This study proposes the improvement of an inventory system, with the objective of reducing the cost and maximizing sales in the pharmacy "The Will of God", because there is empirical management of the inventory by the winemakers, as a result -sequence of this, it is not possible to satisfy customers correctly and sometimes losses are generated due to lack of space in the warehouse or expiration of products. For this reason it was decided to optimize the inventory management by applying the taboo search algorithm obtaining an objective function based on normal probability distributions of 3 products, for this sample of the inventory history from 2017 to 2019 is taken, a sales projection was made for the year 2020; The result is a 14% increase in profit compared to the year 2019 and an inventory system that projects the approximate use of 95% of the products in stock, obtaining optimal values to meet the high demand and maintain inventory control.

**Keywords:** Contingency tables, Chi-square, Migratory flow, Ecuador.

**Fecha de recepción:** Noviembre 20, 2018.

**Fecha de aceptación:** Marzo 5, 2019.

### Introducción

La migración, para la mayoría de las personas es un medio de ayuda para mejorar su vida y la de sus seres queridos ya que encuentra la posibilidad de un mejor empleo, seguridad, estabilidad. Sin embargo, La falta de soluciones hacia la migración ha causado que la misma no sea inspeccionada y se convierta en un fenómeno en constante crecimiento abriendo paso a la sobrepoblación de los países destino y el decrecimiento en el de origen.

Los autores en [1] basan su investigación en los diversos motivos por los que los trabajadores mexicanos deciden migrar

internamente. Cabe precisar que el énfasis se enfoca en los flujos migratorios que se dan entre un estado y otro, esta dimensión representa una parte importante de la delimitación del objeto de estudio. Se implementó el uso de histogramas para registrar de forma gráfica la asociación entre cada variable las cuales son los motivos por los que se decide migrar. El estudio obtiene como resultado, que el individuo puede tener un rango medio de posibilidades de emigrar, dependiendo del comportamiento del mercado de trabajo y cómo evolucione su formación profesional. A pesar de ello este trabajo si logra comprobar la relación entre las variables, sin embargo, el trabajo realizado en este artículo utiliza tablas

<sup>1</sup> Estudiante de Ingeniería en Sistemas Computacionales. Universidad de Guayaquil, Ecuador. E-mail: david.lastreb@ug.edu.ec

<sup>2</sup> Estudiante de Ingeniería en Sistemas Computacionales. Universidad de Guayaquil, Ecuador. E-mail: melanie.paezs@ug.edu.ec

<sup>3</sup> Estudiante de Ingeniería en Sistemas Computacionales. Universidad de Guayaquil, Ecuador. E-mail: olga.lopez@ug.edu.ec

**Como citar:** Lastre, D., Paez Santana, M., & López Tumbaco, O. (2019). Análisis estadístico de tablas de contingencia y chi cuadrado para medir el flujo migratorio en el Ecuador en el 2018. *Ecuadorian Science*, 3(1), 23-30.  
DOI: <https://doi.org/10.26911/issn.2602-8077vol3iss1.2019pp23-30p>.

de contingencia, lo que permite el manejo de mucha más información ya que incluye el uso de variables cualitativas.

La mayor parte de los ingresos de Ecuador provienen de la actividad primaria exportadora y no de las actividades industriales urbanas como se refleja en [2] donde realizaron un estudio sobre los graves problemas socioeconómicos que ha experimentado Ecuador dentro de sus dos grandes ciudades, Guayaquil y Quito, generando un amplio interés en el ámbito académico y político. Esta investigación se realiza mediante el uso de análisis de regresión, el estudio obtiene como resultado que, para frenar el acelerado proceso de urbanización, los responsables de la política pueden fomentar una mayor inversión en capital humano en las ciudades periféricas y promover la generación de fuentes de empleo en el área rural. A pesar de que en el trabajo de José Rafael Alvarado López examina la relación entre las variables que causan la urbanización, este paper aplica el estadístico de chi cuadrado para demostrar si existe una relación entre las variables antes mencionadas.

La Migración se puede clasificar de dos grandes maneras: Migración Interna y Migración Externa. Dentro de [3] se aborda la problemática de las características diferenciales de la migración interna, tanto de la migración de toda la vida y también de la reciente que muestran las personas de 60 años en adelante, residentes en la mayoría de los países de América latina. El fin de esta investigación fue determinar diferencias y semejanzas de características específicas que muestra la población migrante interna con 60 años y más en Cuba, en comparación con diferentes países de América Latina. Para poder desarrollar estos resultados se usan definiciones de Migración reciente, tasas de mortalidad y fecundidad además de indicadores de estadística demográfica y tablas de contingencia. Todo este proceso dio como resultado que, en Cuba, los índices de masculinidad de migrantes y no migrantes revelan el predominio de la población femenina entre los migrantes recientes totales y los que tienen 60 y más años. Sin embargo, el trabajo realizado por Cristina Lopez si analiza las relaciones que existen entre sus variables por tablas de contingencia, pero las resuelve de manera gráfica, dentro de este trabajo las fórmulas y los modelos matemáticos del Chi cuadrado serán prioridad para determinar los resultados finales

En [4] se realiza un estudio sobre la relación causa-efecto entre migración y trabajo infantil, bien sea por desplazamiento interno o movilidad internacional, en este estudio se emplean técnicas estadísticas como tablas de contingencia e histogramas para demostrar la evolución de este en Colombia y también para observar el grado de asociación que poseen las variables que influyen al trabajo infantil. El estudio obtiene como resultado que el trabajo infantil se ha convertido en una opción para muchos niños y jóvenes y esto a su vez en la mayoría de los casos, se convierte en explotación laboral y sexual, trata de menores y, en el peor de los casos, de esclavitud. A pesar de que el trabajo de David Khoudour-Castéras muestra de forma gráfica la relación que tienen las causas de que los padres o tutores de los niños migren, en este paper mediante el uso de RStudio se mostrará una mayor cantidad de datos a relacionar ya que cuenta con una gran variedad de funcionalidades para el tratamiento y el análisis de datos.

## Materiales y métodos

Dentro de esta sección se conceptualizará los materiales y métodos que serán usados para el desarrollo de resultados de este trabajo, para ello se describe un estudio sobre el uso y manejo de tablas de contingencia mediante la prueba de chi-cuadrado el cual

se realizará en RStudio para un mejor desempeño y rapidez al momento de realizar las distintas operaciones estadísticas.

## Tabla de contingencia

Para poder aplicar tablas de contingencia debemos saber que como indica [5] una tabla de contingencia es una tabla de distribución de frecuencias absolutas conjuntas muestrales de dos variables aleatorias clasificadas en categorías, en este caso definimos estas variables como:

**Tabla 1.** Variables para las tablas de contingencia.

Variables
Sub_Cont = Subcontinente de procedencia
Sexo_Mig = Sexo del Migrante
Tip_Nac = Nacionalidad del Migrante
!! Tipo_Mov = Entrada y Salida al país.
Motiv_Viaj = Motivo de Viaje
Año = Año de movilización

A partir de ello [6] Se pretende demostrar que la utilidad metodológica de las tablas de contingencia facilita la comprensión de la relación que guardan dos variables, lo que se aplicara con las variables establecidas en la Tabla 1.

Las tablas de contingencia [7], [8] están compuestas por filas (horizontales), para la información de una variable y columnas (verticales) para la información de otra variable. Estas filas y columnas delimitan celdas donde se vuelcan las frecuencias de cada combinación de las variables analizadas. Por medio de esto se interesa estudiar si existe alguna asociación dentro de la variable que se denomina fila y la variable denominada columna y se calcula la intensidad de dicha asociación. Las tablas de contingencia pueden ser bidimensionales o tridimensionales, existe un método que se encarga de estudiar este tipo de tablas llamado Modelo de Efectos de Columna. Para poder aplicar este método se debe cumplir una restricción: de carácter ordinal en la variable de fila (X) y nominal en la variable de columna (Y). Dichas tablas de contingencia cumplen con el siguiente formato.

**Tabla 2.** Formato de una tabla de contingencia.

	X <sub>1</sub>	X <sub>2</sub>	.....	X <sub>n</sub>
Y <sub>1</sub>	N <sub>1,1</sub>	N <sub>2,1</sub>	.....	Nn.1
Y <sub>2</sub>	N <sub>1,2</sub>	N <sub>2,2</sub>	.....	Nn.2
.....	.....	.....	.....	.....
Totales	∑ N <sub>n,1</sub>	∑ N <sub>n,2</sub>	.....	∑ N <sub>n,n</sub>

La misma tabla puede expresarse en frecuencias relativas o proporciones sin más que dividir cada casilla n<sub>ij</sub> por el total N.

$$N = \sum_{j=1}^n \sum_{i=1}^n n_{ij}$$

Una vez definida la fórmula matemática para la Distribución. Exponencial, utilizaremos el lenguaje de programación, Python, para realizar la simulación de la gráfica que posee dicha distribución.

### Chi Cuadrado

La prueba del Chi Cuadrado se usa para analizar la relación de dependencia entre dos variables cualitativas. [9] Es necesario resaltar que esta prueba indica si existe o no una relación entre las variables, pero no señala el grado o el tipo de relación; es decir, no indica el porcentaje de influencia de una variable sobre la otra o la variable que causa la influencia.

Para lograr asociar estos conceptos planteamos que: Sean X e Y dos características, cualitativas o cuantitativas, con  $i=1, \dots, p$  y  $j=1, \dots, q$  modalidades o categorías, respectivamente, son presentadas en una tabla  $p \times q$ . La prueba del Chi Cuadrado es una medida-resumen que compara los valores ( $n_{ij}$ ) observados en la tabla, con los que teóricamente se obtendría ( $t_{ij}$ ), en el supuesto de que las variables X e Y fuesen independientes.

$$X = \sum_i^p \sum_j^q \frac{(n_{ij} - t_{ij})^2}{t_{ij}}$$

Los valores teóricos  $t_{ij}$  se obtienen mediante:

$$T_{ij} = \frac{n_i * n_j}{N}$$

Este estadístico toma valores comprendidos entre 0 y  $N \cdot \min\{p-1, q-1\}$ , el valor 0 indica que el numerador de la expresión anterior es nulo, por tanto las frecuencias observadas coinciden con las que habría si las variables fuesen independientes; de donde se admite la independencia de X e Y. El hecho de que sus valores dependan tanto del número de elementos de la tabla (N), como del  $n^\circ$  de filas y columnas, hace difícil su interpretación e impracticable la comparación entre tablas. El estadístico Chi-cuadrado permite contrastar la hipótesis de independencia de X e Y, basándose en el conocimiento del comportamiento de Chi-cuadrado bajo la hipótesis de independencia: Modelo Chi-cuadrado con  $(p-1)$   $(q-1)$  grados de libertad.

### Contraste de Independencia para chi-cuadrado

Una vez que calculado el valor experimental y el valor del estadístico teórico.

Se aceptará  $H_0$  si:

$$X_{exp}^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(n_{ij} - t_{ij})^2}{t_{ij}} < x_{\alpha(k-1)(m-1)}^2 \quad (4)$$

Se rechaza  $H_0$  si:

$$X_{exp}^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(n_{ij} - t_{ij})^2}{t_{ij}} > x_{\alpha(k-1)(m-1)}^2 \quad (5)$$

### Hipótesis independientes Nulas y de trabajo

Una hipótesis nula suele ser una afirmación inicial que se basa en análisis previos o el conocimiento especializado [10]. De acuerdo con la hipótesis nula las variaciones en la variable independiente no tienen correspondencia con las variaciones que pudiere haber de la variable dependiente. Es decir que existe "independencia

estadística". Las variaciones que pudiese encontrarse se deberían a factores aleatorios, ajenos a la variable independiente.

La prueba de hipótesis es una regla que se debe seguir para poder elegir entre una hipótesis nula y una hipótesis alternativa. [11] La evidencia de los datos obtenidos puede conducir a no rechazar la hipótesis nula, lo cual no implica que ésta sea cierta se considera que la principal diferencia entre estas dos teorías no radica en los cálculos, sino en las concepciones y el razonamiento subyacente

$H_0$ : Las variables en filas y columnas no están asociadas

$H_1$ : Las variables en filas y columnas están asociadas.

$H_0$  cuando  $x^2_{experimental} > x^2_{critico}$

### Grados de libertad

Dentro de la investigación de [12], [13] indica que la suma de los valores de las desviaciones individuales con respecto a su media es igual a cero, hecho que puede demostrarse. Si se conocen los  $n-1$  de los valores a partir de la media, entonces se conoce el  $n$ -ésimo valor, ya que queda determinado automáticamente debido a la restricción de 3 de que todos los valores de  $n$  sumen cero. El grado de libertad es un estadístico calculado sobre un conjunto de datos que se refiere al número de cantidades independientes que se necesitan en su cálculo, menos el número de restricciones que ligan a las observaciones y el estadístico. El número de grados de libertad del estadístico Chi-cuadrado se calcula de la siguiente forma:

- Se calcula, en primer lugar, el número de sumandos, es decir  $m \times n$ , siendo  $n$  y  $m$  el número de filas y número de columnas en la tabla.
- A esta cantidad se debe restar el número de restricciones impuestas a las frecuencias observadas. Observamos que podemos cambiar todas las frecuencias de la tabla sin cambiar los totales por filas y columnas, excepto los datos en la última fila y columna de la tabla, pues una vez que fijemos todos los valores excepto estos, quedan automáticamente fijados. Por tanto, si la tabla tiene  $m$  filas y  $n$  columnas, el número de grados de libertad es:

$$(k-1) * (m-1) \text{ o } (filas-1) * (col-1)$$

### Cálculo de las frecuencias esperadas

Para el cálculo de las frecuencias esperadas  $f_e$  [14], nos vuelve a colaborar diciendo  $f_e H_0$  que, vienen dadas por la hipótesis nula  $H_0$ , pero no siempre se puede establecer de manera inmediata. Esto solo es posible cuando trabajamos con una variable, pero cuando tenemos cuadros de doble entrada la forma de establecer el valor de la frecuencia esperada de cada celda es el siguiente:

**Tabla 3.** Tabla de contingencia de la cual saldrá la tabla de frecuencias esperadas.

X/Y	$Y_1$	$Y_2$	Marginal $Y1/Y2$
Categoría A	a	b	(a + b)
Categoría B	d	e	(d + e)
Categoría C	g	h	(g + h)
Marginal A/B/C	(a + d + g)	(b + e + h)	N

Se debe aplicar la siguiente fórmula para obtener los valores esperados.

$$e_{i,j} = \frac{f_i \cdot f_j}{n}$$

**Tabla 4.** Tabla general de frecuencias esperadas obtenida una vez ya aplicada la formula a cada celda.

X / Y	Y1	Y2	Marginal Y1/Y2
Categoría A	F.E a	F.E b	(a + b)
Categoría B	F.E d	F.E e	(d + e)
Categoría C	F.E g	F.E h	(g + h)
Marginal A/B/C	(a + d + g)	(b + e + h)	N

### Estadístico Teórico

El estadístico teórico se calcula teniendo en cuenta [15] el grado de libertad y el nivel de significación, el cual se fija de antemano (usualmente entre 0.01 y 0.10, siendo el más usado el de 0.05) luego en la tabla de chi-cuadrado se procede a elegir el número de grado de libertad y el nivel de significancia de 0.05 y el valor será el número que se encuentre en la intersección de ambas, para ello se implementa la formula [10].

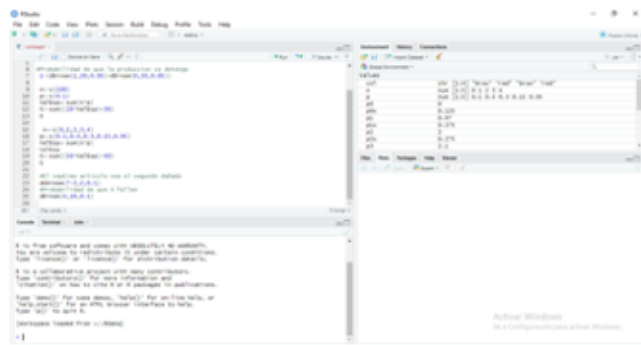
$$X^2_{\alpha(k-1)(m-1)}$$

**Tabla 5.** Ejemplo de tabla de chi-cuadrado.

V/P	0.001	0.0025	0.005	0.01	0.025	0.05
1	10.8274	9.1404	7.8794	6.6349	5.0239	3.8415
2	13.8150	11.9827	10.5965	9.2104	7.3778	5.9915
3	16.2660	14.3202	12.8381	11.3449	9.3484	7.8147
4	18.4662	16.4238	14.8602	13.2767	11.1433	9.4877
5	20.5147	18.3854	16.7496	15.0863	12.8325	11.0705

### Programa RStudio

RStudio es una interfaz de programación basada en el lenguaje estadístico R [9] que es un lenguaje de programación orientado a objetos para calculo estadístico y generar gráficos a partir de ellos. Cabe recalcar que RStudio es un software libre en donde se encuentran diferentes funciones para el análisis de datos, las cuales se implementaron en esta investigación. La consola de RStudio se clasifica de la siguiente manera.



**Figura 1.** Consola de RStudio.

### Para la interpretación de las herramientas en Rstudio:

**Tabla 6.** Cuadro de herramientas de Rstudio.

Cuadro de herramientas	
<b>File</b>	Crea y abre proyectos, importa bases de datos.
<b>Edit</b>	Edición del código: copiar, pegar, ver líneas de código.
<b>Code</b>	Insertar o extraer funciones o líneas de código.
<b>View</b>	Revisa el historial, funciones o variables usadas anteriormente.
<b>Session</b>	Inicia una nueva sesión vinculando una cuenta de correo
<b>Build, Debug, Profiles</b>	Depuran el código y encuentran los márgenes de errores en las líneas que se generaron.
<b>Tools</b>	Se escoge las consolas, la versión que se está usando y descarga paquetes de extensión para la interfaz.

La primera pantalla que se muestra, son las líneas de código en donde se va a programar las funciones que contiene R. Como se ve en (Fig. 1) Se encuentra un código escrito y a su izquierda la numeración correspondiente a cada línea. En la parte superior de la pantalla se encuentra la opción Run que es la que ejecutara línea por línea el código fuente. La segunda pantalla que se muestra en la parte de abajo corresponde a la compilación de esas líneas de código. En la parte superior derecha se encuentra las variables definidas en el código actual y las de códigos anteriores. Debajo de ella la ejecución del código completo

### Caso de estudio

El interés por los estudios sobre migración al extranjero en el país es reciente, a pesar de que el fenómeno como tal ha sido parte de las experiencias de muchas familias ecuatorianas desde muchos años atrás. Es innegable que la crisis financiera y la falta de oportunidades adecuadas de empleo fueron detonantes importantes para la decisión de migrar, pero eso no parece explicar en su totalidad ni la decisión de migrar ni la forma de inserción en los lugares de destino. Muchas de las hipótesis que surgen alrededor de la migración coinciden en señalar que la migración no surge a partir de una decisión racional, economía y estabilidad únicamente, sino que también influyen factores sociales y culturales que no necesariamente responden a la economía que es la responsabilizada por la migración.

### Resultados

#### Variable motivo de viaje vs tipo de movimiento.

**Tabla 7.** Tabla de contingencia de las variables motivo de viaje y tipo de movimiento.

	Entrada	Salida	Total
<b>Estudios</b>	431	5793	6224
<b>Eventos</b>	17663	11032	28695
<b>Negocios</b>	6023	4090	10113
<b>Otros</b>	17808	48654	66462
<b>Residencia</b>	181568	304086	485654
<b>Turismo</b>	277057	125795	402852
<b>Total</b>	<b>500550</b>	<b>499450</b>	<b>1000000</b>

En la presente tabla se muestran los datos recabados y tabulados de la población viajera registrada en relación con el motivo de viaje y el tipo de movimiento es decir si entran o salen del Ecuador,

siendo el motivo residencia el más frecuente en las salidas presenta 304086, por el contrario, el motivo de turismo en entradas presenta 277057.

**Tabla 8.** Tabla de frecuencia esperada obtenida una vez aplicada la formula a cada celda.

	Entrada	Salida	Total
Estudios	3115,4232	3108,5768	6224
Eventos	14363.2828	14331.7177	28695
Negocios	5062.0621	5050.9378	10113
Otros	33267.5541	33194.4454	66462
Residencia	243094.1097	242559.8903	485654
Turismo	201647.5686	201204.4314	402852
<b>Total</b>	<b>500550</b>	<b>499450</b>	<b>1000000</b>

A partir de la tabla anterior una vez aplicado el cálculo correspondiente tal como la tabla nos muestra, los nuevos valores serán las frecuencias esperadas y su suma de valores marginales será igual al total de nuestra tabla de contingencia original los cual nos indica que el cálculo fue concretado con éxito.

A continuación, se plantean las hipótesis a contrastar mediante el chi-cuadrado.

- La variable motivo de viaje es independiente al tipo de movimiento.
- La variable motivo de viaje si depende al tipo de movimiento.

$$\chi^2 = \sum_{ij} \frac{(431 - 3115,4232)^2}{3115,4232} = 2313,049$$

Luego se aplicó el estadístico de contraste de la formula ( ) donde nos indica que es la suma de los valores de.

$$\chi^2_{exp} = \sum_{i=1}^k \sum_{j=1}^m \frac{(431 - 3115,4232)^2}{3115,4232} = 2313,049 + \dots = 1123816$$

Calculamos el valor Teórico de la fórmula (5) pero antes resolvemos la formula (1)

Para hallar el valor de los grados de libertad.

$$(fil - 1) * (col - 1) = (6 - 1) * (2 - 1) = 5$$

chisq.test(Datos1\$mot\_viam, Datos1\$tip\_movi)

Pearsons Chi-squared test

data: Datos1\$tip\_movi and

Datos1\$mot\_viam

X-squared= 1123816, df= 5, p-values < 2.2 e-16

En la Fig2. Test de chi-cuadrado realizado en Rstudio se puede constatar que si coinciden con el cálculo matemático.

$$\chi^2_{0,05,5} = 11.0705$$

La decisión es:  $\chi^2_{exp} > \chi^2_{0,05,5} = 1123816 > 11.0705$  se rechaza mediante el contraste de chi -cuadrado, por ende, la variable motivo de viaje si depende del tipo de movimiento y si se verá afectada por la entrada y la salida

## Variable subcontinente procedencia-destino vs tipo de movimiento

**Tabla 9.** Tabla de contingencia de las variables subcontinente destino-procedencia y tipo de movimiento.

	Entrada	Salida	Total
Africa Central	40	15	55
Africa Meridional	31	51	82
Africa Occidental	24	22	46
Africa Oriental	18	76	94
Africa Septentrional	61	89	150
América Central	24940	24904	49844
América del Septentrional	135767	143113	278880
América del Sur	294428	275759	570187
Antártida	0	5	5
Asia Sudoriental	268	253	521
Asia Central	6	5	11
Asia Meridional	417	294	711
Asia Occidental	482	756	1238
Asia Oriental	2121	3636	5757
Australia y Nueva Zelandia	427	616	1043
Europa Meridional	26807	33926	60733
Europa Occidental	10411	9989	20400
Europa Oriental	854	965	1819
Europa Septentrional	3255	4564	7819
Melanesia	4	8	12
Micronesia	34	30	64
Polinesia	40	44	84
Sin especificar	115	330	445
Total	500550	499450	1000000

En la presente tabla, se presentan los datos tabulados de la población en base al subcontinente de procedencia-destino y el tipo de movimiento siendo el mayor movimiento de salida del Ecuador a países de América del sur.

**Tabla 10.** Tabla de frecuencias esperadas obtenida una vez ya aplicada la formula a cada celda.

	Entrada	Salida	Total
África Central	27.5302	27.4697	55
África Meridional	41.0451	40.9549	82
África Occidental	23.0253	22.9747	46
África Oriental	47.0517	46.9483	94
África Septentrional	75.0825	74.9175	150
América Central	24949.4142	24894.5858	49844
América del Septentrional	139593.384	139286.616	278880
América del Sur	285407.102	284779.89	570187
Antártida	2.50275	2.49725	5
Asia Sudoriental	260.7865	260.2134	521
Asia Central	5.5060	5.4939	11
Asia Meridional	355.89105	355.1089	711
Asia Occidental	619.6809	618.3191	1238
Asia Oriental	2881.6663	2875.3336	5757
Australia y Nueva Zelandia	522.0736	520.9263	1043
Europa Meridional	30399.9031	30333.0968	60733
Europa Occidental	10211.22	10188.78	20400
Europa Oriental	910.50045	908.4995	1819
Europa Septentrional	3913.8004	3905.1995	7819
Melanesia	6.0066	5.9934	12
Micronesia	32.0352	31.9648	64
Polinesia	42.0462	41.9538	84
Sin especificar	222.7447	222.25525	445
Total	500550	499450	1000000

Tomando en cuenta la tabla anterior, una vez aplicada la fórmula de frecuencia esperada a cada uno de los datos podemos determinar que los resultados son los mismos lo cual nos indica que las frecuencias esperadas son iguales al total que se había planteado.

$H_0$ : La variable subcontinente procedencia-destino es independiente al tipo de movimiento.

$H_1$ : La variable subcontinente procedencia-destino si depende al tipo de movimiento.

Aplicamos las componentes del chi-cuadrado de la formula (6).

$$x^2 = \sum_{ij} \frac{(40-27.5302)}{27.5302} = 5.65$$

Luego se aplicó el estadístico de contraste de la formula (7).

$$X_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(40-27.5302)}{27.5302} = 5.65...+=1002797$$

Calculamos el valor Teórico de la fórmula (5) pero antes resolvemos la formula (1)

Para hallar el valor de los grados de libertad.

$$(fil - 1) * (col - 1) = (23 - 1) * (2 - 1) = 22$$

chisq.test(Datos1\$subcont\_prod, Datos1\$tip\_movi)

Pearsons Chi-squared test

data: Datos1\$tip\_movi and Datos1\$subcont\_prod

X-squared= 1002797, df= 22, p-values < 2.2 e-16

En la Fig2.Test de chi-cuadrado realizado en RStudio se puede constatar que si coinciden con el cálculo matemático.

$$X_{0.05,22}^2 = 21.337.$$

La decisión es:  $X_{exp}^2 > X_{0.05,22}^2 = 1002797 > 21.337$  se rechaza mediante el contraste de chi-cuadrado, por ende, la variable de subcontinente de procedencia si depende del tipo de movimiento y si se verán influenciados dependiendo del subcontinente y el tipo de movimiento donde de los pasajeros.

### Variable tipo de nacionalidad vs tipo de movimiento

**Tabla 11.** Tabla de contingencia de las variables tipo de nacionalidad y tipo de movimiento.

	Entrada	Salida	Total
Ecuatoriano	182971	195543	378514
Extranjero	317579	303907	621486
Total	500550	499450	1000000

Como se evidencia en esta tabla los extranjeros son los que más tienen entradas y salidas en el Ecuador.

**Tabla 12.** Tabla de frecuencias esperadas una vez ya aplicada la formula a cada.

	Entrada	Salida	Total
Ecuatoriano	189465.18	189048.81	378514
Extranjero	311084.81	310401.18	621486
Total	500550	499450	1000000

Plantemos nuestras hipótesis que pondremos a prueba, nos permitirá ver la dependencia o independencia de las variables analizadas.

$H_0$ : La variable tipo de nacionalidad es independiente al tipo de movimiento

$H_1$ : La variable tipo de nacionalidad si depende del tipo de movimiento

Aplicamos las componentes del chi-cuadrado de la formula (6).

$$x^2 = \sum_{ij} \frac{(182971-189465.18)}{189465.18} = 222.59$$

Con esta fórmula comparamos la frecuencia obtenida en la tabla de contingencia con la frecuencia que se obtuvo en la tabla de frecuencias esperadas, obteniendo como respuesta 222.59.

La tabla adjunta presenta el promedio de clientes que esperan a ser atendidos especificados por día e intervalos de tiempo, presenta una afluencia normal en el servicio de pago por lo que no hay novedades o problemas en él.

Luego se aplicó el estadístico de contraste de la formula (7).

$$X_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(182971-189465.18)}{27.5302} = 222.59...+=1000827$$

Se procede a calcular el valor experimentar o más conocido como estadístico de contraste el cual nos arrojó un valor de 1000827. Calculamos el valor Teórico de la fórmula (5) pero antes resolvemos la formula (1)

Para hallar el valor de los grados de libertad

$$(fil - 1) * (col - 1) = (2 - 1) * (2 - 1) = 1$$

chisq.test(Datos1\$tip\_naci, Datos1\$tip\_movi)

Pearsons Chi-squared test

data: Datos1\$tip\_naci and Datos1\$tip\_movi  
X-squared= 100827, df= 1, p-values < 2.2 e-16

En la Fig2.Test de chi-cuadrado realizado en RStudio donde podemos constatar que si nos coinciden con el cálculo matemático.

$$X_{0.05,1}^2 = 3.8415$$

La decisión es:  $X_{exp}^2 > X_{0.05,1}^2 = 1000827 > 3.8415$  sea aceptada como hipótesis nula por ende tipo de nacionalidad no depende del tipo de movimiento.

### Variable sexo vs tipo de movimiento

**Tabla 13.** Tabla de contingencia de las variables sexo y tipo de movimiento.

	Entrada	Salida	Total
Hombre	268756	265842	534598
Indefinido	28	28	56
Mujer	231766	233580	465346
Total	500550	499450	1000000

Como se evidencia en esta tabla la variable hombre es la que tiene más entradas y salidas en el Ecuador.

**Tabla 14.** Tabla de frecuencias esperadas una vez ya aplicada la formula a cada celda.

	Entrada	Salida	Total
Hombre	267593.02	267004.47	534598
Indefinido	28.03	27.96	56
Mujer	232928.94	232417.05	465346
Total	500550	499450	1000000

Plantemos nuestras hipótesis que pondremos a prueba, nos permitirá ver la dependencia o independencia de las variables analizadas.

$H_0$  : La variable sexo es independiente al tipo de movimiento.

$H_1$  : La variable sexo si depende del tipo de movimiento.

Aplicamos las componentes del chi-cuadrado de la formula (6).

$$\chi^2 = \sum_{ij} \frac{(268756 - 267593.02)^2}{267593.02} = 5.05$$

Con esta fórmula comparamos la frecuencia obtenida en la tabla de contingencia con la frecuencia que se obtuvo en la tabla de frecuencias esperadas, obteniendo como respuesta 5.05

Luego se aplicó el estadístico de contraste de la formula (7)

$$\chi_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(268756 - 267593.02)^2}{267593.02} = 5.05 \dots + = 21.76$$

Se procede a calcular el valor experimental o más conocido como estadístico de contraste el cual nos arrojó un valor de 21.76.

Calculamos el valor Teórico de la fórmula (5) pero antes resolvemos la formula (1)

Para hallar el valor de los grados de libertad.

$$(fil - 1) * (col - 1) = (3 - 1) * (2 - 1) = 2$$

> chisq.test(Datos1\$sex\_migr, Datos1\$tip\_movi)

Pearsons Chi-squared test

data: (Datos1\$sex\_migr and Datos1\$tip\_movi)

X-squared = 21.76, df= 2, p-value < 2.2 e-16

En la Fig2. Test de chi-cuadrado realizado en RStudio donde podemos constatar que si nos coinciden con el cálculo matemático.

En la Fig2. Test de chi-cuadrado realizado en RStudio se puede constatar que si coinciden con el cálculo matemático

$$\chi_{0.05,2}^2 = 5.9915$$

La decisión es:  $\chi_{exp}^2 > \chi_{0.05,2}^2 = 21.76 > 5.9915$  sea acepta como hipótesis nula por ende sexo no depende del tipo de movimiento.

### Variable año de movimiento vs tipo de movimiento

**Tabla 15.** Tabla de contingencia de las variables sexo y tipo de movimiento.

	Entrada	Salida	Total
2018	500550	499450	1000000

Como se evidencia en esta tabla hubo mas entradas que salidas en el 2018.

## Conclusiones

Se puede decir que Ecuador es un país con un alto número de migrantes que salen del país, como también migrantes que ingresan, ya sea por turismo, residencia, estudios, economía y otros. Podemos saber también que son los extranjeros con un total de 62% los que tienen más entrada y salidas del Ecuador, esto puede entenderse, ya que el país cuenta con un sector turístico que atrae la atención de cientos y miles de personas, y también por su inclinación a la mano de obra internacional, puesto que en muchos trabajos prefieren contratar a personas con estudios realizados en otros países, esto implica un descenso en la tasa de trabajo para las personas de nacionalidad ecuatoriana.

Por consiguiente, nos encontramos con un total de 38% personas que salen del país, con nacionalidad ecuatoriana por motivo de residencia. De esta forma podemos concluir que Ecuador es un país con un gran porcentaje de migrantes buscando un nuevo comienzo, así como nuevas oportunidades tanto para el como para su familia.

## Referencias bibliográficas

- [1] Varela Llamas Rogelio, «Migración interna en Mexico y causas de su movilidad perfiles Latinoamericanos». 2017
- [2] José Rafael Alvarado-López, «Migración interna y urbanización sin eficiencia en países en desarrollo: evidencia para Ecuador,» 2017.
- [3] López-Calleja Hiort-Lorenzen, Cristina; Morejón Seijas, Blanca, «Envejecimiento demográfico y migración interna de personas con 60 años y más de edad en Cuba y países de América Latina Población y Salud en Mesoamérica,» 2015.
- [4] Khoudour-Castéras, David «Efectos de la migración sobre el trabajo infantil en Colombia» 2009.
- [5] Lumeri U, « Los flujos de migración interna a través del analisis de redes» 2017
- [6] Roman P, Gonzales J; «Análisis de la Migración a través de encuestas » 2014
- [7] Sin Nombre, « Tablas de contingencia », 2014
- [8] Sanchez Miguel, « Uso Metodológico de las tablas de contingencia », 2010
- [9] Veletes C, « The R User conference », 2011.
- [10] F. Prof. Nieves, «prueba de chi cuadrado (x) para una sola muestra,» Moreau de Justo, vol. UNIDAD 3, n° 1, pp. 1-9.
- [11] S. Inzunza cazaes y j. Vidal jiménez ramírez, «caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis, » Revista Latinoamericana de Investigación en Matemática Educativa, vol. 16, n° 2, pp. 179-211, 2013.

- [12] J. L. De La Cruz-Oré, «¿Qué significan los grados de libertad?,» Revista Peruana de Epidemiología, vol. 17, n° 2, pp. 1-6, 2013.
- [13] Granada, Universidad de, «El estadístico Chi-cuadrado y contrastes asociados,» Universidad de Granada.
- [14] D. P. Bernal Cortés y D. Cardona Rivas, «caracterización de la mortalidad perinatal en manizales, colombia, 2009-2012,» Revista Hacia la Promoción de la Salud, vol. 19, n° 2, pp. 66-80, 2014.
- [15] M. Zapata Carlos y M. Villegas Sandra, «Reglas de consistencia entre modelos de requisitos de un metodo,» Universidad EAFIT, pp. 40-59, 2006.
- [16] Cevallos Torres Lorenzo, Guijarro Rodríguez Alfonso, López Domínguez Leili, «factores que inciden en el mal uso de la información en trabajos de investigación científica,» Revista Didasc@lia, ISSN 2224-2643.
- [17] Cevallos Torres Lorenzo, Valencia Martínez Nelly, Barro Morales Roosevelt, «Análisis Estadístico Univariado,» Compas Editorial.
- [18] Cevallos Torres Lorenzo, «Diseño de un Sistema de Mejora Continua aplicado a los Cursos de Nivelación Universitaria, basado en el Análisis
- [19] Cevallos Torres Lorenzo, Guijarro Rodríguez Alfonso, Torres Villegas Ignacia, «relación teoría-práctica para la investigación de operaciones: caso práctico en modelos de programación lineal,» Revista Didasc@lia, pp. 66-80, 2014.
- [20] Cevallos Torres, Lorenzo Jovanny, Ortiz Aguilar, Wilber, Muñoz Aveiga., Edelmary de Lourdes « Uso de las tecnologías de la información y la comunicación como estrategia de enseñanza - aprendizaje de la estadística universitaria,» Editorial Academica Universitaria.
- [21] Cevallos Torres Lorenzo, Guijarro Rodríguez Alfonso, Alarcón Cáceres José Mauricio, Delgado Veloz Geomayra Stefania, Barrera Rivera Mirella Katuska, Alvarado Flores Ronald Mauricio, « Análisis estadístico de correlación entre las dosis de eritropoyetina y el nivel de hemoglobina en pacientes con insuficiencia renal crónica,» » Revista Facultad de Ciencias Medicas, pp. 66-80, 2014.