



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

**Análisis multivariado para el
mejoramiento de los programas y
servicios del departamento de
Comunicación Educativa del Museo
Nacional de Antropología**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I A

P R E S E N T A

BRENDA ELIZABETH RINCON ZARZA

Tutor

M. en F. Jorge Luis Reyes García

Ciudad Universitaria, CD. MX. 2023





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno

Rincon
Zarza
Brenda Elizabeth
5555700974
Universidad Nacional Autónoma de México
México
Facultad de Ciencias
Actuaría
402088424

2. Datos del Tutor

M en F
Jorge Luis
Reyes
García

3. Datos del sinodal 1

Dra
Ruth
Selene
Fuentes
García

4. Datos del sinodal 2

Dra
Lizbeth
Naranjo
Albarrán

5. Datos del sinodal 3

M en C
Jaime
Vázquez
Alamilla

6. Datos del sinodal 4

M en C
José Salvador
Zamora
Muñoz

7. Datos del trabajo escrito

Análisis multivariado para el mejoramiento de los programas y servicios del departamento de Comunicación Educativa del Museo Nacional de Antropología
110 p
2023

Agradecimientos

Gracias a la UNAM que me dio amigos, profesores excelentes, educación, desarrollo intelectual y las bases para lograr el éxito profesional.

Gracias a mi tutor Jorge, que siempre ha tenido tiempo y disposición para que pudiera terminar este trabajo, gracias por haber sido un excelente profesor.

Gracias a cada uno de mis sinodales Dra. Ruth Fuentes, Dra. Lizbeth Naranjo, Maestro Jaime Vázquez y Maestro Salvador Zamora, por sus correcciones precisas y tiempo.

Gracias a mi mamá, mi hermano, Alejandra, a mis tías Bertha†, Chela† y a mi papá† que siempre me han brindado su cariño, paciencia y apoyo para cumplir mis objetivos.

Gracias a mis amigos y a todas las personas que me han acompañado en esta etapa de mi vida.

Índice

CAPÍTULO 1. MUSEO NACIONAL DE ANTROPOLOGÍA	1
1.1 HISTORIA DEL MUSEO NACIONAL DE ANTROPOLOGÍA (MNA)	1
1.1.2 RESTRUCTURACIÓN DEL MUSEO NACIONAL	2
1.1.3 ACTUALIDAD DEL MUSEO NACIONAL DE ANTROPOLOGÍA	3
1.4 MISIÓN DEL MUSEO NACIONAL DE ANTROPOLOGÍA	6
1.5 FUNCIÓN DEL MUSEO NACIONAL DE ANTROPOLOGÍA	6
1.7 ORGANIGRAMA DEL MUSEO NACIONAL DE ANTROPOLOGÍA	6
1.8 ÁREAS EDUCATIVAS DEL MUSEO NACIONAL DE ANTROPOLOGÍA	7
1.9 DEPARTAMENTO DE COMUNICACIÓN EDUCATIVA	8
1.10 ESTUDIOS DE PÚBLICO	13
CAPÍTULO 2. MÉTODOS MULTIVARIADOS	18
2.1 COMPONENTES PRINCIPALES	18
2.2 ANÁLISIS FACTORIAL	26
2.3 ANÁLISIS DE CONGLOMERADOS	33
CAPÍTULO 3 ANÁLISIS ESTADÍSTICO Y MULTIVARIADO	53
3.1 INTRODUCCIÓN	53
3.1.2 ENCUESTA Y BASE DE DATOS	53
3.2 ANÁLISIS DESCRIPTIVO DE VARIABLES SOCIODEMOGRÁFICAS	54
3.3 ANÁLISIS ESTADÍSTICOS DE CORRELACIONES	62
3.4 ANÁLISIS POR MÉTODOS MULTIVARIADOS	70
3.4.1 ANÁLISIS FACTORIAL	70
3.4.2 ANÁLISIS DE COMPONENTES PRINCIPALES	72
3.4.3 ANÁLISIS DE CONGLOMERADOS O CLÚSTERES	76
3.5 RESULTADOS	81
3.6 RESULTADOS DE LOS MÉTODOS DE ANÁLISIS MULTIVARIADOS	84
3.7 INSTRUMENTO PROPUESTO	87
CAPÍTULO 4. CONCLUSIONES	91
4.1 PARTE TÉCNICA	91
4.2 ANÁLISIS FACTORIAL	91
4.3 COMPONENTES PRINCIPALES	91
4.4 ANÁLISIS DE CLÚSTERES O CONGLOMERADOS	91
4.5 CONCLUSIÓN GENERAL	92
BIBLIOGRAFÍA	93
ANEXO I	94
ANEXO II	107

Índice de Ilustraciones

Ilustración 1.1.....	2	Figura 3.11.....	61
Ilustración 1.2.....	3	Figura 3.12.....	61
Ilustración 1.3.....	16	Figura 3.13.....	62
Tabla 2.1.....	25	Tabla 3.1.....	70
Ilustración 2.2.....	26	Figura 3.14.....	71
Tabla 2.2.....	41	Tabla 3.2.....	71
Figura 3.1.....	55	Figura 3.15.....	72
Figura 3.2.....	55	Figura 3.16.....	73
Figura 3.3.....	56	Figura 3.17.....	74
Figura 3.4.....	57	Figura 3.18.....	75
Figura 3.5.....	57	Figura 3.19.....	76
Figura 3.6.....	58	Figura 3.20.....	76
Figura 3.7.....	58	Figura 3.21.....	77
Figura 3.8.....	59	Figura 3.22.....	79
Figura 3.9.....	59	Figura 3.23.....	83
Figura 3.10.....	60	Tabla 3.3.....	84
		Tabla 3.4.....	86

Introducción

Justificación e importancia del tema

El trabajo tuvo como finalidad realizar un análisis multivariado a partir de una base de datos que se generó a través de información obtenida por encuestas realizadas en el departamento de Comunicación Educativa del Museo Nacional de Antropología, con el objetivo de determinar patrones sobre el público visitante al área y tomar decisiones para mejorar los programas y servicios que genera y desarrolla este departamento.

Es importante decir que el propósito de la jefatura del Departamento para la realización de la encuesta fue el de conocer al público participante en los programas Talleres Sabatinos y Noche de Museos, que muestra una asistencia distintita a años anteriores y empiezan a generar nuevas dinámicas en los servicios y horarios de atención del área.

Objetivo.

Conocer y analizar los perfiles sociodemográficos y expectativas de los públicos que se acercan al departamento de Comunicación Educativa y la relación que guarda con éste a través de las actividades en que participan, con la aplicación de los métodos multivariados: Análisis factorial, Componentes principales y Análisis de Conglomerados.

Descripción del trabajo.

En el primer capítulo se aborda con brevedad la historia del Museo Nacional de Antropología, desde sus inicios en la Universidad Pontificia en el siglo XIX, hasta la construcción de su actual recinto en 1964, bajo un México independiente y con el firme propósito de conservar y estudiar el patrimonio heredado de las culturas originarias. Así también se describe la historia y organización de las áreas de atención al público, integradas al Museo, la importancia de éstas, sus funciones, cuál es la finalidad de los programas y servicios actuales con los que cuenta. Por último, se explican la necesidad de conocer al público participante en los programas Talleres Sabatinos y Noche de Museos a través del planteamiento de los estudios de público.

El segundo capítulo explica brevemente la historia de los principales métodos estadísticos multivariados y se realiza el marco teórico de los modelos de Componentes principales, Análisis factorial y Análisis de Conglomerados, los parámetros que se deben tomar en cuenta para su aplicación, los resultados gráficos que se generan con estos métodos y su interpretación.

En el tercer capítulo se presentan la construcción y limpieza de la base de datos, obtenida de las encuestas aplicadas en las actividades de Talleres Sabatinos y Noche de Museos llevados a cabo por el departamento, se muestran los resultados obtenidos del análisis descriptivo aplicado a la información y los obtenidos a través de los métodos de análisis multivariado (Análisis factorial, Componentes principales, Análisis de Conglomerados), por

último se propone un instrumento que permita una mejor captación de información para poder ser analizada de una manera más eficiente. Cabe destacar que en este capítulo la información correspondiente a los resultados de correlación y del análisis multivariado no se presenta de manera técnica, sino de manera accesible para quienes no estén relacionados con el lenguaje técnico matemático, los resultados numéricos, matrices de correlación, eigenvectores, eigenvalores, comunalidades y ecuaciones lineales se presenta y explican en el Anexo I.

Finalmente, en el último capítulo se abordan las conclusiones de la aplicación de los métodos estadísticos, se dan las ventajas y desventajas de esta aplicación, así del por qué estas técnicas pueden estar apoyadas por otras disciplinas que ayuden a enriquecer la comprensión de los datos numéricos para el conocimiento de los públicos.

Capítulo 1. Museo Nacional de Antropología

1.1 Historia del Museo Nacional de Antropología (MNA)

En el curso de los años el hombre ha plasmado su huella en distintos lugares del mundo dejando, como herencia a futuras generaciones: ciudades, templos, esculturas, pinturas, artefactos y un sin número de evidencias. Muchas están en museos, sitios por excelencia de resguardo y transmisión del conocimiento de esos objetos convirtiéndose, desde esa perspectiva, en lugares donde el pasado y el presente se reúnen construyendo parte del patrimonio histórico universal.

El actual Museo Nacional de Antropología, institución que expresa el cambiante medio social, económico y político del país inició sus actividades desde 1964, y no siempre ha estado ubicado en el recinto que actualmente se encuentra en el Bosque de Chapultepec. Antes hay una larga historia. Durante los tres siglos de la corona española, el pasado indígena mexicano y sus restos materiales fueron estudiados y saqueados por nativos y extranjeros. En el momento de la conquista, la preocupación indígena por preservar parte de su herencia llevó al ocultamiento de códices y objetos, sin embargo, el afán conservador se vio sobrepasado por la virtual depredación y destrucción de las ciudades indígenas, algunos cuyos tesoros fueron enviados a Europa y otros, los más, destruidos en su propia tierra.

La nación Mexicana, apenas iniciada su independencia, vio nacer el primer museo nacional nombrado Museo Nacional de México el 18 de marzo de 1825, bajo la presidencia de Guadalupe Victoria. Sin embargo hasta el 21 de noviembre de 1831, es cuando el Congreso instituye legalmente al Museo Nacional, compuesto por los ramos de Antigüedades, Productos de la Industria e Historia Nacional y Jardín Botánico. La sede inicial del establecimiento fue dos de las aulas de la Nacional y Pontificia Universidad de México, estas aulas espaciosas estaban llenas de curiosidades de toda descripción pertenecientes a la historia, manualidades, religión y costumbres de la cultura Azteca, una colección de retratos de mexicanos y españoles eminentes. En relación a la naturaleza se mostraban diversos especímenes endémicos de la flora y fauna así como representaciones exactas de minas y de la manera de cavarlas, también se mostraban trabajos en oro y plata; el acervo estuvo expuesto al público en dicho edificio hasta 1866.

El Museo surgió como parte de un espacio público urbano a manera de una institución cultural moderna, que acogió la práctica de la historia natural, la arqueología y la historia.

Entre 1833 y 1854 el Museo Nacional acompañó a la Universidad Pontificia en sus vicisitudes, que culminaron en suspensión en 1857; el edificio, los libros de su biblioteca, sus fondos y sus bienes serían destinados a la creación de la Biblioteca Nacional. Más tarde la Universidad fue restituida y suspendida varias veces hasta que el archiduque Maximiliano de Habsburgo la suprimió definitivamente en 1865.



Ilustración 1.1 Coatlicue y la Piedra de Tizoc en el patio del antiguo Museo Nacional. Fotografía: Archivo Histórico del Museo Nacional de Antropología – CONACULTA-INAH-CANON

1.1.2 Restructuración del Museo Nacional

Al Momento que Maximiliano de Habsburgo declara el cierre de la Nacional Pontificia Universidad de México en 1865, en la que se albergaba el Museo Nacional, también instruye el cambio de sede de las colecciones del museo.

El Museo Público de Historia Natural, Arqueología e Historia fue creado por decreto el 4 de diciembre de 1865. Se estableció en la antigua Casa de Moneda, que era un edificio virreinal, ubicado en la esquina suroeste del Palacio Nacional, y fue inaugurado el 6 de julio de 1866.

Cuando Maximiliano ordenó trasladar el museo a la vieja casa de la calle de La Moneda, lo hizo con el claro objetivo político de su campaña en favor de la dignificación de los pueblos indígenas que habitaban en el territorio nacional.

Tras la caída del emperador Habsburgo, el museo continuó sus actividades; sin embargo, no fue sino hasta el advenimiento del general Porfirio Díaz a la presidencia cuando se inició el proceso de consolidación y fortalecimiento del museo moderno. En 1877, quedó dividido en tres departamentos: Historia Natural, Arqueología, e Historia; se hicieron importantes obras de ornato, alumbrado, estantería y laboratorio. En 1887, se crearon las secciones de Antropología y de Etnografía, el general Díaz inauguró la Galería Monolitos.

En suma, en el panorama cultural de finales del siglo XIX y principios del XX, el Museo Nacional jugó un papel importante; fue el centro de formación y discusión intelectual del país.

Para principios del siglo XX, el edificio que lo albergaba ya no era suficiente. Por acuerdo del Ejecutivo el 28 de enero de 1909, se independizaron las colecciones naturalistas e históricas para formar los museos Nacional de Historia Natural, por un lado y Nacional de Arqueología, Historia y Etnografía, por el otro lado. El primero salió de la Casa de Moneda y el segundo fue preparado para la celebración del primer Centenario de la Independencia. El Museo reabrió, ya reorganizado y con una nueva designación, el 28 de agosto de 1910, sin fiesta ni acto especial.

Entre 1911 y 1913, el movimiento revolucionario no pareció afectar el funcionamiento del museo; por el contrario éste consolidó su organización.

Con la creación de la Secretaría de Educación Pública, en 1921, se estableció con mayor claridad la función de los museos como apoyo al sistema educativo federal y como espacios culturales a través de los cuales se difundiría la ideología del nacionalismo revolucionario producto del movimiento popular de 1910 a 1917. Esta ideología hacía énfasis en el rescate

del pasado indígena, elemento que se consideró esencial para la formación de una conciencia histórica nacional.

Su nombre de Museo Nacional de Arqueología, Historia y Etnografía lo conservaría hasta 1939, cuando las colecciones de historia pasaron al Castillo de Chapultepec. Desde esa fecha disminuyó su importancia como centro de investigación, pues fue creado el Instituto Nacional de Antropología e Historia del cual pasó a formar parte.



Ilustración 1.2 Cour du Musée National. El Palacio de Moneda, en otros tiempos sede del Antiguo Museo Nacional. Óleo sobre tela. Autor: Cleofás Almanza, tomado página web del MNA, archivo histórico.

1.1.3 Actualidad del Museo Nacional de Antropología

De 1944 a 1964, el viejo Palacio de la Moneda llevó como nombre el Museo Nacional de Antropología. Desde esa época ya no incluía la colección de historia, por lo que era incorrecto llamarlo Museo Nacional de Antropología e Historia.

El anterior museo fue creciendo en la medida que los antropólogos mexicanos iban adquiriendo, mediante estudios y excavaciones, nuevas piezas que saturaron las salas y bodegas. Por ello surge la necesidad de crear un nuevo museo, un espacio exclusivo para presentar el México Prehispánico y las culturas indígenas actuales.

El lunes 20 de agosto de 1939 inició sesiones el XXXV Congreso Internacional de Americanistas. La asociación se había ya reunido en México en dos ocasiones anteriores: primero, en 1895 y en 1910.

En la sesión de 1910 Justo Sierra declaró: “Para obtener el título que ambicionamos de Capital Arqueológica del Continente Americano, nos obligamos a cuidar celosamente el tesoro que los siglos nos han donado”¹.

Con este antecedente el Secretario de Educación Pública Jaime Torres Bodet propone al actual presidente de la República Adolfo López Mateos dos proyectos: La construcción de una gran biblioteca y la creación de un verdadero Museo de Antropología. Entre los proyectos le recomendó que escogiera uno y optó por el del museo. El Presidente acogió la idea con beneplácito. Pero antes de tomar la decisión, quiso saber en qué terreno podría hacerse la obra, cuáles serían los méritos del proyecto y cuánto costaría la construcción. Para poder contestar esto Torres Bodet convocó y consultó al arquitecto Pedro Ramírez Vázquez, realizador de la Galería de Historia de Chapultepec quien posteriormente llevó a cabo el proyecto.

¹ Ramírez Vázquez, P. *Museo Nacional de Antropología*, México, 2012, INAH.

En esa época muchas personas acudían regularmente al Bosque de Chapultepec registrando una asistencia dominical de 250 mil personas² (en el 2019 el MNA recibió a un total de 3, 086, 555 visitantes)³ Parecía una buena idea que ese lugar albergara al nuevo museo ya que era el lugar al que por lo común acudía la población.

El terreno que se seleccionó para el museo se ubica en la esquina del Paseo de la Reforma y la Calzada de la Milla ocupado por la Secretaría de Comunicaciones y Transportes. Era un área destinada a la recreación de sus empleados, la zona no estaba arbolada, por lo que no se provocaba ninguna deforestación. El presidente Adolfo López Mateos y el regente Uruchurtu accedieron a ceder este espacio.

El sitio elegido para el nuevo museo, dada su importancia, obedeció a la siguiente reflexión: se requería una gran área para su edificación, con accesos urbanos claros, y un área amplia para estacionamiento que asegurara la asistencia del público.

El Museo fue concebido como un espacio donde el visitante pudiera recibir el mensaje proveniente de las propias obras, reconocidas creaciones artísticas de los pueblos indígenas mexicanos del pasado y el presente. Así también el proyecto originario articulaba la imperativa tarea gubernamental de la educación pública con el conocimiento arqueológico e historiográfico del momento.

Este gran proyecto reunió a un numeroso y calificado grupo de arqueólogos, etnólogos, pedagogos, historiadores, antropólogos físicos, lingüistas, arquitectos, museógrafos que junto con cientos de trabajadores construyeron lo que hoy es el actual Museo Nacional de Antropología.

Además diversos artistas colaboraron al enriquecimiento del discurso museográfico mediante sus murales y óleos. Así el Museo integró en sus colecciones obras de Rufino Tamayo, Nicolás Moreno, Leonora Carrington, Pablo O'Higgins, Rafael Coronel, Raúl Anguiano, Carlos Mérida, Alfredo Zalce, Luis Covarrubias y Jorge González Camarena, entre otros.

La construcción de esta obra monumental se inició en febrero de 1963, y tuvo una duración de diecinueve meses: seis meses de construcción y trece meses en áreas exteriores y en la instalación de las colecciones.

En la etapa de construcción y montaje de las diferentes colecciones que albergarían las veinticinco salas se presentó el problema de poder realizar una museografía que tenía como propósito el atraer la atención de niños, del pueblo o de turistas, a través de un espectáculo que evitase el agobio y el cansancio que es normal en los grandes museos. Por lo que se consideró necesario realizar un cuestionario⁴ para saber qué servicios educativos, de seguridad y de atención al público tenía cada museo, cómo se mostraba la cerámica, cómo se exhibían los documentos, los monolitos y joyas, cómo se conservaba y exhibían los códices y otros documentos.

El proyecto quedó concluido en tiempo y forma para ser inaugurado el 17 de septiembre de 1964 a las nueve horas por el entonces presidente de la República Adolfo López Mateos. Su concepto arquitectónico y diseño museográfico se adelantó a su tiempo por lo cual el Museo Nacional de Antropología es el edificio más emblemático de la arquitectura moderna

² Ramírez Vázquez, P. *Museo Nacional de Antropología*, México, 2012, INAH.

³ *Instituto Nacional de Antropología*, 5 de abril de 2020. https://mna.inah.gob.mx/el_museo.php#la_institucion

⁴ Ramírez Vázquez, P. *Museo Nacional de Antropología*, México, 2012, INAH.

mexicana de la segunda mitad del siglo XX. Ha sido galardonado internacionalmente en numerosas ocasiones, y es reconocido como el más importante de México y de América Latina.

Existen múltiples anécdotas sobre la creación del Museo Nacional de Antropología que, más allá del marco institucional que lo creó y de la historia oficial que lo rodea, nos habla de las expectativas que generó y de su extraordinaria convocatoria pública a lo largo de cincuenta y ocho años.

Anécdota

Traslado de Tláloc a la Ciudad de México.

En las inmediaciones de Coatlinchan cerca de Chapingo, se encontraba un monolito de origen teotihuacano, que resultaba ser una pieza extraordinaria y de enormes proporciones, pues tenía un peso de más de 160 toneladas.

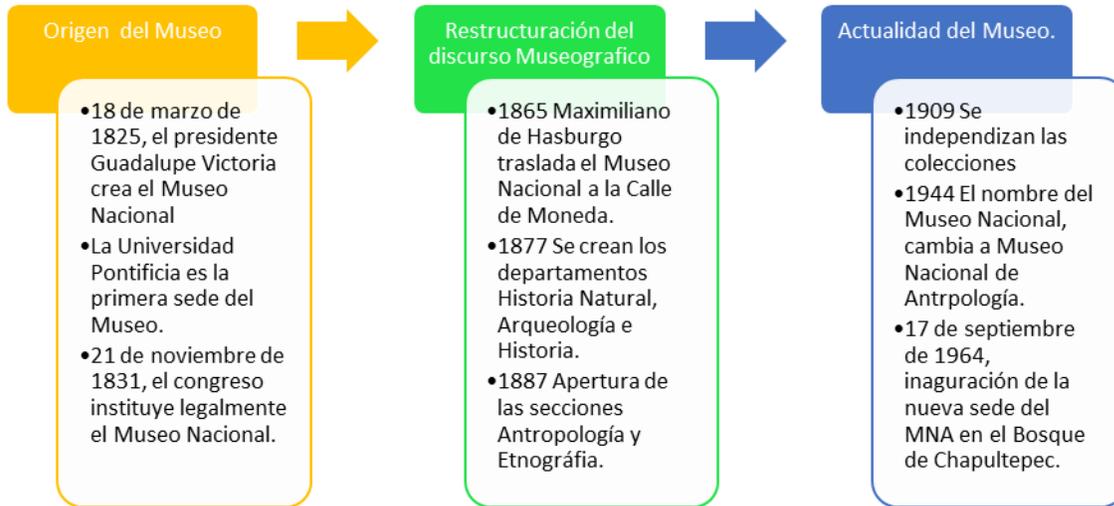
A sugerencia del Presidente López Mateos resultaba la pieza adecuada para ser colocada en la entrada del Museo. Traer de Coatlinchan esta obra no resultaba sencillo; ya que primero era necesario convencer a los habitantes de dicha región para que autorizaran el traslado del monolito a la Ciudad de México; Segundo era resolver el traslado de tan enorme y pesada pieza para que no fuera a ser dañada durante el traslado.

Ambos puntos se resolvieron con facilidad, por lo que sin más inconvenientes se logró el traslado de Tláloc. El suceso reunió a un gran número de capitalinos quienes emocionados esperaron la llegada del Dios de la Lluvia, quien entró al zócalo de la ciudad en medio de un inusual aguacero y fuertes aplausos, el día 16 de abril de 1964⁵.

El MNA es una institución con una historia muy interesante y amplia que comienza al tiempo en que México se convierte en una nación independiente, a lo largo de las décadas se transforma y va adquiriendo un objetivo más específico, esto hace que tenga que emigrar de un recinto a otro. Su importancia y desarrollo en la preservación del patrimonio crean la necesidad de construir un recinto permanente que enaltezca la grandeza cultural de los pueblos originarios del pasado y presente. En resumen su historia puede dividirse en tres etapas a lo largo del tiempo.

⁵ Ramírez Vázquez, P. *Museo Nacional de Antropología*, México, 2012, INAH, p. 66 – 70.

Línea de Tiempo



Línea de Tiempo. Fuente: Creación Propia

1.4 Misión del Museo Nacional de Antropología

El MNA es un legado para las generaciones actuales y futuras, su historia muestra un pensamiento diferente de la sociedad mexicana y el gobierno; valora el pasado, busca conocerlo y cuidarlo. Esta institución que por más de cinco décadas, ha cumplido con la misión de investigar, conservar, exhibir y difundir las colecciones arqueológicas y etnográficas más importantes del país⁶.

1.5 Función del Museo Nacional de Antropología

Las principales funciones del museo son las siguientes: Difusión de la cultura prehispánica y la de los pueblos originarios actuales, por medio de la exposición de piezas del acervo arqueológico y etnográfico, conservación, registro y restauración de las colecciones arqueológicas y etnográficas de México, investigación, publicación y difusión de los diferentes estudios que se llevan a cabo en las colecciones arqueológicas y etnográficas⁷.

1.7 Organigrama del Museo Nacional de Antropología

La estructura del MNA se conforma por una dirección, seis subdirecciones que corresponden a Arqueología, Etnografía, Museografía, Protección de Bienes Culturales. La subdirección Administrativa encargada de los departamentos de Recursos Financieros, Recursos Humanos y Servicios Generales y la Subdirección Técnica que tiene a los Departamentos de Informática, Promoción Cultural y el Departamento de Comunicación Educativa.

⁶ Museo Nacional de Antropología, 5 de abril de 2020, https://mna.inah.gob.mx/el_museo.php#la_institucion

⁷ *Ibíd.*



"Organigrama del Museo Nacional de Antropología" Fuente: Instituto Nacional de Antropología e Historia

1.8 Áreas educativas del Museo Nacional de Antropología.

El Museo Nacional de Antropología se concibió con una clara función educativa, un lugar en donde propios y extraños puedan conocer, admirar, pero sobre todo comprender el enorme legado que como herencia nos dejaron las culturas prehispánicas y del cual aún nos quedan muchos vestigios, y los testimonios culturales de los grupos indígenas actuales.

Su intención didáctica se ve claramente plasmada en sus salas: por eso relaciona los objetos de cada cultura con sus manifestaciones; el medio físico, la vida diaria, las ciudades, la tecnología, la economía, las condiciones sociales y el momento histórico en que las investigaciones la sitúan. Mapas, pinturas, murales, maquetas y dioramas que contextualizan museográficamente el desarrollo de las culturas exhibidas.

A decir de Mario Vázquez el museo "No es un museo de arte, va más allá"⁸ (y esto se comprende claramente en el primer manuscrito sobre el "Programa de funcionamiento del Museo Nacional de Antropología", el cual a la letra dice:

"La función del función del MNA es la de conservar, preservar y dar a conocer en forma didáctica y de acuerdo con el método científico, los materiales que estudia la antropología y su interpretación [...] Como museo de ciencia que es, la presentación de sus materiales debe estar normada fundamentalmente por un mensaje de gran contenido educativo y de positivo valor social. Para que la institución cumpla esta misión, es necesario que su planteamiento permita el desarrollo de toda una serie de actividades complementarias a las funciones esenciales del museo: labores de investigación, servicios al público, tareas específicas de tipo educativo y atención completa a las necesidades de visitantes y personal"⁹.

Con la apertura del Museo Nacional de Antropología la atención al público estuvo dirigida por la Jefatura de Difusión y Relaciones Públicas, quien coordinó los servicios educativos

⁸ Vázquez Mario, *Obra museológica y museográfica*, 2015.

⁹ Vallejo, M. E. *Los servicios educativos del INAH*, México, 2003, INAH.

que se encargaban de la atención a grupos escolares, desde jardín de niños hasta el nivel de secundaria.

Además el área de Difusión tenía a su cargo la selección de guías, que cubrían por una las visitas de nivel bachillerato, así como la de grupos especiales de adultos mexicanos (obreros, trabajadores del Seguro Social, sindicatos, entre otros) y, por otra, al turismo extranjero, con visitas programadas diariamente en varios idiomas. En el año de 1970 estos departamentos se separaron creando sus propias jefaturas.

Actualmente las áreas de atención al público del Museo son dos: Departamento de Promoción Cultural y Departamento de Comunicación Educativa, el primero se encarga de las visitas guiadas a grupos escolares de nivel medio superior, visitas a grupos especiales y extranjeros, y de programar y desarrollar cursos y otros eventos culturales y académicos que imparte personal del departamento de investigadores del Museo y de otras instituciones. El segundo se encarga de atender preferentemente a los niveles de educación básica, preescolar, primaria y secundaria sin desatender la atención a los públicos en general. Siendo el objetivo principal de ambas áreas la difusión y comunicación del patrimonio cultural arqueológico y etnográfico que salvaguarda el museo.

1.9 Departamento de Comunicación Educativa

Antecedentes

El Museo Nacional de Antropología fue el primer Museo en el país en contar con maestros y pedagogos dentro del gran equipo multidisciplinario que planificó su función educativa. Como parte de su estructura, uno de los departamentos que se consideró fundamental desde un principio para apoyar la educación de los niños y jóvenes fue el Departamento de Comunicación Educativa, originalmente nombrado Servicios Educativos.

Los servicios educativos que brindan los museos en México se iniciaron en el Museo Nacional de Historia en el año de 1962 y pronto comenzaron a funcionar también en el viejo Museo Nacional de Antropología, su labor la comenzaron con visitas guiadas principalmente. Es así como en 1963 cuando el proyecto para el nuevo Museo Nacional de Antropología cobra relevancia la función educativa en los museos, baste conocer los espacios arquitectónicos que se construyeron para el desarrollo de esta función.

En vista de mejorar y ampliar los servicios, el equipo de maestros y pedagogos planteó la necesidad de contar con los siguientes espacios y equipamiento: auditorio, salas de usos múltiples, mesas adecuadas para talleres, teatro al aire libre, biblioteca, oficinas y espacios para la recepción de los grupos escolares, instalaciones con las que cuenta actualmente el área. La decoración incluyó dos murales pintados ex profeso; “La ronda en el tiempo” de Fanny Rabel y “La educación del niño en la época mexicana” de Regina Raull.

Las primeras visitas en el MNA, iniciaron en 1964 a cargo de maestras normalistas con especialidad en historia.¹⁰ Así como el desarrollo de múltiples proyectos que el departamento de Comunicación Educativa realizó desde sus comienzos en el año de 1964:

Las funciones de teatro guiñol, en 1965 – 1966, con la obra “El bautismo en los tiempos de los Mexicas” producida, elaborada y presentada por el equipo de servicios educativos en funciones sabatinas y dominicales.

El cineclub infantil en 1968, con la colaboración de la cineteca del INAH.

¹⁰ Ramírez Vázquez, P. *Museo Nacional de Antropología*, México, 2012, INAH, p. 47.

El programa de atención a discapacitados, que inicio en 1975 con un grupo de niños ciegos, que se extendió a niños hipo acústicos.

La intervención por primera vez de artesanos mexicanos, enseñando su arte en los talleres de verano¹¹.

Los servicios educativos del Museo Nacional de Antropología marcaron un hito histórico, ya que sus actividades influyeron en otros museos ampliando la dimensión pedagógica.

A decir de Vallejo, los años de 1999 a 2001 el concepto de funcionamiento de las áreas educativas tuvo un cambio al considerar como prioridades, además de las visitas guiadas tradicionales, los cursos impartidos por maestros y otros apoyos didácticos como: videos, películas, audiovisuales, materiales de afirmación y evaluación, material de divulgación y talleres post - visita, sabatinos, de verano, así como otros proyectos educativos que se realizaron con otras instituciones¹². Actualmente el departamento de Comunicación Educativa presta muchos servicios, cuenta con múltiples apoyos para la realización de visitas escolares y para público en general, imparte talleres vacacionales y sabatinos, tiene a la venta material de divulgación, en el que participaron los curadores de las salas, con el que se apoya a comprender mejor las colecciones de las salas y material de apoyo didáctico.

Objetivo del Departamento de Comunicación Educativa

Promover y difundir estrategias de servicios y comunicación educativa para entablar diálogos entre los diferentes públicos y difundir los acervos patrimoniales que custodia el museo, con la finalidad de propiciar el cuestionamiento, valoración y disfrute de la visita al museo.

Organigrama del Departamento de Comunicación Educativa



"Organigrama del Departamento de comunicación Educativa" Fuente: Creación Propia

El departamento de Comunicación Educativa se integra por una Jefatura encargada de la planeación de proyectos, gestión de los recursos materiales y financieros, coordinar al personal del departamento y prestadores de servicio social.

Las actividades de los gestores culturales y asesores educativos son sustantivas dentro del área ya que se basan en la atención al público visitante, impartición de visitas guiadas, talleres y participación en la planeación y desarrollo de proyectos especiales tendientes a dar respuesta a las necesidades de los públicos que visitan el museo.

¹¹ Vallejo, M. E, Los servicios educativos del INAH, México, 2003, INAH.

¹² *Ibíd.*, p. 84.

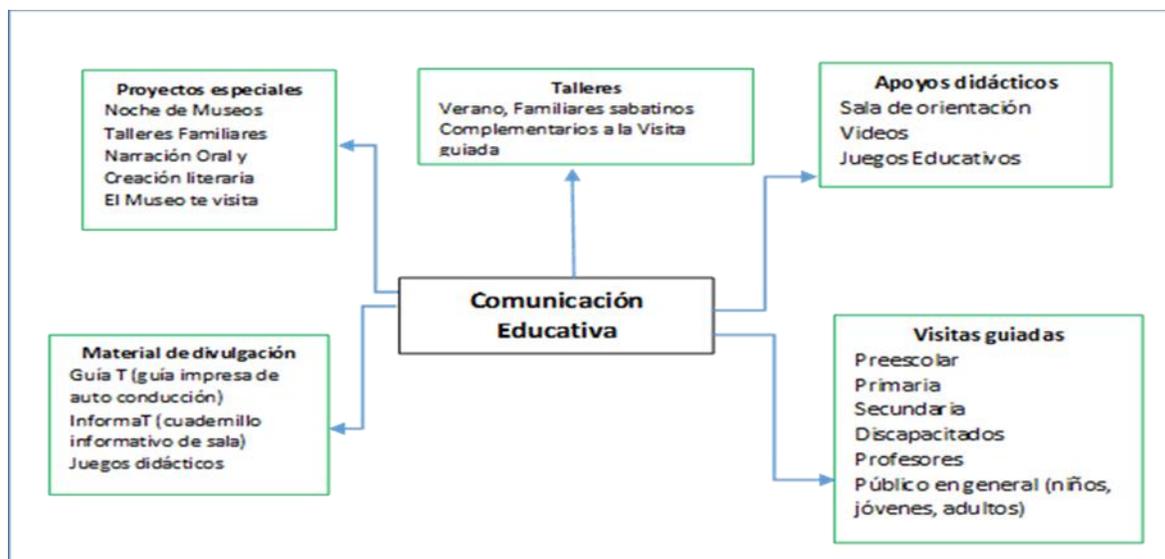
El Departamento de Comunicación Educativa ha contado en toda su historia con el apoyo de prestadores de servicio social de diferentes disciplinas, algunas afines a las tareas sustantivas como la arqueología, antropología, historia, pedagogía, sociología así como de carreras artísticas de danza y teatro. Además de jóvenes de diseño gráfico y de áreas administrativas. Los primeros han apoyado en abrir la cobertura de atención en las visitas guiadas y en actividades complementarias a ellas como los talleres y representaciones teatralizadas. Todos ellos de diferentes universidades públicas y privadas.

Los jóvenes de diseño gráfico contribuyen a dar viabilidad a los materiales informativos y didácticos con el diseño de material atractivo para los diversos públicos. Dicho material se difunde a través de las redes sociales del museo y página web del mismo.

El personal administrativo desarrolla actividades referentes a recursos financieros y controles administrativos del área: inventarios de material utilizados en los talleres, controles de registro de participantes en ellos y otras actividades, manejo de recursos financieros generados por dicho concepto y por último, el personal técnico maneja el equipo de audio y video del auditorio en el que se transmiten videos referentes a los contenidos del museo.

Actividades y Funciones de Comunicación Educativa

Funciones



"Cuadro de actividades y funciones" Fuente: Creación Propia

Funciones del Departamento de Comunicación Educativa

- Elaborar, proponer y analizar los planes, programas y proyectos específicos de servicios y comunicación educativa del Museo, avalados por la Coordinación Nacional de Museos y Exposiciones, a fin de informarle, para su evaluación y seguimiento, para cumplir con lo establecido en las directrices, manuales y lineamientos y las acciones pedagógicas de las áreas educativas del INAH.
- Formular promocionales y programas para ser transmitidos en medios de comunicación, así como materiales impresos electrónicos, e informáticos para públicos diferenciados, para que den a conocer puntualmente el acervo patrimonial del Museo.

- Desarrollar y participar en reuniones y encuentros de especialistas en comunicación educativa y promoción cultural con la finalidad de actualizar los conocimientos de servicios educativos.
- Representar y participar en encuentros con especialistas, en reuniones con la sociedad civil e instituciones públicas para garantizar la presencia del Instituto en el desarrollo educativo de los Museos.
- Programar y coordinar estrategias de servicios y comunicación educativa, promoción cultural y difusión para públicos fuera del museo, con la finalidad de incrementar el rango de acción del museo.
- Recomendar y ejecutar eventos educativos y culturales, visitas guiadas, cursos, talleres, obras teatrales, considerando el perfil de los diferentes públicos, para garantizar una mejor comprensión de las exposiciones.
- Recomendar y ejecutar la producción de materiales didácticos como apoyo a la visita y para la divulgación del patrimonio histórico que resguarda el Museo, con la finalidad de incrementar la información adquirida en la visita.
- Proporcionar a los visitantes escolares de nivel básico los servicios de orientación y atención, con la finalidad de facilitar la visita al Museo.
- Elaborar y aplicar un Programa Anual de Actividades para los diferentes públicos y grupos escolares que realizan visitas guiadas en el Museo, garantizando la calidad de las actividades educativas y culturales que se realizan en él.
- Recomendar en medios de comunicación (prensa, radio, televisión, internet y otros) los contenidos temáticos del Museo, los eventos culturales que organiza y los servicios que presta para facilitar la divulgación de los mismos.
- Analizar las investigaciones y encuestas de carácter museo pedagógico con la finalidad de aplicar los resultados al diseño de estrategias de comunicación educativa.
- Proporcionar los servicios de asesoría pedagógica a estudiantes, profesores y público en general con la finalidad de coadyuvar a procesos de educación formal y no formal.
- Programar, gestionar y desarrollar visitas al Museo para atender a las instituciones privadas, públicas, centros de trabajo y escuelas.
- Recomendar y ejecutar el diseño de estrategias de comunicación y de promoción cultural en la exposición permanente, las exposiciones temporales, al inmueble histórico, con la finalidad de transferir modelos de apropiación patrimonial¹³.

Programas y Proyectos

Programas

Visitas guiadas. Es un programa permanente de atención a grupos escolares de Educación Básica que en coordinación con las autoridades de la SEP y directamente con los profesores interesados en vincular los contenidos escolares con los contenidos disciplinares del museo, programan sus visitas para recibir visitas guiadas a las salas de su interés. En él participan los Gestores Culturales y Asesores Educativos como responsables de la conducción de grupos; a esta labor se suman estudiantes de nivel licenciatura de carreras afines a los contenidos y funciones del museo: Arqueología, antropología, etnología, historia, pedagogía, sociología, teatro, música, danza, quienes previa

¹³ INAH, *Manual de organización del Instituto Nacional de Antropología e Historia*. México, 2017, INAH

capacitación otorgada por el personal del Departamento realizan su servicio social, sumándose a la conducción de grupos.

Talleres Educativos. Los talleres educativos se imparten por lo regular a los grupos que toman visitas guiadas y se dan como complemento o reforzamiento de los contenidos vistos en las salas visitadas. Estos son muy variados, desde pinturas en piezas de cerámica alusivas a las piezas arqueológicas o etnográficas que se pudieron apreciar en salas hasta ejercicios de dibujo, armado y recorte de material, todo relacionado a los contenidos del museo.

Talleres de Verano. Los talleres de verano son actividades especiales que se programan durante un mes en los periodos de vacaciones escolares para que los niños y jóvenes disfruten del museo de una manera más libre y recreativa, fuera de la obligación de las visitas escolares. En ellos pueden conocer de manera más amplia el museo ya que la duración de los talleres es de por lo menos una semana.

Proyectos

El Museo te visita:

Asesores educativos y Gestores culturales del departamento, visitan espacios educativos y comunitarios donde imparten talleres con temáticas del museo, obras de teatro y pláticas entre otras, para promover el conocimiento del patrimonio cultural.

Noches de Museo

Este proyecto de reciente creación, se propone responder a las necesidades de público que por cuestiones laborales principalmente no pueden acudir en los horarios regulares del museo, de tal manera que se les atiende en un horario nocturno, en pequeños grupos fuera de las visitas escolares obligadas. Se dirige especialmente a público joven y adulto, aunque de manera excepcional se ha convocado a público infantil.

Talleres Sabatinos

En respuesta a la dinámica de visitas y necesidades de los públicos en los últimos años, se han enriquecido y diversificado las actividades sabatinas en las que se ha observado una afluencia mayor de familias, es por ello que estos talleres que incluyen visitas guiadas van dirigidos especialmente a ellos.

Narración Oral y Creación Literaria

Actividades con las que propone promover el uso de la lengua oral y escrita vinculada a los contenidos del museo. Consiste en una serie de conferencias, talleres y representaciones teatrales programadas los días sábados encabezadas por invitados hablantes de lenguas indígenas tales como: Tzotzil, Otomí y con la participación también del personal del departamento.

El departamento ha enriquecido su trabajo y ha diversificado sus actividades y funciones conforme se ha incrementado la demanda de sus servicios; es por esta razón que se actualiza y diseñan nuevos materiales para la visita y los talleres con base en técnicas de pedagogía apoyadas en los avances tecnológicos y técnicas de comunicación.

Debido a su actividad este departamento reviste de gran importancia para el Museo, como lo señala Ortiz Lanz "es importante subrayar que los expertos en servicios educativos

constituyen una figura fundamental en materia de difusión de la cultura, pues son ellos quienes se encargan de hacer entendibles los conocimientos que generan los especialistas y los diseñadores de una exposición.”¹⁴

Es en este contexto que cobra importancia el conocimiento de los públicos del museo para el diseño de estrategias comunicativas que logren la apropiación del discurso museológico de los diferentes públicos que visitan el museo.

En tal sentido como lo señala Graciela Schmilchuk (1996) “los estudios de público pretenden ocuparse de toda la gama de comportamientos y actitudes, hábitos culturales y construcciones ligados al modo en que las personas utilizan su tiempo libre en los espacios concebidos para la recreación y la información, ya que esto no es solo competencia del campo cultural, sino que tiene una evidente dimensión política; y lo complementa con lo dicho por Canclini: “El uso del tiempo libre nos construye como ciudadanos, como sujetos sociales “sujetados”, nos impulsa o impide pensar, sentir y actuar sobre la realidad y sobre nosotros mismos ”¹⁵

Esta es la premisa que da origen a este trabajo, cuyo objetivo fundamental es el identificar y clasificar al público visitante y participante de las actividades del Departamento de Comunicación Educativa del MNA.

Las preguntas que dan inicio a este trabajo son: ¿Quién es el público que actualmente visita y participa en las actividades del Departamento de Comunicación Educativa?, ¿Existen métodos estadísticos que permitan una clasificación integral de variables con las cuales sea posible conocer a este público ampliamente?

Para poder contestar estas preguntas es necesario decir, que la investigación y evaluación realizadas sobre este público se basa en parte en las técnicas de los estudios de público, cuyo teoría y desarrollo se han realizado con éxito por investigadores de diferentes partes del mundo como Estados Unidos, Reino Unido, España y México, en diversos museos privados y públicos con discursos disciplinares diversos.

1.10 Estudios de Público

Los estudios de público comenzaron con los estudios a visitantes realizados por Ives Gilman en 1916. La investigación del público se ha configurado como una gran disciplina que ha dado soporte metodológico a este tipo de estudio y cuyos resultados ha permitido organizar una gran cantidad de conocimientos sobre las características, los comportamientos y las experiencias de los visitantes de museos.¹⁶

En los años sesenta y setenta el acercamiento entre museo y población se buscaba a través del aumento de los servicios educativos, de las actividades paralelas o el tratamiento de los temas de exposición¹⁷ Durante los últimos años del siglo XX y la primera década del siglo XXI, se han delimitado con bastante precisión, los perfiles básicos de los visitantes a estos recintos en casi todo el mundo. Se han podido describir sus características

¹⁴ Elda, M. *El Universal*, abril 2002, <https://archivo.eluniversal.com.mx/cultura/22462.html>.

¹⁵ Schmilchuk, G. *Venturas y desventuras de los estudios de público*, México, 1996, CENIDIAP.

¹⁶ Pérez Castellanos, L. *Buenas prácticas en la investigación del público en museos*, 2018.

¹⁷ *Ibid.*

sociodemográficas y psicológicas con precisión, lo que ha permitido tener un buen conocimiento de la audiencia real y potencial de los museos¹⁸.

Las nuevas tecnologías en la informática que han sido un gran apoyo para los métodos estadísticos como el análisis multivariado, hacen posible el estudio de amplios grupos sociales con diversas variables que permiten una observación y conocimiento integral, abriendo posibilidades de una mejora continua en el estudio de públicos.

Al finalizar el capítulo uno, se mencionó la importancia de estudiar al público visitante del Departamento de Comunicación Educativa del MNA, cuyo objetivo fundamental de este trabajo es lograr desarrollar una clasificación precisa de los visitantes y participantes de las actividades del departamento.

Los estudios de público se integran al área de la museología que se dedica a la investigación de los visitantes de los museos, y de otras instituciones afines, desde un punto de vista más amplio, el cual incluye no solo a los visitantes reales sino también a los potenciales e incluso a los llamados no públicos¹⁹.

Los estudios de público se definen como el proceso de obtención de conocimiento sistemático de y sobre los visitantes de museos, actuales y potenciales, con el propósito de incrementar y utilizar dicho conocimiento en la planificación y puesta en marcha de aquellas actividades relacionadas con el público (Committee on Audience Research and Evaluation, Asociación Americana de Museos, 1991)²⁰

La investigación de público es un medio para comprender y obtener información valiosa de los visitantes con el fin de utilizarla en la planificación de acciones, proyectos o programas museísticos. Comprobar mediante investigaciones bien diseñadas si la participación del público en la producción expositiva y museística mejora la interacción entre el público y las colecciones.

Estos estudios pretenden ocuparse de toda la gama de comportamientos y actitudes, hábitos culturales y construcciones imaginarias ligados al modo en que la gente utiliza su tiempo libre en los espacios concebidos para la recreación y la información, ya que esto no es solamente competencia del campo cultural, sino que tiene una evidente dimensión política²¹.

Los problemas que impulsan a realizar este tipo de estudios son variados; diversos espacios institucionales los encargan y financian con el fin de ajustar sus políticas culturales. Los síntomas visibles puntuales que los desencadenan pueden ser la preocupación por la baja afluencia de visitantes en relación con la oferta amplia de algunos museos y con las expectativas de su personal, una mayor afluencia esperada por el museo y la consiguiente dificultad para brindar una atención de calidad; el deseo y capacidad de algunos museos de crecer y ampliar sus públicos, conocer el impacto comunicativo de cierta exposición o de secciones de la misma, etcétera²².

¹⁸ Pérez Castellanos, L. *Buenas prácticas en la investigación de públicos en museos*, 2018.

¹⁹ Pérez Castellanos, L. *Público y museos ¿qué hemos aprendido?*, 2016.

²⁰ *Ibíd.*

²¹ Schmilchuk, G. *Venturas y desventuras de los estudios de público*, México, 1996, CENIDIAP.

²² *Ibíd.*

Los estudios de público tienen un doble nivel de influencia: El primer nivel se plasma en las actividades que involucran directamente a los visitantes y sus experiencias dentro del museo.

Las preguntas más usuales que se formulan son: ¿quiénes son los visitantes?, ¿Cuáles regresan y cuáles no? ¿Cuál es la imagen del museo o de una exposición especial en algunos sectores sociales?, ¿Cuál ha sido el éxito de la exposición en términos cuantitativos y cualitativos?, ¿Qué y cuánto aprende la gente?, ¿cuáles son los mensajes que se ha logrado comunicar y por qué?²³

Estos son solo algunos aspectos parciales de la compleja interacción museo-público y la búsqueda de respuestas ha puesto en práctica técnicas como los cuestionarios escritos, la encuesta, la observación, las entrevistas individuales y grupales²⁴.

El segundo nivel de influencia debe considerarse desde una perspectiva más amplia, es la adopción de un paradigma interpretativo que ayude a una mejora en el conocimiento del público enfatiza el punto de vista de la audiencia en la conceptualización de políticas y estrategias en museos e instituciones culturales (Romanello, 2013). (Pérez, Castellanos Leticia, 2018)

Son estas investigaciones las que buscan ayudar a comprender las condiciones que permiten la visita a museos u otros recintos culturales, ofrecen posibilidades de obtener información aplicable a políticas de cambio, sin disminuir la combinación de métodos y técnicas cuantitativas y cualitativas. (Schmilchuk, 1996)

Características de los estudios de público: Buscan conectar a los usuarios o participantes, estos estudios son para el servicio de la sociedad, promueven y valoran la inclusión social, son estudios sistemáticos de investigación social, se orientan hacia la sociología y psicología, buscan la efectividad, están centrados en el conocimiento, las variables de interés son: motivación, expectativas, experiencias, hábitos de visita, identidad, aprendizaje, integración social, tiene como finalidad la cultura de la evaluación y rendición de cuentas. (Pérez, Castellanos Leticia, 2018)

Es común pensar que los estudios de público y los estudios de mercado son semejantes, esto radica en que ambos buscan contar con información relativa a los destinatarios de las estrategias del museo: sus percepciones e ideas sobre lo que esta institución ofrece, así como el grado de satisfacción, además de que utilizan técnicas comunes, como los sondeos, la observación o las entrevistas. No obstante tienen diferencias claras. Los estudios de mercado tienen un enfoque más comercial, pues se orientan a la venta de marca del museo, la publicidad y los productos que son de mayor interés para cierto tipo de visitantes, todo de manera prospectiva centrada en procesos de planeación en apoyo de la identificación de oportunidades futuras, generación de ganancias y crecimiento de la institución (Tenneson, 2011) (Pérez, Castellanos Leticia, 2016)

Los estudios de mercado son factibles en el museo ya sea para mantener los públicos actuales (consolidación), incrementarlos (cuota de mercado) o investigar sobre nuevos públicos o los actuales desde la perspectiva mercadológica (segmentación y nichos de mercado), mientras que los estudios de públicos buscan informar al personal de la institución, o contribuir de una forma más amplia con sus labores, ya investigando los productos como tales, ya siendo sensibles a los intereses, necesidades, expectativas y

²³ *Ibíd.*

²⁴ *Ibíd.*

motivaciones de los diferentes públicos, y proveyendo evidencias de la eficacia de esos programas más allá de la cuota de ventas, esto es, de acuerdo con la calidad de la experiencia en el museo y sus aportaciones educativas. (Pérez, Castellanos Leticia, 2016)

Los estudios de público como consecuencia de su interdisciplinaridad cubren todo el espectro de experiencias y variables evaluables como es descrito en la siguiente figura (Wells y Butter (2002)). (Pérez, Castellanos Leticia, 2018)

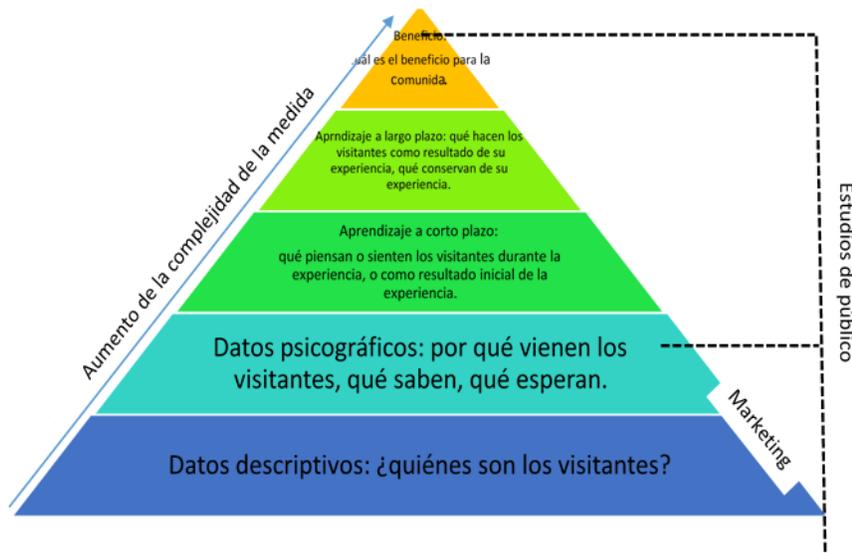


Ilustración 1.3 Jerarquía de la evaluación centrada en el visitante (Wells y Butler, 2002). (Pérez, Castellanos Leticia 2018)

Las áreas de aplicación de los estudios de público son variadas y pueden ser realizados desde los siguientes tipos: diagnóstico, evaluación, monitoreo e investigación.

La investigadora Eloísa Pérez Santos divide las áreas de aplicación en: análisis y captación de públicos, diseño y desarrollo de exposiciones, diseño y desarrollo de programas de actividades, diseño de servicios generales y servicios de atención al visitante. (Pérez, Castellanos Leticia, 2016)

La importancia de llevar a cabo un buen estudio de público consiste en una participación interdisciplinaria y tener muy claros los objetivos que dan origen al estudio además por lo regular son a largo plazo con el propósito de poder realizar evaluaciones que permitan comparar periódicamente los resultados y las estrategias implementadas, como ejemplo, en el año de 1952 el investigador Arturo Monzón realizó un estudio cuyo objetivo era conocer al público visitante al Museo Nacional de Antropología, que permitiera trazar una política adecuada para aumentar el número de visitantes. (Monzón, 1952)

Este estudio se basó en la elaboración de estadísticas referentes al decenio 1942 – 1951 en conjunto, y por separado las del año de 1952, complementadas con una encuesta entre los visitantes, fueron los tres medios principales puestos en práctica para conocer los datos sobresalientes relativos a los visitantes. (Monzón, 1952)

Las estadísticas permitieron obtener el número y proporciones de los visitantes para toda la época estudiada, por años y meses, diversos tipos de visitantes, índices mensuales, los incrementos y, en términos generales, una medición de los resultados de los factores que actuaron para modificar el monto de los visitantes al Museo. (Monzón, 1952)

Se ha explicado brevemente lo que es un estudio de público su origen, definición, características y áreas de aplicación. Por lo que es importante mencionar que este trabajo se basa en algunos lineamientos y características de lo anteriormente definido, en particular el estudio correspondiente será el de diagnóstico, evaluación y clasificación de público, ya que una de las primeras tareas de esta investigación, es el conteo y mapeo de los visitantes, pues esto permitirá un primer acercamiento con el público por estudiar. Posteriormente será el de conocer sus experiencias, motivaciones, aprendizaje, identidad y expectativas que permitan realizar una clasificación.

La forma de obtener la información será a través de encuestas, instrumento que es descrito como una de las técnicas en el estudio de público, a través de las variables que se generen en las encuestas se realizara un análisis para conocer si existe una correlación entre éstas, las cuales tienen una correlación mayor, así como si es posible que estas correlaciones permitan realizar una clasificación de variables y sujetos, en el mejor de los casos una agrupación tanto de variables como de individuos.

Todo esto con la finalidad de poder contestar las preguntas planteadas al inicio de este capítulo, debo decir, lo fundamental que es para esta investigación, el que esta sustentada en un análisis y desarrollo estadístico, que permita una clasificación con métodos ya desarrollados y probados matemáticamente, como son los métodos propuestos para este trabajo: Componentes Principales, Análisis Factorial y Conglomerados.

Los primeros dos métodos permitirán una clasificación de variables y sujetos aportando un panorama general con todas las variables a la vez, así también se podrá corroborar los resultados ya que estos métodos no se pueden corroborar por sí mismos, el método de conglomerados permite realizar una agrupación de variables e individuos lo que permitirá una clasificación integral.

En el siguiente capítulo se describe cada uno de estos métodos, en qué consisten, cuándo surgieron, y su desarrollo matemático

Capítulo 2. Métodos Multivariados

2.1 Componentes Principales

Es un método de análisis multivariado que busca correlaciones importantes entre “n” observaciones con “p” variables numéricas definidas en un conjunto de elementos, permitiendo expresar la información en es nuevas variables que son funciones lineales de las variables originales, llamadas componentes principales, facilitando la observación de “p” variables en dos o tres dimensiones, es decir, permite la reducción de dimensión de variables, ayuda en la agrupación de las observaciones y muestra con facilidad los datos atípicos.

Su origen se encuentra en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson en 1901, aunque la técnica fue desarrollada por Hotelling en 1933. (Everitt, 2001)

Componentes principales son combinaciones lineales de p variables numéricas X_1, X_2, \dots, X_p , definidas en un conjunto de elementos, a partir de estas variables se construyen combinaciones lineales de Y_1, Y_2, \dots, Y_p , variables no correlacionadas entre sí, de tal manera que las primeras variables expresen la mayor proporción de información contenida en las variables originales. (Véliz, 2016)

Su representación es la siguiente:

$$Y_i = u_i X = u_{i1} X_1 + u_{i2} X_2 + \dots + u_{ip} X_p$$

La primera componente es una combinación lineal cuya varianza es la máxima, la segunda componente principal debe ser no correlacionada con la primera componente y debe captar la mayor variabilidad que no ha sido captada por la primera componente, la tercera componente principal debe ser no correlacionada con las dos primeras y debe captar la variabilidad que no han captado las anteriores.

Para que Y_1 capte la mayor variabilidad de los datos, es decir que tenga la varianza máxima los coeficientes que conforman la combinación lineal deben ser los autovectores u_1 de la matriz de covarianza Σ de la variables originales, que correspondan al mayor eigenvalor o autovalor λ_1 .

Es decir, se debe elegir un vector de norma uno, $\|u\| = u_1' u_1 = 1$, para poder obtener la varianza máxima.

$$Var(Y_1) = Var(u_1' X) = u_1' Var(X) u_1 = u_1' \Sigma u_1$$

Para que el segundo componente principal no esté correlacionado con el primero y capte la mayor variabilidad que no ha sido expresada por éste, debe construirse una combinación lineal con coeficientes iguales a los componentes de eigenvectores o autovectores u_2 de la matriz Σ (que es la que se maximizara), cuyo autovalor es λ_2 , que es el siguiente valor del autovalor λ_1 y así sucesivamente hasta obtener los λ_p autovalores de las p componentes principales. (Véliz, 2016)

Las componentes principales se construyen de las combinaciones lineales siguientes:

$$Y_1 = u_{11} X_1 + u_{12} X_2 + \dots + u_{1p} X_p$$

$$Y_2 = u_{21} X_1 + u_{22} X_2 + \dots + u_{2p} X_p$$

$$Y_p = u_{p1}X_1 + u_{p2}X_2 + \dots + u_{pp}X_p$$

Puede escribirse como:

$$Y = U^t X$$

Donde $Y^t = (Y_1, Y_2, \dots, Y_p)$, $X^t = (X_1, X_2, \dots, X_p)$ y

$$U^t = \begin{pmatrix} u_{11} & \dots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{p1} & \dots & u_{pp} \end{pmatrix}$$

Es la transpuesta de la matriz U cuyas columnas corresponden, respectivamente a las coordenadas de los autovectores unitarios u_1, u_2, \dots, u_p de la matriz de covarianza Σ .

Dado que el producto de la matriz U y su transpuesta U^t es igual a la matriz identidad, las variables originales se expresan en términos de las componentes principales como:

$$X = UY$$

La varianza de cada componente principal Y_i es igual al autovalor λ_i , es decir que: $V(Y_i) = \lambda_i$

La suma de los elementos de la diagonal de la matriz de covarianza Σ que es la suma de las varianzas de las variables originales, es igual a la suma de los autovalores de las componentes principales, que se expresan como: $\sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i$, donde σ_{ii} es la varianza de la variable X_i .

Si cada una de las variables originales son estandarizadas, la suma de los autovalores es igual al número de variables originales, si se considera que la información que contienen las variables está medida por la suma de sus varianzas, se tiene que esta información puede ser medida por la suma de los autovalores, si los m primeros autovalores representan una gran parte de la suma de todas las varianzas, será suficiente usar los m primeros componentes principales en lugar de las p variables originales para tener gran parte de la información, logrando así reducir la dimensión.

La correlación entre la variable X_k y la componente principal Y_i está dada por:

$$\text{corr}(Y_i, X_k) = \frac{u_{ik}\lambda_i}{\sqrt{\text{Var}(X_k)\text{Var}(Y_i)}} = \frac{u_{ik}\sqrt{\lambda_i}}{\sigma_k}$$

Donde: σ_k es la desviación estándar de X_k y u_{ik} es el coeficiente de X_k en la combinación lineal que define a Y_i .

Puede ser usada la matriz de correlaciones en lugar de la matriz de covarianzas, los criterios se tratan más adelante, pero en principio al utilizar la matriz de correlaciones equivale a estandarizar las variables iniciales para posteriormente usar la matriz de covarianzas de estas variables.

Construcción Matemática.

Las componentes principales son combinaciones lineales de p variables de la forma:

$$Y_i = u_i'X = u_{i1}X_1 + u_{i2}X_2 + \dots + u_{ip}X_p \quad \text{ó} \quad Y_i = u_i'(X - \mu) \text{ (centradas)}$$

Dónde: $i = 1, 2, \dots, p$

Elegimos un vector de norma uno $\|u_1\| = 1$, es decir $\|u_1\| = 1 = u_1'u_1 = 1$ tal que:

$$\text{Var}(Y_1) = \text{Var}(u_1'X) = u_1'\text{Var}(X)u_1 = u_1'\Sigma u_1$$

Para obtener la varianza máxima.

Construcción de la primera componente.

Sea $F(u)$ una función que se define en igualdad a la varianza como:

$$F(u) = V(Y) = V(u'X) = u'V(X)u = u'\Sigma u$$

Para que la maximización de $F(u)$ tenga solución se impone la siguiente restricción al vector u

$$\lambda\|u\| = u'u = 1$$

Introduciendo esta restricción mediante el multiplicador de LaGrange (es decir restando $F(u)$ menos λ), obtenemos:

$$F(u) = u'\Sigma u - \lambda(u'u - 1)$$

Calculando la matriz jacobiana como función de u , (es decir se deriva respecto al vector u) e igualamos a cero. Entonces:

$$\frac{dF(u)}{du} = 2\Sigma u - 2\lambda u = 0$$

De aquí se obtiene la solución

$$\Sigma u = \lambda u$$

Esto implica que u es un autovector de la matriz Σ y λ es el autovalor correspondiente a el autovector.

Para determinar el valor propio de Σ es el que corresponde a la solución de la ecuación anterior, se multiplica por u' la ecuación anterior.

$$u'\Sigma u = \lambda u'u \Rightarrow u'\Sigma u = \lambda$$

Por lo tanto $V(Y_i) = \lambda_i$, esta es la cantidad que se desea maximizar, λ es el autovalor más grande de la matriz Σ con u el autovector asociado a este autovalor.

Dentro de las condiciones la siguiente componente principal debe tener la mayor varianza del remanente, una vez que se ha calculado la primera componente, y no estar correlacionada con la primera, esto implica que los vectores u_1, u_2 son ortogonales por lo que u_2 tendrá que ser de norma uno, entonces serán ortonormales.

Para garantizar que la segunda componente es de mayor varianza, después de la primera es que la suma de estas dos varianzas sea máxima. Por lo que se tiene que volver a maximizar.

Sea

$$F(u_1, u_2) = u'_1 \Sigma u_1 + u'_2 \Sigma u_2 - \lambda_1 (u'_1 u_1 - 1) - \lambda_2 (u'_2 u_2 - 1)$$

Que incorpora las siguientes restricciones:

$$u'_1 u_1 = 1, u'_2 u_2 = 1 \text{ y } u'_1 u_2 = 0 \text{ por lo tanto se deduce que } u'_1 \Sigma u_2 = 0 \text{ y } \mu u'_1 u_2 = 0$$

Se deriva la función respecto a los vectores u_1 y u_2 , entonces

$$\frac{dF(u_1, u_2)}{du_1} = 2\Sigma u_1 - 2\lambda_1 u_1 = 0$$

$$\frac{dF(u_1, u_2)}{du_2} = 2\Sigma u_2 - 2\lambda_2 u_2 = 0$$

Siendo que $u'_1 u_2 = 0$, ya que son ortonormales, se multiplica la parcial respecto a u_1 por u'_1

Se tiene

$$u'_1 \Sigma u_1 = \lambda_1 \Rightarrow u_1 u'_1 \Sigma u_1 = \lambda_1 u_1 \Rightarrow \Sigma u_1 = \lambda_1 u_1$$

De manera similar, se multiplica la parcial respecto a u_2 por u'_2 , como $u'_2 u_1 = 0$, ya que son ortonormales se tiene:

$$u'_2 \Sigma u_2 = \lambda_2 \Rightarrow u_2 u'_2 \Sigma u_2 = \lambda_2 u_2 \Rightarrow \Sigma u_2 = \lambda_2 u_2$$

Esto implica que u_1 y u_2 deben ser autovectores de Σ . Tomando los autovectores de norma uno se sustituyen en la función objetivo.

Por lo que se obtiene:

$$\lambda_1 u'_1 u_1 + \lambda_2 u'_2 u_2 - \lambda_1 (u'_1 u_1 - 1) - \lambda_2 (u'_2 u_2 - 1) - \mu u'_1 u_2 = \lambda_1 + \lambda_2$$

$$F(u_1, u_2) = \lambda_1 + \lambda_2$$

De modo que λ_1 y λ_2 son los dos eigenvalores más grandes de la matriz Σ y u_1 y u_2 los eigenvectores correspondientes.

La forma general de la j-ésima componente principal se da de la siguiente forma:

$$Y_j = u'_j X$$

Siendo: $j=1,2,\dots,p$, u'_j el eigenvector de la matriz Σ asociado al eigenvalor λ_j , con $\lambda_1 > \lambda_2 > \dots > \lambda_p$.

Propiedades de los componentes principales.

1.- Conservan la variabilidad original de los datos: Es decir la suma de las varianzas de las componentes es igual a la suma de las varianzas de las variables originales

Se tiene por construcción que: $V(Y_1) = \lambda_1$, $V(Y_2) = \lambda_2$, así sucesivamente, la covarianza expresada como $Cov(Y_1, Y_2) = 0$, en general $Cov(Y_i, Y_j) = 0$ para $i \neq j$ $i, j = 1, 2, \dots, p$.

Entonces

$$\text{traza}(\Sigma) = \sum_{i=1}^p V(X_i) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{Var}(Y_i)$$

Es decir que la variable Y_i conjuntamente la misma variabilidad que las variables originales, la suma de la varianza es la misma pero su estructura es diferente.

2.- La proporción de la varianza total explicada por una componente, es el cociente entre su varianza y la suma de los valores propios de la matriz, por lo que se afirma que el i -ésimo componente principal explica una proporción de varianza

Esto se expresa de la siguiente forma:

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$$

3.- Las covarianzas entre el vector de variables originales X y la i -ésima componente principal Y_i , es:

$$\text{Cov}(X, Y_i) = \text{Cov}(X, u'_i X) = u'_i \text{cov}(X, X) = u'_i \Sigma = u'_i \lambda_i = \Sigma a_i = \lambda_i u_i \quad i=1, 2, \dots, p$$

Es decir

$$\text{Cov}(X, Y_i) = \text{Cov}(X_1, X_2, \dots, X_p, Y_i) = \lambda_i u_i = (\lambda_i u_{i1} u_{i2}, \dots, u_{ip})$$

La covarianza entre la i -ésima componente y la j -ésima variable es:

$$\text{Cov}(X_j, Y_i) = \lambda_i u_{ij}$$

Se conoce que $V(X_j) = \sigma_{jj}^2$ y $V(Y_i) = \lambda_i$ entonces:

$$\text{Cov}(X_j, Y_i) = \frac{\text{Cov}(X_j, Y_i)}{\sqrt{V(X_j)V(Y_i)}} = \frac{\lambda_i u_{ij}}{\sqrt{\sigma_{jj}^2 \lambda_i}} = \frac{\sqrt{\lambda_i} u_{ij}}{\sigma_{jj}}$$

Refleja la contribución relativa de cada variable en la componente. Por lo que es necesario analizar el patrón de las u_{ij} de cada componente.

El peso que tiene la variable i en la componente j , está dado por u_{ij} . El tamaño relativo de las u_{ij}

Si se utiliza la matriz de correlación para realizar el análisis de componentes principales, se tiene que $\sigma_{jj}^2 = 1$, entonces $u_{ij}^* = \sqrt{\lambda_j} u_{ij}$, esto es interpretado como el coeficiente de correlación entre la variable j y el componente i .

Matriz de componentes principales propiedades y su estructura de correlación.

Se denota por U a la matriz de componentes principales

$$Y = XU$$

$$U = (u_1, u_2, \dots, u_p) = \begin{pmatrix} u_{11} & \cdots & u_{p1} \\ \vdots & \ddots & \vdots \\ u_{1p} & \cdots & u_{pp} \end{pmatrix}$$

Propiedades de la Matriz U

- 1.- Cada columna de la matriz es un vector propio de la matriz de covarianza Σ .
- 2.- U es ortogonal, es decir: $U'U = UU' = I_p \Rightarrow U' = U^{-1}$
- 3.- La matriz de covarianza por la matriz de eigenvectores U es igual a la matriz de eigenvectores U por la diagonal de eigenvalores es decir $\Sigma U = U\Lambda$, con $\Lambda = \text{diagonal}(\lambda_1, \lambda_2, \dots, \lambda_p)$, y $\lambda_1 > \lambda_2 > \dots > \lambda_p$.

Estructura de correlación

Sea

$$V(Y) = V(XU) = U'V(X)U = U'\Sigma U = U'U\Lambda = \Lambda = \text{diagonal}(\lambda_1, \dots, \lambda_p)$$

Entonces:

$$\text{Cov}(Y_i, Y_j) = 0, \text{ si } i \neq j \text{ y } \text{Var}(Y_i) = \lambda_i \geq \text{Var}(Y_j) = \lambda_j \text{ si } i \leq j$$

Así también $\text{traza}(\Sigma) = \text{traza}(\Sigma U U') = \text{traza}(U \Lambda U') = \text{traza}(\Lambda) = \sum_{j=1}^p \lambda_j$, como la $\text{traza}(\Sigma) = \sum_{j=1}^p \sigma_{jj}^2$ se tiene que $\sum_{j=1}^p \lambda_j$ es una medida de variación total de los datos.

Al aplicar componentes principales la matriz a utilizar es la “Matriz de varianza – covarianza” o “Matriz de correlación”

Matriz de varianza – covarianza

Al utilizar la matriz de varianza covarianza se debe tener en cuenta dos condiciones en los datos:

Las variables medidas tienen que tener las mismas unidades o estar en unidades comparables.

Tener una varianza de tamaño semejante.

Al no tener las variables medidas en las mismas unidades, cualquier cambio en la escala de medición en una o más variables afectará las componentes principales. Si una variable tiene una varianza mayor que las demás, dominará el primer componente principal, sin importar la estructura de covarianza de las variables.

Si no se cumple con las condiciones para un análisis de componentes principales con la matriz de varianza, se puede realizar con la matriz de correlación.

El análisis de componentes principales a través de la matriz de correlación, equivale a realizar un análisis a datos estandarizados en lugar de los datos originales, esto implica asumir que las variables tienen igual importancia dentro del análisis.

Componentes principales con la matriz de correlación.

Al realizar un análisis de componentes principales, primero es estandarizar datos, posteriormente obtener la matriz de correlaciones de la que se partirá para obtener las componentes principales.

Observación: Si las variables están muy poco correlacionadas esta técnica no permitirá hacer una reducción de dimensión, es decir que el número de dimensión real de los datos será la misma que el número de variables medidas.

Elección del número de componentes principales.

Existen varios criterios para seleccionar el número de componentes principales uno de ellos se basa en el porcentaje de varianza total explicada por los componentes, en la práctica se sugiere considerar un número de componentes principales que retenga entre 70% y 90% de la varianza total.

Otro criterio utilizado es cuando las variables originales difícilmente tendrán varianza 1 (Cuando las variables se estandarización), este considera tan sólo los componentes principales cuyos eigenvalores (autovalores) asociados son mayores que 1. Este criterio se conoce como “Criterio o regla de Kaiser” y se justifica aduciendo que cada componente principal debería expresar una cantidad de información mayor a la que explica cada variable. (Véliz, 2016)

Un tercer criterio consiste en la visualización de la gráfica de codo (SCREE), consiste en una gráfica de puntos que representan a los eigenvalores (autovalores) que tienden a nivelarse horizontalmente, cuando están lo suficientemente cercanos a cero y pueden ignorarse, por lo que se elige el número de componentes igual al número de eigenvalores antes que la gráfica se nivele. (Everitt, 2001)

Elección para matriz de varianza o de correlación

Para saber elegir que matriz es la adecuada en un análisis de componentes principales es necesario observar los siguientes criterios:

Matriz de varianza – covarianza

- 1.- Variables medidas en las mismas unidades o en unidades comparables
- 2.- Varianzas de tamaños semejantes.

Matriz de correlación

- 1.- Estandarizar los datos, en lugar de los datos crudos.

La elección de la matriz de varianza depende de que los datos por analizar cumplan los criterios mencionados, de lo contrario se recomienda hacerlo con la matriz de correlación, cabe destacar que realizar el análisis de componentes principales con la matriz de correlación, implica intrínsecamente asumir que todas las variables tiene igual importancia dentro del análisis, supuesto que no siempre puede ser cierto. (Zamora, 2016)

VARIABLES DE MEDIDA EN DIVERSAS ESCALAS

El análisis de componentes principales generalmente se realiza sobre variables continuas, para variables con diversa escalas de medición es posible construir variables dummy's, este proceso tiene la desventaja de incrementar el número de variables dentro del análisis y en ocasiones esto no ayuda a un análisis de componentes, ya que con el aumento de estas variables poco se puede reducir la dimensión. (Zamora, 2016)

Lo más recomendable al tener variables con diversidad de escala, es utilizar una matriz de correlaciones policóricas. Con esta matriz se obtiene la correlación de dos variables de acuerdo a la escala de medición que tengan y se puede aplicar el método de componentes principales. (Zamora, 2016)

Para este análisis la construcción de las variables dummy's, consiste en que el cruce entre cada una de las variables con diversas escalas de medida, permita la obtención de una correlación adecuada, con el fin de construir una matriz de correlación que pueda ser aplicable a los métodos multivariados.

A continuación se muestra una tabla con las diferentes escalas de medición y la correlación que les corresponde.

Escala de medición	Continua	Ordinal	Dicotómica
Continua	Pearson	Policórica	Punto Biserial
Ordinal		Policórica	Policórica
Dicotómica			Tetracórica

Tabla 2.1 Tabla de correlaciones Fuente: Notas Zamora 2016.

Observación de resultados

Los resultados son observados en un tipo de gráfico llamado Biplot. Un biplot es una representación gráfica de la información en una matriz de datos $n \times p$. El "bi" es un reflejo de que la técnica (a través de componentes principales) permite la visualización de información sobre las variables según lo indicado por sus varianzas y covarianzas y las relaciones entre individuos según lo indicado por medidas particulares de distancias interindividuales. (Everitt, 2001)

Interpretación del Biplot.

En el Biplot muestra a los individuos graficados como puntos y las variables como vectores. (Zamora, 2016)

- 1.- Individuos semejantes representaran puntos cercanos en la gráfica
- 2.- Variables cuyo ángulo entre los vectores que las representan sean pequeño, serán variables con una fuerte correlación, como $\cos\theta$ es una función decreciente de 0° a 90° $\cos 0^\circ = 1$, esto implica que los vectores son colinales con correlación = 1 y el $\cos 90^\circ = 0$ entonces su correlación = 0.

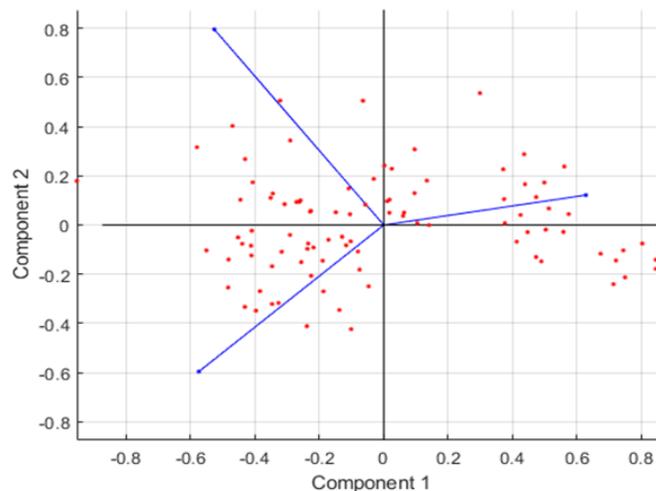


Ilustración 2.2 Gráfica Biplot Fuente: www.mathworks.com/help/examples/stats/win64/ModifyBiplotPropertiesExample_01.p

2.2 Análisis Factorial.

Es una técnica que permite conocer las relaciones de interdependencia o correlación que pueden existir entre grupos de variables numéricas medibles, permite obtener conceptos subyacentes representados con variables referidas como factores. Los factores son variables latentes que no pueden medirse directamente, pero se relacionan con las variables originales mediante combinaciones lineales. (Véliz, 2016)

Su representación es una combinación lineal de la siguiente manera:

$$X_1 = \lambda_1 F + \epsilon_1$$

$$X_2 = \lambda_2 F + \epsilon_2$$

⋮

$$X_p = \lambda_p F + \epsilon_p$$

Estas relaciones indican que es posible expresar la información contenida en las variables iniciales usando tan sólo un factor, pero también puede utilizarse dos factores.

$$X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \epsilon_1$$

$$X_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \epsilon_2$$

⋮

$$X_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \epsilon_p$$

La reducción de la dimensión obtenida al considerar uno o dos factores en lugar de las variables originales, facilita el tratamiento e interpretación de los datos.

Tipos de análisis factorial.

Existen dos tipos de análisis factorial, un primer análisis es conocido como análisis factorial exploratorio; esta es una técnica de exploración de datos, para la cual no se tienen criterios que permitan evaluar la información que contienen los factores encontrados. Con este análisis se obtienen derroteros para formular hipótesis respecto de teorías y de procesos que se encuentran de manera subyacente en las variables iniciales. (Véliz, 2016)

Análisis factorial confirmatorio, permite confirmar las hipótesis planteadas a priori y que están relacionadas con algún proceso de un patrón de correlaciones generado por los valores de las variables originales, este modelo se desarrolla dentro de los modelos de ecuaciones estructurales. (Véliz, 2016)

Modelo matemático.

El análisis factorial se basa en la suposición de que para un grupo de variables iniciales, la varianza de cada una de ellas se puede descomponer en una varianza común a todas las variables, más una varianza debida a errores de medición y propia de la variable. La varianza común es expresada a través de variables latentes (no medibles directamente) llamadas factores comunes y la varianza no común por variables también latentes llamadas factores no comunes. (Véliz, 2016). Se tiene que para p variables numéricas X_1, X_2, \dots, X_p cada una con media 0 y medidas simultáneamente en un conjunto de elementos, existen los factores comunes F_1, F_2, \dots, F_m , con $m < p$ y los factores no comunes $\epsilon_1, \epsilon_2, \dots, \epsilon_p$ que cumplen las relaciones:

$$\begin{aligned} X_1 &= l_{11}F_1 + \dots + l_{1m}F_m + \epsilon_1 \\ X_2 &= l_{21}F_1 + \dots + l_{2m}F_m + \epsilon_2 \\ &\vdots \\ X_p &= l_{p1}F_1 + \dots + l_{pm}F_m + \epsilon_p \end{aligned}$$

Los factores comunes y no comunes tienen media 0, no son correlacionados entre sí, la varianza de cada factor F_i es 1, la varianza de cada ϵ_i es ψ_i^2 y la covarianza entre cada F_i y cada ϵ_j es cero.

Este método utiliza la matriz de covarianzas de las variables originales y se representa de la siguiente forma:

$$\Sigma = L^T L + \Psi$$

Donde L y Ψ son las matrices:

$$L = \begin{pmatrix} l_{11} & l_{21} & l_{p1} \\ \vdots & \vdots & \vdots \\ l_{1m} & l_{2m} & l_{pm} \end{pmatrix} \quad \Psi = \begin{pmatrix} \psi_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \psi_p^2 \end{pmatrix}$$

La matriz L , formada con los coeficientes l_{ij} , se llama matriz de cargas factoriales y es la que prácticamente sirve para expresar la información en las variables originales. La matriz Ψ corresponde a la varianza no explicada por los factores comunes. (Véliz, 2016)

Comunalidades.

Las comunalidades miden la cantidad de información que los factores comunes expresan de cada variable e indican una manera de evaluar la pertinencia de la aplicación del análisis a las variables iniciales. (Véliz, 2016)

A partir de la relación $\Sigma = L^T L + \Psi$ la varianza de cada variable X_i , se tiene:

$$Var(X_i) = l_{i1}^2 + \dots + l_{im}^2 + \psi_i$$

A la suma de cuadrados $l_{i1}^2 + \dots + l_{im}^2$ se llama comunalidad.

Extracción de Factores.

La extracción de factores está referida a los métodos que se aplican para obtención de las cargas factoriales l_{ij} . El método para obtener las cargas factoriales se basa en el uso empírico de los componentes principales, existen también el de máxima verosimilitud y mínimos cuadrados ponderados.

Método de componentes principales.

Utilizando los eigenvalores $\lambda_1, \dots, \lambda_p$ de Σ y los eigenvectores unitarios e_1, \dots, e_p matricialmente se puede escribir:

$$\Sigma = \tilde{L}\tilde{L}^t$$

Donde la i -ésima columna de L es igual al vector $\sqrt{\lambda_i}e_i$; $i = 1, \dots, p$. Si los eigenvalores $\lambda_{m+1}, \dots, \lambda_p$ son iguales a 0, se tendrá que la matriz Σ puede ser factorizada o reconstruida a partir de los m primeros componentes principales.

Si los eigenvalores $\lambda_{m+1}, \dots, \lambda_p$ son pequeños la matriz Σ puede escribirse como:

$$\Sigma = \tilde{L}\tilde{L}^t + \tilde{\psi}$$

Donde \tilde{L} es una matriz cuyas columnas están formadas a partir de los m primeros eigenvectores y eigenvalores de Σ y $\tilde{\psi}$ es una matriz que sus elementos en la diagonal son iguales a $\psi_i = \sigma_{ij} - \sum_{j=1}^m (\tilde{l}_{ij})^2$. (Los elementos que están fuera de la diagonal no son necesariamente iguales a 0).

Los resultados obtenidos se cumplen para la matriz formada con las varianzas y covarianzas muestrales que denotaremos por S , por lo que las propiedades siguientes se aplican a los valores muestrales.

La contribución del primer factor a la varianza muestral s_{ii} es \tilde{l}_{i1}^2 y la varianza total, $s_{11} + s_{22} + \dots + s_{pp}$, es:

$$\tilde{l}_{11}^2 + \tilde{l}_{21}^2 + \dots + \tilde{l}_{p1}^2 = \tilde{\lambda}_1$$

Donde $\tilde{\lambda}_1$ es el mayor eigenvalor de S , de este modo es similar para los otros factores. Modelo revisado por (Véliz, 2016)

Método Máxima Verosimilitud.

Antes de presentar este método es necesario definir algunas expresiones, ya se observó que la matriz de varianza y covarianza de las variables medibles que se expresa en términos del modelo factorial puede asociarse con la matriz de correlaciones. (Zamora, 2016)

Tenemos que:

$$Cov(X_i, X_l) = Cov\left(\sum_{j=1}^k l_{ij}F_j + \epsilon_i, \sum_{j=1}^k l_{lj}F_j + \epsilon_l\right) = \sum_{j=1}^k l_{ij}l_{lj}, \forall i \neq l, i, l = 1, 2, \dots, p$$

Como la varianza y covarianza de las variables medidas pueden ser expresadas en términos del modelo factorial, da la oportunidad que la matriz de correlaciones pueda ser expresada de la siguiente forma:

$$R = \Sigma = L^T L + \Psi$$

Por lo que la matriz de correlaciones que se obtiene del modelo depende de los parámetros del mismo modelo, entonces $\Sigma = \Sigma(\theta)$ y sea R que representa la matriz de correlaciones de los datos.

Por consiguiente se busca un estimador $\Sigma(\theta)$ que pueda minimizar alguna función de distancia entre las dos matrices, la función por minimizar es de la forma:

$$F = G(|\Sigma(\theta) - R|)$$

Donde G es alguna función específica. Los valores en $\Sigma(\theta)$ que minimicen esta función de distancia serán los estimadores de sus parámetros. Dado que Σ se puede descomponer como:

$$\Sigma = \tilde{L}\tilde{L}^t + \tilde{\psi}$$

Los procesos que minimizan esta función de distancia entre estas dos matrices son equivalentes a encontrar los estimadores L y Ψ tales que:

$$R = \hat{\Sigma} \approx \tilde{L}\tilde{L}^t + \tilde{\psi}$$

La función de distancia que se desprende de la verosimilitud del modelo, tiene la forma:

$$F(\Sigma(\theta), R) \propto -\frac{1}{2} \log(\Sigma(\theta)R^{-1}) - traza(\Sigma(\theta)R^{-1})$$

En este método el objetivo es maximizar la verosimilitud, maximizar es equivalente a minimizar el negativo de esta verosimilitud.

Este método requiere que X tenga una distribución normal multivariada, este método es robusto ante desviaciones de la normalidad. Sin embargo es inadecuado en el uso con variables nominales u ordinales.

Mínimos Cuadrados

La función que se minimiza es la siguiente:

$$F(\Sigma(\underline{\theta}), R) = \text{traza}[(R - \Sigma(\underline{\theta}))^2]$$

Se considera como una medida de distancia entre la matriz observada R y la matriz generada por el modelo $\Sigma(\underline{\theta})$. Se minimiza la suma de cuadrados de las diferencias entre estas dos matrices. Los valores que minimice esta función serán los estimadores. (Zamora, 2016)

Mínimos Cuadrados Generalizados

Este método es una generalización de mínimos cuadrados; la función por minimizar es:

$$F(\Sigma(\underline{\theta}), R) = \text{traza} \left[((R - \Sigma(\underline{\theta})) R^{-1})^2 \right]$$

En este caso se pretende minimizar la suma de cuadrados de los elementos de este producto de matices. (Zamora, 2016)

Mínimos Cuadrados Ponderados

En este método el objetivo es minimizar la diferencia entre la matriz generada por el modelo y la estimada por los datos, ponderando las diferencias por una matriz de pesos.

La función que se debe minimizar es la siguiente:

$$F(\Sigma(\underline{\theta}), R) = \text{traza}[(R - \Sigma(\underline{\theta}))\psi^{-1}]^2]$$

Siendo ψ la matriz definida anteriormente. (Zamora, 2016)

Rotación de factores

La interpretación se facilita cuando un grupo de variables están altamente correlacionado con un factor pero tiene baja correlación con los otros factores. Cuando esto no sucede, se acostumbra realizar rotaciones de los factores de tal manera que suceda lo antes expresado.

La razón principal para rotar una solución es clarificar la estructura de las cargas factoriales, si la estructura que muestra las cargas factoriales de la solución inicial son confusas o difíciles de interpretar, una rotación puede proporcionar una estructura más fácil de interpretar.

Algunas rotaciones conservan la perpendicularidad (independencia) de los factores y por ello son llamadas rotaciones ortogonales cuando no conservan la perpendicularidad se llaman oblicuas. (Véliz, 2016)

Rotaciones ortogonales

Uno de los patrones de cargas factoriales más usuales y deseables es la llamada estructura simple de cargas factoriales. Las cargas factoriales presentan una estructura simple si cada variable tiene una gran carga en un solo factor, con cargas cercanas a cero en el resto de

los factores. Las rotaciones que procuran generar una estructura de cargas simple son las rotaciones ortogonales (los nuevos ejes después de la rotación siguen siendo ortogonales). Existen varios métodos para realizar una rotación ortogonal, uno muy utilizado es la rotación varimax propuesta por Kaiser (1985), esta rotación trata de maximizar la dispersión de las carga factoriales para obtener factores que tengan alta correlación con un grupo de variables, pero baja correlación con el resto de las variables. (Véliz, 2016) (Zamora, 2016)

La rotación de los factores facilita la interpretación sin que ello indique la manipulación de la información por parte del investigador. (Véliz, 2016)

Rotaciones Oblicuas

Las rotaciones oblicuas permiten relajar la restricción de ortogonalidad con el fin de ganar en la interpretación de los factores. Con este método los factores resultan correlacionados, aunque generalmente esta correlación es pequeña. Entre las rotaciones oblicuas, se conocen promax y oblimin. (Zamora, 2016)

Número de factores

La determinación del número de factores debe realizarse antes de aplicar alguna rotación, esto facilita la interpretación. Algunos criterios para la elección del número de factores son los siguientes: (Véliz, 2016)

1.- El “Criterio de Kaiser” que se usa cuando la extracción de los factores se realiza con el método de componentes principales y con variables originales estandarizadas. Este criterio solo considera los factores cuyos eigenvalores son mayores que 1. De modo que solo se retienen los factores que explican más varianza que cualquier variable.

2.- El “gráfico de sedimentación”, que es apropiado cuando los factores son extraídos por el método de componentes principales. Este gráfico es una curva que une los puntos (x, y) , donde x indica el orden del factor de acuerdo al tamaño de su eigenvalor, y indica el eigenvalor correspondiente. El número apropiado de factores corresponde al valor x del punto en donde la pendiente de la curva desciende bruscamente. Este criterio es uno de los más usados y está basado en la restitución mínima que con los k primeros factores se puede obtener. (Véliz, 2016)

La restitución esta medida como:

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\sum_{i=1}^p \lambda_i}$$

Entonces, por ejemplo, si se desea restituir un mínimo de 85% de la varianza total, se deben tomar los k primeros factores de tal manera que la fracción de la varianza restituida sea mayor o igual a 0.85.

Importancia de cada factor.

La importancia de cada factor F_i está relacionada con la proporción de la varianza total de la información que este explica.

Cuando los factores se obtienen por componentes principales, la varianza total es igual a la suma de los eigenvalores y el aporte del valor F_i se mide con λ_i por lo que la importancia relativa del factor F_i , respecto todos los factores se determina por $\lambda_i / \sum_{j=1}^p \lambda_j$, si las p variables iniciales son estandarizadas, la importancia del factor F_i se mide con λ_i / p . (Véliz, 2016)

Puntuaciones Factoriales

El análisis factorial se usa en ocasiones como una etapa previa a la realización de otros análisis. De ahí que sea necesario estimar cada factor en cada observación. La operación es directa cuando los factores son extraídos por el método de componentes principales cuando no es posible extraerlos por componentes principales se usa el método de mínimos cuadrados ponderados. (Véliz, 2016)

Este método consiste en la suma de las variables que cargan más en el factor y que previamente han sido estandarizadas.

Validación de resultados

El análisis factorial es una técnica de carácter exploratorio, no cuenta con resultados que permitan comparar los resultados obtenidos. Cuando se dispone de una muestra de suficiente tamaño, ésta se parte en dos mitades aleatorias y en cada una de ellas se realiza el análisis. Si los resultados son parecidos a los que se obtienen con toda la muestra, se puede considerar que se tiene una estabilidad en los factores. (Véliz, 2016)

Número de elementos de la muestra

No existe una regla que indique definitivamente el número de elementos que debe tener una muestra de trabajo. Algunos autores sugieren que el número de sujetos en la muestra debe ser por lo menos igual a 10 veces el número de variables, otros indican que por lo menos debe tener 150 elementos en la muestra, lo que es un hecho es que se debe tener una muestra con un número considerable de elementos. (Véliz, 2016)

Puesto que el análisis factorial parte de descomponer la matriz de correlación, los elementos no redundantes de ésta representa la información necesaria para analizar, esto puede ser explicado de la siguiente manera.

Al tener p variables medidas, el número de elementos no redundantes es $p(p + 1)/2$. Se necesita estimar $p \cdot k$ cargas factoriales totales y p especificidades, entonces se requiere estimar $p(k + 1)$ parámetros del modelo.

Se necesita imponer a este número de parámetros por estimar $k(k - 1)$ restricciones para obtener una solución única. La diferencia entre los parámetros por estimar y las restricciones no debe de exceder el número de elementos no redundantes de la matriz de correlación. (Zamora, 2016)

Por lo que se debe cumplir la siguiente desigualdad:

$$\frac{p(p + 1)}{2} \geq p(k + 1) - \frac{k(k - 1)}{2} \Rightarrow (p - k)^2 \geq p + k$$

A partir de esta desigualdad podemos observar que el mínimo de variables requeridas para extraer un factor es 3. Con cinco variables observadas se puede tener a lo más dos factores, con veinte el número máximo de factores puede ser hasta 14, sin embargo no se busca este número máximo sino aquel que permita explicar lo mejor posible las correlaciones entre las variables medidas. (Zamora, 2016)

Consideraciones generales a seguir en un análisis factorial.

El análisis factorial es una técnica de análisis de datos en donde se exploran las interrelaciones que pueden existir entre un conjunto de variables. Estas variables deberán ser numéricas, aunque también puede aplicarse a variables que tengan otras escalas de medición, utilizando la matriz de correlación Policórica. (Véliz, 2016) (Zamora, 2016)

El primer paso es verificar si es pertinente la aplicación del análisis, ya que la idea es reducir el número de variables y lo óptimo es tener variables correlacionadas. Por lo que se tiene que observar que en la matriz de correlaciones las entradas no sean menores de 0.3, si esto ocurre, no es recomendable utilizar este análisis. (Véliz, 2016)

Existen algunas medidas y pruebas para estudiar la adecuación del análisis entre ellas se tiene: el índice de KMO (Kaiser – Meyer - Olkin) y la prueba de esfericidad de Bartlett.

El índice K M O (Medida de adecuación muestral). La medida está definida por:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} r_{ij \bullet m}^2}$$

Esta prueba es un índice que compara los coeficientes de correlación r_{ij}^2 con los coeficientes de correlación parcial $r_{ij \bullet m}^2$. Esta última correlación es la correlación entre dos variables, eliminando el efecto de las restantes variables incluidas en el análisis. Por lo tanto, si un par de variables está fuertemente correlacionada con el resto, la correlación parcial debe ser pequeña, ya que implica que buena parte de la correlación entre variables puede ser explicada por las otras variables en el análisis. Por el contrario si esta correlación parcial es grande, implica que esta variables tienen poca correlación con el resto, lo que significa una estructura de correlación débil entre el conjunto, y hace cuestionable el análisis factorial. (Zamora, 2016)

Se considera lo siguiente: (Véliz, 2016)

Si $KMO < 0.5$ la aplicación del análisis es inadecuado.

Si $0.5 \leq KMO < 0.7$ se acepta medianamente la aplicación.

Si $KMO \geq 0.7$, se acepta la conveniencia de la aplicación.

La prueba de esfericidad de Bartlett se aplica cuando las variables son normales y evalúa si la matriz de correlaciones es significativamente diferente de la matriz de identidad. Si se rechaza la hipótesis de que la matriz de correlaciones es igual a la matriz de identidad, se acepta la adecuación del análisis. (Véliz, 2016)

En la práctica se considera que si el determinante de la matriz de correlaciones o covarianzas es cercano a 0, se acepta la posibilidad del análisis.

2.3 Análisis de Conglomerados

El análisis de conglomerados o taxonomía numérica comprende una serie de técnicas descriptivas del análisis multivariado. Tiene por objeto la clasificación o agrupamiento de individuos u objetos en clases o conglomerados a partir de mediciones realizadas en ellos, de tal manera que dentro de los grupos se reúnan los elementos más homogéneos y que entre los grupos existe la mayor heterogeneidad. El resultado es la clasificación de entidades a partir de sus atributos en grupos que no son definidos a priori, sino que son descubiertos por el análisis. (Véliz, 2016)

Las técnicas para formar conglomerados se aplican:

En la formación de jerarquías entre los elementos al usar las similitudes entre ellos. Estas jerarquías se estructuran ubicando a los elementos en grupos que aparecen en diferentes niveles, de tal manera que un grupo que está en un nivel superior contiene a grupos que están en niveles inferiores.

En la formación de particiones de elementos de conjuntos.

En la búsqueda de variables o conceptos que influyen en la agrupación de los elementos o variables.

En la exploración de datos con la finalidad de generar hipótesis.

En la detección de datos “aislados”, que distan significativamente de las otras observaciones (outliers). (Véliz, 2016)

En la formación de particiones de elementos de conjuntos.

En la búsqueda de variables o conceptos que influyan en la agrupación de los elementos o variables.

En la exploración de datos con la finalidad

Representación Matricial

La matriz de datos puede representarse como:

$$Y = \begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{pmatrix} = (y_{(1)}, y_{(2)}, \dots, y_{(p)})$$

Donde y'_1 , es un renglón de observaciones del vector y $y_{(j)}$ es una columna correspondiente a la variable. En general se agrupan las n y'_i s observaciones en g clusters. También es posible agrupar las columnas $y_{(j)}, j = 1, 2, \dots, p$.

Dos enfoques comunes para agrupar los vectores de observaciones son los conglomerados y particiones.

Como preámbulo a este tema es necesario definir lo que es una distancia y las diversas medidas útiles para el análisis de conglomerados. (Alvin, 2002)

Medidas de similaridad o disimilaridad

El análisis de conglomerados intenta identificar los vectores de observaciones que son similares y agruparlos en conglomerados, muchas técnicas utilizan un índice de similitud o proximidad entre cada par de observaciones.

Una medida conveniente de proximidad es la distancia entre dos observaciones. Dado que la distancia aumenta a medida que dos unidades se vuelven más lejanas, esto es en realidad una medida de diferencia. (Alvin, 2002)

Similaridad

Un índice de Similaridad para dos elementos X y Y es un número $s(X, Y)$ que cumple con las siguientes propiedades:

1.- $0 \leq s(X, Y) \leq 1$

2.- $s(X, Y) = s(Y, X)$

3.- $s(X, X) = 1$

Como complemento de un índice de Similaridad se define el índice de disimilaridad $\delta(X, Y)$

Como: $\delta(X, Y) = 1 - s(X, Y)$

A continuación se presentan algunos índices de Similaridad que se construyen a partir de una variable dicotómica:

Índice de Rao, como la proporción de copresencias del atributo,

$$s_R = \frac{a}{a + b + c + d}$$

Índice de Sokal y Michener, como la proporción de coincidencias,

$$s_{SM} = \frac{a + b}{a + b + c + d}$$

Índice de Jaccard, como la proporción de apariciones,

$$s_J = \frac{a}{a + b + c}$$

(Véliz, 2016)

Distancia entre dos elementos X y Y

La distancia es un número que indica el alejamiento entre dos elementos. La distancia entre X y Y es un número real $d(X, Y)$ que cumple con las siguientes propiedades:

1.- $d(X, Y) \geq 0$

2.- $d(X, X) = 0$

3.- $d(X, Z) \leq d(X, Y) + d(Y, Z)$ (*desigualdad del triángulo*)

Una distancia es un índice de disimilaridad que a partir de un índice de Similaridad $s(X, Y)$ se puede definir una distancia con:

$$d(X, Y) = \sqrt{2(1 - s(X, Y))}$$

Una distancia comúnmente utilizada es la distancia Euclidiana entre dos vectores. (Alvin, 2002)

Sea $x = (x_1, x_2, \dots, x_p)'$ y $y = (y_1, y_2, \dots, y_p)'$ definida como:

$$d(x, y) = \sqrt{(x - y)'(x - y)} = \sqrt{\sum_{j=1}^p (x_j - y_j)^2}$$

Para ajustar las diferentes varianzas y covarianzas entre las variables p , puede utilizarse la distancia estadística:

$$d(x, y) = \sqrt{(x - y)'S^{-1}(x - y)}$$

Donde S es la matriz de covarianza de la muestra. Una vez realizada la agrupación S puede calcularse como la matriz de covarianza agrupada dentro del clúster. Si calculamos S en la muestra, habrá distorsión de la varianza y covarianza debido a los grupos en los datos. Por lo tanto suele utilizarse la distancia euclidiana, la distancia descrita también es conocida como la distancia de Mahalanobis.

Otra medida de distancia que es sugerida es la métrica Minkowski definida como:

$$d(x, y) = \left[\sum_{j=1}^p |x_j - y_j|^p \right]^{1/p}$$

Si $p = 2$ se transforma en la distancia euclidiana. Si $p = 1$ tenemos la distancia City block que se define como: (Zamora, 2016)

$$d(X, Y) = \sum_{j=1}^p |x_j - y_j|$$

Si $p = \infty$ se tiene la distancia de Chebychev, dada por:

$$D_{\infty} = \max_{1 \leq j \leq p} |x_{1j} - y_{1j}|$$

La escala de medición de las variables es una consideración importante cuando se utiliza la distancia Euclidiana, cambiar la escala puede afectar las distancias relativas entre los elementos. (Alvin, 2002)

La reducción de la dimensión y la formación de los conglomerados

Los conglomerados pueden formarse al reducir previamente la dimensión mediante el análisis factorial o el de componentes principales. Esto facilita la aplicación de los algoritmos; sin embargo los conglomerados que se forman no son fáciles de interpretar y su formación puede ser influenciada por los primeros componentes que explican la mayor variabilidad. La ventaja que ofrecen los componentes aparece cuando su número se reduce a un número tal en el que pueden representarse gráficamente los datos y las variables. (Véliz, 2016)

Técnicas de conglomerados

Algunas técnicas para formar conglomerados son las siguientes:

Técnica jerárquica aglomerativa, con la que se forman grupos sucesivos partiendo de tantos grupos como elementos que se tengan hasta formar un único grupo con todos los elementos.

Técnica no jerárquica, con los que se forma un número preestablecido de grupos. Entre los métodos no jerárquicos destaca el método de k medias que permite la formación de un número k de grupos previamente determinado.

2.3.4 Técnica jerárquica aglomerativa

Los métodos jerárquicos y otros algoritmos de conglomerados representan un buen intento en la agrupación de datos utilizando técnicas computacionales eficientes. Por lo general no es factible examinar todas las posibilidades de agrupación (en conglomerados) para un conjunto de datos.

El número de formas de particionar un conjunto de “n” elementos en “g” (variables) está dado por:

$$N(n, g) = \frac{1}{g!} \sum_{k=1}^g \binom{g}{k} (-1)^{g-k} k^n$$

Esto puede ser aproximado por $g^n/g!$. Por ejemplo, $N(25,10) \cong 2.8 \times 10^{18}$ son el total de posibles combinaciones de conglomerados para un conjunto de 25 elementos y 10 variables. Por lo tanto los métodos jerárquicos y otros enfoques nos permiten buscar una solución razonable sin tener que examinar todos los arreglos posibles.

Los algoritmos jerárquicos de conglomerados implican un proceso; En cada paso de la técnica jerárquica, una observación o un grupo de observaciones se agrupan (o fusionan) en otro grupo o conglomerado. En este proceso, el número de conglomerados se reduce y los propios conglomerados se hacen más grandes. Comenzamos con n conglomerados (elementos individuales) y terminamos con un solo conglomerado que contiene todos los datos.

Un enfoque alternativo, es conocido como método divisivo, comienza con un solo conglomerado (o grupo) que contiene todos los “n” elementos y particiona un grupo en dos grupos en cada paso. El resultado final de este método es “n” conglomerados de un elemento cada uno. Los métodos aglomerantes se utilizan más comúnmente que los métodos divisivos. En cualquiera de los dos tipos de conglomerados con técnicas jerárquicas se debe tomar una decisión en cuanto al número óptimo de conglomerados.

En cada paso de un proceso jerárquico de aglomeración, los dos grupos más cercanos se fusionan formando un nuevo y único conglomerado. Por lo tanto el proceso es irreversible en el sentido que los dos elementos que una vez se agrupan en un conglomerado no se pueden separar posteriormente en el procedimiento, cualquier error temprano no puede ser corregido más tarde. Del mismo modo, en el método jerárquico divisivo, los elementos no se pueden mover a otros conglomerados. Un método opcional es llevar a cabo un procedimiento jerárquico seguido de un procedimiento de partición en el que los elementos se puedan mover de un conglomerado a otro.

Dado que un procedimiento jerárquico de aglomeración combina los dos grupos (o conglomerados) más cercanos en cada paso, se debe considerar la cuestión de medir la similitud o diferencia de dos grupos. Diferentes enfoques para medir la distancia entre grupos dan lugar a diferentes métodos jerárquicos. A continuación se describen estos métodos. (Alvin, 2002)

Métodos de la técnica jerárquica

Vecino más cercano o Ligadura simple

En este método la distancia entre dos clústeres A y B se define como la distancia mínima entre el punto A y el punto B.

$$D(A, B) = \min\{d(y_i, y_j), \text{para } y_i \text{ en } A \text{ y } y_j \text{ en } B\}$$

Donde $d(y_i, y_j)$ es la distancia Euclidiana o alguna otra distancia entre los vectores y_i, y_j . Este método también se denomina "Ligadura simple". En cada paso del método, la distancia es encontrada para cada par de clústeres y los dos clústeres con menor distancia se combinan.

El número de los grupos se reduce en uno, después de combinar dos conglomerados, el procedimiento se repite para el siguiente paso: las distancias entre todos los pares de conglomerados se calcula de nuevo, y el par con la distancia mínima se combina en un solo conglomerado.

El resultado del proceso jerárquico de conglomerado se proyecta gráficamente en un diagrama de árbol que es conocido como dendrograma, el cual muestra el procedimiento jerárquico incluyendo las distancias a las que se combinan los conglomerados.

Vecino más lejano o Enlace Completo

En el método de vecino más lejano también llamado "ligadura completa", la distancia entre dos conglomerados A y B se define como la distancia máxima entre un punto A y un punto B:

$$D(A, B) = \max\{d(y_i, y_j), \text{para } y_i \text{ en } A \text{ y } y_j \text{ en } B\}$$

En cada paso, la distancia se obtiene para cada par de agrupaciones, y los dos conglomerados con la distancia más pequeña se agrupan.

Agrupación Media o Vinculación media.

En este método la distancia entre dos grupos A y B es el promedio de las distancias n_A y n_B

Entre los puntos n_A de A y de n_B de B:

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j)$$

Donde la suma es sobre todos los y_i en A y todos los y_j en B. En cada paso se unen los dos grupos con la distancia más pequeña, medida por la expresión definida para la distancia $D(A, B)$.

Método del Centroide.

En el método del centroide la distancia entre dos grupos A y B se define como la distancia Euclidiana entre la media de los vectores de ambos grupos.

$$D(A, B) = d(\bar{y}_A, \bar{y}_B)$$

Donde \bar{y}_A y \bar{y}_B son los vectores medios del vector de observaciones en A y el vector de observaciones en B, respectivamente, la $d(\bar{y}_A, \bar{y}_B)$ es definida como la distancia Euclidiana. Se definen \bar{y}_A y \bar{y}_B de manera usual como: $\bar{y}_A = \sum_{i=1}^{n_A} y_i / n_A$. Los dos grupos con la distancia más pequeña entre los centroides se agrupan en cada paso.

Después de que los dos grupos A y B se han agrupado, se obtiene el centroide del nuevo grupo AB por medio del promedio ponderado descrito como:

$$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B}$$

Mediana.

Si se combinan dos grupos A y B utilizando el método del centroide, y si A contiene a un mayor número de elementos que B, el nuevo centroide:

$$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B}$$

Puede estar mucho más cercano a \bar{y}_A que a \bar{y}_B .

Para evitar ponderar los vectores medios de acuerdo con el tamaño del conglomerado podemos utilizar la mediana (Punto medio) de la línea (ligadura) que une A y B como punto para calcular nuevas distancias entre grupos:

$$m_{AB} = \frac{1}{2} (\bar{y}_A + \bar{y}_B)$$

Los dos grupos con la distancia más pequeña entre medianas se agrupan en cada paso. Notesé que la mediana definida arriba no es la mediana ordinaria en el sentido estadístico. La terminología surge de una mediana de un triángulo, al conocer la línea de un vértice hacia el punto medio del lado opuesto.

2.3.5.6 Método Ward's.

El método Ward's, también conocido como incremento de la suma de cuadrados, utiliza la distancia (cuadrada) que se encuentra dentro de los grupos y las distancias (cuadrada) entre los grupos (Ward 1963). Si AB es un grupo obtenido mediante la combinación de los grupos A y B, entonces la suma de las distancias dentro del grupo (de los elementos de los vectores medios del grupo) son:

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)' (y_i - \bar{y}_A)$$
$$SSE_B = \sum_{l=1}^{n_B} (y_l - \bar{y}_B)' (y_l - \bar{y}_B)$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})' (y_i - \bar{y}_{AB})$$

Donde:

$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B}$; es la medida que se definió para nuevos grupos por el método del centroide, y n_A , n_B , $n_{AB} = n_A + n_B$ son el número de puntos en A, B y AB respectivamente.

Dado que estas sumas de distancias son equivalentes a las sumas de cuadrados dentro del grupo se denotan por:

$$SSE_A, SSE_B, SSE_{AB}$$

El método Ward's une los dos grupos A y B que minimiza el incremento en SSE definido como:

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$$

Se puede demostrar el incremento I_{AB} de las siguientes formas:

$$\begin{aligned} I_{AB} &= n_A (\bar{y}_A - \bar{y}_{AB})' (\bar{y}_A - \bar{y}_{AB}) + n_B (\bar{y}_B - \bar{y}_{AB})' (\bar{y}_B - \bar{y}_{AB}) \\ &= \frac{n_A n_B}{n_A + n_B} (\bar{y}_A - \bar{y}_B)' (\bar{y}_A - \bar{y}_B) \end{aligned}$$

Por lo tanto al minimizar el aumento de SSE equivale a minimizar la distancia entre los grupos. Si solamente A consiste en y_i y B solamente consiste en y_j , entonces SSE_A y SSE_B son cero y se reduce a:

$$I_{ij} = SSE_{AB} = \frac{1}{2} (y_i - y_j)' (y_i - y_j) = \frac{1}{2} d^2(y_i, y_j)$$

El método de Ward está relacionado con el método del centroide. Si la distancia $d(\bar{y}_A, \bar{y}_B)$ es cuadrada en comparación con I_{ij} , la única diferencia es el coeficiente $n_A n_B / (n_A + n_B)$ del método de Ward's. Por lo tanto el tamaño de las agrupaciones tiene un impacto en el método de Ward, pero no en el método del Centroide. Escrito $n_A n_B / (n_A + n_B)$ en I_{AB} como:

$$\frac{n_A n_B}{n_A + n_B} = \frac{1}{1/n_A + 1/n_B}$$

Se observa que si n_A y n_B incrementan, $n_A n_B / (n_A + n_B)$ incrementa. Describimos el coeficiente como:

$$\frac{n_A n_B}{n_A + n_B} = \frac{1}{1 + n_A/n_B}$$

Vemos que a medida que n_B aumenta con n_A fijo, $n_A n_B / (n_A + n_B)$ aumenta. Por lo tanto en comparación con el método del centroide, el método Ward's tiende a unir grupos más pequeños o grupos de igual tamaño.

Los valores y parámetros requeridos se indican en la siguiente tabla. Para los métodos de centroide, mediana y Ward.

Métodos de Clústeres	α_A	α_B	γ
Ligadura simple (Vecino cercano)	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
Ligadura completa (Vecino lejano)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Ligadura media	$\frac{n_A}{n_A + n_B}$	$\frac{n_B}{n_A + n_B}$	0
Centroide	$\frac{n_A}{n_A + n_B}$	$\frac{n_B}{n_A + n_B}$	0
Mediana	$\frac{1}{2}$	$\frac{1}{2}$	0
Método Ward's	$\frac{n_A + n_C}{n_A + n_B + n_C}$	$\frac{n_B + n_C}{n_A + n_B + n_C}$	0

Tabla 2.2 Valores para los parámetros de los métodos Centroide Mediana y Ward. Fuente: Alvin, 2002

Se ilustrará la elección de los valores de parámetro en el cuadro (anterior) para el método del vecino más cercano (ligadura simple).

Usando $\alpha_A = \alpha_B = \frac{1}{2}$, $\beta = 0$ y $\gamma = -\frac{1}{2}$, como en la primera fila del cuadro la distancia $D(C, AB)$

Es la siguiente:

$$D(C, AB) = \frac{1}{2}D(C, A) + \frac{1}{2}D(C, B) - \frac{1}{2}|D(C, A) - D(C, B)|.$$

SI $D(C, A) > D(C, B)$, entonces $|D(C, A) - D(C, B)| = D(C, A) - D(C, B)$, por lo que $D(C, AB)$ se reduce a

$$D(C, AB) = D(C, B).$$

Por otro lado, si $D(C, A) < D(C, B)$, entonces $|D(C, A) - D(C, B)| = D(C, B) - D(C, A)$, por lo que (C, AB) se reduce a

$$D(C, AB) = D(C, A).$$

Así podemos escribir:

$$D(C, AB) = \min[D(C, A), D(C, B)],$$

Como el equivalente a la definición de distancia para el método del vecino más cercano o ligadura simple. (Alvin, 2002)

Propiedades de los Métodos Jerárquicos

Monotonía (Monotonicidad)

Si un elemento o un grupo se unen a otro grupo a una distancia menor que la distancia previa antes de unir dos grupos, se dice que ha ocurrido una inversión o reversión. La inversión está representada por un cruce en el dendrograma.

Un método jerárquico en donde no se pueden producir reversiones, es llamado monotónico, porque la distancia en cada paso es mayor que la distancia en el paso anterior. A la medida de distancia o el método de agrupación en conglomerados siendo monotónico se le denomina “ultra métrico”.

Los métodos de vinculación única y de vinculación completa son monótonos. Sea d_k la distancia que une dos grupos en el paso $k_t + h$.

Podemos describir los pasos k y $k+1$ en términos de cuatro grupos A, B, C y D. Supongamos $D(A,B)$ es menor que la distancia entre cualquier otro par entre estos cuatro grupos, de modo que A y B se unen en el paso k para formar AB.

Entonces:

$$d_{k+1} = D(C, AB) = \min\{D(A, C), D(B, C)\} > d_k = D(A, B)$$

En d_k , $D(A, C)$ y $D(B, C)$ exceden a $D(A, B)$, esto pasa también en la ligadura completa. Por lo tanto, los métodos de ligadura simple y ligadura completa son monótonos.

Para los métodos del cuadro (tal) que son distintos de la ligadura simple y completa, se toma a $\gamma = 0$ entonces por:

$$D(C, AB) = \alpha_A D(C, A) + \alpha_B D(C, B) + \beta D(A, B) + \gamma |D(C, A) - D(C, B)|$$

Y

$$d_k = D(A, B) < \min\{D(A, C), D(B, C), D(C, D)\}$$

Tenemos:

$$D(C, AB) > (\alpha_A + \alpha_B + \beta)D(A, B).$$

Por lo tanto, se necesita $\alpha_A + \alpha_B + \beta \geq 1$ para la monotonía. Usando este criterio vemos que todos los métodos de la tabla 2.2 (más allá de los dos primeros) son monótonos, excepto el centroide y los métodos de mediana. (Estos dos métodos muestran cruces en el dendrograma). Debido a la falta de monotonía algunos autores no recomiendan los métodos de centroide y mediana. (Alvin, 2002)

Contracción y dilatación

Ahora consideraremos las características de las distancias o proximidades entre los puntos originales. A medida que se forman grupos las propiedades del espacio de distancias pueden ser alteradas. Por lo que es necesario un método de agrupación en conglomerados que no altera las propiedades espaciales, es decir, conserva el espacio, este método fue propuesto por Lance y Williams (1967).

Un método que no conserva el espacio, es decir que no es “espacio – conservador” puede contraer o dilatar el espacio.

Un método de contracción o “espacio-contracción” es cuando el espacio de los grupos recién formados parece acercarse a observaciones individuales, de modo que un elemento individual tiende a unirse a un grupo existente en lugar de unirse con otro elemento individual para formar un nuevo grupo. Esta tendencia es también llamada encadenamiento.

Un método es dilatador del espacio, “espacio-dilatador” si los grupos recién formados parecen alejarse de las observaciones individuales, de modo que los elementos individuales tienden a formar nuevos grupos en lugar de unirse a los grupos existentes. (En este caso los conglomerados parecen ser más distintos de lo que son).

Dubien y Warde (1979) describieron las propiedades espaciales de la siguiente manera.

Suponga que las distancias entre tres grupos satisfacen:

$$D(A, B) < D(A, C) < D(B, C)$$

Entonces un método de conglomerados es de conservación - espacio si:

$$D(A, C) < D(AB, C) < D(B, C)$$

Un método es de contracción – espacio si la primera desigualdad no se mantiene y es de dilatación – espacio si la segunda desigualdad no se cumple.

El método de la ligadura simple es un método de contracción – espacio con tendencia a hacer encadenamientos (Por lo que algunos autores no recomiendan la ligadura simple). La ligadura completa es un método espacio – dilatador con una tendencia a imponer límites de grupo.

Otros métodos jerárquicos se encuentran entre los extremos representados por el de la ligadura simple y ligadura completa. El método del Centroide y el de la ligadura media, son métodos principalmente de contracción de espacio, mientras que el método de Ward’s es también de contracción del espacio.

Cada vez que un método produce reversiones para un conjunto de datos determinado, se puede considerar como un método de contracción del espacio, por ejemplo el método del Centroide conserva el espacio a menos que tenga reversiones, con lo que se convierte en contracción del espacio.

Otras propiedades

El método de la ligadura simple ha sido criticado por varios autores debido a su tendencia de encadenamiento y su sensibilidad a los errores en las distancias entre observaciones. Este método es mejor que otros métodos para identificar grupos que tienen formas curvas, esféricas o elípticas, y es lo suficiente robusto a los valores atípicos en los datos.

El método Ward’s y el método de la ligadura media son relativamente insensibles a los datos atípicos. Por ejemplo en el método de la ligadura media, los datos atípicos tienden a permanecer aislados en las primeras etapas y unirse con otros valores atípicos en lugar de unirse a grandes grupos o a grupos menos compactos.

Esto se debe a dos propiedades del método de la ligadura media:

La distancia media entre dos grupos (distancia euclidiana cuadrada) los puntos de los grupos más dispersos, es decir están más separados.

La distancia media aumenta a medida que aumenta el tamaño de los grupos.

Estas dos propiedades del método de la ligadura media se muestran en una dimensión, donde el grupo A tiene un punto en z_1 y el grupo B tiene dos puntos b_1, b_2 , situados en $z_2 - h$ y $z_2 + h$

La distancia cuadrada media entre A y B es:

$$\begin{aligned}\overline{d^2} &= \frac{1}{2} [(z_1 - z_2 + h)^2 + (z_1 - z_2 - h)^2] \\ &= \frac{1}{2} [(z_1 - z_2)^2 + h^2 + 2h(z_1 - z_2) + (z_1 - z_2)^2 + h^2 - 2h(z_1 - z_2)] \\ &= (z_1 - z_2)^2 + h^2\end{aligned}$$

Por lo tanto la distancia media entre A y B aumenta a medida que la propagación b_1 y b_2 aumenta (es decir, a medida que aumenta h).

El método del Centroide es bastante robusto para los valores atípicos, el enlace completo es algo sensible a los valores atípicos y tiende a producir grupos del mismo tamaño y forma. El método Ward's tiende a producir grupos esféricos del mismo tamaño.

Muchos estudios concluyen en general que los mejores métodos son el método Ward's y el método de ligadura media. Sin embargo, parece haber una interacción entre métodos y conjuntos de datos; es decir, algunos métodos funcionan mejor para ciertos conjuntos de datos y otros métodos funcionan mejor para otros conjuntos de datos. (Alvin, 2002)

Una buena estrategia es probar varios métodos. Si los resultados están de acuerdo en cierta medida, es posible que se puedan haber encontrados algunos grupos naturales en los datos.

Métodos divisivos.

Los métodos jerárquicos de conglomerados hasta este punto tratados, comienzan con "n" elementos y terminan con un único conglomerado que contiene todos los "n" elementos. Un método jerárquico divisivo comienza con un conglomerado único de "n" elementos y lo divide en dos grupos. En cada paso a partir de entonces, uno de los grupos se divide en dos subgrupos. El resultado final de un algoritmo divisivo es "n" conglomerados de un elemento cada uno. Los resultados se pueden mostrar en un dendrograma.

En los métodos divisivos se presenta el mismo inconveniente que se presenta en los métodos tratados de conglomerados, es el de que una vez que se realiza una partición, un elemento no se puede mover a otro grupo al que no pertenece en el momento de la partición. Sin embargo, si los grupos más grandes son de interés, el método divisivo puede ser preferido en el proceso aglomerativo, en el que los grupos más grandes se alcanzan sólo después de un gran número de unión de grupos más pequeños.

Los algoritmos divisivos son generalmente de dos clases: monotéticos y politéticos. En los monotéticos la división de un grupo es en dos subgrupos se basa en una variable, mientras que el politéticos utiliza todas las variables para hacer la división.

Si las variables son binarias (las variables cuantitativas se pueden convertir en variables binarias), el enfoque monotético se puede aplicar fácilmente. La división entre dos grupos se basa en la presencia o ausencia de un atributo. (La variable (atributo) que maximiza una estadística ji cuadrada o una estadística de información.)

Para el enfoque monotético utilizando una variable y , buscamos maximizar la suma entre grupos de cuadrados.

$$SSB = n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2$$

Donde n_1 y n_2 son los tamaños del grupo (con $n_1 + n_2 = n$), \bar{y}_1 y \bar{y}_2 son los grupos de medias y \bar{y} es la media total basada en todas las n observaciones.

La suma de cuadrados SSB puede ser para todas las posibles particiones dentro de dos grupos de tamaño n_1 y n_2 y para cada una de las p variables. La división final debe basarse en las variables que maximiza $SSB / \sum_{i=1}^n (y_i - \bar{y})^2$.

Para el enfoque Politéticos, consideraremos la técnica propuesta por MacNaughton Smith (1964). Para dividir un grupo trabajamos con una división del grupo y el resto. Buscamos el elemento restante cuya distancia promedio (disimilaridad) desde otros elementos en los residuos, menos la distancia promedio de los elementos del grupo dividido es el más grande.

Si la mayor diferencia es positiva, el elemento se desplaza al fragmento del grupo. Si la diferencia más grande es negativa, el procedimiento se detiene y la división es completa.

Podemos iniciar la división del grupo con los elementos que tienen la mayor distancia promedio de los otros elementos del grupo. (Alvin, 2002)

Algoritmo correspondiente a métodos jerárquicos

El algoritmo que corresponde a esta técnica es el siguiente:

Pasos:

- 1.- Se parte de tantos conglomerados como elementos existan.
- 2.- Se calculan las distancias entre los conglomerados iniciales.
- 3.- Con los dos conglomerados más próximos se forma un nuevo grupo.
- 4.- Con los nuevos elementos se procede como en los pasos 2 y 3 hasta obtener un solo grupo formado con todos los elementos.

En el procedimiento para aplicar el paso 4 deberá elegirse una de las distancias entre conglomerados: del vecino más cercano, del vecino más lejano, de los promedios, etc.

Los pasos que se siguen para llevar a cabo la partición de los elementos se representa mediante un dendrograma.

La elección de la distancia dependerá de la naturaleza de los datos.

Desventaja de los métodos jerárquicos.

- 1.- No se puede corregir la mala ubicación temprana de los elementos, pues las etapas que se realizan no se vuelven a repetir.
- 2.- Tienen poca estabilidad. Las soluciones varían significativamente cuando se quitan algunos atributos o cuando se eligen diferentes distancias.
- 3.- Los datos atípicos influyen sensiblemente en los resultados

Métodos no Jerárquicos.

Los métodos no jerárquicos a tratar son los siguientes: Partición, Distribuciones mixtas y Densidad estimada.

Partición

En el enfoque de partición las observaciones son separadas dentro de g conglomerados sin usar un proceso o enfoque jerárquico basado en una matriz de distancias o similitudes entre todos los pares de puntos.

Los métodos que se describen en esta sección son también conocidos como métodos de optimización en lugar de particiones.

Una estrategia atractiva que permita examinar todas las posibles formas de particionar n elementos dentro de g grupos y encontrar el grupo óptimo de acuerdo con algún criterio. Sin embargo el número de posibles particiones es dado por $N(n,g)$ es prohibitivamente grande incluso para valores moderados de n y g . Por lo tanto buscamos técnicas más simples. (Alvin, 2002)

Ahora consideraremos un proceso o enfoque para particiones que usualmente es llamado **“K - medias”**.

Este método permite que los elementos se muevan de un conglomerado a otro o reasignarlos, que no es posible en los métodos jerárquicos.

Primero seleccionaremos g elementos que servirán como semilla. Estos serán remplazados por los centroides de los conglomerados (Vectores medios).

Hay varias formas para seleccionar las semillas: Eligiendo g elementos aleatoriamente (quizás separados por una distancia mínima especificada), eligiendo los primeros puntos g en el conjunto de datos (de nuevo sujetos a un requisito de distancia mínima), Seleccionar los g puntos que están más apartados, encontrar los puntos g con mayor densidad o especificar g puntos espaciados en un patrón de cuadrícula.

Para estos métodos de selección de semillas, se debe especificar el número de conglomerados, g . Alternativamente, se puede especificar una distancia mínima entre las semillas, y luego todos elementos que satisfacen este criterio se eligen como semillas.

Una vez elegidas las semillas, cada punto restante del conjunto de datos se asigna a la semilla más cercana (basada en la distancia Euclidiana). Tan pronto como un grupo tiene más de un miembro, la semilla del conglomerado es remplazada por el centroide.

Después de asignar todos los elementos a los conglomerados, cada elemento se examina para ver si está más cerca al centroide de otro cúmulo que al centroide de su propio cúmulo. Si es así, el elemento se mueve al nuevo conglomerado y se actualizan los dos centroides del conglomerado. Este proceso sigue hasta que no sea posible seguir mejorando.

El procedimiento “k - medias” es algo sensible a la elección inicial de semillas, puede ser aconsejable intentar el procedimiento de nuevo con otra opción de semillas. Si diferentes opciones iniciales de semillas producen grupos finales muy diferentes, o si la convergencia es extremadamente lenta, puede que no haya grupos naturales en los datos.

El método de partición de las “k-medias” se puede utilizar como una posible mejora a las técnicas jerárquicas. Primero agrupamos los elementos usando un método jerárquico y a

continuación, utilizar los centroides de estos grupos como semillas para el método “K-medias”, que permiten que los puntos se reasignan de un grupo a otro. (Alvin, 2002)

Técnicas no jerárquicas: Método de las K medias

Las técnicas no jerárquicas permiten la formación de conglomerados cuyo número K es previamente fijado. Un método particular es el método de la K medias. (Véliz, 2016)

Desarrollo del Proceso:

Se tiene una muestra de N elementos para los cuales están definidas p variables numéricas.

1.- Se forman al azar K grupos y para cada grupo se calculan los centroides o puntos cuyas coordenadas son la medias aritméticas de las p variables.

2.- Usando la distancia euclidiana y para cada elemento se calcula su distancia a cada uno de los centroides, reasignándoles al grupo cuyo centroide el más cercano. Los nuevos centroides de los nuevos grupos formados se recalculan.

3.- Si la distancia entre los centroides iniciales y los nuevos centroides es pequeña o si se ha completado un número fijo de iteraciones, el proceso termina. De otro modo se repite el paso 2.

Al aplicar el método de las K – medias para determinar el número K, se acostumbra usar el índice definido por:

$$F = \frac{B(K) - B(K + 1)}{B(K + 1)/(n - K - 1)}$$

En donde:

$$B(K) = \sum_{k=1}^K \sum_{j=1}^p \sum_{i=1}^{n_k} (x_{ijk} - \bar{x}_{jk})^2$$

K: Es el número de conglomerados

n_k : Es el número de elementos en el k - ésimo conglomerado

x_{ijk} : Es el valor que le corresponde al i – ésimo elemento en la variable j en el k – ésimo conglomerado.

Este índice es llamado pseudo F, expresa la dispersión dentro de los conglomerados para todas las variables.

Si el índice de F para K conglomerados es mayor que 10, se deberá considerar un conglomerado más.

Mezclas de distribuciones

En este método, se asume la existencia de g distribuciones, (usualmente una normal multivariada) y se desea asignar cada uno de los “n” elementos de la muestra a la distribución que más probablemente pertenece. Este método tiene la propiedad de que los puntos se pueden transferir de un conglomerado a otro. (Alvin, 2002)

Definimos la densidad de una mezcla (combinación) de distribuciones “g” como el promedio ponderado

$$h(y) = \sum_{i=1}^g \alpha_i f(y, \mu_i, \Sigma_i)$$

Donde $0 \leq \alpha_i \leq 1, \sum_{i=1}^g \alpha_i = 1$ y $f(y, \mu_i, \Sigma_i)$ es la distribución normal multivariada $N_p(\mu_i, \Sigma_i)$.

Los conglomerados se pueden formar de dos maneras o formas. El primer enfoque consiste en asignar un elemento con vector de observación al conglomerado C_i con mayor valor de estimación de probabilidad.

$$\hat{P}(C_i|y) = \frac{\hat{\alpha}_i f(y, \hat{\mu}_i, \hat{\Sigma}_i)}{h(y)}$$

Donde $\hat{\alpha}_i, \hat{\mu}_i, \hat{\Sigma}_i$ son los estimadores máximos de la función de probabilidad, $h(y)$ se definió anteriormente como la densidad de la combinación con estimaciones insertadas para los parámetros.

La probabilidad descrita como $\hat{P}(C_i|y)$ es la probabilidad estimada de que un elemento con el vector de observación pertenece al conglomerado i -ésimo de C_i . El segundo enfoque consiste en asignar un elemento con el vector de observación al conglomerado con el mayor valor de

$$\ln \hat{\alpha}_i - \frac{1}{2} \ln |\hat{\Sigma}_i| - \frac{1}{2} (y - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (y - \hat{\mu}_i)$$

Para cualquiera de estos enfoques, necesitamos los estimadores $\hat{\alpha}_i, \hat{\mu}_i, \hat{\Sigma}_i$, estos estimadores se obtienen maximizando la función:

$$L = \prod_{j=1}^n h(y)$$

Los resultados son:

$$\hat{\alpha}_i = \frac{1}{n} \sum_{j=1}^n \hat{P}(C_i|y_j), \quad i = 1, 2, \dots, g-1$$

$$\hat{\mu}_i = \frac{1}{n \hat{\alpha}_i} \sum_{j=1}^n y_j \hat{P}(C_i|y_j), \quad i = 1, 2, \dots, g$$

$$\hat{\Sigma}_i = \sum_{j=1}^n (y_j - \hat{\mu}_i)(y_j - \hat{\mu}_i)' \hat{P}(C_i|y_j), \quad i = 1, 2, \dots, g$$

En estas ecuaciones \hat{P} denota la probabilidad posterior estimada de una observación y_j ; perteneciente al componente C_i , es decir,

$$\hat{P}((C_i|y_j) = \frac{\hat{\alpha}_i \text{MVN}(y_j; \hat{\mu}_i, \hat{\Sigma}_i)}{\sum_{j=1}^g p_j \text{MVN}(y_j; \hat{\mu}_j, \hat{\Sigma}_j)}$$

Dónde: MVN es Mezcla Normal Multivariada

Escrito en esta forma se puede ver que las ecuaciones de máxima verosimilitud para los parámetros en una mezcla de normales multivariadas son cercanas análogamente a los de una sola distribución normal, excepto que cada punto de muestra ahora está ponderado por la probabilidad posterior estimada de pertenecer a un componente en particular. (Everitt, 2001)

Las tres ecuaciones intermedias se resuelven iterativamente. Para un valor dado de g , puede comenzarse con estimaciones iniciales para los parámetros y ajustarlos por iteración. Si g no se conoce, podemos comenzar con $g=1$ y luego tratar sucesivamente $g=2$, $g=3$, hasta que los resultados sean satisfactorios. (Alvin, 2002)

El número total de parámetros a estimar es grande. Hay p parámetros en cada μ_i , $\frac{1}{2}p(p+1)$ únicos parámetros en cada Σ_i , y $g-1$ valores de α_i (los restantes $\hat{\alpha}_i$ se encuentran por $\sum_{i=1}^g \hat{\alpha}_i = 1$), para un total de

$$\frac{1}{2}g(p+1)(p+2) - 1$$

Parámetros. Si el tamaño de la muestra “ n ” no es lo suficientemente grande como para estimar todos los parámetros, podría asumirse la matriz común de covarianza Σ , lo que reduce el número de parámetros por $\frac{1}{2}(g-1)p(p+1)$. (Alvin, 2002)

El método de las mezclas es invariante (invariable) a transformaciones lineales de rango completo y es algo robusto a la suposición de normalidad. La técnica funciona mucho mejor si las densidad de g están bien separadas o los tamaños de la muestra son grandes. (Alvin, 2002)

Estimación de densidad

En el método de estimación de densidad o búsqueda de densidad, buscamos regiones de densidad a veces llamada “modas” (modes). No se hace ninguna suposición sobre la forma de la densidad, simplemente se intenta separar las regiones con una alta concentración de puntos de regiones con baja densidad.

Para encontrar regiones de alta densidad, primero elegimos un radio “ r ” y un valor de “ k ”, el número de puntos en un esquema de vecino k – más cercano. Para cada uno de los “ n ” puntos de los datos, se encuentra el número de puntos dentro de una esfera (círculo) de radio r . Un punto se llama un punto denso si al menos “ k ” otros puntos están contenidos en su esfera.

Si un punto denso está a (mayor distancia que) distancia r y a mayor distancia de todos los demás puntos densos, este punto es el núcleo de un nuevo cúmulo (o conglomerado). Si un punto denso está a una distancia r de al menos un punto denso que pertenece a un conglomerado, se agrega al conglomerado. Si el punto denso se encuentra dentro de una distancia r de dos o más conglomerados, estos conglomerados se combinan.

Dos grupos también se combinan si la distancia más pequeña entre sus puntos densos es menor que el promedio de las $2k$ distancias más pequeñas entre los “ n ” puntos originales. El valor de r puede aumentar gradualmente para que más puntos se vuelvan densos. Otra opción es comenzar con el valor especificado de r para cada punto y luego aumentar gradualmente r hasta k . Las observaciones están contenidas en sus esferas. (Alvin, 2002)

Elegir el número de clusters.

En la agrupación jerárquica, podemos seleccionar g conglomerados del dendrograma cortado a través de las ramas en un nivel dado de la medida de distancias utilizada por uno de los ejes.

En este caso deseamos determinar el valor de “ g ” que proporciona el mejor ajuste a los datos.

Un enfoque es buscar grandes cambios en las distancias en las que se forman los clústeres. Mojena (1977) propuso una formalización de este procedimiento. (Alvin, 2002)

Elegir el número de grupos dados por la primera etapa del dendrograma en el que:

$$\alpha_j > \bar{\alpha} + ks_{\alpha}, \quad j = 1, 2, \dots, n$$

Donde $\alpha_1, \alpha_2, \dots, \alpha_n$ son los valores de distancia de las etapas (secciones) con $n, n - 1, \dots, 1$ grupos, $\bar{\alpha}$ y s_{α} son la media y desviación estándar de las α 's, y k es una constante. Mojena (1977) sugirió usar un valor k en un rango de 2.75 a 3.5, pero Milligan y Cooper (1985) recomendaron una $k = 1.25$, sobre la base de un estudio de simulación.

Un índice que se puede utilizar con métodos jerárquicos o de partición es

$$c = \frac{\text{tr}(H)/(g - 1)}{\text{tr}(E)/(n - g)}$$

Dónde: H se define como la matriz cuya diagonal corresponde a la suma de cuadrados y E es la matriz en la que su diagonal corresponde a la suma de cuadrados de los residuales.

Se elige el valor de g que maximiza c . Un enfoque relacionado es elegir el valor de g que minimiza

$$d = g^2 |E|.$$

Para comparar dos soluciones de conglomerados con grupos g_1, g_2 donde $g_2 > g_1$, podemos utilizar la estadística de prueba

$$F = \frac{\text{tr}(E_1) - \text{tr}(E_2)}{\text{tr}(E_2) \left[\left(\frac{n - g_1}{n - g_2} \right) \left(\frac{g_2}{g_1} \right)^{2/p} - 1 \right]}$$

Que tiene una aproximación con la distribución F con $p(g_2 - g_1)$ y $p(n - g_2)$ grados de libertad (Beale (1969)). Las matrices E_1 y E_2 están dentro de las matrices de sumas de cuadrados y productos correspondientes a g_1 y g_2 . La hipótesis es que la solución de conglomerados con g_1 y g_2 son igualmente válidas, si se rechaza la hipótesis implica que la solución con el cluster g_2 es mejor que la solución con g_1 y los conglomerados ($g_2 > g_1$). La aproximación F puede no ser lo suficientemente precisa como para justifica el uso de p -valores. (Alvin, 2002)

Variables de agrupación en clústeres

En algunos casos, puede ser de interés agrupar las variables “ p ” en lugar de las “ n ” observaciones. Para una medida de similitud entre cada par de variables, por lo general se utiliza la correlación.

Puesto que la mayoría de los métodos de agrupación en clusters utilizan diferencias (como las distancias), necesitamos convertir la matriz de correlación $R = (r_{ij})$ en una matriz de diferencias.

Esto se puede hacer se puede hacer reemplazando cada r_{ij} por $1 - |r_{ij}|$ o $1 - r_{ij}^2$.

Utilizando la matriz de diferencias resultante, podemos aplicar un método de agrupación en clusters como una técnica jerárquica para agrupar las variables.

Variables estandarizadas y categóricas

Las medidas de distancia son altamente influenciadas por las unidades de medida de cada variable. Las variables con mayores unidades tienen mayor influencia en la distancia. Para evitar esto se estandarizan previamente las variables quitando las unidades de medida.

Los métodos descritos son válidos cuando las variables son numéricas; cuando las variables son categóricas pueden transformarse en variables cuantitativas usando el análisis de correspondencia. (El análisis de correspondencia es una técnica multivariada análoga al análisis factorial donde las variables que se utilizan son categóricas.) (Véliz, 2016) Pendiente

Validación de los conglomerados

La técnica de agrupamiento utilizada deberá producir conglomerados que internamente sean homogéneos pero heterogéneos entre ellos.

Lo óptimo de esta técnica es que permita:

Producir aglomeraciones significativas que aporten información.

Producir conglomerados que sean insensibles a cambios en alguno de los atributos que se han considerado en el desarrollo de la técnica.

Para verificar la validez de una solución de conglomerados, puede ser posible probar la hipótesis de que no hay conglomerados o grupos en la población de la cual se tomó la muestra en cuestión. (Alvin, 2002)

Por ejemplo, la hipótesis podría ser que la población representa una única distribución unimodal, como la normal multivariada, o que las observaciones surgieron de una distribución uniforme. Gordon revisa las pruebas formales de hipótesis de este tipo relacionadas con la validez del grupo son revisadas por Gordon (1999, Section 7.2) (Alvin, 2002)

Puede ser utilizado un enfoque de validación cruzada para comprobar la validez o estabilidad de un resultado de agrupación en conglomerados. Los datos se dividen aleatoriamente en dos subconjuntos, por ejemplo A y B el análisis del conglomerado se lleva a cabo por separado en cada una de las A y B. Los resultados deben ser similares. Si esto ocurre los conglomerados son válidos. (Véliz, 2016)

Utilice algún método de agrupación en conglomerados para dividir el subconjunto A dentro de grupos. Particiona el subconjunto B en g conglomerados de dos maneras (o formas):

Asigne cada elemento de B al conglomerado A al que esté más cerca utilizando la distancia a los centroides del conglomerado.

Utilice el mismo método de agrupación en conglomerados en B, que utilizó en A.

3.- Compare los resultados de a) y b) en el paso 2.

Interpretación de los resultados

Los conglomerados pueden describirse al analizar los resúmenes numéricos de las variables que se han usados en cada conglomerado construido, esta interpretación se facilita cuando se usan variables auxiliares que no se han utilizado en su construcción “cruzando” las categorías de la variable con los conglomerados si la variable auxiliar es categórica o comparando las medias de la variable en cada conglomerado cuando la variable es numérica. (Véliz, 2016)

Los árboles de clasificación Breiman y otros (1984) son usados para la interpretación de conglomerados. Un árbol de clasificación es una metodología de clasificación supervisada que consta de una variable dependiente categórica y un grupo de variables independientes, permite predecir las categorías de la variable dependiente a partir de las variables independientes. (Véliz, 2016)

Una vez explicada la metodología de estudio y métodos matemáticos, empezaremos con la aplicación de los métodos expuestos a la base de datos generada por encuestas aplicadas a diversos grupos de participantes en las actividades del Departamento de Comunicación Educativa del MNA.

Capítulo 3 Análisis Estadístico Y Multivariado

3.1 Introducción

El contenido de esta sección se divide en cuatro partes. La primera muestra el análisis descriptivo de las variables sociodemográficas de los encuestados, tales como edad, género, nivel educativo, ocupación y alcaldía o estado de residencia; las variables referentes a los canales de información que utilizaron los participantes para conocer las actividades, la frecuencia de participación en las actividades, así como las variables referentes a los motivos de participación, los acompañantes, lo satisfactorio que resultó el evento para los participantes y los horarios de preferencia para asistir a las Noches de Museo. La segunda parte de la sección presenta un análisis sobre las relaciones existentes entre las variables antes mencionadas, las cuales se obtuvieron a través de la matriz de correlaciones, la tercera parte muestra los resultados de la aplicación de los métodos, por último se recomienda un instrumento para la captación de información.

La información contenida en esta sección se obtuvo de los resultados de la encuesta **Talleres Sabatinos 2019 (TS19)** y **Noche de Museos 2019 – 2020 (NM19-20)**, ambas realizadas por el Departamento de Comunicación Educativa del Museo Nacional de Antropología.

3.1.2 Encuesta y base de datos.

La encuesta se formuló como un cuestionario por parte del Departamento de Comunicación Educativa para ser aplicado a los públicos que participaron de dos programas: Noches de Museos (**NM19-20**) y Talleres Sabatinos (**TS-19**). Como parte del personal de servicio social, participé en el diseño y aplicación del instrumento.

La encuesta de Talleres Sabatinos se compuso de trece preguntas las cuales abordaron las siguientes cuestiones y recogieron información de tipo cualitativa:

Datos demográficos de los participantes (Edad, género, escolaridad, ocupación, alcaldía).
Motivos y satisfacciones de los participantes en los Talleres Familiares (Taller, motivo por el que asiste al taller, acompañante, impacto de esta experiencia y satisfacción).
Frecuencia de participación en los talleres y canales de información (primera visita, número de visitas y canal de comunicación).

La encuesta de Noche de Museos se compuso de catorce preguntas las cuales abordaron las siguientes cuestiones que también recogieron información de tipo cualitativa:

Datos demográficos de los participantes (género, edad, escolaridad, ocupación alcaldía).
Motivo de participación, acompañantes y satisfacción con el evento.
Canales de Información, frecuencia de asistencia al evento.
Horarios de preferencia para asistir a las Noches de Museos.

Los encuestados para el caso de los talleres familiares se captaron al azar durante y al finalizar el taller, sumando un total de 138 personas, en tanto que para el caso de los participantes en las Noches de Museos se captaron al inicio de la actividad debido a la dificultad de que el público aceptará quedarse a contestar una vez terminada la visita, de ellos fueron un total de 203 personas.

Con los resultados de las encuestas se formaron dos bases de datos una para cada programa Talleres Sabatinos y Noche de Museos.

Para realizar el procesamiento de datos fue necesario diseñar una escala nominal que categorizó las respuestas dadas y asignó valores numéricos que permitieron obtener una base de datos con variables de tipo numérico. En algunas respuestas se encontró la falta de datos por lo que se asignó el valor NA (No respuesta en inglés No Answer = NA)

Los factores asignados a las respuestas no representan un orden de importancia ni son datos ordenados. De la categorización se derivaron variables de tipo dicotómicas que se definen como: variables que pueden asumir solo dos valores, en este caso para la variable Género, los factores dados fueron Femenino 1, Masculino 2. En la variable Primera Visita SI corresponde a 1 y NO corresponde a 2.

La Edad es una variable de tipo continua, es decir, una variable numérica que tiene un número infinito de valores.

Las variables escolaridad, ocupación, motivo por el que asiste al taller, canal de comunicación, acompañante, tema, procedencia, día que prefiere para participar son de tipo nominal y se asociaron a factores que van del valor 1 al 8.

En la pregunta correspondiente a la Alcaldía de residencia se amplió el rango de respuesta ya que en la Ciudad de México existen 16 alcaldías, además se incluyó el Estado de México como opción, por lo que la escala para esta pregunta fue de 1 a 17.

En el caso para Talleres familiares con la conversión de factores se pudo formar una base de datos numérica, la cual cuenta con trece variables que corresponden a cada una de las preguntas y 138 observaciones que son los participantes.

3.2 Análisis Descriptivo De Variables Sociodemográficas

Los **Talleres Sabatinos** impartidos en los cinco meses ya mencionados, fueron Bordando Historias, Códice Boturini, Códice Dresde, Danza, Estelas de Xochicalco, Instrumentos Musicales y del Día de Niño. En total hubo una participación de 138 personas en los siete talleres. En este programa se encuestó al 100% de los asistentes.

En **Noches de Museos** la encuesta se realizó a partir del mes de marzo de 2019 hasta el mes de febrero de 2020, esta actividad tuvo una participación total de 997 personas, de las cuales se encuestó a 203 participantes, que equivale al 20.36% de los asistentes.

A partir de la encuesta **TS19** se obtuvo que las proporciones muestrales por género fueron 60% mujeres, 28% hombres, 12% no indicaron género. En el caso de la encuesta **NM19-20**, las proporciones muestrales por género fueron 62% mujeres, 35% hombres y 3% no indicó su género.

De acuerdo con la pirámide poblacional para la encuesta **TS19** el rango de edad de los participantes se dio entre 12 a 65 años, los rangos con mayor representación fueron el de 42 a 47 años con 18% de mujeres, y el de 36 a 41 años con 24% de hombres.

En la encuesta de **NM19-20** las proporciones muestrales por género fueron 62% mujeres, 35% hombres y 3% no indicaron su género. Con base a la pirámide poblacional se observa que el rango de edad es de 8 a 82 años, el grupo más amplio de participantes, se da entre los rangos de 22 a 26 años en el que las mujeres representan el 26% de participación, mientras que en el rango de 32 a 36 años los hombres representan el 18% (ver Figura 3.1)¹.

¹ Nota: En el caso de Noche de Museos, el rango de participación comienza de los 12 años en adelante, siendo solo una ocasión en la que participaron niños y es a la que corresponde el rango de 8 a 11 años representada en la gráfica.



Figura 3.1 Distribución de la población por género y edad correspondientes a Talleres Sabatinos y Noche de Museos

En cuanto al nivel educativo los resultados de la encuesta **TS19** muestran que el 7% de la población cuenta con estudios de nivel básico, 19% nivel medio superior, 63% nivel superior y 9% posgrado. Asimismo, para la encuesta de **NM19-20** se observó que el 8% de los encuestados cuentan con estudios de nivel básico, 7% nivel medio superior, 59% nivel superior y 6% otros estudios y posgrado (ver Figura 3.2)².

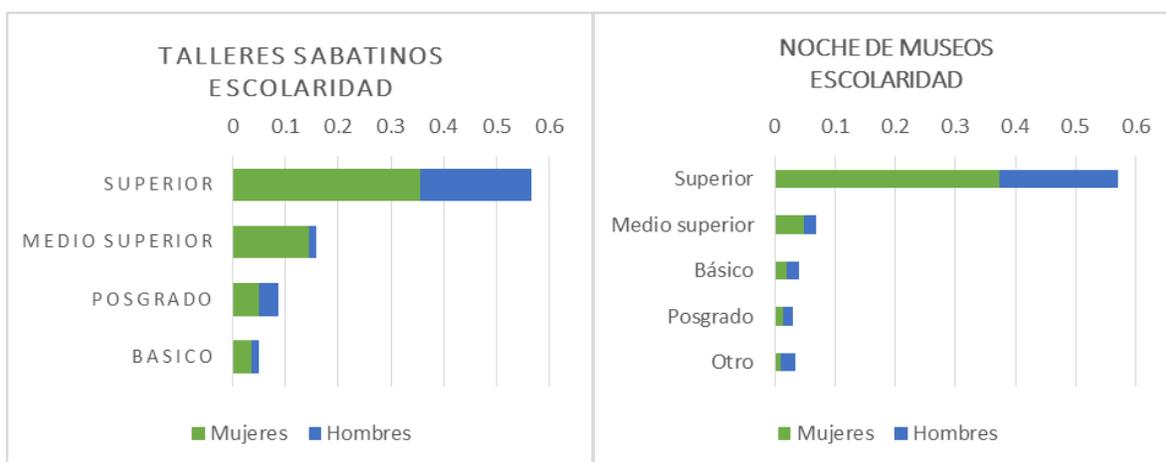


Figura 3.2 Nivel educativo de la población.

² Nota: No se incluye en la gráfica a los participantes que no respondieron su nivel educativo. La categoría de Educación básica incluye Primaria y Secundaria, la de educación Media superior incluye preparatoria, bachillerato; Educación superior incluye licenciatura o profesional, la categoría Otro incluye estudios de nivel técnico y la categoría de posgrado incluye maestría, doctorado, especialidades, etc.

En términos de ocupación los resultados de la encuesta **TS19** indicaron que 49% de los participantes son empleados, 25% estudiantes, 7% jubilados, 24% tiene otra actividad.

En la encuesta de **NM19-20** los resultados fueron los siguientes: 26% son empleados, 14% estudiantes, 1% jubilado, 12% tiene otra actividad (ver Figura 3.3)³.

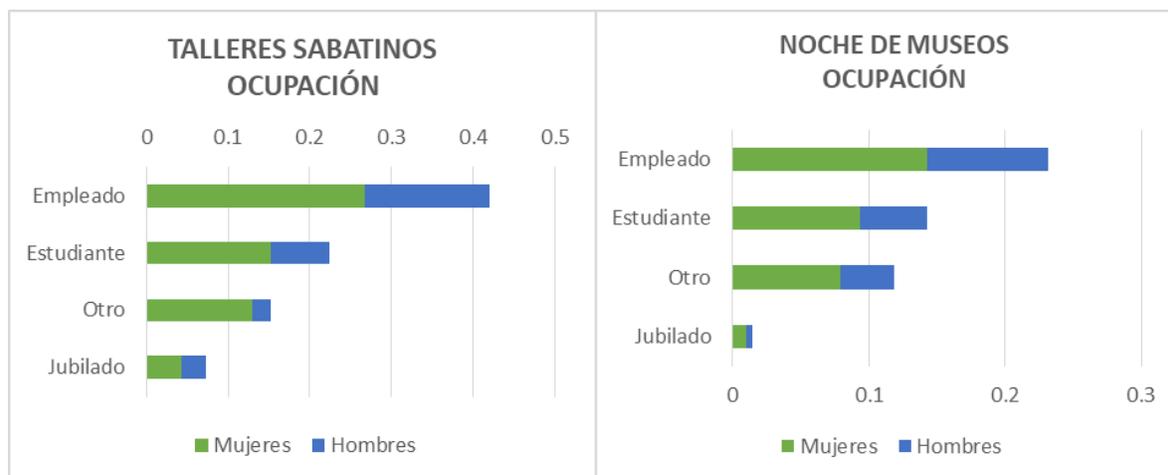


Figura 3.3 Ocupación de los participante encuestados

En lo concerniente a las variables demográficas la última por analizar corresponde a la Alcaldía o estado de residencia. En la encuesta **TS19** se obtuvo que el 14% de participación corresponde a los residentes en la alcaldía Gustavo A. Madero, 7% corresponde a Azcapotzalco y Miguel Hidalgo, mientras que el 12% corresponde al Estado de México. La encuesta **NM19-20**, mostró que 20% de los participantes son residentes del Estado de México, el 8% son residentes de la alcaldía Coyoacán y el 6% corresponde a Cuauhtémoc y Tlalpan.

Análisis Descriptivo De Variables Medios De Información Y Frecuencia De Participación

De acuerdo con los resultados de la encuesta **TS19**, las Redes sociales fue el medio más usado para enterarse de las actividades en el museo, ya que 46% de los participantes las prefirió, seguido del Internet 28%, información a través de familiares 17% y a través de anuncios 7%. Para Noche de museos en la encuesta **NM19-20**, las Redes sociales también fueron las más usadas ya que 31% de los participantes se enteró a través de ellas, seguido del Internet 22% e información por medio de familiares 20% (ver Figura 3.4)⁴.

³ Nota: La actividad de Otro incluye las actividades de comerciante, empresario o cualquier otra actividad que no se relacione con las categorías de empleado, estudiante o jubilado. No se incluye en la gráfica a los que no respondieron sobre su ocupación

⁴ Nota: Las Redes sociales se refieren a Facebook, la categoría de Internet considera a la página del museo o cualquier otro ciberespacio donde fue divulgado el evento, anuncios refiere a volantes que se repartieron entre los visitantes del museo. No se incluye en la gráfica a los que no respondieron.

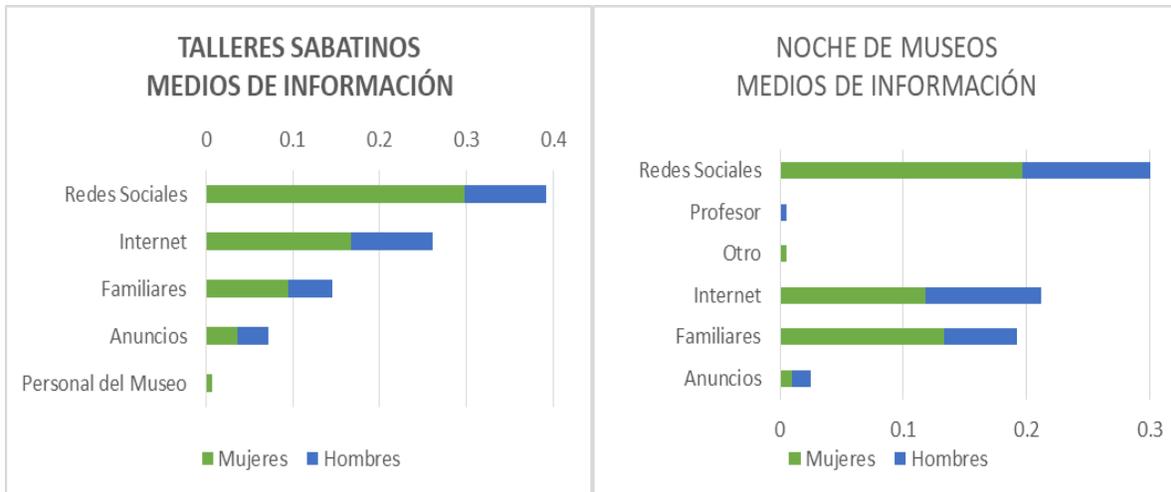


Figura 3.4 Medios de comunicación

La frecuencia de participación indica la asistencia primera o reiterada de los participantes a los eventos, en los resultados de la encuesta de **TS19** se observó que el porcentaje de participantes que asistieron por primera vez fue del 84%, mientras que el porcentaje de quienes asistieron más de una ocasión fue de 16%. En los resultados de la encuesta **NM19-20**, el porcentaje de participantes que asistió por primera vez fue de 79%, para quienes han asistido en otras ocasiones el porcentaje fue de 21% (ver Figura 3.5)⁵.

Análisis descriptivo de las variables tema, motivo de participación acompañante y satisfacción con el evento.

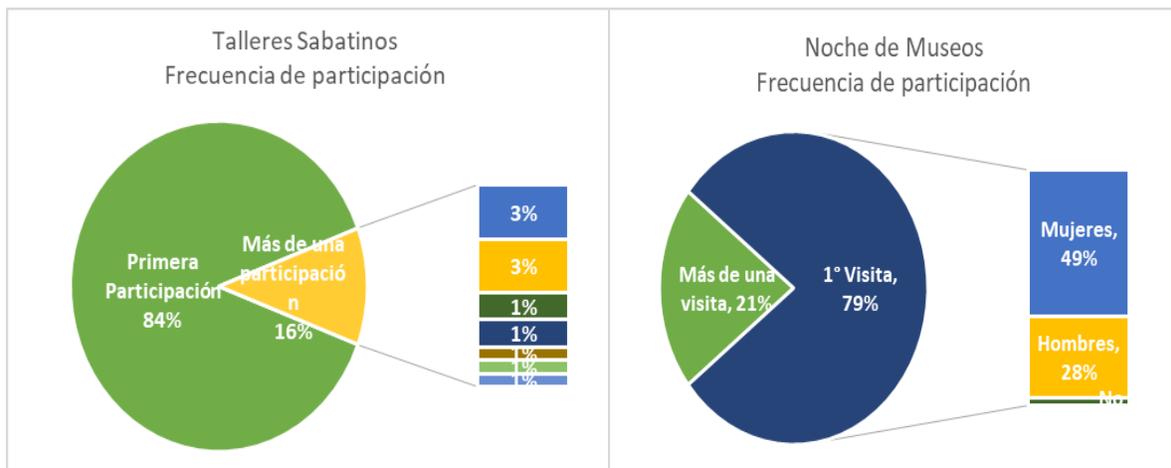


Figura 3.5 Frecuencia de participación en las actividades.

⁵ Nota: No se incluye en la gráfica a los que no contestaron.

Los siguientes resultados corresponden a las variables tema del evento o taller, motivo por el cual se participó en la actividad, quiénes fueron los acompañantes de los participantes y qué tan satisfactorio fue participar en el taller sabatino.

En la encuesta **TS19** la participación por tema tuvo los siguientes resultados: Bordando Historias 2%, Códice Boturini 4%, Códice Dresde 32%, Taller de danza 23%, Estelas de Xochicalco 7%, Instrumentos Musicales 14% y Taller del día del niño 17%. En los resultados de la encuesta **NM19-20**, el tema con mayor participación fue Siempre mujeres con un porcentaje 33%, le sigue Murales con 31% y Guardianes de leyenda con 22%, los temas restantes tuvieron un porcentaje de participación de 9% y 3% (ver Figura 3.6).

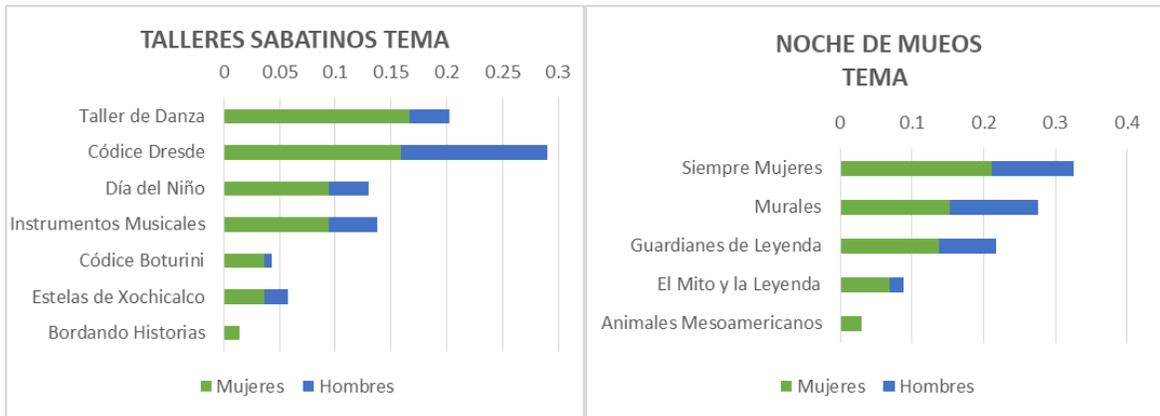


Figura 3.6 Tema de los Talleres Sabatinos y Noche de Museos.

En cuanto al porqué de los participantes para asistir a los talleres sabatinos la encuesta **TS19**, mostró que el aprendizaje fue el motivo con mayor porcentaje 36%, seguido de la convivencia familiar con 25% y el gusto por participar en las actividades con 9%. En la encuesta **NM19-20**, los resultados fueron los siguientes: aprendizaje con 43%, diversión 13% y curiosidad 11% (ver Figura 3.7)⁶.

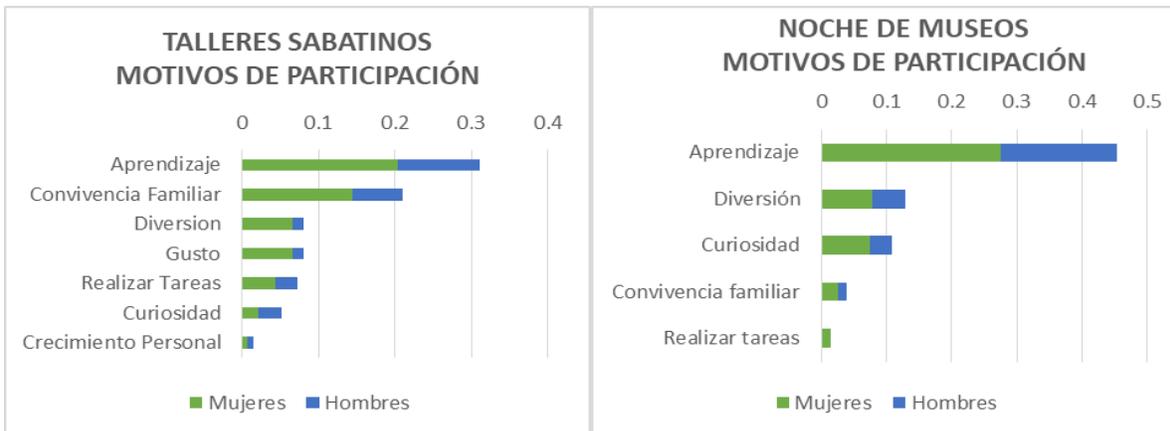


Figura 3.7 Motivos de participación en las actividades.

⁶ Nota: No se incluye a los participantes no respondieron la pregunta.

De acuerdo con los resultados de la encuesta **TS19**, la familia fue el principal acompañante, ya que el 48% de los participantes dijo estar acompañados por ésta, le siguen las personas que asistieron solas 11% y las que asistieron con su pareja 6%. En la encuesta de **NM19-20**, los resultados fueron los siguientes: familia 24%, amigos 22% y pareja 20% (ver Figura 3.8).

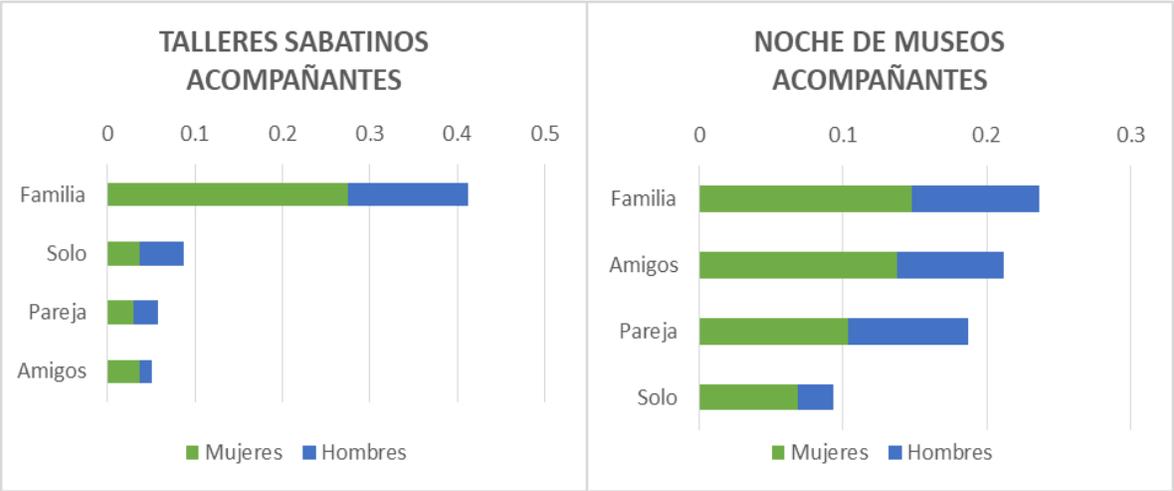


Figura 3.8 Acompañantes de los participantes en las actividades

En la encuesta **TS19**, se preguntó a los participantes en cuál aspecto de su vida habían tenido significado las actividades del taller, los resultados con mayor porcentaje fueron los siguientes: aspecto personal 38% y familiar 20%. En la encuesta de **NM19-20** no se realizó esta pregunta, debido a que la aplicación de la encuesta se realizó antes de comenzar la actividad, por lo que el público no tenía elementos para poder responder (ver Figura 3.9).

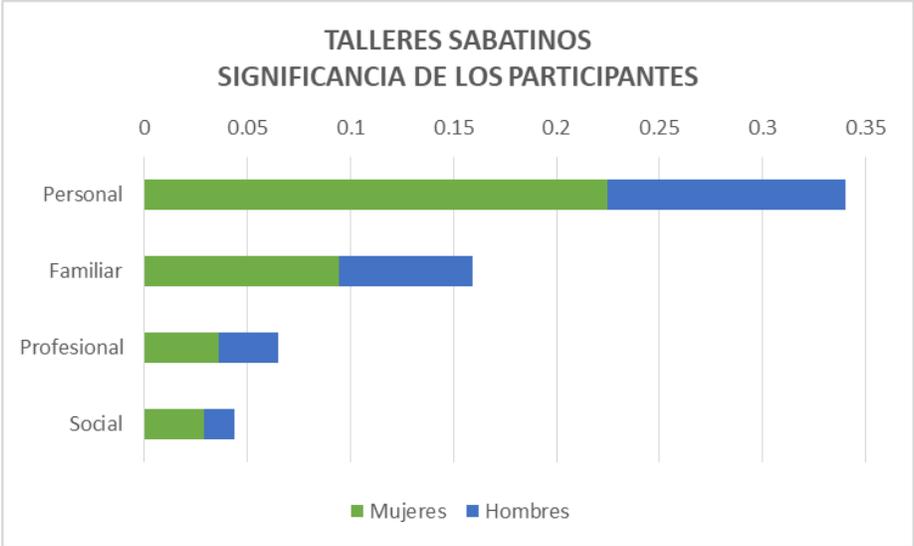


Figura 3.9 Impacto de las actividades en los participantes.

En la encuesta de TS19, se preguntó a los encuestados que indicaran cómo había sido su experiencia en esta actividad, las respuestas obtenidas con mayor porcentaje fueron las siguientes: agradable 45%, y divertida 14% (ver Figura 3.10). A los encuestados de Noche de Museos, no se les preguntó esto, dado que la encuesta se aplicó al iniciar la actividad y por tanto para los participantes no hubiera sido posible la valoración de la actividad.

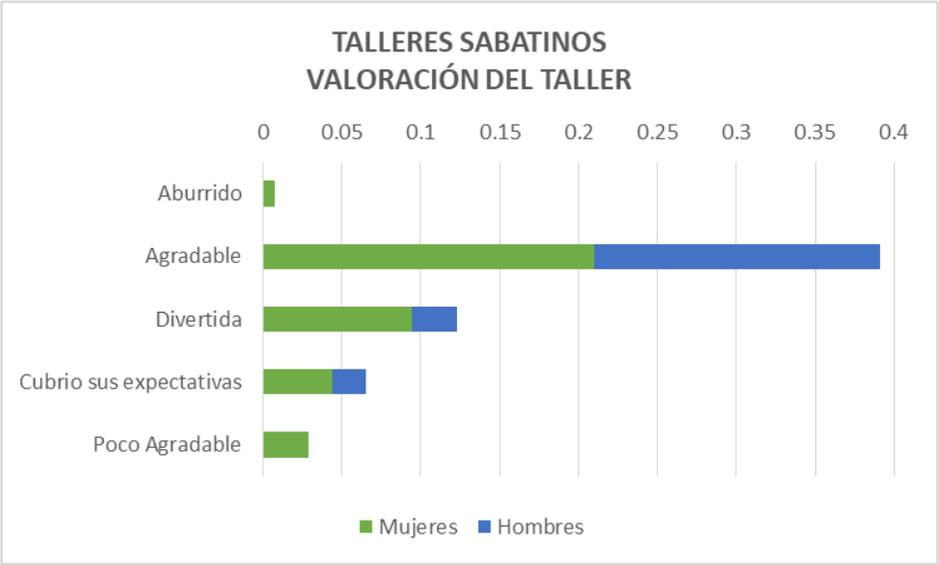


Figura 3.10 Valoración de los talleres por los participantes.

Análisis descriptivo de las variables lugar de procedencia, horario de asistencia, y día de la semana para asistir a la noche de museos.

En esta sección continuamos con el análisis descriptivo de las variables horario, lugar de procedencia y día de la semana que se prefiere para participar en las Noches de Museo. Es importante decir que esta variables solo corresponden a la encuesta de Noche de Museos (NM19-20), pues en la encuesta de Talleres Sabatinos (TS19), no se incluyeron, debido a que estos talleres solo pueden efectuarse el día sábado en el horario establecido, la variable de lugar de procedencia tampoco se aplicó en la encuesta.

Un aspecto importante es conocer desde dónde se trasladan los participantes al museo, los resultados obtenidos de la encuesta **NM19-20**, indicaron que el hogar era el lugar de donde más participantes procedían cuyo porcentaje fue de 33%, le siguen la oficina o lugar de trabajo con 26% (ver Figura 3.11).

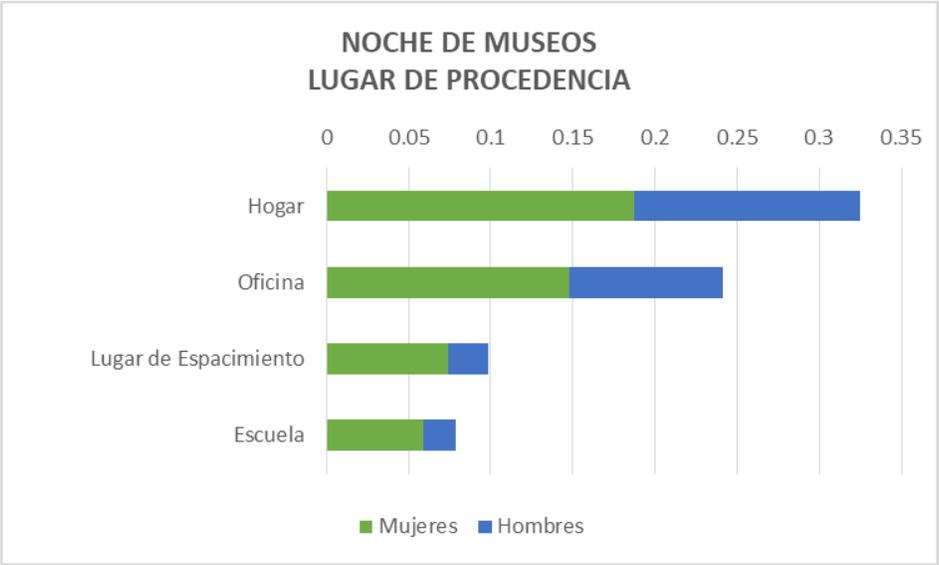


Figura 3.11 Lugar de procedencia de los participantes

El horario de Noche de Museos es extraordinario, ya que no está dentro del horario habitual de operación del Museo, por los que fue interesante conocer el motivo de asistencia de los participantes. Los resultados mayoritariamente que se obtuvieron de la encuesta **NM19-20**, fueron los siguientes: Tiempo del que disponen los participantes 37%, es el único horario que existe para la actividad 25% y el horario permite que los participantes puedan asistir con su familia 6% (ver Figura 3.12).

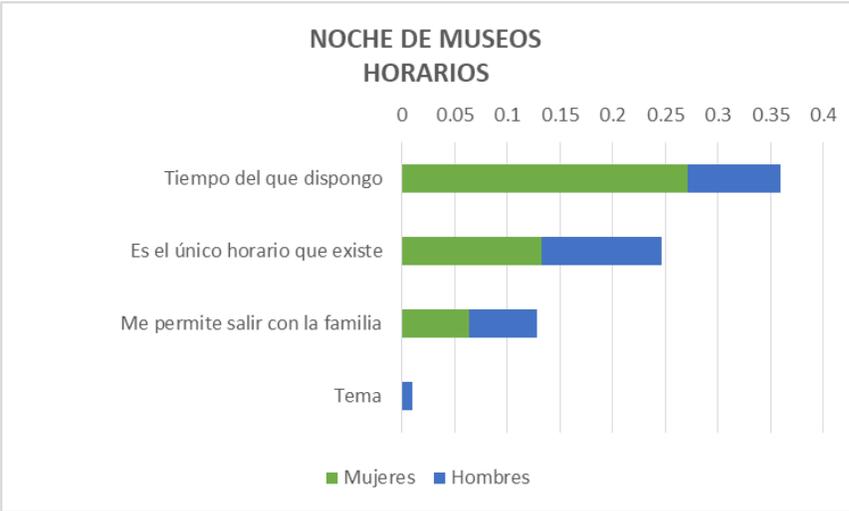


Figura 3.12 Motivos para asistir en el horario de Noche de Museos.

El evento de Noche de Museos se lleva a cabo el último miércoles de cada mes, por lo que es importante conocer si este día es adecuado para el evento o los participantes prefieren otro. Las respuestas obtenidas en la encuesta **NM19-20**, con mayor porcentaje fueron las siguientes: Miércoles 15%, Sábado 13%, y Viernes 10% (ver Figura 3.13).

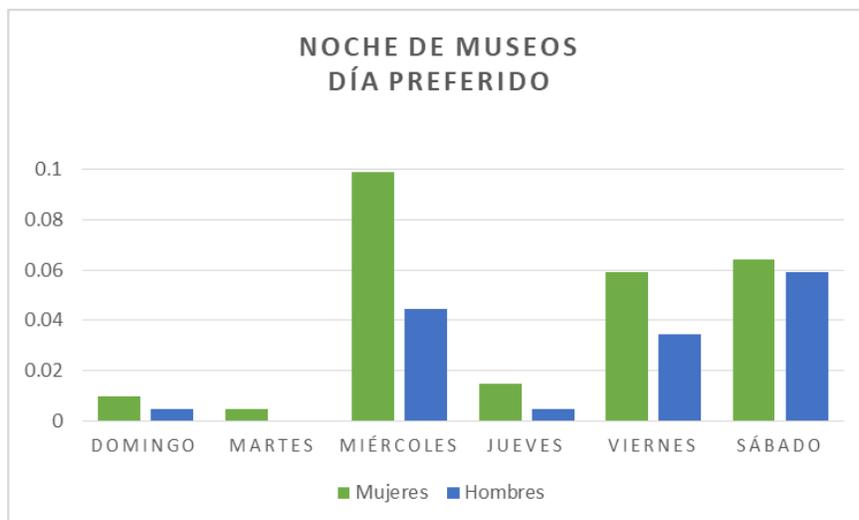


Figura 3.13 Día preferente para asistir a la Noche de Museos.

3.3 Análisis Estadísticos De Correlaciones

La correlación es la existencia de la relación entre las variables que se analizan, en este trabajo el cálculo de la matriz de correlación para TS-19 y MN19-20 se realizó con la correlación correspondiente a la combinación de pares de variables, es decir en base a lo expuesto en el capítulo dos de este trabajo, las correlaciones que se calcularon fueron las siguientes:

Correlación Policórica

- Variables con datos de tipo continuos vs variables con datos de tipo ordinal:
Edad vs Escolaridad, Ocupación
- Variables de tipo ordinal vs variables de tipo dicotómica:
Género vs Escolaridad, Ocupación

Correlación Punto Biserial

- Variable con datos continuos vs variable con datos de tipo dicotómicos:
Edad vs género, primera visita, otros espacios.

Correlación Tetracórica

- Variables con datos dicotómicos vs variables con datos dicotómicos:

Género vs primera visita y otros espacios.

Correlación Poliserial

Esta correlación no fue mencionada en el capítulo dos. Debido a la naturaleza de los datos la utilice para el cálculo de la correlación entre variables con datos que no son del tipo ordinal ni continuo, sino que expresan un factor.

Esta correlación se aplicó a las variables relacionadas a los motivos de asistencia, satisfacción con las actividades, medios de información, horarios y en datos demográficos a la alcaldía. Así también fue utilizada para la variable Edad vs las variables mencionadas.

La matriz de correlación obtenida para Talleres Sabatinos y Noche de Museos se puede ver en anexo I. A continuación se describe la relación encontrada entre las variables.

Relaciones de Talleres Sabatinos

Taller

Los resultados de la matriz de correlación muestran que la variable Taller se relaciona con las variables Número de visitas (NV), Medios de información, nivel de educación y Alcaldía de residencia de los participantes.

La relación de las variables **Taller** y **Número de visitas (NV)** muestra la frecuencia que existe de participación sucesiva o repetida de los asistentes en los diversos talleres.

En la relación de **Taller** y **Medio de información** se observa la preferencia de los asistentes de los diferentes talleres y el medio de información elegido para informarse de las actividades. Se observa que el 63% de los participantes de todos los talleres se informó por las **Redes sociales**, seguido del 28.3% que lo hizo a través del **Internet**. El taller de Códice Dresde fue el que mayor porcentaje mostró en el uso de las Redes sociales; los talleres de danza, Estelas de Xochicalco y del Día del niño, se enterar por medios impresos con un 7.2%; los participantes de los talleres Códice Dresde, Danza, Estelas de Xochicalco, Instrumentos Musicales y Taller del Día del niño se informaron a través de familiares con el 17.4%.

Medio de Información

En la matriz de correlación se obtuvo que la variable **medio de información** se relaciona con las variables: **nivel de escolaridad**, **ocupación** y **taller**. En los datos obtenidos de la encuesta **TS19**, se observó que el 43% de los encuestados con educación superior utilizó las Redes sociales para informarse del evento y un 28% con educación superior se informó por Internet, el 7% restante, que corresponde a los participantes con educación básica se informaron por anuncios, a través de familiares y de internet.

En relación con la variable de **ocupación** se encontró que los participantes con actividad de empleo se informaron por todos los medios aunque mayoritariamente por las Redes sociales cuyo porcentaje fue del 27.5%, seguido de internet 10.9%; los estudiantes se informaron por las Redes sociales con un porcentaje de 10.9%, seguido de internet con 8% y familiares con 6.5%, las personas jubiladas se informaron principalmente por internet 4%, por familiares 2.2% y a través de redes sociales 0.7%.

En la relación con la variable **Taller**, se observa que el 63% de los participantes de todos los talleres se informó por las Redes sociales, seguido del 28.3% que lo hizo a través del

Internet. El taller de Códice Dresde fue el que mayor porcentaje mostró en el uso de las Redes sociales; los talleres de danza, Estelas de Xochicalco y del Día del niño, se enteraron por medios impresos con un 7.2%; los participantes de los talleres Códice Dresde, Danza, Estelas de Xochicalco, Instrumentos Musicales y Taller del Día del niño se informaron a través de familiares con el 17.4%.

Primera visita

De acuerdo con la matriz de correlaciones para la encuesta **TS19**, se obtuvo la relación de la variable **Primera Visita** con las variables: **género** y **ocupación**. En género se observó que el porcentaje de mujeres que asistieron por primera vez a los talleres fue de 48.6% y de varones fue de 23.9%, las personas que no indicaron su género pero asistieron por primera vez tuvieron un porcentaje de 11.6%. En tanto que la mujeres que asistieron más de una vez a los talleres tuvieron un porcentaje del 11.6%, en tanto que los varones que participaron más de una vez tuvieron un porcentaje del 3.6% y los que no indicaron su género y asistieron en otras ocasiones fueron 0.7%. Las mujeres son las que mostraron una mayor frecuencia de participación.

La relación entre **Primera visita** y **ocupación** se dio de la siguiente manera: el porcentaje total de participantes que asistieron por primera vez a los talleres sabatinos fue de 84.1%, de éstos, los participantes cuya actividad es la de **empleado** y participaron por primera vez, el porcentaje fue de 37.7%, el de **estudiantes** fue de 24%, **Jubilados** 5.1% y participantes con **otra actividad** fue de 15.9%. En tanto los participantes cuya ocupación es de empleados y que asistieron en otras ocasiones fue de 11.6%, los **estudiantes** 0.7%, **jubilados** 2.2% y los participantes con **otra actividad** 1.4%. Por lo tanto la participación de las personas con actividad de empleado son las que más veces han participado en los talleres sabatinos, mientras los estudiantes son los que no han participado más de una vez en las actividades.

Acompañante

Las siguientes relaciones que se obtuvieron fue entre la variable **Acompañante** con las variables **Género**, **Motivo** y **Alcaldía**. Con base en los resultados de la encuesta **TS19**, se observó que la relación entre acompañante y género muestra que las **mujeres** que asistieron acompañadas por **familiares** fue el 27.5%, seguido de **amigos** y **solos** con un porcentaje de 3.6% respectivamente y las que asistieron con pareja tuvieron un porcentaje de 2.9%. En el caso de los **hombres**, asistieron con la **familia** el 13.8%, le siguieron, los que asistieron **solos** con el 5.1%, con **pareja** el 2.9% (igual a de las mujeres) y con amigos el 1.4%, Para los participantes que no indicaron su género el 6.5% asistió con la **familia** y el 2.2% **solos**. La familia fue en general con quienes los encuestados asistieron.

Entre las variables **Acompañante** y **Motivo** para participar en las actividades, se encontró con base en los datos de la encuesta que el **aprendizaje** fue la principal razón de asistencia con el 36.2%, seguido de la **convivencia familiar** 24.6% y **gusto** por las actividades 8.7%.

En la relación **Acompañante** y **Alcaldía**, con base en los datos de la encuesta de **TS19** se observó que las alcaldías **Gustavo A. Madero** (13.8%), tuvo como acompañantes de los participantes a **familiares** (9.4%), **pareja** (1.4%), **amigos** (0.7%) y **solos** (2.2%); en **Azacapatzalco** (7.2%), la tuvo como acompañantes a **familiares** (5.1%), **amigos** (0.7%) y quienes asistieron **solos** (1.4%); en la alcaldía **Miguel Hidalgo** (5.8%) los participantes se acompañaron por **familia** (4.3%), **amigos** y **solos** (0.7%) para cada uno. **Iztapalapa** (5.1%) los acompañantes fueron **familia** (4.3%) y participantes **solos** (0.7%). Los participantes residentes del **Estado de México** (12.3%) manifestaron que fueron acompañados por la

familia (8%) y **amigos** (1.4%), las personas que asistieron **solos** tuvieron un porcentaje del (2.9%). Las alcaldías antes mencionadas son la que mayor participación tuvieron, dentro de éstas se observa que la familia (46%) corresponde al principal acompañante de los participantes, seguido de los que asistente solos (10%), mientras que los acompañantes que corresponden a la pareja y amigos tuvieron un (6%) y (5%).

Motivo

Otras relaciones que se obtuvieron fue la de la variable **Motivo** con las variables **Impacto** y **Satisfacción**. De la relación **Motivo** de asistencia con la de **impacto** en algún aspecto de la vida del participante, los datos obtenidos de la encuesta muestran que el **aprendizaje** fue el motivo con mayor porcentaje (26.8%) de éstos el 15.9% contestó que participar en el taller tuvo impacto en el **aspecto personal**, 5.8% en el aspecto **profesional**, y 3.6% en el aspecto **familiar**. El motivo de **Convivencia familiar** (23.2%), de éste el 11.6% contestó que el taller impactó en el **aspecto personal**, **familiar** 10.1% y para el **aspecto social** 1.4%. El motivo de **Realizar tareas** tuvo un porcentaje de 7.2%, de éste el aspecto **familiar** tuvo el 3.6%, el **aspecto personal** 2.2% y el **profesional** y **social** 0.7%, en ambos casos. En resumen se puede decir que el motivo que llevó a las personas a participar en los talleres fue el aprendizaje vinculado al aspecto personal de la mayoría, el motivo de convivencia familiar se relacionó también con el aspecto personal y la intención de realizar tareas se relacionó con la participación de tipo familiar.

En la relación de las variables **Motivo y Satisfacción**, se observó que el **aprendizaje** fue el primer motivo de preferencia (36.2%), los participantes que dijeron que este fue el motivo de su asistencia, contestaron en relación a la satisfacción que fue **agradable** (19.6%), **cumplió sus expectativas** (3.6%), fue **divertido** (2.9%) y le resultó **poco agradable** (0.7%). La **convivencia familiar** fue el segundo motivo por el cual los encuestados participaron en el taller, lo satisfactorio que resultó esta participación se distribuye de la siguiente manera: para los que fue **agradable** tuvo un porcentaje de (15.9%), **divertido** (6.5%), **cumplió expectativas** (0.7%) y **poco agradable** (0.7%). El tercer motivo por el cual se participó en el taller fue el **simple gusto** de asistir, los encuestados que contestaron este motivo no indicaron qué tan satisfactorio fue el taller. El cuarto motivo fue **diversión** y tuvo un porcentaje de 8%, los encuestados que manifestaron asistir por éste contestaron que el taller les fue **agradable y divertido** (2.2%) respectivamente y **cumplió las expectativas** 0.7%. Por último el motivo de **realizar tareas** ocupó el quinto lugar y tuvo un porcentaje de (8.0%), en éste los encuestados manifestaron que el taller les pareció **agradable** (4.3%), **divertido** (2.2%) y **poco agradable** (1.4%). En general los talleres resultaron **agradables** (44.9%), **divertidos** (14.5%), **cumplió expectativas** (7.2%) y **poco agradable** (2.9%).

Satisfacción

Por último se tiene la relación de la variable **Satisfacción** con las variables **Impacto** y **Alcaldía**. La relación **Satisfacción e Impacto**, explica en qué aspectos de la vida de los participantes tuvieron significado las actividades del taller y qué tan agradables fueron éstas. El participar en el taller resultó **agradable** para (43.5%) para ellos, esta actividad tuvo impacto en el **aspecto personal** (26.1%), **familiar** (10.9%), **profesional** (4.3%) y **social** (2.2%). Para quienes resultó **divertido** el taller (14.5%), manifestaron que la actividad tuvo un impacto en el **aspecto familiar** (7.2%), **personal** (5.1%), **social** (1.4%) y **profesional** (0.7%). Por último, para los encuestados que dijeron que el taller **cumplió sus expectativas** (4.3%) dijeron que la actividad realizada impactó en el **aspecto personal** de su vida (4.3%), en el **profesional** (2.2%) y **familiar** (0.7%). Se observó que el (37.7%) de los encuestados

contestó que las actividades tuvieron impacto en el **aspecto personal** de sus vidas, seguido por el impacto en el ámbito **familiar** (19.6%) y por último en el **profesional** (7.2%), siendo estas actividades calificadas mayoritariamente como: **agradables** (43.5%), **divertidas** (14.5%) y que **cubrieron expectativas** (7.2%).

La relación entre **Alcaldía** y **Satisfacción**, se explica por el grado de satisfacción de los participantes y de su lugar de residencia. Con base en los resultados de la encuesta de talleres sabatinos, se observó que el mayor porcentaje de participación entre alcaldías corresponde a **Gustavo A. Madero** (13.8%), le siguen las Alcaldías **Azcapotzalco** (7.2%), **Miguel Hidalgo** (6.5%) e **Iztapalapa** (5.1%). El Estado de México tuvo una participación importante (12.3%). Los encuestados residentes en la alcaldía **Gustavo A. Madero** manifestaron que las actividades del taller resultaron **agradables** (8.7%), **cubrieron sus expectativas** (2.9%), fueron **divertidas** (1.4%) y **poco agradables** (0.7%). Para los residentes en la alcaldía **Azcapotzalco** las actividades fueron: **agradables** (3.6%), **divertidas** (1.4%), **cubrió expectativas**, **poco agradables** y **aburridas** (0.7%) respectivamente; para los residentes en la alcaldía **Miguel Hidalgo** las actividades del taller resultaron: **agradables** (3.6%), **divertidas** y **poco agradables** (1.4%) respectivamente, por último los encuestados residentes en la alcaldía **Iztapalapa** dijeron que las actividades resultaron: **Agradables** y **divertidas** (2.2%) respectivamente y **cubrieron expectativas** (0.7%). Por último, los participantes residentes en el **Estado de México** dijeron que las actividades del taller les resultaron: **agradables** (8.0%), **divertidas** (3.6%) y **cubrieron expectativas** (0.7%). Por lo anterior, en general las actividades resultaron **agradables** (42.8%), **divertidas** (14.5%), **cubrieron expectativas** (7.2%), las menos fueron **poco agradables** (2.9%) y **aburridas** (0.7%).

Hasta aquí se muestran los resultados obtenidos entre las trece variables de la encuesta de **Talleres Sabatinos**, en general se observa que existe una correlación entre doce variables excepto entre las variables **Primera Visita** y **Número de Visitas** (ver matriz de correlación anexo II) en la que no se encontró relación alguna, ello debido a que el 84% de participantes fueron de primera visita, por tanto la pregunta de número de visitas que tenían una escala del 1 al 10 quedó anulada y con ello la probabilidad de relacionar la frecuencia de asistencia a los talleres.

Relación de variables del programa de Noche de Museos

Tema

En la matriz de correlación (que se muestra en el anexo I) se observa una relación entre **Tema** y las variables **Alcaldía** y **Acompañante**.

La relación entre las variables **Tema** y **Alcaldía**, muestra los temas que se dieron en las Noches de Museos y la alcaldía de residencia de los participantes.

La variable **Tema** de la noche de Museos se relaciona con las variables **Acompañante** y **Alcaldía**, con base en los datos de la encuesta se observó que el tema **Siempre Mujeres**

tuvo una participación del 32.5% en la cual los principales motivos de participación fueron: **aprendizaje** (13.3%) y **curiosidad** (4.4%). El tema de **Murales** tuvo una participación del 30.5%, en la cual los participantes manifestaron que los motivos principales de participación fueron: **aprendizaje** (19.2%), **diversión** (4.9%), **curiosidad** (3.4%) y **convivencia familiar** (2.0%). Por último el tema de **Guardianes de Leyendas** con 21.7% de participación mostró que los motivos de fueron: **aprendizaje** (3.9%), **diversión** (3.4%) y **curiosidad** (2.0%).

Mientras que en la relación **Tema y Alcaldía**, los datos de la encuesta de **NM19-20**, mostró que el porcentaje más alto de encuestados son residentes de la alcaldía **Coyoacán** con 8.4%, en ésta la participación en los temas se dio de la siguiente forma: Siempre mujeres (3.9%), Murales y Guardianes de Leyenda con (1.5%) en ambos, Fiestas Tradicionales (1%) y El mito y la leyenda (0.5%), no hubo participantes en Animales Mesoamericanos. **Miguel Hidalgo** fue la segunda alcaldía con mayor porcentaje 7.4%, esta mostró que los participantes asistieron a los temas de Murales (3.4%), Siempre Mujeres y Fiestas Tradicionales (1.5%) respectivamente, el Mito y la leyenda y Animales Mesoamericanos (0.5%) ambos y no hubo participantes en Guardianes de Leyenda. El **Estado de México** mostró una participación importante ya que el porcentaje de residentes de este estado que se encuestó fue 20.2%, los cuales participaron en los temas Murales (9.4%), Siempre mujeres (5.9%), Guardianes de Leyenda (3.0%), El mito y la leyenda (1%), Fiestas tradicionales y Animales Mesoamericanos (0.5%), respectivamente. Otro dato a destacar es que las alcaldías de **Venustiano Carranza, Xochimilco, Tláhuac, Iztacalco y Álvaro Obregón** tuvieron la participación más baja con respecto a otras alcaldías en el evento de Noche de Museos que va del 0.5% al 4.4%. Los temas con participación más baja fueron Animales Mesoamericanos (3%) y Fiesta Tradicionales (3.4%).

Escolaridad

El cálculo de la matriz de correlación para la encuesta de **NM19-20**, mostró una relación entre el **nivel de escolaridad** y las variables **medio de información, Visita a otros espacios, Horario, Tema, Alcaldía y Motivo**. Del total de los encuestados, el nivel de escolaridad con mayor porcentaje fue el de Educación Superior con 59.1%. Los participantes mujeres y hombres con educación superior hicieron uso de los siguientes medios de información: Redes Sociales (25.1%), Internet (página del museo) (18.2%) y Familiares (12.3%). Los participantes con educación Media Superior corresponden al 7.4% de los encuestados, estos se informaron a través de los siguientes medios: Familiares (3.4%), Internet (2%), Redes sociales (1.5%) y Anuncios (0.5%). Los participantes con nivel de Educación Básica fueron 3.9%, ellos se informaron a través de Familiares (2%), Internet y Redes Sociales (1%) respectivamente. Por último los participantes con Otros estudios (3.4%) y Posgrado (3.4%) se informaron por Familia, Redes sociales e Internet.

Entre la variable **escolaridad** de los participantes y la variable **visitó otros espacios culturales** donde se llevaron a cabo las Noches de Museo. El 41.4% de los encuestados dijo que sí había visitado otros espacios donde se realizaron Noches de Museos, entre estos el porcentaje más alto de 29.1% corresponde a personas con Educación Superior, en tanto que el 6% en conjunto de los encuestados con Educación Básica, Media Superior, otros estudios y posgrado también han asistido a otros espacios y 6.4% dijo haber visitado otros espacios aunque no indicó el grado de estudios.

También se encontró una relación entre el nivel de educación de los participantes y el horario de asistencia a las Noches de Museos, hay que recordar que el horario del evento

es de 19:00 a 21: 00 horas. Con base en los datos de la encuesta de **Noche de Museos**, se observó que para el caso de los encuestados con educación superior, que son los que participan mayormente en este evento, se observó que la mayoría asiste en tal horario porque después de sus actividades regulares es el tiempo de que disponen (32.5%) y el segundo motivo, para otros tantos, es porque **es el único horario que existe para este evento** (17.7%), mientras que los participantes con educación básica que asistieron en este horario es porque **les permitió salir con la familia** (2.5%).

Los temas de las Noche de Museos fueron: “Murales”, “Siempre Mujeres”, “Animales Mesoamericanos”, “El Mito y la Leyenda”, “Fiestas Tradicionales” y “Guardianes de Leyenda” se observa que el 59% de participantes son de educación superior, mientras que entre el 3% y 7% se distribuye entre participantes con educación básica, media superior, posgrado y otros⁷. En los resultados obtenidos de la matriz de correlaciones se tiene una vinculación entre el **nivel educativo** de los participantes y la **Alcaldía** o estado donde residen. Los datos de la encuesta de **Noche de Museos** muestran que el 59% de participantes tienen nivel de estudios superior y provienen mayormente de las alcaldías **Coyoacán (8%), Tlalpan (5%), Álvaro Obregón y Cuauhtémoc (4%)** respectivamente. No se recibieron participantes con nivel de educación superior de las alcaldías **Tláhuac y Venustiano Carranza**. En tanto que los participantes con educación básica solo se recibieron de cinco alcaldías cada una con 3% de participación: **Benito Juárez, Iztapalapa, Miguel Hidalgo, Gustavo A. Madero y Tlalpan**. Es de resaltar que del **Estado de México**, el museo recibió la participación del 16% de personas con nivel educativo superior, mientras que de educación básica participo el 0.5%.

Con base en los resultados de la matriz de correlación se obtuvo una relación entre el nivel **educativo de los participantes** y el **motivo** que manifestaron para asistir, ya que 35% con nivel superior afirmó que el motivo de su asistencia fue el aprendizaje, seguido de la curiosidad y diversión con un 20% en ambos casos, en comparación con los motivos de Realizar tareas, y la Convivencia familia cuyos porcentajes fueron del 1% al 2%. También se observó que la participación de personas con educación básica es la más baja 4% y el motivo de participación es el aprendizaje con 3%.

Primera Visita (PV).

La variable **Primera visita** se encontró relacionada con las variables **Horario y Edad**. En los datos de la encuesta de Noche de Museos, la primera participación (78.8%) de los encuestados, se relacionó con sus tiempos disponibles (33%), en tanto que, los que manifestaron haber participado más de una vez (21.2%), menos de la tercera parte (8.9%) acudieron en ese horario porque fue el único existente, lo que presupone un interés del tema o evento y no solo del horario.

En el rango de edad de los participantes que va de 8 a 82 años, sin considerar el género, se observó que el 90% manifestó asistir por primera vez al evento de Noche de Museos, en tanto que el 10% restante dijo haber asistido en más de una ocasión. En los datos totales

⁷ Nota aclaratoria: En esta correlación no se detallan los porcentajes debido a que los datos obtenidos podrían sesgar la información. Vale aclarar que dentro de la metodología de aplicación del instrumento de la encuesta, se definió la aplicación de 20 cuestionarios en cada Noche de Museos independientemente del número de grupo y temas abordados en cada evento, no obstante, por compromisos interinstitucionales el departamento a cargo del evento tuvo que participar en la aplicación de otra encuesta para el mismo evento, por lo que se redujo a 10 el número de encuestas aplicadas y elaboradas por el departamento lo cual impacto los resultados esperados en la variable tema de la que se esperaba conocer las preferencias de los públicos respecto a ellos.

obtenidos, resalta que el 31% que asistió por primera vez se encuentra en la década de los veinte años, le siguen los jóvenes de la década de los treinta con el 19%, posteriormente los de cuarenta y cincuenta años, que alcanzan una participación del 9% y 7%, quedando para las décadas restantes una participación del 9%. En lo que respecta a los participantes que han asistido más de una vez ascienden al 4% y es población que se distribuye entre todas las edades.

Motivo de asistencia

Otra vinculación que se encontró entre variables fue la de **Motivo de asistencia** con **Lugar de procedencia, Alcaldía y Ocupación**.

Otra vinculación que se encontró entre variables fue la de motivo para asistir al evento y el lugar de procedencia del participante. El 46.8% de los encuestados dijo que el motivo para asistir al evento fue el aprendizaje, los cuales provenían de su hogar (21.7%), oficina o lugar de trabajo (12.3%), lugar de esparcimiento (8.9%) y escuela (3.9%); el segundo motivo que llevó a los participantes al museo fue la convivencia familiar con un 4.5%, ellos se trasladaron de la oficina o lugar de trabajo (2.5%) y el hogar (2%). El tercer motivo por el que participaron en el evento fueron las tareas escolares con 1.5% y éstos se trasladaron de diferentes espacios (ver gráfica).

Los datos de la encuesta de **Noche de Museos**, muestran que el motivo de **aprendizaje** tiene el porcentaje del 46.8%, que se distribuye entre los participantes residentes de todas las alcaldías, solo en **Miguel Hidalgo, Coyoacán, Gustavo A. Madero e Iztapalapa** alcanza un porcentaje del 4% al 6%, destaca la participación de los residentes del **Estado de México** con un 13%. Mientras que el porcentaje de participación por motivos de **Pasar tiempo con la familia** y **Realizar tareas escolares** fue de 1% y 2%, distribuido en todas las Alcaldías, de ellas destacan por su menor participación las de **Iztacalco, Xochimilco, Venustiano Carranza y Tláhuac** que van del 0.5% al 1% y que manifestaron varios motivos de participación.

Entre las variables de **Ocupación** y **Motivo** de participación, el 16% de los participantes que se desempeñaban como empleados tuvieron como motivo para asistir el aprendizaje, los estudiantes en su mayoría asistieron por curiosidad y diversión (6% entre ambos), por aprendizaje (2%) y para realizar tareas (1%), los jubilados solo tuvieron como motivo el asistir por aprendizaje(), los participantes con otra actividad asistieron por convivencia familiar (0.5%), curiosidad (1.5%) y aprendizaje (2%). El motivo de aprendizaje está presente en mayor o menor medida en todas las categorías de actividad.

Ocupación

La siguiente relación se dio entre las variables **Ocupación** y **Lugar de procedencia**. Con base en los datos obtenidos de la encuesta de Noche de Museos, el 59% de los participantes con nivel de Educación Superior provienen de su hogar (23%), de la oficina (20%), de algún lugar de esparcimiento (10%) y de la escuela (6%), mientras que los participantes con nivel Medio Superior entre el 2% y 4% proceden del hogar, escuela y lugar de trabajo; en el caso de personas con Educación Básica entre el 3% y 1% procede del hogar y la oficina. Las personas con nivel de posgrado u otros estudios procede del hogar y la oficina (1% a 2.5%). Por lo tanto, las personas con Educación Superior muestran

una mayor variedad del lugar de donde proceden, en comparación con otros participantes con grado de estudio mayor o menor al de nivel superior.

3.4 Análisis por métodos multivariados.

3.4.1 Análisis Factorial.

Talleres sabatinos.

Con la aplicación del método de análisis factorial exploratorio fue posible reducir las trece variables originales a dos factores (Factor 1 y Factor 2), facilitando el análisis del público asistente a Talleres Sabatinos.

La agrupación de las variables en los dos factores se muestra en la siguiente tabla 3.1.

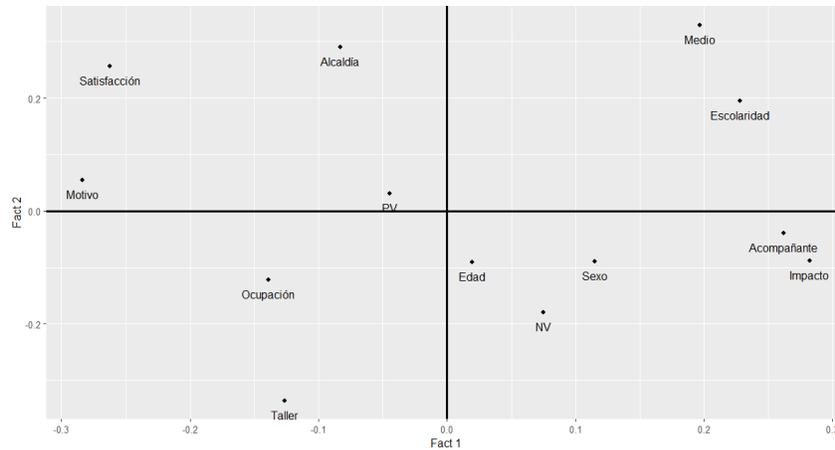
Factor 1	Factor 2
Otros Espacios (OE)	Tema
Acompañante	Medio
Motivo	Satisfacción
Impacto	NV
Escolaridad	Edad
Ocupación	
Sexo	
Primera Visita	

Tabla 3.1 Factores.

El Factor 1 agrupa las variables Otros Espacios, Acompañante, Impacto, Escolaridad, Ocupación Sexo y Primera Visita. En el Factor 2 se agrupan las variables Tema del taller, Medio de información Satisfacción obtenida al participar en la actividad, Número de visitas y Edad de los participantes.

El porcentaje de varianza retenido con el primer y segundo Factor fue de 63%, la Comunalidades más altas en la extracción se observan en las variables Medio (0.65), Impacto (0.62) y en la variable Tema (0.56) (Ver tabla Comunalidades Talleres Sabatinos Anexo I).

La siguiente gráfica 3.14 muestra la distribución las variables en el primer y segundo factor a través de las cargas de factores.



Gráfica 3.14 Distribución de las variables originales de Talleres Sabatinos en base a la carga de factores.

Noche de Museos

El análisis factorial exploratorio permitió reducir las catorce variables originales a dos factores (Factor 1 y Factor 2), facilitando el análisis del público asistente a las Noches de Museos.

La agrupación de las variables en los dos factores se muestra en la siguiente tabla 3.2.

Variables	
Factor 1	Factor 2
Horario	Medio
PV	Escolaridad
Lugar	Alcaldía
Ocupación	OE
Edad	Sexo
Motivo	Día
Tema	Acompañante

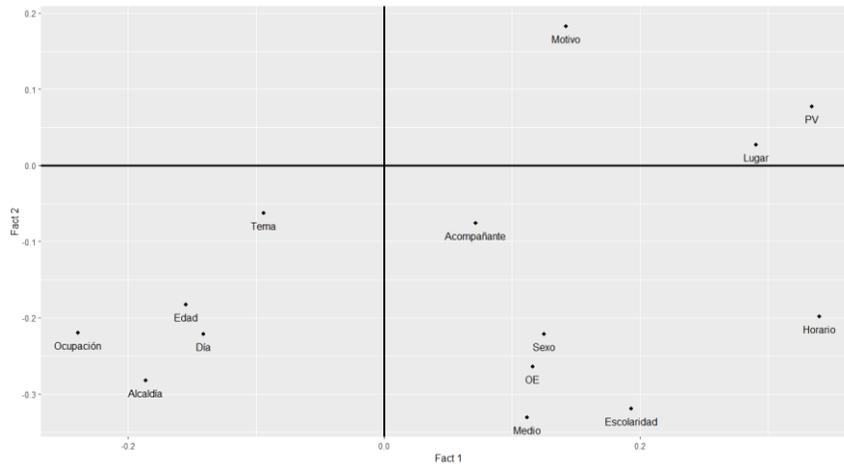
Tabla 3.2 Distribución de las variables originales en las factores.

El Factor 1 agrupa las variables que corresponden a los datos demográficos Ocupación y Edad, Primera Visita, horario de asistencia, lugar de procedencia, Motivo para asistir y Tema de la Noche de Museos. En el Factor 2 se agrupan las variables que se relacionan directamente con los participantes, como son el medio o canal de información, nivel de

escolaridad, Alcaldía de residencia, la participación en otros espacios culturales, el sexo del participante, el día que prefieren para asistir al evento y los acompañantes.

El porcentaje de varianza retenido con el primer y segundo Factor fue de 49%, la Comunalidades más altas en la extracción se observan en las variables Primera visita (0.87), Otro espacios (0.87) y en la variable Motivo (0.58) (Ver tabla Comunalidades Noche de Museos Anexo I).

La siguiente gráfica 3.15 muestra la distribución las variables en el primer y segundo factor a través de las cargas de factores.



Gráfica 3.15 Distribución de las variables originales de Noche de Museos en base a la carga de factores.

3.4.2 Análisis de Componentes Principales.

En esta sección se exponen los resultados obtenidos a través del análisis por el método de Componentes Principales aplicado a las encuestas Talleres sabatinos (TS-19) y Noche de museos (NM19-20), el cual consiste en el cálculo de la matriz de correlaciones, obtención de eigenvalores que se relacionan con la varianza que permite determinar el número de Componentes principales (CP) y gráficas de la distribución de variables y observaciones con relación a las componentes principales. Los datos numéricos obtenidos se muestran en el anexo I.

Distribución de la varianza.

La varianza de las primeras dimensiones muestra la existencia de relaciones fuertes entre las variables y sugiere el número de dimensiones que se deben estudiar.

Dada la información de las encuestas de TS-19, se tiene tres componentes principales cuyos eigenvalores son: 1.737, 1.178 y 1.028, los cuales corresponden a los tres primeros componentes principales. Los valores de porcentaje de la varianza indican la cantidad de información que se mantiene en cada componente principal, al elegir los tres primeros componentes la varianza acumulada mantiene un 78% de información la cual es suficiente para el análisis.

En el caso de la encuesta de **NM19-20**, se obtuvieron cuatro componentes principales, cuyos eigenvalores son: 1.3927, 1.2122, 1.1062 y 1.0664 los cuales corresponden a las cuatro primeras componentes, la varianza a cumulada para las primeras dos componentes mantiene un 49% de información, las tres primeras componentes mantienen un 66% y la cuarta componentes mantiene un 82% de información. La siguiente figura muestra la distribución de la varianza en los componentes principales para **Talleres Sabatinos** y **Noche de Museos** (Ver figura 3.16).

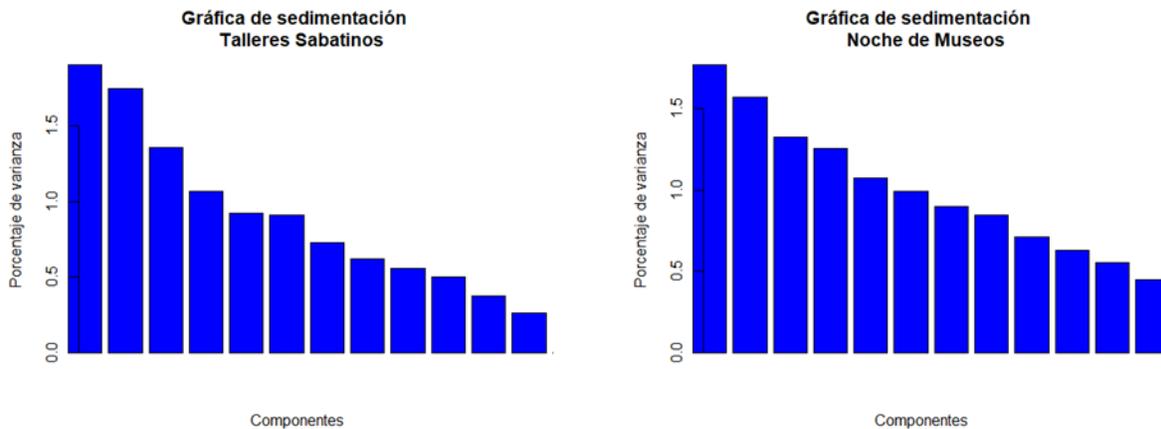


Figura 3.16 Descomposición de la varianza total para talleres Sabatinos y Noche de Museos.

Descripción de la relación entre componentes principales y variables originales de Talleres Sabatinos y Noche de Museos.

Talleres Sabatinos

En **TS-19**, el primer componente (Dim1 correspondiente al eje horizontal de la gráfica, ver Figura 3.17) tiene asociaciones con las variables **Impacto**, **Acompañante** y **NV** (a la derecha del gráfico) y las variables **Satisfacción**, **Motivo** y **Alcaldía** (a la izquierda del gráfico). Este componente principal no se puede interpretar como una medición exacta del bloque de variables que se le asocian, más en cambio puede dar una idea de que las variables asociadas como son **Impacto** y **Acompañante** se relacionan con el participante de manera personal; las variables **Motivo** y **Satisfacción** indica que las actividades cumplen las expectativas de los participantes.

El segundo componente principal (Dim2 corresponde al eje vertical de la gráfica, ver Figura 3.17) se asocia con las variables **Medio** de información y **Escolaridad** (en la parte superior del gráfico) y las variables **Taller** y **Ocupación** (en la parte inferior del gráfico), al igual que el primer componente principal no se puede interpretar como una medición exacta entre el medio de información utilizado para conocer las actividades y el nivel de educación de los participantes, como tampoco qué el tema del Taller esté relacionado a la ocupación de los asistentes.

La tercera componente principal que no se graficó se tiene una asociación indirecta con la variable Primera Visita (PV) y Satisfacción.

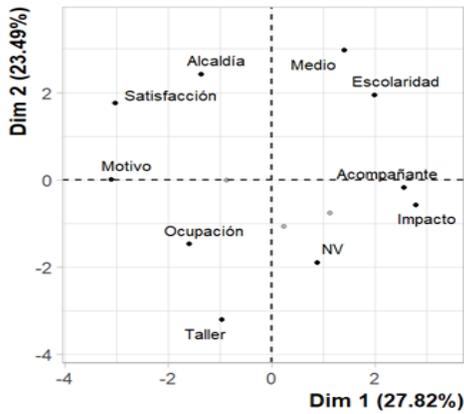


Figura 3.17 Plot de variables y Componentes Principales de TS-19

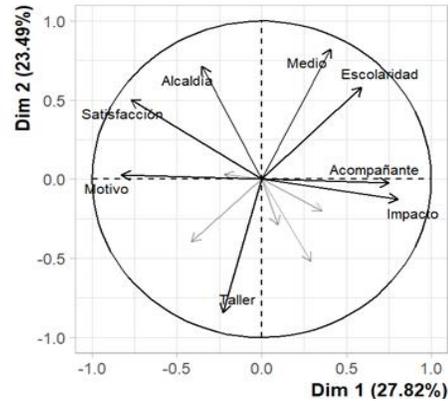


Figura 3.17 Biplot de variables originales y Componentes Principales

Noche de Muesos

El primer componente principal (Dim1 corresponde al eje horizontal de la gráfica, ver Figura 3.18) tiene una asociación con las variables **Escolaridad**, **Ocupación** y **Acompañante** (a la derecha del gráfico parte inferior) y variables como **Motivo** y **Lugar** de procedencia del participante (a la izquierda del gráfico parte superior). Este componente no se puede interpretar como una medición exacta de la relación entre estas variables ya que no es preciso decir que indica esta asociación.

El segundo componente principal (Dim2 corresponde al eje vertical de la gráfica, ver Figura 3.18) tiene asociadas las variables **OE** y **PV** (en la parte superior derecha del gráfico) y las variables **Día** y **Sexo** (en la parte inferior izquierda del gráfico). Este componente no se puede interpretar como una medición exacta entre las variables, ya que no es posible decir qué indica esta asociación, más sin embargo se observó que el género femenino prefiere el día miércoles para esta actividad, en cuanto a la relación con las variables **PV** y **OE** no se encontró un indicador de frecuencia de participación.

Por último el tercer componente principal que no se graficó, presenta una asociación entre las variables Lugar (Lugar de procedencia del participante) y Día (Día que se prefiere para asistir a la actividad).

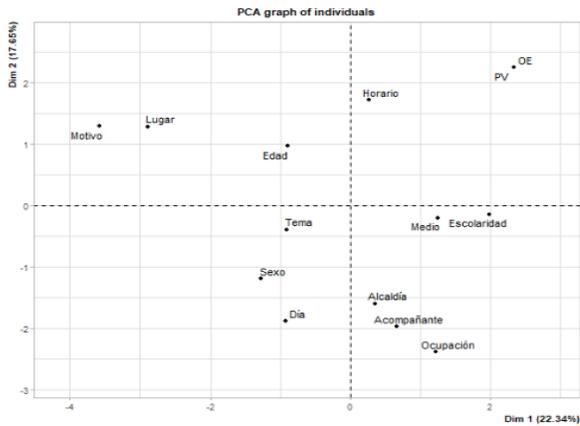


Figura 3.18 Plot variables originales y Componentes Principales NM19-20

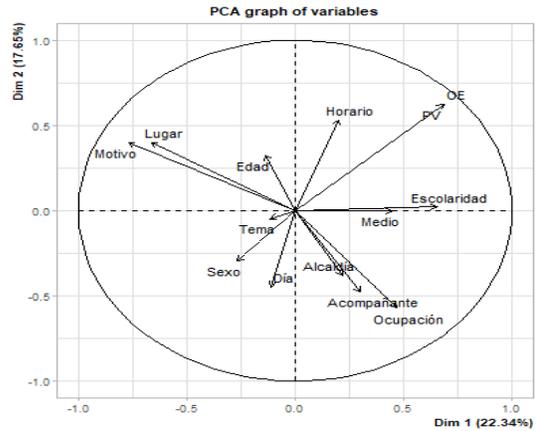


Figura 3.18 Biplot de variables originales de Componentes Principales NM19-20

3.4.3 Análisis de Conglomerados o Clústeres.

El análisis de conglomerados muestra el agrupamiento de las variables y observaciones, que se forma de manera natural, este se aplicó a los resultados de las encuestas **TS-19** y **NM19-20**, los resultados obtenidos se muestran en la gráfica conocida como “dendrograma”, las cuales corresponden a las variables y observaciones, el dendrograma se obtuvo a través del cálculo de la matriz de distancias euclidianas y la obtención del coeficiente de agrupación que indica el método más conveniente a aplicar, el cual resultado ser el método “Ward’s”. La matriz de distancia y los coeficientes de agrupamiento pueden consultarse en el anexo I.

Dendrograma

Talleres Sabatinos

La clasificación realizada para **TS-19** sobre las variables revela 4 clústeres, mientras que para **NM19-20**, revela 3 clústeres (ver Figura 3.19 y 3.20).

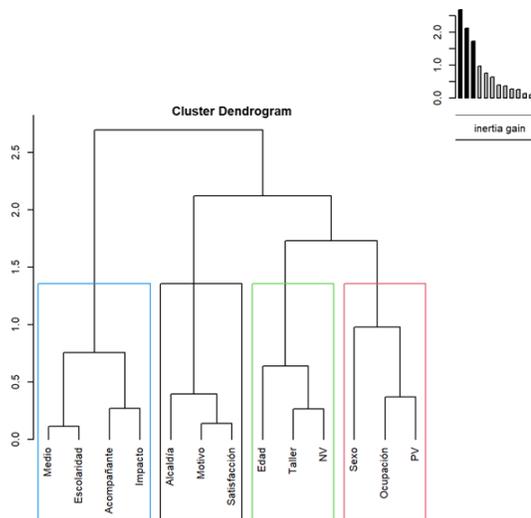


Figura 3.20 Árbol jerárquico TS-19

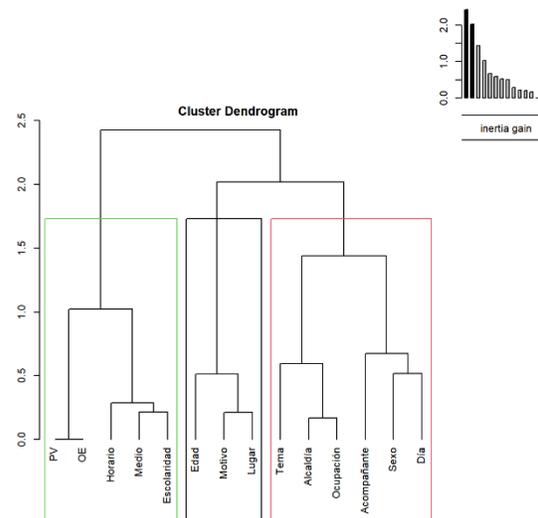


Figura 3.21 Árbol jerárquico NM19-20

En la encuesta de TS-19 la clasificación realizada sobre las variables se indica de la siguiente forma:

El clúster 1 está formado por variables como Motivo, Satisfacción y Alcaldía, este grupo se caracteriza por valores altos para las variables Satisfacción, Motivo y Alcaldía (las cuales se ordenan por las más altas).

El clúster 2 está formado por variables como Ocupación, PV y Sexo y se caracteriza por valores altos para PV.

El clúster 3 está formado por variables como Taller y NV. Este grupo se caracteriza por valores altos para las variables Taller, NV, y Edad.

El clúster 4 está formado por variables como Escolaridad, Medio, Acompañante e Impacto. Este grupo se caracteriza por valores altos para las variables Escolaridad, Impacto, Medio y Acompañante.

En el caso de **NM19-20**, la clasificación de los grupos es la siguiente:

El grupo 1 está formado por individuos como Lugar y Motivo. Este grupo se caracteriza por valores altos para las variables Motivo, Lugar y Edad (las variables se ordenan de la más fuerte).

El clúster 2 está formado por personas como Ocupación, Acompañante, Día y Alcaldía. Este grupo se caracteriza por variables cuyos valores no difieren significativamente de la media.

El clúster 3 está formado por individuos como Escolaridad, PV, OE y Horario. Este grupo se caracteriza por valores altos para las variables OE, PV, Escolaridad, Horario y Medio (las variables se ordenan de la más fuerte).

Dendrograma de observaciones.

En las siguientes gráficas se observa los dendrograma correspondientes a los **Talleres Sabatinos** y **Noches de Museos**. Las siguientes graficas muestras los grupos obtenidos a través de una función en Rstudio, (se agrega en el Anexo II) para el diseño de dendrogramas, estas gráficas no fueron obtenidas a través de la matriz de correlación y dado que se realizaron a través de una función programada no me es posible tener la certeza del cómo se realiza el cálculo de la matriz de distancias, esta función solicita la medida de distancia que en este caso fue la euclidiana, así como con el coeficiente de agrupación más alto que fue el del método de Ward's. Y las calcule por curiosidad (ver Figura 3.21).

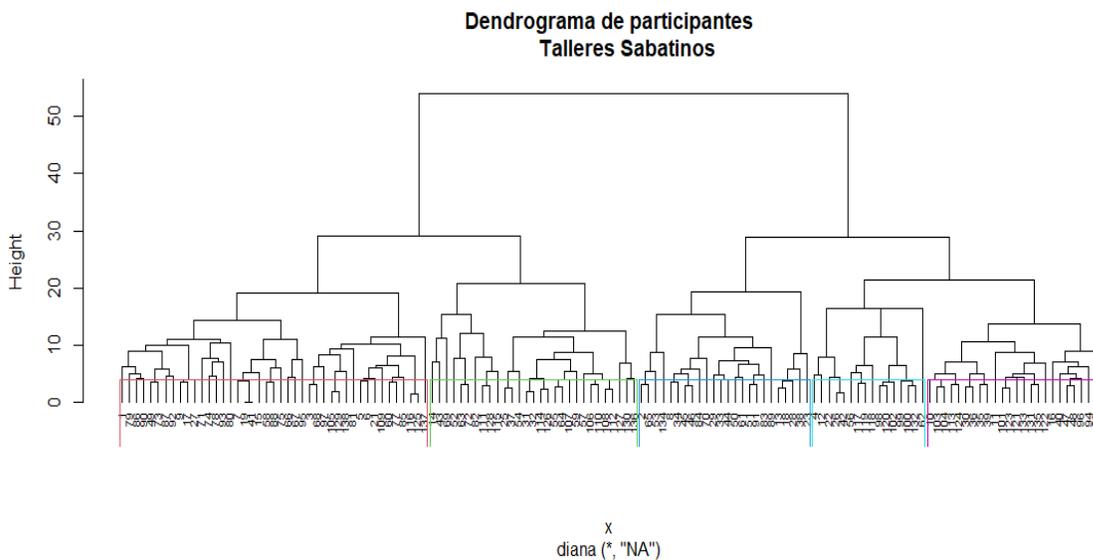


Figura 3.21 Dendrograma observaciones Talleres Sabatinos.

Características de los grupos Talleres Sabatinos.

Grupo I Rojo: Este grupo se conforma con la mayor cantidad de participantes encuestados 43 (31%) los participantes mujeres son 29 y 10 son hombres 4 participantes no indican su sexo. La participación en los talleres se dio de la siguiente manera: Bordando Historias 2 participantes, Códice Dresde 12, Taller de Danza y Estelas de Xochicalco 5, Instrumentos Musicales 3, Taller del Día del Niño 16 participantes. Las edades de los participantes en este grupo son de 38, 39, 40, 42, 43, 44, 45, 46, 47, 48 y 50 años, siendo la edad de 40 años la de mayor participación. La principal ocupación de este grupo es el de empleado, con grado de escolaridad superior, el medio por el cual se enteró de las actividades son las Redes Sociales, para este grupo fue su primera participación en las actividades y el motivo principal fue la Convivencia Familiar, ellos asistieron mayormente con la familia, su experiencia fue agradable y divertida, y todos ellos son residentes del Estado de México.

“En este grupo destacan las variables de Educación y Medios de información, por lo que los participantes de este grupo destacan por que su nivel de educación es Superior y el uso de Internet y Redes sociales son preferentemente sus medios de información”.

Grupo II Verde: Este grupo se conforma por 29 (21%) encuestados, 18 participantes son mujeres y 7 varones, 4 no contestaron su sexo., las edades de los participantes son de 48 a 62, 64 y 65 años, la edad con más participantes es de 50 años. La participación en los talleres se da de la siguiente forma: Bordando Historias 1 participante, Códice Boturini 1, Códice Dresde 6, Taller de Danza 11, Estelas de Xochicalco 2, Instrumentos Musicales 7, Taller del Día del Niño 1 participante, este grupo se conforma de empleados y Jubilados, el grado de escolaridad es Superior; como medio de comunicación utilizan las redes sociales; es su primera participación, el principal motivo para participar es el “Aprendizaje”, la mayoría no especifica con quien asiste y quienes lo hacen dicen que es con la familia, no contestan a pregunta de qué impacto tuvo esta actividad en su vida, la mayoría no contestó el grado de satisfacción y quienes los hicieron dicen que les resultó “Agradable”, no contestaron en que Alcaldía residen.

Grupo III Azul: Este grupo está conformado por 23 (17%) participantes del total de encuestados, los cuales son 9 mujeres y 10 varones y 4 participantes no contestaron su sexo, las edades de los participantes son: 23, 27, 30, 31 a 37 años, siendo la edad de 37 años la de la mayoría de los participantes y las de 31 y 32 años, con menor número de participantes, este grupo tuvo participación en los siguientes Talleres: Códice Boturini 1 participante, Códice Dresde 11, Estelas de Xochicalco 2, Instrumentos Musicales 5, Taller del Día del Niño 4 participantes, ellos tienen ocupación de empleados y estudiantes con grado de estudios “Superior”, medio por el cual se informaron de la actividades son las “Redes Sociales”, fue su primera visita al taller, el motivo de participación fue la “Convivencia Familiar”, “Aprendizaje” y “Curiosidad”, asisten a la actividad con “Familia”, “Amigos” y “Solos”, manifestaron que las actividades tienen un impacto en su vida “Familiar” y “Personal”, la experiencia les fue: “Agradable ” y “Divertida” y residen en las alcaldías Gustavo A. Madero, Azcapotzalco, Álvaro Obregón y Estado de México, mayormente.

Grupo IV Azul claro: Es un grupo con 16 (12%) participantes de los cuales 10 son mujeres, 3 son hombres y 3 no indicaron su sexo, en este grupo las edades son de 12, 15, 16, 18, 19, 20, 21, y 22, y la edad con mayor participantes es de 22 años seguida de la de 18 años; la participación en los talleres se dio de la siguiente manera: 1 participante en Códice Boturini, 5 en Códice Dresde, 8 en Taller de Danza y 2 participantes en Instrumentos Musicales, este grupo estuvo compuesto por “Estudiantes”, con grado de escolaridad “Medio Superior” principalmente, seguido de “Superior” y un solo caso de nivel básico, los participantes se enteraron del evento a través de las “Redes Sociales”, “Familiares” y de la

página del museo; fue su primera visita, el motivo de su participación fue la “Convivencia Familiar” y el “Aprendizaje”; la mayoría no contestó con quien asisten a las actividades, los que contestaron dijeron que con: “Familia”, “Amigos” y “Solos”; No contestaron cómo impacto su vida esta experiencia y quienes lo hacen dicen que en lo personal, la mayoría no contestó cómo fue esta actividad para ellos, quienes lo hicieron dijeron que fue “Agradable”; y la gran mayoría no contestó en qué alcaldía reside.

Grupo V Morado: Este grupo se forma con 23 (17%) participantes, de los cuales 14 son mujeres y 7 son hombres, 2 no especificaron su sexo; las edades de este grupo son: 23, 24, 25, 26, 27, 30, 31 y 33, la edad con mayor participantes es de 25 años seguida de la de 30 años, la participación en los talleres se dio de la siguiente forma: Códice Boturini 2 participantes, Códice Dresde 9, Taller de Danza 8, Instrumentos Musicales 2, Taller del Día del Niño 2 participantes, la ocupación en este grupo está indicada principalmente por “Estudiantes” y “Empleados” y “Otra”; el grado de escolaridad es de “Superior” y “Medio Superior”; los medios de comunicación por los que se enteran fueron “Redes Sociales”, “Página del Museo”, “Familiares” y “Anuncios”; fue su primera visita; el principal motivo de participación fue el “Aprendizaje”; 10 participantes de este grupo no respondió con quienes asistieron al taller, mientras que 8 dijeron asistir con la “Familia”, 3 asistieron con la pareja; la mayoría no contestó cómo impacto su vida la participación en el taller, 6 contestaron que el impacto fue en lo personal; 10 no contestaron qué les pareció el Taller, 6 dijeron que fue agradable, mientras que para 4 se cubrieron sus expectativas. La mayoría de los (11) participantes no contestaron en qué alcaldía residen.

“Las variables que destacan en este grupo son: Edad y Taller,”

Dendrograma

Noche de Museos

La siguiente gráfica dendrograma corresponde a la encuesta **NM19-20**, en la cual se obtuvieron cinco agrupaciones que están delimitadas por semi-recuadros de colores (ver Figura 3.22).

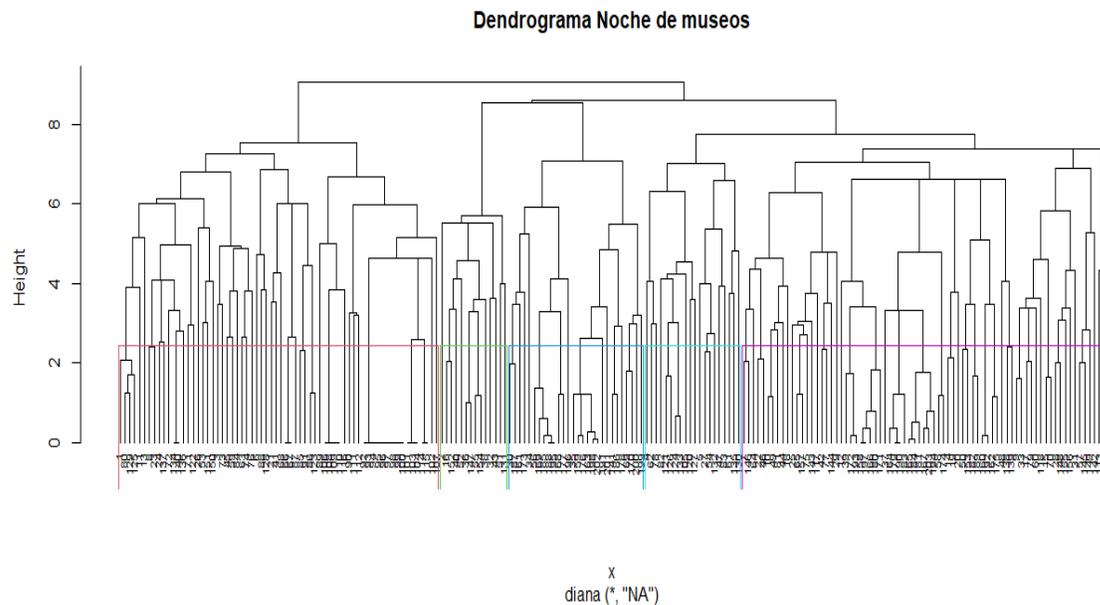


Figura 3.22 Dendrograma 1 Noche de Museos.

Características de los grupos Noche de Museos.

Grupo I Rojo: Este grupo se conforma con 53 (34%) de los encuestados, de los cuales 36 son mujeres y 17 hombres. El rango de edad se encuentra entre 19 a 29 años, el grado de escolaridad de los participantes es de nivel Superior (44), sus ocupaciones son las siguientes: estudiante (14) y empleado (11): El medio por el cual se enteraron de las actividades fueron: por Redes Sociales (21), a través de familiares (15) y por internet (14). Este grupo en su mayoría asiste por primera vez a la Noche de Museos (46) y la mitad dijeron haber participado también en otros espacios (26); Ellos asistieron en este horario porque es el “Tiempo de que disponen” (34); Al evento asistieron acompañados por: amigos (19), pareja (17) y familia (12); y llegaron al museo desde su oficina(16), lugar de esparcimiento (14), escuela (12) y hogar (11); el motivo principal por el que dijeron asistir fue el aprendizaje (30) y la diversión(10); en este grupo la mayoría de los encuestados no contestó qué día de la semana prefieren para asistir al evento, y los pocos que respondieron prefieren el día miércoles (11). La mayoría de los participantes residen en el Estado de México (21). Los temas en los que mayor participación hubo fueron “Murales” (25), seguido de “Siempre Mujeres” (11) y “Guardianes de Leyenda” (10).

Grupo II Verde: Este grupo se conforma por 15 (10%) de los encuestados de los cuales 7 son mujeres, 7 hombres y 1 no especifica su sexo. La edad de los participantes se encuentra en el rango de 20 a 28 años (con un participante de 13 y uno de 18 años). El nivel de escolaridad predominante es el de Superior (11); Las ocupaciones de los participantes son la de estudiante (5) y empleado (4). El medio por el cual se informó de las actividades de Noche de Museos fue por Redes sociales (8), Internet (5) (Página del Museo) y familiares (2). Este grupo asiste por Primera vez a la Noche de Museos (14); La mayoría no ha visitado “Otros espacios” (9). Los participantes manifestaron asistir en este horario por que fue el “Tiempo del que disponían” (8) y fue el único horario existente (5); ellos fueron acompañados al evento principalmente por la pareja (7); y dijeron llegar directamente al museo desde su hogar (6) y oficina (5); asistieron motivados por el aprendizaje (6) y la diversión (6); ellos prefieren asistir a este evento el día sábado (4) y proceden principalmente de las alcaldías de Miguel Hidalgo (4), Iztapalapa (4) y Gustavo A. Madero (3). La participación por Temas se dio en la mayoría en “Siempre Mujeres” (4), le siguieron “El Mito y la leyenda” (3) y “Guardianes de leyenda” (3).

Grupo III azul: Este grupo se conforma por 41 (26%) participantes, los cuales son 23 mujeres y 16 hombres y 2 no indicaron su género. El rango de edad de los participantes es de 32 a 41 años; El nivel de educación predominante es el grado Superior (28); la ocupación de los participantes fue: la de empleado(8) , mientras que la mayoría no contestó (30); El medio por el cual se informaron de las actividades fueron: Redes sociales (19), Internet (10) y a través de familiares (9); La mayoría de los participantes asisten por primera vez a la Noche de Museos (35); Así también es mayor el número de quienes no han visitado Otros Espacios (27); El motivo por el que asistieron en este horario es porque es el tiempo de que disponen (19) y es el único horario que existe (17). Los acompañantes en este grupo fueron: amigos (14), familia (11), personas que asistieron solas (9) y pareja (7); Los lugares de procedencia fueron: el hogar y la oficina (ambos con 20); El motivo de asistencia fue principalmente el aprendizaje (23), la curiosidad y diversión (ambos con 8); Día de la semana preferido para la actividad sábado (11) y miércoles (7); Los participantes residen en las alcaldías Miguel Hidalgo (8), Benito Juárez (7) principalmente; Este grupo tuvo participación en los temas “Murales” (17), “Siempre Mujeres” (11) y el “El Mito y la Leyenda” (7).

Grupo IV Azul claro: Es un grupo con 19 (12%) participantes de los cuales 10 son mujeres y 9 hombres. El rango de edad es de 62 a 69 años (como caso extraordinario hubo participantes de 8 (1) y 9 (2) años). El grado de escolaridad del grupo es Superior (13) y Básico (4); La mayoría de los participantes no contesto su ocupación (11), quienes contestaron dijeron que eran empleados y jubilados (ambos con 3). Los medios por los cuales se informaron de las actividades fueron por Redes Sociales (8), Internet (5) y familiares (4). Es la Primera Visita para la mayoría de este grupo (13), algunos manifestaron haber visitado otros espacios (10) y quienes no lo ha hecho (9). El motivo por el que asistieron en este horario fue: “Es el horario que le permite salir con la familia” (7), “Es el único horario que existe para el evento” (7) y “Es el tiempo del que disponen” (4). Los acompañantes de este grupo fueron: familia (7), Amigos (6) y pareja (5); Ellos dijeron provenir de su hogar (15), escuela (1) y de un lugar de esparcimiento (2). El motivo para asistir al evento fue el aprendizaje (16), curiosidad y realizar tareas (1). El día que prefieren para asistir a Noche de Museos es el Miércoles (5), Sábado (4), Viernes (2) y No contestaron (8). Los participantes de este grupo residen en la Ciudad de México en la alcaldía Gustavo A. Madero (4) y el Estado de México (4) mayoritariamente. El Tema de participación fue “Animales Mesoamericanos” (19).

Grupo V Morado: Este grupo se forma con 28 (18%) participantes de los cuales 17 son mujeres, 11 hombres. El rango de edad es de los 43 a 59 años. El grado de estudio predominante es el Superior (22). Las ocupaciones de los participantes fueron las siguientes: empleado (7), Otra actividad (3) y no contestaron (18). Los medios por los que se enteraron fueron: a través de familiares (11), internet (10) y Redes Sociales (7). La mayoría asistió por primera vez a la Noche de Museos (18), y la mitad ha visitado otros espacios (14). En relación al motivo por el cual asistieron en ese horario la mayoría de los que contestaron dijeron que es el tiempo del que disponen (11), otros porque es el único horario que existe para la Noche de Museos (9) y los demás porque les permite salir con la familia (8). Sus acompañantes fueron: familia (17), pareja (5), amigos (4) y solo (1). Los lugares de procedencia fueron: Hogar (15), Oficina (11), y Lugar de esparcimiento (2). Los motivos por los que asistieron fueron: aprendizaje (18), curiosidad (4), diversión (3), convivencia familiar (2), pasar tiempo con la familia y realizar tareas (1). En este grupo el día que prefieren para la Noche de Museos es: viernes (8) y miércoles (5), no contestaron (13). Las Alcaldía en donde residen son: Coyoacán (5), Azcapotzalco (4), Cuauhtémoc (3) y el Estado de México (11). La participación por Tema se distribuye de la siguiente forma: “Murales” (12), “Siempre Mujeres” (9), “Guardianes de leyendas” (4), “Animales Mesoamericanos” (2) y “El Mito y la leyenda” (1).

3.5 Resultados

A partir de los resultados del análisis estadístico y por el análisis multivariado se identificó a los públicos asistentes a los Talleres sabatinos y a las Noches de Museos. En particular se identificó cuál fue el género más participativo, el grupo de edades de los participantes, nivel educativo, actividad, medios de información utilizados para enterarse de los eventos, motivos de participación, frecuencia de asistencia a estas actividades, alcaldía o estado de residencia de los participantes y los factores latentes de participación.

Público de Talleres Sabatinos

¿Quién es el público que asistió a talleres sabatinos?

El público asistente a talleres sabatinos es diverso, no fue posible obtener un grupo con características particulares que lo definieran. En general puede decirse que en su mayoría los participantes son mujeres adultas, cuyo rango de edad se encuentra entre los 42 a 46 años, tienen estudios de Educación Superior, son empleadas, asisten acompañadas por su familia, (en la encuesta no se preguntó quiénes eran los miembros de la familia por lo tanto se desconoce este dato), el motivo de participación en las actividades fue el aprendizaje, en el que consideraron impactó más en el aspecto personal que en el profesional o social.

Se observó que existe una relación entre el nivel educativo de los participantes y el medio de información que utilizaron para conocer las actividades, ya que la mayoría cuenta con Educación Superior y los medios de comunicación utilizados fueron las redes sociales y el internet, lo que hace suponer que los participantes tienen mayor accesibilidad a éstos por su actividad laboral o de estudio.

Se conoce que las actividades en los Talleres les resultaron agradables a los participantes que en su mayoría son empleados y estudiantes, casi el 90% participaron por primera vez, ello debido probablemente a que los talleres se llevan a cabo el fin de semana, que es el disponible de los empleados y en el caso de los estudiantes son los días en que acuden a realizar tareas.

La participación en los diferentes talleres fue diversa, aunque cabe destacar que el taller con mayor participación fue el de Códice Dresde, en el cual se observó una participación casi igual entre mujeres y hombres, de diversas edades. Otro taller que tuvo una alta participación fue el de danza, en éste los participantes fueron en su mayoría mujeres de diversas edades y actividad laboral.

La participación en los talleres se da mayormente una sola ocasión, ya que los asistentes no regresan, esto queda como tema de investigación ya que en la encuesta no se preguntó las causas o factores que intervienen en esta situación, aunque se puede presumir que es por la distancia de residencia hacia el museo.

Las alcaldías de residencia de los participantes son diversas como Gustavo A. Madero en la que residen un gran número de participantes, seguida de Azcapotzalco y Miguel Hidalgo. Se observó que el Estado de México fue la entidad con mayor número de participantes. Esto es un hecho interesante, dado que el museo se ubica en la alcaldía Miguel Hidalgo y se esperaría una mayor participación de residentes de ésta.

Cabe destacar que la alcaldía de residencia tuvo una relación con el tipo de acompañante ya que los residentes de la Alcaldía Gustavo A. Madero y el Estado de México son los que asistieron mayoritariamente acompañados por familiares, mientras que los de Azcapotzalco se acompañaron por la familia y amigos, en tantos que los de la alcaldía Miguel Hidalgo asistieron solos y con la familia.

Agrupación de los participantes.

La siguiente gráfica muestra las agrupaciones de los participantes, esta permite una visualización de los participantes con base en las variables con mayor correlación, se realizó con las componentes principales obtenidas y agregando las variables con correlación más significativa (ver Figura 3.23).

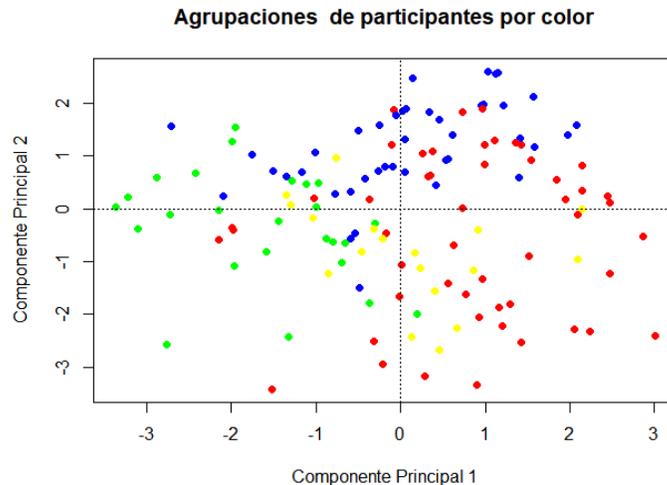


Figura 3.23 Agrupación por color Talleres Sabatinos

La agrupación verde indica a los participantes que tuvieron por motivo el aprendizaje y les gusto el taller, en la agrupación azul están los participantes de primera vez en el taller y se informaron de la actividad a través de las Redes Sociales, la agrupación cuyo color es amarillo indica a los participantes que asistieron acompañados por su familia y tuvieron impacto en el aspecto personal de su vida, por último la agrupación roja indica a participantes que valoraron distintos aspectos a los descritos en las primeras tres agrupaciones.

En conclusión puedo decir que los **Talleres Sabatinos** es una actividad dirigida a todo tipo de público, en el que participan en su mayoría mujeres con su familia, las cuales no están relacionadas profesionalmente con la temática del Museo, pero que posiblemente estén percibiendo esta actividad y el recinto, como un espacio adecuado para el aprendizaje y la convivencia familiar, aunque vale decir que esta encuesta no indagó mayormente sobre los miembros que constituyen a una familia, por lo que se propone una encuesta que permita conceptualizar a los tipos de familia que se atienden en estos talleres.

Público de Noche de Museos

¿Quién es el público que asistió a las Noches de Museos?

El público asistente a las Noches de Museos es mayoritariamente del género femenino, es un público joven, ya que el rango de edad con mayor número de participantes es de 22 a 46 años, en el que destaca la participación de mujeres con edades de 22 a 26 años, con educación predominante en el nivel superior, las principales ocupaciones son empleado y estudiante y asisten acompañados por familia, amigos, parejas y solos, los lugares de procedencia son el hogar y la oficina principalmente.

Los medios de comunicación preferidos, por los cuales se informaron de las actividades fueron las Redes Sociales y el Internet, que se relaciona con el nivel educativo del participante (Como se explicó en Talleres sabatinos).

Los motivos para asistir a la Noche de Museos fueron diversos, el más destacado fue el aprendizaje, así también, se observó que los residentes de las alcaldías Miguel Hidalgo,

Coyoacán, Gustavo A. Madero, Iztapalapa y el Estado de México, tuvieron como principales motivos además del aprendizaje, la curiosidad y la diversión.

Para la mayoría de los asistentes fue la primera participación en la Noche de Museos, esto tuvo relación con el horario de asistencia, ya que como lo manifestaron, la actividad se realizó en un tiempo disponible para que pudieran asistir en su mayoría este público se encuentra en un rango de 22 a 26 años.

Son pocos los participantes que han asistido a otros espacios donde se realizan las Noches de Museos y esto se relaciona con el nivel de escolaridad, puesto que son los participantes con educación superior quienes son los que más han asistido a otros espacios.

En relación con los temas de las Noches de Museos, Siempre Mujeres fue el tema con mayor participación femenina, la mayoría reportó su residencia en la alcaldía Coyoacán, con un nivel de Educación Superior, ocupación de empleadas y estudiantes. El tema de Murales fue otro tema con alta participación, en este fue casi igual la asistencia de hombres y mujeres y, por último el tema de Guardianes de Leyendas que tuvo una presencia semejante de ambos géneros.

La mayoría de participantes en Noche de Museos prefieren que este evento se lleve a cabo los días miércoles en el horario de 19 a 21 horas, tal y como se lleva a cabo en otros recintos.

3.6 Resultados de los Métodos de análisis multivariados.

La aplicación de los métodos de análisis multivariados corroboró los resultados obtenidos con el análisis descriptivo, con estos métodos fue posible encontrar factores latentes, que permitieron una descripción de las características específicas de los encuestados.

Análisis Multivariado Talleres Sabatinos.

Los resultados obtenidos por el método de **análisis factorial** y **componentes principales** para **Talleres Sabatinos**, muestra la agrupación de las variables en dos componentes, en el primer componente se agruparon las variables Motivo por el cual se participa en el taller, Acompañantes de los encuestados, aspecto donde tienen impacto las actividades del taller, qué tan satisfactorio resultó y primera participación. Para el segundo componente las variables que se agruparon son: Número de visitas a los talleres y Medios de información por los cuales los participantes se enteraron de las actividades (ver Tabla 3.3).

1°Componente	2° Componente
Motivo	Número de visitas
Acompañante	Medio
Impacto	
Satisfacción	
Primera Visita	

Tabla 3.3 Agrupamiento de variables en Componentes Talleres Sabatinos

La información obtenida a través del análisis descriptivo con relación a las variables que agrupa el 1° y 2° Componente nos dice que:

1° Componente

El motivo para asistir fue el aprendizaje, seguido de la convivencia familiar.

Los acompañantes fueron principalmente la familia.

Las actividades fueron significativas en el aspecto personal y familiar de los participantes.

Las actividades resultaron agradables y divertidas.

El 84% de los participantes asiste por primera vez y esto se relaciona con el género y la ocupación.

2° Componente

El 16% de los participantes han asistido en más de una ocasión a los talleres.

Los medios de información preferidos por los participantes para enterarse de las actividades fueron las Redes sociales y el Internet.

Esto nos dice que los asistentes prefieren informarse por medio de las redes sociales e internet lo cual se relaciona con el nivel educativo y actividad laboral.

El 60% de los participantes fueron mujeres, que se encuentran en un rango de 42 a 46 años, esto se relaciona con el tipo de acompañante, que mayoritariamente fue la familia, así también con que el motivo de participación en las actividades fue el aprendizaje y la convivencia familiar, resultando para los asistentes agradables y divertidas.

En cuanto a la frecuencia con la que los participantes regresan a los talleres es baja, ya que solo el 16% (22 a 23) de los participantes han asistido más de una vez.

Expuesto lo anterior, se concluye haber encontrado factores latentes que influyen en los asistentes a los talleres sabatinos los cuales son:

- **Nivel educativo y actividad laboral.**
- **Actividades que permitan el aprendizaje y convivencia familiar.**
- **Mujeres que buscan espacios para el aprendizaje y la recreación.**

Análisis multivariado Noche de Museos.

Los resultados obtenidos por el método de **análisis factorial** y **componentes principales** que en ambos casos se igualan en la agrupación de las variables, muestra dos componentes, en el primer componente se agrupan la variables sociodemográficas y las variables relacionadas a la primera visita, el lugar de procedencia, el horario, y la visita a otros espacios.

Quitando las variables demográficas (éstas ya analizadas), tenemos el primer componente con las variables Primera Visita, Otros espacios, Medio de información, Horario de asistencia, Tema de la Noche de Museos; mientras que en el segundo componente se

agrupan las variables que corresponden al Motivo de asistencia, Lugar de procedencia, Acompañantes de los participantes y Día preferido para la actividad (ver Tabla 3.4).

1° Componente	2° Componente
Primera visita	Motivo de asistencia
Otros espacios	Lugar de procedencia
Medio de información	Acompañante
Horario de asistencia	Día preferido para la actividad
Tema de la Noche de Museos	

Tabla 3.4 Agrupamiento de variables en factores.

La información obtenida a través del análisis descriptivo con relación a las variables que agrupa el 1° y 2° Componente nos dice que:

1° Componente

El 79% de los participantes asiste por 1° vez a la Noche de Museos.

El 41% de los asistentes han participado en otros espacios culturales, así también se observa que los participantes con educación superior son los que más asisten a diferentes espacios donde se realizan las Noches de Museos.

Principales medios de información son las **Redes sociales y el Internet.**

La asistencia en el horario de la actividad se corresponde con el **tiempo disponible** de los participantes.

2° Componente

Los motivos son: aprendizaje, curiosidad y diversión.

Los lugares de procedencia son la casa u hogar y el lugar de trabajo.

Los principales acompañantes son: la **familia, amigos y la pareja.**

Los temas con mayor participación son: Siempre mujeres, Murales y Guardianes de Leyenda.

Con base en los componentes, la Noche de Museo es una actividad en la que la mayoría de los participantes ha manifestado asistir por primera vez, esto se relaciona con las actividades de los participantes y el horario, dado que se observa que sus tiempos como empleados y estudiantes son limitados, aunque se observa que un porcentaje de 41%, siguen el programa de Noche de Museos, pues han participado en otros espacios.

La información que se publica a través de las redes sociales es accesible ya que permite a los interesados conocer los temas de noche de museos, horarios y acceso al evento. En este caso estudiado, los participantes han tenido predilección por el tema de **"Siempre mujeres"**, tema que se adapta a las tendencias actuales del reconocimiento a las mujeres y su relación con la visión prehispánica, el otro tema solicitado es el de **"Murales"**, esto puede deberse a la participación del público en un taller, que permite la elaboración de la representación de algún mural del museo, es una actividad interactiva.

Dada las temáticas seleccionadas se puede entender que el público asistente tenga como motivo el conocimiento o aprendizaje no escolarizado, le inspire curiosidad participar en la actividad y tal vez se aprecie como una actividad de aprendizaje y divertida por el horario en el que se realiza y el recinto.

La procedencia de los participantes es principalmente del hogar y el trabajo, esto influye en que los acompañantes sean la familia, los amigos y la pareja. Esto puede estar ligado al horario extraordinario en que se realiza (19:00 horas).

3.7 Instrumento Propuesto

Expuesto lo anterior, se concluye que existen factores latentes, los cuales influyen en la participación de los encuestados a las Noches de Museos, estos son los siguientes:

- Nivel educativo y Actividad laboral
- Actividades que permiten el aprendizaje y la convivencia con seres cercanos.
- Temas de las Noches de Museos, que estén relacionados con las tendencias actuales, así también que involucren actividades didácticas o entretenidas para ellos.
- Horarios alternativos a los horarios de oficinas y escuelas.
- Información de las actividades a través de las Redes sociales y el Internet.

Cabe destacar que además de los factores ya propuestos se puede incluir el hecho que en este programa, como el de talleres sabatinos, la participación mayoritaria corresponde a mujeres jóvenes, lo cual hace suponer que el museo es percibido como un espacio seguro y en el caso de Noche de Museos, para participar en actividades que se realizan en un horario casi nocturno.

En relación a la frecuencia de participación de los asistentes a **Talleres Sabatinos** y **Noches de Museos**, habrá que investigarse más el motivo, por el cual no regresan la mayoría de los participantes.

Dado lo anterior el siguiente paso de este trabajo, es proponer una encuesta que cubra las necesidades planteadas por el Departamento de Comunicación Educativa, que es el de conocer mejor al público que participa en su Talleres Sabatinos y en la Noche de Museos y su información sea eficaz para la aplicación de métodos multivariados y por consecuencia útiles para la planeación de dicho departamento.

Encuesta propuesta

- La formulación de la encuesta se basa en los siguientes pasos:
- Identificación del problema.
- Determinación del diseño de investigación.
- Especificación de las hipótesis.
- Definición de las variables.
- Selección de la muestra.
- Diseño del cuestionario.
- Organización del trabajo de campo.
- Obtención y tratamiento de los datos.

Objetivo de la encuesta: Obtener datos que permitan identificar al público asistente a los Talleres Sabatinos y Noche de Museos.

Diseño de investigación: El diseño de la investigación se basa en un diseño el cual permite la aplicación de un método analítico descriptivo.

Hipótesis

Talleres Sabatinos

- La asistencia a los Talleres es principalmente de mujeres, que asisten con su familia en particular con sus hijos.
- La asistencia de varones es muy baja
- El público participa una sola vez y no regresa a las actividades.

Noche de Museos

- El público asistente a la Noche de Museos, principalmente es de mujeres jóvenes.
- El público a la Noche de Museos asiste por razones de aprendizaje y recreación.

Variables de estudio

- Variables socio demográficas:
Género, edad, nivel de educación, ocupación, lugar de residencia.
- Variables motivo por el cual asisten al evento
Motivo de asistencia, frecuencia de asistencia, cómo eligen el tema, con quiénes asiste
- Variables que permitan conocer cómo es percibido el espacio cultural
Cómo se percibe al museo en cuestión de seguridad, costo de entrada y de las actividades y transporte que utilizan los visitantes para llegar al museo.

Selección de la muestra: La muestra será tomada de manera aleatoria de la información obtenida de las respuestas de las encuestas correspondientes a Talleres sabatinos y Noche de Museos y deberá ser al menos 10 individuos por tema de las actividades realizadas.

Diseño del cuestionario: El diseño del cuestionario se basa en el objetivo a investigar que es el mismo para ambos programas Talleres sabatinos y Noche de Museos, por lo que se diseñó un instrumento que puede ser aplicado a ambos públicos, el cual es el siguiente:

Museo Nacional de Antropología

Departamento de Comunicación Educativa

Lugar de aplicación: Museo Nacional de Antropología.

Ciudad de México

Fecha: _____.

No. de cuestionario: _____.

Público asistente

Este cuestionario tiene como objetivo conocer al público que asiste y participa en las actividades del departamento, conocer los motivos por los cuales asiste, la frecuencia a la que asiste a las actividades y cuál es su opinión del costo del museo.

Datos generales:

Sexo: _____; Edad: _____ años; Estado civil: _____.

Lugar de residencia: CDMX / Estado de México/ Otro estado;

Alcaldía de la CDMX: _____.

En las siguientes preguntas marque con una X una sola respuesta.

Escolaridad:

Primaria	Secundaria	Medio Superior	Superior
Especialidad	Posgrado	Educación técnica	Sin escolaridad

Ocupación:

Profesionista	Técnico	Trabajador de la educación	Trabajador del arte y/o deporte
Trabajador del sector público, privado o social	Artesano	Operador de maquinaria	Conductor de transporte
Comerciante	Trabajador en servicios domésticos	Trabajador en servicios de protección, vigilancia y fuerzas armadas	Trabajador en servicios de la salud
Estudiante	Empresario	Investigador	Trabajador con ocupación no clasificada

ITÉM	Respuesta 1	Respuesta 2	Respuesta 3	Respuesta 4	Respuesta 5
1.- ¿Qué tema eligió?	Tema1	Tema2	Tema3	Tema4	Tema5
2.- ¿Por qué eligió este tema?	Aprender de la cultura	Realizar la actividad en el taller	Realizar una tarea escolar	Por convivencia familiar	Otro
¿Quién eligió el tema?	Usted	Familia	Amigos	La escuela donde estudia mi hijo o yo / la institución donde laboro	Compañeros
4.- ¿Con quién o quiénes asiste a la actividad?	Solo	Familia (Hijos / Padres / Pareja / Parientes)	Amigos	Compañeros de escuela o trabajo	Otro, ¿cuál?
5.- ¿Cómo se enteró de las actividades?	Redes sociales / Pagina del museo	Recomendación de personas que han participado en las actividades	Por la institución donde estudian mis hijos o yo	Por la institución donde laboro	Otro, ¿cuál?
6.- Con cuánta frecuencia participa en las actividades?	Es la primera y única vez en el año	Es la segunda vez en el año	Son más de dos veces en el año	No, lo se	Otro, ¿cuál?
7.- ¿Considera que el museos es un lugar :	Seguro para asistir solo o con la familia	Un lugar seguro y accesible para asistir con niños	Un lugar que propicia la convivencia familiar o social	Ninguno de los anteriores	Otro, ¿Cuál?
9.- ¿Cómo considera que es el costo de entrada y actividades en el museo?	Adecuada a su presupuesto	Excesiva para su presupuesto	Bajo para su presupuesto	No le cuesta	Otra, ¿Cuál?
10.- El traslado desde su lugar de procedencia al museos es:	De fácil acceso en vehículo privado	De fácil acceso en transporte público	De fácil acceso caminando o en bicicleta	Ninguna de las anteriores	Otra, ¿Cuál?

Organización del trabajo

La encuesta puede ser aplicada por personal de servicio social, el encuestador debe conocer el objetivo que se tiene para realizar la encuesta, y poder comunicárselo al encuestado así como aclarar cualquier duda.

La aplicación de la encuesta no debe tardar más de 3 minutos, se deben contestar todos los campos y se debe aplicar al inicio de la actividad cuando se considera que los asistentes pueden contestar.

Debe haber un responsable para entregar y recibir las encuestas a los aplicadores, el cual tendrá a su cargo la captura de los datos.

Obtención y Tratamiento de los datos.

La obtención de los datos será a través de las respuesta de la encuesta aplicada, capturada en Excel de forma completa y ordenada de acuerdo con la secuencia de las preguntas, una vez obtenida la información el siguiente punto es el análisis estadístico y posteriormente el análisis multivariado que puede ser aplicado por Excel Rstudio o cualquier otro software de aplicación estadística.

Una vez realizado el análisis lo siguiente es la elaboración de un reporte con los resultados, el cual debe ser de forma clara y sencilla para ser entendido por cualquier persona.

Capítulo 4. Conclusiones

4.1 Parte técnica

El desarrollo de la parte técnica fue en suma complicado, debido a la naturaleza de las encuestas, ya que los resultados de éstas, no tuvieron una escala de medición uniforme y numérica, por lo que fue necesario convertirlos a factores numéricos que permitieran la aplicación de los métodos de análisis multivariado. Debido a lo anterior los resultados, en los criterios para la aplicación de los métodos apenas resultaron suficientes, la matriz de correlaciones indicó valores menores de 0.30, por lo tanto se puede decir que no existe correlación entre las variables. Por lo que se requiere una mejor estructuración de la encuesta y una aplicación por un tiempo más largo, para tener resultados más consistentes y de mayor profundidad.

4.2 Análisis Factorial

La aplicación del análisis factorial, fue solo de tipo exploratorio, ya que la información de ambas encuestas, no fue propicia para cumplir con el criterio de Test de esfericidad de Bartlett, pues en ambos casos fue menor al valor de significancia, el otro criterio fue el índice KMO en el cual se obtuvo que para TS-19 fue: 0.51 y para NM19-20 fue: 0.50, lo que indicó que el análisis era aceptable medianamente.

Como resultado del análisis factorial exploratorio, se encontraron factores latentes, que con la información obtenida del análisis descriptivo, fueron de ayuda para la descripción del público estudiado.

4.3 Componentes Principales

En el método de componentes principales, no fue posible la obtención de grupos naturales que permitieran identificar características específicas de los participantes, más sin en cambio, fue posible corroborar los resultados obtenidos con el análisis factorial, ya que la agrupación de las variables fue idéntica al exploratorio factorial, lo que permitió la obtención de factores latentes.

He de decir, que se siguieron todos los criterios, para la aplicación del método y estos estuvieron en los criterios mínimos, es decir que la aplicación del método era medianamente aceptable, por lo tanto se aplicó.

4.4 Análisis de Clústeres o Conglomerados.

Con el método de Conglomerados, fue posible obtener una agrupación de los participantes en Talleres Sabatinos y Noche de Museos, esta agrupación se realizó siguiendo los criterios de agrupación jerárquica, en la que se aplicó el Método Ward's, resultando el más conveniente ya que el coeficiente de agrupamiento obtenido fue de 0.5562 lo que indica una estructura de agrupación fuerte en comparación con la agrupación por vecino más cercano, lejano y promedio. De entre los métodos de análisis multivariado aplicados a este estudio, este fue el que permitió una agrupación de manera más rápida, natural y más explicativa.

4.5 Conclusión general.

Como conclusión general referente a la parte técnica de la aplicación de los métodos, la naturaleza de la información fue difícil y obligó a buscar alternativas del cálculo de la correlación, por lo cual se calcularon correlaciones de tipo Policórica, Tetracórica, y Punto Biserial, esto permitió el cálculo de la matriz de correlación la cual fue la base para la aplicación de los métodos multivariados, los cuales se aplicaron, no siendo posible obtener resultados óptimos, pero si resultados eficientes, es decir que con la aplicación de los métodos se logró obtener información que complementara la obtenida a través del análisis descriptivo.

Por lo anterior concluyo que los métodos multivariados aplicados en este trabajo, son métodos que se pueden aplicar a bases de datos que contengan factores y no solo datos de naturaleza numérica, pero esta base tiene que tener factores de proporciones adecuadas, permitiendo así un mejor cálculo en las correlaciones, para la eficacia de los métodos multivariados.

En general este trabajo representó un reto desde generar un instrumento que pudiera dar respuesta a las situaciones planteadas de conocer al público y conocer las causas de las nuevas dinámicas de asistencia, el manejo de la información obtenida y el procesamiento de la misma hasta comprender y dar a conocer de forma sencilla los resultados.

Reconozco que este tipo de trabajos en el que se ve involucrado el estudio de aspectos sociales, personales de individuos requiere un periodo de estudio de mediano y largo plazo, así como el acompañamiento de otras disciplinas como la sociología, psicología y pedagogía que enriquezcan el análisis de los datos estadísticos, que en este caso se hizo en base a mi experiencia y conocimiento.

Bibliografía

- Alvin, C. R. (2002). *Methods of Multivariate Analysis*. Canada: Wiley Interscience.
- Everitt, B. S. (2001). *Applied Multivariate Data Analysis*. London: John Wiley & Sons Ltd.
- García Canclini, N. (1991). *El consumo cultural en México*. Ciudad de México: Grijalbo, CNCA.
- INAH. (2017). *Manual de Organización del Instituto Nacional de Antropología e Historia*. Ciudad de México: INAH.
- Monzón, A. (1952). *Bases para incrementar el público que visita el Museo Nacional de Antropología*. Ciudad de México: INAH.
- Olivé Negrete, J. (1995). *INAH una historia*. Ciudad de México: INAH.
- Pastor Homs, M. I. (1992). *El Museos y la comunicación en la comunidad*. Barcelona: Ediciones ceac.
- Pérez Castellanos, L. (s.f.). *Público y Museos ¿qué hemos aprendido?*
- Pérez Castillo, L. (2018). *Buenas prácticas en la investigación del público en museos*.
- Pérez, Castellanos Leticia. (2016). Público y museos :¿qué hemos aprendido? *Estudios sobre públicos y museos*, 21- 42.
- Ramírez Vázquez, P. (2012). *Museo Nacional de Antropología*. Ciudad de México: INAH.
- Ramírez Vázquez, P. (2012). *Museo Nacional de Antropología* . Ciuda de México: INAH.
- Schmilchuk, G. (1996). *Venturas y desventuras de los estudios de público*. México: CENIDIAP.
- Serra Puche, M. C. (1997). El Museo Nacional de Antropología. *Arqueología Mexicana* 6.
- Vallejo, M. E. (2003). Los servicios educativos del INAH. En M. E. Vallejo, *Una Historia de cincuenta años* (págs. 75-86). Ciudad de México: INAH.
- Vázquez, M. (2015). *Obra museológica y museográfica*.
- Véliz, C. (2016). *Análisis multivariante métodos estadísticos multivariantes para la investigación*. Ciudad de México: CENGAGE Learning.
- Zamora, S. (2016). *Notas de Clases Análisis Multivariado*. Ciudad de México .

Anexo I

En este anexo se presenta el resultado numérico obtenido del cálculo de la matriz de correlación y los métodos de análisis multivariado.

Matriz de correlación Talleres Sabatinos.

Corr. Sabatinos	Taller																	
Taller	1.000																	
Edad	0.124	1.000																
Sexo	0.036	-0.025	1.000															
Ocupación	0.071	-0.069	-0.175	1.000														
Escolaridad	-0.297	0.088	0.021	-0.201	1.000													
Medio	-0.462	-0.158	-0.084	-0.272	0.337	1.000												
PV	-0.174	-0.108	0.241	0.290	-0.058	0.035	1.000											
NV	0.137	0.137	-0.074	-0.055	-0.028	-0.105	0.000	1.000										
Motivo	0.017	-0.069	-0.095	0.120	-0.090	-0.234	0.065	-0.113	1.000									
Acompañante	-0.136	0.030	0.224	0.106	0.199	0.128	-0.048	-0.002	-0.243	1.000								
Impacto	-0.060	-0.048	-0.011	-0.061	0.182	0.141	-0.101	0.165	-0.270	0.222	1.000							
Satisfacción	-0.096	-0.078	-0.213	0.032	-0.071	0.150	-0.035	-0.165	0.351	-0.305	-0.345	1.000						
Alcaldía	-0.269	0.067	-0.122	-0.114	0.086	0.150	0.058	-0.141	0.045	-0.069	-0.214	0.314	1.000					

Matriz de correlación Noche de Museos.

Matriz de correlación

Corr. NM	Tema																		
Tema	1.0000																		
Edad	0.0550	1.0000																	
Sexo	0.0480	0.0280	1.0000																
Escolaridad	0.1081	-0.0493	0.2047	1.0000															
Ocupación	-0.0535	0.0960	0.0704	-0.1335	1.0000														
Medio	0.0107	0.0556	0.0359	0.2134	-0.0236	1.0000													
PV	-0.0974	-0.2765	0.0081	-0.0197	-0.0337	0.0387	1.0000												
OE	-0.0567	0.1133	-0.1288	0.2351	0.0039	0.1283	-0.0071	1.0000											
Horario	-0.1006	0.0048	0.1363	0.1918	0.0105	0.1052	0.2844	0.1899	1.0000										
Acompañante	-0.1154	0.0269	0.0325	0.0551	0.2098	0.0962	0.2291	-0.0284	0.0131	1.0000									
Lugar	-0.0203	-0.0163	0.2382	-0.0711	-0.3010	0.0428	0.1653	0.0353	0.1890	-0.0931	1.0000								
Motivo	-0.0539	-0.0183	0.0387	0.0070	0.0046	-0.0741	0.0988	-0.0648	0.0506	-0.0051	0.0603	1.0000							
Día	-0.1077	0.1013	0.1211	-0.0238	0.2138	0.0747	-0.0760	-0.0104	-0.1006	-0.0101	0.0771	-0.1042	1.0000						
Alcaldía	0.1389	-0.0228	0.0756	0.0076	0.2601	0.0875	-0.0784	0.0129	0.0023	-0.1222	-0.0814	-0.1711	0.0991	1.0000					

La matriz de correlación para ambos casos, se calculó con base en las correlaciones Policórica, Tetracórica y Punto Biserial.

Análisis factorial exploratorio.

Test de esfericidad de Bartlett

Talleres Sabatinos

Test de esfericidad de Bartlett	Chi - cuadrado	200.9759
	G. libertad	78
	p - valor	0.000000

Noche de Museos

Test de esfericidad de Bartlett	Chi - cuadrado	Inf
	G. libertad	91
	p - valor	0.000000

Índice de Kaiser – Meyer – Olkin

Test KMO mide la correlación y la correlación parcial entre dos variables

Rangos:

Si $KMO < 0.5$ la aplicación del análisis es inadecuado.

Si $0.5 \leq KMO < 0.7$, se acepta medianamente la aplicación.

Si $KMO \geq 0.7$, se acepta la conveniencia de la aplicación del análisis.

Índice KMO Talleres Sabatinos = 0.50; El análisis se acepta medianamente la aplicación.

Medidas de adecuación

<i>Tema</i>	<i>Edad</i>	<i>Sexo</i>	<i>Escolaridad</i>	<i>Ocupación</i>	<i>Medio</i>	<i>Día</i>
0.50	0.50	0.50	0.50	0.50	0.50	0.50
<i>PV</i>	<i>OE</i>	<i>Horario</i>	<i>Acompañante</i>	<i>Lugar</i>	<i>Motivo</i>	<i>Alcaldía</i>
0.50	0.50	0.50	0.50	0.50	0.50	0.50

Índice KMO Noche de Museos = 0.50; El análisis se acepta medianamente la aplicación.

Medidas de adecuación

<i>Tema</i>	<i>Edad</i>	<i>Sexo</i>	<i>Escolaridad</i>	<i>Ocupación</i>	<i>Medio</i>	<i>Día</i>
0.50	0.50	0.50	0.50	0.50	0.50	0.50
<i>PV</i>	<i>OE</i>	<i>Horario</i>	<i>Acompañante</i>	<i>Lugar</i>	<i>Motivo</i>	<i>Alcaldía</i>
0.50	0.50	0.50	0.50	0.50	0.50	0.50

Tabla de componentes y eigenvalores TS-19

Componente	Eigenvalores	% de Varianza	% de Varianza Acumulada
1	2.215	17.04	17.04
2	2.056	15.82	32.86
3	1.469	11.30	44.16
4	1.173	9.02	53.18
5	1.072	8.25	61.42
6	0.986	7.58	69.01
7	0.864	6.65	75.65
8	0.700	5.39	81.04
9	0.652	5.02	86.05
10	0.568	4.37	90.42
11	0.552	4.24	94.67
12	0.421	3.24	97.91
13	0.272	2.09	100.00

Tabla de componentes y eigenvalores NM19-20

Componentes	Eigenvalores	% de Varianza	% de Varianza Acumulada
1	2.370	16.93	16.93
2	1.773	12.66	29.59
3	1.478	10.56	40.15
4	1.343	9.60	49.75
5	1.121	8.01	57.76
6	1.065	7.61	65.37
7	0.956	6.83	72.19
8	0.884	6.31	78.50
9	0.746	5.33	83.83
10	0.718	5.13	88.96
11	0.597	4.26	93.22
12	0.503	3.59	96.82
13	0.446	3.18	100.00
14	0.000	0.00	100.00

Tablas correspondientes a los componentes y eigenvalores para Talleres sabatinos y Noche de Museos.

Comunalidades

Variables	Inicial	Extracción
Tema	1.00	0.56
Edad	1.00	0.04
Sexo	1.00	0.10
Ocupación	1.00	0.16
Escolaridad	1.00	0.42
Medio	1.00	0.65
PV	1.00	0.01
NV	1.00	0.16
OE	1.00	0.41
Motivo	1.00	0.34
Acompañante	1.00	0.42
Impacto	1.00	0.62
Satisfacción	1.00	0.39

Comunalidades Talleres Sabatinos

Variables	Inicial	Extracción
Tema	1.00	0.001
Edad	1.00	0.075
Sexo	1.00	0.014
Escolaridad	1.00	0.349
Ocupación	1.00	0.358
Medio	1.00	0.141
PV	1.00	0.878
OE	1.00	0.878
Horario	1.00	0.191
Acompañante	1.00	0.198
Lugar	1.00	0.310
Motivo	1.00	0.580
Día	1.00	0.068
Alcaldía	1.00	0.101

Comunalidades Noche de Museos

Carga de factores.

Variables	Factores o Componentes	
	1	2
Tema	-0.126	-0.336
Edad	0.020	-0.090
Sexo	0.115	-0.089
Ocupación	-0.139	-0.122
Escolaridad	0.228	0.196
Medio	0.197	0.329
PV	-0.045	0.032
NV	0.075	-0.179
OE	-0.284	0.055
Motivo	0.262	-0.039
Acompañante	0.282	-0.088
Impacto	-0.262	0.257
Satisfacción	-0.083	0.291

Carga de factores Talleres Sabatinos

Variables	Factores o Componentes	
	Factor 1	Factor 2
Tema	-0.035	-0.015
Edad	0.131	0.241
Sexo	-0.033	-0.112
Escolaridad	0.542	-0.237
Ocupación	0.137	-0.582
Medio	0.350	-0.136
PV	0.885	0.308
OE	0.885	0.308
Horario	0.334	0.282
Acompañante	0.136	-0.423
Lugar	-0.142	0.538
Motivo	-0.425	0.632
Día	-0.072	-0.251
Alcaldía	0.115	-0.297

Carga de factores Noche de Museos

Componentes Principales

Eigenvalores y Varianza.

Los eigenvalores se relacionan con la varianza explicada y permiten determinar el número de componentes principales adecuados. Las siguientes Tablas () muestran la desviación estándar que equivale al valor del eigenvalor de cada componente principal (PC), el número de componentes principales a elegir se determina por los valores de la desviación estándar que sean mayores a uno.

En la encuesta de Talleres Sabatinos se tiene tres componentes principales cuyos eigenvalores son (1.737, 1.178, 1.028) los cuales corresponden a los tres primeros componentes principales. Los valores de porcentaje de la varianza indican la cantidad de información que se mantiene en cada componente principal (CP), al elegir los tres primeros componentes la varianza acumulada mantiene un 78% de información la cual es suficiente para el análisis.

Resumen de las Componentes Principales							
Componente Principal	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Desviación estándar	1.737	1.178	1.028	0.861	0.645	0.505	0.355
Proporción de varianza %	43%	20%	15%	11%	6%	4%	2%
Varianza acumulada en %	43%	63%	78%	89%	95%	98%	100%

Tabla de distribución de varianza de los Componentes principales de Talleres Sabatinos.

Para la encuesta de Noche de Museos se obtuvieron 4 componentes principales, cuyos eigenvalores son (1.3927, 1.2122, 1.1062 y 1.0664), los cuales corresponden a las cuatro primeras componentes, la varianza acumulada mantiene un 82% de información con una

pérdida de 18% del total de información. Las tres primeras componentes contienen el 66% de información con una pérdida del 44%, por lo que no es recomendable, ya que hay una pérdida considerable de información.

Componentes Principales	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Desviación estándar	1.3927	1.2122	1.1062	1.0664	0.8054	0.7624	1.49E-17
Proporción de varianza %	28%	21%	17%	16%	9%	8%	0%
Varianza acumulada en %	28%	49%	66%	82%	92%	100%	100%

Tabla de distribución de varianza de los Componentes principales Noche de Museos.

Correlación entre componentes principales y variables.

El primer componente (CP1) tiene una correlación alta y directa con las variables Impacto (0.844), Acompañante (0.764) y una correlación alta e indirecta con las variables Motivo (-0.856), Satisfacción (-0.842). La segunda componente principal (PC2) tiene correlación alta y directa con las variables: Número de Visitas (NV), y correlación alta e indirecta con la variable Medio (-0.84), la tercera componente principal (PC3) tiene correlación alta e indirecta con la variable Primera Visita (PV).

Correlación Componentes Principales vs Variables			
Variables	PC1C	PC2	PC3
Medio	0.254	-0.843	0.108
PV	-0.223	0.031	-0.922
NV	0.408	0.694	0.231
Motivo	-0.856	0.194	0.057
Acompañante	0.764	-0.229	-0.184
Impacto	0.844	0.009	0.133
Satisfacción	-0.842	-0.322	0.295

Tabla Correlación de Componentes Talleres Sabatinos.

Noche de Museos.

El primer componente (CP1) tiene una correlación alta y directa con las variables Acompañante (0.826), Medio (0.528) y una correlación alta indirecta con la variable Motivo (-0.557), el segundo componente (CP2) tiene correlación alta y directa con las variables Horario (0.661), y una correlación alta e indirecta con la variable Tema (-0.785) en la componente tres (CP3) la variables Lugar (0.808) y Día (0.596) tienen una correlación alta y directa.

Correlación Componentes Principales vs Variables			
Variables	CP1	CP2	CP3
Tema	-0.428	-0.785	-0.236
Medio	0.528	0.253	-0.211
Horario	-0.413	0.661	0.140
Acompañante	0.826	0.257	-0.130
Lugar	-0.316	0.171	0.808
Motivo	-0.557	0.290	-0.281
Día	0.462	-0.416	0.596

Tabla de correlaciones Noche de Museos

Eigenvectores o autovalores.

Los valores que corresponden a los eigenvectores permiten estimar las componentes principales. Estos datos sirven para el cálculo de coordenadas, que se construyen a través del siguiente modelo matemático.

Talleres Sabatinos.

Matriz de eigenvectores o autovectores de los componentes principales							
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Taller	-0.3084	-1.4951	0.6740	-0.5978	0.0370	-0.9303	0.3318
Edad	0.4315	-0.0380	0.5089	-0.3361	0.3695	-0.2371	0.1709
Sexo	0.4218	-0.1215	-0.9479	-0.5260	0.1369	-0.3881	0.2295
Ocupación	-0.5833	-0.5474	-0.8542	-0.5458	-0.0307	-0.2343	-0.2288
Escolaridad	0.8167	0.7876	0.3469	-0.0424	-0.1238	0.0251	0.3134
Medio	0.7437	2.5360	0.3897	1.1777	0.0717	0.2006	0.4736
PV	-0.8165	-0.1967	-2.8864	1.0831	-0.1949	0.0322	-0.1201
NV	1.2929	-2.4228	0.8514	1.6122	0.8051	0.5212	-0.0525
Motivo	-3.1222	-0.7357	0.3446	-0.8261	-0.7785	1.0931	0.3628
Acompañante	2.5046	0.6858	-0.4005	-1.6054	0.8901	0.6611	-0.3652
Impacto	2.8375	-0.1468	0.6177	0.1740	-1.7326	-0.1076	-0.4015
Satisfacción	-3.0837	0.9981	1.1681	0.3790	0.2542	-0.1877	-0.7884
Alcaldía	-1.1346	0.6966	0.1876	0.0534	0.2960	-0.4482	0.0745

Modelo Matemático.

Las componentes principales son:

$$Y_1 = -0.3084Z_{r1} + 0.4315Z_{r2} + 0.4218Z_{r3} - 0.5833Z_{r4} + 0.8167Z_{r5} + 0.7437Z_{r6} \\ - 0.8165Z_{r7} + 1.2929Z_{r8} - 3.1222Z_{r9} + 2.5046Z_{r10} + 2.8375Z_{r11} \\ - 3.08372Z_{r12} - 1.1346Z_{r13}$$

$$Y_2 = -1.4951Z_{r1} + -0.0380Z_{r2} + -0.1215Z_{r3} - 0.5474Z_{r4} + 0.7876Z_{r5} + 2.5360Z_{r6} \\ - 0.1967Z_{r7} - 2.4228Z_{r8} - 0.7357Z_{r9} + 0.6858Z_{r10} - 0.1468Z_{r11} \\ + 0.9981Z_{r12} + 0.6966Z_{r13}$$

Cada r corresponde al número de observación de los datos, siendo $r = 1, \dots, 138$. En esta relación la variable Z_i es la estandarización de la variable X_i , la cual corresponde a la información contenida en la base.

La variable Z_i se obtiene de la siguiente formula:

$$Z_i = \frac{X_i - \bar{x}_i}{\sigma_i}$$

Donde:

X_i : Corresponde al valor de la observación en la encuesta.

\bar{x}_i : Es el promedio de los datos.

σ_i : Es la varianza de los datos.

Con los datos estimados en la matriz de eigenvectores o autovectores y la normalización de los datos originales de acuerdo con la formula anteriormente presentada fue posible la aplicación de las ecuaciones lineales Y_1, Y_2 de las cuales se obtuvieron los sitios en el nuevo sistema de ejes (Componentes principales) Estos sitios son las coordenadas de los datos originales y son conocidos como valores de las componentes principales y corresponden a los resultados de las Y_{ij} , donde $i = 1$ o 2 y $j = 1, \dots, 138$.

Noche de Museos.

Matriz de eigenvectores o autovectores de las componentes principales							
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Tema	-0.30727	-0.64790	-0.21334	0.33005	-0.14466	-0.10874	0.54645
Medio	0.37880	0.20831	-0.19099	0.52645	0.62772	-0.27045	0.17979
Horario	-0.29691	0.54526	0.12682	0.40591	-0.17218	0.53414	0.34459
Acompañante	0.59312	0.21172	-0.11793	-0.23269	-0.49048	-0.17194	0.51498
Lugar	-0.22726	0.14072	0.73051	0.06101	-0.04338	-0.60421	0.15569
Motivo	-0.40029	0.23929	-0.25376	-0.59523	0.42874	-0.08425	0.41581
Día	0.33208	-0.34298	0.53850	-0.19233	0.35934	0.47749	0.29662

Modelo Matemático.

Las componentes principales son:

$$Y_1 = -0.3072Z_{r1} + 0.3788Z_{r2} - 0.2969Z_{r3} + 0.5931Z_{r4} - 0.2272Z_{r5} - 0.4003 + 0.3321Z_{r7}$$

$$Y_2 = -0.6479Z_{r1} + 0.2083Z_{r2} + 0.5453Z_{r3} + 0.2117Z_{r4} + 0.1407Z_{r5} + 0.2393 - 0.3429Z_{r7}$$

Cada r corresponde al número de observación de los datos, siendo $r = 1, \dots, 203$. En esta relación la variable Z_i es la estandarización de la variable X_i , la cual corresponde a la información contenida en la base.

La variable Z_i se obtiene de la siguiente formula:

$$Z_i = \frac{X_i - \bar{x}_i}{\sigma_i}$$

Donde:

X_i : Corresponde al valor de la observación en la encuesta.

\bar{x}_i : Es el promedio de los datos.

σ_i : Es la varianza de los datos.

Con los datos estimados en la matriz de eigenvectores o autovectores y la normalización de los datos originales de acuerdo con la fórmula anteriormente presentada fue posible la aplicación de las ecuaciones lineales Y_1, Y_2 de las cuales se obtuvieron los sitios en el nuevo sistema de ejes (Componentes principales) Estos sitios son las coordenadas de los datos originales y son conocidos como valores de las componentes principales y corresponden a los resultados de las Y_{ij} , donde $i = 1$ o 2 y $j = 1, \dots, 203$.

Los valores de Y_1 y Y_2 con respecto a cada observación, corresponden a las coordenadas asociadas a cada componente de acuerdo con los modelos matemáticos presentados para Talleres Sabatino y Noche de Museos, con esto es posible reducir el espacio de 13 y 14 dimensiones originales (correspondiente a las variables) a dos dimensiones, facilitando la visualización y análisis de los datos.

Las siguientes gráficas muestran los valores de Y_1 y Y_2 , obtenidos para Talleres Sabatinos y Noche de Museos. Es importante aclarar que en cada tabla se presentan algunas de las observaciones ya que la cantidad de información es demasiada y resulta difícil la visualización de esta, por lo que se eligieron algunos de los primeros y últimos datos.

Talleres Sabatinos.

Valores de los componentes principales Y1 y Y2		
Observaciones	Y1	Y2
1	-0.1726	-1.1843
2	3.3959	-0.6941
3	-0.9116	-0.9084
4	-0.5330	-1.2852
5	1.0184	1.3458
6	-0.8040	1.8115
7	1.3181	-0.2074
8	0.8842	1.2728
9	-0.3139	-1.4601
10	-1.9586	0.4615
101	0.1297	-0.8842
102	-0.2223	-2.0057
103	-1.9945	-0.3928
104	2.2717	2.5799
105	0.7658	0.4354
106	2.4442	0.1901
107	1.3819	-2.4295
108	-1.6045	-2.3577
130	2.3330	-0.4929
131	-0.6050	1.3089
132	0.8849	0.6380
133	0.4108	0.6906
134	-0.0510	0.9639
135	-0.2000	-0.6322
136	0.5787	-4.1034
137	-2.2489	1.5349
138	-2.3762	2.2350

Noche de Museos.

Valores de las componentes principales Y1 y Y2		
Observaciones	Y1	Y2
1	5.21	6.08
6	5.29	3.91
9	1.65	1.76
10	0.52	4.66
11	3.37	5.21
12	1.02	4.22
13	1.39	4.21
15	2.45	2.53
16	1.05	2.08
17	0.38	2.15
18	1.20	4.59
19	0.28	3.67
20	-0.86	4.03
21	1.89	4.09
43	0.77	1.83
44	0.47	4.50
45	3.09	2.92
46	0.34	2.29
47	3.71	2.73
48	0.66	1.84
49	0.48	4.33
57	3.51	1.67
58	-1.35	0.37
84	2.07	4.97
85	1.64	2.98
86	3.51	2.12
87	2.90	2.79
88	5.35	2.71

Análisis de Conglomerados o Clústeres.

En análisis comienza con el cálculo de la matriz de distancias euclidianas entre las variables.

Las Tablas 1 y 2, muestran la matriz de distancias euclidianas en las variables de la encuesta Talleres Sabatinos y Noche de Museos, los elementos cuya distancia es menor se agruparan.

Talleres Sabatinos.

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Taller	Edad	Sexo	Ocupación	Escolaridad	Medio	PV	NV	Motivo	Acompañante	Impacto	Satisfacción	Alcaldía
1 Taller	0												
2 Edad	1.412	0											
3 Sexo	1.615	1.604	0										
4 Ocupación	1.479	1.482	1.729	0									
5 Escolaridad	2.106	1.485	1.554	1.927	0								
6 Medio	2.310	1.611	1.737	2.034	1.009	0							
7 PV	1.835	1.696	1.270	1.205	1.684	1.667	0						
8 NV	1.364	1.358	1.580	1.593	1.622	1.789	1.592	0					
9 Motivo	1.605	1.711	1.784	1.393	1.881	1.995	1.494	1.767	0				
10 Acompañante	1.865	1.314	1.275	1.606	1.268	1.477	1.638	1.535	2.051	0.000			
11 Impacto	1.778	1.495	1.542	1.771	1.301	1.481	1.759	1.291	2.090	1.175	0.000		
12 Satisfacción	1.897	1.715	1.983	1.647	1.811	1.631	1.657	1.910	1.059	2.136	2.189	0.000	
13 Alcaldía	2.040	1.619	1.772	1.751	1.470	1.358	1.510	1.808	1.476	1.791	1.938	1.095	0.000

Matriz de distancias Talleres Sabatinos.

Noche de Museos.

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Tema	Edad	Sexo	Escolaridad	Ocupación	Medio	PV	OE	Horario	Acompañante	Lugar	Motivo	Día
1 Edad	1.2657	0.0000											
2 Sexo	1.4117	1.4407	0.0000										
3 Escolaridad	2.0948	1.8533	1.7048	0.0000									
4 Ocupación	1.5000	1.2660	1.6779	2.2208	0.0000								
5 Medio	1.7737	1.7525	1.6040	1.0944	1.9109	0.0000							
6 PV	1.8465	2.0692	1.6706	1.5721	2.0062	1.4829	0.0000						
7 OE	1.8281	1.7244	1.6285	1.1205	2.0688	1.1402	1.5402	0.0000					
8 Horario	1.9023	2.0666	1.9449	1.2741	2.2374	1.3960	1.0883	1.2632	0.0000				
9 Acompañante	1.8959	1.5742	1.5156	1.5974	1.4460	1.3536	1.5552	1.5870	1.7741	0.0000			
10 Lugar	1.7741	1.8075	1.5991	1.4041	2.3976	1.5005	1.3539	1.3263	1.2970	1.7607	0.0000		
11 Motivo	1.7318	1.9157	1.6766	1.5134	1.9704	1.6304	1.2593	1.5671	1.4195	1.7051	1.4778	0.0000	
12 Día	1.5630	1.3598	1.3133	1.7363	1.3882	1.4838	1.5785	1.6872	1.8050	1.3700	1.5415	1.6946	0.0000
13 Alcaldía	1.3750	1.5328	1.5633	1.7371	1.4753	1.4618	1.8569	1.5645	1.6927	1.5890	1.8054	1.7020	1.4872

Matriz de distancias Noche de Museos 1

Una vez que se han calculado las distancias, el siguiente punto es decidir cuál de los cuatro métodos: Vecino más cercano, Vecino más lejano, Promedio y Ward's es el más conveniente para ser aplicado. Para esto es necesario calcular el coeficiente de

aglomeración que mide la cantidad de estructura de agrupamiento cuyos valores se encuentran entre el intervalo (0,1) esto ayuda a decidir que método tiene una estructura de agrupación fuerte, así el método cuyo coeficiente de agrupación este cercano a 1 será el más eficaz.

Coeficiente de Agrupación

Talleres Sabatinos

Coeficientes de agrupamiento			
Vecino más cercano	Vecino más lejano	Promedio	Ward
0.1413	0.4660	0.3179	0.5562

Coeficientes de agrupación Talleres Sabatinos.

Los resultados muestran que el coeficiente de agrupación cercano a 1 es el coeficiente correspondiente al método Ward's con un valor de 0.5562, por lo tanto es el método a aplicar.

Noche de Museos.

Coeficientes de agrupamiento			
Promedio	Vecino más cercano	Vecino más lejano	Ward
0.2936	0.1067	0.4661	0.6009

Coeficientes de agrupación Noche de Museos.

En Noche de Museos el coeficiente de agrupación cercano a 1, corresponde al método Ward's con un valor de 0.6009.

El método Ward consiste en minimizar la varianza total del grupo por lo que agrupa un par de clústeres con una distancia mínima, es decir forma grupos de una manera que minimiza la perdida de información asociada con cada grupo. **Fuente especificada no válida.**

Anexo II

A continuación se muestra el código utilizado en el programa Rstudio, con el cual se realizó el análisis para la información obtenida de las encuestas Talleres Sabatinos y Noches de Museos.

```
## Código en Rstudio para análisis de Encuestas elaborada por el Departamento de Comunicación Educativa del MNA.
```

```
library(readxl)
```

```
library(dplyr)
```

```
library(rgl)
```

```
library(corrplot)
```

```
library(graphics)
```

```
library(scatterplot3d)
```

```
library(car)
```

```
library(mvtnorm)
```

```
library(psych)
```

```
library(bpca)
```

```
library(mnormt)
```

```
library(purrr)# Cargamos la base de datos
```

```
setwd("C:/Users/brend/OneDrive/Desktop/Descriptivos encuestas")
```

```
n <- read_excel(path = "Noche de Museos 1.xlsx", sheet = "Cualitativos1")
```

```
attach(n)
```

```
# Correlaciones
```

```
## El cálculo de las correlaciones se realiza con el código siguiente para cada una de las variables.
```

```
##### Correlaciones Tema #####
```

```
M1 <- as.data.frame(cbind(Tema,Edad))
```

```
names(M1)<-c("Tema","Edad")
```

```
poliserial1<-polyserial(M1,M1)
```

```
poliserial1
```

```
M2 <- as.data.frame(cbind(Tema,Sexo))
```

```
names(M2)<-c("Tema","Sexo")
```

```
poliserial2<-polyserial(M2,M2)
```

```
poliserial2
```

```
## Una vez calculadas las correlaciones lo siguiente es formar la matriz de correlaciones.
```

```

datos <- c(1, poliseriales1[1,2], poliseriales2[1,2], poliseriales3[1,2], poliseriales4[1,2], policorica5$rho[1,2], poliseriales6[1,2], poliseriales7[1,2], poliseriales8[1,2], poliseriales9[1,2], poliseriales10[1,2], policorica11$rho[1,2], policorica12$rho[1,2], poliseriales13[1,2], poliseriales14[1,2], 1, poliseriales15[1,2], poliseriales16[1,2], poliseriales17[1,2], poliseriales18[1,2], poliseriales19[1,2], poliseriales20[1,2], poliseriales21[1,2], poliseriales22[1,2], poliseriales23[1,2], poliseriales24[1,2], poliseriales25[1,2], poliseriales26[1,2], 1, poliseriales27[1,2], poliseriales28[1,2], poliseriales29[1,2], poliseriales30[1,2], poliseriales31[1,2], poliseriales32[1,2], poliseriales33[1,2], poliseriales34[1,2], poliseriales35[1,2], poliseriales36[1,2], poliseriales37[1,2], poliseriales38[1,2], poliseriales39[1,2], poliseriales40[1,2], poliseriales41[1,2], poliseriales42[1,2], poliseriales43[1,2], poliseriales44[1,2], poliseriales45[1,2], poliseriales46[1,2], poliseriales47[1,2], poliseriales48[1,2], poliseriales49[1,2], poliseriales50[1,2], policorica51$rho[1,2], policorica52$rho[1,2], poliseriales53[1,2], poliseriales54[1,2], poliseriales55[1,2], policorica55$rho[1,2], poliseriales56[1,2], poliseriales57[1,2], poliseriales58[1,2], poliseriales59[1,2], poliseriales60[1,2], poliseriales61[1,2], poliseriales62[1,2], poliseriales63[1,2], poliseriales64[1,2], poliseriales65[1,2], poliseriales66[1,2], poliseriales67[1,2], poliseriales68[1,2], poliseriales69[1,2], poliseriales70[1,2], poliseriales71[1,2], poliseriales72[1,2], poliseriales73[1,2], poliseriales74[1,2], poliseriales75[1,2], poliseriales76[1,2], poliseriales77[1,2], poliseriales78[1,2], poliseriales79[1,2], poliseriales80[1,2], poliseriales81[1,2], poliseriales82[1,2], poliseriales83[1,2], poliseriales84[1,2], poliseriales85[1,2], poliseriales86[1,2], poliseriales87[1,2], poliseriales88[1,2], poliseriales89[1,2], poliseriales90[1,2], poliseriales91[1,2], 1)

```

```
## Determinante de Matriz de Correlaciones ##
```

```
D <- det(MC)
```

```
#### Análisis Factorial ####
```

```
#Test de Esfericidad de Bartlett
```

```
# Busca contrastar la hipótesis nula de que la matriz de correlación es igual a una matriz de identidad
```

```
Bart <- cortest.bartlett(MC, n = nrow(n))
```

```
Bart$statistic
```

```
Bart <- bartlett.test(n, g = MC)
```

```
#Test de Bartlett no se aplica por que las variables no son normales
```

```
## Test KMO mide la correlación y la correlación parcial entre dos variables.
```

```
TKMO <- KMO(MC)
```

```
## Eigenvectores y Eigenvalores
```

```
aucor=eigen(MC)
```

```
## Porcentajes de varianza acumulada
```

```
Prop.Var=aucor$values/sum(aucor$values)*100
```

```
cumProp.var=cumsum(aucor$values/sum(aucor$values)*100)
```

```
porc=data.frame(Comp=1:14,Autovalor=round(aucor$values,3),Porc.Var=round(Prop.Var,3),Acum.Porc.Var=round(cumProp.var,3))
```

```
## Cálculo de Comunalidades Matriz de Autovalores
```

```
mautov=matrix(diag(aucor$values),ncol=14,nrow=14)
```

```
## Matriz de carga de factores
```

```
lamda=aucor$vectors%*%sqrt(mautov)
```

```
hi=lamda%*%t(lamda)
```

```
## Comunalidad
```

```
hi2=diag(hi)
```

```
hi2
```

```
## Especificidad
```

```
him=lamda[,1:2]%*%t(lamda[,1:2])
```

```

## Comunalidad
hi2m=diag(him)hi2m
## Matriz de Comunalidades
data.frame(variable=names(n[,1:14]),inicial=round(hi2,3),extraccio=round(hi2m,3))
## Varianza Total
sum(aucor$values)
## Gráfico de sedimentación
plot(1:14,aucor$values,type="l",xlab="Componetes",ylab="autovalores", main = "Gráfico de sedimentación")
plot(aucor$values, type = "b", lty=1, col = "blue", main = "Gráfica de sedimentación", xlab = "Componentes",
ylab = "Eigenvalores",
      xlim = c(1,14))
data.frame(variable=names(n[,1:14]),comp1=round(-lamda[,1],3),comp2=round(lamda[,2],3))
## Matriz de puntuaciones factoriales
B=solve(MC)%*%lamda
## Matriz de puntuaciones factoriales de componentes
data.frame(variable=names(n[,1:14]),Fac1=round(-B[,1],3),Fac2=round(B[,2],3))
## Estandarización de variables
estdn=data.frame(n [1:14])
#Estimación factorial del individuo.
Fp=data.matrix (estdn)%*%(B)
head(Fp)
## Gráfico de correlaciones y variables
library(ggplot2)
B1=data.frame(B)
ggplot(B1,aes(-B1[,1],B1[,2],label=rownames(B1)))+geom_point()+geom_text(vjust = 2)+xlab("Fact
1")+ylab("Fact 2")+geom_hline(yintercept=0,size=1)+geom_vline(xintercept=0,size=1)
## Factoriales
grap.fact=data.frame(y1=-Fp[,1],y2=Fp[,2],lab=1:203,grupo=n[,14])
ggplot(grap.fact,aes(y1,y2,label=lab,color= rainbow(203)))+geom_point()+geom_text(vjust = 2)+xlab("Fact
1")+ylab("Fact 2")+geom_hline(yintercept=0,size=1)+geom_vline(xintercept=0,size=1)

## Calculo de las componentes principales de las variables originales estandarizadas
Acp <-prcomp(MC, scale = TRUE)
summary(Acp)
Acp$rotation

```

```

tabla= data.frame(Comp=1:7,vr, pvr, porcentajev)

plot(Acp, col = "blue", main = "Distribución de varianza", xlab = "Componentes Principales", ylab =
"Eigenvectores")

## Varianza explicada para cada componente

vr <- Acp$sdev

pvr <- vr/sum(vr)

porcentajev <- cumsum(pvr)*100

plot(Acp$sdev, type = "b", col= "darkblue", xlab = "Componentes Principales", ylab = "Eigenvectores", main =
"Gráfica de sedimentación") ## Gráfica de segmentación

plot(Acp[1:88,], col = "blue", main = "Distribución de varianza", xlab = "Componentes Principales")

## Correlación entre las variables originales y las componentes principales

oc <- Acp$rotation%*%diag(Acp$sdev)

## Calculo de las componentes principales como combinación lineal de las variables estandarizadas

Comp <- predict(Acp)

## Calculo de la componente principal

pc_1 <- apply(Acp$rotation[,1]*base1[1:88,], 1, sum) ##Obtenemos los valores de Zi
pc_2 <- apply(Acp$rotation[,2]*base1[1:88,], 1, sum) ##Obtenemos los valores de Zi

## Gráficas

plot(pc_1, pc_2,main="Componentes Principales y observaciones",col=rainbow(88),
ylab="Componente Pricipal 2",xlab="Componente Principal 1",
xlim=c(-4,4), ylim=c(-5,4))

text(pc_1, pc_2, labels=1:88,col=rainbow(88))

abline(h=0,lty=0)
abline(v=0,lty=0)

#### Análisis de Conglomerados ####

# Matriz de disimilitud euclidean, "maximum", "manhattan", "canberra", "binary" or "minkowski"

d<- dist(MC, method = "euclidean",diag = FALSE, upper = FALSE, p = 2)

# Hierarchical clustering using Complete Linkage

hc1 <- hclust(d, method = "complete" )

hc1$ac

# Plot para obtener el dendrograma

plot(hc1, cex = 0.6, hang = -1, main = "Dendrograma VariablesNoche de Museos")

rect.hclust(hc1, k = 5, border = 2:10)i

```

ⁱ Código análisis factorial y conglomerados se obtuvo de [RPubs - Análisis de Cluster en R](#)