

Modelo estadístico para asociar variables que permitan determinar el grado de dependencia lineal

Statistical model to associate variables for determining the degree of linear dependence

Xavier Álvarez¹, Steeven Ambuludí², y María Tola Romero³

RESUMEN

En los últimos años el flujo de migraciones ha aumentado de manera exponencial, esto podría generar conflictos en los países que acogen a los migrantes, es por ello que mediante el presente trabajo de investigación y a través de una base de datos obtenida del Instituto Nacional de Estadística y Censos (INEC) acerca de las entradas y salidas de migrantes de este país, se consideró una muestra de un millón de personas, mediante este análisis se busca determinar el porcentaje de extranjeros que ingresan al Ecuador y también los países de procedencia más frecuentes de los inmigrantes, para esto se implementa un análisis estadístico implementando tablas de contingencias y la prueba de chi cuadrado, realizado en el lenguaje de programación R. Este trabajo arrojó como resultados que el 32.2% de migrantes que ingresan al país son extranjeros, también que los países con mayor recurrencia de salida de migrantes hacia este país son: Venezuela, Colombia, Estados Unidos y Perú. Además, se encontró que estas migraciones están relacionadas a factores como turismo y residencia. En conclusión, en este trabajo se pudo verificar que las tablas de contingencia siguen siendo un buen instrumento estadístico para determinar la dependencia que pueda existir entre dos variables determinadas.

Palabras clave: Inmigrantes, Análisis estadístico, Tablas de contingencia, Relación de variables.

ABSTRACT

In recent years the flow of migration has increased exponentially, this could generate conflicts in the countries that host migrants, which is why through this research work and through a database obtained from the National Statistics Institute and Censuses (INEC) about the entry and exit of migrants from this country, a sample of one million people was considered, this analysis seeks to determine the percentage of foreigners entering Ecuador and also the most frequent countries of origin of For this, a statistical analysis is implemented by implementing contingency tables and the chi-square test, carried out in the programming language R. This work showed that 32.2% of migrants entering the country are foreigners, also that the Countries with the highest recurrence of migrants leaving to this country are: Venezuela, Colombia, United States, and Peru. In addition, it was found that these migrations are related to factors such as tourism and residence. In conclusion, in this work, it was possible to verify that the contingency tables are still a good statistical instrument to determine the dependence that may exist between two determined variables.

Keywords: Immigrants, Statistical analysis, Contingency tables, Variable relationships.

Fecha de recepción: Septiembre 5, 2019.

Fecha de aceptación: Enero 14, 2020.

Introducción¹ 2

En la actualidad las crisis económicas y los conflictos sociales por los que atraviesan algunos países suelen ser los factores más influyentes al momento de que una persona decida migrar lejos de su hogar buscando seguridad, estabilidad económica entre

otras cosas. Son tantas las migraciones que se han dado en los últimos años hacia este país que la población nativa se ve afectada puesto que aumenta la mano de obra barata y reduce la plaza de empleos. Es por ello, que a través de esta investigación se busca dar a conocer el porcentaje de extranjeros que ingresan a Ecuador y los países de procedencia más usuales de los mismos, así también la relación que existe entre dichas variables.

Trabajos Relacionados

Vásquez-De Kart et al. (Vásquez-De Kartow, Castillo-Durán, & Lera M., 2015) realizan un análisis estadístico para ver la relación de las variables sociodemográficas con la tasa de inmigración de población pediátrica en países de América Latina. Este análisis fue realizado mediante el uso de Excel en el cual se aplicó el método de correlación lineal de Pearson y la prueba de Fisher para evaluar las diferencias en la tasa de migración. Obtuvieron como

¹ Estudiante de Ingeniería en Sistemas Computacionales. Universidad de Guayaquil, Ecuador. E-mail: xavier.alvarezp@ug.edu.ec

² Estudiante de Ingeniería en Sistemas Computacionales. Universidad de Guayaquil, Ecuador. E-mail: steeven.ambuludic@ug.edu.ec

³ Estudiante de Ingeniería en Sistemas Computacionales. Universidad de Guayaquil, Ecuador. E-mail: maria.tolar@ug.edu.ec

Como citar: Álvarez, X., Ambuludí, S., & Tola Romero, M. (2020). Modelo estadístico para asociar variables que permitan determinar el grado de dependencia lineal. *Ecuadorian Science Journal*. 4(1), 8-13.
DOI: <https://doi.org/10.46480/esj.4.1.51>

resultado que la tasa neta de migración fue positiva solo para 5 países de América Latina: Costa Rica, Panamá, Venezuela, Chile y Argentina.

De acuerdo con el trabajo de Guevara Santana (Guevara Santana, 2017) el flujo migratorio se puede observar de forma externa e interna y es analizado desde cualquier punto de vista tanto político, social o económico. En este trabajo se hace un análisis para observar la relación que existe entre las políticas migratorias y el retorno de los migrantes ecuatorianos. Para llevar a cabo este análisis los investigadores emplearon el método de correlación de Pearson y el estadístico de prueba T de student mediante el cual obtuvieron como resultado que existe una relación significativa entre las políticas migratorias y el retorno de migrantes ecuatorianos.

Como observamos en los autores Vásquez-De Kartzow (Vásquez-De Kartzow et al., 2015) y Guevara Santana (Guevara Santana, 2017), utilizaron el método de correlación de Pearson para el analizar la relación de sus variables, sin embargo, este método solo nos sirve para medir variables cuantitativas por eso en este trabajo se hará uso de tablas de contingencia ya que nos ayuda a medir tanto variables cuantitativas como cualitativas. Además, el trabajo de Vásquez-De Kartzow (Vásquez-De Kartzow et al., 2015) está hecho en Excel y este será realizado en R.

Parra Vera (Parra Vera, 2016) busca analizar el impacto económico de las migraciones en Ecuador y Bolivia con la finalidad de observar los orígenes de las tendencias migratorias y su relación con la naturaleza esperada de las remesas de los emigrantes. Para ello utiliza un análisis estadístico de regresión lineal múltiple y una reducción factorial por método de componentes principales mediante el uso del software IBM-SPSS versión 24. Obtuvieron que, en el caso de Bolivia la naturaleza de las remesas de los migrantes es oportunista tanto en el análisis con Estados Unidos y España. De similar manera en Ecuador, pero solo en el caso de España. Sin embargo, a pesar de que la herramienta SPSS es muy buena para el análisis estadístico de datos en nuestra investigación se usara el software R debido a que SPSS presenta limitaciones en sus funciones a la hora de realizar problemas complejos mientras que R es actualmente el software más potente y completo del mercado en cuanto al análisis estadístico.

En Nieves (Nieves, 2015) el autor realiza una investigación sobre la correlación que existe entre la migración y el desarrollo humano. Las migraciones son un tema delicado en los países receptores donde es vista como uno de los principales problemas y es explotada por algunos grupos políticos. Mediante este trabajo obtuvieron como resultado que si en lugar de contratar a un experto europeo para desarrollar un trabajo de cooperación en un país empobrecido, se contratara a un nacional en la diáspora, con un nivel de cualificación equivalente, se favorecerá evidentemente el retorno. Desde el punto de vista de la dependencia de los territorios periféricos, el subdesarrollo es la causa de las migraciones, que a su vez refuerzan las condiciones estructurales de pobreza y dependencia en las que se encuentra el país de origen.

Materiales y Métodos

En este apartado se detalla los materiales y métodos que se implementaran a lo largo de este trabajo, para dar a conocer el porcentaje de extranjeros que ingresan al Ecuador, se

implementaran tablas de contingencia utilizando la prueba de chi cuadrado.

Estadística

Para Cevallos Torres et al. (Cevallos Torres, MARTINEZ, AMERICA, & BARROS MORALES, 2017), la estadística es la ciencia en donde se obtienen, analizan, procesan y muestran datos referentes a un fenómeno que presentan variabilidad o incertidumbre para su estudio metódico, con objeto de deducir las características de una población objetivo y poder actuar de esa forma sobre los mismos, tomar decisiones u obtener conclusiones.

Tablas de contingencia

Se sabe que la información proporcionada por una tabla bidimensional puede expresarse en términos diversos: frecuencias absolutas conjuntas, relativas conjuntas, condicionadas de una variable a valores de la otra. Además, puede derivarse el comportamiento unidimensional de las variables implicadas mediante las distribuciones marginales Nieves (Nieves, 2016).

La tabla bidimensional recibe el nombre de tabla de contingencia cuando las características en estudio no son cuantitativas.

La tabla de contingencia es un medio particular de representar simultáneamente dos caracteres observados en una misma población, si son discretos o continuos reagrupados en clases. Los dos caracteres son x e y, el tamaño de la muestra es n. Las modalidades o clases de x se escribirán C1, ..., Cr, las de y, d1, ... ds. Se denota:

1. n_{kk} el efectivo conjunto de C_h y d_k : es el número de individuos para los cuales x toma el valor C_h y el valor d_k
2. $n_h = \sum_{k=1}^r n_{hk}$ el efectivo marginal de C_h : es el número de individuos para los cuales x toma el valor C_h .
3. $n_k = \sum_{h=1}^r n_{hk}$ el efectivo marginal de d_k : es el número de individuos para los cuales y toma el valor d_k .

Se representan estos valores en una tabla de doble entrada, llamada tabla de contingencia:

Tabla 1. Tabla de contingencia general.

x/y	d1	dk	ds	Total
C1	n11	n1k	n1s	n1
.....
Ch	nh1	nhk	nhS	nh
.....
Cr	nr1	nrk	nrS	nr
Total	n1	nk	ns	n

Prueba de chi cuadrado

Según Manzano (Manzano, 2014), la prueba chi-cuadrado de contingencia sirve para comprobar la independencia de frecuencias entre dos variables aleatorias, X e Y.

Las hipótesis contrastadas en la prueba son:

- **Hipótesis nula:** X e Y son independientes.
- **Hipótesis alternativa:** X e Y no son independientes (No importa cuál sea la relación que mantengan ni el grado de esta.

La condición de independencia es: X e Y son independientes si y sólo si para cualquier pareja de valores x e y la probabilidad de que X tome el valor x e Y el valor y, simultáneamente, es igual al producto de las probabilidades de que cada una tome el valor correspondiente.

$$X \text{ e } Y \text{ son independientes} \Leftrightarrow \forall x, y f(x, y) = f(x) \cdot f(y)$$

Por tanto, todo lo que necesitamos serán unas estimas de las funciones de probabilidad de ambas variables por separado (f(x) y f(y)) y de la función de probabilidad conjunta (f(x,y)). Empezaremos la prueba tomando una muestra de parejas de valores sobre la que contaremos la frecuencia absoluta con la que aparece cada combinación de valores (xi, yj) o de grupos de valores (i,j) (Oij). La tabla siguiente, en la que se recogen estos datos, es en realidad nuestra estimación de la función de probabilidad conjunta multiplicada por el número total de datos (T).

Para obtener las estimas de las funciones de probabilidad marginales debemos sumar por filas y por columnas los valores de las frecuencias conjuntas. Las sumas de filas (Fi) son, en cada caso, el número de veces que hemos obtenido un valor de X (xi) en cualquier combinación con distintos valores de Y, es decir, son nuestra estima de la función de probabilidad de X multiplicada por el número total de observaciones; análogamente, las sumas de columnas (Cj) son nuestra estima de la función de probabilidad de Y multiplicada por el número total de observaciones.

El número total de observaciones lo podemos obtener como la suma de todas las frecuencias observadas o, también, como la suma de las sumas de filas o de las sumas de columnas:

$$T = \sum_{ij} O_{ij} = \sum_i F_i = \sum_j C_j \quad (1)$$

Así pues, si las variables fueran independientes debería cumplirse que:

$$\forall i, j \frac{O_{ij}}{T} = \frac{F_i}{T} \cdot \frac{C_j}{T} = \frac{F_i \cdot C_j}{T^2} \quad (2)$$

Podemos convertir la ecuación anterior a frecuencias absolutas multiplicando por T:

Si X e Y son independientes, Oij debe ser igual a $\frac{F_i \cdot C_j}{T}$ y, por tanto, bajo la hipótesis de independencia $\frac{F_i \cdot C_j}{T}$, es el valor esperado de Oij (Eij). Si las variables son independientes, es decir, si las frecuencias Eij son realmente los valores esperados de las frecuencias Oij, se puede calcular un parámetro que depende de ambas que tiene distribución chi-cuadrado.

$$\sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx \chi^2 \quad (3)$$

Por otra parte, si las variables no son independientes, las diferencias entre las series de frecuencias observadas y esperadas serán mayores que las atribuibles al efecto del azar y, al estar

elevadas al cuadrado en el numerador de la expresión anterior, ésta tenderá a ser mayor que lo que suele ser el valor de una variable chi-cuadrado.

$$\sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \gg \chi^2$$

Por tanto, el parámetro anterior ser el estadístico de la prueba de hipótesis y la región crítica se encontrar siempre en la cola derecha de la distribución chi-cuadrado. Nuevamente, esta prueba será siempre de una sola cola. Estadístico (5) :raste.

$$\chi^2 \approx \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Se acepta la hipótesis nula si $\chi^2 < \chi^2_{1-\alpha, \nu}$, el percentil 1 - α de la distribución chi-cuadrado con ν grados de libertad.

Tal como ocurría en la prueba anterior lo corriente es que queramos demostrar que dos variables son independientes, es decir, que, habitualmente, nos veremos obligados a colocar nuestra hipótesis en la hipótesis nula.

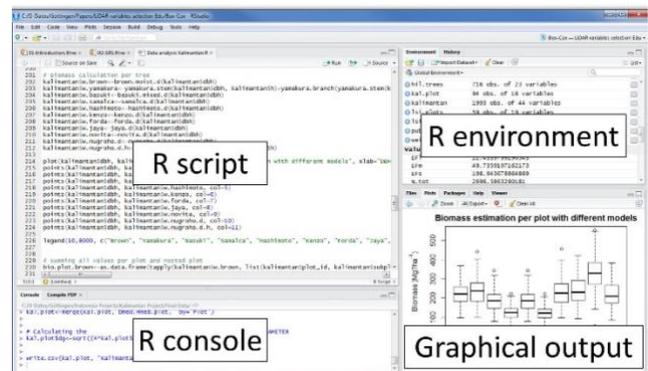
Prueba de chi cuadrado

Para Ruiz-Ruano & Puga (Ruiz-Ruano & Puga, 2016), R es un lenguaje Orientado a Objetos: bajo este complejo termino se esconde la simplicidad y flexibilidad de R. Es un lenguaje interpretado (como Java) y no compilado (como C, C++, Fortran, Pascal...), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir ejecutables.

Como segunda medida, la sintaxis de R es muy simple e intuitiva. Por ejemplo, una regresión lineal se puede ejecutar con el comando lm (y, x). Para que una función sea ejecutada en R debe estar siempre acompañada de paréntesis, inclusive en el caso que no haya nada dentro de los mismos (por ej., ls ()). Si se escribe el nombre de la función sin los paréntesis, R mostrara el contenido (código) mismo de la función.

El entorno RStudio está definido por una serie de ventanas que permiten realizar distintas tareas. Las más importantes son:

Figura 1. Ventana del programa RStudio.



Consola R: Esta ventana interactiva permite escribir directamente código R, y muestra también la salida del código que ejecuta. El código escrito en la consola R no se guarda directamente, aunque se puede acceder a las órdenes anteriores con la flecha hacia arriba.

Entorno R (environment): En este espacio se muestra el conjunto de objetos que R tiene cargados en memoria. Esto incluye conjuntos de datos, variables, vectores, funciones, etc. También hay otras solapas que permiten acceder entre otros, a la historia: el código ejecutado recientemente.

Salida gráfica: Este espacio tiene varias solapas. La solapa plots muestra los gráficos creados con R. Hay otras solapas: files permite explorar los archivos en la carpeta de trabajo, incluido cambiar ésta. La solapa Packages permite comprobar qué paquetes de R están instalados, cargarlos e instalar nuevos. La pestaña help permite acceder a la ayuda de R integrada. Por último, la pestaña viewer contiene los resultados producidos por R que corresponden a otro tipo de objetos, como ficheros html.

Editor R (Script): Cuando abrimos por primera vez esta ventana no aparece. Como su nombre sugiere, aquí tienes espacio para editar código. Los ficheros de código de R son siempre ficheros de texto, que pueden tener la extensión .r (sólo código) o .rmd (código combinado con generación de documentos). Para ejecutar códigos seleccionamos las líneas de código y hacemos Ctrl + Enter o hacemos clic en el botón run. En este espacio también se abren pestañas para visualizar objetos de R cuando hacemos clic en la ventana del entorno R.

Caso de estudio

El tema de la migración es de interés en los últimos tiempos debido al alto índice migratorio que ha surgido en los últimos años, Ecuador se ha convertido en un punto de escala para los viajeros, ya sea por estudios, turismo, trabajo, etc. Los altos índices migratorios representan la necesidad social de cada individuo que busca mejorar sus oportunidades de vidas, o solo hacer viaje turístico dentro del Ecuador Cevallos-Torres & Botto-Tobar (Cevallos-Torres & Botto-Tobar, 2019).

Población

Para este trabajo la población de estudio, se tomó una muestra de más de 1 millón de personas. Esta información fue proporcionada por la base de datos que tiene por nombre Control de entradas y salidas de los migrantes en Ecuador, la misma que fue obtenida a través del Instituto Ecuatoriano de Estadística y Censo (INEC).

Por lo que para el análisis estadístico de este estudio se tomó en consideración las siguientes variables:

Tabla 2. Variables utilizadas.

Nombre de las variables					
Edad	Tipo de movimiento	Tipo de nacionalidad	Sexo	Motivo de viaje	Medio de transporte

Resultados y Discusiones

En esta parte se mostrará los resultados obtenidos mediante el análisis estadístico con tabla de contingencias por medio la prueba de chi cuadrado.

Algoritmo 1:

```
tablaPaises<-table(Base.de.datos.estadistica.Final$tip_movi,Base.de.datos.estadistica.Final$tip_naci)
print(totPercents(tablaPaises))
```

Tabla 3. Porcentajes entre el tipo de movimiento de migración y los tipos de persona.

Ecuatoriano	Extranjero	TOTAL
-------------	------------	-------

Entrada	17.9	32.2	50.1
Salida	19.3	30.7	49.9
TOTAL	37.1	62.9	100.0

De acuerdo con los resultados de los porcentajes de la tabla de contingencia se puede declarar que de la base de datos solo un 32.2% pertenece a los extranjeros que ingresan al país.

Algoritmo 2:

```
tablaContingenciaI<-table(BaseDatosReducida$nac_migr,BaseDatosReducida$mot_viam)
print(totPercents(tablaContingenciaI))
```

Tabla 4. Porcentajes entre la nacionalidad de los extranjeros y el motivo de viaje.

	Estudios	Eventos	Negocios	Otros	Residencia	Turismo	Total
Colombia	0.0	0.5	0.6	0.8	1.8	15.3	19.1



Estados Unidos	0.1	0.4	0.1	0.5	1.2	9.0	11.3
Otros	0.0	2.4	1.1	3.0	4.2	22.2	32.9
Perú	0.0	0.2	0.2	0.4	0.5	8.0	9.3
Venezuela	0.0	0.2	0.1	0.7	1.1	25.3	27.4
TOTAL	0.2	3.7	2.1	5.4	8.8	79.8	100.0

Hipótesis:

H0: La nacionalidad de los extranjeros es independiente del motivo de viaje.

H1: La nacionalidad de los extranjeros no es independiente del motivo de viaje.

Algoritmo 3:

```
PruebaI<-chisq.test(tablaContingenciaI, correct=FALSE)
```

Figura 2. Histograma de los países con mayor frecuencia.

```
print(PruebaI)
```

Prueba de Chi-Cuadrado en R

data: tablaContingenciaI

X-squared = 73602, df = 920, p-value < 2.2e-16

Se puede verificar el resultado aplicando la fórmula matemática de chi-cuadrado:

Figura 3. Histograma de los países con mayor frecuencia.

$$X_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(41 - 79.07)^2}{79.07} = 18.33 + \dots = 73602$$

Luego se obtiene el grado de libertad aplicando su respectiva fórmula matemática:

$$df = (185 - 1) * (6 - 1) = 184 * 5 = 920$$

Por último, se encuentra el valor crítico a utilizar:

$$X_{0.05,920}^2 = 658.09$$

De acuerdo con los resultados de la prueba de chi-cuadrado se puede verificar que la hipótesis nula (**H₀**) es rechazada debido a que el valor obtenido, 73602, es mayor al valor crítico que corresponde a 658.09. También se puede afirmar que el porcentaje de extranjeros que entran al Ecuador y vienen por motivo turístico es de 79.8%, un 8.8% de extranjeros entran al Ecuador por motivo de residencia, y solo un 0.2%, 3.7%, 2.1%, 5.4% de extranjeros entran al Ecuador por motivos de estudios, eventos, negocios, y otros, respectivamente. Y por último se puede afirmar que el mayor porcentaje de extranjeros que vienen al Ecuador son de Venezuela con un 27.4%, seguido con un 19.1% que corresponde a los extranjeros que son de Colombia, y con un 11.3% correspondiente a los extranjeros que llegan a Ecuador desde Estados Unidos.

Algoritmo 4:

```
tablaContingencia2<table(BaseDatosReducida$mot_viam,BaseDa
tosReducida$sex_migr)
print(totPercents(tablaContingencia2))
```

Dependencia entre el sexo de los extranjeros y el motivo de viaje

Tabla 5. Porcentajes entre el sexo de los extranjeros y el motivo de viaje.

	Hombre	Mujer	TOTAL
Estudios	0.1	0.1	0.2
Eventos	2.8	0.9	3.7
Negocios	1.7	0.3	2.1
Otros	3.8	1.6	5.4
Residencia	4.8	4.1	8.8
Turismo	43.4	36.4	79.8
TOTAL	56.5	43.5	100.0

Hipótesis:

H₀: El sexo de los extranjeros es independiente respecto al motivo de viaje.

H₁: El sexo de los extranjeros no es independiente respecto al motivo de viaje.

Algoritmo 5:

```
Prueba2<-chisq.test(tablaContingencia2, correct=FALSE)
print(Prueba2)
```

Prueba de Chi-Cuadrado en R

data: tablaContingencia2
X-squared = 3964.8, df = 5, p-value < 2.2e-16

Se puede verificar el resultado aplicando la fórmula matemática d e chi-cuadrado:

$$X_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(169 - 233.96)^2}{233.96} = 18.04 + \dots = 3964.8$$

Luego se obtiene el grado de libertad aplicando su respectiva fórmula matemática:

$$df = (6 - 1) * (2 - 1) = 5 * 1 = 5$$

Por último, se encuentra el valor crítico a utilizar:

$$X_{0.05,5}^2 = 11.07$$

De acuerdo con los resultados de la prueba de chi-cuadrado se puede verificar que la hipótesis nula (**H₀**) es rechazada debido a que el valor obtenido, 3964.8, es mayor al valor crítico que corresponde a 11.07. Y también se puede declarar que el porcentaje de hombres que entran al Ecuador es mayor al porcentaje de mujeres que llegan al país.

Algoritmo 6:

```
tablaContingencia3<table(BaseDatosReducida$e
dad,BaseDatosReducida$mot_viam)
print(totPercents(tablaContingencia3))
```

Tabla 6. Porcentajes entre la edad de los extranjeros y el motivo de viaje.

	Estudi os	Event os	Negoc ios	Otr os	Residen cia	Turis mo	Total
1 a 10	0.0	0.0	0.0	0.0	1.0	6.4	7.4
11 a 20	0.0	0.0	0.0	0.1	1.0	7.6	8.7
21 a 30	0.0	1.0	0.0	1.1	1.4	23.5	27.0
31 a 40	0.0	1.0	0.9	1.0	2.0	16.1	21.0
41 a 50	0.0	1.0	1.0	0.3	2.0	9.6	1
51 a 60	0.0	0.7	0.2	1.0	0.4	7.6	3.9 9.9
61 a 70	0.0	0.0	0.0	0.0	1.0	6.4	7.4
71 a 80	0.1	0.0	0.0	0.0	0.0	2.5	2.6
81 a 90	0.0	0.0	0.0	0.0	0.0	0.1	0.1
91 a 100	0.0	0.0	0.0	2.0	0.0		2.0
TOTAL	0.2	3.7	2.1	5.4	8.8	79.8	100.0

Hipótesis:

H₀: La edad de los extranjeros es independiente respecto al motivo de viaje.

H₁: La edad de los extranjeros no es independiente respecto al motivo de viaje.

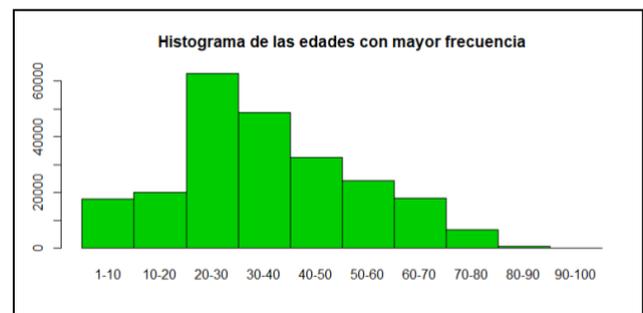


Figura 4. Histograma de las edades con mayor frecuencia.

Algoritmo 7:

```
Prueba3<-chisq.test(tablaContingencia3, correct=FALSE)
```

```
print(Prueba3)
```

Prueba de Chi-Cuadrado en R

```
data: tablaContingencia3
X-squared = 12810 df = 485, p-value < 2.2e-16
```

Se puede verificar el resultado aplicando la fórmula matemática de chi-cuadrado:

$$X_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(0 - 2.58)^2}{2.58} = 2.48 + \dots = 12810$$

Luego se obtiene el grado de libertad aplicando su respectiva fórmula matemática:

$$df = (98 - 1) * (6 - 1) = 97 * 5 = 485$$

Por último, se encuentra el valor crítico a utilizar:

$$X_{0.05,485}^2 = 553.13$$

De acuerdo con los resultados de la prueba de chi-cuadrado se puede verificar que la hipótesis nula (**H₀**) es rechazada debido a que el valor obtenido, 12810, es mayor al valor crítico que corresponde a 553.13. Y también se puede afirmar que el mayor porcentaje de extranjeros que vienen al Ecuador se encuentra entre un rango de 20 y 60 años.

Algoritmo 8:

```
tablaContingencia4 <- table(BaseDatosReducida$mot_viam, BaseDa
tosReducida$via_tran)
print(totPercents(tablaContingencia4))
```

Dependencia entre la vía de transporte y el motivo de viaje

Tabla 7. Porcentajes entre la vía de transporte y el motivo de viaje.

	Vía Aérea	Vía fluvial	Vía marítima	Vía terrestre	Total
Estudios	0.1	0.1	0.2	0.0	0.2
Eventos	2.8	0.9	3.7	0.0	3.7
Negocios	1.7	0.3	2.1	0.0	2.1
Otros	3.8	1.6	5.4	1.31	5.4
Residencia	4.8	4.1	8.8	1.2	8.8
Turismo	43.4	36.4	79.8	48.7	79.8
TOTAL	56.5	43.5	100.0	51.2	100.0

Hipótesis:

H₀: La vía de transporte es independiente respecto al motivo de viaje.

H₁: La vía de transporte no es independiente respecto al motivo de viaje.

Algoritmo 9:

```
Prueba4 <- chisq.test(tablaContingencia4, correct=FALSE)
print(Prueba4)
```

Prueba de Chi-Cuadrado en R

```
data: tablaContingencia4
X-squared = 64576, df = 15, p-value < 2.2e-16
```

Se puede verificar el resultado aplicando la fórmula matemática de chi-cuadrado:

$$X_{exp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(399 - 190.47)^2}{190.47} = 228.30 + \dots = 64576$$

Luego se obtiene el grado de libertad aplicando su respectiva fórmula matemática:

$$df = (6 - 1) * (4 - 1) = 5 * 3 = 15$$

Por último, se encuentra el valor crítico a utilizar:

$$X_{0.05,15}^2 = 25$$

De acuerdo con los resultados de la prueba de chi-cuadrado se puede verificar que la hipótesis nula (**H₀**) es rechazada debido a que el valor obtenido, 64576, es mayor al valor crítico que corresponde a 25. Y también se puede afirmar que el mayor porcentaje de extranjeros que vienen al Ecuador prefieren usar una vía terrestre con un 51.2%, seguido de un 46% que prefieren usar una vía aérea, un 2.6% que optan por usar una vía marítima, y, por último, de un 0.2% que deciden usar una vía fluvial.

Conclusiones

Según los resultados obtenidos mediante el análisis de la prueba de Chi-Cuadrado se logró conocer que solo un 32.2% de los extranjeros son los que ingresan al país y que los países con mayor el porcentaje de extranjeros que migran hacia Ecuador son Venezuela con un 27.4%, Colombia con un 19.1%, y Estados Unidos con un 11.3%. También se logró determinar los motivos más frecuentes por los cuales los extranjeros vienen a nuestro país; entre los cuales se destaca el turismo, la residencia, eventos sociales y otros. Por otra parte, se observa las principales vías de transporte por las cuales ingresan los extranjeros, que son las de vía terrestre y aérea, respectivamente. Y finalmente, se demostró que la mayoría de los extranjeros que ingresan se encuentran en un rango de edad de 20 a 60 años.

Referencias Bibliográficas

- Cevallos-Torres, L., & Botto-Tobar, M. (2019). Case study: Logistical behavior in the use of urban transport using the monte carlo simulation method. In *Problem-Based Learning: A Didactic Strategy in the Teaching of System Simulation* (pp. 97–110). Springer.
- Cevallos Torres, L. J., MARTINEZ, V., AMERICA, N., & BARROS MORALES, R. L. (2017). Análisis Estadístico Univariado. grupo Compás-Universidad de Guayaquil.
- Guevara Santana, D. J. (2017). *Análisis de la aplicación de políticas para retorno de migrantes en el Ecuador, Periodo 2012-2015*. UNIVERSIDAD DE GUAYAQUIL.
- Manzano, V. (2014). Chi cuadrado de pearson para dos variables nominales.
- Nievas, F. (2015). *Informe de Gestión*.
- Nievas, F. (2016). Pruebas de Chi Cuadrado para una sola muestra.
- Parra Vera, S. R. (2016). ANÁLISIS DE LA MIGRACIÓN EN LA COMUNIDAD ANDINA Y SUS MOTIVACIONES ECONÓMICAS. Universidad de Guayaquil.
- Ruiz-Ruano, A. M., & Puga, J. L. (2016). R COMO ENTORNO PARA EL ANÁLISIS ESTADÍSTICO EN EVALUACIÓN PSICOLÓGICA. *Papeles del Psicólogo*, 37(1), 74–79.
- Vásquez-De Kartzow, R., Castillo-Durán, C., & Lera M., L. (2015). Migraciones en países de América Latina. Características de la población pediátrica. *Revista Chilena de Pediatría*, 86(5), 325–330. <https://doi.org/10.1016/j.rchipe.2015.07.007>

