



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICA

Análisis Multivariado de datos provenientes de la población asistida por la organización sin fines de lucro Acción Solidaria.

Trabajo Especial de Grado presentado ante
la ilustre Universidad Central de Venezuela
por la **Br. Brenda M. Gómez C.** para op-
tar al título de Licenciado en Matemática.

Tutores: Dra. Mairene Colina.

MSc. José Benito Hernández.

Caracas, Venezuela

Mayo, 2012.

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado “**Análisis Multivariado de datos provenientes de la población asistida por la organización sin fines de lucro Acción Solidaria**”, presentado por la **Br. Brenda M. Gómez C.**, titular de la Cédula de Identidad **V-19.371.794**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciado en Matemática**.

Dra. Mairene Colina
Tutor

MSc. José Benito Hernández
Co-tutor

MSc. Angie Pineda
Jurado

Dedicatoria

Este trabajo especial de grado es sólo un paso más para obtener un título, pero lo que lleva consigo son experiencias y aprendizajes a lo largo de una carrera. Quiero dedicarle este trabajo a Dios, a mi familia , especialmente a mi madre Maria Trinidad por ser una gran trabajadora, ser una gran ejemplo de lucha y a seguir, ser el apoyo de vida, a mi padre Hugo por ser un hombre que me ha enseñado siempre el buen actuar, mi tía Helina por ser la persona que siempre estuvo en mi formación, a mis hermanos Sofía y Carlos, y a mis amigos porque han sido el pilar de fortaleza, sacrificio y confianza durante toda mi vida.

Finalmente, quiero dedicar este trabajo a una de las personas que siempre está presente en múltiples formas en mi vida, y que siempre habrán cosas que me recordarán a ella, mi tía Ruth Gómez (1944 - 2007) (Q.E.P.D).

Agradecimiento

Principalmente quiero agradecer a Dios por todo en mi vida, a mis tutores Mairene Colina y José Benito Hernández por haberme ayudado en todo momento y por confiar en mí. Quiero agradecer también a la organización sin fines de lucro, Acción Solidaria por haberme facilitado los datos de las asesorías, pilar fundamental en esta tesis.

Mis experiencias de vida antes de la universidad me hicieron pensar que te encuentras solo como individuo, pero debo decir que a lo largo de mi carrera me he encontrado con gente maravillosa, gente que me ha ayudado de diferentes formas, es por eso que este agradecimiento no estaría completo sin dar gracias a Nancy Torres, Sherezade Rivas, Reyfel Mendoza, Diarelys Escobar, Jesús Gonzáles, Francisco Fernández, Yorley Acevedo, Leimar Torres, Henry Cerezo, Randy Alzate, German Miranda, Gabriel Hernández, Antonietta Imbriano, Witney Castillo, Edwin Pin y Henry Navarro, todos ustedes saben que de alguna u otra forma estuvieron ahí.

Hay algunos con los cuales no compartí un aula, pero quiero agradecerles por ser incondicionales Sofía Gómez, Daniela Adarmes, Rafael Marín, Josu Ortega, Mónica De Santis, Cristina Yepez y en especial a Sergio Barrantes.

Sigo agradeciendo a todos, pero hay un grupo en particular que quiero dar infinitas gracias y es el grupo de profesores que me dieron clases, debo decir que ha sido todo un honor y me ha llenado de sabiduría a lo largo de los años, en especial el profesor y amigo Jose Gregorio Gómez, pues ha sido una gran apoyo y un ser de admirar.

A mi familia, amigos, compañeros y profesores, infinitas gracias.

Índice general

Introducción	1
Capítulo 1. Preliminares	4
1. Matriz de Datos	5
2. Estimaciones Insesgadas	7
3. Elementos de la matriz- Autovalores y Autovectores	9
Capítulo 2. Análisis de Componentes Principales (ACP)	12
Capítulo 3. Análisis Factorial (AF)	15
Capítulo 4. Análisis y Resultados	21
1. Histogramas Múltiples	43
2. Análisis y Resultados del ACP y AF	46
3. Conclusiones	54
Apéndice A.	56
Bibliografía	60

Introducción

Actualmente existe una enfermedad que afecta a las personas sin distinguir raza, género, edad, orientación sexual y/o creencia, la cual es el VIH/SIDA. A partir de 1981 se identificaron los primeros casos de SIDA en los Estados Unidos desde entonces casi 30 millones de personas han sido infectadas en el mundo. En América Latina los primeros casos se comenzaron a detectar en 1983.

El VIH son las iniciales de **V**irus de **I**mmunodeficiencia **H**umana, este es un Virus que ataca a nuestro sistema inmunológico y además sólo afecta a los seres humanos, lo que descarta la posibilidad de que afecte a los animales o insectos. El VIH es un retrovirus que causa eventualmente el SIDA. Se llama retrovirus porque por sí solo no puede replicarse y para hacerlo necesita enzimas que se encuentran en las células del sistema inmunológico. Cualquier ser humano, bajo ciertas condiciones, puede contraer el VIH. El VIH no es contagioso sino transmisible.

El SIDA son las iniciales de **S**índrome de **I**mmunodeficiencia **A**dquirida, es un conjunto de signos y síntomas que aparecen cuando nuestro sistema inmunológico está deficiente, es decir, debilitado por el ataque continuo del VIH. Además es adquirido, no se hereda.

Los síntomas iniciales de la infección pueden ocurrir entre la segunda y la cuarta semana después del momento de la infección. Los síntomas pueden ser variados y no todas las personas presentan el mismo cuadro. Entre ellos están manifestaciones como ganglios inflamados, fiebre y dolor de garganta, diarrea, dolor de cabeza, vómitos y náuseas, sudoraciones nocturnas y agotamiento.

Una vez que una persona adquiere el VIH, TODOS sus fluidos contienen VIH. Sin embargo los únicos fluidos capaces de transmitir el virus son: la sangre, el semen, el líquido pre-seminal, los fluidos vaginales y la leche materna. Esto ocurre porque éstos poseen la calidad necesaria para que el virus pueda sobrevivir y que además esté en concentraciones

adecuadas para que ocurra la infección. La transmisión del VIH ocurre cuando cualquiera de estos fluidos infectados entra al torrente sanguíneo de otra persona.

Como ya mencionamos anteriormente el VIH/SIDA puede afectar a cualquier ser humano, en Venezuela, específicamente existen varios casos a ser estudiados, es por eso que la organización Acción Solidaria nos brindó un amplio archivo de datos sobre asesorías para hacer un estudio a profundidad de los mismos.

ACCIÓN SOLIDARIA, AcSol, es una Asociación Civil sin fines de lucro, fundada en Octubre de 1995. Su misión es contribuir a reducir el impacto social de la epidemia del VIH/SIDA, en Venezuela y otros países de habla hispana, ofreciendo respuesta integrales en las áreas de prevención, atención e incidencia pública, enmarcadas dentro del reconocimiento, respeto, protección y atención a los derechos humanos de las personas que viven con VIH o SIDA, sus parejas y familias. Sus prioridades son:

- a) Promover que quienes vivan con VIH o SIDA cuenten con los servicios y recursos necesarios para llevar una vida activa, integrada y sin estigmatización o discriminación.
- b) Frenar la expansión del VIH, evitando que más personas sean afectadas, a través de estrategias preventivas que permitan disminuir la incidencia del VIH.

Para contribuir con esta organización se hará un estudio sobre los datos provenientes de la población asistida por la organización, dichos datos son asesorías de los años 2008, 2009 y 2010. Se tomaron las siguientes características: Tipo de Asesoría, Persona que recibe la asesoría, Género, Orientación, Nivel de instrucción, Estado civil, Edad, Exámenes previos y Donde obtuvo la información, a fin de determinar cuáles son las más resaltantes y dar una información acerca de los que acuden a esta organización y en que forma se puede llegar a ser efectivo en la comunicación, difusión y prevención del VIH/SIDA. Nuestro marco teórico se basa en el uso de procedimientos estadísticos y análisis multivariante, tales como histogramas, histogramas múltiples, Análisis de Componentes Principales y Análisis Factorial.

El análisis multivariante es un conjunto de técnicas estadísticas, cuya finalidad es analizar simultáneamente un conjunto de datos multivariantes, en el sentido de que existen varias

variables medidas para cada individuo u objeto. Su razón de ser radica en tratar de dar una mejor explicación del fenómeno de estudio, obteniendo información que los métodos estadísticos univariados y bivariados son incapaces de conseguir. Es por eso que se hará uso de estos métodos en este trabajo.

Es por toda esta iniciativa que surge el interés de estudio de una base de datos que más que una simple información es una realidad que afecta indudablemente a la población venezolana. Nuestro aporte es de gran significancia, al saber cuáles mecanismos de difusión, quiénes son los que más acuden a la organización, entre otros, nos darán una vista panorámica sobre la situación actual, así se podrá manejar desde diferentes ópticas, ya sea mediante insumos, propaganda u otras soluciones que se plantearán en referencia a los resultados obtenidos del análisis de esta base de datos.

Este trabajo se encuentra estructurado de la siguiente manera: en el primer capítulo trataremos algunos conceptos estadísticos de nuestro interés, a partir del segundo capítulo se describe el procedimiento matemático que transforma un conjunto de variables correlacionadas en un conjunto menor de variables no correlacionadas (componentes principales), seguidamente se estudiará el proceso analítico que se basa en la matriz de correlaciones (Capítulo 3) y para finalizar se expondrán los histogramas obtenidos, los resultados de los análisis y las conclusiones de este trabajo (Capítulo 4).

Capítulo 1

Preliminares

El análisis multivariante es un conjunto de técnicas estadísticas, cuya finalidad es analizar simultáneamente un conjunto de datos multivariantes, en el sentido de que existen varias variables medidas para cada individuo u objeto. Su razón de ser radica en tratar de dar una mejor explicación del fenómeno de estudio, obteniendo información que los métodos estadísticos univariados y bivariados son incapaces de conseguir.

Los datos de variables múltiples se presentan en todas las ramas de las ciencias.

A continuación se definirán algunos conceptos vinculados al análisis multivariado.

DEFINICIÓN 1.1. Una **unidad experimental** es cualquier objeto o concepto que se puede medir o evaluar de alguna manera.

La medición y evaluación de unidades experimentales es una actividad principal de la mayoría de los investigadores. Algunos ejemplos de unidades experimentales son: gente, animales, insectos, compañías, árboles, países, etc.

DEFINICIÓN 1.2. Los **datos de variables múltiples** son la información obtenida cuando se evalúa más de un atributo ó característica de cada unidad experimental.

DEFINICIÓN 1.3. La **variable** es el atributo o característica de cada unidad experimental.

Dentro de las variables podemos encontrar las **variables continuas**, que aparecen en un intervalo, pero no se miden con exactitud y las **variables discretas** que pueden ser *numéricas*, indicando cantidad exacta y *no numéricas* que no se pueden ordenar ni cuantificar.

Los métodos multivariados son extraordinariamente útiles para ayudar a los investigadores a hacer que tengan sentido conjuntos grandes, complicados y complejos de datos, que constan de una gran cantidad de variables, medidas en números grandes de unidades experimentales.

Entre los objetivos del análisis de datos multivariados encontramos:

- (1) Resume grandes cantidades de datos por medio de relativamente pocos parámetros.
- (2) Proporcionar métodos cuya finalidad es el estudio de un conjunto de datos multivariantes que el análisis estadístico uni y bidimensional es incapaz de conseguir.
- (3) Ayuda al analista o investigador a tomar decisiones óptimas en el contexto en que él se encuentra, teniendo en cuenta la información aportada por el conjunto de datos analizados.

1. Matriz de Datos

Sea p la cantidad de variables numéricas de respuesta que se están midiendo y N el número de unidades experimentales sobre las cuales se están midiendo las variables. Se usa la variable x_{rj} para identificar el valor de la j -ésima variable respuesta en la r -ésima unidad experimental para $r = 1, 2, \dots, N$ y $j = 1, 2, \dots, p$.

Los x_{rj} se pueden disponer en una matriz llamada **matriz de datos**, de modo que x_{rj} es el elemento en el r -ésimo renglón y la j -ésima columna de la propia matriz. La matriz de datos es de orden $N \times p$ y se denota por X . Bajo este contexto,

$$(1.1) \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix}_{N \times p}$$

representa una matriz de datos.

Los datos del j -ésimo renglón de X se escriben en un vector columna, se denota por x_j . En este caso

$$(1.2) \quad x_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{bmatrix}.$$

Donde N es el número de unidades experimentales y $j = 1, 2, \dots, p$.

DEFINICIÓN 1.4. Se dice que un vector de variables aleatorias $\begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix}$ tiene una **distribución normal multivariada**, si

$$ax' = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = a_1x_1 + \cdots + a_px_p$$

$$= \sum_{i=1}^p a_ix_i$$

tiene una distribución normal univariada para todos los conjuntos posibles de valores seleccionados para los elementos en el vector a .

DEFINICIÓN 1.5. La **media de un vector de variables aleatorias** x , se denota por μ y la **matriz de covarianzas de** x se denota por Σ .

Así,

$$\mu = \mathbb{E}(x) = \begin{bmatrix} \mathbb{E}(x_1) \\ \mathbb{E}(x_2) \\ \vdots \\ \mathbb{E}(x_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

y la varianza muestral de la i -ésima variable se define por:

$$\sigma_{ii} = Var(x_i) = \mathbb{E}[(x_i - \mu_i)^2] \text{ para } i = 1, 2, \dots, p;$$

la covarianza entre la i -ésima y j -ésima variable viene dada de la siguiente manera:

$$\sigma_{ij} = Cov(x_i, x_j) = \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)] \text{ para } i \neq j = 1, 2, \dots, p.$$

Por tanto la matriz formada por el arreglo de los σ_{ii} y σ_{ij} será la matriz:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix},$$

es decir la matriz de covarianzas de x .

DEFINICIÓN 1.6. El **coeficiente de correlación** entre x_i y x_j se denota por ρ_{ij} y se define por

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

El coeficiente de correlación proporciona una medida de la asociación lineal entre dos variables.

DEFINICIÓN 1.7. La **matriz de correlación** para un vector aleatorio se denota por P , y se define por:

$$P = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}.$$

2. Estimaciones Insegadas

DEFINICIÓN 1.8. Sea x_1, x_2, \dots, x_p una muestra aleatoria de una distribución normal multivariada como en la expresión (1.2), que tiene un vector de medias, μ , una matriz de covarianzas, Σ , y una matriz de correlación, P . **Las estimaciones insegadas de μ y Σ** se expresan por,

$$\hat{\mu} = \frac{1}{p} \left(\sum_{t=1}^p x_t \right) = \frac{x_{1r} + x_{2r} + \dots + x_{pr}}{p}$$

y

$$\hat{\Sigma} = \frac{1}{p-1} \left[\sum_{t=1}^p (x_t - \hat{\mu})(x_t - \hat{\mu})' \right]$$

respectivamente, con $1 \leq r \leq N$.

Observación: La distribución normal multivariada $N_p(\mu, \Sigma)$ es una generalización de la normal univariante, es decir que las distribuciones de cada variable son normales, de forma que cada $X_i \sim N(\mu_i, \sigma_i^2)$, con media μ_i y varianza σ_i^2 , para $i = 1, 2, \dots, p$.

Note que:

$$\hat{\mu} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_p \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & \cdots & \hat{\sigma}_{pp} \end{bmatrix}$$

en donde

$$\hat{\sigma}_{ii} = \widehat{Var}(x_i) = \frac{1}{p-1} \sum_{t=1}^p (x_{ti} - \bar{x}_i)^2$$

y

$$\hat{\sigma}_{ij} = \widehat{Cov}(x_i, x_j) = \frac{1}{p-1} \sum_{t=1}^p (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j).$$

Los estimadores de los coeficientes de correlación ρ_{ij} suelen tomarse como $r_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}$, aún cuando r_{ij} no sea una estimación insesgada de ρ_{ij} , para cualquier i y j . Aunque r_{ij} no sea una estimación insesgada de ρ_{ij} , si tiene la propiedad de que $-1 \leq r_{ij} \leq 1$, para toda $i \neq j$.

La matriz de correlación de la muestra se denota por \mathbf{R} y se define por:

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}.$$

Datos estandarizados y/o valores Z

A veces los datos son más fáciles de comprender y de comparar cuando las variables respuesta se estandarizan, de modo que se midan en unidades comparables. Esto suele hacerse mediante la eliminación de las unidades de medición en conjunto. Defínase

$$Z_{rj} = \frac{x_{rj} - \hat{\mu}_j}{\sqrt{\hat{\sigma}_{jj}}},$$

para $r = 1, 2, \dots, N$ y $j = 1, 2, \dots, p$.

La variable Z_{rj} se llama valor Z para la j -ésima variable de respuesta de la r -ésima unidad experimental y

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{Np} \end{bmatrix}$$

se conoce como la matriz de valores de Z .

3. Elementos de la matriz- Autovalores y Autovectores

Casi todos los métodos multivariados dependen de funciones de los elementos de una matriz de covarianzas de la muestra, por lo común denotada por $\hat{\Sigma}$, o bien, de los elementos de una matriz de correlaciones de la muestra, normalmente denotada por \mathbf{R} . Las funciones más importantes de los elementos de una matriz, para los fines de los análisis de datos multivariados, son:

- (1) Su traza.
- (2) Su determinante.
- (3) Sus autovalores y autovectores.

Traza y determinante

En cada una de las definiciones siguientes, sea

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Note que Σ es una matriz cuadrada $p \times p$. Asimismo, Σ es una matriz simétrica. Las matrices de varianzas-covarianzas y las de correlaciones siempre son simétricas.

DEFINICIÓN 1.9. La **traza de** Σ , denotada por $tr(\Sigma)$, se define por

$$tr(\Sigma) = \sum_{i=1}^p \sigma_{ii} = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}.$$

Por lo tanto, la traza de una matriz cuadrada es igual a la suma de sus elementos en la diagonal que van desde la esquina superior izquierda hasta la esquina inferior derecha.

DEFINICIÓN 1.10. El **determinante de una matriz cuadrada** Σ , denotado por $|\Sigma|$, se define por $|\Sigma| = \sum_{j=1}^p \sigma_{1j} \Sigma_{1j}$, en donde $\Sigma_{1j} = (-1)^{1+j} |\Sigma_{1j}|$ y $|\Sigma_{1j}|$ es la matriz obtenida a partir de Σ al eliminar su primer renglón y su j -ésima columna. El determinante de una matriz de 1×1 se define como el valor de la única entrada de esa matriz. Por lo tanto, $det(\sigma_{11}) = \sigma_{11}$.

Autovalores y Autovectores

DEFINICIÓN 1.11. Los **autovalores** (también llamados raíces características) de Σ son las raíces de la ecuación polinomial definida por $|\Sigma - \lambda I| = 0$.

Observación: Si Σ es una matriz simétrica, sus autovalores son números reales y se pueden ordenar desde el más grande hasta el más pequeño. En este caso, los autovalores de Σ se denotarán por

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p.$$

De este modo, λ_1 es el autovalor más grande de Σ , λ_2 es el segundo autovalor más grande de Σ , \dots , y λ_p es el autovalor más pequeño de Σ .

DEFINICIÓN 1.12. Cada autovalor de Σ tiene un vector no cero correspondiente a una columna de números llamada **autovector** (también conocido como vector característico) que satisface la ecuación matricial

$$\Sigma a = \lambda a.$$

Debido a que Σ tiene p autovalores, tendrá p autovectores. Denotemos por a_1, a_2, \dots, a_p , los autovectores de Σ correspondientes a los autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$, respectivamente.

Nota. Los autovectores de una matriz no son únicos, ya que con frecuencia se normalizan de modo que $a_i' a_i = 1$, para $i = 1, 2, \dots, p$. Cuando dos autovalores de Σ no son iguales, sus autovectores correspondientes serán ortogonales entre sí (es decir, $a_i' a_j = 0$, cuando $\lambda_i \neq \lambda_j$). Cuando dos autovalores de Σ son iguales, los autovectores correspondientes siempre se elegirán como ortogonales entre sí aunque, en este caso, los autovectores no quedan determinados de manera única. Algunas propiedades de los autovalores de una matriz son:

$$tr(\Sigma) = \sum_{i=1}^p \lambda_i,$$

$$|\Sigma| = \prod_{i=1}^p \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_p.$$

Por tanto, la traza de una matriz simétrica es igual a la suma de sus autovalores y el determinante de una matriz de ese tipo es igual al producto de sus autovalores.

Análisis de Componentes Principales (ACP)

El análisis de componentes principales es un procedimiento matemático que transforma un conjunto de variables, respuestas, correlacionadas en un conjunto menor de variables no correlacionadas llamadas componentes principales.

Los objetivos del ACP son:

- (1) Reducir la dimensionalidad del conjunto de datos.
- (2) Identificar nuevas variables significativas subyacentes.

Cálculo de las componentes principales

Para realizar el cálculo de componentes principales, se considera una serie de variables (x_1, x_2, \dots, x_p) sobre un grupo de objetos ó individuos, y se trata de calcular a partir de ellas, un nuevo conjunto de variables y_1, y_2, \dots, y_p ; incorrelacionadas entre sí.

DEFINICIÓN 2.1. La **primera variable o componente principal** se define por $y_1 = a_1'x$, en donde a_1 se elige de modo que la varianza de $a_1'x$ se maximice sobre todos los vectores a_1 que satisfagan $a_1'a_1 = 1$.

DEFINICIÓN 2.2. En líneas generales, **la j -ésima** ($j = 1, 2, \dots, p$) **componente principal** se expresa por $y_j = a_j'x$, en donde a_j se elige de modo que $a_j'a_j = 1$ y de tal forma que la varianza de $a_j'x$ sea un máximo entre todas esas combinaciones lineales de x que no están correlacionadas con las $j - 1$ componentes principales.

El valor máximo de la varianza de $a_j'x$ entre todas las combinaciones lineales de x que no están correlacionadas con y_1 a través de y_{j-1} es igual a λ_j , en otras palabras, el j -ésimo autovalor más grande de Σ , y este máximo ocurre cuando a_j es un autovector de Σ correspondiente al autovalor λ_j y que satisface $a_j'a_j = 1$. De este modo, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ denotan los autovalores ordenados de Σ y a_1, a_2, \dots, a_p , denotan los autovectores normalizados correspondientes a Σ .

Calificaciones de las componentes principales

Las calificaciones de componentes principales (valores de las variables componentes principales) para cada unidad experimental en el conjunto de datos es la que proporciona las ubicaciones de las observaciones en un conjunto de datos con respecto a sus ejes componentes principales, es decir con respecto a los autovectores de la matriz de correlaciones X asociados a los autovalores de ésta.

Sea x_r , el vector de variables medidas para la r -ésima unidad experimental. Entonces el valor (calificación) de la j -ésima variable componente principal, para la r -ésima unidad experimental, es $y_{rj} = a'_j x_r$, para $r = 1, 2, \dots, N$ y $j = 1, 2, \dots, p$.

Los **vectores de carga de componentes** serán $c_j = \sqrt{\lambda_j} a_j$.

Los **estimadores** de los λ_i y los a_i siempre se toman como los autovalores y autovectores correspondientes de $\hat{\Sigma}$. Éstos se denotan por $\hat{\lambda}_i$ y \hat{a}_i respectivamente. Las calificaciones de componentes principales se estiman por $\hat{y}_{rj} = \hat{a}'_j x_r$, para $r = 1, 2, \dots, N$ y $j = 1, 2, \dots, p$.

Determinación del número de componentes principales

Cuando se lleva a cabo un ACP, se necesita determinar la dimensionalidad real del espacio en el que caen los datos, es decir, el número de componentes principales que tienen varianzas mayores que cero.

Método 1: Si se desea tomar en cuenta $\gamma\%$ de la variabilidad total en las variables originales. En uno de los métodos para estimar la dimensionalidad (d) se considera $V = (\lambda_1 + \lambda_2 + \dots + \lambda_k) / \text{tr}(\Sigma)$ para valores sucesivos de $k = 1, 2, \dots, p$. Entonces d se estima por el menor de los valores de k , en el que, por primera vez, sobrepasa $\gamma\%$. La variabilidad dependerá a menudo del tipo de población de la que está tomando la muestra.

Método 2: En este método se pretende estimar d , mediante el uso de una gráfica SCREE de los autovalores. Una gráfica SCREE se construye al situar el valor de cada autovalor contra el recíproco. Es decir, se sitúan las parejas $(1, \hat{\lambda}_1), (2, \hat{\lambda}_2), \dots, (p, \hat{\lambda}_p)$. Cuando los puntos de la gráfica tienden a nivelarse, estos autovalores suelen estar suficientemente cercanos a cero

como para que puedan ignorarse. Por este método se supone que la dimensionalidad del espacio de datos es la que corresponde al autovalor grande más pequeño.

Nota: Cuando las variables no se presentan con fundamentos iguales, es decir con unidades como medidas diferentes, es necesario aplicar métodos ACP a los datos estandarizados, lo que se hace al calcular los autovalores y autovectores de la matriz de correlaciones. La estandarización o tipificación de los datos implica expresar cada valor de los datos en términos de su distancia (en desviaciones estándar) a la media.

En resumen; el **Análisis de Componentes Principales (ACP)** comprende un procedimiento matemático que transforma un conjunto de variables correlacionadas en un conjunto menor de variables no correlacionadas llamadas "componentes principales". La finalidad fundamental es analizar el conjunto de datos de variables múltiples de forma tal, que se pueda realizar una división en subgrupos de las unidades experimentales.

Al realizar un ACP, se desea determinar la dimensionalidad real del espacio en el que caen los datos, es decir el número de variables que tienen varianzas mayores que uno.

Los valores numéricos se pueden observar en la matriz de varianzas, también con la gráfica Scree o Gráfico de Varianzas se obtiene una solución más inmediata.

Se puede realizar una gráfica utilizando también la matriz de autovalores valiéndonos de la matriz de correlación P y analizarla de la misma forma, la cual se hace con la gráfica de varianzas, pero muchas veces este procedimiento no es necesario.

Análisis Factorial (AF)

El **Análisis Factorial (AF)** es una técnica de interdependencia con la cual se trata de estudiar el conjunto de interrelaciones que se dan entre conjunto de variables; es aplicable a variables tipificadas o estandarizadas.

El AF maneja comandos similares al ACP, pero con unas variantes. El proceso analítico se basa en la matriz de correlaciones entre las variables originales. Para que el análisis factorial sea adecuado es necesario que las variables estén correlacionadas ya que se espera determinar los factores subyacentes que explican tales correlaciones. Existen dos contrastes estadísticos que pueden realizarse para analizar si es adecuado aplicar el análisis factorial: Test de esfericidad de Bartlett y la Medida de adecuación muestral (KMO).

El Análisis Factorial (AF) trata de explicar p variables observables X_1, X_2, \dots, X_p mediante $m + p$ ($m < p$) variables no correlacionadas no observables $F_1, \dots, F_m; V_1, \dots, V_p$ denominadas factores, de acuerdo al modelo lineal siguiente:

$$\begin{aligned} X_1 &= \lambda_{11}F_1 + \dots + \lambda_{1m}F_m + \eta_1V_1 \\ X_2 &= \lambda_{21}F_1 + \dots + \lambda_{2m}F_m + \eta_2V_p \\ &\vdots \\ X_p &= \lambda_{p1}F_1 + \dots + \lambda_{pm}F_m + \eta_pV_p \end{aligned}$$

Las variables no observables F_1, \dots, F_m se denominan **factores comunes**, pues influyen en mayor o menor medida sobre cada una de las variables. Las variables no observables V_1, \dots, V_p se denominan **factores únicos**, pues su efecto específico influye sobre una variable.

- El número de factores comunes m es menor al número de variables p .

- Los factores comunes se obtendrán de modo que el primero sea el que explique la mayor parte de la variabilidad de la muestra de datos, el segundo sea ortogonal (esté no correlacionado al primero) y sea el que explique la mayor parte de la variabilidad no explicada por el primero y así sucesivamente.
- Cada uno de los coeficientes λ_{ij} se denomina carga factorial y representa la correlación lineal entre la variable X_i y el factor F_j . Un coeficiente grande en valor absoluto indica que la variable i y el factor j están estrechamente relacionados.

La matriz de coeficientes que recoge la información de la relación entre las variables y factores comunes es $\Lambda_{p \times m}$ y se denomina **matriz de patrón factorial**. La matriz que recoge los coeficientes de los factores únicos o **saturaciones**, es $D_{p \times p}$, así:

$$\Lambda = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pm} \end{bmatrix}; D = \begin{bmatrix} \eta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \eta_p \end{bmatrix}.$$

El primer objetivo del análisis factorial es detectar los factores subyacentes que explican las correlaciones entre un conjunto de variables y los factores, éste se consigue determinando la matriz Λ .

A partir del modelo factorial podemos cambiar del espacio \mathbb{R}^p al espacio \mathbb{R}^m , pues podemos expresar los factores comunes mediante combinaciones lineales de las variables observadas, mediante las relaciones:

$$\begin{aligned} F_1 &= \omega_{11}X_1 + \cdots + \omega_{p1}X_p + \varepsilon_1 \\ F_2 &= \omega_{12}X_1 + \cdots + \omega_{p2}X_p + \varepsilon_2 \\ &\vdots \\ F_m &= \omega_{1m}X_1 + \cdots + \omega_{pm}X_p + \varepsilon_m \end{aligned}$$

Los coeficientes ω_{ij} representan el **peso** de la variable X_i en el factor F_j . La matriz de coeficientes que permite obtener **puntuaciones de los factores** es $W_{m \times p}$:

$$W = \begin{bmatrix} \omega_{11} & \cdots & \omega_{p1} \\ \vdots & \ddots & \vdots \\ \omega_{1m} & \cdots & \omega_{pm} \end{bmatrix}.$$

Las variables $\varepsilon_1, \dots, \varepsilon_m$ recogen los **errores** cometidos al expresar los factores comunes a partir de las variables, ya que en el análisis factorial, se asume, que sólo una parte de la variabilidad es explicada por los factores comunes, es decir, existe otra porción que es explicada por un factor único, por lo tanto las variables ε_i son combinaciones lineales de los factores únicos.

Test de Esfericidad de Bartlett

El Test de esfericidad de Bartlett se utiliza para contrastar la hipótesis nula de incorrelación lineal entre las variables originales, es decir:

$$H_0 : R = I$$

contra

$$H_1 : R \neq I,$$

donde R es la matriz de correlaciones de la muestra e I es la matriz de identidad.

El estadístico de prueba para el contraste de hipótesis se define como:

$$EB = -(n - 1 - \frac{1}{6}(2p + 5)) \ln |R|,$$

donde n es la dimensionalidad y p el número de variables.

Debemos rechazar la hipótesis nula para que el análisis factorial resulte adecuado. La hipótesis nula nos dice que la matriz de correlaciones (R) es igual a la matriz identidad (I), es decir que las variables no están correlacionadas.

La hipótesis nula se rechaza para grandes valores de EB, ya que si no existe correlación significativa entre las variables $|R| = 1$ y $\ln |R| = 0$, luego cuánto más grande sea el estadístico

mayor es la probabilidad de alejarnos de la formulación de incorrelación lineal entre las variables.

Medida de adecuación muestral (KMO)

Esta prueba de bondad está basada en los coeficientes de correlación parcial entre pares de variables que denotaremos por σ_{ij}^* ; si existe correlación entre las variables, dichos coeficientes (que representan las correlaciones entre cada par de variables, tras haber eliminado la influencia del resto) debe ser casi nulo, por lo cual la medida:

$$KMO = \frac{\sum_{i \neq 0} \sigma_{ij}^2}{\sum_{i \neq j} \sigma_{ij}^2 + \sum_{i \neq 0} \sigma_{ij}^{*2}}$$

se aproxima a uno. En función del valor de KMO se determina la idoneidad del análisis factorial. Según Kaiser:

- $0,9 < KMO \leq 1$ son valores excelentes,
- $0,8 < KMO \leq 0,9$ son valores buenos,
- $0,7 < KMO \leq 0,8$ son valores aceptables,
- $0,6 < KMO \leq 0,7$ son valores regulares,
- $0,5 < KMO \leq 0,6$ son valores malos,
- $KMO \leq 0,5$ valores inaceptables.

En el análisis de factores comunes, cada variable no sólo es función de los factores comunes, sino también de factores únicos; esto significa que con las factores comunes sólo se explica una parte de la variabilidad de las variables, aquella que es común a todas.

Bajo este planteamiento, como las unicidades son distintas de cero:

$$R_{xx} = \Lambda \Lambda^t + DD^t \Rightarrow R_{xx} - DD^t = \Lambda \Lambda^t$$

$$\Lambda \Lambda^t = \begin{bmatrix} 1 - \eta_1^2 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{p2} & \cdots & 1 - \eta_p \end{bmatrix}, \quad D = \begin{bmatrix} \eta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \eta_p \end{bmatrix},$$

$$R_{xx} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}.$$

La forma en la que se determina Λ , ya sea a partir de R_{xx} o de $R_{xx} - DD^t$, requiere del cálculo de los autovalores y autovectores de la matriz correspondiente, de modo que:

$$\Lambda = V\Psi^{\frac{1}{2}}$$

donde:

- V matriz de autovectores de R_{xx} o $R_{xx} - DD^t$ en sus columnas.
- Ψ matriz diagonal con autovalores en cada elemento de la diagonal.

Determinación del número de factores (m) Para ello existen criterios diferentes:

- **Determinación a priori:** el investigador fija previamente el valor de m , independiente de los datos.
- **Autovalores mayores que 1:** Se eligen los autovectores asociados a autovalores mayores que 1. Existen dos motivos para justificar este criterio:
 - (1) La media de los valores propios es la unidad y por ello se eligen valores mayores que la media.

$$Tr(\Psi) = \lambda_1 + \dots + \lambda_p = Tr(R_{xx}) = p$$

y

$$\bar{\lambda} = \frac{\lambda_1 + \dots + \lambda_p}{p} = \frac{p}{p} = 1.$$

- (2) La varianza de cada factor se expresa como el autovalor al cual está asociado, por lo tanto se eligen autovalores con valor mayor que uno. Autovalores menores que uno no son mejores que ninguna de las variables originales.

- **Porcentaje de varianza acumulada:** todos los factores están ordenados en forma descendente de acuerdo a la varianza explicada por cada uno de ellos.

La variabilidad de los datos es $Tr(\Psi) = \lambda_1 + \dots + \lambda_p$.

Por tanto la parte de la variabilidad total explicada por cada factor se expresa como:

$$\frac{\lambda_1}{p}, \frac{\lambda_2}{p}, \dots, \frac{\lambda_m}{p}, \dots, \frac{\lambda_p}{p}.$$

Luego, la varianza total acumulada explicada por los primeros m factores es

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{p}.$$

De lo anterior, se elige un porcentaje de varianza a ser explicada por los factores, normalmente se elige mayor a un 60 %.

Análisis y Resultados

En este trabajo se hizo un estudio sobre las asesorías que realiza la organización sin fines de lucro, Acción Solidaria. La información registrada es de los años 2008, 2009 y 2010. El Centro de Información Nacional de VIH/SIDA toma en cuenta distintos aspectos para la Asesoría, de los cuales nosotros seleccionamos algunos. Los aspectos a continuación son las variables que se están midiendo y las personas que recibieron asesorías por AcSol son nuestras unidades experimentales.

(1) **Tipo de Asesoría:**

- (a) Pre-Test y Post-Test.
- (b) Información General.
- (c) Información PVVIH/SIDA.

(2) **Persona que recibe la asesoría**

- (a) P.V.VIH/SIDA.
- (b) Familiar.
- (c) Pareja.
- (d) Compañero de trabajo, estudios, etc.
- (e) Público en general.

(3) **Género**

- (a) Masculino.
- (b) Femenino.

(4) **Orientación**

- (a) Ht. (Heterosexual).
- (b) Hm. (Homosexual).
- (c) Bs. (Bisexual).

(5) **Nivel de Instrucción**

- (a) A (Profesional).

- (b) B (Semiprofesional).
- (c) C (Trabajador).
- (6) **Estado Civil**
 - (a) Soltero(a).
 - (b) Casado(a).
 - (c) Divorciado(a).
 - (d) En Pareja.
 - (e) Viudo(a).
- (7) **Edad**
 - (a) Menor de 11 años.
 - (b) Entre 12 y 18 años.
 - (c) Entre 19 y 25 años.
 - (d) Entre 26 y 45 años.
 - (e) Entre 46 y 65 años.
 - (f) Mayor a 66 años.
- (8) **Exámenes Previos**
 - (a) Todos.
 - (b) Algunos.
 - (c) Ninguno.
- (9) **Donde obtuvo la información**
 - (a) Medios/ TV.
 - (b) Radio.
 - (c) Cine.
 - (d) Marcalibros.
 - (e) Prensa.
 - (f) Tríptico.
 - (g) Amigos y/o Familiares.
 - (h) ONG's.
 - (i) C.I.N.V.I.H.
 - (j) Otros: web, médico, etc.

Cada variable registra una característica en particular y a cada registro correspondiente se le asignó un valor numérico dependiendo de la información suministrada por los asistentes. Así la variable *género*, por ejemplo, tiene dos posibles respuestas masculino y femenino, a las cuales se les asigna el valor 1 y el valor 2 respectivamente; por tanto cuando se suministra la información por el asistente se coloca el valor correspondiente y de manera análoga se procede con las distintas variables.

En este trabajo fueron aplicadas diferentes técnicas, las cuales serán tratadas cada una por separado. El mismo ha sido elaborado mediante dos programas de computación **Matlab** y **R**, los cuales son de gran utilidad para hacer cálculos estadísticos, y en otros ámbitos también. Comenzaremos con la realización de histogramas por año e histogramas múltiples, todo esto con la ayuda de **Matlab**.

Histogramas-Asesorías del año 2008

Tenemos una matriz original llamada **datos08.xls**, que se encuentra ya tabulada en MS Excel. Ésta matriz posee en la primera columna el año y en la segunda los trimestres y en las siguientes columnas se encuentran los distintos aspectos de la asesoría antes mencionados. Los datos por cada fila registran la información de una persona con respecto a las distintas variables, como se mencionó anteriormente, a estos datos se les asignó un valor numérico, sólo se admitió una información respuesta en cada variable.

Como primer dato podemos decir que durante el año 2008, Acción Solidaria recibió y atendió a 648 personas en las asesorías. La matriz **datos08.xls** fue importada al Workspace de Matlab y fue identificada como "data". Este mismo procedimiento fue realizado con los datos de los años 2009 y 2010.

A continuación los histogramas obtenidos:

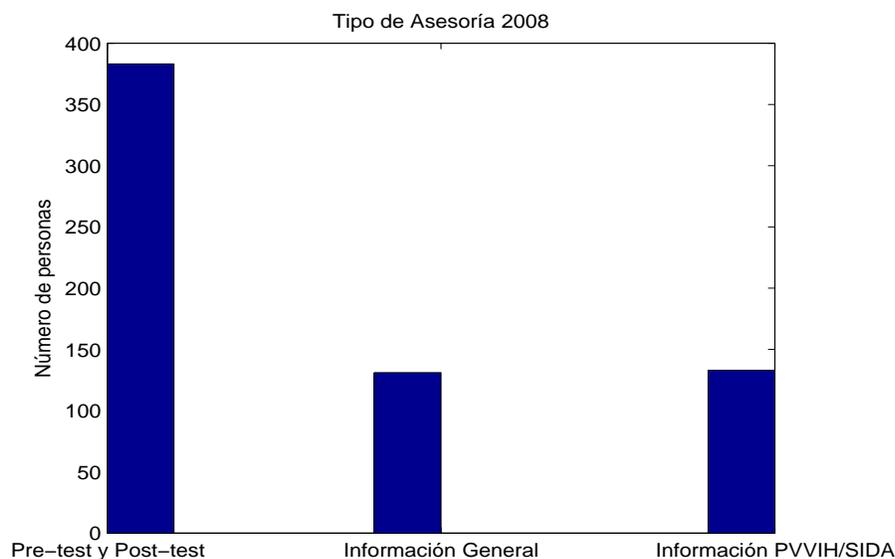


FIGURA 4.1. Histograma de frecuencias para el tipo de asesoría 2008

En la **figura 4.1** se observan los tres tipos de asesorías que puede recibir la persona y son: Pre-Test y Post-Test, Información General, Información P.V.VIH/SIDA. Se ve reflejado que existe un gran número de personas interesadas en realizarse las pruebas de Pre-Test y Post-Test. De las 648 asesorías durante el año 2008, 370 estaban interesadas en las pruebas, es decir un 57 % de los asistentes; mientras que un 20 % de los asistentes buscaban información general y sobre el PVVIH/SIDA.

Dentro de las personas que reciben asesorías se encuentran: P.V.VIH/SIDA, familiar, pareja, compañero de trabajo, estudios, etc, y público en general. Se observa en la **figura 4.2** que un 54 % de las personas que reciben asesorías es el público en general (350 personas), en segundo lugar encontramos que 200 personas con P.V.VIH/SIDA reciben asesoría, lo cual equivale al 31 % de la población asistida; esto es sumamente importante a la hora de estudiar cuantas personas con P.V.VIH/SIDA son tratadas y llevar un registro de las mismas.

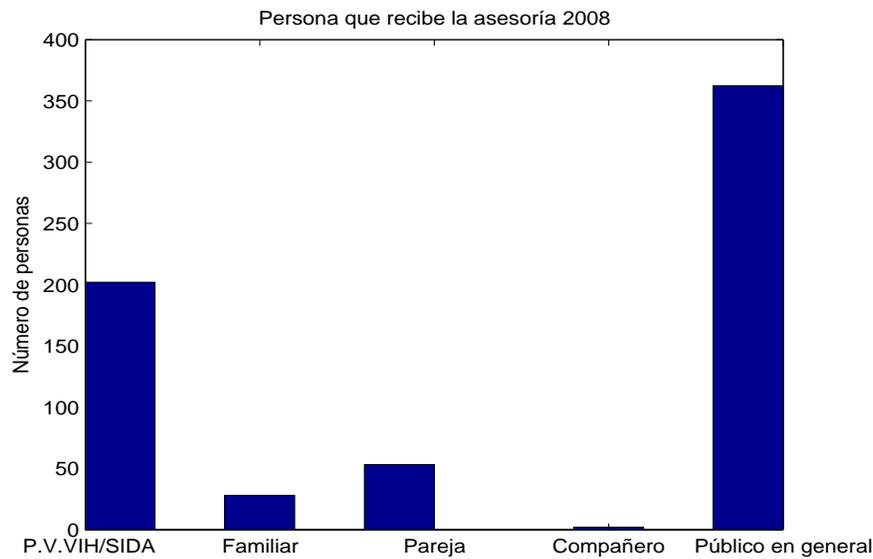


FIGURA 4.2. Histograma de frecuencias para las personas que reciben asesorías 2008

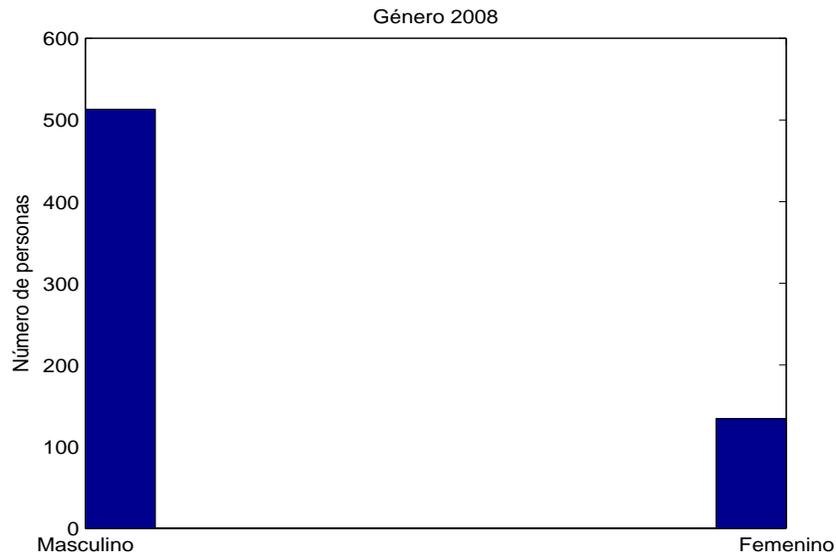


FIGURA 4.3. Histograma de frecuencias para el Género 2008

Evidenciamos en la **figura 4.3** que un 77% de los asistentes son hombres, esto puede ser debido a múltiples factores, tales como: requisitos laborales, están más informados al respecto, entre otros, mientras que un 20% de las personas asistentes son mujeres.

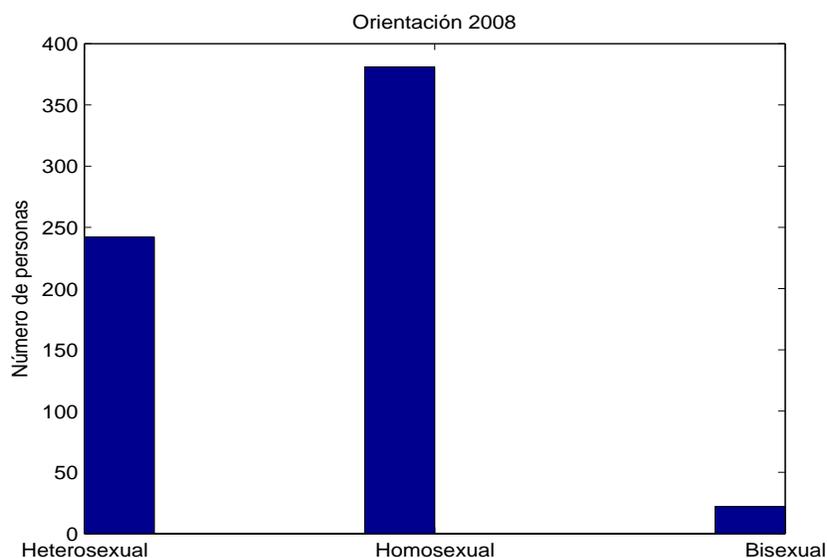


FIGURA 4.4. Histograma de frecuencias para el tipo de Orientación sexual 2008

Los tipos de orientación son Ht. (Heterosexual), Hm. (Homosexual), Bs. (Bisexual) y se reflejan en la **figura 4.4** en ese respectivo orden. Se observa que hay una gran cantidad de personas homosexuales que reciben asesorías (57%) y en segundo lugar se encuentran las personas heterosexuales (37%) que participan en las asesorías, más adelante se estudiará con profundidad que relación tiene esto con las otras variables anteriores, como género y persona que recibe la asesoría.

Los niveles de instrucción tomados en cuenta para la asesoría son: A (Profesional), B (Semiprofesional), C (Trabajador) y se observa en la **figura 4.5** que la mayoría son personas semiprofesionales, superior a las 530 personas lo cual equivale al 82% de los asistidos en el año 2008, las cuales poseen un técnico o son estudiantes universitarios, entre otros, mientras que las personas profesionales y trabajadores que reciben asesoría es un número muy reducido de 50 personas (8%).

El estado civil de la persona es de suma importancia, pues eso ayudará a orientarla de una forma u otra y a su vez a tener un apoyo. Dentro de los estados civiles encontramos los siguientes: Soltero(a), Casado(a), Divorciado(a), En Pareja y Viudo(a).

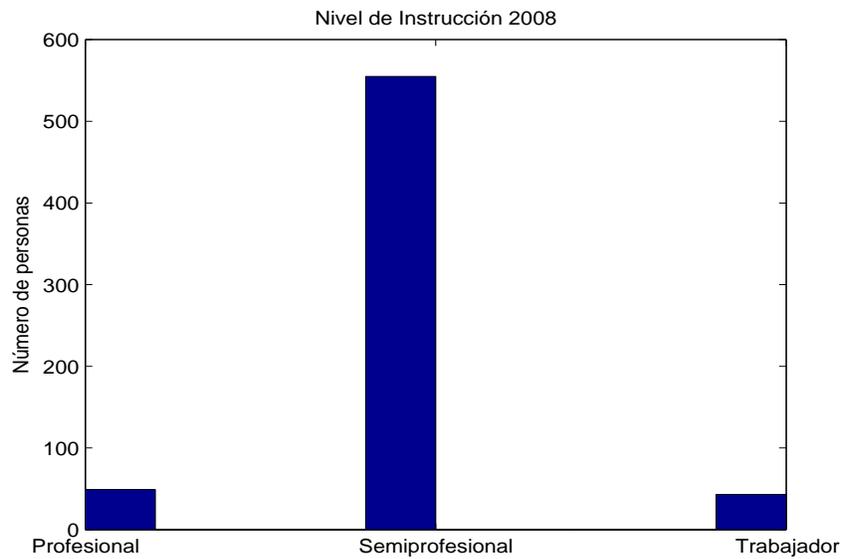


FIGURA 4.5. Histograma de frecuencias para el nivel de Instrucción 2008

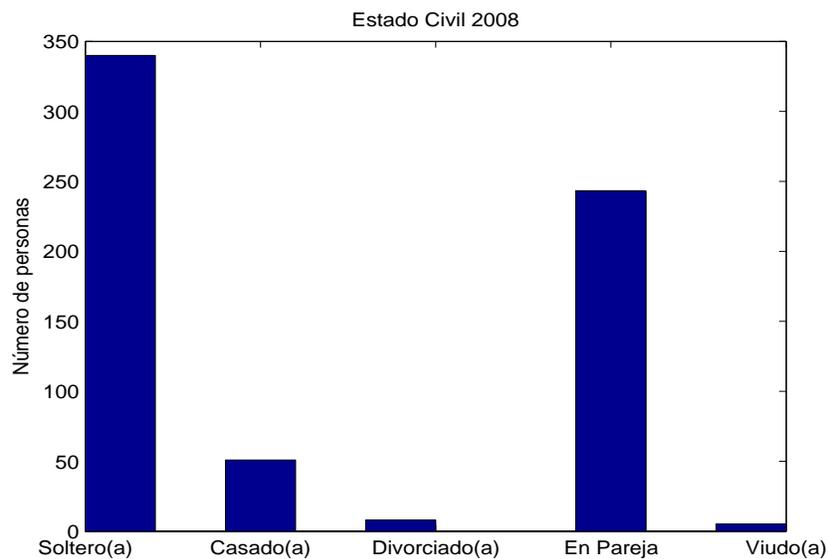


FIGURA 4.6. Histograma de frecuencias para el estado civil 2008

En la **figura 4.6** observamos que 340 personas (52 %) que recibieron asesorías son solteras, y a su vez, casi 250 personas (39 %) viven en pareja. Se ve reflejado que tanto las personas divorciadas como viudas no acuden a las asesorías.

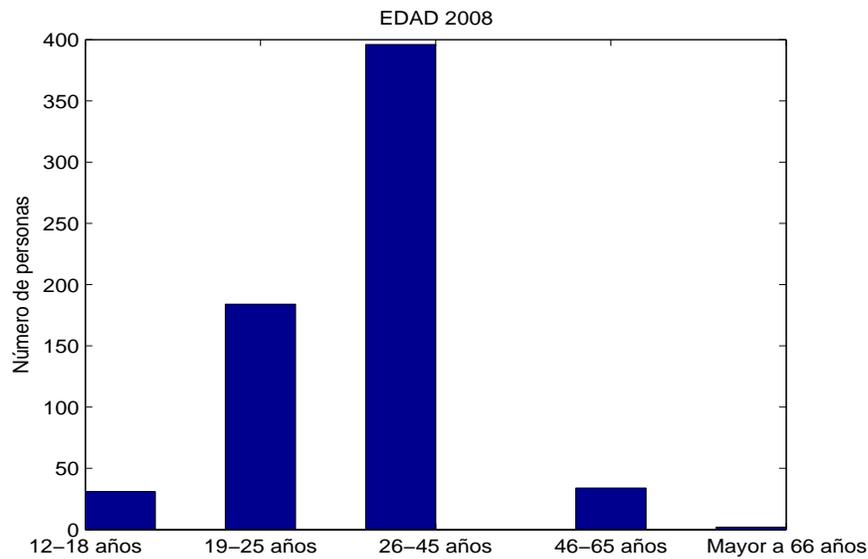


FIGURA 4.7. Histograma de frecuencias para los intervalos de edad 2008

El grupo de edades estudiadas en las asesorías se refleja en la **figura 4.7**, y se clasificaron de la siguiente forma: menor de 11 años, entre 12 y 18 años, entre 19 y 25 años, entre 26 y 45 años, entre 46 y 65 años, mayor a 66 años, evidenciándose que los jóvenes entre 26 y 45 años son los que más reciben asesorías, casi 400 personas dentro del rango antes mencionado son asistidas en la institución de Acción Solidaria, lo cual es equivalente al 62% de la población asistida. A su vez observamos que pocas personas mayores a 46 años acuden a la organización (sólo un 6%), lo mismo que menores de 18 años.

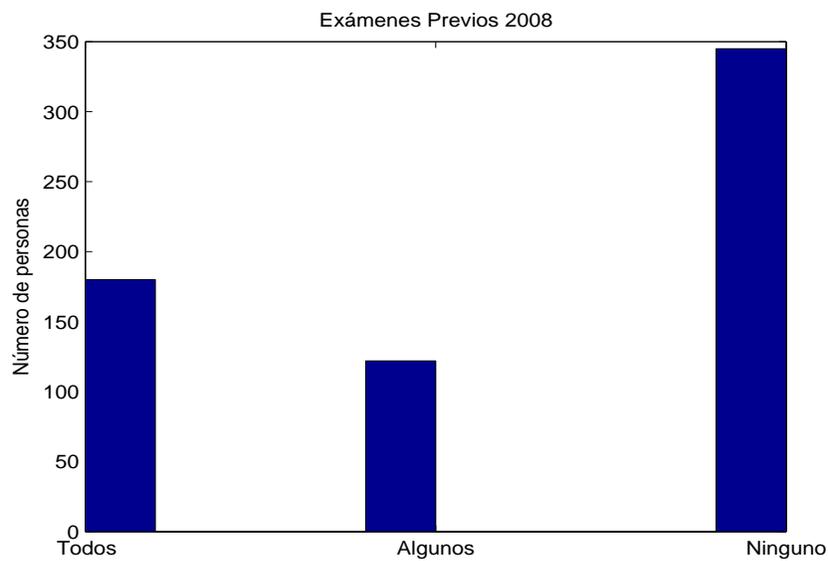


FIGURA 4.8. Histograma de frecuencias sobre la cantidad de exámenes previos realizados 2008

Los exámenes previos que se ha realizado una persona son de gran importancia para seguir un control de su progreso, y además para determinar si contienen el virus P.V.VIH/SIDA, pues éste no se detecta inmediatamente, sino hay un período en el cual el virus se empieza a manifestar. En la **figura 4.8** clasificamos los exámenes previos de la siguiente forma: Todos, Algunos, Ninguno. Las personas que poseen P.V.VIH/SIDA ya se han realizado todos los exámenes previos, pues de esa forma saben en cual condición se encuentran, mientras que las personas que no tienen un examen previo indica que es la primera vez que recurren a una institución (en este caso Acción Solidaria) y por lo general buscan mas información. Se observa que 340 personas (52 %) aproximadamente no tiene ningún examen previo, mientras que aproximadamente 170 personas (26 %) ya se han realizado todos los exámenes previos.

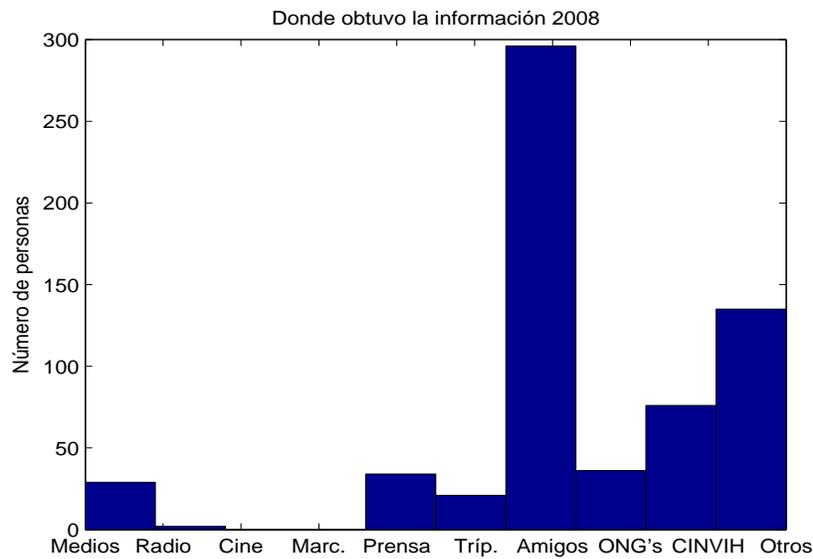


FIGURA 4.9. Histograma de frecuencias sobre dónde obtuvo la información 2008

El medio por el cual se obtuvo la información sobre la institución Acción Solidaria es de gran importancia para reforzar estos canales de comunicación, más aún para brindarle a la institución un índice de cual medio se debería trabajar más. En la **figura 4.9** se clasificó así: Medios/ TV, Radio, Cine, Marcalibros, Prensa, Tríptico, Amigos y/o Familiares, ONG's, C.I.N.V.I.H, Otros: web, médico, etc. Se evidenció que el 46% de personas que recibieron asesorías obtuvieron información mediante un amigo y/o familiar, y como segunda opción se obtuvo que un 22% de los asistentes obtienen información con otros medios, es decir ya sea vía web, médico, entre otras opciones.

Histogramas-Asesorías del año 2009

En el año 2009, Acción Solidaria atendió a 696 personas, es decir un 7% más que el año 2008. Estos datos fueron analizados y a continuación los histogramas obtenidos:

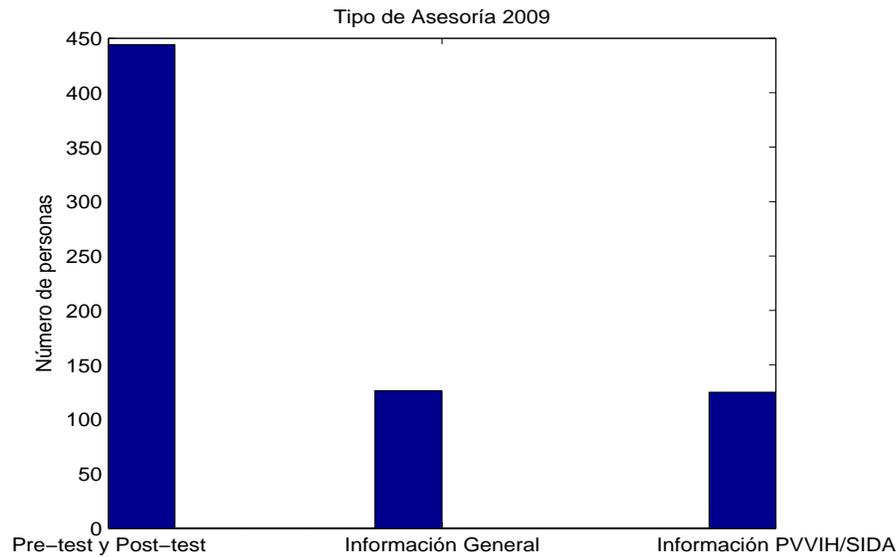


FIGURA 4.10. Histograma de frecuencias para el tipo de asesoría 2009

En la **figura 4.10** se observan los tres tipos de asesorías que puede recibir la persona y son: Pre-Test y Post-Test, Información General, Información PVVIH/SIDA. Se ve reflejado que existe un gran número de personas interesadas en realizarse las pruebas de Pre-Test y Post-Test. De las 696 asesorías durante el año 2009, aproximadamente el 63% de las personas asisten por la asesoría de los test, mientras que el 19% y 18% de las personas buscaban información general y sobre el PVVIH/SIDA respectivamente, un número similar al año pasado (2008).

Las personas que reciben asesorías son: P.V.VIH/SIDA, Familiar, pareja, compañero de trabajo, estudios, etc, y público en general. Se observa en la **figura 4.11** que mayormente las personas que reciben asesorías es el público en general (63%), en segundo lugar encontramos que aproximadamente 170 personas con P.V.VIH/SIDA recibieron asesoría, es decir un 24%, lo cual es un número significativo de personas con esta condición.

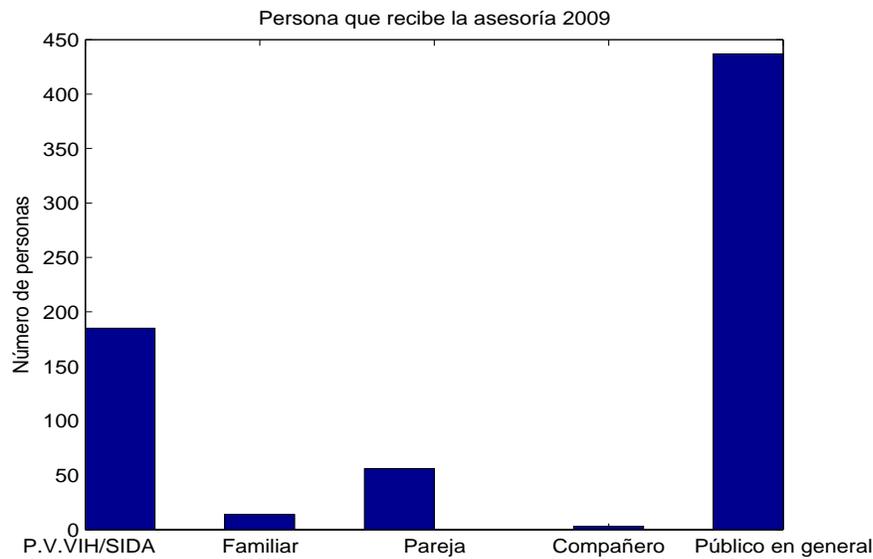


FIGURA 4.11. Histograma de frecuencias para las personas que reciben asesorías 2009

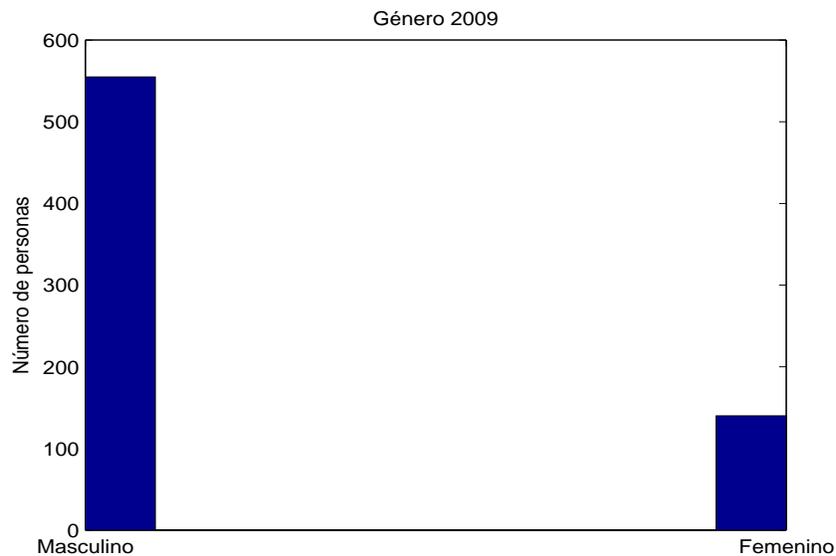


FIGURA 4.12. Histograma de frecuencias para el Género 2009

En la **figura 4.12** se observa que el 80 % de los que recibieron asesorías son hombres, es decir 60 % más que la cantidad de mujeres que recibieron asesoría, lo cual también se vio reflejado en el año 2008.

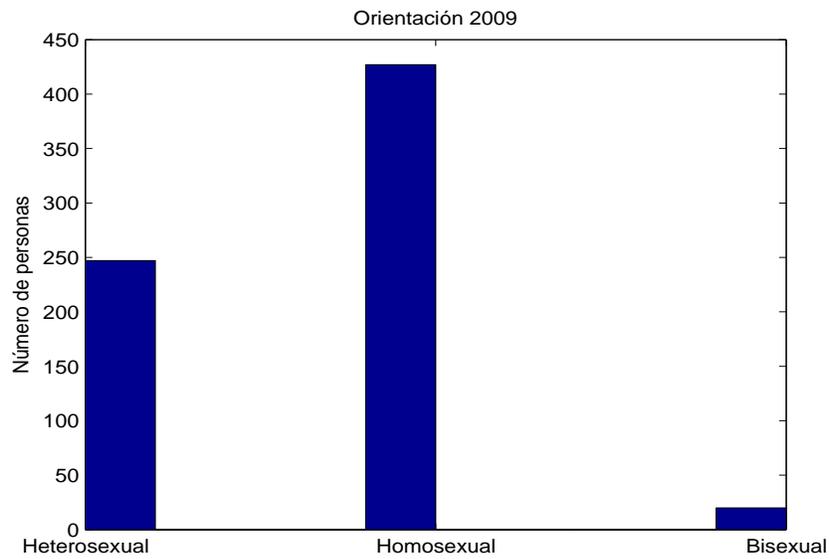


FIGURA 4.13. Histograma de frecuencias para el tipo de Orientación sexual 2009

En la **figura 4.13** se refleja un 62 % de personas con orientación sexual homosexual que recibieron asesorías y de segundo lugar se encuentran las personas heterosexuales que participaron en las asesorías, alrededor del 36 % de personas, en último lugar están las personas cuya orientación es bisexual, registrando un 3%. Estos datos tienen gran similitud con el año 2008, y lo que llama la atención es que en el año 2009 existe una mayor cantidad de personas registradas cuya orientación es homosexual comparada con el 2008.



FIGURA 4.14. Histograma de frecuencias para el nivel de Instrucción 2009

El nivel de instrucción que poseen la mayoría de las personas que reciben asesoría es el Semiprofesional registrando un 86 %, tal como se ve en la **figura 4.14**, el número es superior que en el año 2008, porcentualmente 4 % más, mientras que el número de personas que son profesionales y trabajadores representan el 4 % y el 9 % de los asistentes respectivamente.

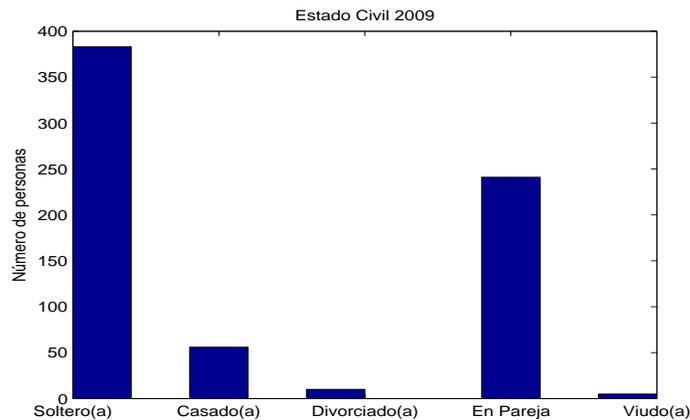


FIGURA 4.15. Histograma de frecuencias para el estado civil 2009

Observamos en la **figura 4.15** que el 55 % de las personas son solteras, mientras que el 36 % de las personas están en una relación de pareja. Tal comportamiento es muy similar al año 2008. Teniendo una cantidad menor de personas que estén casadas, divorciadas o viudas representando el 7 %, 1 % y 0.6 % de las personas asistentes respectivamente.

Las personas que más reciben asesorías tienen entre 26 y 45 años (representando el 62 % de los asistentes), tal como se muestra en la **figura 4.16**. En este caso particular, también un grupo joven entre los 19 y 25 años recibió asesoría y encontramos que representan el 29 % de las personas. Los demás grupos de edades tienen una cantidad inferior de personas asistidas, inferior al 6 % de personas en cada grupo.

En la **figura 4.17** observamos que mayormente tenemos una población de personas que no se han realizado ningún examen previamente (57 %), por tanto puede ser que es la primera vez que asisten a la institución. En cuanto a la cantidad de personas que han realizado todos los exámenes previos, encontramos que no supera el 20 % de personas.

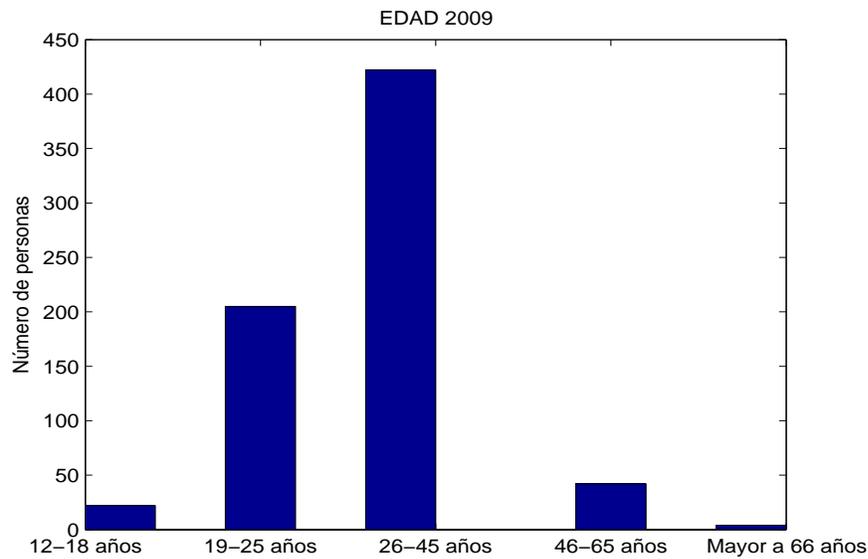


FIGURA 4.16. Histograma de frecuencias para los intervalos de edad 2009

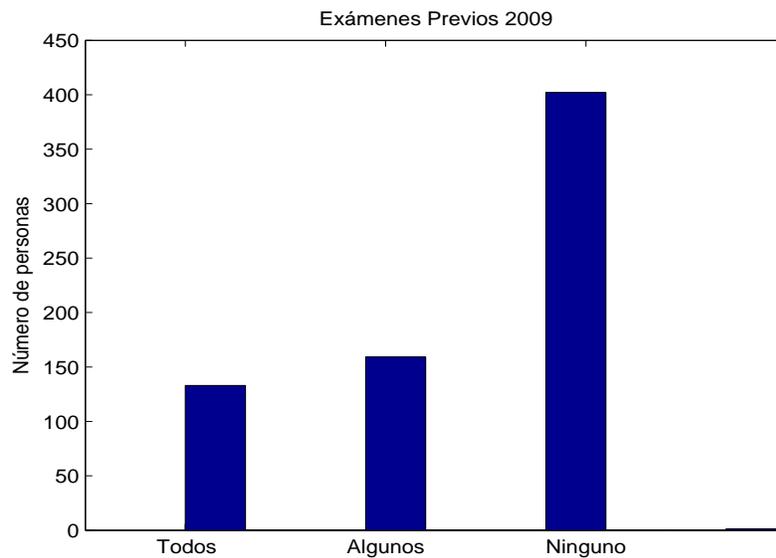


FIGURA 4.17. Histograma de frecuencias sobre la cantidad de exámenes previos realizados 2009

En el año 2009, el medio de comunicación que las personas más utilizaron para informarse fue la referencia de Amigos y/o familiares, obteniendo el 45% de personas (**figura 4.18**).

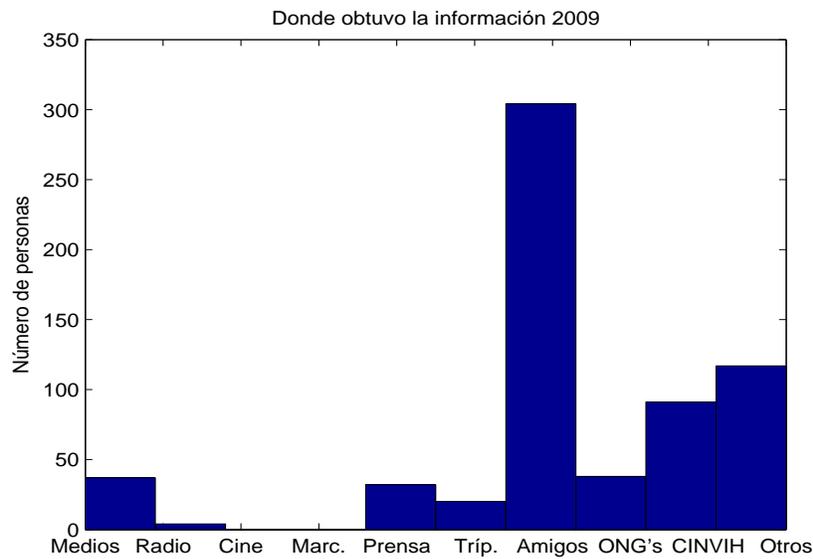


FIGURA 4.18. Histograma de frecuencias sobre dónde obtuvo la información 2009

Al igual que el año 2008, las personas utilizan otras fuentes, tales como internet, médico, entre otros (18 %). Y situándose de tercer lugar C.I.N.V.I.H (14 %). El cine y los marcalibros no son un medio de comunicación que fue utilizado para informarse por las personas que recibieron asesoría en el año 2009.

Histogramas-Asesorías del año 2010

En el año 2010, Acción Solidaria atendió a 604 personas, es decir un 15% menos que en el 2009. Estos datos fueron analizados y a continuación los histogramas obtenidos:

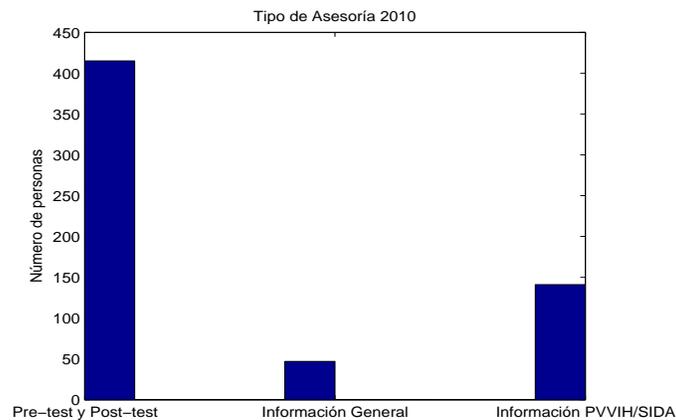


FIGURA 4.19. Histograma de frecuencias para el tipo de asesoría 2010

Las asesorías que más reciben las personas es de pruebas (**figura 4.19**), el 70% de las 604 asistidas en el año 2010. Las asesorías de información P.V.VIH/SIDA representan el 23% y la información general tiene una cantidad menor que no supera el 8%.

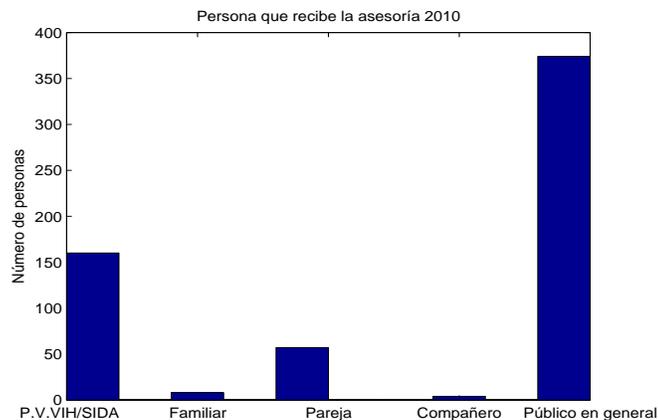


FIGURA 4.20. Histograma de frecuencias para las personas que reciben asesorías 2010

Las personas que reciben asesorías son el público general con una cantidad de 370 personas es decir el 60 %. Mientras que las personas con PVVIH/SIDA también reciben asesoría y se registra una cantidad del 25 % de las personas. Véase **figura 4.20**

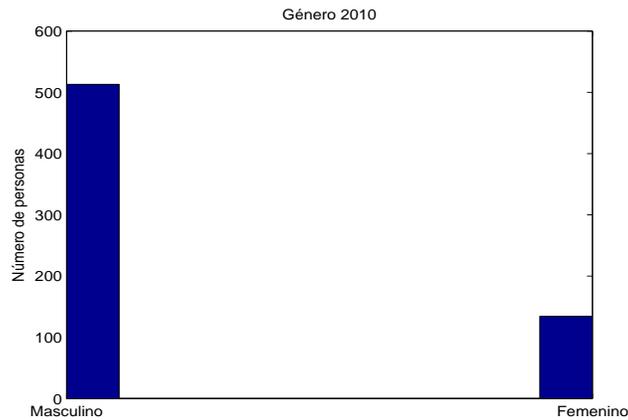


FIGURA 4.21. Histograma de frecuencias para el Género 2010

Al igual que en años anteriores, tenemos que mayormente son hombres los que asisten a la institución Acción Solidaria y reciben asesoría. Observamos un 60 % más de la cantidad de hombres a la de mujeres que reciben asesorías. (**figura 4.21**). Este comportamiento es similar al de los años 2008 y 2009.

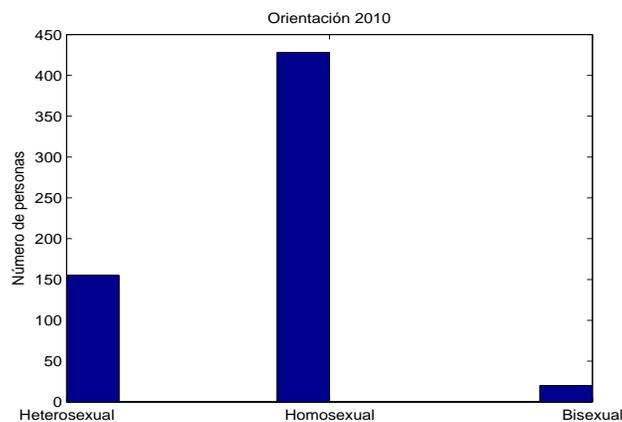


FIGURA 4.22. Histograma de frecuencias para el tipo de Orientación sexual 2010

Se observa que el 71 % de personas (**figura 4.22**) tienen orientación homosexual, tal y como se evidenció en los años 2008 y 2009, pero lo que si podemos notar es que la cantidad de personas que tiene orientación heterosexual representa el 25 %, número inferior a años anteriores. Esto nos indica que en años más recientes las personas de orientación heterosexual han disminuido sus visitas de asesorías a la institución, es una situación que puede llegar a repetirse.

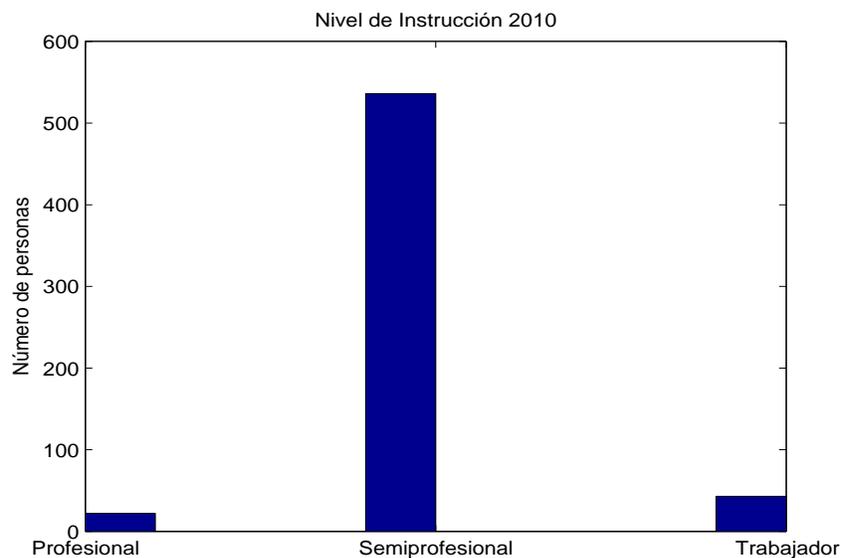


FIGURA 4.23. Histograma de frecuencias para el nivel de Instrucción 2010

El nivel de instrucción que tiene la mayoría de las personas que reciben asesorías es semiprofesional representando un 91 % (**figura 4.23**), y un número muy inferior de personas profesionales (3 %) y trabajadores (5 %). Podemos inferir que las personas semiprofesionales están más interesados por recibir asesorías, mientras que los profesionales y trabajadores dedican su tiempo a otras actividades o no llegan a informarse en AcSol.

En la **figura 4.24** en particular tenemos que el estado civil de la mayoría de las personas que reciben asesorías es soltero(a), y lo que también se ve reflejado es el hecho de que hay un gran número de personas que viven en pareja, ambas condiciones destacan más que los demás estados civiles, representando el 48 % y 45 % de las personas que recibieron asesorías respectivamente.

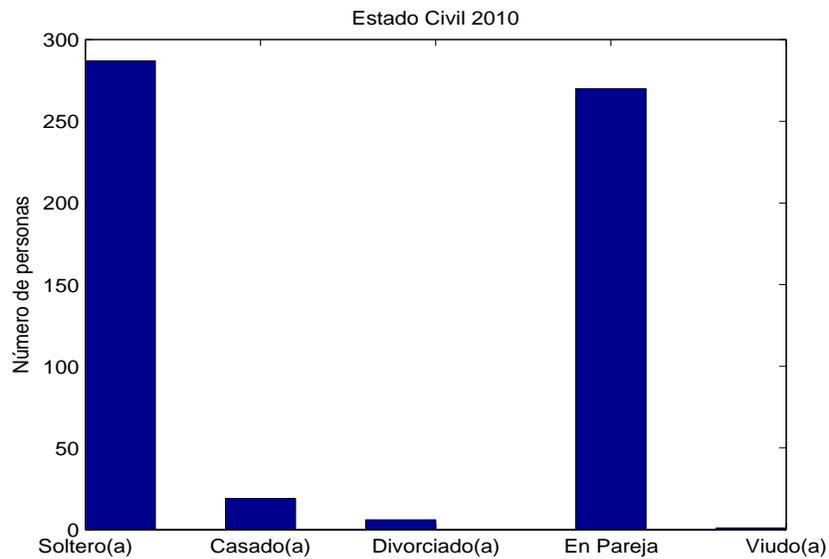


FIGURA 4.24. Histograma de frecuencias para el estado civil 2010

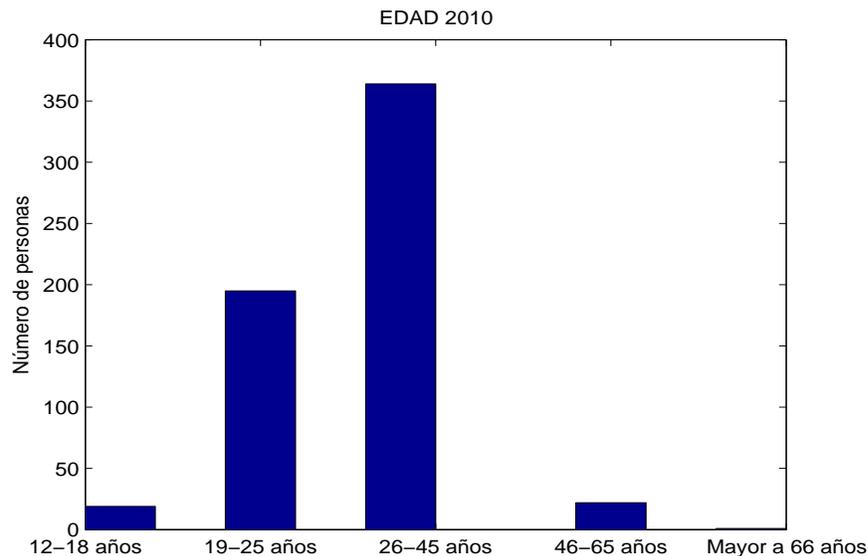


FIGURA 4.25. Histograma de frecuencias para los intervalos de edad 2010

En la **figura 4.25** la mayor cantidad de personas que asisten a las asesorías son jóvenes que tienen edades entre 26 y 45 años. Se observa que el 60 % de los asistentes se encuentran en el grupo antes mencionado, seguidamente el 33 % de las personas se encuentran el rango de edades entre 19 y 25 años.

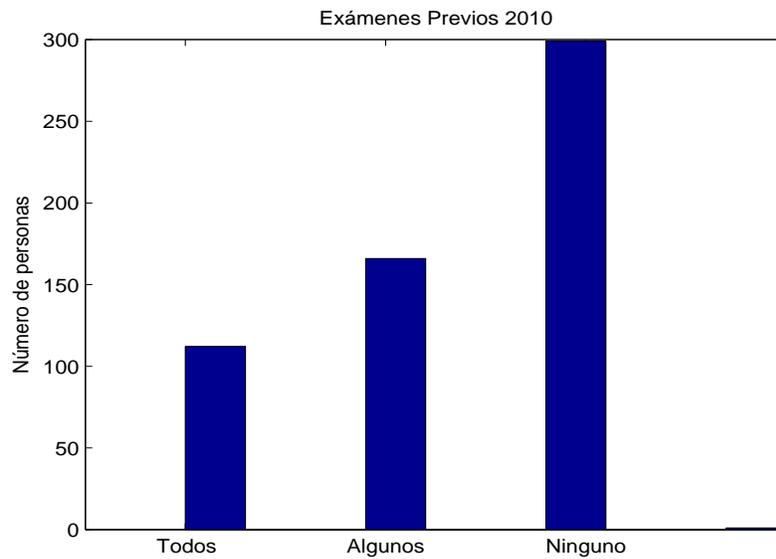


FIGURA 4.26. Histograma de frecuencias sobre la cantidad de exámenes previos realizados 2010

Los exámenes previos que pueden realizarse son todos, algunos o ninguno; en ese respectivo orden se encuentran en la **figura 4.26**, donde vemos reflejado que la cantidad de personas que no se han realizado ningún examen es superior al 50 %, y en segundo lugar encontramos a las personas que se han hecho algunos exámenes (28 %). Cabe destacar que el 25 % de los asistentes son personas con VIH/SIDA, tal como se evidencia en **figura 4.20** y los que se han realizado todos los exámenes (portadores del VIH/SIDA) representan el 20 %.

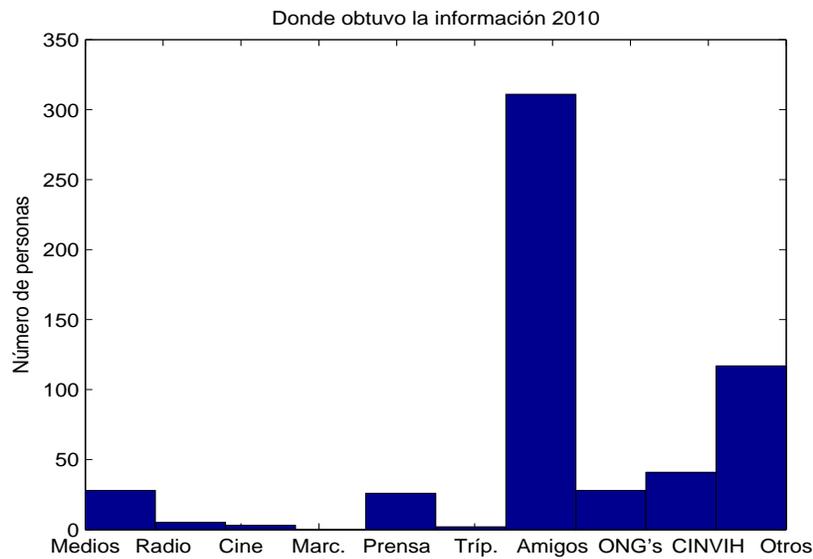


FIGURA 4.27. Histograma de frecuencias sobre dónde obtuvo la información 2010

La mayoría de las personas que recibieron asesorías durante el año 2010 obtuvieron información sobre la institución Acción Solidaria mediante un amigo y/o familiar mostrando un incremento porcentualmente de 8% con respecto al año 2009 y al igual que en años anteriores se tiene que el segundo medio por el cual se enteraron es otros: web, médico, etc. (figura 4.27)

1. Histogramas Múltiples

Histogramas Múltiples 2008, 2009, 2010

Los siguientes histogramas múltiples se realizaron para ver las características resaltantes en un solo bloque, es decir, todos los análisis realizados anteriormente, característica por característica, son reflejados en un solo histograma múltiple por año.

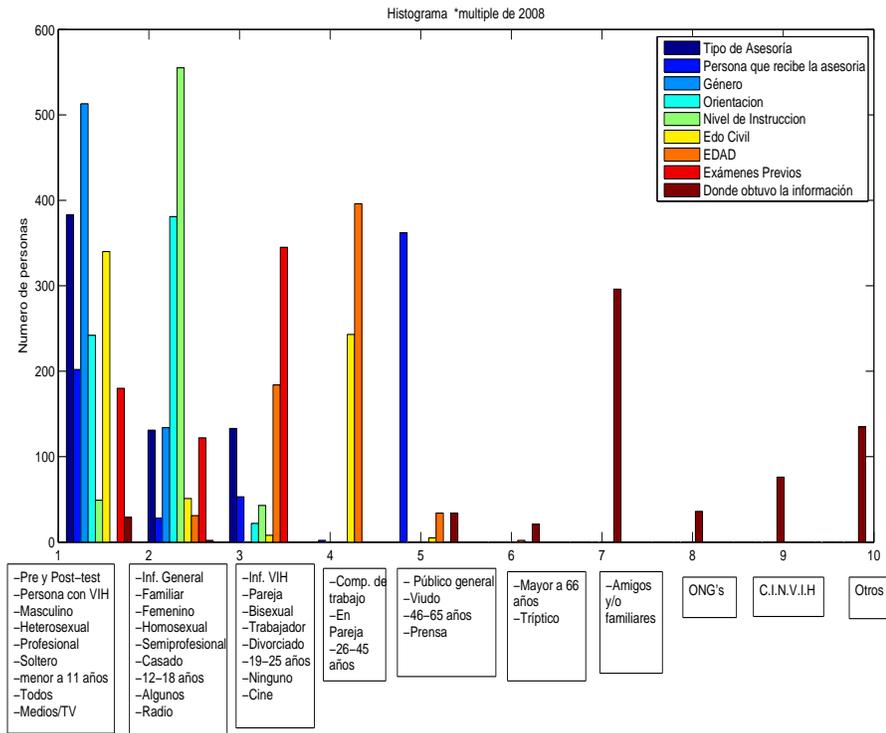


FIGURA 4.28. Histograma Múltiples 2008

En los histogramas múltiples lo que se desea observar son las diferentes variables reflejadas en contraste con las otras, es así como vemos como se agrupan mayormente en una determinada característica. En el año 2008 (figura 4.28), observamos lo siguiente: la mayoría de las personas vienen por el Pre-test y Post-test, la personas que mayormente reciben asesoría es el público general, los hombres son los que más acuden a la institución, la orientación sexual es homosexual, el nivel de instrucción es semiprofesional, el estado civil es soltero, las personas que tienen entre 26 y 45 años son los que más asesoría reciben, además hay una

gran cantidad de personas que no se han hecho algún exámen previamente y para finalizar, las personas encuentran información mayormente mediante sus amigos y/o familiares.

Toda esta información obtenida se ve reflejada en los análisis realizados previamente. Dado que la orientación sexual homosexual y el rango de edad entre 26 y 45 años son las características más resaltantes, podemos inferir que esto deberá estar estrechamente relacionado. Es muy importante tomar en cuenta estos resultados, para así atender de una manera más eficiente a los asistentes a las asesorías.

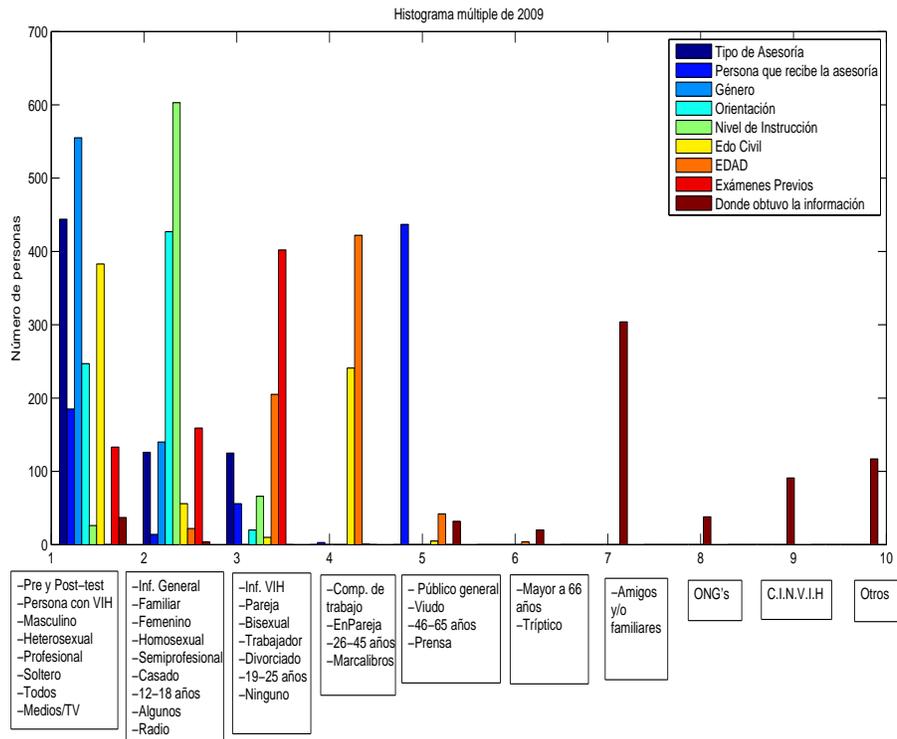


FIGURA 4.29. Histograma Múltiples 2009

En el histograma del 2009 (figura 4.29) tenemos que el tipo de asesoría que recibieron la mayoría de las personas es sobre pruebas. Se ve reflejado una gran cantidad de hombres que recibieron asesorías en el año 2009. Además observamos otras características resaltantes tales como el hecho de que la mayoría de los asistentes hayan sido solteros, cuya orientación sexual predominante sea homosexual, semiprofesionales aproximadamente 600 personas de los asistentes. Teniendo esta información se podrían dirigir esfuerzos en brindar asesorías para aquellos sectores que no resaltaron en los análisis.

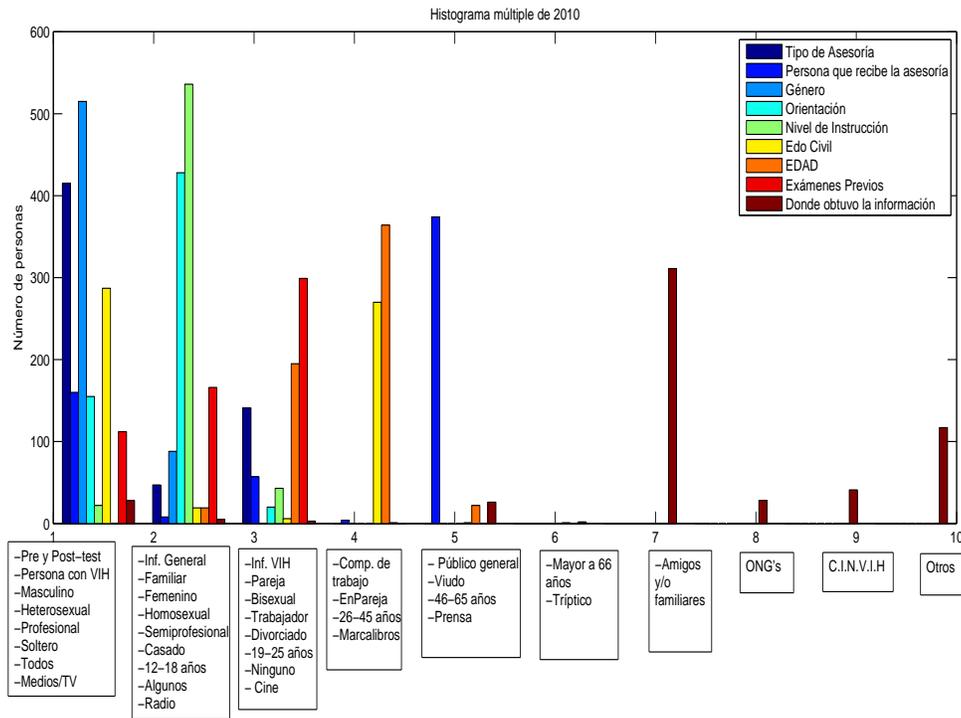


FIGURA 4.30. Histograma Múltiples 2010

En el histograma múltiple del año 2010 (**figura 4.30**) encontramos que el tipo de asesoría al cual más recurren es el de Pre-test y Post-test, se registra que el público general es el que más asesoría recibe, mayormente hombres, la orientación con más índice es homosexual, el nivel de instrucción predominante es semiprofesional, el estado civil de los que acuden a las asesorías es soltero aunque el estado en pareja posee una cantidad muy similar al estado civil soltero registrandose más de 250 personas. Las edades de los que reciben asesorías oscilan entre 26 y 45 años, dato que se registró también en los años 2009 y 2010. Encontramos una gran cantidad de personas sin exámenes previos, y para finalizar encontramos que mayormente las personas obtienen información sobre la institución mediante un amigo y/o familiar. Estas características antes mencionadas se repiten en los años 2008 y 2009.

2. Análisis y Resultados del ACP y AF

Las matrices de datos de los años 2008, 2009 y 2010 fueron tabuladas a Excel. Estos datos presentaron inicialmente mucha información faltante. Un dato faltante se denomina NAN, esto se debe a sus siglas en inglés Not A Number, lo que se traduce como no hay ó no es un número. Para cubrir estos espacios faltantes se hizo una suma total de los datos existentes de cada tipo y los faltantes, teniendo una relación directa por característica. Luego de esto, proporcionalmente se completaron las tablas de datos.

Los datos de las asesorías del año 2008 se encuentran guardados en la matriz de datos 'Datos08.xls' que se encuentra ya tabulada en MS Excel. Ésta fue importada al programa **Matlab** para poder ser trabajada. Esta Matriz posee bastante información, pues las primeras 2 columnas son los registros sobre el año y trimestre, vamos a utilizar las demás características que mencionamos al inicio de nuestros análisis. Se tomó en cuenta desde la columna 3 a la 11, es decir : tipo de asesoría, persona que recibe asesoría, género, orientación, nivel de instrucción ,estado civil, edad, exámenes previos y donde obtuvo la información.

Se procede a calcular las matrices de varianza-covarianza, correlación, medias, desviación estándar, autovalores y demás elementos del análisis de componentes principales con el programa computacional **Matlab**.

<i>Correlaciones</i>	Tipo de asesoría	Persona que recibe asesoría	Género	Orientación	Nivel de Instrucción	Edo. Civil	Edad	Exámenes previos	Donde obtuvo la información
Tipo de Asesoría	1	-0.767	-0.001	0.005	0.039	-0.142	0.184	-0.644	0.037
Persona que recibe asesoría	-0.767	1	-0.062	-0.015	-0.065	-0.0007	-0.205	0.755	-0.054
Género	-0.001	-0.062	1	-0.509	0.032	0.074	0.069	0.061	0.052
Orientación	0.005	-0.015	-0.509	1	-0.061	-0.058	-0.116	-0.117	0.034
Nivel de Instrucción	0.039	-0.065	0.032	-0.061	1	0.001	0.006	0.002	0.090
Edo. Civil	-0.142	0.0007	0.074	-0.058	0.001	1	0.056	0.006	0.002
Edad	0.184	-0.205	0.069	-0.116	0.006	0.056	1	-0.132	-0.087
Exámenes previos	-0.644	0.755	0.061	-0.117	0.002	0.006	-0.132	1	-0.055
Donde obtuvo la información	0.037	-0.054	0.052	0.034	0.090	0.002	-0.087	-0.055	1

CUADRO 1. Matriz de correlaciones del año 2008

Al observar la matriz de correlaciones correspondiente al año 2008 (**cuadro 1**), encontramos fuertes correlaciones entre algunos pares de variables, tales como el tipo de asesoría

y la personas que recibe la asesoría. El coeficiente de correlación entre ellas es de -0.767 . También el tipo de asesoría con los exámenes previos arrojó un coeficiente de correlación de -0.644 . Adicionalmente la persona que recibe asesoría y los exámenes previos demuestran una fuerte correlación de 0.755 , así podemos agrupar estas características en un grupo de alta correlación. En donde estarían el tipo de asesoría, persona que recibe la asesoría y los exámenes previos. A su vez, el género y la orientación tienen una correlación de -0.509 , lo cual también es alto, y estarían ubicados en un segundo grupo.

	Tipo de asesoría	Persona que recibe asesoría	Género	Orientación	Nivel de Instrucción	Edo. Civil	Edad	Exámenes previos	Donde obtuvo la información
Varianzas	4.456	4.004	2.060	0.449	0.352	0.253	0.219	0.138	0.096

CUADRO 2. Varianzas del año 2008

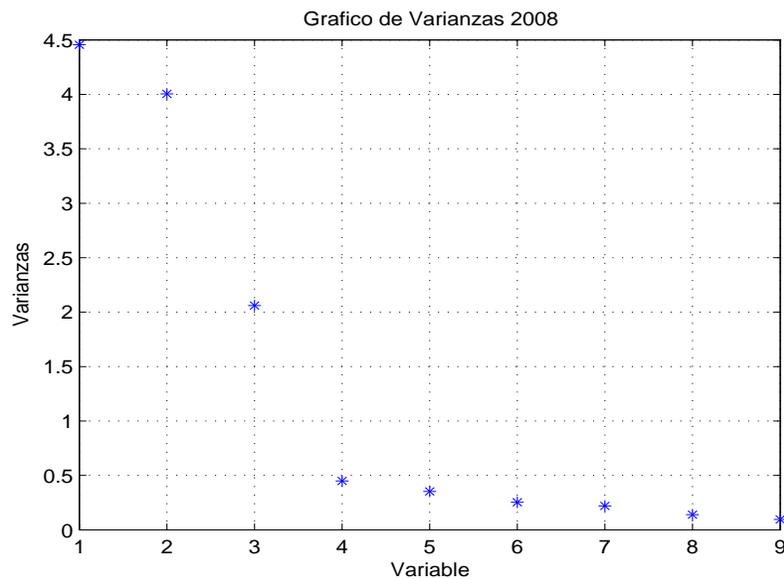


FIGURA 4.31. Gráfico de varianzas 2008

Al realizar un ACP, se desea determinar la dimensionalidad real del espacio en el que caen los datos, es decir el número de variables que tienen varianzas mayores que 1. Al observar la gráfica se hallan tres componentes cuyas varianzas se encuentran superiores a 1, así a partir

de la cuarta unidad, cada uno de los siguientes se acerca a cero. Lo que podemos concluir de esta gráfica, hasta los momentos, es que tres componentes explicarían bien los datos.

A continuación, se calculó la matriz de autovalores valiéndonos de la matriz de correlación R se realizó el gráfico. Los autovalores obtenidos son:

	Tipo de asesoría	Persona que recibe asesoría	Género	Orientación	Nivel de Instrucción	Edo. Civil	Edad	Exámenes previos	Donde obtuvo la información
<i>Autovalor</i>	0.184	0.327	0.482	0.822	0.948	1.008	1.126	1.580	2.518

CUADRO 3. Autovalores del año 2008

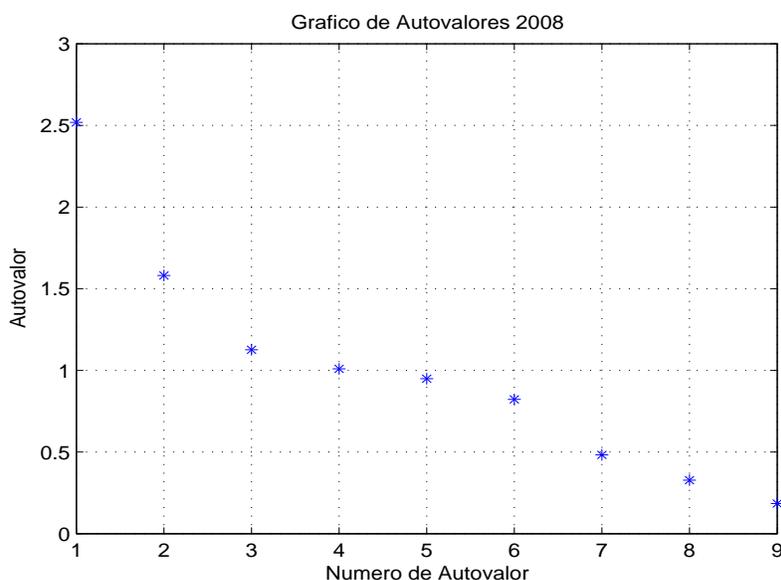


FIGURA 4.32. Gráfico de Autovalores 2008

En el gráfico de autovalores (**figura 4.32**) se muestra claramente tres autovalores superiores al valor 1 y un autovalor muy cercano a este valor. Como se observó en el **cuadro 3** los autovalores superiores al valor 1 son 2,518, 1,580, 1,126, 1,008. Más adelante veremos que éste último autovalor no explica lo suficiente para ser utilizado, tal como se observa en el **cuadro 5**.

	Tipo de asesoría	Persona que recibe asesoría	Género	Orientación	Nivel de Instrucción	Edo. Civil	Edad	Exámenes previos	Donde obtuvo la información
<i>Impj</i>	0.015	0.027	0.040	0.068	0.078	0.083	0.093	0.131	0.209

CUADRO 4. Carga de componentes del año 2008

Cada componente incide de una u otra manera a la hora de hacer un análisis, para saber de que manera inciden calculamos: $Imp_j = \frac{autovalor_j}{traza(\Sigma)}$, donde Σ es la matriz de varianzas-covarianzas, es decir vamos a medir la importancia de la j-ésima componente principal.

Se observa en el **cuadro 4** que las dos últimas componentes tienen un valor numérico de mayor importancia que las otras componentes, y se encuentran en el orden de aparición de la matriz de datos.

A continuación presentamos el **cuadro 5**, en el cual se ven reflejadas la desviación estándar, la proporción de varianza y la proporción acumulativa. Este análisis fue realizado con el programa **R**.

Importancia Componentes	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8	Comp. 9
Desviación Estándar	1.59	1.257	1.061	1.004	0.974	0.9068	0.6943	0.5727	0.4296
Proporción de varianza	0.28	0.176	0.125	0.112	0.105	0.0914	0.0536	0.0364	0.0205
Proporción acumulativa	0.28	0.455	0.581	0.693	0.798	0.8895	0.9431	0.9795	1.000

CUADRO 5. Importancia de Componentes del año 2008

Con esta información obtenida, consideramos tres componentes principales, con las cuales se explica un 58% de la variabilidad de los datos.

Adicionalmente se hizo el gráfico de las cargas de las tres componentes principales, obteniendo cuales características explican más a cada componente.

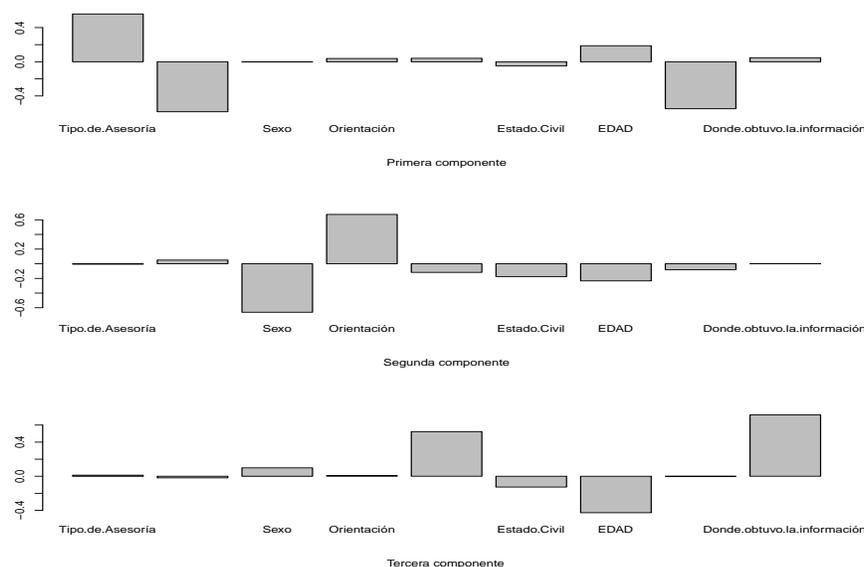


FIGURA 4.33. Gráfico de la carga de las tres componentes principales 2008

El tipo de asesoría, la persona que recibe la asesoría y los exámenes previos se encuentran resaltantes en la primera componente de la (figura 4.33), es decir éstas tres características mencionadas aportan una mayor carga a esta componente. Mientras que el sexo (género) y la orientación sexual aportan una carga significativa a la segunda componente. El nivel de instrucción, la edad y donde obtuvo la información explican y aportan más a la tercera componente.

Las componentes indican, como mencionamos anteriormente, la dimensionalidad del espacio en el cual estamos trabajando, en particular, tres componentes agrupan las características de toda la data, y lo que se observa son tres grupos. La (figura 4.33) nos indica cual es el aporte de carga de cada característica en relación a la componente.

Finalmente los gráficos de componentes por pares son los siguientes:

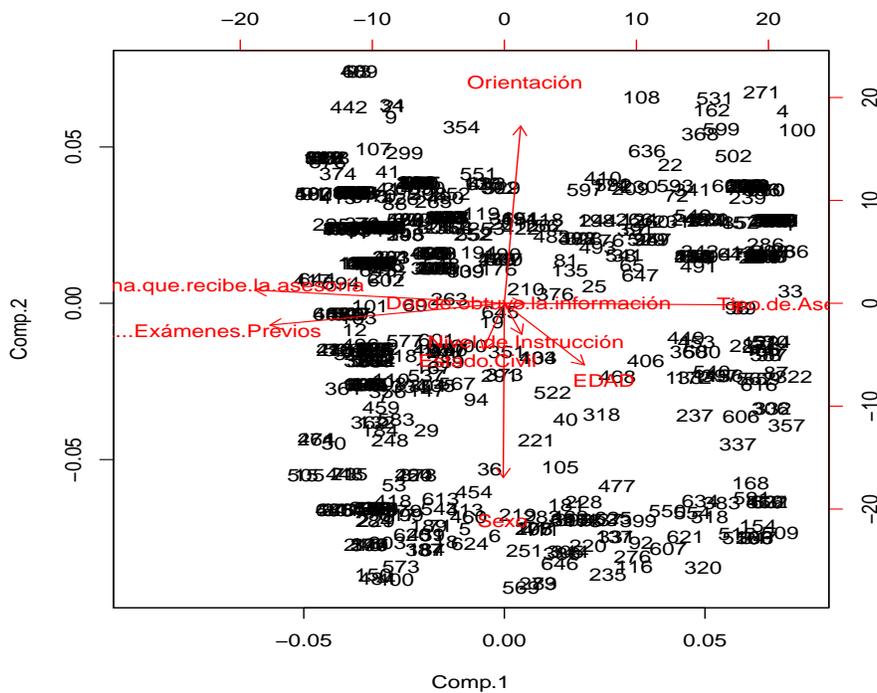


FIGURA 4.34. Gráfico de componente 1 vs. componente 2

En la figura 4.34, observamos como se encuentran distribuidos los datos. Encontramos dos grandes grupos diferenciados. Un grupo de personas que están caracterizadas negativamente por la componente 1 y positivamente por la componente 2, las cuales tienen como características la orientación y personas que recibe la asesoría. Mientras que existe otro grupo

que tiene como variables que los caracterizan el sexo (género), exámenes previos y estado civil.

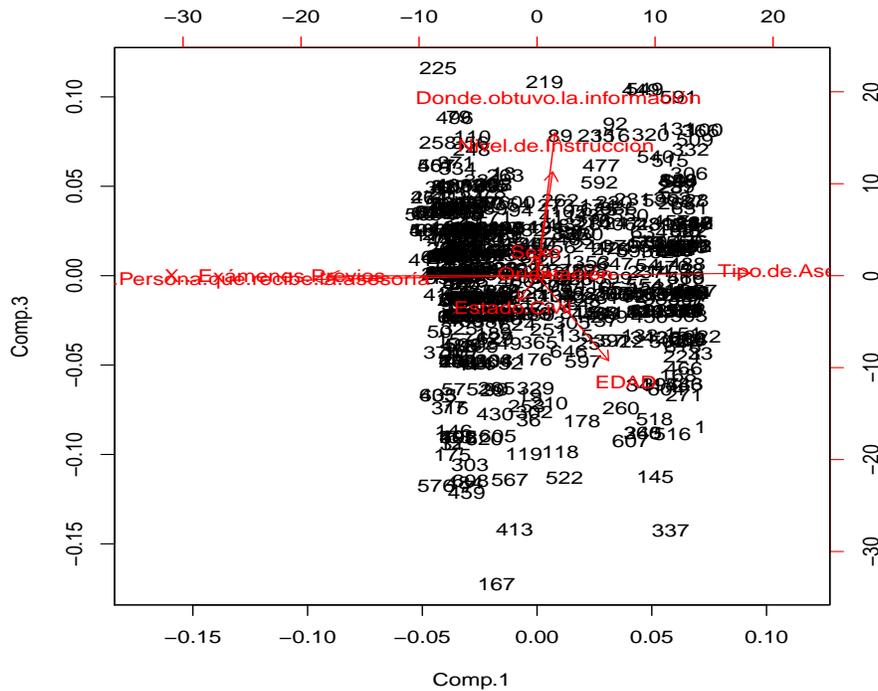


FIGURA 4.35. Gráfico de componente 1 vs. componente 3

Un grupo en el cuadrante superior izquierdo se destaca en la **figura 4.35**, teniendo como variables que lo caracterizan Persona que recibe la asesoría, Exámenes previos y Nivel de Instrucción. Adicionalmente observamos un grupo en el cuadrante inferior izquierdo caracterizado por las variables Persona que recibe la asesoría, Exámenes previos y Estado civil.

En la **figura 4.36** existen dos grupos fuertemente relacionados, que están caracterizados positivamente con las variables tipo de asesoría y orientación. Sin embargo la diferencia que hay entre estos grupos, es que un grupo está relacionado positivamente con las variables donde obtuvo la información y nivel de instrucción, tal como se observa en el cuadrante superior derecho, mientras que el segundo grupo tiene una relación negativa con estas variables (Cuadrante inferior derecho).

Para realizar este análisis primero utilizaremos el Test de esfericidad de Bartlett. La hipótesis nula es que se tiene un conjunto de variables no correlacionadas. Al encontrar el valor del estadístico de prueba EB frente al χ^2 en el programa **Matlab** se obtiene que es $EB = 1459,780 > 50,998 = \chi^2$, por lo cual se ha rechazado la hipótesis nula, de donde vemos que se tiene un conjunto de variables correlacionadas.

Por otra parte la Medida de adecuación muestral KMO ha sido creada como una función particular dentro del programa **Matlab**. En función del valor de KMO se determina la idoneidad del análisis factorial, tal como se mencionó en el Capítulo 3. Se obtuvo el siguiente resultado: "Kaiser-Meyer-Olkin Measure of Sampling Adequacy: 0.6595".

Es decir, este valor se encuentra en el siguiente rango $0,6 < KMO \leq 0,7$ que son valores regulares o mediocres. A pesar de que la prueba de esfericidad de Bartlett indique que se puede realizar el análisis factorial, el KMO arroja una confiabilidad poca idónea, ya que no proporciona ninguna información relevante, por tanto no se procede a realizar este análisis.

3. Conclusiones

Las asesorías registradas en los años 2008, 2009 y 2010 han brindado información sobre las características de los que acuden a la organización sin fines de lucro Acción Solidaria. Observamos a lo largo de estos años que se mantiene un patrón de comportamiento de los asistentes. Resaltaremos la característica predominante en cada variable. El tipo de asesoría por el cual la población acude es para la realización de Pre-test y Post-Test aproximadamente el 60 %, siendo en su mayoría Público General representada en un rango alrededor de 55 %.

El género masculino acude más a las asesorías aproximadamente el 80 % de los asistentes, es decir el 40 % más a la cantidad de mujeres registradas. La orientación que más resalta es la homosexual en todos los años, sin embargo en el año 2010 hubo un aumento porcentual del 15 % con respecto al año 2008. Es importante destacar que el nivel de instrucción de la mayoría de los asistentes es semiprofesional (82 %).

Las personas solteras reciben mayormente asesorías, pero cabe de destacar que el año 2010 el número de solteros y las personas en pareja manifiestan porcentajes cercanos de 48 % y 45 %, siendo el estado civil soltero(a) predominante. Acción Solidaria atiende público de todas las edades, sin embargo el registro predominante de las asesorías es de jóvenes con edades entre 26 y 45 años, representando aproximadamente el 62 % de las personas que acuden a las asesorías. Resalta el hecho de que la mayoría de los asistentes no poseen exámenes previos (entre 54 % y 60 %) y de donde obtuvieron información es en su mayoría mediante amigos y/o familiares representando el 45 % o más de las personas que recibieron asesorías.

Recordemos también que se realizaron en este trabajo histogramas múltiples, donde se refleja la información anterior.

Al realizar el análisis de componentes principales, encontramos varios puntos de interés, tales como fuertes correlaciones que se evidencian entre el tipo de asesoría, persona que recibe la asesoría y exámenes previos, tal como se mencionó previamente, además que el género y la orientación manifiestan grandes correlaciones (**cuadro 1**). Cuando se realizó la

importancia de las variables (**cuadro 4**), o mejor dicho la carga que cada una aporta, se observó que la variable donde obtuvo la información aporta un 0.209, es decir un valor alto con respecto a las otras variables. También consideramos tres componentes principales, con las cuales se explica un 58 % de la variabilidad de los datos según **cuadro 5**, y se realizó un gráfico (**figura 4.33**) donde se indica cual es el aporte de carga de cada característica en relación a cada componente.

En este exhaustivo análisis se realizó el gráfico de componentes por pares, en donde se observan los grupos formados por las diferentes características.

Finalmente de forma práctica podemos decir que hay un gran número de personas que no se han realizado algún examen previo, y se debe enfatizar en los mecanismos de difusión de la información, ya que en su mayoría son recomendados por amigos y/o familiares, pero a su vez ha tenido gran auge la web esto puede ser de gran utilidad en una era tecnológica como la que se vive actualmente, este medio debe ser explotado al máximo a fin de llegar hasta personas de niveles de instrucción tales como profesionales, trabajadores y obreros.

Apéndice A

A continuación los comandos utilizados en Matlab;

Histogramas de cada característica-2008.

```
figure
A=data(:,3);
hist(A)
title('Tipo de Asesoría 2008')
ylabel('Número de personas');
set(gca,'xtick',1:3,'xticklabel','Pre-test y Post-test','Información General','Información PVVIH/SIDA')
```

```
figure
B=data(:,4);
hist(B)
title('Persona que recibe la asesoría 2008')
set(gca,'xtick',1:5,'xticklabel','P.V.VIH/SIDA','Familiar','Pareja','Compañero','Público en general')
ylabel('Número de personas');
```

```
figure
C=data(:,5);
hist(C)
title('Género 2008')
set(gca,'xtick',1:2,'xticklabel','Masculino','Femenino')
ylabel('Número de personas');
```

```
figure
D=data(:,6);
hist(D)
title('Orientación 2008')
set(gca,'xtick',1:3,'xticklabel','Heterosexual','Homosexual','Bisexual')
ylabel('Número de personas');
```

```
figure
E=data(:,7);
hist(E)
title('Nivel de Instrucción 2008')
```

```
set(gca,'xtick',1:3,'xticklabel','Profesional','Semiprofesional','Trabajador')
ylabel('Número de personas');
```

```
figure
F=data(:,8);
hist(F)
title('Estado Civil 2008')
set(gca,'xtick',1:5,'xticklabel','Soltero(a)','Casado(a)','Divorciado(a)','En Pareja','Viudo(a)')
ylabel('Número de personas');
```

```
figure
G=data(:,9);
hist(G)
title('EDAD 2008')
set(gca,'xtick',1:6,'xticklabel','Menor de 11 años','12-18 años','19-25 años','26-45 años','46-65 años','Mayor a 66 años')
ylabel('Número de personas');
```

```
figure
H=data(:,10);
hist(H)
title('Exámenes Previos 2008')
set(gca,'xtick',1:3,'xticklabel','Todos','Algunos','Ninguno')
ylabel('Número de personas');
```

```
figure
I=data(:,11);
hist(I)
title('Donde obtuvo la información 2008')
set(gca,'xtick',1:10,'xticklabel','Medios','Radio','Cine','Marc.','Prensa','Tríp.','Amigos','ONG's','CINVIH','Otros')
ylabel('Número de personas');
```

Histogramas Múltiples

```
X=data(:,[ 3 4 5 6 7 8 9 10 11]);
figure
hist(X)
title('Histograma múltiple de 2008')
ylabel('Número de personas');
legend('Tipo de Asesoría','Persona que recibe la asesoría','Género','Orientación','Nivel de Instrucción','Edo Civil','EDAD','Exámenes Previos','Donde obtuvo la información')
```

Pca sin NAN

```
P=data(2:648,3:11);
S=cov(P);
R=corr(P);
medias=mean(P);
desviaciones=std(P);
```

```
[coef,score,varianzas]=princomp(P);
```

```
figure
plot(varianzas,'*')
xlabel('Variable');
ylabel('Varianzas');
title('Grafico de Varianzas 2008');
grid on
```

```
figure
autovalores=sort(eig(R),'descend');
plot(autovalores,'*')
xlabel('Numero de Autovalor');
ylabel('Autovalor');
title('Grafico de Autovalores 2008');
grid on
```

Comandos utilizados para el FA.

```
P=data(2:648,3:11);
S=cov(P);
R=corr(P);
medias=mean(P);

[N,p]=size(P);
EB=-((N-1-(1/6)*(2*p+5))*log(abs(det(R))));
chi2=chi2inv(0.95,p*(p-1)/2);
KMO(P);

autovalores=sort(eig(R));
T=trace(S);
Impj=(autovalores)/T;
```

Comandos en R para el análisis de componentes principales

```
library(Rcmdr)
data(Datos)
```

```
cov(Datos[,3:11])
pca.datosj-princomp(Datos[,3:11],cor=T)
names(pca.datos)
options(digits=3)
summary(pca.datos)
loadings(pca.datos)
par(mfrow=c(3,1))
barplot(loadings(pca.datos)[,1],sub="Primera componente")
barplot(loadings(pca.datos)[,2],sub="Segunda componente")
barplot(loadings(pca.datos)[,3],sub="Tercera componente")

par(mfrow=c(3,1))
biplot(pca.datos, choices = c(1,2))
biplot(pca.datos, choices = c(1,3))
biplot(pca.datos, choices = c(2,3))
```

Nota: Los comandos utilizados para los años 2009 y 2010 se proceden de forma análoga.

Bibliografía

- [1] Data de las asesorías de los años 2008, 2009 y 2010 de la organización sin fines de lucro Acción Solidaria.
- [2] TAPIA, JESÚS M, (2007.) *Introducción al Análisis de Datos Multivariantes*. UNELLEZ. VENEZUELA. PP. 1-20.
- [3] DALLAS, JOHNSONN, (2004) *Métodos Multivariados Aplicados al Análisis de Datos*. INTERNATIONAL THOMSON EDITORES, S.A DE C.V. MÉXICO. PP.1-155.
- [4] PROGRAMAS COMPUTACIONALES MATEMÁTICOS **Matlab** Y **R**.