





Gráficos poligonales para el estudio de la correlación de unidades estadísticas

| Polygonal graphics for the study of the correlation of statistical units |

 **David A. Alfaro Alfaro**
david.alfaro.alfaro@mep.go.cr
Ministerio de Educación Pública
Grecia, Costa Rica

 **Luis G. Alfaro Alpizar**
luis.alfaro.alpizar@mep.go.cr
Ministerio de Educación Pública
Grecia, Costa Rica

 **Ana M. Vásquez Monge**
ana.vasquez.monge@mep.go.cr
Ministerio de Educación Pública
Grecia, Costa Rica

Recibido: 22 Febrero 2021

Aceptado: 23 Agosto 2021

Resumen: Este artículo contiene un análisis de las correlaciones entre múltiples unidades estadísticas de un estudio, mediante el uso de Gráficos Poligonales, los cuales están hechos mediante un complemento creado en Excel en cuya programación se utilizan comandos que filtran la información y crean el gráfico deseado. Este enfoque se basa en un nuevo procedimiento de representación mediante polígonos donde se evidencia el número de unidades estadísticas que coinciden con las mismas características en las variables analizadas. Los resultados obtenidos facilitan el análisis de metadatos en estudios estadísticos utilizando una hoja de cálculo.

Palabras Clave: correlación de unidades estadísticas, estadística, metadatos, multivariable

Abstract: This article contains an analysis of correlations between multiple statistical units of a study, through the use of Polygonal Graphs, which are generated from an add-in created in Excel, where programming commands are used to filter the information creating the desired graph according needs. This approach is based on a new procedure, showing data analysis represented in polygons graphs, where the number of statistical units that coincide with the same characteristics of the analyzed variables. These results obtained, facilitate the analysis of metadata in statistical studies using a spreadsheet.

Keywords: units statistical correlation, statistics, big data, multivariable—

1. Introducción

La correlación se define como “correspondencia o relación recíproca entre dos o más cosas o series de cosas” (Real Academia Española, s.f.) y en este artículo se estudiará la correlación entre unidades estadísticas, es decir, las que coinciden con las mismas características a través de múltiples variables.

Se propone una nueva forma de gráfica para poder evidenciar las relaciones entre múltiples unidades estadísticas de forma correlacionada, basado en la hoja de calculo de Excel usando comandos de tabla

dinámica, ya que “es una herramienta avanzada para calcular, resumir y analizar datos que le permite ver comparaciones, patrones y tendencias en ellos.” (Microsoft, 2021).

Para ello se construyen gráficos en forma triangular, cuadrada, pentagonal, etc., dependiendo de la cantidad de variables que se desea analizar de manera correlacionada a las unidades estadísticas. El vértice del gráfico corresponde a cada una de las variables en cuestión y en cada arista se ubicarán el dato que corresponde a la cantidad de unidades estadísticas que cumplen con las variables correspondientes a los vértices que la forman.

En el polígono, la ubicación de las variables estará determinada por el orden con el cual sean seleccionadas, así, si se desea que dos variables estén en vértices consecutivos deberán seleccionarse una seguida de la otra en la hoja de cálculo, en el momento de construir el gráfico Poligonal.

Esta nueva forma de analizar los datos se espera que ayude a las personas a procesar de manera mucho más fácil, rápida y eficaz los metadatos o grandes volúmenes de información provenientes de procesos estadísticos, cuya información se pueda ubicar en una hoja de cálculo, específicamente en Microsoft Excel 2007 (v12.0) o posterior, esto por cuanto la programación de los gráficos fue desarrollada en esta herramienta.

2. Marco Teórico

Los análisis a partir de la media aritmética, mediana, desviación estándar, varianza, etc., son exclusivos para variables cuantitativas, pues utilizan fórmulas. Según Acuña y Chinchilla (2015) “en notación de sumatoria la media aritmética se escribe $\bar{X} = \frac{1}{n} \sum_{i=1}^m X_i$ ” (p.9), la cual es imposible de utilizar en variables cualitativas al igual que la mediana que “es el valor que está en el centro de todos los valores si estos se ordenan” (Acuña y Chinchilla, 2015, p.11) y solamente se pueden ordenar los valores numéricos de menor a mayor únicamente aplicable a variables cuantitativas, además “la desviación estándar es: $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ ” (AyC, 2015, p.20) donde evidentemente se realizan operaciones matemáticas para su cálculo, al igual que la varianza que “es el cuadrado de la desviación estándar y se denota s^2 ” (AyC, 2015, p.20) por lo que a las variables cuantitativas se les puede analizar de diversas maneras.

Sin embargo las cualitativas carecen de procesos que permitan un verdadero estudio de las mismas, únicamente la moda que es el dato con mayor frecuencia absoluta, y el conteo. A pesar del poco análisis que se le pueda realizar a las variables cualitativas, estas no carecen de importancia pues se utilizan para identificar los gustos, preferencias y características de una población, cuyo estudio es relevante para la toma de decisiones en una pequeña o mediana empresa, organización o gobierno.

En dichos estudios las tablas dinámicas toman protagonismo pues permiten analizar grandes volúmenes de información de manera rápida, de variables tanto cualitativas como cuantitativas, sin embargo, con ellas solo se pueden observar las relaciones entre dos variables en forma tabular, o se pueden observar resultados de los datos tipo filtro de variable en variable.

Debido a lo anterior, las tablas dinámicas no permiten observar la relación existente entre más de dos variables analizadas limitando la visualización completa de sus interacciones. Tanto la tabla dinámica como el filtrado de datos depende de la elección de variables que se haga para el estudio según el interés que se tenga, pero no permite la visualización de datos significativos presentes en la información, visibles solo si dichas variables son elegidas.

El Gráfico Poligonal de correlación viene a suplir dicha falencia, ya que se basa en la estructura de construcción de tablas dinámicas, pero muestra no solo el dato solicitado, sino que también la frecuencia absoluta y relativa de las unidades estadísticas que cumple con la o las características de las variables seleccionadas.

La programación del complemento que permite construir los Gráficos Poligonales de correlación se realizó en la versión de Microsoft Excel 2017 bajo ambiente Windows y fue adaptado para que funcione en Microsoft Excel para Mac versión 16.5, permitiendo que cualquier persona que desee analizar información pueda utilizarla.

3. Metodología

El complemento elaborado para la creación del gráfico poligonal está compuesto de ocho módulos de programación y existen dos comandos que el usuario debe utilizar para la creación y eliminación del gráfico, que son: el comando Grafico y EliminarGrafico.

El comando Grafico, se utiliza para la creación del gráfico, este comando crea una nueva hoja con los datos filtrados y el gráfico correspondiente, una vez ejecutado el comando, se le solicita al usuario que ingrese el número de variables a utilizar que van de 3 variables (mínimo) a 8 variables (máximo), luego se le solicita que indique con 1 si el usuario desea seleccionar las variables a graficar, o 2 si la selección de las variables las realiza el programa automáticamente utilizando las variables que contiene mayor número de datos repetidos.

Si se elige la opción 1 se le solicita que seleccione una celda que esté dentro del rango de datos a analizar (esto para ubicar al programa dentro del rango de datos que se van a utilizar), una vez hecho esto el programa le muestra un resumen de cada uno de las variables analizadas y por último le solicita que indique el tamaño para el gráfico que va de 50 pixeles (como mínimo) a 200 pixeles (como máximo).

El comando EliminarGrafico, se utiliza para eliminar la hoja del gráfico que fue creada con el comando anterior.

Para la instalación del complemento, se debe descargar el archivo GraficosPoligonales.xlam de la dirección https://tecdigital.tec.ac.cr/revistamatematica/Articulos/RevistaDigital_V22_n2_2022_Alfaro/GraficosPoligonales.xlam

Una vez descargado se instala el complemento en Microsoft Excel utilizando los siguientes pasos:

1. Seleccione la opción "Insertar".
2. En la barra de opciones, seleccione el icono "Complementos" luego, "Mis complementos" y luego la opción "Administrar otros complementos".
3. En la ventana que aparece y teniendo seleccionada la opción Complementos, indique en la opción "Administrar": "Complementos de Excel" y luego el botón "Ir"
4. Una vez que aparece la ventana de "Complementos", seleccione "Examinar" y seleccione el archivo descargado (Gráficos poligonales.xlam) y luego el botón "Aceptar".
5. Cuando aparece el mensaje con la pregunta ¿Desea copiar 'GraficosPoligonales.xlam' en la carpeta Addins?, responda con "Si" y luego el botón "Aceptar".
6. Ahora que ya está instalado el archivo de complementos para gráficos poligonales debemos agregar las opciones en la barra de menú Insertar, para ello:
 - a) Seleccione el menú "Archivo" y luego la opción: "Opciones".
 - b) En la ventana que aparece seleccione: "Personalizar cinta de opciones".
 - c) En esta misma ventana en la opción: "Comandos disponibles en":, seleccione: "Macros".
 - d) En la sección "Pestañas principales" que aparece a la derecha de la ventana, seleccione la opción "Gráficos" y luego el botón "Nuevo Grupo".

- e) Seleccione la opción “Nuevo grupo (personalizada)” y luego el botón “Cambiar nombre”.
 - f) Seleccione un icono y cambie el nombre por Gráfico Poligonal y “Aceptar”.
 - g) Estando seleccionada la opción “Gráfico Poligonal (personalizada)”, seleccione en la sección de la izquierda: ‘GraficosPoligonales.xlam’!Grafico.Grafico y luego el botón “Agregar”.
 - h) Utilice el botón “Cambiar nombre”, para cambiar el nombre de ‘GraficosPoligonales.xlam’!Grafico.Grafico por: Crear Gráfico Poligonal, luego seleccione un Símbolo para identificarlo y por último el botón “Aceptar”.
 - i) Estando seleccionada la opción “Gráfico Poligonal(personalizada)”, seleccione en la sección de la izquierda: Eliminar Grafico y luego el botón Agregar.
 - j) Utilice el botón Cambiar nombre, para cambiar el nombre de “Eliminar Grafico” por: Eliminar Gráfico Poligonal, luego seleccione un símbolo para identificarlo y por último el botón “Aceptar”.
 - k) Utilice el botón “Aceptar” para cerrar el cuadro de diálogo actual.
7. Una vez realizados los pasos anteriores, ya podemos utilizar las opciones de crear gráfico poligonal y eliminar gráfico poligonal que aparecen en el menú Insertar.

Dentro de las restricciones del complemento para la creación de los gráficos poligonales existen:

1. Solamente permite crear un gráfico poligonal por libro de cálculo(datos).
2. Para evitar errores a la hora de crear los gráficos poligonales, es importante que la información de los datos que se van a utilizar para el gráfico, esté organizada en rangos de datos consecutivos en los que no hayan celdas vacías entre los mismos.
3. En cuanto a la cantidad de datos que se han utilizado para comprobar los resultados de los gráficos, estos han contenido más de 12000 registros(filas) y más 250 campos(columnas). Es importante aclarar que mientras mas datos tiene la base de datos, más tiempo tarda el programa en realizar el análisis de los mismos y por ende en graficar los resultados.
4. Se permite crear gráficos poligonales de 3, 4, 5, 6 y 8 variables.

4. Análisis de Resultados

Para ejemplificar el trabajo que realizan los Gráficos Poligonales de correlación de unidades estadísticas se utiliza una base de datos proporcionada por el Instituto Nacional de Estadística y Censos (INEC), llamada “Encuesta Continua de Empleo, I Trimestre 2019”, aplicada a una muestra de nueve mil viviendas que equivale a 25696 personas mayores de 15 años la cual evalúa 307 variables de las cuales para ejemplificar la construcción de los Gráficos Poligonales se toman solamente: Relación parentesco, Estado conyugal, Sexo, País nacimiento, Grupo edad, Educación título, Lugar de nacimiento, y Posee seguro. De los 25696 datos estadísticos proporcionados se utilizaron 1000 ya que mientras más grande sea la base de datos, más tiempo tarda el programa en realizar el análisis de los mismos (dependiendo de la memoria RAM y procesador de la computadora que se utilice) y por ende en graficar los resultados.

Los gráficos poligonales de correlación presentados contienen información numérica en diferentes colores. Siendo el blanco el correspondiente a la cantidad de encuestados que cumplen solamente una característica (vértices del polígono), de color verde la cantidad de unidades estadísticas que cumplen con dos de las variables (aristas del polígono), además en color amarillo la cantidad de unidades estadísticas que satisfacen tres variables a la vez (en la zona interna al gráfico) y en color rojo la

cantidad de datos estadísticos que cumplen con todas las características seleccionadas de interés para el estudio (en el centro del gráfico). Estas cantidades se pueden apreciar tanto en valores absolutos como en valores relativos.

Todos estos datos numéricos dentro del Gráfico Poligonal muestran el grado de correlación entre las unidades estadísticas en estudio, a continuación, se brindan varios ejemplos donde se evidencia lo anteriormente explicado.

A continuación se muestran la Figura 1a y Figura 1b donde con solo cambiar una de las variables, el valor central del gráfico es significativamente mayor en la Figura 1b.

En la Figura 1a, el 334 en color blanco corresponde a las personas que cumplen la condición de ser “Jefe o jefa”, el 535 blanco indica la cantidad de personas que cumplen la condición de ser “Mujer”, y el 299 blanco representa las personas que cumplen la condición de estar “casado(a)” según la base de datos proporcionada por el INEC.

El 146 de color verde corresponde a las personas que cumplen la condición de ser “Jefe o jefa” y además ser persona “casado(a)”, el 132 de color verde indica la cantidad de “Mujeres” que son “Jefas”, y el 150 de color verde representa a las “Mujeres” que están “Casadas”.

Con color rojo se muestra que 11 personas cumplen con las tres características seleccionadas en este caso ser “Mujer”, “Jefa” y estar “casada”.

En la Figura 1b, el 334 en color blanco corresponde a las personas que cumplen la condición de ser “Jefe o jefa”, el 535 blanco indica la cantidad de personas que cumplen la condición de ser “Mujer”, y el 299 blanco representa las personas que cumplen la condición de estar “soltero(a)”.

El 58 de color verde corresponde a las personas que cumplen la condición de ser “Jefe o jefa” y además ser persona “soltero(a)”, el 132 de color verde indica la cantidad de “Mujeres” que son “Jefas”, y el 150 de color verde representa a las “Mujeres” que están “Solteras”.

Con color rojo se muestra que 42 personas cumplen con las tres características seleccionadas, en este caso ser “Mujer”, “jefa” y estar “soltera”.

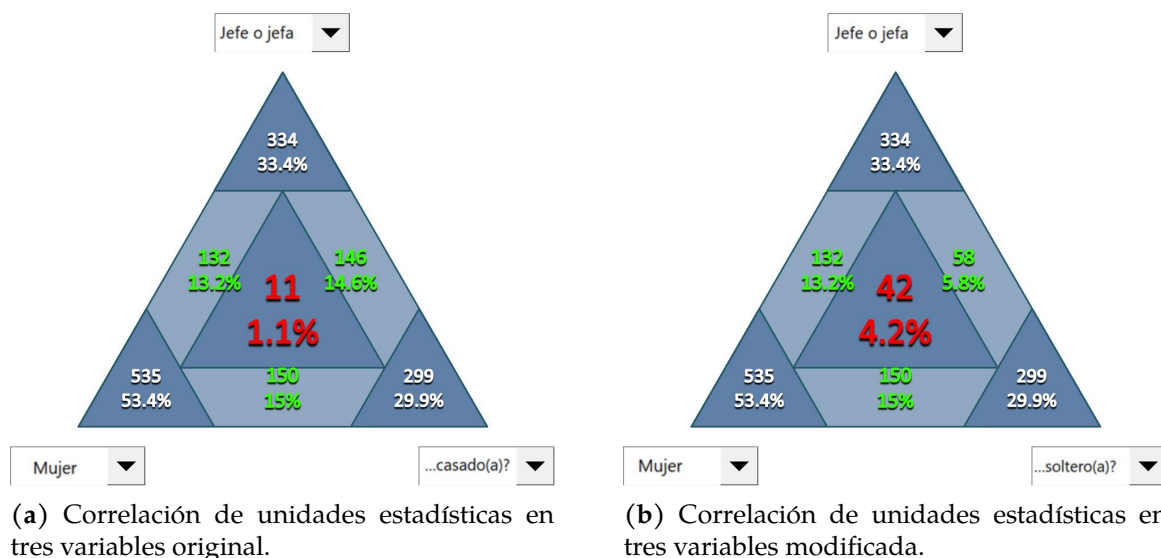


Figura 1: Correlación de unidades estadísticas en tres variables.

Al modificar la variable de estado conyugal de persona casada a persona soltera se evidencia que es aproximadamente cuatro veces mayor el índice de mujeres jefas solteras que el de mujeres jefas casadas, varió de 11 en la Figura 1a a 42 en la Figura 1b.

Lo anterior revela que cambiar la característica de una variable en la pestaña desplegable modifica de

inmediato los datos contenidos dentro del Gráfico Poligonal, permitiendo observar diferentes correlaciones ya sean con índices mayores o menores.

En la Figura 2, se analiza la correlación de unidades estadísticas tomando tres variables cualitativas (Nacionalidad, Sexo, y Educación título) y una cuantitativa (Grupo edad). A diferencia de las Figuras 1a y 1b, se observan los índices de color amarillo que corresponden a las personas que cumplen la variable del vértice donde se encuentra y los dos adyacentes, así 93 personas de mil encuestadas son “Mujeres” “costarricenses” entre los “30 y 44” años; 13 personas son “Mujeres” entre los “30 y 44 años” con título de “Bachillerato”, 27 personas son “Mujeres” “costarricenses” con el título de “Bachillerato”, y 23 personas son “costarricenses” entre los “30 y 44 años” con el título de “Bachillerato”.

Asimismo, en la Figura 2, se muestran algunas de las características correspondientes para la variable “Educación título” que pueden ser seleccionadas en la lista desplegable, en donde al elegir otra diferente cambiarían de inmediato los datos numéricos de correlación y la lista se repliega quedando visible únicamente la característica elegida.

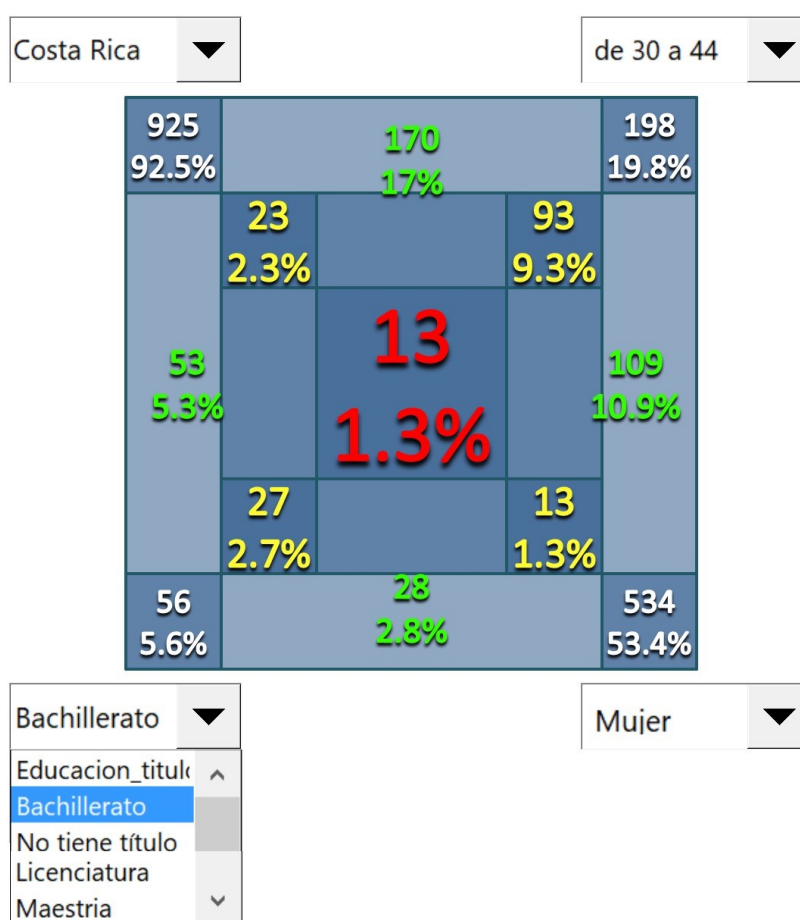


Figura 2: Correlación de unidades estadísticas en cuatro variables.

A mayor número de variables analizadas será mayor la cantidad de datos correlacionados que se pueden observar. Por lo que entre más interno se encuentre el dato dentro del Gráfico Poligonal, este posee la información de la correlación existente entre una mayor cantidad de unidades estadísticas.

En la Figura 3, se observa que dentro del gráfico hay 19 datos agrupados en 4 colores diferentes, dependiendo de el grado de correlación entre las unidades estadísticas, particularmente en color rojo dos personas de mil son “jefes”, nacidos en “el mismo cantón”, entre los “15 y 29” años, “Hombres”, que “si” poseen seguro y están “casados”.

En esta figura se muestra la correlación de unidades estadísticas en seis variables, pero podrían ser muchas más de acuerdo al interés del investigador.

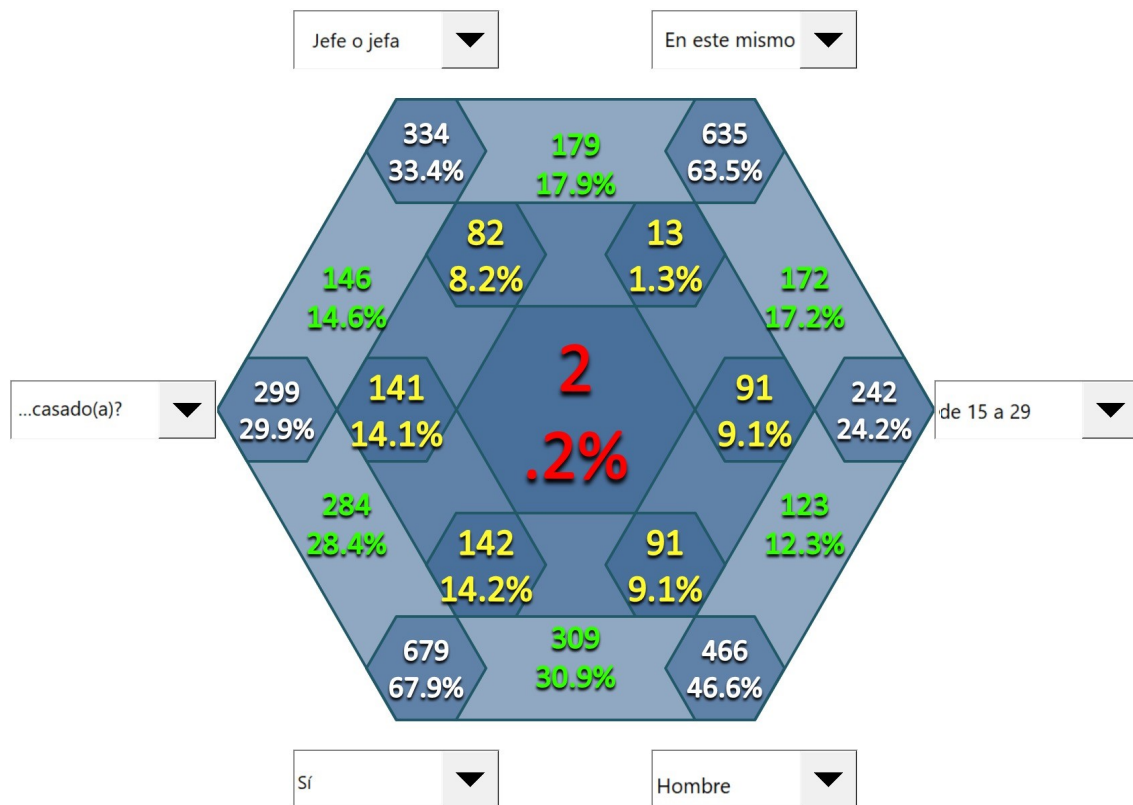


Figura 3: Correlación de unidades estadísticas en seis variables.

5. Conclusiones

1. El análisis de metadatos a través de Gráficos Poligonales le permite al usuario combinar las diferentes características de una variable, para encontrar las mejores o más altas correlaciones que le facilite tomar las mejores decisiones según los resultados que vislumbre.
2. Al elegir una diferente característica en una variable en la pestaña desplegable modifica de inmediato los valores absolutos y relativos dentro del Gráfico Poligonal, permitiendo al investigador observar diferentes correlaciones según su interés ya sean con índices mayores o menores.
3. En la Figura 3 se observa que solo dos personas de mil son hombres casados entre los 15 y 29 años, jefes, que viven en el mismo cantón donde trabajan, y que poseen seguro; y esto es una población meta muy pequeña, visible gracias a los Gráficos poligonales. La población meta es importante por ejemplo para definir los beneficiarios de programas de bien social, requisitos que debe cumplir un solicitante para créditos bancarios, otorgamiento de becas estudiantiles, características deseables en personas candidatas a un empleo, etc, y la utilización de estos gráficos ayudaría considerablemente para determinar el impacto que se esperaría tengan dichos proyectos.
4. En la Figura 1a, se obtuvo un valor central de 11 mujeres casadas jefas, y en la Figura 1b un valor central de 42 mujeres solteras jefas, por lo que estamos hablando de 52 mujeres jefas, es decir, se evidencia como las mujeres que ostentan un puesto de jefatura solo tienen un 20,7% (11 de 53) de probabilidades de conseguirlo si están casadas, frente al 79,3% (42 de 53) de las solteras. Gracias al análisis de los gráficos poligonales es fácil evidenciar ese dato, lo cual puede dirigir futuras investigaciones sobre las razones que lo generan.

6. Bibliografía

- [1] Real Academia Española. (s.f). Correlación. En Diccionario de la lengua española. <https://dle.rae.es/correlaci%C3%B3n>.
- [2] Instituto Nacional de Estadística y Censos (INEC). (2019). Encuesta Continua de Empleo, I Trimestre 2019. <http://sistemas.inec.cr/pad5/index.php/catalog/246/study-description>
- [3] Microsoft. (s.f.) Crear una tabla dinámica para analizar datos de una hoja de cálculo. <https://support.microsoft.com/es-es/office/crear-una-tabla-din%C3%A1mica-para-analizar-datos-de-una-hoja-de-c%C3%A1lculo-a9a84538-bfe9-40a9-a8e9-f99134456576>
- [4] Acuña, R. y Chinchilla, Jorge. (2015). Estadística y Probabilidad para profesores de Matemática. Instituto Tecnológico de Costa Rica. https://www.drea.co.cr/sites/default/files/Contenido/Sesion%20N%C2%B04_Folleto%20con%20actividades_Actualizado%20al%2027%20junio.pdf